PTO/AIA/15 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	01-007706US
First Named Inventor	Stephen Turner
Title	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
Express Mail Label No.	N/A - filed via EFS Web

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

1. [✓] **Fee Transmittal Form** (PTO/SB/17 or equivalent)
2. [ ] **Applicant asserts small entity status.** See 37 CFR 1.27
3. [ ] **Applicant certifies micro entity status.** See 37 CFR 1.29. Applicant must attach form PTO/SB/15A or B or equivalent.
4. [✓] **Specification** [*Total Pages* 106] Both the claims and abstract must start on a new page. *(See MPEP § 608.01(a) for information on the preferred arrangement)*
5. [✓] **Drawing(s)** (35 U.S.C. 113) [*Total Sheets* 24]
6. [ ] **Inventor's Oath or Declaration** [*Total Pages* ____] *(including substitute statements under 37 CFR 1.64 and assignments serving as an oath or declaration under 37 CFR 1.63(e))*
   a. [ ] Newly executed (original or copy)
   b. [✓] A copy from a prior application (37 CFR 1.63(d))
7. [✓] **Application Data Sheet** ** See note below.* See 37 CFR 1.76 (PTO/AIA/14 or equivalent)
8. **CD-ROM or CD-R** in duplicate, large table, or Computer Program (*Appendix*)
   [ ] Landscape Table on CD
9. **Nucleotide and/or Amino Acid Sequence Submission** *(if applicable, items a. – c. are required)*
   a. [ ] Computer Readable Form (CRF)
   b. [ ] Specification Sequence Listing on:
      i. [ ] CD-ROM or CD-R (2 copies); or
      ii. [ ] Paper
   c. [ ] Statements verifying identity of above copies

ADDRESS TO:

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

ACCOMPANYING APPLICATION PAPERS

10. [✓] **Assignment Papers** (cover sheet & document(s))
    Name of Assignee: Pacific Biosciences of California, Inc.
11. [✓] **37 CFR 3.73(c) Statement** [✓] **Power of Attorney** *(when there is an assignee)*
12. [ ] **English Translation Document** *(if applicable)*
13. [✓] **Information Disclosure Statement** (PTO/SB/08 or PTO-1449)
    [ ] Copies of citations attached
14. [ ] **Preliminary Amendment**
15. [ ] **Return Receipt Postcard** *(MPEP § 503) (Should be specifically itemized)*
16. [ ] **Certified Copy of Priority Document(s)** *(if foreign priority is claimed)*
17. [ ] **Nonpublication Request** Under 35 U.S.C. 122(b)(2)(B)(i). Applicant must attach form PTO/SB/35 or equivalent.
18. [✓] **Other:** Certification and Request for Prioritized Examination

***Note:** (1) Benefit claims under 37 CFR 1.78 and foreign priority claims under 1.55 **must** be included in an Application Data Sheet (ADS).
(2) For applications filed under 35 U.S.C. 111, the application must contain an ADS specifying the applicant if the applicant is an assignee, person to whom the inventor is under an obligation to assign, or person who otherwise shows sufficient proprietary interest in the matter. See 37 CFR 1.46(b).

19. CORRESPONDENCE ADDRESS

[✓] The address associated with Customer Number: 57770 **OR** [ ] Correspondence address below

Name					
Address					
City		State		Zip Code	
Country		Telephone		Email	

Signature	/David C. Scherer, Ph.D./	Date	December 19, 2016
Name (Print/Type)	David C. Scherer, Ph.D.	Registration No. (Attorney/Agent)	56,993

PTO/SB/17 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

FEE TRANSMITTAL

Complete if known	
Application Number	Not Yet Assigned
Filing Date	December 19, 2016
First Named Inventor	Stephen Turner
Examiner Name	Not Yet Assigned
Art Unit	Not Yet Assigned
Practitioner Docket No.	01-007706US

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($) 6740.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☑ Deposit Account Deposit Account Number: 57770 Deposit Account Name: Pacific Biosciences

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☑ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☑ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17 ☑ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**U ($)**	**S ($)**	**M ($)**	**Fees Paid ($)**
Utility	280	140*	70	600	300	150	720	360	180	1600.00
Design	180	90	45	120	60	30	460	230	115	
Plant	180	90	45	380	190	95	580	290	145	
Reissue	280	140	70	600	300	150	2,160	1,080	540	
Provisional	260	130	65	0	0	0	0	0	0	

* The $140 small entity status filing fee for a utility application is further reduced to $70 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	**Undiscounted Fee ($)**	**Small Entity Fee ($)**	**Micro Entity Fee ($)**
Each claim over 20 (including Reissues)	80	40	20
Each independent claim over 3 (including Reissues)	420	210	105
Multiple dependent claims	780	390	195

Total Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
18	-20 or HP =	0	x 80	=	0.00

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims		**Extra Claims**	**Fee ($)**		**Fee Paid ($)**
1	-3 or HP =	0	x 420	=	0.00

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	**Fee Paid ($)**
780	0.00

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		**Extra Sheets**		**Number of each additional 50 or fraction thereof**		**Fee ($)**		**Fee Paid ($)**
98	- 100 =	0	/ 50 =	0	(round up to a whole number) x	400.00	=	0.00

4. OTHER FEE(S) **Fees Paid ($)**

Non-English specification, $130 fee (no small or micro entity discount) _____

Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity) _____

Other (e.g., late filing surcharge): Track 1 Prioritization Fee and Processing Fee 4140.00

SUBMITTED BY

Signature	/David C. Scherer, Ph.D./	Registration No. (Attorney/Agent) 56,993	Telephone 650-521-8127
Name (Print/Type)	David C. Scherer, Ph.D.		Date December 19, 2016

APPLICATION DATA SHEET

APPLICATION INFORMATION

Application Type::	Regular
Title::	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
Attorney Docket Number::	01-007706US
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets ::	24
Small Entity?::	No
Petition included?::	No
Licensed US Govt. Agency::	No
Contract or Grant Numbers::	Not Applicable
Secrecy Order in Parent Application?::	No

INVENTOR INFORMATION

Inventor 1

Primary Citizenship Country::	United States
Given Name::	Stephen
Middle Name::	
Family Name::	Turner
City of Residence::	Seattle
State or Province of Residence::	Washington
Country of Residence::	United States
Street of mailing address::	4216 NE 113th Street
City of mailing address::	Seattle
State or Province of mailing address::	Washington
Postal or Zip Code of mailing address::	98125

Inventor 2

Primary Citizenship Country::	United States
Given Name::	Jon
Middle Name::	
Family Name::	Sorenson
City of Residence::	Alameda
State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	1725 Nason Street
City of mailing address::	Alameda
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94501

Inventor 3

Primary Citizenship Country::	United States
Given Name::	Kenneth
Middle Name::	Mark
Family Name::	Maxham
City of Residence::	Redwood City
State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	241 Harrison Avenue
City of mailing address::	Redwood City
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94062

Inventor 4

Primary Citizenship Country::	United States
Given Name::	John
Middle Name::	
Family Name::	Eid
City of Residence::	San Francisco
State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	52 Sheridan Street, #2
City of mailing address::	San Francisco
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94103

DOMESTIC PRIORITY INFORMATION

Application::	Continuity Type::	Parent Application::	Parent Filing Date
This Application	Continuation of	14/708,603	05/11/15
14/708,603	Continuation of	14/091,961	11/27/13
14/091,961	Continuation of	12/982,029	12/30/10
12/982,029	An application claiming the benefit under 35 USC 119(e)	61/099,696	09/24/08
12/982,029	An application claiming the benefit under 35 USC 119(e)	61/139,402	12/19/08
12/982,029	Continuation-in-part of	12/413,226	03/27/09

Initial 12/19/16

FOREIGN PRIORITY INFORMATION

Country::	Application number::	Filing Date::	Priority Claimed::

APPLICANT INFORMATION

Applicant is assignee?::	Yes
Applicant Name::	Pacific Biosciences of California, Inc.
Street of Mailing Address::	1380 Willow Road
City of mailing address::	Menlo Park
State or Province of mailing address::	California
Postal or Zip code of mailing address::	94025

CORRESPONDENCE INFORMATION

Correspondence Address::	Pacific Biosciences of California, Inc. 1380 Willow Road Menlo Park, CA 94025
Phone number::	(650) 521-8127
Fax number::	(650) 323-9420
E-Mail address::	dscherer@pacb.com
Customer Number::	57770

REPRESENTATIVE INFORMATION

Representative Name::	David C. Scherer, Ph.D.
Registration Number::	56,993

Signature:: /David C. Scherer, Ph.D./
David C. Scherer, Ph.D.

Initial 12/19/16

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation application of U.S. Patent Application No. 14/708,603, filed May 11, 2015, which is a continuation application of U.S. Patent Application No. 14/091,961, filed November 27, 2013, now U.S. Patent No. 9,057,102, which is a continuation application of U.S. Patent Application No. 12/982,029, filed December 30, 2010, now U.S. Patent No. 8,628,940, which (1) claims the benefit of U.S. Provisional Application No. 61/099,696, filed September 24, 2008; (2) claims the benefit of U.S. Provisional Application No. 61/139,402, filed December 19, 2008; and (3) is a continuation-in-part application of U.S. Patent Application No. 12/413,226, filed March 27, 2009, now U.S. Patent No. 8,143,030, the full disclosures of all of which are incorporated herein by reference in their entireties for all purposes.

[0002] This application is also related to U.S. Provisional Application No. 61/072,160, filed March 28, 2008, U.S. Patent Application No. 12/383,855, filed March 27, 2009, now U.S. Patent No. 8,236,499, and U.S. Patent Application No. 12/413,258, filed March 27, 2009, now U.S. Patent No. 8,153,375, all of which are incorporated herein by reference in their entireties for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0003] Not Applicable.

BACKGROUND OF THE INVENTION

[0004] The use of optically detectable labeling groups, and particularly those groups having high quantum yields, e.g., fluorescent or chemiluminescent groups, is ubiquitous throughout the fields of analytical chemistry, biochemistry, and biology. In particular, by providing a highly visible signal associated with a given reaction, one can better monitor that reaction as well as any potential effectors of that reaction. Such analyses are the basic tools of life science research in genomics, diagnostics, pharmaceutical research, and related fields.

[0005] Such analyses have generally been performed under conditions where the amounts of reactants are present far in excess of what is required for the reaction in question. The result of this excess is to provide ample detectability, as well as to compensate for any damage caused by the detection system and allow for signal detection with minimal impact on the reactants. For example, analyses based on fluorescent labeling groups generally require the use of an excitation radiation

source directed at the reaction mixture to excite the fluorescent labeling group, which is then separately detectable. However, one drawback to the use of optically detectable labeling groups is that prolonged exposure of chemical and biochemical reactants to such light sources, alone, or when in the presence of other components, e.g., the fluorescent groups, can damage such reactants. The traditional solution to this drawback is to have the reactants present so far in excess that the number of undamaged reactant molecules far outnumbers the damaged reactant molecules, thus minimizing or negating the effects of the photo-induced damage.

[0006] A variety of analytical techniques currently being explored deviate from the traditional techniques. In particular, many reactions are based on increasingly smaller amounts of reagents, e.g., in microfluidic or nanofluidic reaction vessels or channels, or in "single molecule" analyses. Such low reactant volumes are increasingly important in many high throughput applications, such as microarrays. The use of smaller reactant volumes offers challenges to the use of optical detection systems. When smaller reactant volumes are used, damage to reactants, such as from exposure to light sources for fluorescent detection, can become problematic and have a dramatic impact on the operation of a given analysis. In other cases, other reaction conditions may impact the processivity, rate, fidelity, or duration of the reaction, including salt or buffer conditions, pH, temperature, or even immobilization of reaction components within observable reaction regions. In many cases, the effects of these different reaction or environmental conditions can degrade the performance of the system over time. This can be particularly detrimental, for example, in real-time analysis of reactions that include fluorescent reagents that can expose multiple different reactions components to optical energy. In addition, smaller reactant volumes can lead to limitations in the amount of signal generated upon application of optical energy.

[0007] Further, in the case of sequencing-by-synthesis applications, an additional challenge has been to develop ways to effectively sequence noncontiguous portions of a template nucleic acid on a single molecule. This challenge is exacerbated in template nucleic acids that contain highly repetitive sequence and/or are hundreds or thousands of nucleotides in length, such as certain genomic DNA fragments. The difficulty in generating such noncontiguous reads from a single template has hampered efforts to construct consensus sequences for long templates, for example, in genome sequencing projects.

[0008] As such, methods and systems that result in enhanced reaction performance, such as an increase in processivity, rate, fidelity, or duration of a reaction of interest, would provide useful improvements to the methods and compositions currently available. For example, methods, devices,

and systems that increase reaction performance by, e.g., mitigating to some extent photo-induced damage in a reaction of interest and/or increasing various other performance metrics for the reaction would be particularly useful.

BRIEF SUMMARY OF THE INVENTION

[0009] In a general sense, the methods provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0010] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one

reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion. The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0011] In certain aspects, the present invention is generally directed to methods, devices, and systems for enhancing the performance of illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. In certain preferred embodiments, illuminated reactions comprise one or more fluorescent or fluorogenic reactants. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In some aspects, enhancing the performance of an illuminated reaction means increasing the processivity, rate, fidelity, and/or duration of the reaction. For example, enhancing the performance of an illuminated reaction can involve reducing or limiting the effects of photo-induced damage during the reaction. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction.

[0012] In certain aspects, methods of the invention useful for characterizing an analytical reaction comprise preparing a reaction mixture and initiating the analytical reaction therein, subjecting the reaction mixture to at least one detection period and at least one non-detection period during the course of the analytical reaction, collecting data during both the detection period(s) and the non-detection period(s), and combining the collected data to characterize the analytical reaction. In certain embodiments, the analytical reaction comprises an enzyme that exhibits an improvement in performance as compared to its performance in the analytical reaction under constant illumination, and such improvement may be related to various aspects of enzyme activity, e.g., processivity, fidelity, rate, duration of the analytical reaction, and the like. In certain embodiments, stop or pause points are used to control the activity of the enzyme, and such stop or pause points may comprise elements such as large photolabile groups, strand-binding moieties, non-native bases, and others well known in the art. In certain preferred embodiments, the one or more detection periods are illuminated periods and the one or more non-detection periods are non-illuminated periods. In certain preferred embodiments, a plurality of analytical reactions disposed on a solid support are characterized, preferably in a coordinated fashion as described elsewhere herein.

[0013] In certain preferred embodiments, the analytical reaction is a sequencing reaction that generates sequence reads from a single nucleic acid template during the detection period(s) but not during the non-detection period(s). For example, the analytical reaction can comprise at least two or more detection periods and can generate a plurality of noncontiguous reads from the single nucleic acid template. In some embodiments, the single nucleic acid template is at least 100 bases in length and/or comprises multiple repeat sequences. In certain embodiments, the sequencing reaction comprises passage of the single nucleic acid template through a nanopore, and in other embodiments the sequencing reaction comprises primer extension by a polymerase enzyme.

[0014] The analytical may optionally be a processive reaction monitored in real time, i.e., during the course of the processive reaction. In preferred embodiments, such a processive reaction is carried out by a processive enzyme that can repetitively execute its catalytic function, thereby completing multiple sequential steps of the reaction. For example, a processive polymerization reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.

[0015] In certain preferred embodiments, the analytical reaction comprises at least one component comprising a detectable label, e.g., a fluorescently labeled nucleotide. In certain embodiments, the labeled component is present throughout the course of the analytical reaction, i.e., during both the detection and the non-detection periods. The method may further comprise an optical system to collect the data during the detection period, but optionally not to collect the data during the non-detection period.

[0016] In certain aspects, methods of the invention comprise providing a substrate having a reaction mixture disposed thereon and illuminating the reaction mixture on the substrate with an excitation illumination for multiple, noncontiguous periods during the course of the reaction, thereby subjecting the reaction mixture to intermittent excitation illumination. In some embodiments, the reaction mixture comprises first reactant and a second reactant, wherein an amount of photo-induced damage to the first reactant occurs as a result of interaction between the first reactant and the second reactant under excitation illumination. In certain embodiments, the method further comprises monitoring a reaction between the first and second reactants during illumination and collecting the data generated therefrom. In some embodiments, the reaction is a primer extension reaction and/or the first reactant is a polymerase enzyme. In certain embodiments, the second reactant is a fluorogenic or fluorescent molecule.

[0017] In yet another aspect, the methods are useful for mitigating photo-induced damage in an illuminated reaction by subjecting the illuminated reaction to intermittent illumination rather than constant illumination. For example, certain methods of the invention monitor a reaction mixture comprising at least one enzyme and a fluorescent or fluorogenic substrate for the enzyme, wherein interaction of the enzyme and the substrate under excitation illumination can result in altered activity of the enzyme, e.g if such excitation illumination is present over an extended period of time. Such methods can comprise directing intermittent excitation illumination at a first observation region for a first period that is less than a photo-induced damage threshold period under the intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions. As such, certain aspects of the invention lengthen a photo-induced damage threshold period for an analytical reaction through intermittent inactivation of the excitation illumination source since the photo-induced damage threshold period under intermittent illumination is longer than the photo-induced damage threshold period under constant illumination.

[0018] In a related aspect, the invention also provides methods of performing an enzyme reaction, comprising providing an enzyme within a first observation region, contacting the enzyme with a fluorescent or fluorogenic substrate for the enzyme, and directing an excitation radiation at and detecting signals from the first observation region for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0019] In further aspects, the invention provides methods of monitoring a primer extension reaction, comprising providing a polymerase enzyme within a first observation region, contacting the polymerase with at least a first fluorescent or fluorogenic nucleotide analog, and monitoring a fluorescent signal emitted from the first observation region in response to illumination with excitation radiation for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0020] In addition, the invention provides methods for generating a plurality of noncontiguous sequence reads from a single nucleic acid template molecule. Such methods generally comprise preparing a reaction mixture comprising the template molecule, a polymerase enzyme, and a set of differentially labeled nucleotides or nucleotide analogs, wherein the set comprises at least one type of nucleotide or nucleotide analog for each of the natural nucleobases (A, T, C, and G). The polymerization reaction is initiated, the polymerase begins processive incorporation of the labeled nucleotides or nucleotide analogs into a nascent nucleic acid strand, and during such incorporation the reaction is monitored by optical means to detect incorporation events, thereby generating a first sequence read. In a subsequent step, the labeled nucleotides or analogs are replaced with unlabeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed without detecting incorporation events. Subsequently, the unlabeled nucleotides or analogs are replaced with labeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed once again with real time detection of incorporation events, thereby generating a second sequence read that is noncontiguous to the first sequence read. The substitution of labeled for unlabeled, and unlabeled for labeled, nucleotides and nucleotide analogs can be repeated multiple times to generate a plurality of noncontiguous sequence reads, each of the plurality generated during a period when the labeled nucleotides or nucleotide analogs are being incorporated into the nascent strand and such incorporation is being detected in real time.

[0021] In certain aspects, devices of the invention can comprise a solid support (e.g., substrate) having an observation region, a first reactant immobilized within the observation region, and a second reactant disposed within the observation region, and a means for subjecting the observation region to at least one illuminated period and at least one non-illuminated period. In certain embodiments, interaction between the first and second reactants under excitation illumination causes photo-induced damage to the first reactant, and further wherein the photo-induced damage is reduced by subjecting the observation region to intermittent illumination. In some embodiments, the first reactant is an enzyme (e.g., a polymerase), the second reactant (e.g., a nucleotide) has a detectable label (e.g., fluorescent label), and/or the observation region is within a zero-mode waveguide. The means for subjecting the observation region to one or more illuminated and non-illuminated periods may comprise, e.g., a laser, laser diode, light-emitting diode, ultra-violet light bulb, white light source, a mask, a diffraction grating, an arrayed waveguide grating, an optic fiber, an optical switch, a mirror, a lens, a collimator, an optical attenuator, a filter, a prism, a planar waveguide, a wave-plate, a delay line, a movable support coupled with the substrate, and a movable illumination source, and the like. The device may further comprise a means for collecting the data during the illuminated period(s), such as an optical train, e.g., operably coupled to a machine comprising machine-readable medium onto which such data may be written and stored.

[0022] In further aspects, the invention provides systems for performing intermittent detection of an analytical reaction comprising reagents for the analytical reaction disposed on a solid support, a mounting stage configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the solid support detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other, and a data processing system operably coupled to the optical train. In certain preferred embodiments, the analytical reaction is a sequencing reaction and/or the solid support comprises at least one zero-mode waveguide.

[0023] In still other aspects, the invention provides systems for analyzing an illuminated reaction that is susceptible to photo-induced damage when illuminated for a period longer than an photo-induced damage threshold period, comprising a solid support having reagents for the reaction disposed thereon, a mounting stage supporting the solid support and configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the

solid support to illuminate the portion of the solid support and detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, and a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other. In some embodiments, the illuminated reaction is a sequencing reaction, e.g., a nucleotide sequencing-by-synthesis reaction. In certain embodiments, the solid support comprises at least one optical confinement, e.g., a zero-mode waveguide.

[0024] The invention provides methods of performing analytical reactions, e.g., processive analytical reactions, that include preparing a reaction mixture comprising reaction components, at least one of which is a detectable component that is detectable during one or more detection periods, and at least one of which is a clocking component that is detectable during one or more non-detection periods during the analytical reaction. The methods further comprise initiation the analytical reaction and maintaining conditions that allow the analytical reaction to proceed while subjecting it to at least one detection period and at least one non-detection period, both in the presence of the clocking component and the detectable component. In certain embodiments, the detectable component emits a detectable signal is response to excitation illumination during the detection period, but not during the non-detection period when a clocking signal is emitted from the clocking component. The detectable signal is collected during the detection period and the clocking signal is detected during the non-detection period, e.g., using an optical system. Optionally, the clocking signal can also be collected during the detection period and the non-detection period. In certain preferred embodiments, detection data is collected in read time during the detection period, non-detection data is collected in real time during the non-detection period, and the detection data and non-detection data are both used to characterize the analytical reaction. In some embodiments, the transition between the detection period and the non-detection period does not involve substitution and/or addition of reaction components during progression of the analytical reaction, and in other embodiments the transition does involve substitution and/or addition of reaction components, e.g., via a reaction mixture exchange. In some preferred embodiments, a plurality of analytical reactions are disposed on a solid support, subjected to intermittent illumination, monitored to collect data, and characterized based upon the data so collected.

[0025] The detectable component and clocking component are typically linked to discrete molecules in the analytical reaction. For example, the detectable component can be linked to a first subset of nucleotide analogs and the clocking component can be linked to a second subset of

nucleotide analogs in the analytical reaction mixture. Alternatively, both the detectable component and the clocking component can be linked to a single molecule, e.g., a single nucleotide or nucleotide analog, in the analytical reaction. The detectable component and clocking component can both comprise detectable labels (e.g., luminescent, fluorescent, or fluorogenic labels, including, e.g., quantum dots), and in some embodiments, different detectable labels, e.g. having different absorption peaks.

[0026] In certain preferred embodiments, an analytical reaction performed according to the invention comprises at least one enzyme, e.g., a polymerase, ligase, ribosome, nuclease, and/or kinase. In some embodiments, pause or stop points are engineered into the analytical reaction to control activity of the enzyme. Various aspects of the analytical reaction can be changed by being subjected to at least one detection period and at least one non-detection period, such aspects including but not limited to processivity, fidelity, rate, and duration, e.g. of enzyme activity.

[0027] In certain preferred embodiments, the analytical reaction is a sequencing reaction comprising a single nucleic acid template that generates sequence reads during the detection period by detecting the detectable component, and does not generate sequence reads during the non-detection period by suspending detection of the detectable component. Such a sequencing reaction typically comprises at least two or three detection periods and generates a plurality of noncontiguous sequence reads from the single nucleic acid template. In some embodiments, the template comprises multiple repeat or complementary sequences. In some embodiments, the sequencing reaction comprises passage of the single nucleic acid or a nascent strand complementary thereto through a nanopore. In some preferred embodiments, the sequencing reaction comprises primer extension by a polymerase enzyme and the detectable component is linked to a nucleotide or nucleotide analog. In some embodiments, the clocking component is linked to the polymerase enzyme, and optionally can be a multi-component label, e.g, a FRET label.

[0028] In certain aspects, the invention provides methods of mitigating photo-induced damage during an illuminated reaction that include preparing a reaction mixture having first and second reactants, where interaction of the reactants under excitation illumination can cause photo-induced damage to the first reactant. The illuminated reaction is subjected to intermittent excitation illumination characterized by periods of maximal illumination followed by periods of modified but not absent illumination. The intermittent excitation illumination reduces the amount of photo-induced damage to the first reactant during the illuminated reaction as compared to the illuminated reaction under constant maximal excitation illumination, thereby mitigating photo-induced damage

to the first reactant. In certain preferred embodiments, the illuminated reaction is a primer extension reaction. In certain preferred embodiments, the first reactant is an enzyme, e.g., a polymerase or ligase enzyme. In certain preferred embodiments, the second reactant comprises a fluorescent or fluorogenic molecule. In certain embodiments, the modified excitation illumination is illumination with a lower intensity excitation illumination than the maximal excitation illumination. In certain embodiments, a set of illumination sources provides the maximal excitation illumination and a subset of the set of illumination sources provides the modified excitation illumination.

[0029] In other aspects, the invention provides a method of sequencing a template nucleic acid that includes subjecting the template to methylation to generate at least one methylated base, subjecting the methylated base to base excision to generate at least one abasic site in the template, annealing a primer to the template nucleic acid, contacting the template with a polymerase enzyme to promote extension of the primer in a template-dependent manner, monitoring the extension of the primer in real time to generate a nucleotide sequence read complementary to the template, extending the primer until the abasic site is encountered by the polymerase, at which time the polymerase pauses on the template, and reinitiating primer extension by facilitating abasic site bypass by the polymerase. The monitoring, extending, and reinitiating steps are repeated until a desired number of nucleotide sequence reads is generated and collected, and subsequently analyzed to determine the sequence of the template nucleic acid. In certain embodiments, the contacting step occurs during a detection period or a detection period immediately follows the contacting step. In certain embodiments, a detection period ends and a non-detection period begins prior to one or more pauses of the polymerase on the template. In certain embodiments, a non-detection period is terminated simultaneous with or immediately following one or more reinitiation steps. In some embodiments, the reinitiating step comprises introduction of a pyrene to the polymerase, where the polymerase incorporates the pyrene into the nascent strand opposite and, therefore, "pairing with" an abasic site in the template. In certain preferred embodiments, the template is circular and the polymerase pauses at the same abasic site multiple times during the primer extension reaction. In other embodiments, the method further comprises terminating the monitoring when a desired length of the nucleotide sequence read is collected, e.g., by removing or modifying excitation illumination. Optionally, the desired length can be less than a length of the template nucleic acid. Additionally, the monitoring can be reinitiated subsequent to or simultaneous with the reinitiating of primer extension.

[0030] In yet further aspects, the invention provides a method of performing an illuminated reaction that includes preparing a reaction mixture comprising multiple optically detectable components that are distinguishable from one another based upon their individual signal emissions, initiating the illuminated reaction, and maintaining conditions that allow the illuminated reaction to proceed while subjecting the reaction mixture to at least one maximal illuminated period and at least one modified illuminated period during the illuminated reaction. In preferred embodiments, at least a portion of the optically detectable components are detectable during both the maximal and modified illuminated periods. In certain embodiments, the maximal illuminated period is characterized by a first excitation radiation intensity and the modified illuminated period is characterized by a second excitation radiation intensity that is less than the first excitation radiation intensity. In certain preferred embodiments, all of the optically detectable components are detectable during both the maximal and modified illuminated periods, but are distinguishable from one another during the maximal illuminated period, but are not distinguishable during the modified illuminated period. In certain embodiments, the maximal illuminated period comprises exposing the reaction mixture to a set of excitation radiation wavelengths and the modified illuminated period comprises exposing the reaction mixture to a subset of the set of excitation radiation wavelengths. In certain preferred embodiments, all of the optically detectable components are detectable and distinguishable during the maximal illuminated period, but only a subset of the optically detectable components are detectable during the modified illuminated period.

[0031] In some embodiments, the illuminated reaction is initiated during a modified illuminated period and subsequently subjected to a maximal illuminated period, where data collected during the modified illuminated period is used in the statistical analysis of data collected during the maximal illuminated period. For example, an illuminated reaction that is a polynucleotide sequencing reaction can generate sequence read data during a modified illuminated period that is subsequently used to construct a sequence scaffold for assembly of sequence read data collected during a maximal illuminated period. Additionally or optionally, the illuminated reaction is a template-directed sequencing reaction and sequence read data collected during a modified illuminated period is used to determine a rate of translocation of a polymerase during the modified illuminated period.

[0032] Some embodiments of the invention comprise performing a plurality of illuminated reactions, each of which is exposed to the set of excitation radiation wavelengths during the maximal illuminated period, but is exposed to a different subset of the set of excitation radiation

wavelengths during the modified illuminated period, such that a distinct subset of optically detectable components are detectable during the modified illuminated period for each of the plurality of illuminated reactions. In other words, for two such illuminated reactions, although all optically detectable components are detectable during their respective maximal illuminated periods, only a subset of the optically detectable components is detectable in each reaction, and the subset detectable in the first reaction is preferably different from the subset detectable in the second reaction.

[0033] In certain aspects, the invention provides methods for performing paired-end sequencing on a single template molecule. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion. A first linker ligated to the first terminal portion of the nucleic acid molecule connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and a second linker ligated to the second terminal portion of the nucleic acid molecule connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion. A template nucleic acid molecule is thereby formed comprising the double-stranded nucleic acid molecule with both the first linker and the second linker ligated thereto. The template molecule is subjected to a sequencing process in which sequence reads are generated for the first terminal portion and the second terminal portion, but sequence reads are not generated for the intermediate portion, even if the intermediate portion is processed during the sequencing process, e.g., by a polymerase. In some embodiments, the first linker and second linker are identical, and in other embodiments they are different from one another, i.e., not identical. In certain embodiments, the first and second linkers comprise complementary regions and can be hybridized to one another prior to one or both of the ligating steps. In some cases, hybridized linkers that are ligated to the ends of a double-stranded nucleic acid molecule are separated prior to subjecting the molecule to a sequencing reaction, and in some cases the hybridized linkers remain hybridized during at least a portion of the sequencing reaction. For example, in a template-directed sequencing reaction, a polymerase capable of strand displacement separates the hybridized linkers as it sequences the template. In certain preferred embodiments, the sequencing process comprises at least one detection period (e.g., an illuminated period) and at least one non-detection period (e.g., a non-illuminated period) such that the intermediate portion of the template molecule is subjected to the sequencing process during the non-detection period. In some embodiments, the template is fragmented after ligation to remove the intermediate portion. The

sequencing process can generate redundant sequence data from one or both of the first terminal portion and the second terminal portion, and/or can generate sequence data from an additional portion of the template molecule that is noncontiguous with the first terminal portion and the second terminal portion. In preferred embodiments, the sequencing process involves circularizing the template molecule by separating the complementary strands of the template molecule and using the complementary strands in template-directed nascent strand synthesis catalyzed by a single polymerase enzyme. Optionally, the template molecule can comprise a primer binding site, a registration sequence, and/or a synthesis blocking moiety. The primer binding site, a registration sequence, or synthesis blocking moiety can be present in one or both of the linkers, or can be located elsewhere within the template molecule. In some cases, the synthesis blocking moiety is selected from the group consisting of an abasic site, a nick, a synthetic linker, a non-native nucleotide or analog thereof, a primer, a large photolabile group, a strand-binding moiety, a damaged base, and a modified base. The synthesis blocking moiety can permanently or temporarily block progression of the sequencing process, e.g., by interfering with the activity of an enzyme, e.g., a polymerase enzyme. In certain preferred embodiments, the synthesis blocking moiety is an abasic site, e.g., introduced by a DNA glycosylase.

[0034] In some aspects, the invention provides methods for generating a nucleic acid construct for analytical reactions. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion; providing a first stem-loop linker hybridized to a second stem-loop linker; ligating the first stem-loop linker to the first terminal portion of the nucleic acid molecule, wherein the first stem-loop linker connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and ligating the second stem-loop linker to the second terminal portion of the nucleic acid molecule, wherein the second stem-loop linker connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion, thereby generating the nucleic acid construct. Optionally, the nucleic acid construct can be subjected to fragmentation after the ligating of steps c and d, wherein the fragmentation removes the intermediate portion from the nucleic acid construct and introduces two double-stranded termini. The method can further include ligating the two double-stranded termini to one another. In some embodiments, one of the stem-loop linkers comprises a primer binding site, registration sequence, or a synthesis blocking moiety that is absent from the other stem-loop linker.

[0035] In further aspects, the invention includes a single template nucleic acid molecule comprising a duplex region; a first linker linking termini at a first end of the duplex region; a second linker linking termini at a second end of the duplex region, wherein a region of the first linker is complementary to a region of the second linkers. Optionally, the single template molecule comprises the first and second linkers hybridized with one another. In some embodiments, the duplex region is separated or melted apart to transform the single template nucleic acid molecule into a topologically single-stranded, circular nucleic acid molecule. Further, the invention provides a composition comprising a single, optically resolvable polymerase enzyme in association with a single-stranded circular nucleic acid molecule, wherein the single-stranded circular nucleic acid molecule comprises first, second, third, and fourth regions, and further wherein the first region is complementary to the second region, and the third region is complementary to the fourth region, and further wherein the regions are ordered on the single-stranded circular nucleic acid molecule as follows: first region, third region, second region, fourth region.

[0036] In still further aspects of the invention, machine-implemented methods for transforming nucleotide sequence read data into consensus sequence data, wherein the nucleotide sequence read data is generated by sequencing a target region of a template nucleic acid multiple times, and the consensus sequence data is representative of a most likely actual sequence of the template nucleic acid. Such machine-implemented methods can comprise various steps, such as a) mapping the nucleotide sequence data to a target sequence using a local alignment method that produces a set of local alignments comprising an optimal local alignment and sub-optimal local alignments, b) enumerating the set of local alignments, c) constructing a weighted directed graph wherein each local alignment in the set of local alignments is represented as a node, thereby generating a set of nodes in the weighted directed graph, d) drawing edges between pairs of nodes in the weighted directed graph if the pair represents a potential reconstruction of the template nucleic acid, e) assigning weights to the edges drawn in step d, wherein a given weight for a given edge represents the log-likelihood that a given pair of nodes connected by the given edge is truly a reconstruction of the template nucleic acid, f) finding the shortest path to each node in the weighted directed graph, thereby generating a set of shortest paths for the weighted directed graph, g) ranking the set of shortest paths to determine the best assignment, and h) storing the results of steps a-g on a machine-readable medium. In certain embodiments, the steps of the machine implemented methods are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Also provided are

computer program products comprising a computer usable medium having computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the machine-implemented methods of the invention, and machine-readable medium on which the results of the method steps are stored. The invention further includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods.

[0037] In certain aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme (e.g., a polymerase) during a time t, where time t occurs during a non-detection period. Such a method comprises, in certain embodiments, developing a probability model *p(v)* to describe an observed distribution of enzyme velocities during one or more detection periods; sampling velocities from *p(v)*; summing and recording the velocities sampled in step b to produce a sum that is an estimate of x/τcorr; and repeating the sampling, summing, and recording M times to generate a distribution of sums that are estimates of x/τcorr, with the distribution of sums being the distribution of the distance x travelled by an enzyme during a time t. Preferably, at least some of the steps are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Optionally, the enzyme is a polymerase enzyme. In some embodiments, multiple enzymes are observed simultaneously and the probability model *p(v)* is determined independently for each of the multiple enzymes. In certain preferred embodiments,

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}.$$

[0038] In further aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme during a time t, where time t occurs during a non-detection period. In some embodiments, the method comprises estimating a distribution of local rates p(v), making independent identically distributed draws of N = t/τcorr velocities from from p(v); summing the velocities; recording the velocities summed in c) as an estimate of x/τcorr; and repeating b-d M times, e.g., where M is preferably at least 1000. Optionally, p(v) is determined using a Hidden Markov Model or the autocorrelation function $\langle \delta v(t)\delta v(t+\Delta)\rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$. The invention further

includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods, as well as a machine-readable medium on which the results of the steps of the methods are stored.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions, whether illumination is initiated before (A) or after (B) initiation of the reaction.

[0040] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination, including depictions of multiple reactions arrayed on a solid support (A) and prophetic data (B) from certain embodiments of the invention.

[0041] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions on a solid support (A) using intermittent illumination and a mask (B). A graph (C) depicts prophetic data from certain embodiments of the invention.

[0042] Figure 4 provides additional embodiments of masks for use in the methods of the invention, including a mask that allows illumination of columns of reactions (A) and a mask that allows illumination of every other reaction in a row and column (B).

[0043] Figure 5 illustrates an aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Figure 5A illustrates a solid support comprising four quadrants, each quadrant containing a different sample. Figure 5B illustrates a mask design for selective illumination of the substrate. Figures 5C and 5D demonstrate various positions of the mask on the solid support.

[0044] Figure 6 provides an illustration of paths in a sequence alignment matrix representing sequencing data from a SMRTbell™ template.

[0045] Figure 7 illustrates a hypothetical directed graph.

[0046] Figure 8 provides data from single-molecule sequencing-by-synthesis reactions. Figure 8A provides data from a two-minute interval beginning at initiation of the reactions, i.e., from 0-120 seconds. Figure 8B provides data from a second two-minute interval from 300-420 seconds. Figure 8C provides data from a third two-minute interval from 600-720 seconds.

[0047] Figure 9 schematically illustrates one embodiment of a system for use with the methods, devices, and systems of the invention.

[0048] Figure 10 provides a graphical representation of rates of polymerase activity on different portions of a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0049] Figure 11 provides a graphical representation of the average rate of polymerase translocation over a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0050] Figure 12 provides a distribution of the physical coverage of a template nucleic acid achieved during a sequencing reaction utilizing intermittent illumination, with A showing mapping to a reference sequence with sequence reads (and portions thereof) that do not map to the reference excluded and B showing a similar mapping that further includes sequence reads corresponding to insertions in the template that are absent from the reference sequence.

[0051] Figure 13 provides a distribution of the physical coverage provided by sequence reads generated during sequencing reactions utilizing intermittent illumination across an approximately 40 kb template nucleic acid.

[0052] Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described herein and a reference sequence.

[0053] Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system.

[0054] Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters $P(S \rightarrow P) = 1/24$; $P(P \rightarrow S) = 1/11$; and $p(v) \sim \text{Gamma}(48, 0.25)$. Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a non-detection period.

[0055] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period.

[0056] Figure 18 provides an exemplary strategy for selectively reducing the size of a duplex fragment within a SMRTbell™ template.

[0057] Figure 19 provides an illustrative example of nucleic acid templates having nicks.

DETAILED DESCRIPTION OF THE INVENTION

[0058] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. All publications mentioned herein are incorporated herein by reference for the purpose of

describing and disclosing devices, formulations and methodologies which are described in the publication and which might be used in connection with the presently described invention.

[0059] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polymerase" refers to one agent or mixtures of such agents, and reference to "the method" includes reference to equivalent steps and methods known to those skilled in the art, and so forth. Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either both of those included limits are also included in the invention.

[0060] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention. Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

I. General

[0061] In a general sense, the methods, devices, and systems provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a

single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0062] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion.

[0063] The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. For example, such an alteration in reaction conditions during the second stage may result in an increase in reaction rates, e.g., speeding up the progression of a template nucleic acid through

a nanopore; or may reduce exposure of reaction components to harmful radiation or other reaction condition related to detection of the products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0064] "Intermittent detection," as used herein, generally refers to a means of monitoring a reaction that is carried out intermittently during the course of the reaction. Intermittent detection may refer to intermittent use of one or more monitoring methods, but does not necessarily mean that all means of monitoring a given reaction are intermittently halted. For example, monitoring of one or more nucleotide incorporations to generate nucleotide sequence reads may be intermittently halted while other aspects of a sequencing reaction are constantly monitored, e.g., temperature, reaction time, pH, etc. In certain embodiments, intermittent detection is achieved by intermittent or differential illumination of a given reaction, e.g., a reaction that uses an illumination system to detect reaction products and/or progression. Although various aspects of the invention are described herein in terms of embodiments using intermittent illumination, it should be understood that where applicable intermittent detection by other means (e.g., electrochemical, radiochemical, etc.) can be utilized in the methods of the invention. Likewise, a stage of a reaction during which an intermittent detection method is active may be referred to as a "detection period" and a stage of a reaction during which an intermittent detection method is inactive may be referred to as a "non-detection period." In illuminated reactions, such periods may also be referred to as "illuminated periods" and "non-illuminated periods," respectively, although it is to be understood that the term "non-illuminated period" included periods in which illumination may be present but altered as compared to illumination during an "illuminated period." For example, a non-illuminated period may be characterized by a complete absence of illumination, or a modification of illumination, including but not limited to changes in wavelength, frequency, intensity, and/or number of illumination sources. Alternatively or additionally, reaction components that are excited by the illumination source(s) may be modified or removed from a reaction mixture to create a non-illuminated period. For example, a fluorescent dye detected during an illuminated period may be removed from the reaction mixture, e.g., by buffer exchange, thereby producing a non-illuminated period during which time the fluorescent dye cannot be detected even if the excitation illumination is present. In a further example, a non-illuminated period can indicate a period during an illuminated reaction

during which a type of illumination-based detection that occurs during an illuminated period is not occurring, e.g., the identity of fluorescently labeled nucleotides incorporated into a nascent strand is not being detected or recorded.

[0065] In certain aspects, the present invention is generally directed to improved methods, devices, and systems for performing illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In certain preferred embodiments, illuminated reactions comprise one or more fluorogenic or fluorescent components. In accordance with certain methods of the invention, such illuminated analyses are subjected to intermittent detection (e.g., data collection) for one or more aspects of the data typically collected for a given reaction. For example, aspects of the data typically collected for nucleotide sequencing reactions include nucleotide sequence data, read quality data, signal to background ratios, reaction rates and durations, measures of the fidelity of the reaction, reaction times, and the like. In certain preferred embodiments, nucleotide sequence data is iteratively collected during an ongoing sequencing reaction to generate nucleotide sequence reads for at least two or more noncontiguous regions of a template nucleic acid molecule. Such iterative sequence data acquisition may be achieved in various ways depending on the sequencing technology in use. For example, in sequencing methods that utilize luminescent components that generate a signal indicative of the identity of a base position, iterative sequence data collection may be achieved by removing or altering an illumination source (or a reaction relative to an illumination source), substituting the luminescent components for unlabeled components that do not generate signal, or otherwise interrupting signal acquisition in the experimental system.

[0066] In certain preferred embodiments, such illuminated reactions are illuminated for an amount of time that permits the effective performance of the analysis. Traditionally, illuminated reactions are illuminated from initiation through completion, and the time during which reaction data may be reliably collected is dictated by the progression (as measured by, e.g., processivity, rate, fidelity, duration, etc.) of the reaction under constant illumination. Some reactions are sensitive to such constant illumination, which can reduce their performance (e.g., processivity), and thereby prevent collection of data from later stages of the reaction, i.e., stages that would otherwise occur if the reaction were carried out with no illumination. The present invention provides methods for

performing illuminated reactions comprising subjecting the reactions to intermittent illumination. Such intermittent illumination can increase performance (e.g., processivity, rate, fidelity, duration, etc.) of the reactions, thereby allowing generation of data that cannot be collected under constant illumination, such as data from later stages of an ongoing reaction whose progression is compromised under constant illumination. For example, in sequencing-by-incorporation reactions the use of intermittent excitation illumination can increase processivity, which has the benefit of providing sequence reads more distal from the polymerase binding/initiation site than such reactions subjected to constant exposure to excitation illumination.

[0067] Further, it is an object of the instant invention to provide sequence data from noncontiguous regions of a nucleic acid template in a single reaction. Other commercially available platforms have attempted to achieve such noncontiguous sequence data through, e.g., complex cloning and sequencing strategies. The present invention provides a clear advantage over such strategies by providing a simple and economical solution that is applicable across various platforms, and is particularly applicable to illuminated, single-molecule sequencing-by-incorporation reactions.

[0068] In preferred embodiments, illuminated reactions for use with the instant invention are nucleic acid sequencing reactions, e.g., sequencing-by-incorporation reactions. In preferred embodiments, such an illuminated reaction analyzes a single molecule to generate nucleotide sequence data pertaining to that single molecule. For example, a single nucleic acid template may be subjected to a sequencing-by-incorporation reaction to generate one or more sequence reads corresponding to the nucleotide sequence of the nucleic acid template. For a detailed discussion of such single molecule sequencing, see, *e.g.*, U.S. Patent Nos. 6,056,661, 6,917,726, 7,033,764, 7,052,847, 7,056,676, 7,170,050, 7,361,466, 7,416,844; Published U.S. Patent Application Nos. 2007-0134128 and 2003/0044781; and M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, SCIENCE 299:682-686, January 2003 Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations, all of which are incorporated herein by reference in their entireties for all purposes. In some embodiments, a plurality of single nucleic acid templates are analyzed separately and often simultaneously to generate a plurality of sequence reads corresponding to the nucleotide sequences of the plurality of nucleic acid templates. In certain preferred embodiments, the plurality of nucleic acid templates includes at least two nucleic acid templates that comprise identical nucleotide sequences such that analysis of the two nucleic acid templates generates overlapping sequence reads. In certain preferred embodiments, at least one of

the nucleic acid templates is configured to provide redundant sequence data in a single sequence read, e.g., via duplications, sense and antisense sequences, and/or circularization.

[0069] Certain aspects of the invention are directed to methods, devices, and systems for generating a sequence scaffold for a nucleic acid template, e.g., chromosome, genome, or portion thereof. A sequence scaffold as used herein refers to a set of sequence reads that extends across at least a portion of a nucleic acid template. In some embodiments, such a sequence scaffold is used to generate a consensus sequence for the nucleic acid template. In some embodiments, the nucleic acid template is very large, e.g., at least about 100, 1000, 10,000, 100,000, or more bases or base pairs in length. In some embodiments, the sequence scaffold and/or consensus sequence is based on at least 1-, 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of at least a portion of the nucleic acid template. In some preferred embodiments, the portion of the nucleic acid is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the entire length of the nucleic acid template.

[0070] In certain aspects, the invention is particularly suitable for sequencing nucleic acid templates interspersed with repetitive elements. Such repetitive elements present major logistical and computational difficulties for assembling fragments produced by sequencing strategies, especially those with read-lengths that are too short to encompass unique reads outside the repeat region. For example, the human T-cell receptor locus contains a five-fold repeat of a trypsinogen gene that is 4 kbp long and that varies 3 to 5% between copies. Therefore, a sequencing strategy that cannot provide nucleotide sequence information that spans at least 20 kb for a single molecule containing the locus will have difficulty providing consensus sequence for the locus. Further, Alu repeats (~300 bp retrotransposons) are also problematic because they cluster and can constitute up to 50-60% of the template sequence, with copies varying from 5-15% between each other. The human genome contains an estimated one million Alu repeats and 200,000 LINE elements (average length ~1000 bp), representing roughly 10% and 5% of the entire genome, respectively. In certain embodiments, the present methods facilitate efficient and accurate sequence determination for long templates comprising such repetitive sequences, in part because the present methods do not rely solely on sequence overlap to generate consensus sequences, but also include information related to the expected location of the polymerase on the template nucleic acid, thereby linking a particular sequence read to a particular location on the template nucleic acid. This greatly facilitates accurate assembly of sequence reads to generate sequence scaffolds and/or consensus sequences.

[0071] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads in an illuminated sequencing-by-incorporation reaction that are distal from one another (i.e., noncontiguous) on a single nucleic acid template by removing the excitation illumination during the course of the reaction, and subsequently reinitiating the excitation illumination. Sequence reads are generated only during the periods of time when the excitation illumination is present, resulting in a "gap" between the sequence reads from a single template nucleic acid that corresponds to the time during which the excitation illumination was absent but the incorporation of nascent nucleotides continued "in the dark." As such, the number of sequence reads generated for a given template nucleic acid is equal to the number of periods during which the excitation illumination is present.

[0072] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads from a plurality of nucleic acid templates comprising identical nucleotide sequences. In some embodiments, the multiple sequence reads are not all from the same region of the nucleic acid templates. In some embodiments, there is overlap between the multiple sequence reads. In some embodiments, a single sequence read is generated from each of the plurality of nucleic acid templates, and in other embodiments multiple noncontiguous sequence reads are generated from each of the plurality of nucleic acid templates. In certain preferred embodiments, the multiple noncontiguous sequence reads from each of the plurality of nucleic acid templates together extend across the nucleic acid templates such that they can be combined to provide a consensus sequence for the identical nucleotide sequence in the nucleic acid templates. In some embodiments, the consensus sequence is based on at least 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of the identical nucleotide sequence. In some embodiments, the identical nucleotide sequence represents at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the nucleic acid template.

[0073] Certain aspects of the invention are directed to methods, devices, and systems for reducing or limiting the effects of photo-induced damage during illuminated reactions, particularly reactions that employ fluorescent or fluorogenic reactants. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction. Without being bound to a particular theory or mechanism of operation, some illuminated reactions are subject to photo-induced damage that can hinder progression of the reaction, e.g., via damage to reaction components, such as enzymes, cofactors, templates, etc. As such, the illumination of the illuminated reaction can directly or

indirectly negatively impact progression of the reaction, and such an impact can be measured based on various characteristics of the reaction progression, e.g., processivity, rate, fidelity, duration, etc. The present invention provides methods for subjecting an illuminated reaction to intermittent exposure to illumination, which reduces the amount of photo-induced damage at a given time during the reaction, allowing the reaction to proceed further than it does when constantly exposed to the illumination.

[0074] In some embodiments, the methods herein may further comprise the addition of one or more photo-induced damage mitigating agents (e.g., triplet-state quenchers and/or free radical quenchers) to the illuminated reaction. Such photo-damage mitigating agents are generally known to those of skill in the art. Further discussion of photo-induced damage and related compounds, compositions, methods, devices, and systems are also provided in U.S. Pub. No. 20070161017, filed December 1, 2006; and U.S.S.N. 61/116,048, filed November 19, 2008, which are incorporated by reference herein in their entireties for all purposes.

II. Intermittent Illumination of Analytical Reactions

[0075] Certain aspects of the invention are generally directed to improved methods for performing illuminated analyses. The terms "illuminated analysis" and "illuminated reaction" are used interchangeably and generally refer to an analytical reaction that is occurring while being illuminated (e.g., with excitation radiation), so as to evaluate the production, consumption, and/or conversion of luminescent (e.g., fluorescent) reactants and/or products. As used herein, the terms "reactant" and "reagent" are used interchangeably. As used herein, the terms "excitation illumination" and "excitation radiation" are used interchangeably. In certain embodiments, the illuminated reaction is a sequencing reaction, e.g., a sequencing-by-incorporation reaction. In certain embodiments, the illuminated reaction is designed to analyze a single molecule, e.g., by ensuring the molecule is optically resolvable from any other molecule being analyzed and/or in the reaction mixture. In certain embodiments, one or more components of the reaction are susceptible to photo-induced damage directly or indirectly elicited by an excitation radiation source. In certain preferred embodiments, an illuminated reaction is subjected to intermittent excitation radiation during the course of the illuminated reaction. In certain preferred embodiments, a sequencing-by-incorporation reaction is subjected to intermittent excitation radiation during the course of a polymerization reaction to generate a plurality of noncontiguous sequence reads from a single nucleic acid template.

[0076] In certain aspects, the methods herein provide benefits over methods currently used for sequencing large template nucleic acids, such as human genomes. For example, the traditional shotgun sequencing approach entails sequencing nucleic acid fragments and analyzing the resulting sequence information for overlap and similarity to known sequences to construct the complete sequence of the template nucleic acid. One disadvantage to the shotgun approach is that assembly may be difficult if the template nucleic acid comprises numerous repeated sequences, and the inability to assemble a genomic sequence in repeat regions leads to gaps in the assembled sequence. (See, e.g., Myers, G.; "Whole-Genome DNA Sequencing" in Computing in Science and Engineering; Vol 1, Issue 3; pgs. 33-43; May/Jun 1999.) One method of resolving these gaps is to sequence fragments large enough to span the repeat regions, but sequencing large fragments can be difficult and time-consuming. Another approach to spanning a gap is to determine the sequence of two ends of a large fragment which has known spacing and orientation, and this approach is generally termed paired end sequencing (see, e.g., Smith, M. W. et al., (1994) Nature Genetics 7:40-47; and U.S. Pub. No. 2006/0292611, filed June 6, 2006, both of which are incorporated by reference herein in their entireties for all purposes). This method is limited by the requirement for information about the spacing and orientation of the ends of the long fragment, and/or complex sample preparation of the nucleic acid template. The present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.) or complex sample preparation, thereby allowing economical, efficient, and effective de novo sequencing or resequencing of long template nucleic acids.

[0077] In certain aspects, the methods herein provide various strategies for achieving intermittent illumination of illuminated reactions. Essentially, at least one type of illumination (e.g., excitation illumination) is present for at least one time period ("illuminated period") and absent during at least one other time period ("non-illuminated period") during an illuminated reaction. As described above, the term "non-illuminated" indicates a change in illumination including, but not limited to a complete absence of illumination. For example, a non-illuminated period may also be characterized by a different illumination source or intensity than an illuminated period, or by a change in reaction components, e.g., detectable labels. In general, at least one type of data collected during an illuminated period (e.g., nucleotide sequence data) is not collected during a non-illuminated period. An absence of the illumination may be due to, e.g., inactivation of the illumination source (e.g., laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb,

and/or a white light source), removal of the illuminated reaction from the illumination source (or vice versa), or may be due to blockage of the illumination from the reaction, as discussed below. Modifications to the illumination may be due to, e.g., adjustment of the intensity of an illumination source, or a substitution of one illumination wavelength and/or frequency for another. Further, components detectable during an illuminated period may be removed from the reaction mixture during a non-illuminated period, e.g., a fluorescently labeled nucleotide may be replaced with an unlabeled nucleotide. Knowledge of the rate of the reaction and the time during which the illumination is absent is used to estimate the progress of the reaction during the non-illuminated period. For example, if a reaction proceeds such that one molecule is incorporated into a macromolecule per second, and the illumination is absent for 20 seconds, it can be estimated that 20 molecules were incorporated during the non-illuminated period. This information is useful during data analysis to provide context for the reaction data collected during the illuminated period(s). For example, in a sequencing-by-incorporation reaction the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. Further the processivity of an enzyme during a non-illuminated period (or other type of non-detection period) can be manipulated or adjusted by methods known to those of skill in the art. In particular, the kinetics of replication by a polymerase enzyme can be altered by changing the chemical environment in which it operates, and such methods are further described, e.g., in U.S. Patent Application Nos. 12/414,191, filed March 30, 2009; 12/537,130, filed August 6, 2009; and U.S. Patent Application No. [unassigned], attorney docket no. 105-006301US, entitled "Engineering Polymerases and Reaction Conditions for Modified Incorporation Properties," filed September 4, 2009, the disclosures of all of which are incorporated herein by reference in their entireties for all purposes. For example, methods are provided for adjusting the enzyme activity, and these methods find particular relevance in the instant invention when used to enhance accuracy during detection periods, and to enhance processivity during non-detection periods. Information regarding enzyme translocation rate and processivity is useful for positioning the sequence reads for

a single template nucleic acid relative to one another in the construction of a sequence scaffold and/or consensus sequence for the template nucleic acid.

[0078] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions. A reaction mix is prepared at step 100. In process A shown on the left, illumination of the reaction 105 is begun prior to initiation of the reaction 110, which allows "illumination data" to be collected at initiation. (In an alternative embodiment, illumination may commence simultaneously with initiation of the reaction.) "Illumination data" as used herein refers to data collected during an illuminated period, e.g., the length of the illuminated period and luminescent signal(s) from the reaction product. At least one non-illuminated period 115 occurs during the course of the reaction, followed by at least one additional illuminated period 120. Multiple additional non-illuminated and illuminated periods may follow. During the illuminated periods (105 and 120), illumination data is collected 175. During the non-illuminated period(s), non-illumination data is collected 180. As used herein, "non-illumination data" refers to data collected during a non-illuminated period, e.g., the length of the non-illuminated period can be monitored. In process B shown on the right, the reaction is initiated 155 during a first non-illuminated period 150. At least one illuminated period 160 occurs during the course of the reaction, optionally followed by at least one additional non-illuminated period 165. Multiple additional illuminated and non-illuminated periods may follow. As for process A, illumination data is collected 175 during the illuminated period(s) 160, and non-illumination data is collected 180 during non-illuminated periods (155 and 165).

[0079] One benefit provided in certain embodiments of the invention is that the reaction need not be further manipulated after initiation (aside from the control of illumination). For example, the method can be used to analyze reaction mixtures without the need for buffer changes, addition of further reaction components, or removal of detectable components, e.g., light-activatable components such as fluorophores. For example, in a sequencing-by-incorporation reaction, labeled nucleotides may be present throughout the life of the reaction, even when the reaction is not generating nucleotide sequence data (e.g., during a non-illuminated period). This provides clear advantages over methods that require additional handling of the reaction after initiation, which tend to not only be expensive and time-consuming, but which also provide opportunities for contamination of the reaction. For example, illumination can be reinitiated at any time during the reaction at the whim of the ordinary practitioner by simply activating the illumination. In certain preferred embodiments, the concentration of labeled nucleotides or nucleotide analogs in the

reaction mixture is greater than the concentration of unlabeled nucleotides in the reaction mixture throughout the course of the reactions, and may represent at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the total nucleotides in the reaction mixture. Methods for ensuring a high ratio of labeled versus unlabeled nucleotides in a reaction mixture are known in the art and certain preferred embodiments are provided in U.S. Patent Pub. Nos. 2006/0063264, 2006/0194232, and 2007/0141598, which are incorporated herein by reference in their entireties for all purposes.

[0080] In embodiments in which a sequencing-by-incorporation reaction is subjected to intermittent illumination, the sequence reads collected during the illuminated periods are arranged in order and separated from one another by an estimated number of nucleotides incorporated into the nascent strand during the intervening non-illuminated periods. The resulting gapped read can then be used to assess certain characteristics of the template nucleic acid. When multiple identical template nucleic acids are subjected to such a sequencing-by-incorporation reaction, the resulting set of gapped reads can be combined to create a sequence scaffold and/or a consensus sequence for the template nucleic acid.

[0081] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods can be used to provide additional data during the course of the reaction, e.g., data from illuminated or non-illuminated periods. In certain preferred embodiments, such alternative labeling methods may comprise using labels that are incorporated into a product of the reaction. For example, in sequencing-by-incorporation reactions that use nucleotides comprising labeled terminal phosphates (*e.g.*, the gamma phosphate as in dNTP, or terminal phosphates on nucleotide analogs with a greater number of phosphate groups) to identify the nucleotides incorporated into a nascent polynucleotide, the reaction mixture may also include nucleotides comprising a base-linked label. During the reaction, these "base-labeled nucleotides" will be incorporated into the nascent strand, but unlike the terminal phosphate labels removed during incorporation, the base-linked labels are not cleaved from the nucleotide upon incorporation by the polymerase, resulting in a nascent strand that comprises the base-linked labels. The concentration of such base-labeled nucleotides can be adjusted in the reaction mixture to promote their incorporation into the nascent strand at a predictable rate, e.g., based on the known sequence of the template or the average frequency of a given nucleotide. The presence and/or rate of incorporation of the base-linked labels into the nascent strand can provide a measure of the length of the nascent strand generated (and, therefore, the distance traveled by the polymerase along the

template nucleic acid) during the reaction by subjecting the reaction to excitation illumination that excites the base-linked label (but preferably not the non-base-linked labels), and detecting the signal emitted. The excitation of the base-linked labels preferably occurs as a pulse during or immediately following a non-illuminated period, and is otherwise absent during the reaction. The strength of the signal is indicative of how many labels are present in the nascent strand, thereby providing a measure of the processivity of the polymerase for a given period during the ongoing reaction, e.g. during one or more illuminated or non-illuminated periods. Since the base-linked labels remain in the nascent strand, it is beneficial to minimize the amount of time those fluorophores are subjected to excitation illumination to mitigate the potential of photo-induced damage to the reaction components. As such, in preferred embodiments, the excitation illumination wavelength for the base-labeled nucleotides is different than that of other fluorescent labels in the reaction.

[0082] This method can be modified in various ways. For example, the base-labeled nucleotides may also comprise a terminal phosphate label so that their incorporation can be monitored in the same manner during an illuminated period as the non-base-labeled nucleotides. There may be a single type of base-labeled nucleotide in a reaction mixture, or multiple types may be present, e.g., each type carrying a different nucleobase. The concentration of base-labeled nucleotides in the reaction mix may be varied, although it is preferred that the ratio of base-labeled nucleotides to non-base-labeled nucleotides be relatively low. For example, in a reaction mixture comprising a single type of base-labeled nucleotide (e.g., base-labeled dATP), it is preferred that the ratio of base-labeled dATP to non-base labeled dATP be less than 1:8, and more preferably 1:10 or less. The low concentration of base-labeled nucleotides is preferred in order to minimize sterically induced polymerase stalling when incorporating multiple base-labeled nucleotides in a row. In some embodiments, the optimal ratio is pre-determined using capillary electrophoresis for any specific base-labeled nucleotide and likely homopolymer sequence prevalence. In certain preferred embodiments, at least 50, 75, 100, 125, or 150 base-labeled nucleotides are incorporated into the nascent strand during a single non-detection period. The base-labeled nucleotides may be present throughout the reaction, or may be washed in during non-illuminated periods and washed out after the pulse of excitation illumination. The reaction mixture comprising base-linked nucleotides being washed in may also include unlabeled nucleotides for incorporation during a non-detection period. During a subsequent illuminated period, a reaction mixture comprising terminal phosphate-labeled nucleotides replaces the reaction mixture comprising base-linked nucleotides and unlabeled nucleotides. This protocol is one embodiment of the methods of the invention in which a non-

detection period is not necessarily a non-illuminated period because in this case illumination may be present, but no incorporation of nucleotides is be detected.

[0083] Alternatively or in addition, a low concentration of a fifth terminal phosphate labeled nucleotide can be present in the sequencing reaction, wherein the label has a different excitation wavelength than the other labels in the reaction mixture. For example, a small proportion of one nucleotide analog, e.g., dA6P, can be labeled with the "fifth label." During non-detection periods when the sequence of incorporation of nucleotides is not being monitored, the reaction site is illuminated by excitation radiation specific for the fifth label, and this fifth label excitation radiation can be inactivated during the detection periods. Emissions detected upon incorporation of the nucleotide analog comprising the fifth label are used to "clock" the pace of the polymerase during the non-detection period, e.g., based upon the known or estimated frequency of the complementary nucleotide in the template strand. The fifth label can be chosen such that the excitation and emission radiation are less likely or unlikely to cause photo-induced damage to reaction components, e.g. by choosing a label with a long excitation wavelength (e.g., toward the red end of the visible spectrum), a label that has a low propensity for entering into a triplet state, and/or a label that has a low propensity to form a radical. Since the fifth label is being excited when other labels are not, there is no requirement for optimal spectral separation from other labels in the reaction mixture. Further, since the fifth label is not being used for sequencing, other optimizations are also not necessary, e.g., related to branching, accuracy, and the like. Various types of labels can be used as a fifth label of the invention including, but not limited to, organic and non-organic dye fluorophores. For example, latex nanoparticles or quantum dots are particularly suitable due to their lower propensity for photo-induced damage of certain analytical reaction components. In certain preferred embodiments, a quantum dot label has an emission spectrum within the same spectral window as the labels that are used to identify the sequence of base incorporations into the nascent strand ("sequencing labels") but an excitation spectrum that does not overlap those of the sequencing labels to allow detection of the fifth label emissions using the same optical system as is used to detect the sequencing label emissions.

[0084] This method can be modified in various ways. For example, more than one small subset of a nucleotide analog can be labeled with a fifth label, and in certain embodiments, a small subset of each nucleotide analog present in the reaction mixture is labeled with the fifth label. Further, there may be a plurality of additional labels present in the reaction, each of which is present on a small subset of a single type of nucleotide analog, e.g., sixth, seventh, and eighth labels. By

increasing the number of types of nucleotide analogs labeled with fifth (or sixth, seventh, eighth) labels, their frequency of incorporation is likewise increased, which improves the translocation rate calculation for the polymerase during the non-detection periods. Alternatively, each type of nucleotide analog can comprise both a sequencing label that is specific for the cognate base in the nucleotide, as well as a fifth label for clocking the polymerase. The sequencing labels are excited and detected during the detection periods and the fifth labels are excited and detected during the non-detection periods. Since every nucleotide analog is labeled with a fifth base, each incorporation event can be counted during the non-detection period and the exact rate of incorporation can be determined. Both the sequencing and fifth labels may be bound to the same or different linkers on the nucleotide analogs. In certain preferred embodiments, a linker on a nucleotide analog positions the fifth label within an illumination zone to allow excitation, but far from an enzyme (e.g., polymerase) to mitigate photo-induced damage related to excitation of and/or emission from the fifth label.

[0085] In some embodiments, the fifth label is also excited by an illumination during the detection periods. The availability of the clocking function during the detection period can be used during sequence analysis to identify positions in the resulting sequence read where a signal was not detected (resulting in an apparent "missing base" in the read) and to distinguish between true insertions and branching events in which two signals are detected for a single incorporation event.

[0086] In yet further embodiments, assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid is facilitated by using "non-illuminated periods" characterized by modified excitation illumination rather than a complete absence of excitation illumination (which can also be termed "low-illuminated periods"). For example, in some embodiments a lower intensity excitation illumination is used during the non-illuminated periods that excites one or more of the labels that are excited during the illuminated periods. As such, unlike various strategies described above, no fifth label is necessary. The lower intensity excitation illumination results in emissions that are lower intensity but still intense enough to identify an emission signal over background counts, though typically not intense enough to be used to identify the particular label generating the emission signal. For example, if label "A" and label "B" are in a reaction mixture, during an illuminated period the intensity of the signal emissions from each are high enough that the artisan can distinguish from which label a particular signal originates by the wavelength and/or frequency of the signal. However, during a low-illuminated period the artisan can only identify that a signal emission occurs, but is unable to distinguish the originating label

because its particular wavelength and/or frequency cannot be accurately determined. The decrease in excitation illumination intensity provides both a mitigation of photo-induced damage to reaction components within the observation volume while allowing the practitioner to count the emissions, and therefore the incorporations, during the non-illuminated period.

[0087] In other embodiments, multiple excitation illumination sources are used during an illuminated period, and a first subset of these illumination sources is removed during a non-illuminated period, while a second subset remains. The illumination sources that remain during the non-illuminated period may be present in the same manner as during the illuminated period, or various aspects may be altered, e.g., intensity may be reduced. For example, if labels A and B present in a reaction mixture are excited by a first illumination source and labels C and D present in the reaction mixture are excited by a second illumination source, removal of the first illumination source during the non-illuminated period results in an inability to detect labels A and B, while C and D are still detectable. Such an incomplete data set can be used to clock the progress of the reaction during the non-illuminated period(s). Further, it can also be used in various ways to facilitate the statistical analysis of data collected during the illuminated period(s). For example, for nucleotide sequencing applications (as described elsewhere herein) the incomplete data set(s) collected during non-illuminated period(s) can be used during assembly of a sequence scaffold. For example, during *de novo* sequence assembly a collection of sequences (contigs) are generated, but the order of the contigs relative to the template nucleic acid is not always apparent. The scaffolding process uses extra information to determine the correct order of the contigs. So, if only two bases are identifiable in the non-illuminated periods, the incomplete sequence reads comprising only incorporation of these two bases can be aligned to modified versions of the contigs assembled from data collected during an illuminated period, but in which the two bases not detected during the non-illuminated periods have been removed. Once the order of the contigs has been determined, the incorporation data for the two bases not detected during the non-illuminated periods is restored and the assembly of the contigs is complete. This method can be modified in various ways. For example, the practitioner may choose which illumination sources to remove during the non-illuminated periods based on various characteristics, such as their propensity to cause photo-induced damage to one or more reaction components, the propensity of the corresponding emission signal to cause photo-induced damage to one or more reaction components; their energy consumption; and wear-and-tear on the source device. Further, as described elsewhere herein, rather then removing an illumination source, reaction components that are excited by the illumination source may be

removed from the reaction mixture during the non-illuminated period, necessarily rendering them undetectable. For example, one or more fluorescently labeled nucleotide analogs may be replaced with unlabeled nucleotide analogs during the non-illuminated periods.

[0088] In certain aspects, the invention provides advantages to performing intramolecular redundant sequencing, in which a template nucleic acid is used to generate multiple copies of a sequence read of interest, whether by virtue of multiple copies of the complement being present in the template, repeated replication of the template, or a combination thereof. For example, a first stage of a template-dependent sequencing reaction on a single-stranded circular template can comprise a non-illuminated period during which the template is completely replicated at least one time to generate at least one incomplete sequence read for a sequence complementary to the template. The first stage is followed by a second stage comprising an illuminated period during which the template is replicated multiple times to generate multiple complete sequence reads for the complementary sequence. The incomplete reads generated in the first stage can be used to construct a scaffold for assembly of the complete sequence reads generated in the second stage. Further, incomplete sequence reads can also be used to clock the progress of the reaction during the non-illuminated periods by providing a count of the detectable reaction components and combining that information with known or estimated characteristics of the template, e.g., nucleotide composition or sequence.

[0089] The subset of signal emissions detectable in the non-illuminated periods as compared to the number detectable in the illuminated periods is not limiting and may be chosen based upon the non-illumination data desired by the ordinary practitioner and/or other considerations, such as mitigation of photo-induced damage to extend readlength. For example, to lower the likelihood of photo-induced damage, the ordinary practitioner may choose to remove the illumination source that is most damaging, e.g., has the highest frequency. In certain embodiments, multiple sequencing reactions may be performed for a single amplified template, each with a different combination of illumination sources and/or detectable components. Alternatively or additionally, multiple replicate reactions can also be performed for one or more of the combinations of illumination sources and/or detectable components. The combination of data from multiple different and/or replicate reactions performed on a single template provides myriad benefits during statistical analysis. As noted above, data can be combined to facilitate assembly of contigs generated during illuminated periods. Data from non-illuminated periods can also provide value in assessing the quality of the sequence reads generated during the illuminated periods.

[0090] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods used to provide additional data during the course of the reaction can comprise using labels that are incorporated into an enzyme of the reaction. For example, FRET labels can be used to label portions of a polymerase enzyme such that the conformational change between the open and closed states of the enzyme change the FRET value. For example, a FRET-based system can be used to monitor the kinetics of opening and closing of the finger subdomain of DNA polymerase, as described in Allen, et al. (2008) Protein Science 17:401-408, incorporated herein by reference in its entirety for all purposes. In certain preferred embodiments, a closed conformation produces a FRET signal because the donor and acceptor are close to one another, and an open conformation silences the signal because there is no energy transferred between the donor and acceptor. By monitoring the emission from the FRET pair, each incorporation event can be monitored during non-detection periods, and optionally or additionally during detection periods. In certain preferred embodiments, the FRET donor is GFP (excitation at 484 nm; emission at 510 nm), and the FRET acceptor is YFP (excitation at 512 nm; emission at 529 nm). Methods for monitoring polymerase activity using FRET labels are known in the art, e.g., in WO/2007/070572 A2, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

[0091] A given reaction may experience one or a plurality of illuminated periods or non-illuminated periods, but preferably experiences at least two illuminated periods. For example, a given reaction providing nucleotide sequence information from a single template nucleic acid may have at least about 2, 3, 5, 10, 20, 50, or 100 illuminated periods with intervening non-illuminated periods. In an embodiment employing multiple periods of illumination and/or non-illuminated, the periods may be the same for both, e.g., 100 seconds "on" and 100 seconds "off." Alternatively, the illuminated periods may be longer or shorter than the non-illuminated periods. For example, in certain embodiments, a non-illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent illuminated period; or an illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent non-illuminated period. Further, each illuminated period may be the same or different from each other illuminated period, and each non-illuminated period may be the same or different from each other non-illuminated period. For example, some embodiments generate a smaller number of long reads, and other embodiments generate a larger number of short reads. It will be understood that the number and length of the

illuminated and non-illuminated periods is limited only by the experimental system in use and the data acquisition goals of the ordinary practitioner. In some embodiments, a nucleotide sequence read generated during a single illuminated period comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, a region of a nucleic acid template processed during a non-illuminated period during a single reaction comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, the set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises at least about 40, 60, 80, 100, 1000, 10,000, 25,000, 50,000, 100,000, 250,000, 500,000, or 1,000,000 nucleotide sequence positions from a single nucleic acid template. In some embodiments, a set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises multiple reads of at least a portion of the nucleotide sequence positions from a single nucleic acid template.

[0092] As noted above, the present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.). However, such information, if available, may also be useful to the ordinary practitioner in determining an optimal periodicity for illuminated and non-illuminated periods during a sequencing reaction, especially when sequencing repetitive sequences. For example, if a genomic region is known to contain five adjacent copies of a one kilobase nucleotide sequence (i.e., five "repeat regions"), it would be beneficial to keep the non-illuminated periods short enough to be able to confidently map the resulting sequence reads to the correct repeat region. If a non-illuminated period were too long, the natural variation in translocation rate of the polymerase would make it difficult to assign a sequence read to a particular repeat region, especially those farther from the binding/initiation site of the polymerase. In a further example, if the "copies" each had a few mutations that could be used to distinguish them from each other, it would be beneficial to keep the illuminated periods long enough to increase the chance one of these mutations would be included in a resulting sequence read, thereby allowing the unambiguous assignment of the read to a particular repeat region. If the illuminated period were too short the sequence reads from two different repeat regions could be identical, making mapping the sequence read challenging. (Another way to mitigate these difficulties would be to incorporate pause or stop points into the template nucleic acid, as discussed below.)

[0093] Essentially, the practitioner may design the number of and lengths of time for each illuminated and non-illuminated period to best suit the illuminated reactions being analyzed and the invention is not limited in this regard. In certain embodiments, a practitioner may wish to increase the processivity of a polymerase thereby extending the length of the template nucleic acid processed in a sequencing reaction to be, e.g., at least 2-, 3-, 4-, 6-, 8-, 10-, or 20-fold, thereby generating sequence data much farther away from the polymerase binding/initiation site than would be achieved under constant illumination. In certain embodiments, a practitioner of the instant invention may wish to focus on data from one or more stages of an ongoing reaction, such as stages for which more data is required for analysis. In the case of sequencing-by-synthesis, one or more particular regions of a template nucleic acid may need to be resequenced. Some traditional methods require that new template nucleic acids be prepared to bring a region requiring resequencing closer to the initiation point of the sequencing reaction, or require preparation of multiple new templates if multiple regions to be resequenced. In contrast, the methods herein allow the practitioner to subject a template identical to the previously sequenced template (e.g., from a large genomic DNA sample preparation) to a sequencing reaction wherein illuminated periods are timed to illuminate the sample only when the polymerase is incorporating nucleotides into the nascent strand at the one or more particular regions requiring resequencing. This advantage substantially lowers the time and resources required for such resequencing operations, therefore providing a significant advantage over traditional methods.

[0094] The instant invention contemplates various means for providing non-illuminated periods during illuminated reactions. In some embodiments, the illumination source is turned off during the ongoing reaction to create one or more non-illuminated periods. In some embodiments, the illumination source remains on during the course of the reaction, but the illuminated reaction is removed from the system for a period of time. In some embodiments, the illumination source remains on during the course of the reaction, but the illumination is blocked to create one or more non-illuminated periods. For example, a movable mask may be manually or mechanically positioned between the illumination source and the illuminated reaction to block the illumination during non-illuminated periods and removed to allow exposure to the illumination during illuminated periods. Such a mask may also be dynamically controlled, such as a thin film transistor display (e.g., an LCD mask). Masks for blocking illumination and manufacture thereof are well known to those of ordinary skill in the art and need no further elaboration herein.

[0095] One aspect of the present invention is multiplexing of large numbers of single-molecule analyses. For a number of approaches, e.g., single molecule methods as described above, it may be desirable to provide the reaction components in individually optically resolvable configurations, such that a single reaction component or complex can be individually monitored. Providing such individually resolvable configurations can be accomplished through a number of mechanisms. For example, by providing a dilute solution of complexes on a substrate surface suited for immobilization, one will be able to provide individually optically resolvable complexes. (See, e.g., European Patent No.1105529 to Balasubramanian, et al., the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Alternatively, one may provide a low density activated surface to which complexes are coupled. (See, e.g., Published International Patent Application No. WO 2007/041394, the full disclosure of which is incorporated herein by reference in its entirety for all purposes). Such individual complexes may be provided on planar substrates or otherwise incorporated into other structures, e.g., zero-mode waveguides or waveguide arrays, to facilitate their observation.

[0096] In some embodiments, a plurality of illuminated reactions are carried out simultaneously, e.g., on a solid support. In some preferred embodiments, a solid support comprises an array of reaction sites. In preferred embodiments, the reaction sites on a solid support are optically resolvable from each other. In further preferred embodiments, each of the reaction sites on a solid support contains no more than a single reaction to be interrogated. For example, in a sequencing-by-incorporation embodiment, each reaction site preferably has no more than one polymerase and no more than one nucleic acid template. The reaction sites may be confinements (e.g., optical and/or physical confinements), each with an effective observation volume that permits resolution of individual molecules present at a concentration that is higher than one nanomolar, or higher than 100 nanomolar, or on the order of micromolar range. In certain preferred embodiments, each of the individual confinements yields an effective observation volume that permits resolution of individual molecules present at a physiologically relevant concentration, *e.g.*, at a concentration higher than about 1 micromolar, or higher than 50 micromolar range or even higher than 100 micromolar. In addition, for purposes of discussion herein, whether a particular reagent is confined by virtue of structural barriers to its free movement, or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined."

[0097] As used herein, a solid support may comprise any of a variety of formats, from planar substrates, e.g., glass slides or planar surfaces within a larger structure, e.g., a multi-well

plates such as 96 well, 384 well and 1536 well plates or regularly spaced micro- or nano-porous substrates, or such substrates may comprise more irregular porous materials, such as membranes, aerogels, fibrous mats, or the like, or they may comprise particulate substrates, e.g., beads, spheres, metal or semiconductor nanoparticles, or the like. The solid support may comprise an array of one or more zero-mode waveguides or other nanoscale optical structures.

[0098] As used herein, "zero-mode waveguide" refers to an optical guide in which the majority of incident radiation is attenuated, preferably more than 80%, more preferably more than 90%, even more preferably more than 99% of the incident radiation is attenuated. As such high level of attenuation, no significant propagating modes of electromagnetic radiation exist in the guide. Consequently, the rapid decay of incident electromagnetic radiation at the entrance of such guide provides an extremely small observation volume effective to detect single molecules, even when they are present at a concentration as high as in the micromolar range. The fabrication and application of ZMWs in biochemical analysis, and methods for calling bases in sequencing-by-incorporation methods are described, e.g., in U.S. Patent Nos. 7,315,019, 6,917,726, 7,013,054, 7,181,122, and 7,292,742, U.S. Patent Pub. No. 2003/0174992, and U.S. Patent Application No. 12/134,186, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[0099] A set of reactions (e.g., contained on a solid support) may comprise identical or different components. For example, a single template nucleic acid may be analyzed in all reactions in the set, or a plurality of template nucleic acids may be analyzed, each present in only one or a subset of the set of reactions. In preferred embodiments, template nucleic acids comprising the same nucleotide sequence are analyzed in a plurality of reactions sufficient to provide adequate redundant nucleotide sequence data to determine a consensus sequence for the template nucleic acids. A number of sequence reads that will provide adequate nucleotide sequence data will vary, depending, e.g., on the quality of the template nucleic acid and other components of the reaction, but in general coverage for a template nucleic acid or portion(s) thereof is at least about 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage. Further, the numbers and lengths of illuminated and non-illuminated periods for a given reaction in the set of reactions may be the same or different than those for other reactions in the set. In some embodiments, a mixture of different periodicities are used for a set of reactions comprising the same template nucleic acid. This strategy can be beneficial for providing nucleotide sequence reads from varying regions of the template sequence, thereby increasing the likelihood of overlapping sequence reads between individual reactions. These

overlapping sequence reads can facilitate construction of a more robust sequence scaffold than could be constructed were the reactions all subjected to the same periodicity of illuminated and non-illuminated periods.

[00100] Methods of controlling polymerase progress and/or synchronizing polymerases in different reactions are also useful in analysis (e.g., mapping, validation, etc.) of nucleic acid reads farther from the initial binding site of the polymerase. During detection periods earlier in the reaction (i.e., closer to the time at which the polymerase began to process the template nucleic acid, such as during a first illuminated period), the position of a polymerase on the template can be estimated with generally good accuracy based on the known translocation rate of the polymerase under a given set of reaction conditions. As the duration of the reaction increases, however, the natural variation in polymerase translocation rate makes it more difficult to accurately determine the exact position of the polymerase on a template using estimation based on translocation rate alone; and through each subsequent illuminated period such estimations of polymerase position become less accurate, making subsequent analysis and mapping of the sequence reads to the template more difficult. Methods of regulating the position of the polymerase on the template allow more accurate determinations the polymerase's position. For example, causing the polymerase to pause or stop at a given location on the template during a non-illuminated period and reinitiating the polymerization during or immediately prior to a subsequent illuminated period provides a way to reorient the subsequently generated read with the template sequence, allowing easier consensus sequence determination and mapping analyses. Further, such pause/stop points can provide a means of controlling what regions of the template are processed during the illuminated periods by restricting where the polymerase will reinitiate on the template, thereby allowing a practioner of the instant invention to target one or more particular regions of a template for analysis during one or more detection periods during the course of an analytical reaction. Such methods are also useful to synchronize a set reactions being monitored simultaneously. For example, a plurality of reactions, each comprising a single polymerase/template complex, may be synchronized by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that show less spreading (i.e., less variation in the position on the template from which the sequence reads are generated) in the later stages of the reactions than would otherwise be observed without such regulation.

[00101] Various methods can be used to control or monitor the progress of a polymerase on a template nucleic acid. For example, as noted above, one may employ a reaction stop or pause point

within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the single-stranded portion that was not used in priming. Reaction stop or pause points can be engineered into a portion of the template for which the nucleotide sequence is unknown (e.g., a genomic fragment), but is preferably located within a portion for which the nucleotide sequence is known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, certain preferred sequencing templates (e.g., SMRTbell™ templates, described elsewhere herein) are closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, and one or both of these linkers can comprise a stop or pause point to control the passage of the polymerase through them. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00102] By way of example, at a selected time following initiation of polymerization the reaction may be subjected to a non-illuminated period. The incorporation of a synthesis blocking moiety coupled to the template nucleic acid at a position encountered by the polymerase during the non-illuminated period will cause the polymerase to pause. An example of an engineered pause point is a known sequence on the template nucleic acid where a primer sits and blocks progression of a polymerase that is actively synthesizing a complementary strand. The presence of the primer by itself could introduce a pause in the polymerase sequencing or the primer could be chemically modified to force a full stop (and synchronization of multiple polymerases in multiple reactions). The chemical modification could be subsequently removed (for example, photo-chemically) and the polymerase would subsequently continue along the template nucleic acid. In some embodiments, multiple primers could be included in a reaction to introduce multiple pause or stop points along the template nucleic acid. Other methods for inducing a reversible pause (stop) in synthesis are known in the art and include, e.g., reversible sequestering of required cofactors (e.g., Mn^{2+}, one or more nucleotides, etc.). Once sufficient time has passed that the polymerase is paused at the blocking group, illumination is reintroduced and the blocking group removed. This allows control of the position on the template nucleic acid at which the polymerase will begin generating nucleotide sequence data during the illuminated period. A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the template nucleic acid that inhibit

polymerase mediated replication, strand-binding moieties that prevent processive synthesis, non-native nucleotides included within the primer and/or the template, and the like. Such reaction stops/pause points are useful in providing more certainty about the relationship of the reads to each other. For example, since the exact position on a template nucleic acid at which each sequence read begins would be known, the resulting reads could be better mapped relative to one another for construction of a sequence scaffold and/or consensus sequence. Further description of these and other methods for regulating the progress of a polymerase on a template are provided, e.g., in U.S.S.N. 61/099,696, U.S. Patent Pub. No. 2006/0160113, and U.S. Patent Pub. No. 2008/0009007, all of which are incorporated by reference herein in their entireties for all purposes.)

[00104] By way of example, a sequencing reaction may be initiated on a template comprising a non-native base in the absence of the complement to the non-native base, which would not impact the overall sequence determination of other portions of the template that are complementary to native bases. By starving the reaction for the complement to the non-native base, one can prohibit synthesis, and thus, the sequencing process, until the non-native base complement is added to the mixture. This can provide a "hot start" capability for the system and/or an internal check on the sequencing process and progress that is configurable to not interfere with sequence analysis of the regions of interest in the template, which would be complementary to only native bases. In some embodiments, the non-native base complement in the sequence mixture is provided with a detectably different label than the complements to the four native bases in the sequence, and the production of incorporation-based signals associated with such labels provides an indication that the polymerase has initiated or reinitiated. Although described as the "non-native base" it will be appreciated that this may comprise a set of non-natural bases that can provide multiple control elements within the template structure. In certain embodiments, two different non-native bases are included within the template structure, but at different points, to regulate procession of the sequencing process, e.g., allowing controlled initiation and a controlled stop/start position later in the sequence, e.g., prior to a subsequent illuminated period. For example, the complement to the first non-native base can be added to initiate sequencing immediately prior to the start of a first illuminated period. During a first non-illuminated period following the first illuminated period, the polymerase encounters the second non-native base, e.g., at a nucleotide position near but upstream of a nucleotide region desired to be sequenced in a second illuminated period. Sequencing would stop until the complement to the second non-native base is added to the reaction mixture. Likewise, multiple such non-native bases could be incorporated into the template to effectively target the

polymerase to multiple regions of interest for which sequence data is desired. Further, in applications in which multiple identical templates are being sequenced, this would allow a resynchronization of the various sequencing reactions and the data generated therefrom.

[00105] Methods of controlling polymerase progress in different stages of a sequencing reaction are also useful for not only creating "condition-dependent" non-detection periods (during which time illumination may or may not be present), but also for minimizing the amount of time required for traversing a given length of template during a non-detection period (whether or not illumination is present). In order to reliably detect incorporation events, non-natural reagent conditions are typically used to limit polymerization during detection periods to approximately 1-5, or about 3 bases per second. In certain embodiments, replacement of Mg^{2-} ions with Mn^{2+} ions serves to stabilize and slow the translocation of the polymerase. When magnesium and, optionally, native nucleotides (e.g., lacking fluorescent labels) are used, the rate of translocation and/or processivity of the polymerase may increase up to two orders of magnitude. Use of such "rapid translocation" conditions during the non-detection periods can provide myriad benefits, including but not limited to a more rapid polymerization rate, an increased processivity (e.g., due to decreased stalling and misincorporation), and an overall savings due to reduced use of expensive labeled nucleotide analogs and/or reagents that mitigate oxidative stress.

[00106] In certain embodiments, a protocol for intermittent detection comprises alternating reaction mixtures, where a first reaction mixture used during the detection periods is optimized for sequence read generation, and a second reaction mixture used during the non-detection periods is optimized for processivity and/or rapid polymerization. For example, when reagents for optimal sequence read generation are present, DNA synthesis rate is low, and there is a fluorescence signal associated with each incorporation event. After replacing the reaction mixture optimized for sequence read generation with the reaction mixture optimized for processivity and/or rapid polymerization, the polymerase rapidly advances across the template. In certain embodiments, a flow cell is used to deliver and switch between the two (or more) reaction mixtures during the course of the reaction.

[00107] In an exemplary embodiment, a first reaction mixture comprises fluorescently-labeled nucleotide analogs and manganese ions that restrict polymerization to a rate appropriate for high fidelity detection of nucleotide incorporation. The first reaction mixture can also include additional agents for mitigation of photo-induced damage of various components of the reaction mixture. A second reaction mixture comprises natural nucleotides and an appropriate magnesium

ion concentration for rapid synthesis of the nascent strand complementary to the template. A first detection period of a sequencing reaction is initiated by introduction of the first reaction mixture, and a sequence read is generated based upon synthesis of the nascent strand during the detection period. After a predetermined time interval a sufficient quantity of the second reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, thereby initiating a first non-detection period. As noted above, the lack of labeled nucleotides in the second reaction mixture alone can produce the non-detection period, since there will be no signal emitted coincident with incorporation of the native nucleotides, but in certain embodiments illumination may also be removed, e.g., to further mitigate photo-induced damage during the non-detection period. At a time appropriate to initiate a second detection period, a sufficient quantity of the first reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, and detection of incorporation event is reinitiated. The cycle of reaction mixture exchange is repeated to generate multiple detection and non-detection periods.

[00108] A flow cell for reaction mixture exchange preferably has two inputs that are gated such that only a single reaction mixture flows into a reaction site or plurality of reaction sites, e.g., on a substrate. A single out-flow line may be used to remove reaction mixtures from the reaction site(s) to a single collection vessel, or multiple collection vessels may be used, one for each type of reaction mixture used. Further, accurate estimation of the distance a polymerase translocates during a non-detection period is important for bioinformatics applications. This estimation is complicated if the time for reaction mixture exchange is slow. As such, the flow is preferably at a sufficient rate that the time for exchange is significantly less than the time spent in the presence of either reaction mixture alone.

[00109] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination. In this embodiment, sixteen sequencing-by-incorporation reactions are performed on single nucleic acid templates (each of which comprises the same nucleotide sequence) with the timing of the illuminated and non-illuminated periods the same for all sixteen reactions. In A, the sixteen reactions are shown disposed on sixteen reaction sites on a solid support and are numbered for convenience. A representation of the illumination data is shown in B, with bars extending across the graph indicative of illumination data collected during illuminated periods for each reaction. In this illustrative example, each reaction is subjected to three illuminated periods, each followed by a non-illuminated period, resulting in three noncontiguous sequence reads

for each reaction, i.e., three noncontiguous reads per template molecule sequenced. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. During the first illuminated period, the sequence reads generally overlap, but the natural variation of polymerase translocation rate over the set of reactions results in a "spreading" of the sequence reads as the reaction proceeds through the second and third illuminated periods with increasing variation in the exact position of each polymerase on the template at the beginning and end of each illuminated period. As such, the earlier illumination data provides better redundancy ("oversampling") of sequence information over a relatively narrow portion of the template nucleic acid, while the later illuminated periods provide less redundant sequencing data over a broader region of the template nucleic acid. The timing of the non-illuminated periods between the illuminated periods and the known or calculated rate of incorporation are used to determine approximate spacing between the resulting sequence reads, providing context for building a sequence scaffold or consensus sequence. It is important to note that although shown disposed on a solid support in A, the data shown in B could also have been generated from reactions not disposed on a solid support nor performed simultaneously and the methods are generally not so limited. Further, as described above, the spreading of the sequence reads from later stages of the reactions can be mitigated by synchronizing the reactions, e.g., by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that provides better redundancy (i.e., more overlap in the positions on the template from which the sequence reads are generated), especially in the later stages of the reactions.

[00110] Using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be reset to change the regions of the template for which nucleotide sequence data is generated. As described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the

template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest. Alternatively, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be kept the same to provide a greater-fold coverage of one or more regions of interest in the template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which are incorporated by reference herein in their entireties for all purposes. A specific embodiment is also provided in the Exemplary Applications section herein.

[00111] Another exemplary embodiment of an analysis of a plurality of illuminated reactions using intermittent illumination comprises a first illuminated period that is initiated at different times over the plurality of reactions. For example, the illuminated period for a first reaction may start at 0 seconds, the illuminated period for a second reaction may start at 5 seconds, the illuminated period for a third reaction may start at 10 seconds, and so forth. Additionally or alternatively, a first subset of reactions may begin at a first time, a second subset may begin at a second time, etc. The first illuminated period continues for a given length of time, followed by a non-illuminated period and a subsequent second illuminated period. Optionally, a plurality of non-illuminated periods and illuminated periods follow the first illuminated period. Staggered start times can provide staggered data sets (e.g., two or more sequence reads) for the plurality of reactions, allowing multiple different stages of the overall reaction to be interrogated in different reactions. Preferably, the staggered data sets overlap to an extent that allows further analysis and validation of the reaction data. For example, a sequencing-by-incorporation reaction subjected to such an embodiment of the invention would preferably have sufficient overlap between sequence reads from different individual reactions to allow construction of a sequence scaffold and/or consensus sequence for a template nucleic acid.

[00112] A mask for use with a solid support (e.g., an array of confinements) can be designed to allow illumination of one or more portions of the solid support while blocking illumination to other portions of the solid support. For example, a mask may comprise one or more windows that allow excitation illumination to pass through the mask. Such a mask may be physically moved over

the surface of the solid support (or the solid support can be moved relative to the mask), e.g., to selectively allow excitation illumination to reach a subset of confinements in an array. For example, a mask that allows 10% of reaction sites to be illuminated could be used to increase the sequencing scaffold coverage by sliding the illumination area (the area being subjected to excitation illumination) back and forth across the solid support. The 10% of reactions would cover certain regions of the nucleic acid template for any given time period (and therefore region of sequence in the template). In certain embodiments, an automated mask that selectively controls the timing of illumination of reactions on a solid support during the course of the reaction/acquisition may be used rather than a mask that must be physically moved.

[00113] The timing of the illuminated and non-illuminated periods for a set of reactions on a solid support may be the same or may vary, and may be synchronized or random. In certain embodiments in which the excitation illumination source is turned on and off, the timing of the illuminated and non-illuminated periods for the set of reactions will be identical. In other embodiments, for example, those that comprise use of a mask, the timing of the illuminated and non-illuminated periods for the set of reactions can vary so that while a subset of the reactions are illuminated, another subset of the reactions are not illuminated. Various exemplary and nonlimiting embodiments of masks that may be used with a set of reactions on a solid substrate are provided in Figures 3-5, as described below. In certain embodiments, the illuminated/non-illuminated status of each reaction may be random across the solid support, e.g., to remove any experimental bias potentially introduced by actively selecting which reactions to illuminate at a given time, as long as the sequence reads being generated at the illuminated reactions and the time at which these reactions are not illuminated are able to be assigned to a particular reaction. For ease of discussion, the action of both illuminating and collecting emission signals from a reaction of interest, or a particular region on a solid support in which a reaction of interest is taking place, is referred to as "interrogating" that reaction and/or that region. A region being so interrogated is termed an "observation region."

[00114] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination and a mask. As in Figure 2, an array of reactions on a solid support 310 is provided containing sixteen reaction sites, numbered for convenience (A). In B, a mask 320 is provided with a single window 330 to allow passage of illumination to a subset of reactions on the solid support. Window 330 is wide enough to allow illumination of at least two columns of reaction sites on solid support 310. As in Figure 2, a representation of the illumination

data is shown in C, with bars extending across the graph indicative of illumination data collected for each reaction. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. When the sequencing reaction is initiated at all positions on solid support 310, the window 330 is positioned to allow illumination to only reactions 1, 5, 9, and 13, and these four reactions provide sequence reads 350 for the earliest stage of the reactions. The window 330 is subsequently moved to provide an illuminated period for reactions 2, 6, 10, and 14 while still continuing the illuminated period for reactions 1, 5, 9, and 13. The illumination data for reactions 2, 6, 10, and 14 provides sequence reads 360, which partially overlap sequence reads 350 for reactions 1, 5, 9, and 13. The window 330 is moved again to provide illuminated periods for reactions 3, 7, 11, and 15 while still continuing the illuminated period for reactions 2, 6, 10, and 14, but removing illumination from reactions 1, 5, 9, and 13. The illumination data for 3, 7, 11, and 15 results in sequence reads 370, which partially overlap sequence reads 360 for reactions 2, 6, 10, and 14. A fourth position of the mask 320 initiates an illuminated period for reactions 4, 8, 12, and 16 while continuing illumination of reactions 3, 7, 11, and 15, but ending the illuminated period for reactions 2, 6, 10, and 14. Sequence reads 380 correspond to sequence reads from reactions 4, 8, 12, and 16. Finally, the window is moved to end the illuminated period for reactions 3, 7, 11, and 15 while continuing the illuminated period for reactions 4, 8, 12, and 16. Repeating the above process allows a second read to be generated from each reaction, and this second read is noncontiguous with the first read. For example, reactions 1, 5, 9, and 13 correspond to reads 350 and, later in the reaction, reads 355. The two reads generated in a single reaction do not overlap and are separated by a length of nucleotides that was incorporated during the non-illuminated period between the two illuminated periods.

[00115] The mask can optionally be passed over the substrate additional times to generate additional reads until the reactions are complete or no longer provide reliable data, such as when the total illumination time (computed by summing the times for the multiple illuminated periods) has surpassed a photo-induced damage threshold period. Further, the mask may be passed back and forth, or may pass over the solid support in only one direction, e.g., always left to right, or vice versa.

[00116] Further, unlike the data shown in Figure 2B which has gaps in the sequence coverage for the template nucleic acid, the strategy provided in this embodiment results in at least two-fold coverage across the entire template nucleic acid (Figure 3C), although at a lower-fold redundancy.

The portion of the template covered by only reads 380 and reads 355 has the least-fold redundancy, and in some instances a gap in coverage may be present in this region due to the movement of the mask 320 from the far right to the far left of the solid support 310. Of course, oversampling by adding replicate reactions to the set of reactions, or using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the coverage of a nucleic acid template, thereby providing more data for construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which have been previously incorporated by reference herein. The natural variation of polymerase translocation rate over the set of reactions is also apparent in this prophetic example as the spreading of the sequence reads and decreasing overlap between reads from reactions in adjacent columns in the later stages of the reactions as compared to the earlier stages.

[00117] Figure 4A provides an embodiment of a mask similar to that provided in Figure 3 except that it comprises three windows allowing multiple nonadjacent columns of reaction sites to be illuminated simultaneously. Figure 4B provides an embodiment of a mask comprising twelve windows, each of which allows illumination of a single reaction site on a solid support. The windows are oriented in the mask to allow illumination of every other reaction in each row and every other reaction in each column. It will be understood that these mask designs are merely exemplary and nonlimiting embodiments as it is well within the abilities of the ordinary practitioner to determine an appropriate mask design depending on the experimental design or the illuminated reactions to be interrogated.

[00118] Figure 5B illustrates yet another aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Four different samples are disposed on a solid support, one in each quadrant 510, 520, 530, and 540 (A). A mask 550 is used that comprises two windows 560 that allow multiple rows of reaction sites to be illuminated simultaneously (B). A first position of this mask over a solid support in which two reactions in each quadrant are illuminated is demonstrated in C. A second position of the mask allowing illumination of the previously non-illuminated reactions is demonstrated in D. The mask may be moved back and forth as indicated by the double-arrow to provide multiple illuminated and non-illuminated periods for each reaction containing one of the four samples.

[00119] The present invention is also useful for redundant interrogation of reactions or portions of a solid support of interest. In certain aspects, sequential interrogation of different observation regions may be repeated a number of times, e.g., more than 2, 5, 10, 50, 100, 500, 1000, or even more than 10,000 times. In general, this method of stepping the observation region to another, preferably adjacent region, and repeating the interrogation process is generally referred to as a "step and repeat" process, and may be performed by various methods, including but not limited to moving the incident light and the solid support relative to one another and moving a mask across the surface of the solid support, as described above. Although described as a "step and repeat" method, in some embodiments where the observation region is moved across a substrate, that movement is not step-wise and iterative, but instead constitutes a continuous motion, substantially continuous motion, or stepped movement, or an iterative motion whereby each iterative step interrogates a new region that overlaps with some portion of the previously interrogated region. In particular, a substrate may be moved continuously relative to an optical system, whereby the observation region moves continuously across the substrate being interrogated (in a "scan mode").

[00120] The present invention is optionally combined with an optical system that provides illumination and/or collection of emitted illumination. Preferably, the optical system is operatively coupled to the reaction sites, e.g., on a solid support. One example of a particularly preferred optical system is described in U.S.SN. 11/201,768, filed August 11, 2005, and incorporated herein by reference in its entirety for all purposes. Optical systems are described further below.

[00121] In some embodiments, one or both of the solid support and optical system are moved during interrogation. For example, a solid support being interrogated may be held stationary while the optical system is moved, or the solid support may be moved relative to a stationary optical system. Such movement may be accomplished using any of a variety of manipulation hardware or robotic set-ups, e.g., a stepper/feeder apparatus, and are well known in high performance printing technologies and in the semiconductor industry. For example, robotic systems may be used to pick up and re-orient a given solid support in order to interrogate different regions of the solid support, or make a previously unaccessible region (e.g., blocked by clips, support structure, or the like) of the solid support accessible. Such robotic systems are generally available from, e.g., Beckman, Inc., Tecan, Inc., Caliper Life Sciences, and the like.

[00122] In addition to the foregoing, it will be appreciated that the reagents in a given reaction of interest, including those reagents for which photo-induced damage is being mitigated in accordance with the invention, may be provided in any of a variety of different configurations. For

example, they may be provided free in solution, or complexed with other materials, e.g., other reagents and/or solid supports. Likewise, such reagents may be provided coupled to beads, particles, nanocrystals or other nanoparticles, or they may be tethered to larger solid supports, such as matrices or planar surfaces. These reagents may be further coupled or complexed together with other reagents, or as separate reagent populations or even as individual molecules, e.g., that are detectably resolvable from other molecules within the reaction space. As noted above, whether a particular reagent is confined by virtue of structural barriers to its free movement or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined." Further examples of such confined reagents include surface immobilized or localized reagents, e.g., surface immobilized or associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflection microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. For example, in some preferred embodiments, one or more reagents in an assay system are confined within an optical confinement. Such an optical confinement may be an internal reflection confinement (IRC) or an external reflection confinement (ERC), a zero-mode waveguide, or an alternative optical structure, such as one comprising porous film with reflective index media or a confinement using index matching solids. More detailed descriptions of various types of optical confinements are provided, e.g., in International Application Publication No. WO/2006/083751, incorporated herein by reference in its entirety for all purposes.

[00123] The invention is generally applicable to any of a variety of optical assays that require substantial illumination and/or photoactivated conversion or excitation of chemical groups, e.g., fluorophores. For example, the compositions and methods provided herein may be used with fluorescence microscopy, optical traps and tweezers, spectrophotometry, fluorescence correlation spectroscopy, confocal microscopy, near-field optical methods, fluorescence resonance energy transfer (FRET), structured illumination microscopy, total internal reflection fluorescence microscopy (TIRF), etc. The methods provided herein may be particularly useful in assays that are negatively impacted, directly or indirectly, by prolonged exposure to illumination. Of particular interest are those assays that are impaired by the generation and/or accumulation of triplet-state forms or free radicals during illumination.

[00124] One particularly apt example of analyses that benefit from the invention are single-molecule biological analyses, including, inter alia, single molecule nucleic acid sequencing analyses, single molecule enzyme analyses, hybridization assays (e.g., antibody assays), nucleic acid hybridization assays, and the like, where the reagents of primary import are subjected to prolonged illumination with relatively concentrated light sources (e.g., lasers and other concentrated light sources, such as mercury, xenon, halogen, or other lamps) in an environment where photoconversion/excitation is occurring with its associated generation of products. In certain embodiments, the methods, compositions, and systems are used in nucleic acid sequencing processes that rely on detection of fluorescent or fluorogenic reagents. Examples of such sequencing technologies include, for example, SMRT™ nucleic acid sequencing (described in, e.g., U.S. Patent Nos. 6,399,335, 6,056,661, 7,052,847, 7,033,764, 7,056,676, 7,361,466, 7,416,844, the full disclosures of which are incorporated herein by reference in their entirety for all purposes), non-real-time, or "one base at a time" sequencing methods available from, e.g., Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), Clonal Single Molecule Array™, and SOLiD™ sequencing. (See, e.g., Harris, et al. (2008) Science 320 (5872):106-9, incorporated by reference herein in its entirety for all purposes.) Such prolonged illumination can negatively impact (e.g., by introducing photo-induced damage) these reagents and diminish their effectiveness in the desired reaction.

III. Prevention of Photo-induced Damage

[00125] The methods provided herein are particularly useful in analyses that utilize very limited concentrations of reactants, such as single molecule detection/monitoring assays. As will be appreciated, in such reagent limited analyses, any loss, degradation, or depletion of a critical reagent will dramatically impact the analysis by further limiting the reagent, which not only can adversely effect the detectable signal, but may also directly impact the reaction being monitored, e.g., by changing its rate, duration, or product(s). For example, photo-induced damage can include a photoinduced change in a given reagent that reduces the reactivity of that reagent in the reaction, e.g., photobleaching of a fluorescent molecule, which diminishes or removes its ability to act as a signaling molecule. Also included in the term photo-induced damage are other changes that reduce a reactant's usefulness in a reaction, e.g., by making the reagent less specific in its activity in the reaction. Likewise, photo-induced damage includes undesired changes in a reagent that are caused by interaction of that reagent with a product of another photoinduced reaction, e.g., the generation

of singlet oxygen during a fluorescence excitation event, which singlet oxygen may damage organic or other reagents, e.g., proteins. Photo-induced damage also includes downstream effects of damage to reactants, such as irreversible interactions between damaged reactants and other critical components of the reaction, e.g., reactive proteins or enzymes. For example, damage to an enzyme that catalyzes a reaction being monitored may cause a reduction in the rate of the reaction, in some cases stopping it altogether, or may reduce the duration or fidelity of the reaction.

[00126] As suggested by the foregoing, photo-induced damage generally refers to an alteration in a given reagent, reactant, or the like, that causes such reagent to have altered functionality in a desired reaction, e.g., reduced activity, reduced specificity, or a reduced ability to be acted upon, converted, or modified, by another molecule, that results from, either directly or indirectly, a photo-induced reaction, e.g., a photo-induced reaction creates a reactant that interacts with and causes damage to one or more other reactants. Typically, such photoreaction directly impacts either the reactant of interest, e.g., direct photo-induced damage, or impacts a reactant within one, two or three reactive steps of such reactant of interest. Further, such photoreaction can directly impact the reaction of interest, e.g., causing a change in rate, duration, processivity, or fidelity of the reaction.

[00127] The amount of time an illuminated analysis may be carried out before photo-induced damage so substantially impacts the reactants to render the analysis non-useful is referred to as the "photo-induced damage threshold period." A photo-induced damage threshold period is assay-dependent, and is affected by various factors, including but not limited to characteristics of enzymes in the assay (e.g., susceptibility to photo-induced damage and the effect of such damage on enzyme activity/processivity), characteristics of the radiation source (e.g., wavelength, intensity), characteristics of the signal-generating molecule (e.g., type of emission, susceptibility to photo-induced damage, propensity to enter triplet state, and the effect of such damage on the brightness/duration of the signal), similar characteristics of other components of the assay. It can also depend on various components of the assay system, e.g., signal transmission and detection, data collection and analysis procedures, etc. It is well within the abilities of the ordinary practitioner to determine an acceptable photo-induced damage threshold period for a given assay, e.g., by monitoring the signal decay for the assay in the presence of a photodamaging agent and identifying a period for which the signal is a reliable measure for the assay. In terms of the invention, the photo-induced damage threshold period is that period of illuminated analysis during which such photo-induced damage occurs so as to reduce the rate or processivity of the subject reaction by at least

10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over the same reaction in the absence of such illumination. It is an object of the invention to increase the photo-induced damage threshold period, thereby increasing the amount of time reactions can proceed toward completion with minimal damage to the reactants, thereby lengthening the time in which the detectable signal is an accurate measure of reaction progression.

[00128] In some contexts, a "photo-induced damaged" reaction may be subject to spurious activity, and thus be more active than desired. In such cases, it will be appreciated that the photo-induced damage threshold period of interest would be characterized by that period of illuminated analysis during which such spurious activity, e.g., as measured by an increase in reaction rate, or an increase in non-specific reaction rate, is no more than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over a non-illuminated reaction. In one non-limiting example, where a nucleic acid polymerase, by virtue of a photodamaging event, begins to incorrectly incorporate nucleotides during template directed synthesis, such activity would impact the photo-induced damage threshold period as set forth above. In this case, the methods, devices, and systems of the invention would increase the photo-induced damage threshold period, thus increasing the amount of time the reaction could proceed before the above-described spurious activity occurred.

[00129] With reference to nucleic acid analyses, it has been observed that in template-directed synthesis of nucleic acids using fluorescent nucleotide analogs as a substrate, prolonged illumination can result in a substantial degradation in the ability of the polymerase to synthesize the nascent strand of DNA, as described previously, e.g., in U.S. Published Patent Application No. 20070161017, incorporated by reference herein in its entirety for all purposes. Damage to polymerase enzymes, template sequences, and/or primer sequences can significantly hinder the ability of the polymerase to process longer strands of nucleic acids. For example, reduction in the processivity of a polymerase leads to a reduction in read lengths for sequencing processes that identify sequence constituents based upon their incorporation into the nascent strand. As is appreciated in the art of genetic analysis, the length of contiguous reads of sequence directly impacts the ability to assemble genomic information from segments of genomic DNA. Such a reduction in the activity of an enzyme can have significant effects on many different kinds of reactions in addition to sequencing reactions, such as ligations, cleavages, digestions, phosphorylations, etc.

[00130] Without being bound to a particular theory or mechanism of operation, it is believed that at least one cause of photo-induced damage to enzyme activity, particularly in the presence of

fluorescent reagents, results from the direct interaction of the enzyme with photo-induced damaged fluorescent reagents. Further, it is believed that this photo-induced damage of the fluorescent reagents (and possibly additional damage to the enzyme) is at least partially mediated by reactive intermediates (e.g., reactive oxygen species) that are generated during the relaxation of triplet-state fluorophores. One or both of the photo-induced damaged fluorescent reagents and/or reactive intermediates may be included in the overall detrimental effects of photo-induced damage.

[00131] In certain aspects, the invention is directed to methods, devices, and systems that reduce the amount of photo-induced damage to one or more reactants during an illuminated reaction, e.g., thereby improving the reaction, e.g., by increasing the processivity, rate, fidelity, processivity, or duration of the reaction. In particular, methods are provided that yield a reduction in the level of photo-induced damage and/or an increase in the photo-induced damage threshold period as compared to such reactions in the absence of such methods, devices, and systems. In particular embodiments, such methods comprise subjecting an illuminated reaction to periods of non-illuminated during the course of the reaction, as described above, or by temporarily removing components of the reaction mixture that are believed to cause such damage, as described below.

[00132] As generally referred to herein, limited quantity reagents or reactants may be present in solution, but at very limited concentrations, e.g., less than 200 nM, in some cases less than 10 nM and in still other cases, less than 10 pM. In preferred aspects, however, such limited quantity reagents or reactants refer to reactants that are immobilized or otherwise confined within a given area or reaction site (e.g., a zero-mode waveguide), so as to provide limited quantity of reagents in that given area, and in certain cases, provide small numbers of molecules of such reagents within that given area, e.g., from 1 to 1000 individual molecules, preferably between 1 and 10 molecules. As will be appreciated, photo-induced damage of immobilized reactants in a given area will have a substantial impact on the reactivity of that area, as other, non-damaged reactants are not free to diffuse into and mask the effects of such damage. Examples of immobilized reactants include surface-immobilized or -localized reagents, e.g., surface-immobilized or -associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflectance microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface

imaging, or the like. Various types of solid supports upon which one or more reactants can be immobilized are described above.

[00133] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is interrogated for one or more illuminated periods that cumulatively are less than a photo-induced damage threshold period, as set forth elsewhere herein. Such interrogation may occur coincident with or independent of interrogation of additional observation regions on a solid support containing the first observation region. In accordance with the present invention, the observation region typically includes confined reagents (e.g., enzymes, substrates, etc.) that are susceptible to photo-induced damage, and may include an area of a planar or other solid support upon which confined reagents are immobilized. Alternatively or additionally, the observation region may include a physical confinement that constrains the reagents that are susceptible to photo-induced damage, including, e.g., microwells, nanowells, planar surfaces that include hydrophobic barriers to confine reagents.

[00134] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is intermittently interrogated under constant illumination by virtue of intermittent presence of detectable components of the reaction, wherein the presence of such detectable components has the potential to directly or indirectly cause photo-induced damage to one or more other reaction components. For example, a buffer comprising detectable components of a reaction can be temporarily replaced with a buffer comprising non-detectable versions of the same components of the reaction, thereby interrupting data acquisition for the reaction. When data acquisition is to be recommenced, the buffer comprising detectable component is substituted for the buffer comprising non-detectable components. This substitution of reaction components may be repeated multiple times to generate multiple sets of data collected at noncontiguous stages of the reaction. For example, such a substitution can occur at least about 2, 4, 6, 8, or 10 times during the course of the reaction.

[00135] In certain preferred embodiments, the detectable components are fluorescently-labeled components that can be damaged by exposure to excitation illumination, and can further cause damage to other reaction components, as described above. For example, a sequencing-by-incorporation reaction can be initiated in the presence of fluorescently-labeled nucleotides whose incorporation is indicative of the nucleotide sequence of the nascent strand synthesized by a polymerase, and by complementarity, of the template nucleic acid molecule. At a selected time point during the ongoing reaction, the labeled nucleotides can be removed and replaced with

unlabeled nucleotides, for example, by buffer exchange. After a period of time during which data acquisition has been interrupted by the absence of signal from the ongoing reaction, the labeled nucleotides can be reintroduced to reinitiate data acquisition. The labeled nucleotides may be removed and reintroduced multiple times and for various lengths of time, as preferred by the ordinary practitioner. In this way, multiple noncontiguous sequence reads can be generated from a single nucleic acid molecule in real time.

[00136] The methods herein slow the accumulation of photo-induced damage to one or more reagents, and may therefore indirectly mitigate the impact of photo-induced damage in an ongoing reaction of interest. By way of example, methods that reduce exposure of a critical enzyme component to illumination radiation (e.g., by subjecting the reaction to periods of non-illumination or by temporarily removing a component of the reaction responsible for such damage) do not necessarily prevent the photo-induced damage to the enzyme component, but rather extend the photo-induced damage threshold period by slowing the accumulation of photo-induced damage in the reaction mixture. Measurements of reduction of photo-induced damage as a result of implementation of intermittent illumination may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction subjected to constant illumination. Likewise, measurements of reduction of photo-induced damage as a result of temporary removal of reaction components responsible for such damage may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction in which such components are present throughout. Further, characterization of a reduction in photo-induced damage generally utilizes a comparison of reaction rates, durations, or fidelities, processivities, e.g., of enzyme activity, and/or a comparison of the photo-induced damage threshold period, between a reaction mixture subjected to such the methods and/or systems of the invention and a reaction mixture not so subjected.

[00137] In the case of the present invention, implementation of the methods, devices, and systems of the invention generally results in a reduction of photo-induced damage of one or more reactants in a given reaction, as measured in terms of "prevented loss of reactivity" in the system. Using methods known in the art, the amount of prevented loss of activity can at least 10%, preferably greater than 20%, 30%, or 40%, and more preferably at least 50% reduction in loss of reactivity or increase in processivity, and in many cases greater than a 90% and up to and greater than 99% reduction in loss of reactivity or increase in processivity. By way of illustration, and purely for the purpose of example, when referring to reduction in photo-induced damage as a measure of enzyme activity in the presence and absence of intermittent illumination, if a reaction

included a reaction mixture having 100 units of enzyme activity that would, under constant illumination, yield a reaction mixture having only 50 units of activity, then a 10% reduction in photo-induced damage would yield a final reaction mixture of 55 units (e.g., 10% of the 50 units otherwise lost, would no longer be lost). Further, use of the invention is expected to increase the performance (e.g., processivity, duration, fidelity, rate, etc.) of a reaction whose performance is negatively impacted by constant exposure to illumination by at least about 2-, 5-, 10-, 20-, 30-, 50-, 80-, 100-, 500-, or 1000-fold over that achieved by the reaction under constant illumination. For example, it is a specific object of the instant invention to increase the processivity of a polymerase enzyme in a sequencing reaction to allow collection of data across a longer length of the template.

[00138] With regards to sequencing applications, the methods herein facilitate the scaffolding of nucleic acid sequences in reactions susceptible to photo-induced damage. For example, if the sequencing device has 1000 base pair average readlength under constant illumination, one could subject the reaction to illuminated periods timed to allow approximately 100 nucleotides to be incorporated into the nascent strand of read, followed by non-illuminated periods timed to allow approximately 1000 nucleotides to be incorporated "in the dark." The sequence reads resulting from this experimental design would comprise about ten sequence reads of about 100 nucleotides each separated by gaps of about 1000 nucleotides each. If a plurality of sequencing reactions were carried out in this manner, and the illuminated periods were staggered appropriately, the reads from the plurality of reactions could be combined to provide nucleotide sequence data for the entire template nucleic acid. This would potentially allow sequence scaffolds to be built much more easily than can be done with short-read systems, enabling structural analysis of previously impossible-to-sequence sections of highly repetitive DNA, given the sequencing system is capable of long reads in the absence of photodamage.

IV. Software and Algorithm Implementations

[00139] The methods herein may operate with numerous methods for sequence alignment including those generated by various types of known multiple sequence alignment (MSA) algorithms. For example, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read using a reference sequence. In some embodiments in which a reference sequence is known for the region containing the target sequence, the reference sequence can be used to produce an MSA using a variant of the center-star algorithm. Alternatively, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each

read relative to every other read without using a reference sequence ("*de novo* assembly routines"), e.g., PHRAP, CAP, ClustalW, T-Coffee, AMOS make-consensus, or other dynamic programming MSAs. Depending on the sequence-generating methods used, the determination of sequence alignment may also involve analysis of read quality (e.g., using TraceTuner™, Phred, etc.), signal intensity, peak data (e.g., height, width, proximity to neighboring peak(s), etc.), information indicative of the orientation of the read (e.g., 5'→3' designations), clear range identifiers indicative of the usable range of calls in the sequence, and the like. Additional algorithms and systems for sequence alignment are well know to those of skill in the art, and are described further, e.g., in G. A. Churchill, M. S. Waterman (1992) "The Accuracy of DNA Sequences: Estimating Sequence Quality," *Genomics* 14: 89-98; M. Stephens, et al. (2006) "Automating sequence-based detection and genotyping of SNPs from diploid samples," *Nat. Genet.*, **38**: 375-381; J. Hein (1989) *Mol. Biol. Evol.*, **6**: 649-668; U.S.S.N. 12/134,186, filed June 5, 2008; and U.S.S.N. 61/116,439, filed November 20, 2008.

[00140] A standard sequence alignment problem in the context of DNA sequencing is to align the sequence of a relatively short fragment (<2 kilobases) to a large target sequence. The assumption is made that this fragment represents a contiguous portion of DNA to be mapped to a single location on the reference sequence. (A "contiguous portion" to be mapped to a single location may contain small insertions and/or deletions and still be considered contiguous in this context.) With the further development of nucleic acid sequencing technologies (e.g., from Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), and Applied Biosystems, Inc. (Foster City, CA)) and mate-pair sequencing protocols (see, e.g., U.S. Patent Pub. No. 2006/0292611 A1, which is incorporated by reference herein in its entirety for all purposes), the alignment problem has been extended to align two fragments coming from the same read to the reference sequence using some knowledge of the expected mate-pair configuration (distance and orientation).

[00141] With regards to mate-paired reads, mapping two fragments with a distance constraint and orientation constraint has been treated by various short-read mapping algorithms, e.g., SOAP (Li, et al. (2008) Bioinformatics, **24**, 713-714); SOAPdenovo; and Maq, a set of programs that map and/or assemble fixed-length Solexa/SOLiD reads (SourceForge, Inc.). While these algorithms can handle simple cases of mate-pair alignment, which generally treat the specific problem of only two reads coming from a mate-paired sequence and use the distance constraint as a hard filter (i.e., if two reads are within *x* bp of each other and in the correct orientations, report them as a mate-pair hit), the methods provided herein are more general and can handle much more complex data sets,

including those with multiple reads, those for which a reference sequence is or is not present, potential non-template sequence (e.g., adapter regions or linker portions described below), and complex distance and orientation constraints. Other programs are also available that attempt to generalize on top of the mapping and aligning performed by the programs described above. These include, e.g., Breakdancer, variationhunter, GASV, etc., which can handle more complex mappings, e.g., by clustering.

[00142] Real-time single molecule sequencing presents opportunities for obtaining much more complex sequence fragments from a single DNA sequencing read. Two examples are the reading of multiple discontiguous sequence fragments from a single long stretch of DNA using a pulsed or intermittent detection system (e.g., intermittent illumination) as described herein and the contiguous reading of forward, reverse and adapter fragments from a circular templates (SMRTbell™ templates; see e.g., U.S.S.N. 61/099,696, filed September 24, 2008; U.S. Patent Application No. 12/383,855, filed March 27, 2009 and U.S. Patent Application No. 12/413,258, filed March 27, 2009, all of which are incorporated by reference herein in their entireties for all purposes). Further, methods for sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.

[00143] Certain aspects of the invention provide methods for optimally aligning such sequences to a reference sequence using knowledge of the molecular configuration and/or sequencing protocol used to generate the related sequence reads. In particular, methods are provided to address the general problem of mapping multiple fragments to a reference sequence with variable distance and orientation constraints.

[00144] Beginning with raw sequence data generated by a nucleic acid sequencing instrument (step 1), the sequence data is mapped to a target sequence (step 2) using a local alignment method which produces sub-optimal local alignments as well as the optimal alignment, for example, the Smith-Waterman algorithm. Another, more flexible example of a local alignment method is a chaining method using a method for aligning very short fragments to the target sequence (e.g., kmer-indexing, suffix trees, suffix arrays, etc.) and chaining the resulting hits back into longer chains of significant matches (see, e.g. D. Gusfield, Algorithms on Strings, Trees, and Sequences,

Cambridge University Press: Cambridge, UK, 1997, which is incorporated by reference herein in its entirety for all purposes). The chains do not necessarily need to be refined by dynamic programming in order to be useful for the following algorithm, permitting a very fast algorithm. In certain embodiments, dynamic-programming refinement of the chain might improve the power (area under the ROC curve) of the algorithm.

[00145] The target sequence consists of the potential hypotheses for the molecular template in question. In the example of nucleic acid sequencing methods using iterative illumination for sequencing a shotgun fragment from a linear DNA sequence, the potential hypotheses are both orientations of the genome (since we do not know the original orientation of the fragment). In the example of sequencing of a SMRTbell™ template (e.g., see Example 1 herein), the hypotheses include both orientations of the genome and known adapter sequences. The parameters determining how many hits are reported for each local fragment can be varied to change the specificity and sensitivity of this algorithm. Figure 6 shows what these hits might look like for a SMRTbell™ template (represented as paths in the sequence alignment matrix, which is often called the dynamic-programming matrix, although it isn't necessary to use dynamic programming to find these paths).

[00146] After the potential local alignments have been enumerated, a weighted directed graph is constructed with each local alignment represented as a node in the graph (step 3). The edges are drawn between nodes if they represent a potential reconstruction of the original molecular template using knowledge of the expected molecular configuration. The directed connection of an alignment path *A* to an alignment path *B* is interpreted as "The target sequence represented by *B* could follow the target sequence represented by *A* in the original molecule." For example, if a linear single-stranded DNA molecule is being sequenced by a method that uses iterative illumination, then fragments from opposite orientations would not be expected to be connected (unless the linear single-stranded DNA molecule also included oppositely oriented sequences, e.g., as in the case of a linearized SMRTbell™ template.) In general, fragments that represent the same stretch of the sequencing read but that align to different regions on the target sequence would not be connected. Aside from these examples, the rules for connecting nodes should be fairly loose to permit exploration of weak possibilities that gain significance when all the evidence (e.g. all the sequence reads) are considered. The assignment of edge weights handles the proper weighting of the likelihood of these edges, and the speed of the algorithm can be tuned by optimizing the pruning of highly unlikely edges. As usual this represents a tradeoff between speed and sensitivity.

[00147] Weights are assigned to connections ($A \rightarrow B$) in the graph representing the log-likelihood that target fragment A is followed by target fragment B in the original molecule.

$$w(A \rightarrow B) = -\log P(B|A)$$

The conditional probability $P(B|A)$ encodes the knowledge of the possible molecular configurations and the alignment significance of B.

$$P(B|A) = f(B)g(A,B)$$

where f is a measure of alignment significance (either theoretical or empirically obtained) and g encodes the physical constraints representing the allowed molecular configurations.

[00148] For example, in the context of sequencing using iterative illumination the following may be known: the time between the end of one fragment and the beginning of the next fragment is 200 seconds. If the polymerase incorporates bases with an average rate of 4bp/sec with a standard deviation of 1bp/sec, it can be hypothesized that the probability of target fragment 2 following target fragment 1 is determined by the distance between these fragments on the target and a normal probability:

$$g(A,B) = \frac{1}{\sqrt{2\pi(200)}} \exp\left[-(d-800)/2(200)^2\right]$$

[00149] In a SMRTbell™ template example, knowledge of the expected insert size and the observed distance and orientation between fragments would be used to weight the likelihood that these two fragments could come from a correctly generated SMRTbell™ template. This weight could include the expected rate of the polymerase as well and rules for the orientation of fragments with respect to each other and their distance apart in the original read. For example, while it may be expected that two forward fragments mapping to the same region in the target genome potentially come from multiple passes around a SMRTbell™ template molecule, those fragments would not be expected to be immediately adjacent in sequencing time. The weighting function would account for the proper amount of expected time between such fragments (i.e. the elapsed time would be expected to be long enough to include two adapter sequences and a reverse sequence).

[00150] In general, the weighting function could be arbitrarily complex and tuned to empirically observed relationships between sequencing fragments given the available knowledge (distance between fragments on the target sequence, sequencing time between fragments, expected length of the template, etc.). For example, the empirical probability distributions might be observed to exhibit longer tails than a Gaussian probability model might predict. The use of a conditional log-

likelihood for the assignment of edge weights is motivated by the following logic. In a graph of possible local alignments it is desirable to find a highly likely path that best explains the observed data. Consider a path through three nodes *A*, *B*, and *C*, with *P(ABC)* being the probability that *ABC* is the correct assignment:

$$P(ABC) = P(C|AB)P(B|A)P(A)$$
$$\approx P(C|B)P(B|A)P(A)$$

where the last approximation is justified by the observation that the constraints between allowable assignments to the target sequence are typically local in nature. Generalizing this formula for a path $a_1,\ldots,a_N$ and taking the negative logarithm of both sides gives

$$-\log P(a_1 \ldots a_N) = -\sum_{i=1}^{N-1} \log P(a_{i+1}|a_i) - \log P(a_1)$$

[00151] It is apparent that the edge weights are additive if we use log-likelihood and we can use standard shortest-path algorithms for directed graphs to find the optimal path. A hypothetical directed graph is illustrated in Figure 7. This graph corresponds to the situation depicted by the alignments pictured above. Heavier lines correspond to more likely paths with the optimal path shown in blue. Dashed lines represent forbidden transitions. Not all paths are considered in the illustration to avoid clutter in the presentation. The general formula listed above includes a "one-body" term $P(a_1)$ for the starting node in each path that weights the probability that this initial alignment is correct. To accommodate this probability in a path-finding algorithm we add a pseudo-source *s* to the graph which connects to every possible node (not shown in the graph above). The edge weight connecting the pseudo-source with a node a_i is $-\log P(a_i)$. This allows the use of a conventional single-source shortest-path algorithm starting from the pseudo-source. The desired probability $P(a_i)$ can come from a measure of alignment significance (theoretical or empirically determined) or could be set uniformly across all alignments to allow the path logic to determine the best path assignment, independent of the relative value of the starting points. It is anticipated that a threshold will be required here to only allow edges between the pseudo-source and nodes for highly likely alignments; otherwise the shortest path algorithm in the next step will not give the desired path.

[00152] After construction of the weighted directed graph, the shortest path to each node is determined (step 4). The graph is directed and acyclic (DAG) so we can use the standard shortest-path DAG algorithm (see T.H. Cormen, CE Leiserson, RL Rivest, Introduction to Algorithms, MIT Press: Cambridge, Massachusetts, 1990). This algorithm scales as O(V+E) and should be very

quick for these graphs. After the shortest path to each node is determined, the paths need to be ranked to declare the best assignment. It is suggested that the best metric would be a measure which rewards paths that explain more of the sequenced read (longer paths) with high likelihood. One such metric would be the normalized negative log-likelihood: dividing the total weight of the path by the number of bases in the sequenced read explained by this path. For more complicated graphs or edge-weight assignments, Dijkstra's algorithm, the Bellman-Ford algorithm, or the A* algorithm could be applied. Other algorithms that may also be used include, but are not limited to the Floyd-Warshall algorithm.

[00153] For noisy sequence data it is likely that the local alignments found in step 2 will occasionally overlap with each other in the sequenced read even though it is physically impossible for such overlaps to occur in a perfect system (unless there has been a rearrangement relative to the reference genome). As such, some amount of slack must be allowed in the edge assignment logic in step 3 to account for not knowing the precise boundaries of each local alignment. Once the best physical model explaining the observed read is determined, the boundaries of the local alignments can be refined to reflect the physical necessity that each base in the sequenced read can only be represented in one local alignment. It is also desirable to explain all of the bases in between the local alignments that haven't been assigned in the graph. One straightforward approach to refinement would be to construct the perfect model of the sequence and to realign the sequenced read to this sequence. This refinement algorithm would preserve physical constraints (each base in the sequenced read can only be explained by one location in the template) and would assign all bases between the extremal nodes in the optimal path.

[00154] Certain aspects of the software and algorithm implementations described herein may be varied or altered without departing from the spirit and scope of the invention. For example, with regards to algorithm seeding, many algorithms can be applied for the original determination of sub-optimal local alignments (step 2). Conventional examples include FASTA, BLAST, or Smith-Waterman. It is expected that the best benefit will be obtained from using short-sequence alignment algorithms (suffix array, suffix tree, Boyer-Moore, Rabin-Karp, kmer-indexing, and the like) followed by chaining to establish regions of significant matches. An advantage of the algorithm described here is that it does not require dynamic-programming refinement of the resulting chains and therefore can be quite fast, however it is expected that using dynamic programming to refine the chains in step 2 could increase the power of the algorithm.

[00155] With regards to graph construction, there will be advantages to tuning the logic of edge assignments to keep the size of the graph manageable. It is possible that steps 2 and 3 might be combined to in a greedy fashion to focus the potentially slow step 2 into productive areas of the graph. For example, if a particularly strong hit is found early in step 2, then it may be beneficial to search for sub-optimal hits only in this local vicinity, knowing that this strong hit should be in the final solution. Tuning of the graph construction might include thresholds, below which edges are not created. Further, there are multiple parameters (minimum chain length, minimum probability for edge assignment, relative weighting of length vs. accuracy, etc.) which can be exposed and tuned in this algorithm to maximize the sensitivity and specificity of the algorithm for a given scenario.

[00156] With regards to determination of the distance a polymerase travels between reads, various strategies are provided that are more sophisticated than estimation based upon the rate of incorporation and the time between detection (e.g., illuminated) periods. In certain embodiments, the distribution of the base pair distance travelled by a polymerase during a non-detection period is called $p(x)$. The distribution of enzyme velocities, $p(v)$, is estimated by aligning observed reads to a reference sequence, and this distribution is represented as the number of reference bases per unit time. There is a length of time, τ, over which measurement of the instantaneous rate is not independent. While this method of determining the distance the polymerase travels during a non-detection period should not be overly sensitive to non-independent estimation of the polymerase rate, it is likely to strive for independent measurements of the rate. The distance τ can be estimated from an exponential fit to the auto-correlation function $\langle \delta v(t) \delta v(t - \Delta t) \rangle$, and $v(t)$ tabulated across the aligned sequence at increments of τ.

[00157] Where multiple single polymerase enzymes are being observed simultaneously, e.g., each being optically resolvable from every other on a single array, the $p(v)$ for each is preferably determined independently for each enzyme. Further, information regarding rare but extended events, such as polymerase "stalling" on the template, can be measured across a larger data set. For example, the statistics of stalls can be determined by aggregating rate measurements across an entire array. Where a stall distribution is characterized by a "long tail" corresponding to multi-exponential behavior of IPD distribution, such a distribution of polymerization rates can be extended for stalls longer than the observed reaction by fitting the long-tail behavior to an appropriate functional form, e.g., using a single-exponential parametric model or other physically motivated model (e.g., multi-exponential, stretched exponential, power-law, etc.) In certain preferred embodiments, the following representation of a "per-enzyme" $p(v)$ is used:

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}$$

where *f(v)* is an interpolating function designed to retain information about the zero-velocity tail of the global *p(v)* distribution while taking the estimate of the polymerase velocity dynamics(e.g., the dominant high velocity mode) from the specific enzyme. Such an interpolating function is:

$$f(v) = \frac{1 + erf(v / v_0)}{2}$$

where v_0 is a scale parameter to be chosen based on experience (but optionally fixed). Alternatively, the average of the empirical $p_{enzyme}(v)$ and $p_{array}(v)$ can be used. This approach can be motivated by a Bayesian approach to density estimation. Other kernel density and Bayesian methods can be suggested. Alternatively or in addition, the robustness of *p(v)* to conditions and daily phenomenon can be explored and used to estimate *p(v)* more globally, e.g., using one or more weekly control experiments.

[00158] Given the lack of a known reference sequence for *de novo* assembly, several alternative ways to formulate *p(v)* are provided as follows. For example, in a first embodiment a control template (essentially a proxy reference sequence) can be subjected to sequencing, e.g., in the same reaction as the *de novo* sample or in an identical reaction. The observed velocity for the sequencing reactions would be measured based upon alignments of the reads from the control template to its known sequence. Typically, a per-enzyme correction would not be available for the *p(v)* and *p(v)* would default to an array-averaged *p(v)*. In a second embodiment, a previously determined *p(v)* from experiments using a known reference sequence can be used, e.g., where the previous experiments were performed under the same conditions as the *de novo* experiments. In a third embodiment, *p(v)* is estimated by using quality information/metrics to screen for the most likely "true" calls, and restricting the estimates of *v* to regions containing those calls. In a fourth embodiment, where error is low, the called base rate and reference base rate converge to the same rate, and measurements of p(v) without knowledge of the reference become substantially reliable. Further, even if they do not fully converge, they can still be used to accurately infer *p(v)*,as long as the called base rate is predictably higher/lower than the reference base rate. Yet further, the measurement of *p(v)* when a reference sequence is not available can benefit from a detailed look at the probability model which is available from an algorithm like a CRF. That is, *p(v)* can be tabulated using a weighted sum over paths through a CRF probability model.

[00159] As will be clear to the ordinary practioner based upon the teachings herein, this framework extends naturally to the measurements of other potentially systematic variations in $p(v)$ across an array, e.g., even where a single reaction mixture is applied to the entire array. For example, the local temperature of the reaction environment can vary systematically across an array of reactions. The average and variation in the rates of polymerase enzymes on the array would likely have a dependence on this hidden variable. Where the functional form of the temperature dependence is known, the measurement of $p(v)$ can be stabilized across the array by modeling a de novo $p(v)$ as $p_{cond}(v) + p_{x,y}(v)$ where x,y are geometrical variables defining the location on the array. Further, $p(v)$ has been found to be somewhat variable over time. As such, in certain embodiments a model of $p(v;t)$ is developed using an appropriate model for the evolution of $p(v)$ over time.

[00160] Once a representative distribution of velocities $p(v)$ has been obtained for a given read from a given reaction, the expected travel distance in the non-detection period can be expressed as:

$$p\left(\frac{x}{\tau}\right) = IL\left[L[p(v)]^{t/\tau}\frac{1-L[p(v)]}{s}\right]$$

where L[] and IL[] stand for the Laplace and inverse Laplace transform, respectively. A similar result is derived in Svoboda, et al. (PNAS 91:11782 (1994)) and readily follows from considering the pdf of a sum of random variables. Optionally, in certain embodiments density estimation techniques (e.g., kernel density estimation, etc.) are useful when modeling $p(v)$ since they can smooth the resulting numerical calculations in the Laplace and inverse Laplace transform.

[00161] Knowledge of the complete distribution has several advantages over the commonly applied Gaussian approximation. For example, knowledge of the complete distribution of insert lengths is very desirable when using a Bayesian framework approach to detect structural variation. (See, e.g., Bashir, et al. (2008) PLoS Comput. Biol. 4:51; Hormozdiari, et al. (2009) Genome Res. 19:1270; and Lee, et al. (2008) Bioinformatics 24:59.) While Bashir, et al. does not strictly follow a Bayesian approach, the geometric approach described in the paper can be straightforwardly modified to incorporate an actual posterior instead of the boxcar posterior assumed in the paper. Further, during mapping of noncontiguous reads to a genome where they are expected to be concordant (*i.e.*, not a structural variation), it is useful to consider the known distribution when judging the significance of the resulting alignments between the observed reads and the genomic sequence. In addition, when clustering noncontiguous reads that scaffold contigs in a *de novo*

assembly, a path of Bayesian significance can be followed that is very similar to that followed in the structural variation case discussed *supra*.

[00162] In further embodiments, the determination of the distance a polymerase travels between reads is performed using an algorithm based on a simulation approach rather than the exact analytical result used in the algorithm described above. This method relies on Monte Carlo sampling from a distribution, which allows a better extension to arbitrary empirical distributions. It also lacks the difficult computations of numerical Laplace and inverse Laplace transforms, and permits calculation of distances traveled during non-detection periods when the underlying kinetic processes have multi-phasic kinetics, e.g., the presence of long stalls.

[00163] This approach aims to calculate the distribution of the distance x travelled by an enzyme during a time t during which it was not being observed (e.g., during a non-detection period). In some embodiments, a distribution of local rates, $p(v)$, is estimated, where the definition of "local" is set by the correlation length of the rate autocorrelation function, e.g.:

$$\langle \delta v(t)\delta v(t+\Delta)\rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$$

Given a local rate distribution and an assumption that independent identically distributed (i.i.d.) draws can be made from this distribution, one approach to calculating the distribution is as follows. First, draw $N = t/\tau_{corr}$ velocities from $p(v)$; and subsequently sum them and record them as an estimate of x/τ_{corr}. Repeat the process M times, with the optimal choice of M dependent on the desired level of precision for estimation of the $p(x)$ distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000.

[00164] In some embodiments in which the enzyme system is not well explained by a single kinetic process or cycle (as in the case of observed stalling behavior), above-described rate autocorrelation function and the i.i.d. assumption will be violated. As such, a probability model having a richer structure is preferably used. One such probability model is a Hidden Markov Model (HMM). Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system. Where the kinetics of the pausing state can be well described by a single-exponential, this model is expected to describe the observed distribution of local velocities. The single-exponential assumption is implicit in the state structure of the model since the amount of time spent in the pause state will be a geometric distribution with mean p/(1-p) [*i.e.*, the observed stall times will have to be added to this model]. If the stall kinetics are multi-phasic, then more

"dark states" will have to be added to this model. Further, the model shown in Figure 15 can actually be treated as a Markov Model and not a Hidden Markov Model without much loss of generality because the "pause" state is not actually hidden due to the fact that the data collected during the pause state is highly distinguishable from the data collected during the sequencing state. As such, the general HMM apparatus is not necessary. The model in Figure 15 can be used to simulate the distribution of local velocities when there is a long-term pause or stall phase present in the reaction data kinetics. S0 is the start state, and there is no explicit end state since this model is used as a generative model and it is assumed that it is run forward for a prescribed number of steps. The qualities P(P→S) and P(S→P) represent exit from a stalled state and entry into a stalled state, respectively. These qualities can be measured by an EM algorithm or they can be quickly estimated by physical observables. $P_{P\to S} = 1/\left(1+\frac{\tau_{stall}}{\tau_{corr}}\right)$ and P(S→P) is the frequency of stall starts per τ_{corr}. (Example parameters are τ_{stall}=80 seconds; τ_{corr} = 10 seconds; and P(S→P) = 1/24.) The simulation estimate of *p(x)* can now be produced using the procedure outlined above in which N = t/τ_{corr} velocities are drawn from *p(v)*; and they are subsequently summed and recorded as an estimate of x/τ_{corr}. The process is repeated M times, with the optimal choice of M dependent on the desired level of precision for estimation of the *p(x)* distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000. Figure 16 shows exemplary simulated applications of this method. Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters P(S→P) = 1/24; P(P→S) = 1/11; and *p(v)*~Gamma(48,0.25). Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a 1300 second non-detection period, which is calculated by sampling 2000 estimates from the HMM model.

[00165] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period. The influence of pause/stall behavior can be seen in the heavy-left tailing of both distributions.

[00166] While the simulation method in which i.i.d. draw assumption is valid is more general and can treat arbitrary *p(v)* and more complex models for non-sequencing states, the two-state model using the HMM can be treated analytically. The result of this is:

$$p(x/\tau_{corr}) = \sum_{N_S=0}^{N} \pi_{N_S}(x) p_N(N_S)$$

where $\pi_{N_S}(x)$ is the distribution of the sum of N_S variables drawn from $p(v)$. For the general case, this distribution is given by the Laplace transform approach presented above. For $p(v)$~Normal(μ,σ), this distribution is distributed as Normal($N_S\mu,\sqrt{N_S}\sigma$). For $p(v)$~Gamma(k,θ), this distribution is distributed as Gamma(N_Sk,θ). $P_N(N_S)$ is the number of cycles spent in the sequencing state if we observe N cycles from the Markov process in Figure 15. The expression for this is described in Pedler, et al. (1971) J. Appl. Prob. 8:381, which is incorporated herein by reference in its entirety for all purposes.

[00167] As will be clear to one of ordinary skill in the art upon review of the teachings herein, these methods can be readily extended to the non-detection period estimations of procession by other cyclical biological reactions, such as the action of reverse transcriptase or the synthesis of proteins by a ribosome complex, e.g., and certain preferred embodiments of such reactions are further described in U.S.S.N. 12/767,673, filed April 26, 2010; and U.S.S.N. 12/813,968, filed June 11, 2010, the disclosures of which are incorporated herein by reference in their entireties for all purposes. Further, the simulation model described above is not restricted to simple two-state kinetics, and the use of $p(v)$ is not restricted to analytical models. In fact, in certain embodiments, empirical estimates are preferably used.

[00168] Although useful in certain preferred embodiments of the invention, certain algorithms as presented above do not easily handle the case where the template does not match a physically-motivated expected model. A relevant example of such a case is when the template contains a genomic structural variation (SV), such as translocation, whereby two fragments which are correctly adjacent in the template are located very far apart in the reference genome. Such structural variation cases are best handled in the context of the current algorithm by reporting the confidence of an observed path and reporting situations when no physically expected path seems to fit the observed data. In general, the detection of structural variation requires the presence of multiple highly significant local alignments which can be identified as significantly overturning the null hypothesis of matching the genomic ordering of fragments with their own individual merit. Nevertheless, with molecular redundant sequencing such as SMRTbell™ template sequencing the current algorithm can be adapted to improve the ability to identify an SV event. Such a modification could be a feedback approach which allows modification of the linking constraints in step 3 to allow very far separations on the target sequence when the individual alignments are very significant. Only one such highly-significant pair would be needed to enable the rescue of less significant partial matches that support the same SV hypothesis.

[00169] The software and algorithm implementations provided herein are particularly suited for transforming sequence read data generated from various sequencing technologies (e.g., sequencing-by-synthesis, intramolecular redundant sequencing, Sanger sequencing, capillary electrophoretic sequencing, pyrosequencing, ligase-mediated sequencing, etc.) into consensus sequence data that provides a representation of the actual nucleotide sequence of the template nucleic acid that was subjected to the sequencing reaction(s) from which the sequence read data was generated. The software and algorithm implementations provided herein are preferably machine-implemented methods. The various steps recited herein are preferably performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. The results of these methods are preferably stored on a machine-readable medium, as well. Further, the invention provides a computer program product comprising a computer usable medium having a computer readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention.

[00170] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data).

[00171] While described with reference to certain specific applications above, it will be understood that these methods are also applicable to other types of complex data sets, and the invention should not be limited to only the specific examples provided herein. Other applications of the instant methods will be clear to those of ordinary skill in the art and are considered to be additional aspects of the instant invention.

V. Devices and Systems

[00172] The invention also provides systems that are used in conjunction with the compositions and methods of the invention in order to provide for intermittent detection of analytical reactions. In particular, such systems typically include the reagent systems described herein, in conjunction with an analytical system, e.g., for detecting data from those reagent systems.

For example, a sequencing reaction may be subjected to intermittent illumination, and the sequencing system may include the system components provided with or sold for use with commercially available nucleic acid sequencing systems, such as the Genome Analyzer System available from Illumina, Inc., the GS FLX System, available from 454 Life Sciences, or the ABI 3730 System available from Life Technologies, Inc.

[00173] In certain preferred embodiments, reactions subjected to intermittent illumination are monitored using an optical system capable of detecting and/or monitoring interactions between reactants at the single-molecule level. Such an optical system achieves these functions by first generating and transmitting an incident wavelength to the reactants, followed by collecting and analyzing the optical signals from the reactants. Such systems typically employ an optical train that directs signals from the reactions to a detector, and in certain embodiments in which a plurality of reactions is disposed on a solid surface, such systems typically direct signals from the solid surface (e.g., array of confinements) onto different locations of an array-based detector to simultaneously detect multiple different optical signals from each of multiple different reactions. In particular, the optical trains typically include optical gratings or wedge prisms to simultaneously direct and separate signals having differing spectral characteristics from each confinement in an array to different locations on an array based detector, *e.g.*, a CCD, and may also comprise additional optical transmission elements and optical reflection elements.

[00174] An optical system applicable for use with the present invention preferably comprises at least an excitation source and a photon detector. The excitation source generates and transmits incident light used to optically excite the reactants in the reaction. Depending on the intended application, the source of the incident light can be a laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source. Further, the excitation light may be evanescent light, e.g., as in total internal reflection microscopy, certain types of waveguides that carry light to a reaction site (see, e.g., U.S. Application Pub. Nos. 20080128627, 20080152281, and 200801552280), or zero-mode waveguides, described below. Where desired, more than one source can be employed simultaneously. The use of multiple sources is particularly desirable in applications that employ multiple different reagent compounds having differing excitation spectra, consequently allowing detection of more than one fluorescent signal to track the interactions of more than one or one type of molecules simultaneously. A wide variety of photon detectors or detector arrays are available in the art. Representative detectors include but are not limited to optical reader, high-efficiency photon detection system, photodiode (*e.g.* avalanche photo diodes

(APD)), camera, charge couple device (CCD), electron-multiplying charge-coupled device (EMCCD), intensified charge coupled device (ICCD), and confocal microscope equipped with any of the foregoing detectors. For example, in some embodiments an optical train includes a fluorescence microscope capable of resolving fluorescent signals from individual sequencing complexes. Where desired, the subject arrays of optical confinements contain various alignment aides or keys to facilitate a proper spatial placement of the optical confinement and the excitation sources, the photon detectors, or the optical train as described below.

[00175] The subject optical system may also include an optical train whose function can be manifold and may comprise one or more optical transmission or reflection elements. Such optical trains preferably encompass a variety of optical devices that channel light from one location to another in either an altered or unaltered state. First, the optical train collects and/or directs the incident wavelength to the reaction site (e.g., optical confinement). Second, it transmits and/or directs the optical signals emitted from the reactants to the photon detector. Third, it may select and/or modify the optical properties of the incident wavelengths or the emitted wavelengths from the reactants. In certain embodiments, the optical train controls an on/off cycle of the illumination source to provide illuminated and non-illuminated periods to one or more illuminated reaction sites. Illustrative examples of such optical transmission or reflection elements are diffraction gratings, arrayed waveguide gratings (AWG), optic fibers, optical switches, mirrors (including dichroic mirrors), lenses (including microlenses, nanolenses, objective lenses, imaging lenses, and the like), collimators, optical attenuators, filters (*e.g.*, polarization or dichroic filters), prisms, wavelength filters (low-pass, band-pass, or high-pass), planar waveguides, wave-plates, delay lines, and any other devices that guide the transmission of light through proper refractive indices and geometries. One example of a particularly preferred optical train is described in U.S. Patent Pub. No. 20070036511, filed August 11, 2005, and incorporated by reference herein in its entirety for all purposes.

[00176] In a preferred embodiment, a reaction site (e.g., optical confinement) containing a reaction of interest is operatively coupled to a photon detector. The reaction site and the respective detector can be spatially aligned (*e.g.*, 1:1 mapping) to permit an efficient collection of optical signals from the reactants. In certain preferred embodiments, a reaction substrate is disposed upon a translation stage, which is typically coupled to appropriate robotics to provide lateral translation of the substrate in two dimensions over a fixed optical train. Alternative embodiments could couple the translation system to the optical train to move that aspect of the system relative to the substrate. For

example, a translation stage provide a means of removing a reaction substrate (or a portion thereof) out of the path of illumination to create a non-illuminated period for the reaction substrate (or a portion thereof), and returning the substrate at a later time to initiate a subsequent illuminated period. An exemplary embodiment is provided in U.S. Patent Pub. No. 20070161017, filed December 1, 2006.

[00177] In particularly preferred aspects, such systems include arrays of reaction regions, e.g, zero-mode waveguide arrays, that are illuminated by the system, in order to detect signals (e.g., fluorescent signals) therefrom, that are in conjunction with analytical reactions being carried out within each reaction region. Each individual reaction region can be operatively coupled to a respective microlens or a nanolens, preferably spatially aligned to optimize the signal collection efficiency. Alternatively, a combination of an objective lens, a spectral filter set or prism for resolving signals of different wavelengths, and an imaging lens can be used in an optical train, to direct optical signals from each confinement to an array detector, *e.g.*, a CCD, and concurrently separate signals from each different confinement into multiple constituent signal elements, *e.g.*, different wavelength spectra, that correspond to different reaction events occurring within each confinement. In preferred embodiments, the setup further comprises means to control illumination of each confinement, and such means may be a feature of the optical system or may be found elsewhere is the system, e.g., as a mask positioned over an array of confinements. Detailed descriptions of such optical systems are provided, e.g, in U.S. Patent Pub. No. 20060063264, filed September 16, 2005, which is incorporated herein by reference in its entirety for all purposes.

[00178] The systems of the invention also typically include information processors or computers operably coupled to the detection portions of the systems, in order to store the signal data obtained from the detector(s) on a computer readable medium, e.g., hard disk, CD, DVD or other optical medium, flash memory device, or the like. For purposes of this aspect of the invention, such operable connection provide for the electronic transfer of data from the detection system to the processor for subsequent analysis and conversion. Operable connections may be accomplished through any of a variety of well known computer networking or connecting methods, e.g., Firewire®, USB connections, wireless connections, WAN or LAN connections, or other connections that preferably include high data transfer rates. The computers also typically include software that analyzes the raw signal data, identifies signal pulses that are likely associated with incorporation events, and identifies bases incorporated during the sequencing reaction, in order to convert or transform the raw signal data into user interpretable sequence data (See, e.g., Published

U.S. Patent Application No. 2009-0024331, the full disclosure of which is incorporated herein by reference in its entirety for all purposes).

[00179] Exemplary systems are described in detail in, e.g., U.S. Patent Application No. 11/901,273, filed September 14, 2007 and U.S. Patent Application No. 12/134,186, filed June 5, 2008, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00180] Further, as noted above, the invention provides data processing systems for transforming sequence read data into consensus sequence data. In certain embodiments, the data processing systems include machines for generating sequence read data by interrogating a template nucleic acid molecule. In certain preferred embodiments, the machine generates the sequence read data using a sequencing-by-synthesis technology, as described elsewhere herein, but the machine may generate the sequence read data using other sequencing technologies known to those of ordinary skill in the art, e.g., pyrosequencing, ligation-mediated sequencing, Sanger sequencing, capillary electrophoretic sequencing, etc. Such machines and methods for using them are available to the ordinary practioner.

[00181] The sequence read data generated is representative of the nucleotide sequence of the template nucleic acid molecule only to the extent that a given sequencing technology is able to generate such data, and so may not be identical to the actual sequence of the template nucleic acid molecule. For example, it may contain a deletion or a different base at a given position as compared to the actual sequence of the template, e.g., when a base call is missed or incorrect, respectively. As such, it is beneficial to generate redundant sequence read data, and the methods described herein provide manipulations and computations that transform redundant sequence read data into consensus sequence data that is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Redundant sequence read data comprises multiple reads, each of which includes at least a portion of sequence read that overlaps with at least a portion of at least one other of the multiple reads. As such, the multiple reads need not all overlap with one another, and a first subset may overlap for a different portion of the template nucleic acid sequence than does a second subset. Such redundant sequence read data can be generated by various methods, including repeated sequencing of a single nucleic acid template, sequencing of multiple identical nucleic acid templates, or a combination thereof.

[00182] In another aspect, the data processing systems can include software and algorithm implementations provided herein, e.g. those configured to transform redundant sequence read data into consensus sequence data, which, as noted above, is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Further, the transformation of the redundant sequence read data into consensus sequence data identifies and negates some or all of the single-read variation between the multiple reads in the redundant sequence read data. As such, the transformation provides a representation of the actual nucleotide sequence of the nucleic acid template from which redundant sequence read data is generated that is more accurate than a representation based on a single read.

[00183] The software and algorithm implementations provided herein are preferably machine-implemented methods, e.g., carried out on a machine comprising computer-readable medium configured to carry out various aspects of the methods herein. For example, the computer-readable medium preferably comprises at least one or more of the following: a) a user interface; b) memory for storing redundant sequence read data; c) memory storing software-implemented instructions for carrying out the algorithms for transforming redundant sequence read data into consensus sequence data; d) a processor for executing the instructions; e) software for recording the results of the transformation into memory; and f) memory for recordation and storage of the resulting consensus sequence read data. In preferred embodiments, the user interface is used by the practitioner to manage various aspects of the machine, e.g., to direct the machine to carry out the various steps in the transformation of redundant sequence read data into consensus sequence data, recordation of the results of the transformation, and management of the consensus sequence data stored in memory.

[00184] As such, in preferred embodiments, the methods further comprise a transformation of the computer-readable medium by recordation of the redundant sequence read data and/or the consensus sequence data generated by the methods. Further, the computer-readable medium may comprise software for providing a graphical representation of the redundant sequence read data and/or the consensus sequence read data, and the graphical representation may be provided, e.g., in soft-copy (e.g., on an electronic display) and/or hard-copy (e.g., on a print-out) form.

[00185] The invention also provides a computer program product comprising a computer-readable medium having a computer-readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and

optionally also providing storage for the results of the methods of the invention. In certain preferred embodiments, the computer program product comprises the computer-readable medium described above.

[00186] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data), such as the computer-readable medium described above.

[00187] As shown in Figure 9, the system 900 includes a substrate 902 that includes a plurality of discrete sources of chromophore emission signals, e.g., an array of zero-mode waveguides 904. An excitation illumination source, e.g., laser 906, is provided in the system and is positioned to direct excitation radiation at the various signal sources. This is typically done by directing excitation radiation at or through appropriate optical components, e.g., dichroic 108 and objective lens 910, that direct the excitation radiation at the substrate 902, and particularly the signal sources 904. Emitted signals from the sources 904 are then collected by the optical components, e.g., objective 910, and passed through additional optical elements, e.g., dichroic 908, prism 912 and lens 914, until they are directed to and impinge upon an optical detection system, e.g., detector array 916. The signals are then detected by detector array 916, and the data from that detection is transmitted to an appropriate data processing system, e.g., computer 918, where the data is subjected to interpretation, analysis, and ultimately presented in a user ready format, e.g., on display 920, or printout 922, from printer 924. As will be appreciated, a variety of modifications may be made to such systems, including, for example, the use of multiplexing components to direct multiple discrete beams at different locations on the substrate, the use of spatial filter components, such as confocal masks, to filter out-of focus components, beam shaping elements to modify the spot configuration incident upon the substrates, and the like (See, e.g., Published U.S. Patent Application Nos. 2007/0036511 and 2007/095119, and U.S. Patent Application No. 11/901,273, all of which are incorporated herein by reference in their entireties for all purposes.)

VI. Exemplary Applications

[00188] The methods and compositions of the invention are useful in a broad range of analytical reactions in which one or more aspects of a detection method are detrimental to one or more aspects of the analytical reaction, such as rate, duration, fidelity, processivity, and the like. In such cases, intermittent detection at least partially mitigates the detrimental effect while allowing collection of data from stages of the analytical reaction that were previously uncollectable. As noted above, illuminated reactions are one example of analytical reactions that benefit from the compositions and methods described herein, particularly those using photoluminescent or fluorescent reagents, and particularly such reactions where one or more of the reaction components that are susceptible to photo-induced damage are present at relatively low levels. One exemplary application of the methods and compositions described herein is in single molecule analytical reactions, where the reaction of a single molecule (or very limited number of molecules) is observed in the analysis, such as observation of the action of a single enzyme molecule. In another aspect, the present invention is directed to illuminated reactions for single molecule analysis, including sequencing of nucleic acids by observing incorporation of nucleotides into a nascent nucleic acid sequence during template-directed polymerase-based synthesis. Such methods, generally referred to as "sequencing-by-incorporation" or "sequencing-by-synthesis," involve the observation of the addition of nucleotides or nucleotide analogs in a template-dependent fashion in order to determine the sequence of the template strand. See, e.g., U.S. Patent Nos. 6,780,591, 7,037,687, 7,344,865, 7,302,146. Processes for performing this detection include the use of fluorescently labeled nucleotide analogs within a confined observation region, e.g., within a nanoscale well and/or tethered, either directly or indirectly to a surface. By using excitation illumination (i.e., illumination of an appropriate wavelength to excite the fluorescent label and induce a detectable signal), the fluorescently labeled bases can be detected as they are incorporated into the nascent strand, thus identifying the nature of the incorporated base, and as a result, the complementary base in the template strand.

[00189] In particular aspects, when an analysis relies upon a small population of reagent molecules, damage to any significant fraction of that population will have a substantial impact on the analysis being performed. For example, prolonged interrogation of a limited population of reagents, e.g., fluorescent analogs and enzymes, can lead to photo-induced damage of the various reagents to the point of substantially impacting the activity or functionality of the enzyme. It has been shown that prolonged illumination of DNA polymerases involved in synthesis using fluorescent nucleotide analogs results in a dramatic decrease in the enzyme's ability to synthesize

DNA, often measured as a reduction in processivity. Without being bound to any theory of operation, it is believed that in some cases a photo-induced damage event affects the catalytic region of the enzyme thus affecting either the ability of the enzyme to remain complexed with the template, or its ability to continue synthesis. In general, the methods, devices, and systems of the present invention can increase performance and/or selectively monitor one or more stages of an illuminated reaction by subjecting the reaction to intermittent illumination.

[00190] One particularly preferred aspect of the invention is in conjunction with the sequencing by incorporation of nucleic acids within an optical confinement, such as a zero-mode waveguide. Such reactions involve observation of an extremely small reaction volume in which one or only a few polymerase enzymes and their fluorescent substrates may be present. Zero-mode waveguides, and their use in sequencing applications are generally described in U.S. Patent Nos. 6,917,726 and 7,033,764, and preferred methods of sequencing by incorporation are generally described in Published U.S. Patent Application No. 2003-0044781, the full disclosures of which are incorporated herein by reference in their entireties for all purposes, and in particular for their teachings regarding such sequencing applications and methods. Briefly, arrays of zero-mode waveguides ("ZMWs"), configured in accordance with the present invention may be employed as optical confinements for single molecule DNA sequence determination. In particular, as noted above, these ZMWs provide extremely small observation volumes at or near the transparent substrate surface, also termed the "base" of the ZMW. A nucleic acid synthesis complex, e.g., template sequence, polymerase, and primer, which is immobilized at the base of the ZMW, may then be specifically observed during synthesis to monitor incorporation of nucleotides in a template dependent fashion, and thus provide the identity and sequences of nucleotides in the template strand. This identification is typically accomplished by providing detectable label groups, such as fluorescent labeling molecules, on the nucleotides. In some instances, the labeled nucleotides terminate primer extension, allowing a "one base at a time" interrogation of the complex. If, upon exposure to a given labeled base, a base is incorporated, its representative fluorescent signal may be detected at the base of the ZMW. If no signal is detected, then the base was not incorporated and the complex is interrogated with each of the other bases, in turn. Once a base is incorporated, the labeling group is removed, e.g., through the use of a photocleavable linking group, and where the label was not the terminating group, a terminator, upon the 3' end of the incorporated nucleotide, may be removed prior to subsequent interrogation. In other more preferred embodiments, the incorporation of a labeled nucleotide does not terminate primer extension and the processive

incorporation of multiple labeled nucleotides can be monitored in real time by detecting a series of fluorescent signals at the base of the ZMW. In some such embodiments, the label is naturally released upon incorporation of the labeled nucleotides by the polymerase, and so need not be released by alternative means, e.g., a photocleavage event. As such, a processive sequencing reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. (Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.)

[00191] In accordance with the present invention, sequencing reactions may be carried out by only interrogating a reaction mixture, e.g., detecting fluorescent emission for one or more illuminated periods before excessive photo-induced damage has occurred. In general, the methods described herein are implemented in a manner sufficient to provide beneficial impact, e.g., reduced photo-induced damage and/or extension of the photo-induced damage threshold period, but are not implemented in such a manner to interfere with the reaction of interest, e.g., a sequencing reaction. The present invention also contemplates alternative methods of and compositions for mitigating the impact of photo-induced damage on a reaction, as described above and in, e.g., U.S.S.N. 61/116,048, filed November 19, 2008. Such alternative methods and compounds can be used in combination with the compositions and methods provided herein to further alleviate the effects of species that can be generated during an illuminated reaction.

[00192] Another method of mitigating the impact of photo-induced damage on the results of a given reaction provides for the elimination of potentially damaging oxygen species using means other than the use of the photo-induced damage mitigating agents described above. In one example, dissolved oxygen species may be flushed out of aqueous systems by providing the reaction system under different gas environments, such as by exposing an aqueous reaction to neutral gas environments, such as argon, nitrogen, helium, xenon, or the like, to prevent dissolution of excess oxygen in the reaction mixture. By reducing the initial oxygen load of the system, it has been observed that photo-induced damage effects, e.g., on polymerase mediated DNA synthesis, is markedly reduced. In particularly preferred aspects, the system is exposed to a xenon atmosphere. In particular, since xenon can be induced to form a dipole, it operates as a triplet-state quencher in addition to supplanting oxygen in the aqueous system. (See, e.g., Vierstra and Poff, Plant Physiol.

1981 May; 67(5): 996–998) As such, xenon would also be categorized as a quencher, as set forth above.

[00193] Although described in terms of zero-mode waveguides, it will be appreciated that a variety of selective illumination strategies may be employed to selectively interrogate different regions of a solid support over time, e.g., so as to only damage molecules within certain selected regions of a substrate while not damaging molecules in other selected regions of the substrate. In certain embodiments, such methods can involve using a directed light source (e.g., a laser) to illuminate only selected regions; changing the illumination angle of the light source; or refocusing the illumination, e.g., by passing the illumination through an optical train that alters the shape of the incident light on the solid support. These and further examples of alternative methods of mitigating photo-induced damage which can be used in combination with methods and systems of the invention described herein are provided in U.S. Patent Pub. No. 20070036511, filed August 11, 2005; U.S. Patent No. 6,881,312; U.S.S.N. 61/116,048, filed November 19, 2008; and U.S. Patent Pub. No. 20070161017, filed December 1, 2006, all of which are incorporated herein by reference in their entireties for all purposes, and in particular for disclosure related to these methods of mitigating photo-induced damage.

[00194] As noted above, using templates that allow repeated sequencing (e.g., circular templates, SMRTbell™ templates, etc.) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated and/or increase the fold-coverage of the sequence reads for one or more regions of interest in the template, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, in certain preferred embodiments, templates sequenced by the methods described herein are templates comprising a double-stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double-stranded or otherwise internally complementary. Such templates may, for example, comprise a double-stranded portion comprised of two complementary sequences and two single-stranded linking portions (e.g., oligos or "hairpins") joining the 3' end of each strand of the double-stranded region to the 5' end of the other strand (sometimes referred to as "SMRTbell™" templates). In certain embodiments, double-stranded portions for use in such templates are PCR-amplified. Optionally, restriction sites are incorporated within the PCR primers such that subsequent digestion of the amplified products with appropriate restriction enzymes generates double-stranded portions containing known

overhang sequences on either end, which are then ligated to hairpin adapters containing a complementary overhang to generate the SMRTbell™ templates.

[00195] These template molecules are particularly useful as nucleotide sequence data generated therefrom comprises both sense and antisense nucleotide sequences for the double-stranded portion, and the circular conformation of the template enables repeated sequencing (e.g., using a polymerase capable of strand-displacement) provides duplicative or redundant sequence information. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining sequence data for each segment from the complementary sequences, as well as sequence data within each segment, by repeatedly sequencing that segment. Iterative illumination is useful in such sequencing applications, e.g., to focus nucleotide sequence data collection on stages of the sequencing reaction most of interest, such as the stages during which nucleotide sequence data is being generated from a strand of the (previously) double-stranded portion. Iterative illumination may also allow additional "rounds" of sequencing the template by virtue of the reduction in photo-induced damage to reaction components, as described elsewhere herein, thereby providing more complete and robust nucleotide sequence data for future analysis, e.g., sequence scaffold construction and/or consensus sequence determination. Further, as described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest, e.g., from the ends of the double-stranded portion of a SMRTbell™ template.

[00196] In addition to providing sense and antisense sequence data within a single template molecule that can be sequenced in one integrated process, the presence of the single-stranded linking portions also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., the sense strand, is completed and the other begins, e.g., the antisense strand. Such registration sequences provide a basis for alignment sequence data from

multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. Additional aspects of and uses for registration sequences, e.g., for molecular redundant sequencing, are further described in U.S. Patent Publication No. 20090029385, which is incorporated herein by reference in its entirety for all purposes.

[00197] In certain embodiments, such a sequencing process begins by priming the template nucleic acid within one of the linking portions and allowing the polymerase to proceed along the strand of the double-stranded portion of the template that is immediately downstream of the primed linking portion when the double-stranded portion is melted or denatured. The sequence process proceeds around the second linking portion and proceeds along the complementary strand of the (now previously) double-stranded portion of the template. Because the template is circular, this process can continue to provide multiple repeated sequence reads from the one template. Thus, sequence redundancy comes from both the determination of complementary sequences (sense and antisense strands of the double-stranded portion), and the repeated sequencing of each circular template. The ongoing sequencing reaction is subjected to multiple illuminated and non-illuminated periods to generate at least two or more sequence reads per pass around the template. The illuminated periods are preferably timed to allow generation of nucleotide sequence data for selected regions of the template. For example, it may be beneficial to only generate nucleotide sequence data for the complementary strands of the double-stranded portion, or segments thereof. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing. Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post-synthesis.

[00198] One may optionally employ various means for controlling initiation and/or progression of a sequencing reaction, and such means may include the addition of specific sequences or other moieties into the template nucleic acid, such as binding sites, e.g., for primers or proteins. Various methods of incorporating control elements into an analytical reaction, e.g. by integrating stop or pause points into a template, are discussed elsewhere herein and are further described in related application, U.S. Application No. 12/413,258, filed March 27, 2009, which is incorporated herein by reference in its entirety for all purposes.

[00199] In certain embodiments, a reaction stop or pause point may be included within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g.,

on the linking portion that was not used in priming. By way of example, following initial sequencing from the original priming location, e.g., from the single-stranded linking portion used in priming synthesis through a first portion of the sense strand (e.g., the 3' end), the data acquisition may be switched off and the polymerase allowed to proceed around the template, e.g., through the remainder of the sense strand to the other linking portion. The incorporation of a synthesis blocking moiety coupled to this linking portion will allow control of reinitiation of the polymerase activity at the 3' end of the antisense strand. One would thereby obtain paired-end sequence data for the overall (previously) double-stranded segment, with sequence data from one end coming from the sense strand and sequence data from the other end coming from the antisense strand. This template construction and sequencing methodology is particularly useful in the case of long double-stranded segments, especially given the short read lengths generated by some sequencing technologies.

[00200] A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the nucleobase portion of one or more bases in the single-stranded portion that inhibit polymerase-mediated replication; strand-binding moieties that prevent processive synthesis; non-native nucleotides included within the primer and/or template; and the like. The use of strand-binding moieties includes, but is not limited to, reversible, specific binding of particular proteins to recognition sequences incorporated into the template (or primer bound thereto) for this purpose. In certain embodiments, such control sequences may include binding sites for transcription factors, e.g., repressor binding regions provided within the linking portion(s). For example, the lac repressor recognition sequence is bound by the lac repressor protein, and this binding has been shown to block replication in a manner reversible by addition of appropriate initiators, such as isophenylthiogalactoside (IPTG) or allolactose.

[00201] In some embodiments, primer recognition sequences and/or additional control sequences may also be provided for control of initiation and/or progression of polymerization, e.g., through a hybridized probe or reversibly modified nucleotide, or the like. (See, e.g., U.S. Patent Application No. 2008-0009007, the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Such probes include but are not limited to probes at which a polymerase initiates polymerization, probes containing various types of detectable labels, molecular beacons, TaqMan® probes, Invader® probes (Third Wave Technologies, Inc.), or the like, that can be used for various purposes, e.g., to provide indications of the commencement and/or progress of synthesis.

[00202] An engineered pause point (reversible or irreversible) can include one or more non-native (non-natural) or fifth bases that do not pair with any of the four native nucleoside polyphosphates in the synthesis reaction, e.g., in the template and/or oligonucleotides probe(s), and/or that exhibit a distinct kinetic signature during template-dependent synthesis at such a base. Upon encountering such a base, the polymerase pauses until the complement to the non-natural base is added to the reaction mixture. Likewise, an engineered pause point could include a "damaged" base that causes a stop in replication until repair enzymes are added to the mixture. For example, a template having a pyrimidine dimer would cause the replication complex to pause, and addition of the photolyase DNA repair enzyme would repair the problem location and allow replication, and sequencing to continue. In yet further embodiments, a combination of modification enzymes could be used to engineer a set of modified bases on a template, e.g., a combination of glycosylases, methylases, nucleases, and the like. (Further information on sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.)

[00203] As noted elsewhere herein, stop or pause points can be engineered into various portions of the template, e.g., portions for which the nucleotide sequence is unknown (e.g., a genomic fragment) or known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, SMRTbell™ templates are topologically closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, such that hybridization of the regions of internal complementarity produces a double-stranded portion within the template. One or both of the linkers can comprise a stop or pause point to modulate polymerase activity. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00204] In other embodiments, an abasic site is used as a synthesis blocking moiety or pause point until addition of a non-natural "base," such as a pyrene, which has been shown to "base-pair" with an abasic site during DNA synthesis. (See, e.g., Matray, et al. (1999) Nature 399(6737):704-8,

which is incorporated herein by reference in its entirety for all purposes.) Where a permanent termination of sequencing is desired, no non-natural analog is added and the polymerase is permanently blocked at the abasic site. DNA (or RNA) glycosylases create abasic sites that are quite different from the normal coding bases, A, T, G, and C (and U in RNA). A wide variety of monofunctional and bifunctional DNA glycosylases that have specificity for most common DNA or RNA adducts, including 5-methylcytosine, are known in the art, with different glycosylases capable of recognizing different types of modified DNA and/or RNA bases. The molecular structures of many glycosylases have been solved, and based on structural similarity they are grouped into four superfamilies. The UDG and AAG families contain small, compact glycosylases, whereas the MutM/Fpg and HhH-GPD families comprise larger enzymes with multiple domains. As an example, four enzymes have been identified in *Arabidopsis thaliana* in the plant pathway for cytosine demethylation. Additionally, other enzymes are also known to recognize 5-methyl cytosine and remove the methylated base to create an abasic site. Further, various enzymes are known to methylate cytosine in a sequence-specific manner. As such, a combination of a cytosine-methylase and an enzyme that creates an abasic site from a methylated cytosine nucleotide can be used to create one or more abasic sites in a template nucleic acid. The size of the recognition site of the methylase and the base composition of the template determine how frequently methylation occurs, and therefore, the number of abasic sites created in a given template nucleic acid, allowing the ordinary practioner to choose a methylase with a recognition site that produces a desired spacing between modified nucleotides. For example, if the recognition site is three bases long, then on average an abasic site is expected every 64 bases; if the recognition site is four bases long, then on average an abasic site is expected every 256 bases; if the recognition site is six bases long, then on average an abasic site is expected every 4096 bases; and so forth. Of course, templates with a higher GC content would be expected to have more frequent abasic site formation, and templates with lower GC content would be expected to have less frequent abasic site formation.

[00205] Uracil-DNA glycosylases can also be used to introduce abasic sites into a template nucleic acid comprising deoxyuridine nucleotides. This strategy has the advantage of allowing the practitioner to choose the locations of the abasic sites within a DNA template since deoxyuridine nucleotides are not generally found in DNA. Various methods of inserting deoxyuridine nucleotides into a DNA template may be used, and different methods will be preferred for different applications. In certain embodiments, one or more site-specific deoxyuracils are incorporated during standard phosphoramidite oligonucleotide synthesis. To place uracils at indeterminate positions in a

DNA, replacing a portion of the deoxythymidine triphosphate with deoxyuridine triphosphate will result in an amplimer with random U sites in place of T sites after polymerase chain reaction. In other embodiments, deoxyuridine nucleotides are engineered into the template, e.g., by ligation of a synthetic linker or adaptor comprising one or more deoxyuridine nucleotides to a nucleic acid sequence to be sequenced. In certain preferred embodiments, deoxyuridine nucleotides are incorporated into the linker portions of a SMRTbell™ template.

[00206] To subsequently introduce abasic sites prior to sequencing, the deoxyuridine nucleotide-containing template is subjected to treatment with uracil-DNA glycosylase, which removes the one or more uracil bases from the deoxyuridine nucleotides, thereby generating one or more abasic sites in the template. Alternatively, since the deoxyuridine nucleotide can be recognized as a template base and paired with deoxyadenosine during template-dependent nascent strand synthesis, the synthesis-blocking abasic site can instead be introduced after initiation of the sequencing reaction, e.g., at a time chosen by the practitioner. For example, the reaction can be initiated with a deoxyuridine-containing template, and uracil-DNA glycosylase can subsequently be added to block the polymerase and halt the reaction after the reaction has proceeded for a given time. As such, termination of the reaction is optional rather than required.

[00207] While uracil-DNA glycosylase activity is useful for introducing abasic sites into a template as described above, this activity can be problematic during the preparation of such templates. As such, strategies are typically implemented during preparation and manipulation of uracil-containing DNA, e.g., using molecular biology enzymes, to avoid uracil-DNA glycosylase activity, in particular, due to the *E. coli* UDG enzyme. Since a majority of standard molecular biology enzymes are overexpressed and subsequently purified from an *E. coli* host, UDG activity can be a contaminating activity that is often not monitored by the enzyme manufacturer's quality control procedures. To mitigate contaminating UDG activity, a commercially available UDG inhibitor, also known as uracil glycosylase inhibitor or UGI (e.g., from New England Biolabs, Ipswich, MA) can be included in molecular biology reactions. This is a small protein inhibitor from the *B. subtilis* bacteriophase PBS1 that binds reversibly to E. coli UDG to inhibit its catalytic activity. UGI is also capable of dissociating UDG from a DNA molecule. Alternatively, UDG activity can be inhibited without exogenous protein using a chemical inhibitor of the enzyme, such as an oligonucleotide containing a 1-aza-deoxyribose base, a transition state analog for the UDG enzyme. This and other cationic nitrogenous sugars have been used for mechanistic studies of UDG

activity and show potent inhibition activity. (See, e.g., Jiang et al. Biochemistry, 2002, 41 (22), pp 7116–7124.)

[00208] In certain applications, UDG activity needs to be inhibited temporarily, and subsequently enabled to remove create an abasic site as described above. In some embodiments, a DNA purification that removes proteins is employed, e.g., including a phenol-chloroform extraction with subsequent ethanol precipitation, a silica-based column approach (e.g., QiaQuick columns from Qiagen and similar products), and/or a PEG/sodium chloride precipitation (e.g., AMPure beads from Beckman Coulter). Alternatively or additionally, a commercially-available UDG enzyme that is not inhibited by UGI is added when abasic site formation is desired. For example, the *A. fulgidus* UDG is from a thermophilic organism and cannot be inhibited by the same bacteriophage protein as is the *E. coli* UDG enzyme. In certain preferred embodiments, UDG-inhibition is employed during template preparation, and inhibition-resistant UDG activity is added at a subsequent time to trigger the creation of abasic sites at deoxyuridine nucleotides, e.g., immediately prior to or during an ongoing reaction.

[00209] In some preferred embodiments, one or more abasic sites are engineered into a linker or adapter sequence within a sequencing template molecule. Abasic sugar residues serve as efficient terminators of polymerization for many polymerases, e.g., Φ29. 1',2'-dideoxyribose is the most common synthetic "abasic site". In other embodiments, a synthetic linker is incorporated into a linker or adaptor. For example, an internal spacer (e.g., Spacer 3 from Biosearch Technologies, Inc.) or other carbon-based linker can be used in lieu of a sugar-base nucleotide. Similar to an abasic nucleotide, the polymerase will be blocked upon encountering these moieties in the template nucleic acid.

[00210] In certain embodiments, synthesis blocking moieties are nicks in the template nucleic acid. Nicking enzymes (e.g., nicking endonucleases) are known in the art and can be used to specifically nick the template prior to or during a template-directed sequencing reaction. The use of site-specific nicking endonucleases allows the practitioner to incorporate a recognition sequence at a particular location within the template nucleic acid, and such nicking endonucleases are commercially available, e.g., from New England Biolabs, Inc. For example, a linker or adapter can be synthesized with a nicking endonuclease recognition sequence, ligated to a nucleic acid molecule to be sequenced, and can be specifically nicked either before or during a subsequent sequencing reaction. Nicks can also be introduced by ligating duplex segments that lack either a terminal 3'-hydroxy (e.g., have a dideoxynucleotide at the 3'-terminius) and/or 5'-phosphate group on one

strand. The ligation results in covalent linkage of the phosphodiester backbone on one strand, but not on the other, which is therefore effectively "nicked." In certain embodiments, a SMRTbell™ template is constructed using a duplex (or "insert") nucleic acid molecule lacking a 5'-phosphate group at one or both termini. Upon ligation of the hairpin or stem-loop adaptors at each end, nicks are created at one or both ligation site(s), depending on whether the duplex lacked a 5'-phosphate at one or both ends, respectively. In other embodiments, a SMRTbell™ template is constructed using one or two stem-loop adaptors lacking a 3'-hydroxy group at the terminus (e.g., comprising a 2',3'-dideoxynucleotide rather than a 2'-deoxynucleotide). Upon ligation of one or two stem-loop adaptors lacking a 3'-hydroxy group, one or two nicks are created at the ligation site(s), depending on whether one or two adaptors lacked the 3'-hydroxy group, respectively. In both cases, a nick is created in the template nucleic acid, and a primer bound to one of the adaptors provides an initiation site for the polymerase, which will process the template until encountering a nick, at which point the polymerase will terminate the reaction by dissociation from the template. Regardless of how a nick is created, the position of a nick relative to the initiation site for the polymerase determines how much of the template will be sequenced. For example, Figure 19A provides an illustrative example of an embodiment in which a nick is present on a first strand of a duplex portion at a position distal to the adaptor containing the primer binding site. The first strand is processed by a polymerase, but the complementary strand is not processed because the polymerase dissociates at the nick site. An alternative embodiment is shown in Figure 19B, in which a nick is present on the strand complementary to the first strand at a position proximal to the adaptor containing the primer binding site. In this case both the first and complementary strands, as well as the adaptor not containing the primer binding site, are processed by the polymerase prior to dissociation. The position of the primer binding site also determines how much of the template is processed by the polymerase. Figure 19C provides a template having a primer binding site at a position from which a polymerase would process a significant portion of the adaptor prior to entering the duplex portion. An additional advantage to using a 3'-dideoxynucleotide at a nick is that it prevents the use of the nick as a polymerse initiation site, since strand extension requires a 3-hydroxy group. As such, the resulting nick would not compete with a primer site for initiation of nascent strand synthesis by the polymerase. Having a single, known site of initiation on a template molecule is beneficial, e.g., for subsequent mapping of a read generated in such a reaction. In certain preferred embodiments, a nick site both lacks a 5'-phosphate group and comprises a 3'-dideoxynucleotide.

[00211] In certain preferred embodiments, modification and base excision is performed prior to introduction of a template nucleic acid to a reaction site, e.g., a zero-mode waveguide. As noted above, the choice of recognition site for the methylase depends on how far apart the practitioner wishes point of synthesis initiation to be on the template. For example, after initiating the template-dependent sequencing reaction, the sequence of nucleotide incorporations into the nascent strand is monitored for a desired sequence read, which may extend from the initiation point to the pause point, or may end before the polymerase reaches the pause point. In some preferred embodiments, as described elsewhere herein, the monitoring is suspended by modifying or removing an illumination source, e.g., by moving the illumination source or a substrate comprising the reaction site. Synthesis of the nascent strand will continue until the pause site is reached, whether or not the reaction is being actively monitored. When the reaction is to be reinitiated, reaction components are added that allow bypass, e.g., pyrene, polymerase, etc., and these can be subsequently removed (e.g., by buffer exchange) to allow additional pauses at other pause sites on the template.

[00212] In certain embodiments using pyrosequencing-based technologies (e.g., as developed by 454 Life Sciences), abasic sites can be introduced into a set of amplified template nucleic acids and synthesis initiated. Since all templates in the set are identical, they will comprise the same number of abasic sites in the same positions. During the course of the synthesis reaction, the synchronous incorporation of nucleotides into the nascent strands is monitored until either an abasic site is reached (at which point the synthesis is paused) or until the incorporation becomes asynchronous (which increases the background noise and decreases reliability of the sequence read). In the latter case, the practioner may opt to speed up the reaction, e.g., by adding all nucleotides at one time, to extend all nascent strands to the first abasic site in the templates. When synthesis is to be reinitiated, reaction components are added that allow bypass of the abasic site, e.g. one or more pyrenes. A wash step may be performed to remove nucleotides and/or polymerases from the reaction sites prior to such addition. Further, in some cases, a different polymerase may be used for pyrene incorporation as is used for sequencing-by-synthesis reactions. In certain preferred embodiments, the reaction mixture comprising the pyrene for abasic site bypass allows readthrough of the abasic site, but no further on the template. Subsequent addition of sequencing reaction mixture allows the sequencing-by-synthesis reaction to recommence and incorporation of nucleotides into the nascent strand to be monitored. Alternatively or additionally, the practioner need not wait until an abasic site is reached to suspend detection and, optionally, speed up the reaction to bring all nascent strands to a given abasic site, but can choose to do this before a reaction

has become asynchronous, e.g., after desired sequence data has been collected for a particular region of interest in a template nucleic acid.

[00213] In certain embodiments using ligation-based technologies (e.g., the SOLiD™ System developed by Life Technologies), a pause site can be engineered by using an oligonucleotide that cannot participate in the ligation reaction and that is complementary to a desired location on the set of identical template nucleic acids, e.g., on a bead. When the serial ligation reaction hits the position recognized by this polynucleotide, the reaction cannot proceed and any reactions that have become asynchronous will "catch up." The user can then unblock the oligo (e.g., using chemical treatment or photo-cleavage) and reinitiate the sequencing reaction.

[00214] In some cases, it may be desirable to provide endonuclease recognition sites within the template nucleic acid. For example, inclusion of such sites within a circular template can allow for a mechanism to release the template from a synthesis reaction, i.e., by linearizing it, and allowing the polymerase to run off the linear template, and/or to expose the template to exonuclease activity, and thus terminate synthesis through removal of the template. Such sites could additionally be exploited as control sequences by providing specific binding locations for endonucleases engineered to lack cleavage activity, but retain sequence specific binding, and could therefore be used to block progression of the polymerase enzyme on a template nucleic acid.

[00215] In some cases, nicking sites, e.g., sites recognized by nicking endonucleases, may be included within a portion of the template molecule, and particularly within a double-stranded portion of the template, e.g., in a double-stranded segment of a SMRT bell™ or in the stem portion of an exogenous hairpin structure. Such nicking sites provide one or more breaks in one strand of a double-stranded sequence and can thereby provide one or more priming locations for, e.g., a strand-displacing polymerase enzyme. A variety of nicking enzymes and their recognition sequences are known in the art, with such enzymes being generally commercially available, e.g., from New England Biolabs.

[00216] In certain embodiments, methods for intermittent detection described herein are useful in "paired-end" sequencing applications in which sequence information is generated from two ends of a template nucleic acid but not for at least a portion of the intervening portion of the template. Typically, paired-end sequencing applications provide sequence data for only the two ends of a nucleic acid template, but the present invention also allows generation of additional sequence reads that are noncontiguous with the sequence reads from the ends of the template. In certain preferred embodiments, a duplex fragment (e.g., genomic fragment) is ligated to a single-

stranded linker that connects the 3' end of the sense strand to the 5' end of the antisense strand, or that connects the 5' end of the sense strand to the 3' end of the antisense strand. In either orientation, separation of the two strands of the duplex fragment results in a single-stranded linear template nucleic acid that contains the linker in between the sense and antisense strands. Subsequent sequencing can involve intermittent detection that generates sequence reads for only the portions of the sense and antisense strands that are of interest, e.g., one or both of the ends. In certain embodiments, both sense and antisense strands may be sequenced at both ends to provide redundancy in the sequence data. Sequence reads recognized as being from the linker portion of the template (e.g., based on the known linker sequence or specific registration sequences encoded therein) can be used to orient the alignment of the sequence reads from the sense and antisense portions of the template, providing context for determining the sequences of the ends of the duplex fragment and subsequent sequence scaffold construction and/or mapping. In certain embodiments, pause or stop points may be incorporated into the linker to control the processing of the template by the polymerase, and therefore may be used to synchronize the detection periods to ensure generation of sequence reads from particular regions of template. Further, additional detection periods can be included that are timed to provide sequence reads from portions of the sense and/or antisense strand that are noncontiguous with the end regions.

[00217] In a related embodiment, paired-end sequencing may be accomplished by using a nucleic acid template that has linkers connecting the sense and antisense strands of a duplex fragment at both ends, such that separation of the strands of the duplex fragment provides a single-stranded circular template that contains a linkers in between each end of the sense and antisense strands of the original duplex fragment. Such a template molecule would allow a strand-displacing polymerase to proceed around the template multiple times, thereby potentially generating redundant sequence data from both ends of both strands of the original duplex fragment. As noted elsewhere herein, such redundancy is useful for determination of consensus sequences and/or construction of sequence scaffolds. As the polymerase enzyme processes the template, detection periods can be timed (e.g., based on knowledge of the rate at which the polymerase processes the template, which is dependent not only on the polymerase but also on the sequence of the template itself) to generate nucleotide sequence reads from the regions of the template corresponding to one or both ends of the sense and antisense strands, and can also include detection periods to generate additional reads from other, noncontiguous regions of the duplex fragment, as well. Although such timing can be used to determine the appropriate periodicity of the detection periods, at later stages of the reaction (e.g., as

the polymerase repeatedly proceeds around the template), the exact location of reinitiation of sequence read generation becomes more approximate. Incorporation of pause or stop points into one or both linkers to regulate the processing of the template by the polymerase may be used to synchronize the detection periods regardless of the total distance travelled by the polymerase around the template. This strategy more reliably ensures generation of sequence reads from selected regions of template, e.g, the ends of the sense and antisense portions and, optionally, regions in between and noncontiguous with the end regions regardless of the number of passes of the polymerase around the template nucleic acid, especially in later stages of the reaction. Further, the known sequence of one or both of the linkers can be used to orient sequence reads from the sense and antisense portions for consensus sequence determination and/or mapping.

[00218] In some such embodiments, a duplex fragment inserted between two hairpin linkers may be much larger than desired, increasing the difficulty of limiting nucleotide sequence read data to particular regions of the fragment. The size of the duplex fragment ligated to the two hairpin linkers can be selectively reduced to retain the regions attached to the linkers and to lose a central portion of the duplex fragment. One particularly preferred strategy, illustrated in Figure 18, comprises hairpin linkers (1802, 1804) having a regions of cross-complementarity (1806, 1808), such that the two linkers 1802 and 1804 can anneal to each other in a manner that does not interfere with ligation to a duplex fragment 1810. Duplex fragment 1810 comprises ends 1812 and 1814, as well as a long central region 1816, which is not shown but is understood to be between the two curvy lines. Once end 1812 is ligated to linker 1802 and end 1814 is ligated to linker 1804, the construct is subjected to fragmentation, which removes the central region 1816 of the duplex fragment 1810, producing construct 1818 having ends 1820 and 1822. After fragmentation, the ends of the portions of the duplex fragment still associated with the annealed linker pair (ends 1820 and 1822) are ligated together to produce construct 1824, which can then be treated (e.g., with heat, gentle denaturation, primer invasion, changing salt concentration, etc.) to separate cross-complementary regions 1806 and 1808 from one another, e.g., to generate a circular single-stranded nucleic acid molecule. Alternatively, the separation may occur during the course of the subsequent reaction, e.g., by polymerase-mediated strand displacement. Yet further, where the hybridized cross-complementary regions are long enough to undergo a complete DNA turn, an additional reaction component (e.g., helicase, topoisomerase, polymerase, etc.) may be needed to unwind the duplex and allow separation. As such, the resulting "mate-pair" construct has only the ends of the original duplex fragment ligated together and capped with adaptors that link the 5' end of each

strand of the duplex with the 3' end of the other strand of the duplex, and denaturation of the duplex produces a closed, single-stranded circular construct.

[00219] Fragmentation of the duplex fragment can be performed by a variety of known methods. For example, fragmentation can be performed enzymatically (e.g., using restriction enzymes or other nucleases) or mechanically, by shearing or sonication. The type of fragmentation chosen will determine various characteristics of the resulting construct, e.g., how large a central region is removed and the types of ends remaining (e.g., blunt, 5' overhang, 3' overhang, random, identical on both ends, etc.). Optionally, the ends can be modified after fragmentation to facilitate the subsequent ligation step. Although not shown in Figure 18, it is expected that the ligation of the duplex fragment to the hybridized linkers will be a two-step process, with one end being ligated first and unimolecular kinetics favoring ligation of the second end to the second linker. The cross-complementary regions of the linkers can be designed to produce varying levels of complementarity, and therefore varying strengths of the hybridization. For example, a longer or higher GC content in a cross-complementary region lends a higher stability to the linker:linker interaction, but separation of the hybridized linkers requires a more severe treatment, e.g., higher temperature, more stringent conditions, etc. As such the cross-complementary regions should be engineered to produce a stable linker:linker interaction that is disruptable under conditions that are not destructive to the overall construct. Further the linkers can vary in regions apart from the cross-complementary regions For example, one linker can have a primer binding site that the other lacks, which would provide a single polymerase initiation site in the final construct. Other sequence characteristics described herein (e.g., pause sites, registrations sequences, etc.) can also be included in one or both linker regions. If topological constraints limit the subsequent processing of the resulting contruct, e.g., during template-directed nascent strand synthesis, these can be addressed by addition of a reaction component (e.g., a helicase or topoisomerase) to resolve the topological constraint. As such, the methods can be used to add asymmetric linkers to duplex polynucleotides, whether or not the duplex is to be selectively reduced in size, or not, as long as the asymmetric linkers can cross-hybridize to one another.

[00220] Although in preferred embodiments, the two linkers to be ligated to a single duplex fragment are hybridized to one another prior to ligation, in some embodiments they are instead hybridized after the initial ligation reaction, and where topological constraints inhibit such a post-ligation hybridization a reaction component (e.g., topoisomerase) may be included to relieve such constraints. In certain embodiments, the hybridized linkers are separated prior to addition of

reaction components for a subsequent reaction, and in other embodiments the hybridized linkers are not separated until after the addition of reaction components for a subsequent reaction. For example, a polymerase enzyme may bind to a primer annealed to a linker before or after separation of the linker from a second linker. In fact, it may be benefical in some embodiments to postpone separation of the linkers, e.g., where compaction of the nucleic acid construct is beneficial, such as when the construct must be loaded into a confinement of some kind, e.g., a nanowell, optical confinement, etc.

[00221] In some embodiments, the methods further include separation of single linker constructs from hybridized linker pair constructs. This can be accomplished by an exonuclease treatment after ligation of the duplex fragment to the linkers, which would degrade any constructs having an unannealed end. Alternatively, it may be desirable to remove the single linkers prior to ligation, for example using a size separation methodology or by allowing them to bind to oligonucleotides that are complementary to the cross-complementary regions and bound to a column or magnetic beads. (The cross-complementary regions of the hybridized linker pairs will not be available for binding to the oligonucleotides. Other methods known in the art can also be used to separate single linkers from hybridized linker pairs.

[00222] Interestingly, the use of the sense/antisense nucleic acid templates described above would represent a unidirectional processing of a template to provide paired-end sequence data, as opposed to the more traditional bi-directional processing of a linear template molecule. Further, unlike traditional approaches, these methods for paired-end sequencing involve processing, chemically or otherwise, of not just the regions at the ends, but also regions in between the ends, and in some embodiments comprising processing of the entire template. For example, a polymerase incorporates nucleotides into a nascent strand for each position of the template (thereby "processing" each position of the template), yet the sequencing data generated is limited to specific regions of the template that are of particular interest to the practitioner, such as the end regions. As such, in certain embodiments the duplex fragment is not further reduced in size after ligation to a linker pair, and the entire duplex fragment is processed by the polymerase.

[00223] In certain embodiments, methods for intermittent detection described herein are useful in analysis systems that employ nanopores. A nanopore is a small pore in an electrically insulating membrane that can be used for single molecule detection. In general, a nanopore functions as a Coulter counter for much smaller particles, and can take various forms, e.g., a protein channel in a lipid bilayer or a pore in a solid-state membrane. The detection principal is based on

monitoring the ionic current of an electrolyte solution passing through the nanopore as a voltage is applied across the membrane. For example, passage of a polynucleotide molecule (e.g., DNA, RNA, etc.) through a nanopore causes changes in the magnitude of the current through the nanopore, with each nucleotide obstructing the nanopore to a different, characteristic degree. As such, the pattern of variations in the current passing through the nanopore as the polynucleotide is drawn through may be monitored and analyzed to determine the nucleotide sequence of the polynucleotide. A polynucleotide may be drawn through the nanopore by various means, e.g., by electrophoresis, or using enzyme chaperones to guide the polynucleotide through the nanopore. For additional discussion of methods of fabrication and use of nanopores, see, e.g., U.S. Patent No. 5,795,782; Kasianowicz, J.J., et al. (1996) *Proc Natl Acad Sci USA* 93(24):13770-3; Ashkenas, N., et al. (2005) *Angew Chem Int Ed Engl* 44(9):1401-4; Winters-Hilt, S., et al. (2003) *Biophys J* 84:967-76; Astier, Y., et al. (2006) *J Am Chem Soc* 128(5):1705-10; Fologea, D., et al. (2005) *Nano Lett* 5(10):1905-9; Deamer, D.W., et al. (2000) *Trends Biotechnol* 18(4):147-51; and Church, G.M. (2006) *Scientific American* 294(1):52, all of which are incorporated by reference herein in their entireties for all purposes. In some embodiments, intermittent detection of nucleic acid sequence data from a nanopore may be achieved by modifying the progress of the polynucleotide through the nanopore so that progress is sped up during non-detection periods and progress is slowed to allow sequence determination during detection periods. The rate of passage of the polynucleotide through the nanopore may be modified by various methods, including but not limited to increasing an electrophoretic field carrying the polynucleotide (e.g., by increasing the voltage, changing the conductivity of the reaction mixture, and the like), or changing various reaction conditions to alter the speed at which a protein chaperone carries the polynucleotide. Further, in embodiments utilizing a processive exonuclease to feed individual bases through the nanopore, the kinetics of the exonuclease may be modified based on the known biochemical characteristics of the exonuclease.

[00224] In diagnostic sequencing applications, it may be necessary only to provide sequence data for a small fragment of DNA, but do so in an extremely accurate sequencing process. For such applications, shorter target segments may be employed, thus permitting a higher level of redundancy by sequencing multiple times around a smaller circular template, where such redundancy provides the desired accuracy. Thus, in some cases, the double stranded target segment may be much shorter, e.g., from 10 to 200, from 20 to 100 or from 20 to 50 or from 20 to 75 bases in length. For purposes of the foregoing, the length of the target segment in terms of bases denotes the length of one strand of the double stranded segment. In such applications, various methods for

intermittent detection described herein may be used to analyze the sequence of the template, thereby targeting the sequence data to the portion(s) of the template of particular interest to the diagnostician, and/or improving various aspects of the reaction performance, e.g., by virtue of the reduction of photo-induced damage to one or more reaction components.

[00225] It is to be understood that the above description is intended to be illustrative and not restrictive. It readily should be apparent to one skilled in the art that various embodiments and modifications may be made to the invention disclosed in this application, including but not limited to combinations of various aspects of the invention, without departing from the scope and spirit of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. All publications mentioned herein are cited for the purpose of describing and disclosing reagents, methodologies and concepts that may be used in connection with the present invention. Nothing herein is to be construed as an admission that these references are prior art in relation to the inventions described herein. Throughout the disclosure various patents, patent applications and publications are referenced. Unless otherwise indicated, each is incorporated by reference in its entirety for all purposes.

[00226] Although described in some detail for purposes of illustration, it will be readily appreciated that a number of variations known or appreciated by those of skill in the art may be practiced within the scope of present invention. Unless otherwise clear from the context or expressly stated, any concentration values provided herein are generally given in terms of admixture values or percentages without regard to any conversion that occurs upon or following addition of the particular component of the mixture. To the extent not already expressly incorporated herein, all published references and patent documents referred to in this disclosure are incorporated herein by reference in their entirety for all purposes.

[00227] The following non-limiting examples are provided to further illustrate the invention.

VI. Examples of Intermittent Illumination of a Single Molecule Sequencing-by-Synthesis Reaction

Example I

[00228] A nucleic acid template was provided that comprised a double-stranded region and two single-stranded linker portions at each end. The first linker portion connected the 3' end of the

sense strand with the 5' end of the antisense strand, and the second linker portion connected the 3' end of the antisense strand with the 5' end of the sense strand. This template was designed to form a single-stranded circle of approximately 500 bases when the double-stranded region was opened (e.g., by heat denaturization, helicase activity, etc.), and is sometimes referred to as a SMRTbell™ template. A plurality of this nucleic acid template was incubated with polymerases, primers, and other reaction components to allow formation of polymerase-template complexes. (See, e.g., Korlach, J., et al. (2008) Nucleosides, Nucleotides and Nucleic Acids, 27:1072-1083; and Eid, J. (2009) Science 323:133-138.) The complexes were immobilized in zero-mode waveguides in a reaction mixture containing all necessary buffer and nucleotide analog components for carrying out sequencing-by-synthesis reactions with the exception of a cognate starting base and a metal dication. A Smith-Waterman algorithm was used to perform the alignment of the known sequence of the template with the sequence reads generated in the reaction, and the positions of the sequence reads is graphically illustrated in Figure 8.

[00229] Acquisition of the data shown in Figure 8 was collected as follows. Illumination of the array of zero-mode waveguides was initiated with laser excitation (532 nm and 641 nm laser lines) at t = -5 seconds, and the missing cognate starting base and metal dication (manganese metal) were added at t = 0 seconds to simultaneously initiate the sequencing-by-synthesis reactions in all zero-mode waveguides. The reactions were monitored under illumination for 120 seconds at which time the illumination was removed; the sequencing reads generated during that stage of the reaction are shown in Figure 8A as a function of the template position to which each read maps. At 295 seconds illumination was resumed and data acquisition was reinitiated at 300 seconds and maintained for another 120 second interval; the sequencing reads during this second illuminated period are shown in Figure 8B. At 595 seconds illumination was resumed and data acquisition was reinitiated at 600 seconds and maintained for another 120 second interval; the sequencing reads during this third illuminated period are shown in Figure 8C.

[00230] As expected, the longer the amount of time before the sequence data is collected (that is, the later the illuminated period), the further into the template the alignments shift, and this shift is a rough function of time since initiation of the reaction. Further, the distribution of sequence reads generated during each subsequent illuminated period becomes more dispersed than the previous illuminated period(s). Further, due to the circular nature of the template, Figure 8C clearly shows that some polymerases have passed completely around the substrate and are beginning to

generate sequence reads from a second pass around the template, thereby generating redundant sequence information for a single template nucleic acid.

Example II

[00231] As in Example I, a SMRTbell™ template was used. For templates of defined sequence, PCR was used to generate 3 or 6 kb DNA inserts for the double-stranded region in the SMRTbell™ templates using a standard PCR methodology. For genomic and other biological samples, a DNA fragmentation protocol was used that generates DNA fragments distributed around 3 or 6 kb. Generation of fragments in these ranges was done using a HydroShear® (Genomic Solutions®) device with settings recommended by the manufacturer. The random genomic DNA fragments were enzymatically treated to generate blunt ends. Both the PCR products and randomly generated DNA fragments were phosphorylated and then immediately put into a ligation reaction with a blunt hairpin adapter. The products were purified through two size selection steps using reduced volumes of AMPure® magnetic beads (Agencourt®) to remove hairpin dimers and other short products. (Fabrication of SMRTbell™ templates is further described elsewhere herein.)

[00232] The system components used for polynucleotide sequencing using intermittent detection are comparable to single-molecule sequencing applications under constant illumination, which are described, e.g., in Eid, et al. (2009) Science 323:133-138. Specifically, the immobilization and sequencing buffer compositions, nucleotide analogs identity and concentration, polymerase, ZMWs, surface treatment and instrumentation were identical to the standard methodology. Modifications to the SMRTbell™ template DNA and polymerase binding and immobilization and data acquisition protocols are as follows.

[00233] A binding solution was prepared by incubation of 3 or 6 kb DNA SMRTbell™ templates (1-10 nM) with a 10-fold excess of DNA polymerase (10-100 nM, respectively) in 10 mM MOPS (pH 7.5), 10 mM KOAc, 100 mM DTT & 0.05% Tween-20 for 2 hours at 30°C, followed by 1 hour at 37°C and subsequent storage at 4°C prior to immobilization on the ZMWs. Immediately prior to immobilization, the binding solution was diluted in the standard immobilization solution (50 mM MOPS (pH 7.5), 75 mM KOAc, 5 mM DTT, 0.05% Tween-20) to the desired final concentration, typically 0.1 to 1 nM, and incubated for 30 to 60 minutes at 22°C. Post-immobilization chip preparation and sequencing initiation were identical to the standard methods.

[00234] The data acquisition protocol was similar to the standard application with coordinated modifications to the collection timing and ZMW positioning. In the standard acquisition procedure, a single long acquisition (~10 minutes) is performed for each ZMW. In the intermittent illumination acquisition procedure, multiple short acquisitions (~3 minutes) of sequence reads (also termed "strobe reads") were performed for each ZMW (during "detection periods") with an interval between each acquisition period during which no acquisition of sequence reads was performed ("non-detection period"). The duration of the interval between each acquisition of sequence reads was determined based upon a desired distance (i.e., number of nucleotide positions) between each sequence (or strobe) read, the polymerization rate of the polymerase, and the SMRTbell™ template insert size.

[00235] SMRTbell™ templates were generated as described above for AC223433, a fosmid clone comprising a sequence of an approximately 40 kb region of *Homo sapiens* chromosome 15. The reference sequences used to map the sequence reads generated in the sequencing reactions were the publically available sequences of *Homo sapiens* chromosome 15 (Hg18; NCBI Build 36.1) and fosmid AC223433 (NCBI GenBank accession number). Table 1 shows the number of statistically significantly mapped sequence reads for several types of intermittent illumination sequencing reactions. The number of mappable "looks" is equivalent to the number of mappable sequence reads generated during detection periods for a single template molecule. For example, a "mapped 1-look read" means, for a single template molecule, only a single detection period generated a sequence read that could be mapped to the reference sequence.

Table 1: Summary of Sequencing Results

Mapping Reference	Mapped 1-look reads	Mapped 2-look reads	Mapped 3-look reads	Mapped 4-look reads
Human chr15 (Hg18)	13834	1289	127	4
Fosmid	15253	1571	158	5

[00236] Deviations in the expected time span for a set of sequencing reads from a single sequencing reaction are indicative of genomic events such as genomic rearrangements, e.g., insertions, deletions, etc. Figures 10 and 11 illustrate this point. Specifically, the time and distance travelled along the template (based upon the reference sequence) by the polymerase was computed within and between the sequence reads generated during the detection (illuminated) periods. These calculations were used to detect unexpected variations, indicating possible genomic events in the

template as compared to the reference sequence. Figure 10 provides a plot that illustrates the normalized average time it took for the polymerase to traverse a region of the template based on the length of that region in the *Homo sapiens* chromosome 15 reference sequence. The sequence reads are fit to a diagonal having a slope equal to the average speed for sequencing reads. Deviations from the regressed diagonal indicate genomic events (for example, structural variants), and the slope of the sequence reads around such deviations indicate the relative size of the genomic event (e.g., in the case of insertions/deletions). For example, if the time for the polymerase to traverse a region was unexpectedly long, this indicated the polymerase actually traversed a longer region than was expected based on the reference sequence. The two distinct off-diagonal deviations (upper right hand corner) with higher slope indicated that an insertion had occurred in the reference sequence, and this was verified by comparison to the known fosmid sequence.

[00237] Figure 11 shows the average time it took the polymerase to traverse the template. For each mapped read, starting and ending times and positions were determined and used to compute the distance traversed by the polymerase between sequence reads. Based on these determinations, an average time across any particular region of the human reference sequence was computed. Regions that were traversed by the polymerase more slowly have peaks of higher ΔT, and were indicative of insertions in the template relative to the *Homo sapiens* chromosome 15 reference sequence. The insertions identified were the same insertions identified above.

[00238] Intermittent illumination-based sequencing reactions across fosmid sequence AC223433 showed significant sequence read coverage across the insertion events. The distribution of the physical coverage is shown in Figure 12, which illustrates examples of three-look strobes (i.e., sequencing reactions having three detection/illuminated periods) that span or intersect the insertion events. Figure 12A shows the mapping of the strobe sequence reads to the *Homo sapiens* chromosome 15 reference sequence, where the sequence reads generated from the insert sequences in the template are excluded. Arrows indicate the locations of the insertions. Figure 12B shows a similar mapping with the sequence reads generated from the insert sequences indicated with brackets. A number of sequence reads flank the insertions, connect the two insertions, or clarify sequence within (or at the boundaries of) the insertion sequences. Such flanking and connecting sequence reads are useful for predicting and detecting genomic events, anchoring them to genomic references, and scaffolding for de novo assembly of novel sequences. In particular, there are 30 and 38 "3-look" reads that intersect the two regions of insertion of (1192 bp and 6879 bp, respectively). These sequence reads facilitated mapping of the insertions to the human reference sequence, which

would have been extremely difficult, if not impossible, with commercially available short-read sequencing technologies. Further the sequence of the smaller insertion was a highly repetitive sequence, which would also have made mapping difficult with certain short-read technologies.

[00239] Figure 13 illustrates the sequence coverage obtained across the fosmid sequence, showing all two-, three-, and four-look strobe sequence reads spanning the sequence that are mappable to the known AC223433 fosmid sequence.

[00240] A consensus sequence was derived from the set of mappable sequence reads generated in these sequencing reactions. Strobe sequence reads were combined with sequence reads generated under constant illumination and assembled based on the human reference sequence (Hg18). High quality reads surrounding the (suspected) insertion sites, as well as high quality reads that did not map to the reference sequence, were extracted and assembled with a "de novo" greedy suffix tree assembler; the resulting contigs were mapped to the Hg18 reference sequence. Contigs spanning the (suspected) insertion sites were identified and fed back into the "de novo" assembler, and the resulting contigs were manually edited using standard techniques and placed back into the derived reference guided assembly. The final consensus sequence was a hybrid of a reference guided assembly and attempts at de novo assembly of novel insert sequences. Alignments to reference sequences were performed and plotted. Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described above and the fosmid reference sequence, and this plot confirmed a high degree of alignment between the two sequences. This dot plot was generated using Gepard 1.21 ("GEnome PAir – Rapid Dotter," available from the Munich Information Center for Protein Sequences (MIPS)) with a word size of 7. Nucleic acid dot plots are widely used in the art and are further described, e.g., in Krumsiek et al. (2007) Bioinformatics 23(8):1026-8; Maizel et al. (1981) Proc Natl Acad Sci USA 78:7665; Pustell, et al. (1982) Nucleic Acids Res 10:4765; and Quigley, et al. (1984) Nucleic Acids Res 12:347, all of which are incorporated herein by reference in their entireties for all purposes.

What is claimed is:

1. A method of determining a nucleotide sequence of a region of interest in a polynucleotide, the method comprising:

introducing a polynucleotide comprising a region of interest to a sequence analysis system comprising a nanopore and an enzyme chaperone;

monitoring variations in a parameter of the nanopore of the sequence analysis system during passage of the polynucleotide through the nanopore, wherein the enzyme chaperone regulates the rate of passage of the polynucleotide through the nanopore;

analyzing the monitored variations of the nanopore of the sequence analysis system to determine a nucleotide sequence of the polynucleotide, wherein the nucleotide sequence comprises redundant sequence information for the region of interest; and

determining a consensus sequence for the region of interest based on the redundant sequence information.

2. The method of claim 1, wherein the nanopore comprises a protein channel.

3. The method of claim 1, wherein the nanopore is in a membrane.

4. The method of claim 3, wherein the membrane is a lipid bilayer.

5. The method of claim 3, wherein the membrane is a solid-state membrane.

6. The method of claim 3, wherein a voltage is applied across the membrane.

7. The method of claim 6, wherein the parameter in the monitoring step is the ionic current of an electrolyte solution passing through the nanopore.

8. The method of claim 1, further comprising changing reaction conditions to alter the speed at which the enzyme chaperone guides the polynucleotide through the nanopore.

9. The method of claim 1, wherein the polynucleotide comprises complementary sequences of the region of interest, wherein the redundant sequence information comprises the nucleotide sequence of the complementary sequences.

10. The method of claim 9, wherein the complementary sequences are linked by an oligonucleotide.

11. The method of claim 10, wherein the oligonucleotide comprises a registration sequence.

12. The method of claim 1, wherein the polynucleotide comprises double-stranded DNA.

13. The method of claim 12, wherein the polynucleotide is greater than 75% double-stranded DNA.

14. The method of claim 12, wherein the polynucleotide is greater than 90% double-stranded DNA.

15. The method of claim 1, wherein the polynucleotide comprises a double-stranded portion comprising complementary sequences of the region of interest, wherein the redundant sequence information comprises the nucleotide sequence of the complementary sequences.

16. The method of claim 15, wherein the complementary sequences are linked by an oligonucleotide.

17. The method of claim 16, wherein the oligonucleotide comprises a registration sequence.

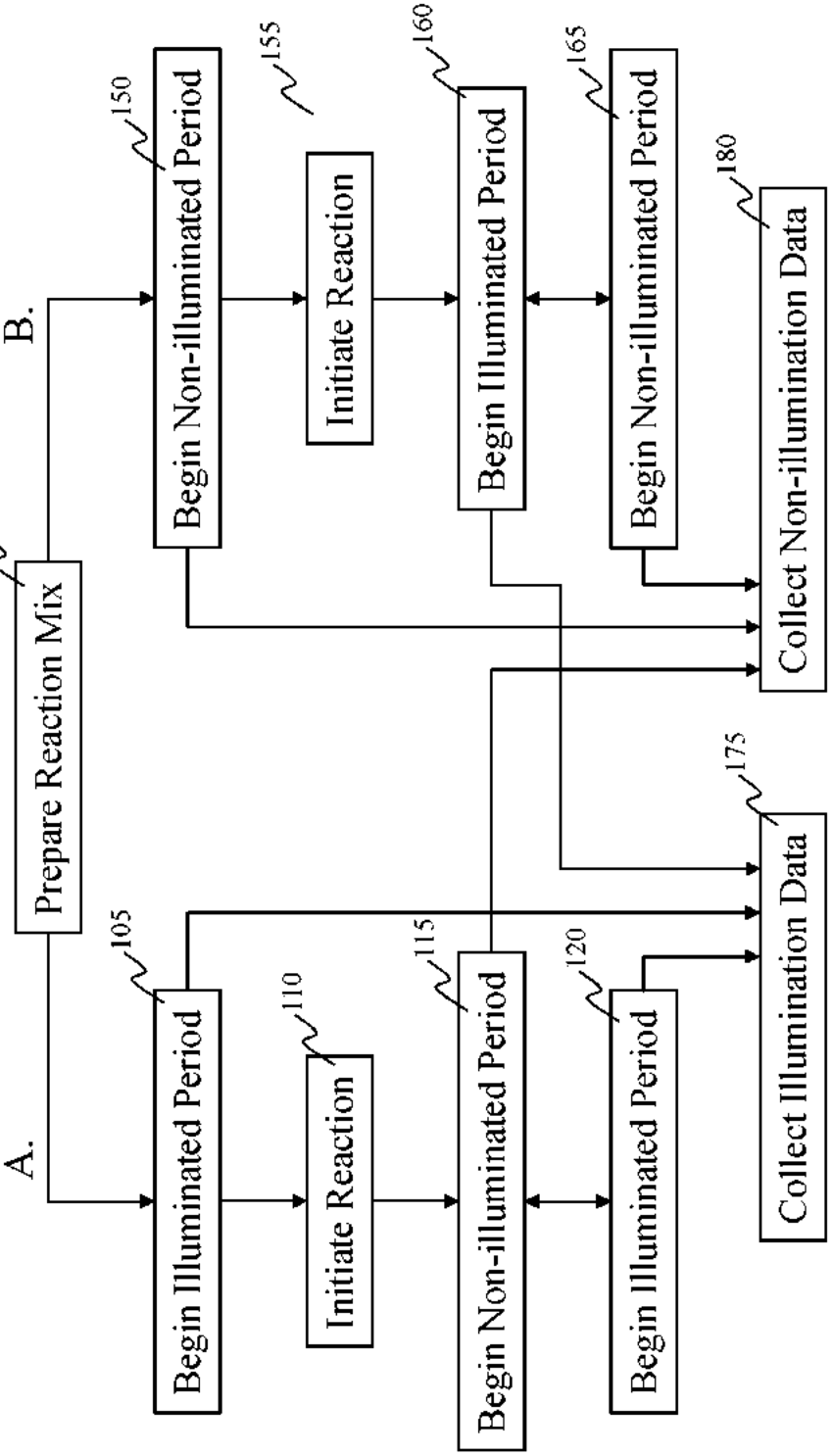
18. The method of claim 1, wherein the polynucleotide comprises multiple repeats of the region of interest, wherein the redundant sequence information comprises the nucleotide sequence of the multiple repeats.

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

ABSTRACT

Methods, devices, and systems for performing intermittent detection during analytical reactions are provided. Such methods facilitate collection of reaction data from disparate reaction times. Further, such methods are useful for reducing photo-induced damage of one or more reactants in an illuminated analytical reaction at a given reaction time. In preferred embodiments, the reaction mixture is subjected to at least one illuminated and non-illuminated period and allowed to proceed such that the time in which the reaction mixture is illuminated is less than a photo-induced damage threshold period.

Figure 1


Prepare Reaction Mix
100
A.
B.
Begin Illuminated Period
105
Initiate Reaction
110
Begin Non-illuminated Period
115
Begin Illuminated Period
120
Begin Non-illuminated Period
150
Initiate Reaction
155
Begin Illuminated Period
160
Begin Non-illuminated Period
165
Collect Non-illumination Data
180
Collect Illumination Data
175

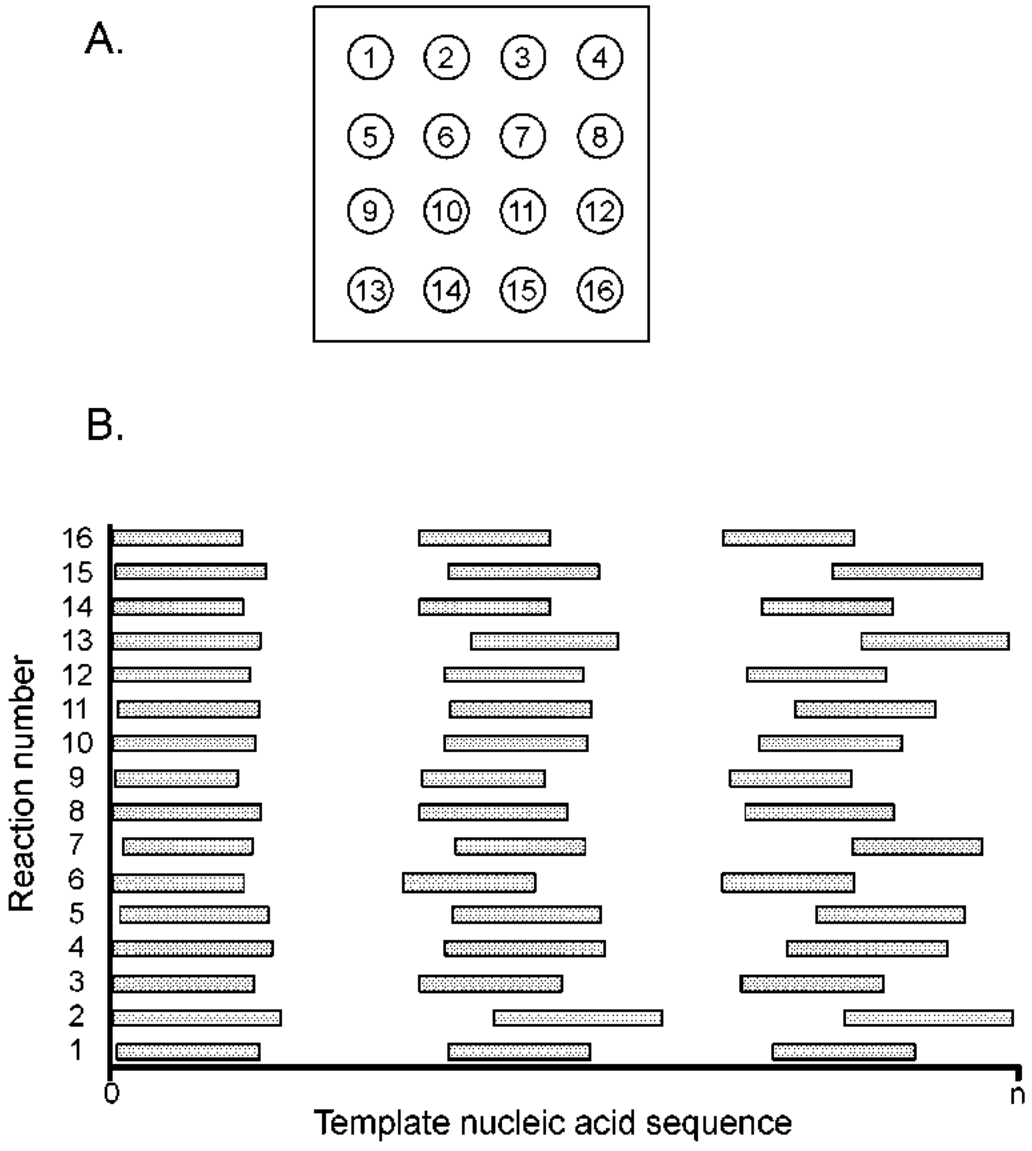

A.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
B.
Reaction number
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
n
Template nucleic acid sequence

Figure 2

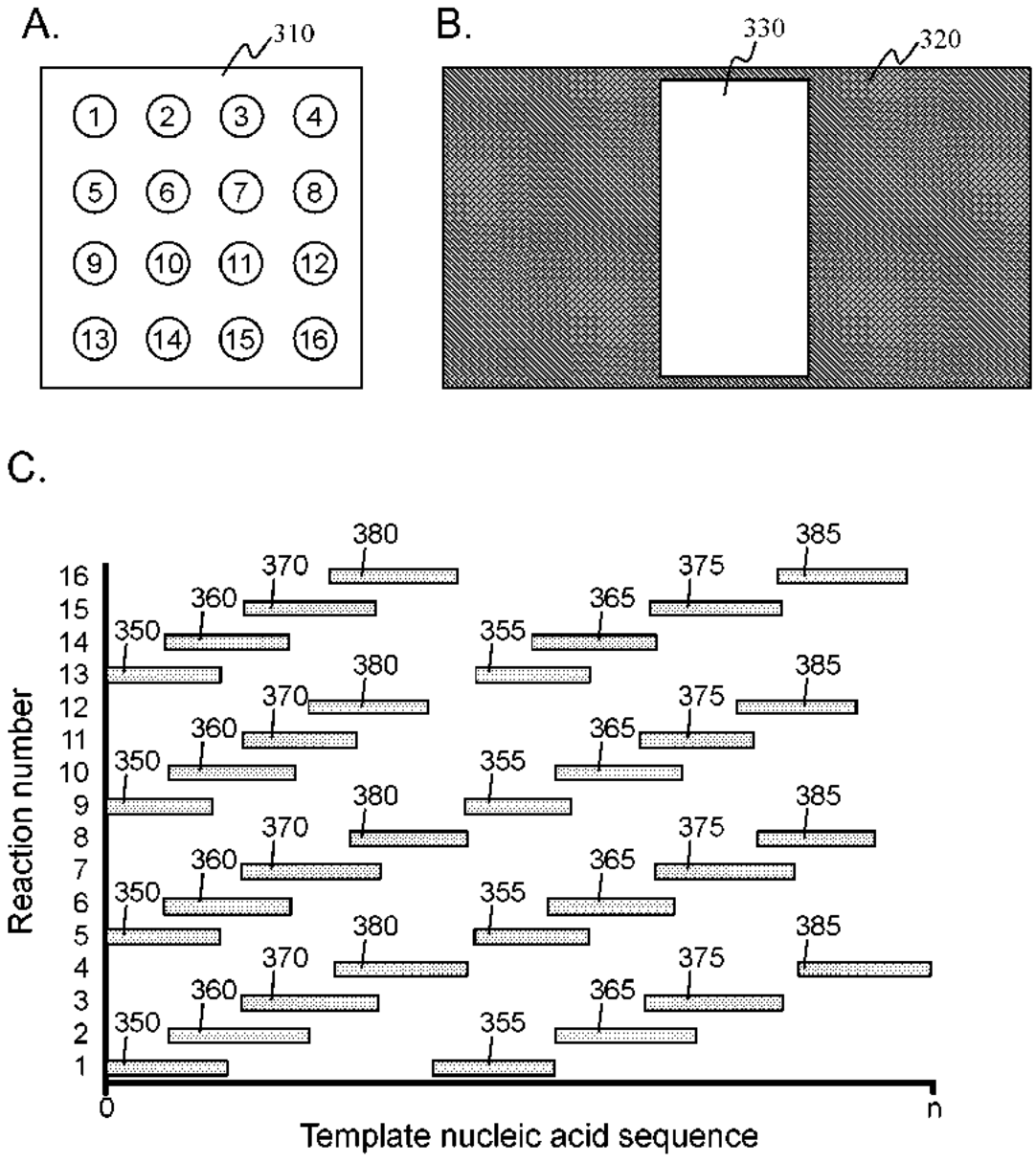
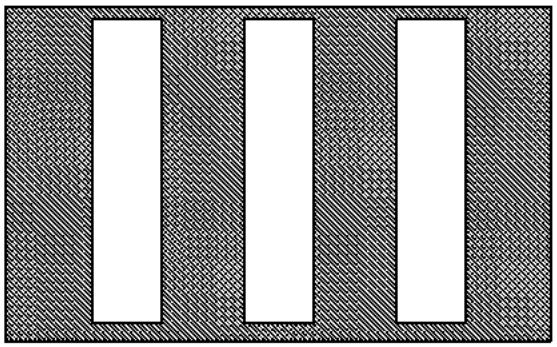

A.
310
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
B.
330
320
C.
Reaction number
Template nucleic acid sequence
0
n
350
355
360
365
370
375
380
385

Figure 3

A.



B.

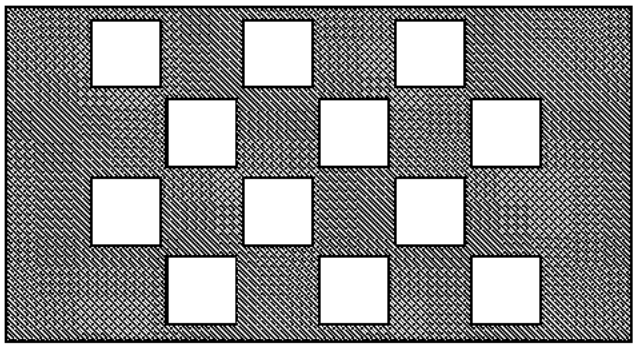


Figure 4

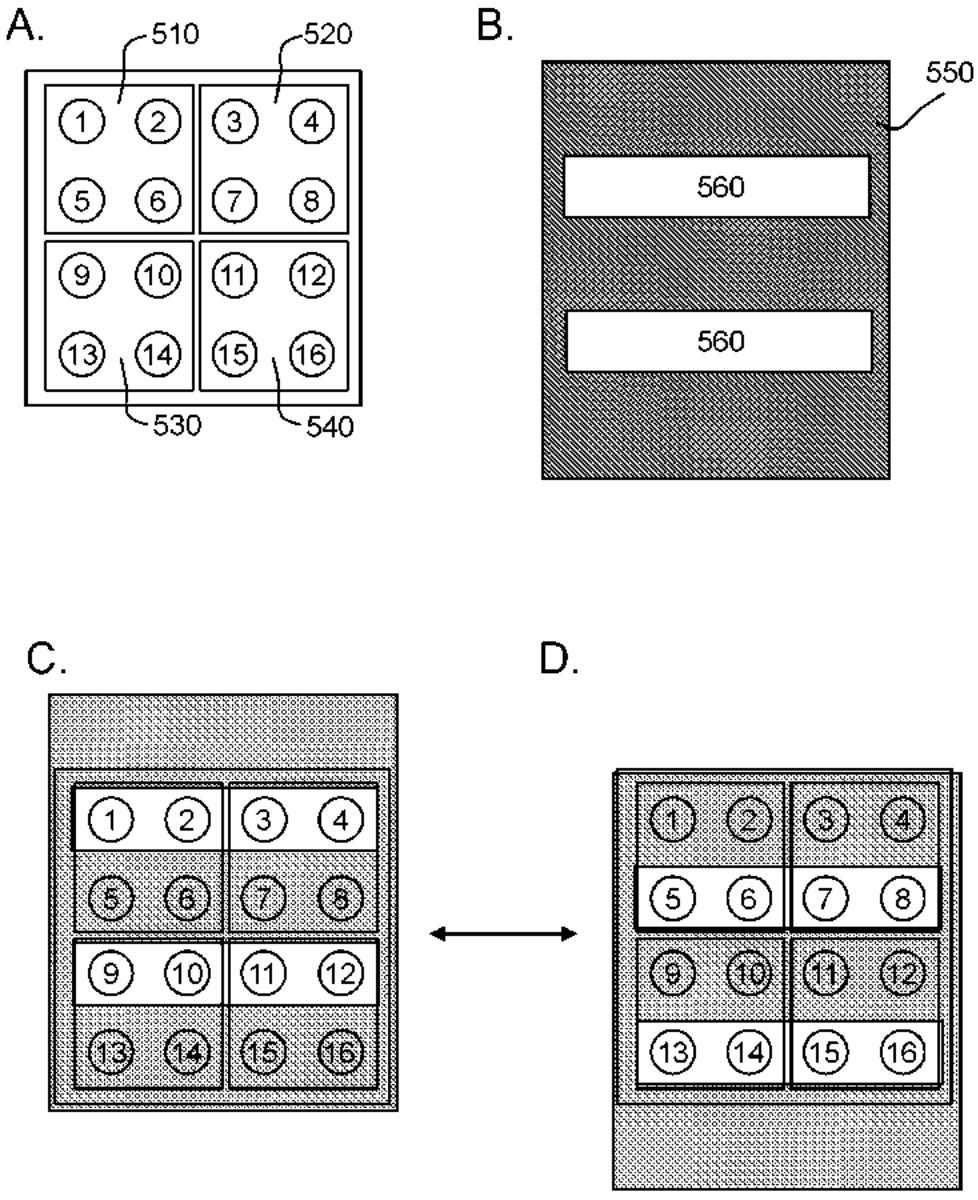


Figure 5

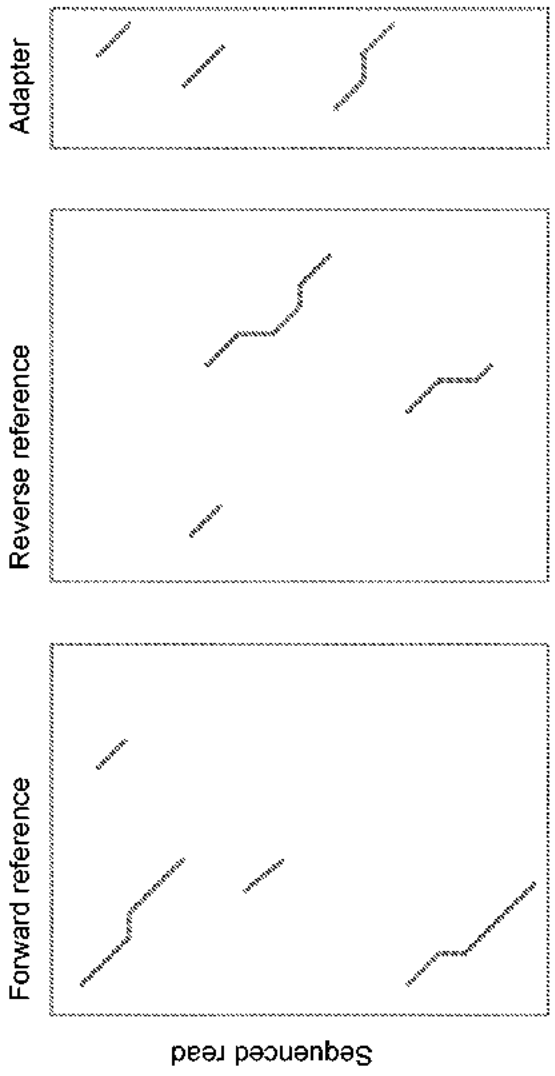
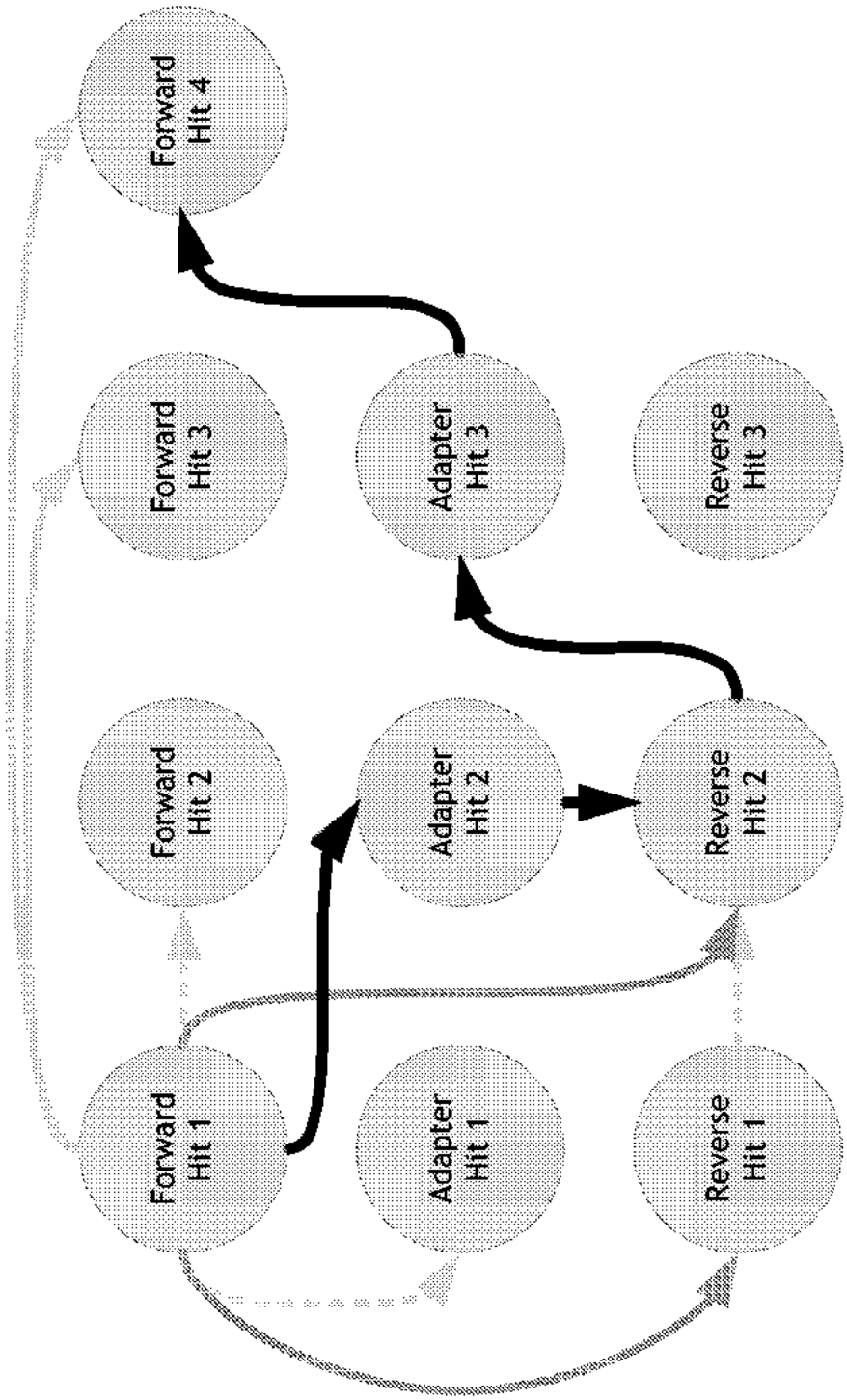

Adapter
Reverse reference
Forward reference
Sequenced read

Figure 6

Figure 7


Forward Hit 4
Forward Hit 3
Adapter Hit 3
Reverse Hit 3
Forward Hit 2
Adapter Hit 2
Reverse Hit 2
Forward Hit 1
Adapter Hit 1
Reverse Hit 1

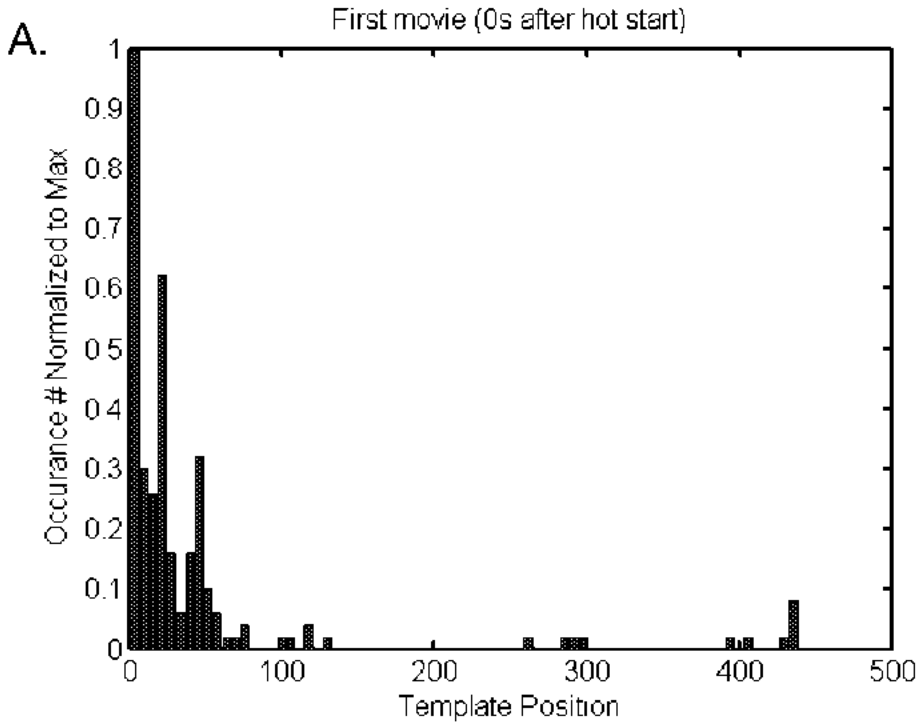

A.
First movie (0s after hot start)
Occurance # Normalized to Max
Template Position
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500

Figure 8

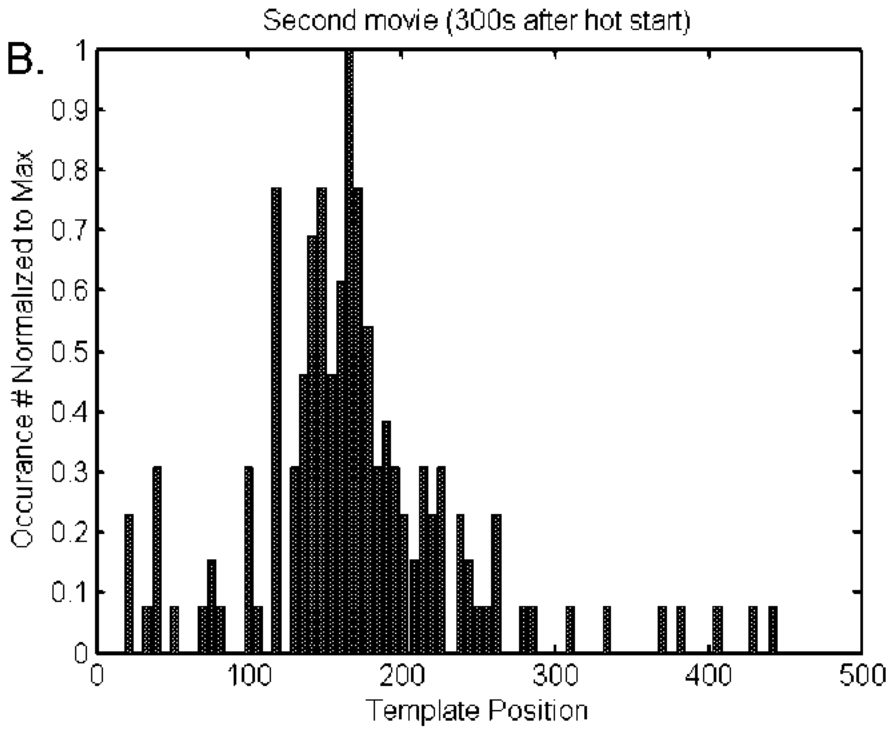


Figure 8, cont.

C.

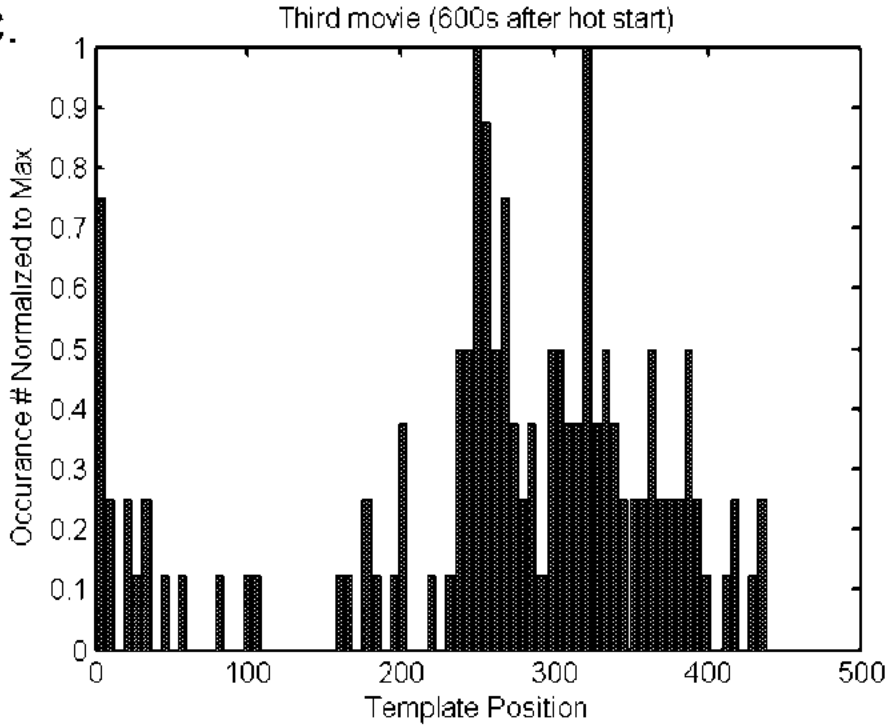


Figure 8, cont.

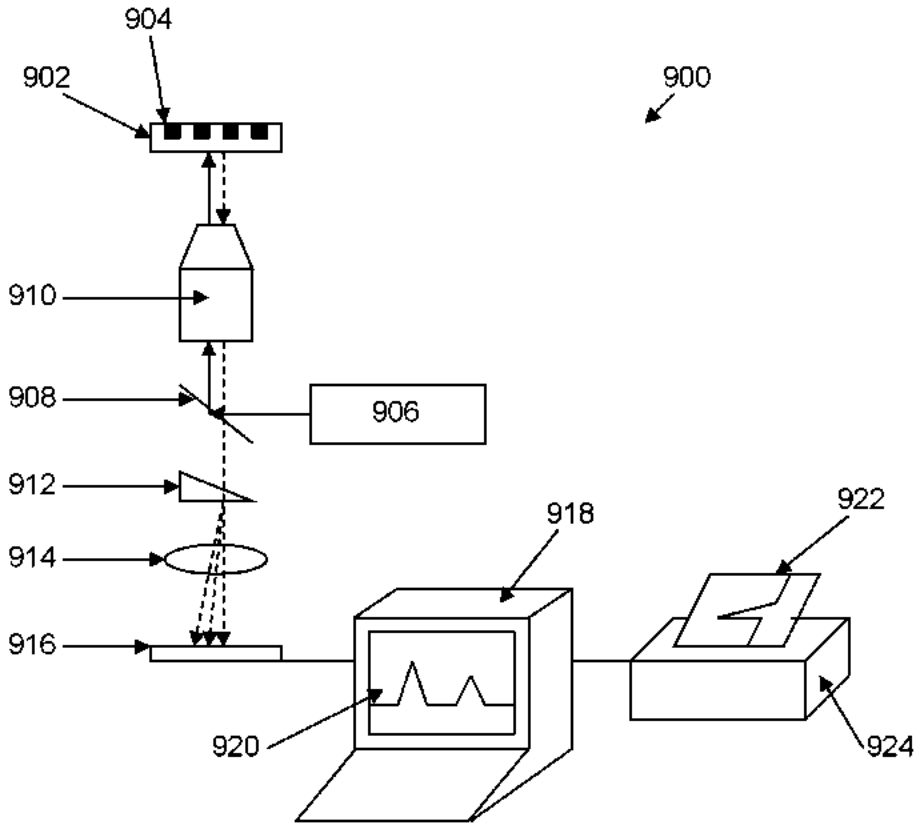


Figure 9

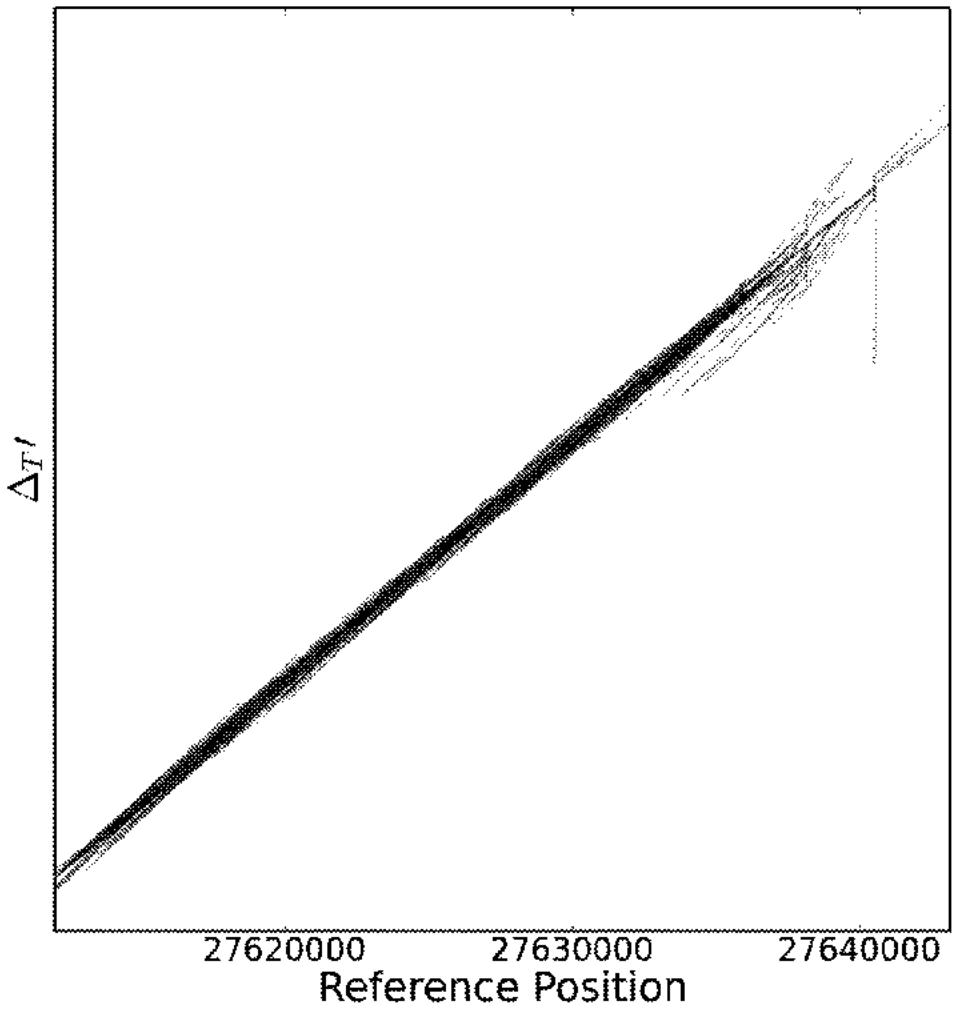


Figure 10

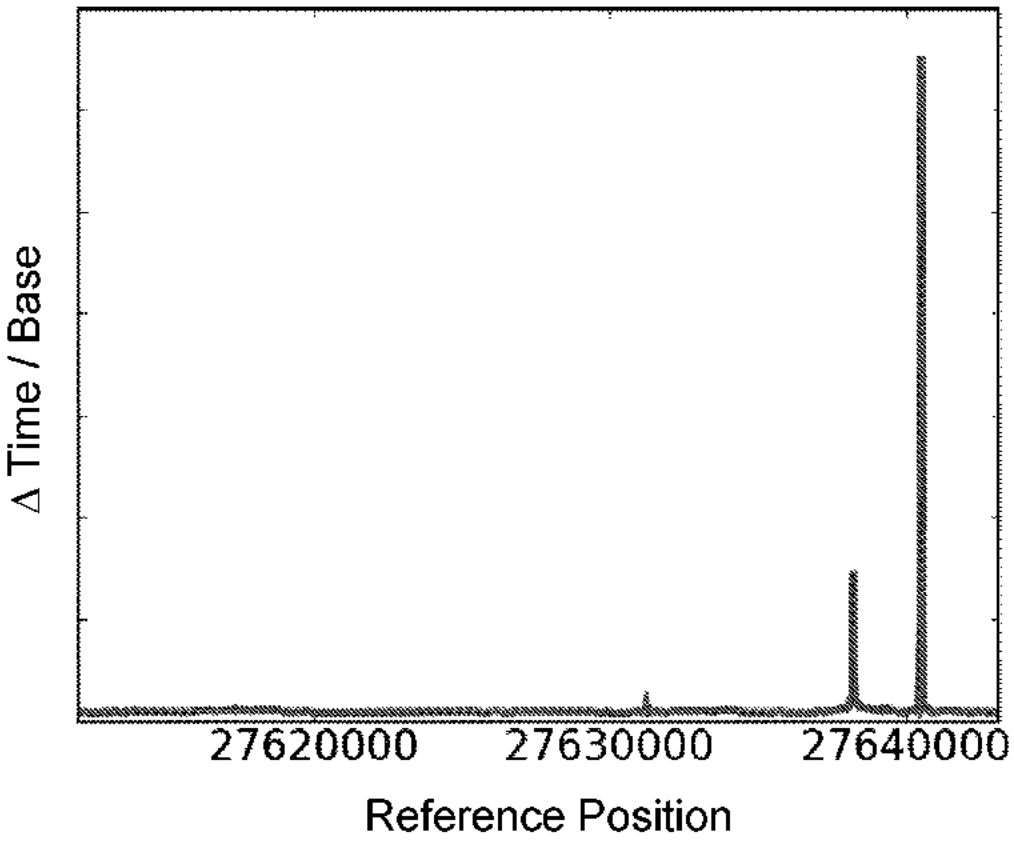

Δ Time / Base
27620000
27630000
27640000
Reference Position

Figure 11

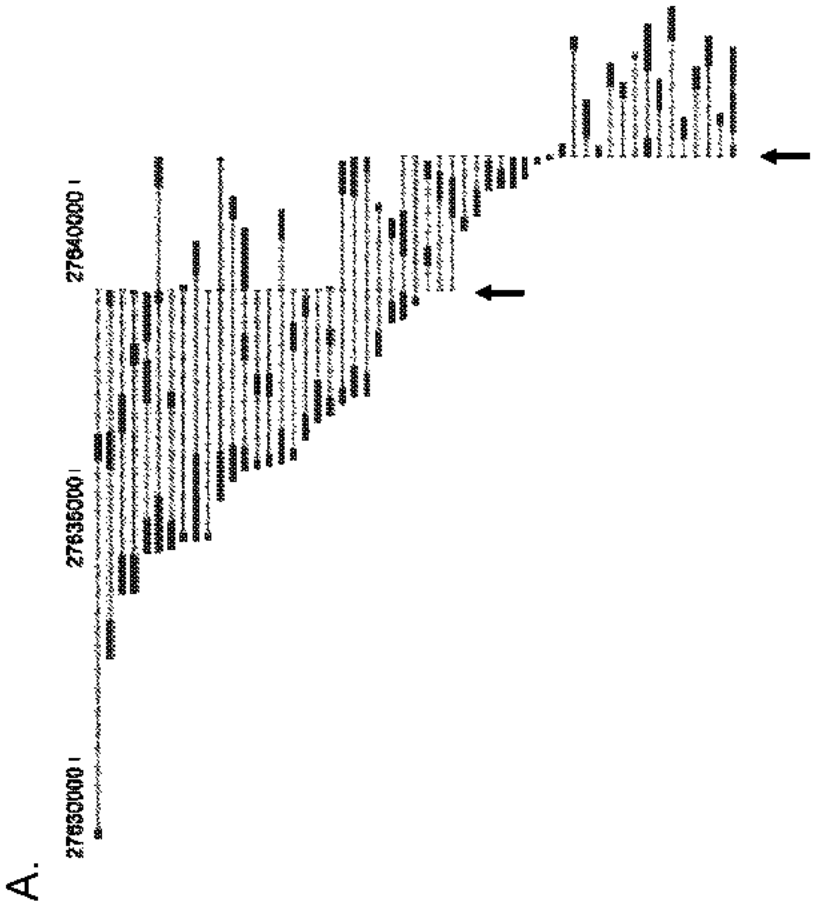


Figure 12

B.

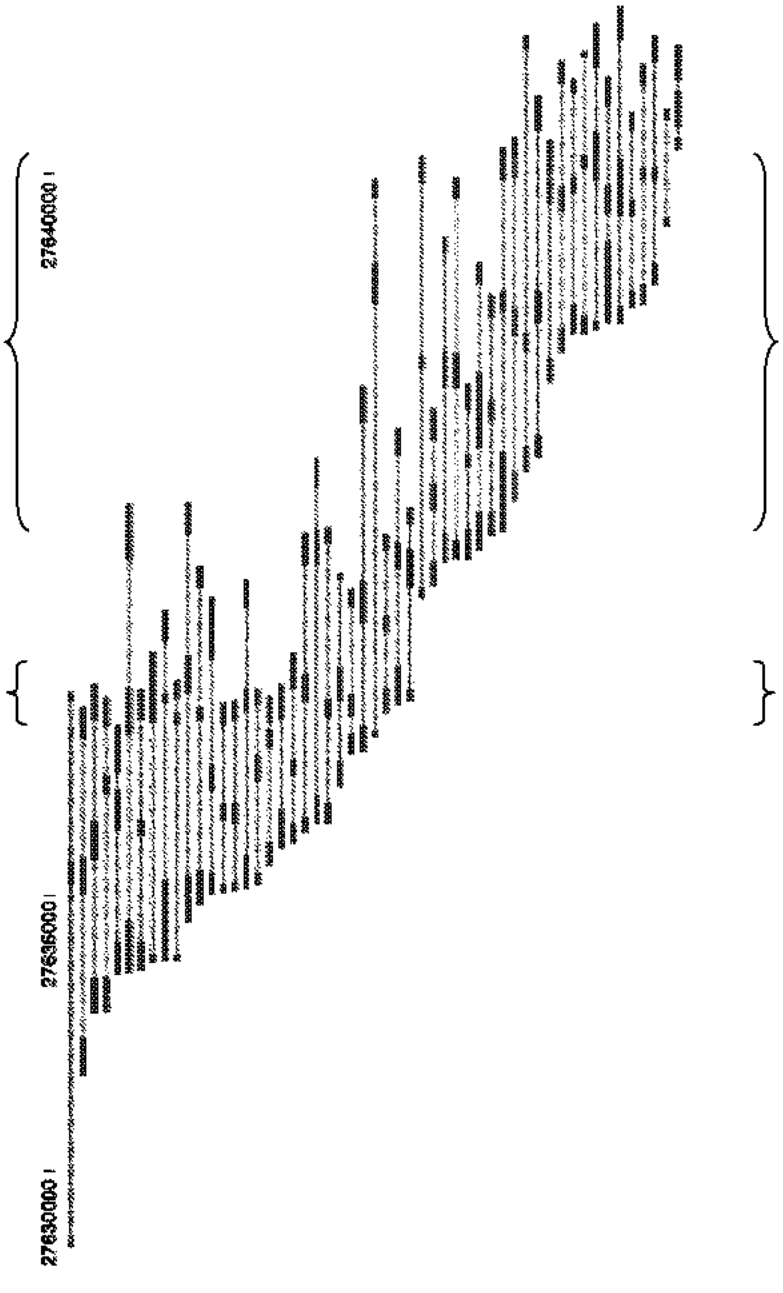


Figure 12, cont.

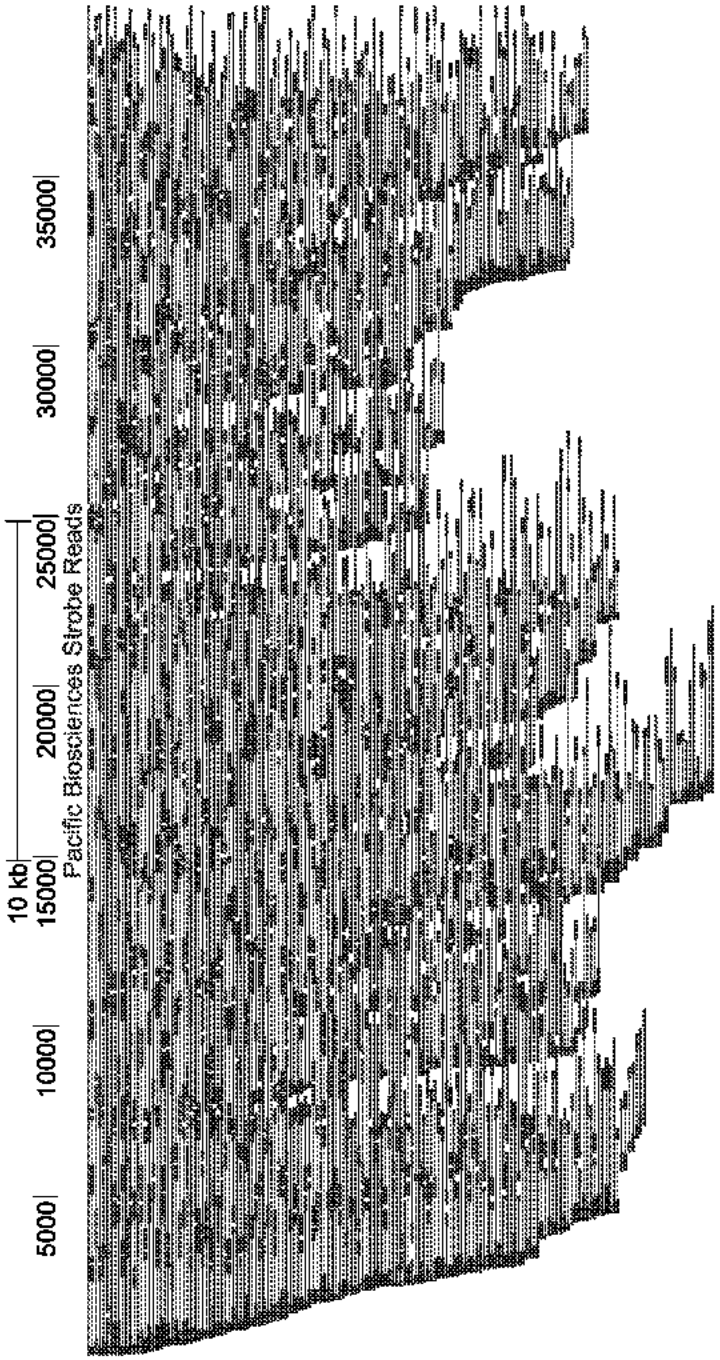

10 kb
Pacific Biosciences Strobe Reads
5000
10000
15000
20000
25000
30000
35000

Figure 13

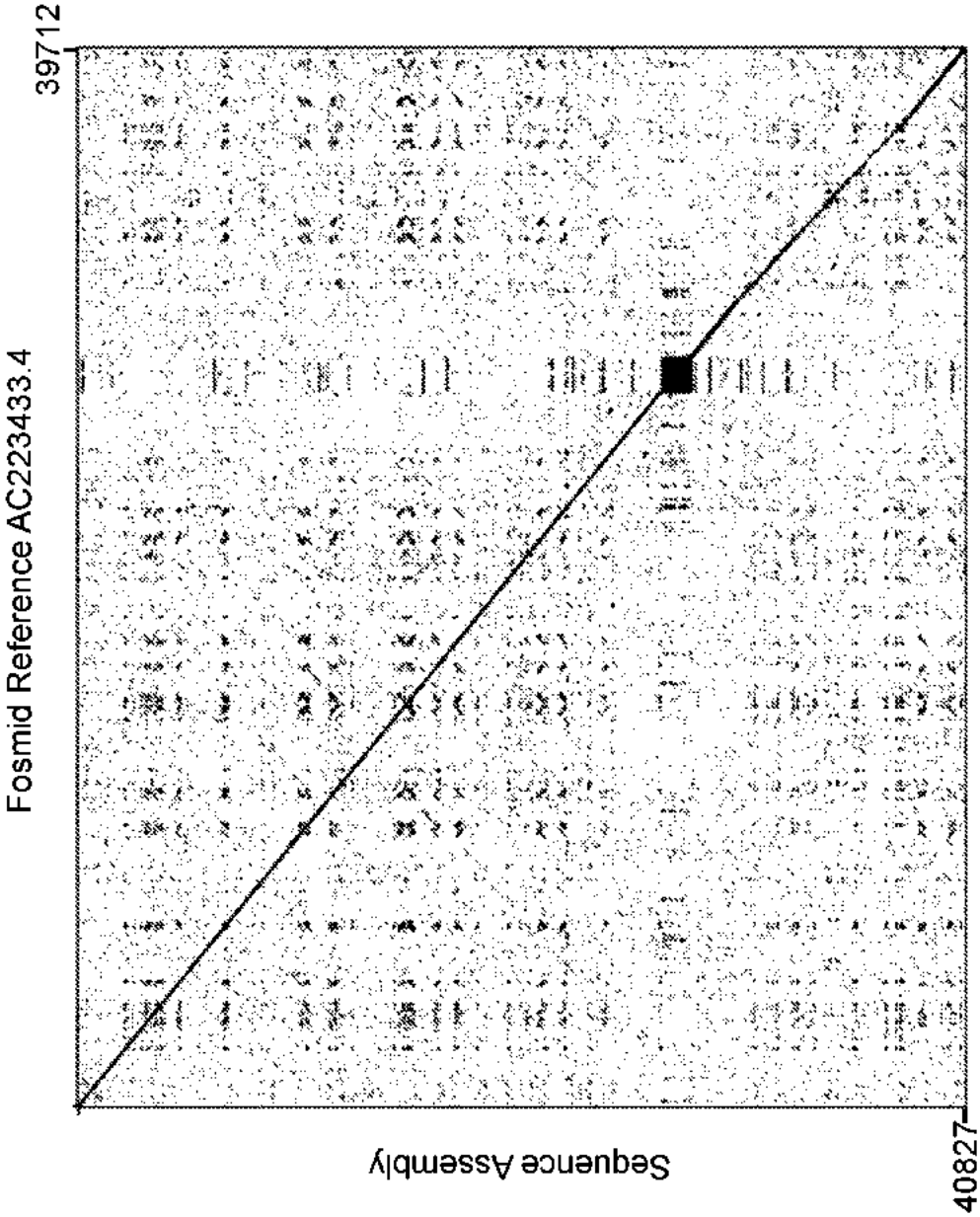

39712
Fosmid Reference AC223433.4
Sequence Assembly
40827

Figure 14

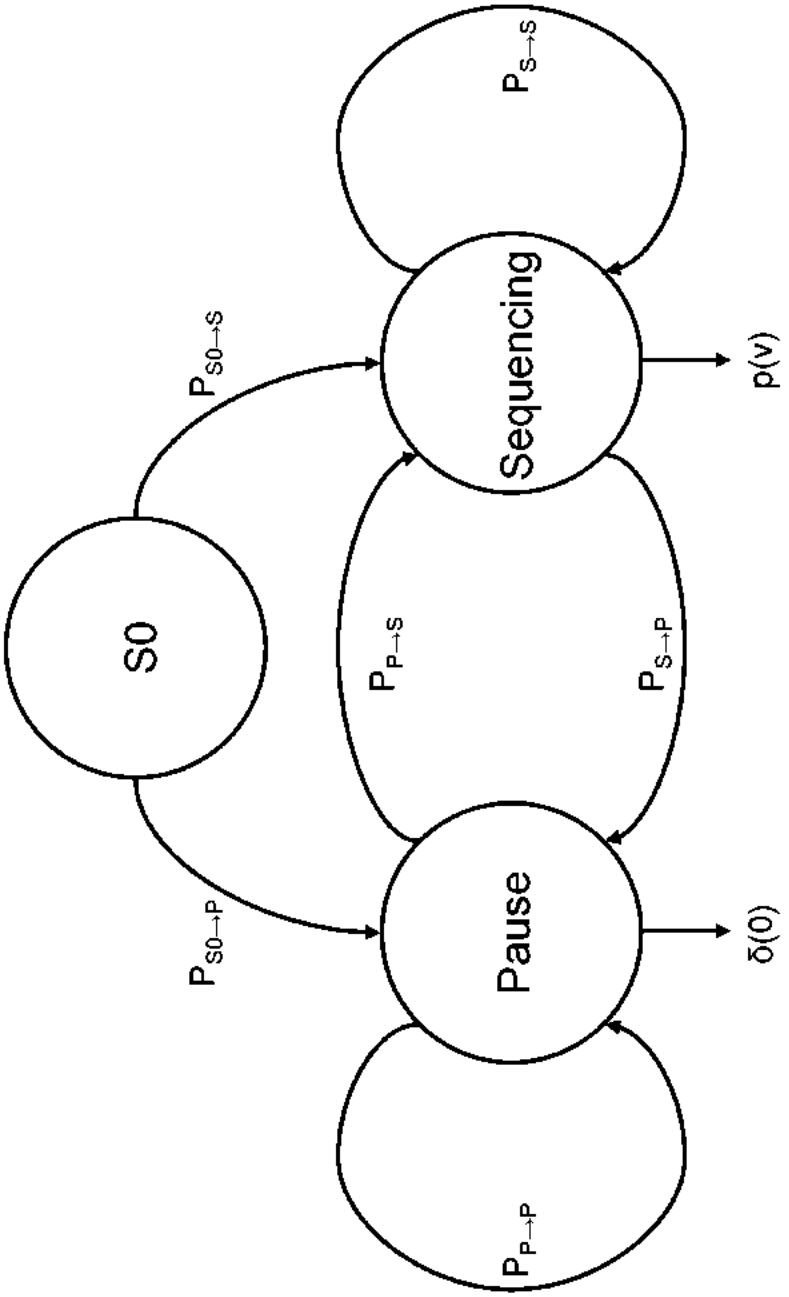

S0
Sequencing
Pause
$P_{S0\rightarrow S}$
$P_{S0\rightarrow P}$
$P_{S\rightarrow S}$
$P_{S\rightarrow P}$
$P_{P\rightarrow S}$
$P_{P\rightarrow P}$
p(v)
δ(0)

Figure 15

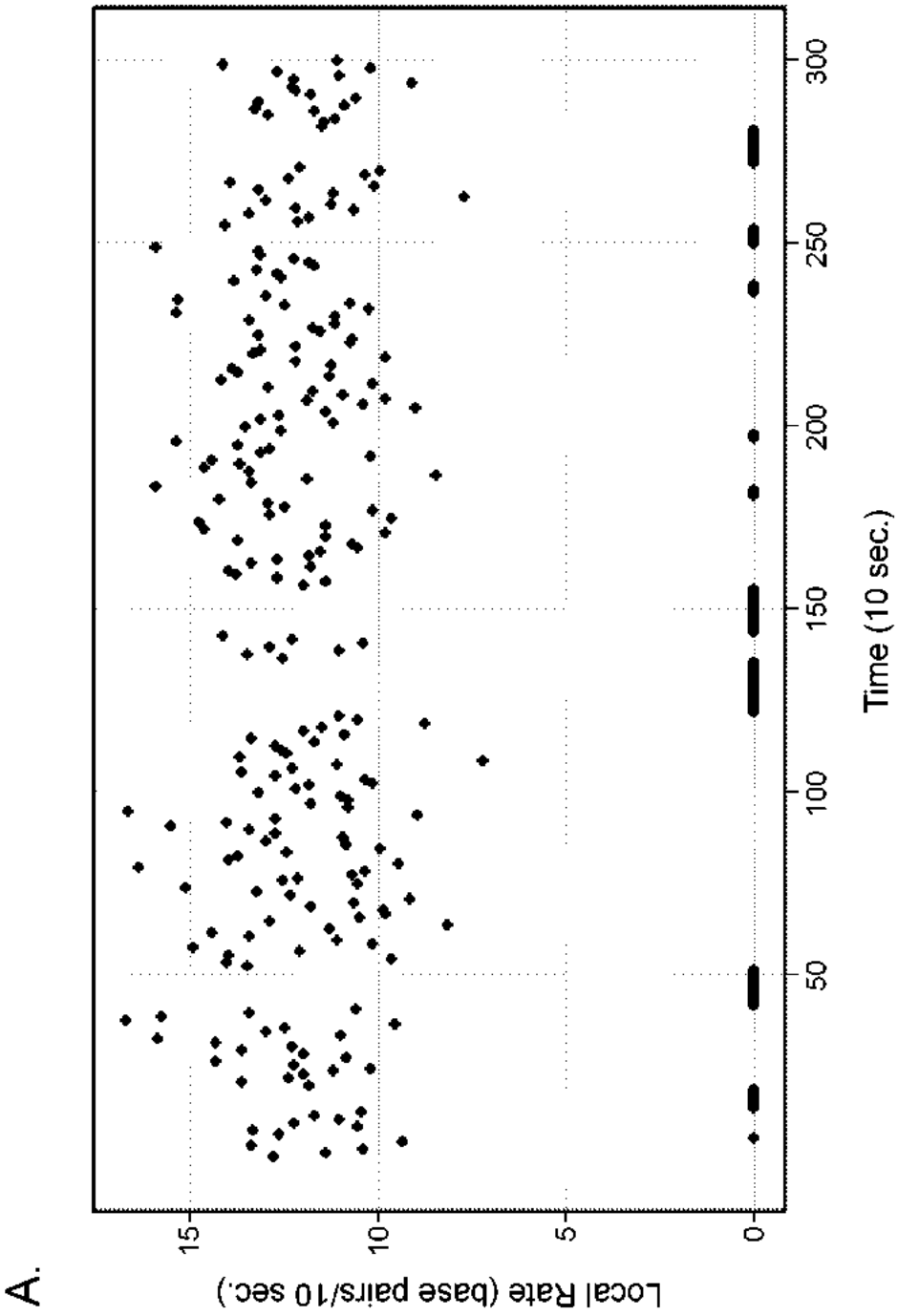

A.
Local Rate (base pairs/10 sec.)
0
5
10
15
50
100
150
200
250
300
Time (10 sec.)

Figure 16

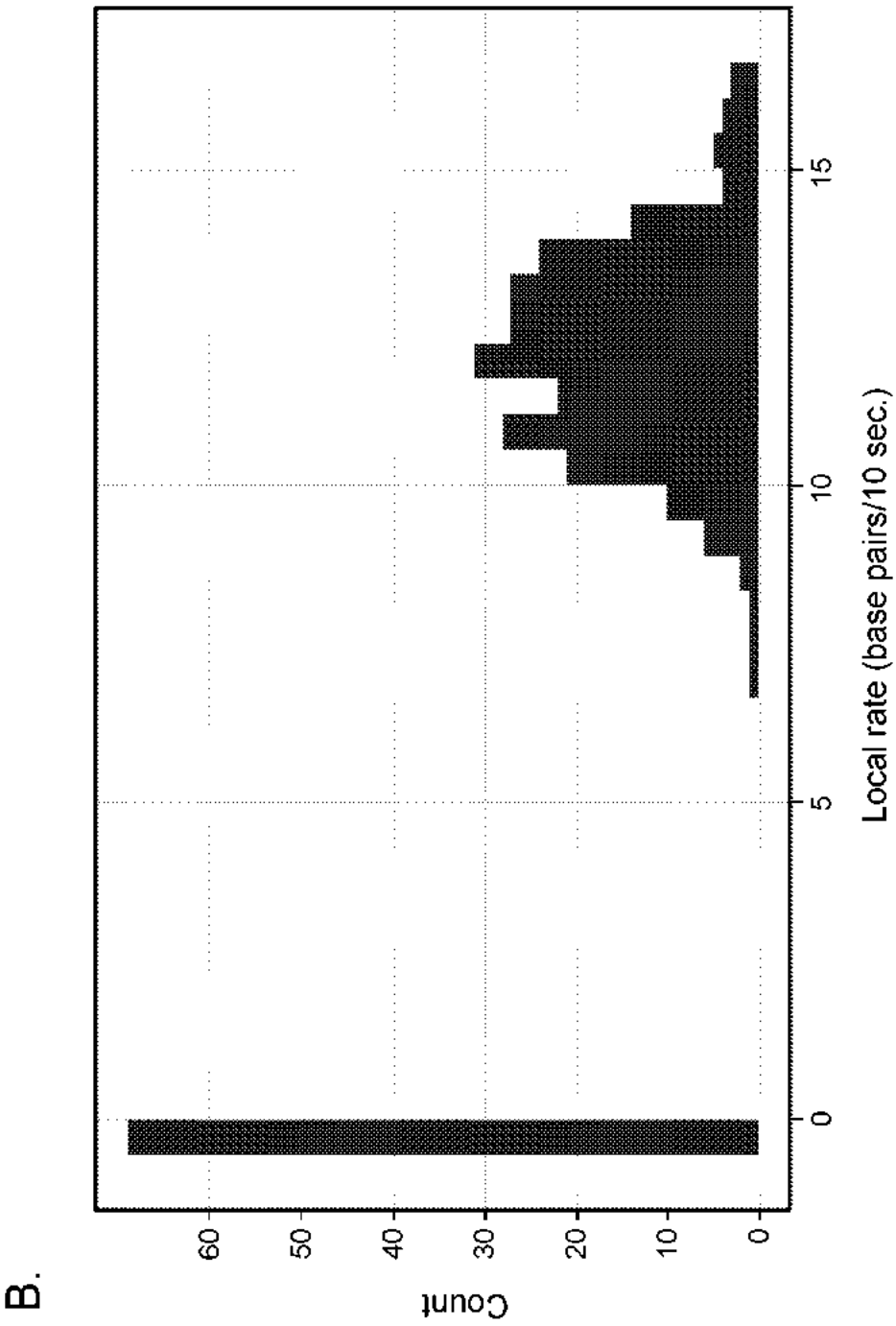


Figure 16, cont.

C.

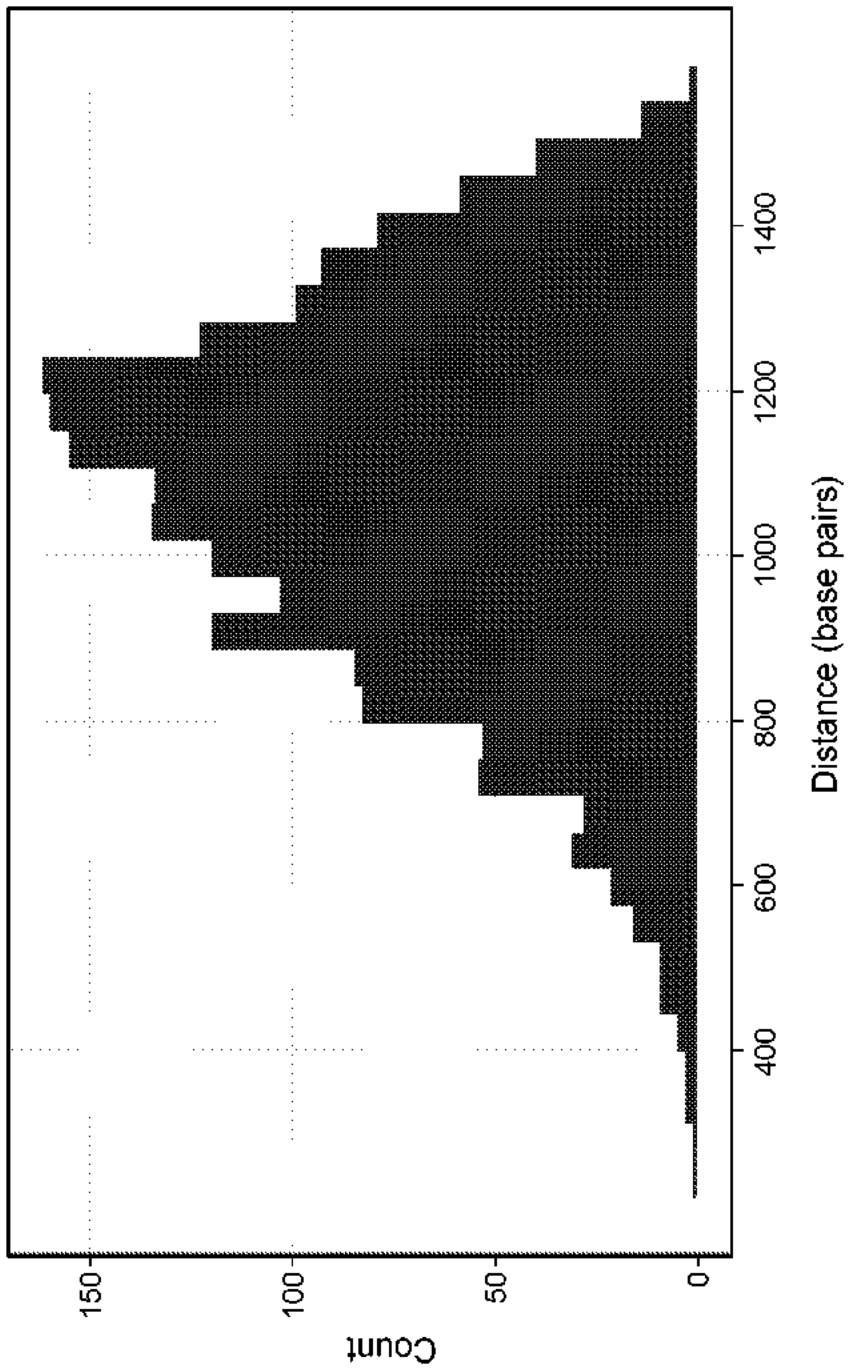


Figure 16, cont.

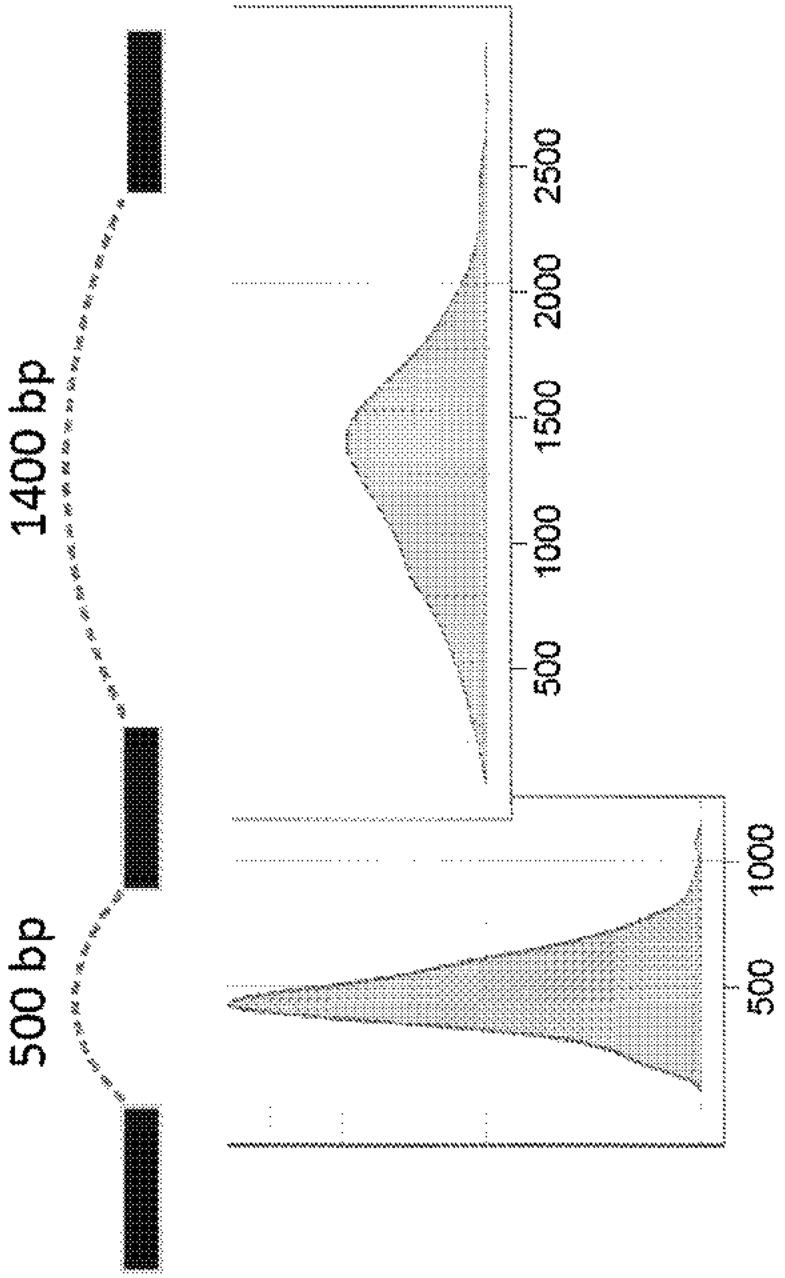


Figure 17

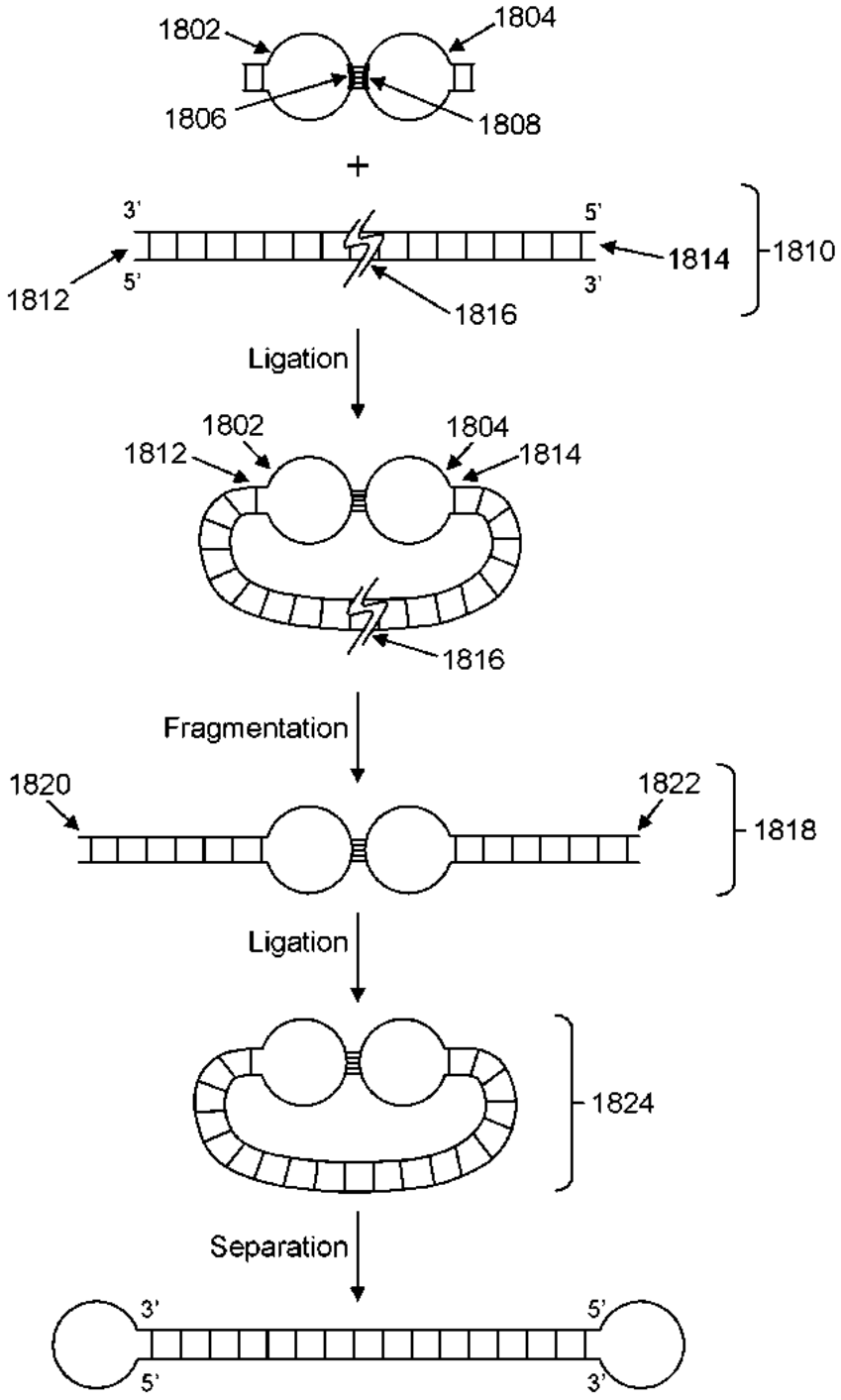

1802
1804
1806
1808
+
3'
5'
1814
1810
1812
5'
3'
1816
Ligation
1802
1804
1812
1814
1816
Fragmentation
1820
1822
1818
Ligation
1824
Separation
3'
5'
5'
3'

Figure 18

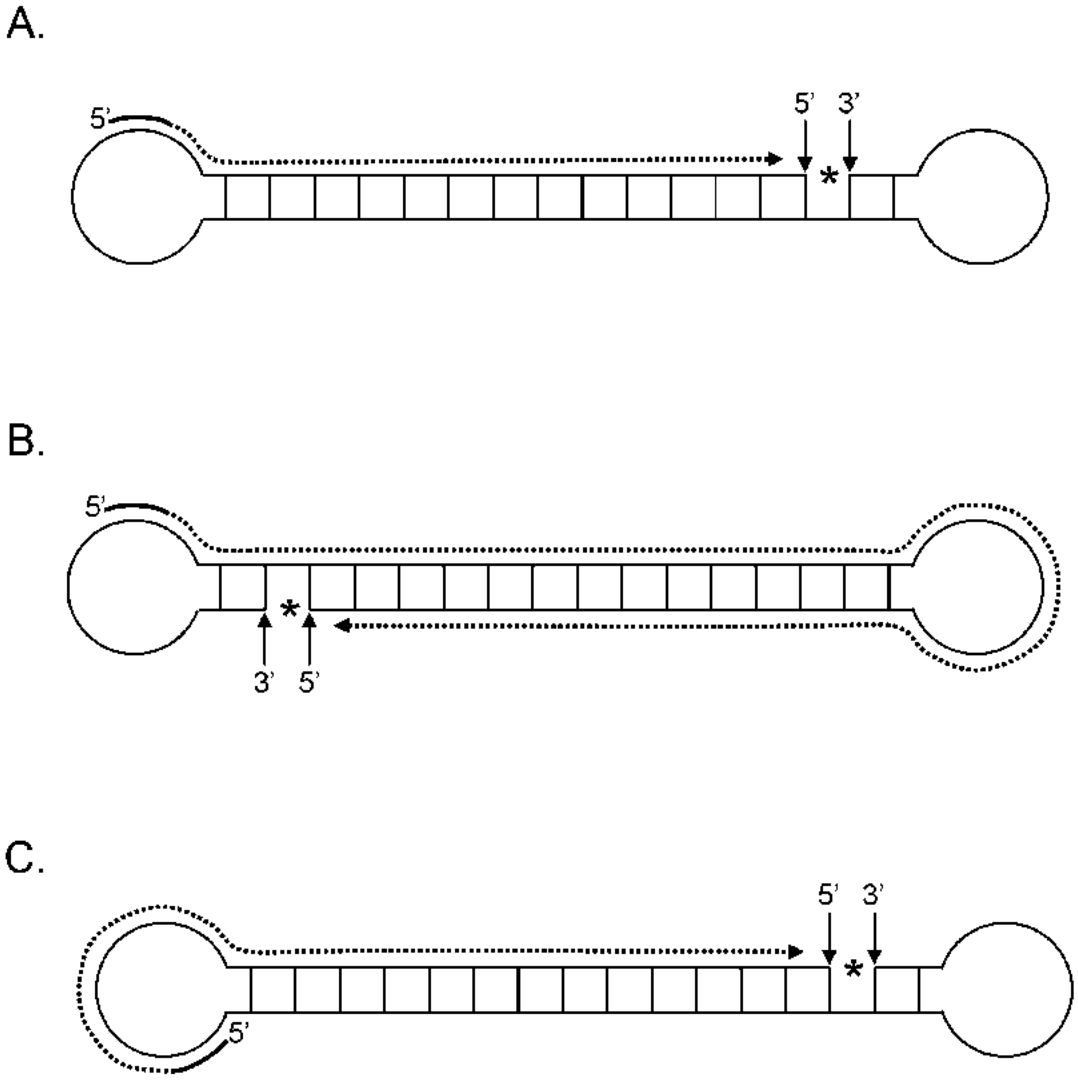


Figure 19

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

As the below named inventor(s), I/we declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States or PCT Application No. 14/091,961, filed on November 27, 2013,

☐ as amended on _____ (if applicable)

The above-identified application was made or authorized to be made by me/us.

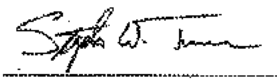
I/we believe that I/we am/are the original inventor/original joint inventors of a claimed invention in the application.

I/we acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me/us to be material to patentability as defined in 37 CFR 1.56.

I/we hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

FULL NAME OF INVENTOR(S)

Inventor one: Stephen Turner Date: 11 December 2013

Signature: 

Inventor two: Jon Sorenson Date:

Signature:

Inventor three: Kenneth Mark Maxham Date:

Signature:

Inventor four: John Eid Date:

Signature:

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

As the below named inventor(s), I/we declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States or PCT Application No. 14/091,961, filed on November 27, 2013,

☐ as amended on _____ (if applicable)

The above-identified application was made or authorized to be made by me/us.

I/we believe that I/we am/are the original inventor/original joint inventors of a claimed invention in the application;

I/we acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me/us to be material to patentability as defined in 37 CFR 1.56.

I/we hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

FULL NAME OF INVENTOR(S)

Inventor one: Stephen Turner Date: _____

Signature: _____

Inventor two: Jon Sorenson Date: 01/14/14

Signature: _____

Inventor three: Kenneth Mark Maxham Date: _____

Signature: _____

Inventor four: John Eid Date: _____

Signature: _____

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

As the below named inventor(s), I/we declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States or PCT Application No. 14/091,961, filed on November 27, 2013,

☐ as amended on _____ (if applicable)

The above-identified application was made or authorized to be made by me/us.

I/we believe that I/we am/are the original inventor/original joint inventors of a claimed invention in the application;

I/we acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me/us to be material to patentability as defined in 37 CFR 1.56.

I/we hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

FULL NAME OF INVENTOR(S)

Inventor one: Stephen Turner Date: _____

Signature: _____

Inventor two: Jon Sorenson Date: _____

Signature: _____

Inventor three: Kenneth Mark Maxham Date: 12/17/2013

Signature: K. Mark Maxham

Inventor four: John Eid Date: _____

Signature: _____

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

As the below named inventor(s), I/we declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States or PCT Application No. 14/091,961, filed on November 27, 2013,

☐ as amended on _____ (if applicable)

The above-identified application was made or authorized to be made by me/us.

I/we believe that I/we am/are the original inventor/original joint inventors of a claimed invention in the application;

I/we acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me/us to be material to patentability as defined in 37 CFR 1.56.

I/we hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

FULL NAME OF INVENTOR(S)

Inventor one: Stephen Turner Date: _____

Signature: _____

Inventor two: Jon Sorenson Date: _____

Signature: _____

Inventor three: Kenneth Mark Maxham Date: _____

Signature: _____

Inventor four: John Eid Date: 1/8/2014

Signature: _____

PTO/AIA/80 (07-12)
Approved for use through 11/30/2014. OMB 0651-0035
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

■ Practitioners associated with Customer Number: 57770

OR

☐ Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number	Name	Registration Number

As attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignments documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

■ The address associated with Customer Number: 57770

OR

☐	Firm or Individual Name			
	Address			
	City		State	Zip
	Country			
	Telephone		Email	

Assignee Name and Address: Pacific Biosciences of California, Inc.
1380 Willow Road
Menlo Park, CA 94025

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/AIA/96 or equivalent) is required to be Filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of The practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee

Signature	[signature]	Date	2016-12-19
Name	Robert H. Reamey, Ph.D.	Telephone	650-521-8000
Title	Vice President, Intellectual Property - Pacific Biosciences of California, Inc.		

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PTO/AIA/96 (08-12)
Approved for use through 01/31/2013. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Pacific Biosciences of California, Inc.

Application No./Patent No.: Not Yet Assigned Filed/Issue Date: December 19, 2016

Titled: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Pacific Biosciences of California, Inc., a corporation

(Name of Assignee) (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. [✓] The assignee of the entire right, title, and interest.

2. [ ] An assignee of less than the entire right, title, and interest (check applicable box):

   [ ] The extent (by percentage) of its ownership interest is _____%. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   [ ] There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. [ ] The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. [ ] The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. [✓] An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel 032084, Frame 0166, or for which a copy thereof is attached.

B. [ ] A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

   1. From: _____ To: _____

      The document was recorded in the United States Patent and Trademark Office at

      Reel _____, Frame _____, or for which a copy thereof is attached.

   2. From: _____ To: _____

      The document was recorded in the United States Patent and Trademark Office at

      Reel _____, Frame _____, or for which a copy thereof is attached.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

4. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

5. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

6. From: _____ To: _____

The document was recorded in the United States Patent and Trademark Office at

Reel _____, Frame _____, or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☑ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

Signature	Date: 2016-12-19
Robert H. Reamey, Ph.D.	Vice President, Intellectual Property-Pacific Biosciences
Printed or Typed Name	Title or Registration Number

UNITED STATES PATENT AND TRADEMARK OFFICE
UNDER SECRETARY OF COMMERCE FOR INTELLECTUAL PROPERTY AND
DIRECTOR OF THE UNITED STATES PATENT AND TRADEMARK OFFICE

JANUARY 30, 2014

PTAS

JACQUELINE L LIM
1380 WILLOW ROAD
MENLO PARK, CA 94025

502657767

UNITED STATES PATENT AND TRADEMARK OFFICE
NOTICE OF RECORDATION OF ASSIGNMENT DOCUMENT

THE ENCLOSED DOCUMENT HAS BEEN RECORDED BY THE ASSIGNMENT RECORDATION BRANCH OF THE U.S. PATENT AND TRADEMARK OFFICE. A COMPLETE COPY IS AVAILABLE AT THE ASSIGNMENT SEARCH ROOM ON THE REEL AND FRAME NUMBER REFERENCED BELOW.

PLEASE REVIEW ALL INFORMATION CONTAINED ON THIS NOTICE. THE INFORMATION CONTAINED ON THIS RECORDATION NOTICE REFLECTS THE DATA PRESENT IN THE PATENT AND TRADEMARK ASSIGNMENT SYSTEM. IF YOU SHOULD FIND ANY ERRORS OR HAVE QUESTIONS CONCERNING THIS NOTICE, YOU MAY CONTACT THE ASSIGNMENT RECORDATION BRANCH AT 571-272-3350. PLEASE SEND REQUEST FOR CORRECTION TO: U.S. PATENT AND TRADEMARK OFFICE, MAIL STOP: ASSIGNMENT RECORDATION BRANCH, P.O. BOX 1450, ALEXANDRIA, VA 22313.

RECORDATION DATE: 01/29/2014 — REEL/FRAME: 032084/0166
NUMBER OF PAGES: 7

BRIEF: ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).

ASSIGNOR:
TURNER, STEPHEN — DOC DATE: 12/11/2013

ASSIGNOR:
SORENSON, JON — DOC DATE: 01/14/2014

ASSIGNOR:
MAXHAM, KENNETH MARK — DOC DATE: 12/17/2013

ASSIGNOR:
EID, JOHN — DOC DATE: 01/08/2014

ASSIGNEE:
PACIFIC BIOSCIENCES OF CALIFORNIA, INC.
1380 WILLOW ROAD
MENLO PARK, CALIFORNIA 94025

APPLICATION NUMBER: 14091961 — FILING DATE: 11/27/2013
PATENT NUMBER: — ISSUE DATE:
TITLE: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

ASSIGNMENT RECORDATION BRANCH
PUBLIC RECORDS DIVISION

Attorney Docket No.: 01-007704US

ASSIGNMENT OF PATENT APPLICATION

JOINT

WHEREAS,

Stephen Turner of 810 5th Street, Kirkland, WA 98033, USA;

Jon Sorenson of 1725 Nason Street, Alameda, CA 94501, USA;

Kenneth Mark Maxham of 241 Harrison Avenue, Redwood City, CA 94062, USA, and

John Eid of 52 Sheridan Street, #2, San Francisco, CA 94103, USA,

hereinafter referred to as "Assignor(s)," is/are the inventor(s) of the invention described and set forth in the below-identified patent application:

Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
Filing Date:	November 27, 2013
Application No.:	14/091,961; and

WHEREAS, Pacific Biosciences of California, Inc., located at 1380 Willow Road, Menlo Park, CA 94025, hereinafter referred to as "ASSIGNEE," is desirous of acquiring an interest in the invention and application and in any U.S. Letters Patent and Registrations which may be granted on any patent application claiming priority from the same;

For good and valuable consideration, receipt of which is hereby acknowledged by Assignor(s), Assignor(s) has/have assigned, and by these presents does/do assign to Assignee all right, title and interest in and to the invention and application and to all foreign counterparts (including patent, utility model and industrial designs), and in and to any Letters Patent and Registrations which may hereafter be granted on any patent application claiming priority from the same in the United States and all countries throughout the world, and to claim the priority from the application as provided by the Paris Convention. The right, title and interest is to be held and enjoyed by Assignee and Assignee's successors and assigns as fully and exclusively as it would have been held and enjoyed by Assignor(s) had this Assignment not been made, for the full term of any Letters Patent and Registrations which may be granted thereon, or of any division, renewal, continuation in whole or in part, substitution, conversion, reissue, prolongation or extension thereof.

Assignor(s) further agree(s) that Assignor(s) will, without charge to Assignee, but at Assignee's expense, (a) cooperate with Assignee in the prosecution of U.S. Patent applications and foreign counterparts on the invention and any improvements, (b) execute, verify, acknowledge and deliver all such further papers, including applications and instruments of transfer, and (c) perform such other acts as Assignee lawfully may request to obtain or maintain Letters Patent and Registrations for the invention and improvements in any and all countries, and to vest title thereto in Assignee, or Assignee's successors and assigns.

IN TESTIMONY WHEREOF, Assignor(s) has/have signed his/her/their name(s) on the date indicated.

12/11/13
Dated — Stephen Turner

Dated — Jon Sorenson

Dated — Kenneth Mark Maxham

Dated — John Eid

IN TESTIMONY WHEREOF, Assignor(s) has/have signed his/her/their name(s) on the date indicated.

Dated	Stephen Turner
01/14/14 Dated	Jon Sorenson
Dated	Kenneth Mark Maxham
Dated	John Eid

IN TESTIMONY WHEREOF, Assignor(s) has/have signed his/her/their name(s) on the date indicated.

Dated	Stephen Turner
Dated	Jon Sorenson
12/17/2013 Dated	 Kenneth Mark Maxham
Dated	John Eid

IN TESTIMONY WHEREOF, Assignor(s) has/have signed his/her/their name(s) on the date indicated.

Dated	Stephen Turner
Dated	Jon Sorenson
Dated	Kenneth Mark Maxham
1/8/2014 Dated	John Eid

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

December 19, 2016

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
Attorney Docket No.: 01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Stephen Turner et al.

Application No.: To be Assigned

Filed: December 19, 2016

For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Examiner: To be Assigned

Art Unit: To be Assigned

Confirmation No: To be Assigned

INFORMATION DISCLOSURE STATEMENT UNDER 37 CFR § 1.97 AND 1.98

COMMISSIONER FOR PATENTS
P. O. BOX 1450
Alexandria, CA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Sir:

This information Disclosure Statement is submitted:

(a) [X] **Under 37 CFR 1.97(b)**
(Within three months of filing national application or date of entry of national application or before mailing date of First Office Action on the merits whichever occurs last).

(b) [ ] **Under 37 CFR 1.97(c) together with either a:**
[ ] Statement under CFR 1.97(e) or
[ ] $180.00 fee under 37 CFR 1.17(p)
(After the CFR 1.97(b) time period, but before a final action, notice of allowance, whichever occurs first).

(c) [ ] **Under 37 CFR 1.97(d) together with: a**
Statement under 37 CFR 1.97(e), and
$180.00 fee as set forth in 37 CFR 1.17(p).
(After a final action or notice of allowance, whichever occurs first, but before payment of the issue fee).

[ ] **STATEMENT UNDER 37 CFR 1.97(e)**

The undersigned certifies that:

[ ] Each item of information contained in the Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the statement, or

[ ] No item of information contained in the Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart foreign application , and to the knowledge of the undersigned after making reasonable inquiry, was known to any individual designated in 37 CFR §1.56 more than three months prior to the filing of the Information Disclosure Statement.

[X] PRIOR APPLICATIONS

All references in the enclosed PTO Form SB08A which are required to be included in this submission, were disclosed in prior Patent Application No.: 14/708,603, filed May 11, 2015, and, as such, copies thereof are not included pursuant to the provisions of 37 CFR 1.98(d).

[ ] FOREIGN LANGUAGE DOCUMENTS

A concise explanation of the relevance of foreign language patents, foreign language publications and other foreign language information listed on PTO form 1449, as presently understood by the individual(s) designated in 37 CFR 1.56 most knowledgeable about the content is given on the attached sheet, or where a foreign language patent is cited in a search report or other action by a foreign patent office in a counterpart foreign application, an English language version of the search or action which indicates the degree of relevance found by the foreign office is listed on form PTO 1449 and is enclosed herewith.

[ ] FEE AUTHORIZATION

Please charge to Deposit Account No. 50-4427 the sum of $ 0.00 at anytime during the pendency of this application, please charge any fees required or credit any overpayment to Deposit Account No.50-4427.

Respectfully Submitted,

December 19, 2016
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF CALIFORNIA
1380 Willow Road
Menlo Park, CA 94025
Phone: (650) 521-8127
Fax: (650) 323-9420
Email: dscherer@pacb.com

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	
Examiner Name	Not Yet Assigned
Attorney Docket Number	01-007706US

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5001050		1991-03-19	Blanco et al.	
	2	5198543		1993-03-30	Blanco et al.	
	3	5350686		1994-09-27	Jhingah	
	4	5470724		1995-11-28	Ahern	
	5	5547839		1996-08-20	Dower et al.	
	6	5576204		1996-11-19	Blanco et al.	
	7	5648245		1997-07-15	Fire et al.	
	8	5674683		1997-10-07	Kool	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	9	5674716		1997-10-07	Tabor et al.	
	10	5714320		1998-02-03	Kool	
	11	5854033		1998-12-29	Lizardi	
	12	6210891	B1	2001-04-03	Nyren et al.	
	13	6210896		2001-04-03	Chan	
	14	6255083		2001-07-03	Williams	
	15	6261808		2001-07-17	Auerbach	
	16	6369038		2002-04-09	Blumenfeld et al.	
	17	6451563		2002-09-17	Wittig et al.	
	18	6498023	B1	2002-12-24	Abarzua	
	19	6787308		2004-09-07	Balasubramanian	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	20	6917726		2005-07-12	Levene et al.	
	21	7013054		2006-03-14	Levene et al.	
	22	7033764		2006-04-25	Korlach et al.	
	23	7045362	B2	2006-05-16	Hartwich et al.	
	24	7052847		2006-05-30	Korlach et al.	
	25	7056661		2006-06-06	Korlach et al.	
	26	7056676		2006-06-06	Korlach et al.	
	27	7170050		2006-06-06	Korlach et al.	
	28	7181122		2007-02-20	Levene et al.	
	29	7229799		2007-06-12	Williams et al.	
	30	7282337		2007-10-16	Harris et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	31	7292742		2007-11-06	Levene et al.	
	32	7361466		2008-04-22	Korlach et al.	
	33	7368265		2008-05-06	Brenner et al.	
	34	7416844		2008-08-26	Korlach et al.	
	35	7476503		2007-01-30	Turner et al.	
	36	7485424		2009-02-03	Korlach et al.	
	37	7601495		2009-10-13	Chen et al.	
	38	7601499		2009-10-13	Berka et al.	
	39	7700287	B2	2010-04-20	Chen et al.	
	40	7754429	B2	2010-07-13	Rigatti et al.	
	41	7767400	B2	2010-08-03	Harris et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	42	6849404	B2	2005-02-01	Park et al.	
	43	8143030	B2	2012-03-27	Maxham et al.	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20010030290	A1	2001-10-18	Stern	
	2	20030044781	A1	2003-03-06	Korlach et al.	
	3	20030096253		2003-05-22	Nelson et al.	
	4	20030143550	A1	2003-07-31	Green et al.	
	5	20030190647		2003-10-09	Odera	
	6	20030207279	A1	2003-11-06	Crothers et al.	
	7	20030213771	A1	2003-11-20	Ohshita et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	8	20030215862		2003-11-20	Parce et al.	
	9	20040048300		2004-03-11	Sood et al.	
	10	20040152119		2004-08-05	Sood et al.	
	11	20040203008	A1	2004-10-14	Uemori et al.	
	12	20040224319		2004-11-11	Sood et al.	
	13	20040259082		2004-12-23	Williams	
	14	20050176035	A1	2005-08-11	Crothers et al.	
	15	20060061754	A1	2006-03-23	Turner et al.	
	16	20060292611	A1	2006-12-28	Berka et al.	
	17	20070062934	A1	2007-03-22	King	
	18	20070161017	A1	2007-07-12	Eid et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	19	20070178482	A1	2007-08-02	Lezhava et al.	
	20	20070269825	A1	2007-11-22	Wang et al.	
	21	20080026393	A1	2008-01-31	Mindrinos et al.	
	22	20080233575		2008-09-25	Harris et al.	
	23	20090197257	A1	2009-08-06	Harris	
	24	20090269771		2009-10-29	Schroeder	
	25	20090305248		2009-12-10	Lander et al.	
	26	20020197618	A1	2002-12-26	Sampson	
	27	20080009007		2008-01-10	Lyle et al.	
	28	20080176241		2008-07-24	Eid et al.	
	29	20090005252	A1	2009-01-01	Drmanac et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	30	20090087850		2009-04-02	Eid et al.	
	31	20090233291	A1	2009-09-17	Chen et al.	
	32	20081218184	A1	2008-09-11	White et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	199106678	WO	A1	1991-05-16	SRI International		☐
	2	199416090	WO	A1	1994-07-21	Molecular Tool, Inc.		☐
	3	199627025	WO	A1	1996-09-06	Rabani		☐
	4	199905315	WO	A2	1999-02-04	Medical Biosystems Ltd.		☐
	5	2007003017	WO	A1	2007-01-11	Biochip Innovations PTY		☐
	6	2007070572	WO		2007-06-21	US Govt as represented by DHHS		☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	7	2008058282	WO		2008-05-15	Complete Genomics Inc.		☐
	8	2009124255	WO	A2	2009-10-08	Helicos Biosciences Corp.		☐
	9	1225234	EP	B1	2007-11-07	Agilent Technologies		☐
	10	1907573	EP	B1	2010-01-27	Braslavsky		☐
	11	2007010263	WO	A2	2007-01-25	Swerdlow et al.		☐
	12	2008064905	WO	A2	2007-06-07	Eid et al.		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	BASHIR, A. et al., "Evaluation of paired-end sequencing strategies for detection of genome rearrangements in cancer" Plos CompBiol (2008) 4(4):1-14	☐
	2	EID, et al., "Real-time DNA sequencing from single polymerase molecules" Science (2009) 323(5910):133-138	☐
	3	HARRIS, T.D. et al., "Single-molecule DNA sequencing of a viral genome" Science (2008) 320:106-109	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	4	HONG, Y.S. et al., "Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing" Mol Genet Genomics (2003) 268:720-728	☐
	5	KEANE, T. et al., "Assessing assemblability of reads from new sequencing platforms" Wellcome Trust Poster, Page 1	☐
	6	KOONIN et al. "Computer-assisted dissection of rolling circle DNA replication" Biosystems (1993) 30(1-3):241-268	☐
	7	KORBEL, J.O. et al. "Paired-end mapping reveals extensive structural variation in the human genome" Science (2007) 318:420-426	☐
	8	KUHN et al., "Rolling-circle amplificaiton under topological constraints" Nucl Acids Res (2002) 30(2):574-580	☐
	9	LEVENE et al., "Zero-mode waveguides for single-molecule analysis at high concentrations" Science (2003) 299 (5607):682-686	☐
	10	MATRAY, T.J. et al. "A specific partner for abasic damage in DNA" Nature (1999) 399:704-708	☐
	11	MYERS, G. "Whole-genome DNA sequencing" IEEE (May-June 1999) pp. 33-43	☐
	12	NOVICK "Contrasting lifestyles of rolling-circle phages and plasmids" Trends Biochem Sci (1998) 23(11):434-438	☐
	13	REIFENBERGER, J. et al., Advances in Genome Biol and Tech (AGBT) (2009) Abstract February 4-7, 2009	☐
	14	REIFENBERGER, J. et al., Biophys Soc 53rd Ann Meeting (2009) Abstract, February 28, 2009	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	15	SMITH, M. et al., "Genomic sequence sampling: a strategy for high resolution sequence-based physical mapping of complex genomes" Nature Genetics (1994) 7:40-47	☐
	16	SPINELLA et al., "Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles" Nucl Acids Res (1999) 27(18):e22-e22	☐
	17	VELCULESCU et al. "Serial analysis of gene expression" Science (1995) 270(5235): 484-487	☐
	18	VOLIK, S. et al., "End-sequence profiling: sequence-based analysis of aberrant genomes" PNAS (2003) 100 (13):7696-7701	☐
	19	WILEY, G. et al., "Methods for generating shotgun and mixed shotgun/paired-end libraries for the 454 DNA sequencer" Current Protocols in Human Genomics (2009) Chapter 18; Unit 18.1, pp. 1-21	☐
	20	Technology Spotlight: Illumina Sequencing Technology, current of 10/08/08, pp 1-4	☐
	21	HORMOZDIARI, et al. "Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. " Genome Research (2009) 19:1270-1278	☐
	22	LEE, et al., "A robust framework for detecting structural variations in a genome," Bioinformatics (2008) 24:i59-i67.	☐
	23	MARGULIES, et al., "Genome sequencing in microfabricated high-density picolitre reactors," Nature (2005), 437:376-382.	☐
	24	PEDLER, "Occupation Times for Two State Markov Chains," Journ Appl Probability (1971), 8(2):381-90.	☐
	25	SVOBODA, et al., "Fluctuation analysis of motor protein movement and single enzyme kinetics," PNAS (1994), 91:11782-86.	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	26	International Search Report and Written Opinion dated 4/29/2010 for related case PCT/US2009/005169.	☐
	27	International Search Report and Written Opinion dated 10/27/2009 for related case PCT/US2009/001930	☐
	28	International Preliminary Report on Patentability dated 10/7/2010 for related case PCT/US2009/001930	☐
	29	International Search Report and Written Opinion dated 11/17/2009 for related case PCT/US2009/001926	☐
	30	METZKER, M.I., "Emerging Technologies in DNA Sequencing," Genome Research (2005) 15:1767-1776	☐
	31	International Preliminary Report on Patentability dated 4/7/2011 for related case PCT/US2009/005169	☐
	32	International Search Report and Written Opinion dated 11/3/2009 for related case PCT/US2009/001927	☐
	33	International Preliminary Report on Patentability dated 4/7/2011 for related case PCT/US2009/001927	☐
	34	Supplementary European Search Report dated March 20, 2012 for related case EP 09816557.4	☐
	35	First Exam Report dated 7/18/13 of related EP 09816557.4	☐
	36	Second Exam Report dated 4/9/2015 of related EP 09816557.4	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	37	Third Exam Report dated 3/30/16 of related EP 09816557.4	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David C. Scherer, Ph.D./	Date (YYYY-MM-DD)	2016-12-19
Name/Print	David C. Scherer, Ph.D.	Registration Number	56,993

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Filer:	David Christopher Scherer/Jacqueline Lim
Attorney Docket Number:	01-007706US

Filed as Large Entity

Filing Fees for Track I Prioritized Examination - Nonprovisional Application under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
UTILITY APPLICATION FILING	1011	1	280	280
UTILITY SEARCH FEE	1111	1	600	600
UTILITY EXAMINATION FEE	1311	1	720	720
REQUEST FOR PRIORITIZED EXAMINATION	1817	1	4000	4000
Pages:				
Claims:				
Miscellaneous-Filing:				
PROCESSING FEE, EXCEPT PROV. APPLS.	1830	1	140	140

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**5740**

Electronic Acknowledgement Receipt	
EFS ID:	27838385
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	19-DEC-2016
Filing Date:	
Time Stamp:	18:11:47
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$5740
RAM confirmation Number	122016INTEFSW00006164504427
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	TrackOne Request	01007706_2016-12-19_ReqPrioritizedExam.pdf	423332 b931c8f23dcf2a15442b1bf4dc7c06e042397a8ffb	no	1
Warnings:					
Information:					
2	Transmittal of New Application	01007706_2016-12-19_AppTrans.pdf	600215 fc2102aefa9825edea7200de18bfedfa1739341e	no	1
Warnings:					
Information:					
3	Fee Worksheet (SB06)	01007706_2016-12-19_FeeTrans.pdf	592748 771cc03c0dbdde593b9837163457580696485c55	no	1
Warnings:					
Information:					
4	Application Data Sheet	01007706_2016-12-19_ADS.pdf	53364 521dd8000c25ea938fa558a49bf56f0943256e86	no	3
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
5		01007706_2016-12-19_App.pdf	466929 59be590daa6590ac2a3db3026177fa126fa344900	yes	106

Multipart Description/PDF files in .zip description

Document Description	Start	End
Specification	1	103
Claims	104	105

	Abstract		106		106
Warnings:					
Information:					
6	Drawings-only black and white line drawings	01007706_2016-12-19_Figs.pdf	1246021 aa1138a3a74418c7846d31c3076777le11ac446a	no	24
Warnings:					
Information:					
7	Oath or Declaration filed	01007706_2016-12-19_Parent Declaration.pdf	1433616 cf41c66078e35799f4de4767e8f6405e49ec8128	no	4
Warnings:					
The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing					
Information:					
8	Power of Attorney	01007706_2016-12-19_POA.pdf	1096040 0ca9ab15dea24ea0a590c51597f2243deac3eadf	no	1
Warnings:					
The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing					
Information:					
9	Miscellaneous Incoming Letter	01007706_2016-12-19_Stmnt373c.pdf	1631837 40f173c20e6ca5f04d344a65e13557d529c61be2	no	2
Warnings:					
The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing					
Information:					
10	Miscellaneous Incoming Letter	01007706_2016-12-19_ParentAssignment.pdf	1280639 29f0d27133c9a0fa93c0a127c068a60ecd0fe8ad	no	7
Warnings:					
The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing					
Information:					

11	Transmittal Letter	01007706_2016-12-19_IDS.pdf	82182 12f8a839723bdcf2cb5569a7dba3a571cba47674	no	2
Warnings:					
Information:					
12	Information Disclosure Statement (IDS) Form (SB08)	01007706_2016-12-19_SB08A.pdf	1455328 05a60af5faf763948572f3a9d94fd2492f3ca4870	no	14
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
13	Fee Worksheet (SB06)	fee-info.pdf	38459 714226d13c925a7823c0043abcc3243d8d2064802	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	10400710		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRACK1.REQ
Document Description: TrackOne Request

PTO/AIA/424 (04-14)

CERTIFICATION AND REQUEST FOR PRIORITIZED EXAMINATION UNDER 37 CFR 1.102(e) (Page 1 of 1)

First Named Inventor:	Stephen Turner	Nonprovisional Application Number (if known):	Not Yet Assigned
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS		

APPLICANT HEREBY CERTIFIES THE FOLLOWING AND REQUESTS PRIORITIZED EXAMINATION FOR THE ABOVE-IDENTIFIED APPLICATION.

1. The processing fee set forth in 37 CFR 1.17(i)(1) and the prioritized examination fee set forth in 37 CFR 1.17(c) have been filed with the request. The publication fee requirement is met because that fee, set forth in 37 CFR 1.18(d), is currently $0. The basic filing fee, search fee, and examination fee are filed with the request or have been already been paid. I understand that any required excess claims fees or application size fee must be paid for the application.

2. I understand that the application may not contain, or be amended to contain, more than four independent claims, more than thirty total claims, or any multiple dependent claims, and that any request for an extension of time will cause an outstanding Track I request to be dismissed.

3. The applicable box is checked below:

I. ☑ Original Application (Track One) - Prioritized Examination under § 1.102(e)(1)

i. (a) The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a). This certification and request is being filed with the utility application via EFS-Web.

---OR---

(b) The application is an original nonprovisional plant application filed under 35 U.S.C. 111(a). This certification and request is being filed with the plant application in paper.

ii. An executed inventor's oath or declaration under 37 CFR 1.63 or 37 CFR 1.64 for each inventor, **or** the application data sheet meeting the conditions specified in 37 CFR 1.53(f)(3)(i) is filed with the application.

II. ☐ Request for Continued Examination - Prioritized Examination under § 1.102(e)(2)

i. A request for continued examination has been filed with, or prior to, this form.
ii. If the application is a utility application, this certification and request is being filed via EFS-Web.
iii. The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a), or is a national stage entry under 35 U.S.C. 371.
iv. This certification and request is being filed prior to the mailing of a first Office action responsive to the request for continued examination.
v. No prior request for continued examination has been granted prioritized examination status under 37 CFR 1.102(e)(2).

Signature /David C. Scherer, Ph.D./	Date December 19, 2016
Name (Print/Typed) David C. Scherer, Ph.D.	Practitioner Registration Number 56,993

Note: *This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

☐ *Total of _____ forms are submitted.

SCORE Placeholder Sheet for IFW Content

Application Number: 15383965 Document Date: 12/19/2016

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

- Drawings – Other than Black and White Line Drawings

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

To access the documents in the SCORE database, refer to instructions below.

At the time of document entry (noted above):

- Examiners may access SCORE content via the eDAN interface.
- Other USPTO employees can bookmark the current SCORE URL (http://Score.uspto.gov/ScoreAccessWeb/).
- External customers may access SCORE content via the Public and Private PAIR interfaces.

Form Revision Date: September 30, 2013

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

December 20, 2016

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
Inventor: Stephen Turner, et al.	Examiner: Not Yet Assigned
Application No.: 15/383,965	Confirmation No.: 8144
Filed: December 19, 2016	Art Unit: Not Yet Assigned
For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS	PRELIMINARY AMENDMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INTRODUCTORY REMARKS

This Preliminary Amendment is filed prior to an action on the merits in this matter. Applicant respectfully requests entry of the amendments and remarks provided herein.

Amendments to the Specification begin on page **2** of this paper.

Amendments to the Drawings begin on page **6** of this paper.

Remarks/Arguments begin at page 7 of this paper.

AMENDMENTS TO THE SPECIFICATION

Please add the following paragraphs after paragraph **[0057]** and renumber the subsequent paragraphs in the application accordingly. No new matter has been added.

[0058] Figures 20A and 20B illustrate two exemplary embodiments of template constructs used in the present invention.

[0059] Figures 21A and 21B schematically illustrate redundant or consensus sequencing using the constructs shown in Figures 20A and 20B.

Please insert the following paragraphs after paragraph [00110] of the specification as filed. These will become new paragraphs [00113]-[00117]. Please renumber the subsequent paragraphs in the application accordingly. No new matter has been added.

[00113] The present invention provides novel template configurations and methods for exploiting these compositions in template directed sequencing processes. While these compositions and methods have utility across all of the various template directed processes described herein, for ease of discussion, they are being primarily discussed in terms of preferred single molecule, real-time sequencing processes, in which they provide myriad benefits. In particular, the present invention is generally directed to nucleic acid sequences that employ improved template sequences to improve the accuracy of sequencing processes. For example, in at least one aspect, the template compositions of the invention are generally characterized by the presence of a double stranded segment or a pair of sub-segments that are internally complementary, i.e., complementary to each other. In particular contexts, the target nucleic acid segment that is included within a template construct will typically be substantially comprised of a double stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double stranded or otherwise internally complementary.

[00114] Examples of template configurations of the invention that are partially and completely contiguous are schematically illustrated in Figure 20A and 20B, respectively. In particular, as shown in Figure 20A, a partially contiguous template sequence 200 is shown which includes a double stranded portion, comprised of two complementary segments 202 and 204, which, for example, represent a target sequence or portion thereof. As shown, the 3' end of segment 202 is linked to the 5' end of segment 204 by linking oligonucleotide 206, providing a single stranded portion of the template, and yielding a partially contiguous sequence. By comparison, as shown in Figure 20B, a completely contiguous template sequence 210 is shown. Sequence 210 includes a double stranded portion again comprised of two complementary segments 212 and 214. As with the partially contiguous sequence of Figure 20A, the 3' end of segment 212 is joined to the 5' end of segment 214 via oligonucleotide 216 in a first single stranded portion. In addition, the 5' end of segment 212 is joined to the 3' end of segment 214 via linking oligonucleotide 218, providing a second single stranded portion, and yielding a completely contiguous or circular template sequence.

[00115] In addition, the templates of the invention, by virtue of their inclusion of double stranded segments, provide consensus through the identification of both the sense and antisense strand of such sequences (in both the partially and completely contiguous configurations).

[00116] By way of example, and with reference to Figures 20A and 20B, with respect to a partially contiguous template shown in Figure 20A, obtaining the entire sequence, e.g., that of segments 202, 204 and 206 provides a measure of consensus by virtue of having sequenced both sense strand, e.g., segment 202, and the antisense strand, e.g., segment 204. In addition to providing sense and antisense consensus within a single template molecule that can be sequenced in one integrated process, the presence of linking segment 206 also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., 202, is completed and the other begins, e.g., 204. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. The progress of sequencing

processes is schematically illustrated in Figure 21A. In particular, as shown, a sequencing process that begins, e.g., is primed, at the open end of the partially contiguous template, proceeds along the first or sense strand, providing the nucleotide sequence (A) of that strand, as represented in the schematic sequence readout provided. The process then proceeds around the linking oligonucleotide of the template, providing the nucleotide sequence (B) of that segment. The process then continues along the antisense strand to the A sequence, and provides the nucleotide sequence (A'), which provides consensus data for the sense strand as its antisense counterpart. As noted, because the B sequence may be exogenously provided, and thus known, it may also provide a registration sequence indicating a point in the sequence determination at which the data transitions from sense to antisense strands.

[00117] With respect to completely contiguous or circular template sequences configured in accordance with the invention, the consensus potential is further increased. In particular, as with the partially contiguous sequences shown in Figure 20A, the completely contiguous sequences also provide sense and antisense consensus. In addition, such templates provide for the potential for iterative sequencing of the same molecule multiple times, by virtue of the circular configuration of the template. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining consensus for each segment from the complementary sequences, as well as consensus within each segment, by repeatedly sequencing that segment. This is schematically illustrated in Figure 21B, again with a representative illustration of a sequence readout provided. As shown, a sequencing process that is primed at one end, e.g., primed within one linking oligonucleotide sequence, e.g., linking oligonucleotide 218 of Figure 20, proceeds along the first or sense strand 214, again providing the nucleotide sequence A of that strand. The sequence process then proceeds around the first linking oligonucleotide, e.g., linking oligonucleotide 216 from Figure 20, to provide the nucleotide sequence B of that segment of the template. Proceeding along the antisense strand, e.g., segment 212 of Figure 20B), provides the nucleotide sequence A', which is again, complementary to sequence A. The sequencing process then continues around the template providing the nucleotide sequence for the other linking

oligonucleotide, e.g., linking oligonucleotide 218 of Figure 20B, where the illustrated sequencing process began, providing nucleotide sequence C. Because the template is circular, this process can continue to provided multiple repeated sequence reads from the one template, e.g., shown as providing a second round of the sequence data A-B-A'-C-A-B-A'. Thus, sequence redundancy comes from both the determination of complementary sequences A and A', and the repeated sequencing of each segment. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing. Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post synthesis.

AMENDMENT TO DRAWINGS

Please add Figure 20A, 20B, 21A and 21B submitted herein to Figures 1-19 in the application as filed. These figures are referred to in paragraphs [00113] - [00117] described above in the previous section. No new matter has been added.

REMARKS

Amendments to the Specification

The application as filed, in paragraph [00110], states that "Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which are incorporated by reference herein in their entireties for all purposes." Applicants note that in addition to incorporation by reference, the subject application claims priority to U.S.S.N. 61/099,696.

Applicants have added paragraphs [0058]-[0059]. These paragraphs provide text derived from paragraphs [0009]–[0010] of the U.S.S.N. 61/099,696 application as filed (i.e., descriptions of figures 2A, 2B, 3A, and 3B), which was incorporated by reference into the subject application in the excerpt above. The figure numbers have been changed to reflect new numbering in the subject application. No new matter has been added.

Applicants have added paragraphs [00113]-[00117]. These paragraphs provide text derived from paragraphs [0028]–[0035] of the U.S.S.N. 61/099,696 application as filed, which was incorporated by reference into the subject application in the excerpt above. The figure numbers have been changed to reflect new numbering in the subject application. No new matter has been added.

Amendments to the Drawings

Figures 20A, 20B, 21A and 21B provided herein are copies of figures 2A, 2B, 3A and 3B of the U.S.S.N. 61/099,696 application as filed, which was incorporated by reference into the subject application in the excerpt above. The figure numbers have been changed to reflect new numbering in the subject application. No new matter has been added.

CONCLUSION

Based upon the foregoing amendments and remarks, Applicant believes the instant application is in condition for allowance and action toward that end is respectfully requested. If the Examiner believes a telephone conference would expedite the prosecution of this application, the Examiner is encouraged to contact Applicant's undersigned representative at (650) 521-8127.

Respectfully submitted,

December 20, 2016
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF
CALIFORNIA, INC.
1380 Willow Road
Menlo Park, CA 94025

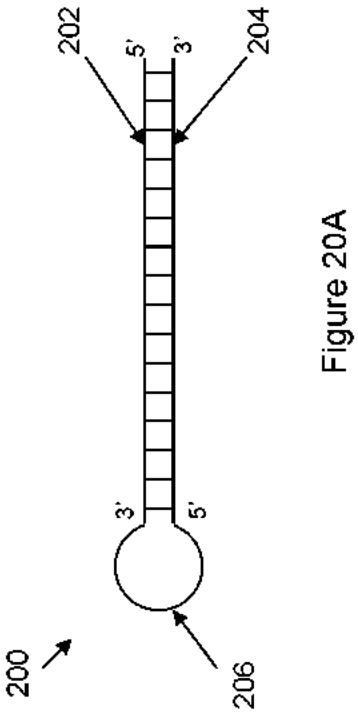


Figure 20A

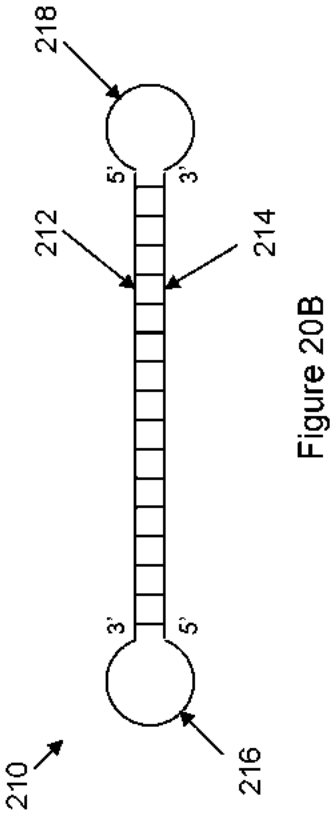


Figure 20B

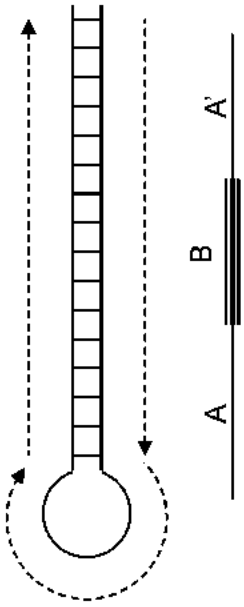


Figure 21A

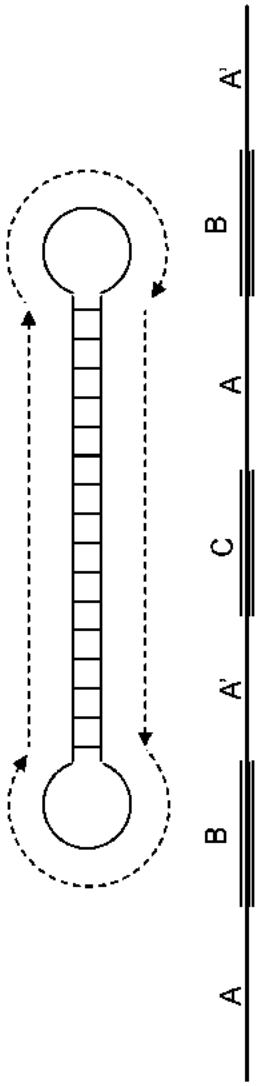


Figure 21B

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation application of U.S. Patent Application No. 14/708,603, filed May 11, 2015, which is a continuation application of U.S. Patent Application No. 14/091,961, filed November 27, 2013, now U.S. Patent No. 9,057,102, which is a continuation application of U.S. Patent Application No. 12/982,029, filed December 30, 2010, now U.S. Patent No. 8,628,940, which (1) claims the benefit of U.S. Provisional Application No. 61/099,696, filed September 24, 2008; (2) claims the benefit of U.S. Provisional Application No. 61/139,402, filed December 19, 2008; and (3) is a continuation-in-part application of U.S. Patent Application No. 12/413,226, filed March 27, 2009, now U.S. Patent No. 8,143,030, the full disclosures of all of which are incorporated herein by reference in their entireties for all purposes.

[0002] This application is also related to U.S. Provisional Application No. 61/072,160, filed March 28, 2008, U.S. Patent Application No. 12/383,855, filed March 27, 2009, now U.S. Patent No. 8,236,499, and U.S. Patent Application No. 12/413,258, filed March 27, 2009, now U.S. Patent No. 8,153,375, all of which are incorporated herein by reference in their entireties for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0003] Not Applicable.

BACKGROUND OF THE INVENTION

[0004] The use of optically detectable labeling groups, and particularly those groups having high quantum yields, e.g., fluorescent or chemiluminescent groups, is ubiquitous throughout the fields of analytical chemistry, biochemistry, and biology. In particular, by providing a highly visible signal associated with a given reaction, one can better monitor that reaction as well as any potential effectors of that reaction. Such analyses are the basic tools of life science research in genomics, diagnostics, pharmaceutical research, and related fields.

[0005] Such analyses have generally been performed under conditions where the amounts of reactants are present far in excess of what is required for the reaction in question. The result of this excess is to provide ample detectability, as well as to compensate for any damage caused by the detection system and allow for signal detection with minimal impact on the reactants. For example, analyses based on fluorescent labeling groups generally require the use of an excitation radiation

source directed at the reaction mixture to excite the fluorescent labeling group, which is then separately detectable. However, one drawback to the use of optically detectable labeling groups is that prolonged exposure of chemical and biochemical reactants to such light sources, alone, or when in the presence of other components, e.g., the fluorescent groups, can damage such reactants. The traditional solution to this drawback is to have the reactants present so far in excess that the number of undamaged reactant molecules far outnumbers the damaged reactant molecules, thus minimizing or negating the effects of the photo-induced damage.

[0006] A variety of analytical techniques currently being explored deviate from the traditional techniques. In particular, many reactions are based on increasingly smaller amounts of reagents, e.g., in microfluidic or nanofluidic reaction vessels or channels, or in "single molecule" analyses. Such low reactant volumes are increasingly important in many high throughput applications, such as microarrays. The use of smaller reactant volumes offers challenges to the use of optical detection systems. When smaller reactant volumes are used, damage to reactants, such as from exposure to light sources for fluorescent detection, can become problematic and have a dramatic impact on the operation of a given analysis. In other cases, other reaction conditions may impact the processivity, rate, fidelity, or duration of the reaction, including salt or buffer conditions, pH, temperature, or even immobilization of reaction components within observable reaction regions. In many cases, the effects of these different reaction or environmental conditions can degrade the performance of the system over time. This can be particularly detrimental, for example, in real-time analysis of reactions that include fluorescent reagents that can expose multiple different reactions components to optical energy. In addition, smaller reactant volumes can lead to limitations in the amount of signal generated upon application of optical energy.

[0007] Further, in the case of sequencing-by-synthesis applications, an additional challenge has been to develop ways to effectively sequence noncontiguous portions of a template nucleic acid on a single molecule. This challenge is exacerbated in template nucleic acids that contain highly repetitive sequence and/or are hundreds or thousands of nucleotides in length, such as certain genomic DNA fragments. The difficulty in generating such noncontiguous reads from a single template has hampered efforts to construct consensus sequences for long templates, for example, in genome sequencing projects.

[0008] As such, methods and systems that result in enhanced reaction performance, such as an increase in processivity, rate, fidelity, or duration of a reaction of interest, would provide useful improvements to the methods and compositions currently available. For example, methods, devices,

and systems that increase reaction performance by, e.g., mitigating to some extent photo-induced damage in a reaction of interest and/or increasing various other performance metrics for the reaction would be particularly useful.

BRIEF SUMMARY OF THE INVENTION

[0009] In a general sense, the methods provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0010] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one

reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion. The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0011] In certain aspects, the present invention is generally directed to methods, devices, and systems for enhancing the performance of illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. In certain preferred embodiments, illuminated reactions comprise one or more fluorescent or fluorogenic reactants. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In some aspects, enhancing the performance of an illuminated reaction means increasing the processivity, rate, fidelity, and/or duration of the reaction. For example, enhancing the performance of an illuminated reaction can involve reducing or limiting the effects of photo-induced damage during the reaction. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction.

[0012] In certain aspects, methods of the invention useful for characterizing an analytical reaction comprise preparing a reaction mixture and initiating the analytical reaction therein, subjecting the reaction mixture to at least one detection period and at least one non-detection period during the course of the analytical reaction, collecting data during both the detection period(s) and the non-detection period(s), and combining the collected data to characterize the analytical reaction. In certain embodiments, the analytical reaction comprises an enzyme that exhibits an improvement in performance as compared to its performance in the analytical reaction under constant illumination, and such improvement may be related to various aspects of enzyme activity, e.g., processivity, fidelity, rate, duration of the analytical reaction, and the like. In certain embodiments, stop or pause points are used to control the activity of the enzyme, and such stop or pause points may comprise elements such as large photolabile groups, strand-binding moieties, non-native bases, and others well known in the art. In certain preferred embodiments, the one or more detection periods are illuminated periods and the one or more non-detection periods are non-illuminated periods. In certain preferred embodiments, a plurality of analytical reactions disposed on a solid support are characterized, preferably in a coordinated fashion as described elsewhere herein.

[0013] In certain preferred embodiments, the analytical reaction is a sequencing reaction that generates sequence reads from a single nucleic acid template during the detection period(s) but not during the non-detection period(s). For example, the analytical reaction can comprise at least two or more detection periods and can generate a plurality of noncontiguous reads from the single nucleic acid template. In some embodiments, the single nucleic acid template is at least 100 bases in length and/or comprises multiple repeat sequences. In certain embodiments, the sequencing reaction comprises passage of the single nucleic acid template through a nanopore, and in other embodiments the sequencing reaction comprises primer extension by a polymerase enzyme.

[0014] The analytical may optionally be a processive reaction monitored in real time, i.e., during the course of the processive reaction. In preferred embodiments, such a processive reaction is carried out by a processive enzyme that can repetitively execute its catalytic function, thereby completing multiple sequential steps of the reaction. For example, a processive polymerization reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.

[0015] In certain preferred embodiments, the analytical reaction comprises at least one component comprising a detectable label, e.g., a fluorescently labeled nucleotide. In certain embodiments, the labeled component is present throughout the course of the analytical reaction, i.e., during both the detection and the non-detection periods. The method may further comprise an optical system to collect the data during the detection period, but optionally not to collect the data during the non-detection period.

[0016] In certain aspects, methods of the invention comprise providing a substrate having a reaction mixture disposed thereon and illuminating the reaction mixture on the substrate with an excitation illumination for multiple, noncontiguous periods during the course of the reaction, thereby subjecting the reaction mixture to intermittent excitation illumination. In some embodiments, the reaction mixture comprises first reactant and a second reactant, wherein an amount of photo-induced damage to the first reactant occurs as a result of interaction between the first reactant and the second reactant under excitation illumination. In certain embodiments, the method further comprises monitoring a reaction between the first and second reactants during illumination and collecting the data generated therefrom. In some embodiments, the reaction is a primer extension reaction and/or the first reactant is a polymerase enzyme. In certain embodiments, the second reactant is a fluorogenic or fluorescent molecule.

[0017] In yet another aspect, the methods are useful for mitigating photo-induced damage in an illuminated reaction by subjecting the illuminated reaction to intermittent illumination rather than constant illumination. For example, certain methods of the invention monitor a reaction mixture comprising at least one enzyme and a fluorescent or fluorogenic substrate for the enzyme, wherein interaction of the enzyme and the substrate under excitation illumination can result in altered activity of the enzyme, e.g if such excitation illumination is present over an extended period of time. Such methods can comprise directing intermittent excitation illumination at a first observation region for a first period that is less than a photo-induced damage threshold period under the intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions. As such, certain aspects of the invention lengthen a photo-induced damage threshold period for an analytical reaction through intermittent inactivation of the excitation illumination source since the photo-induced damage threshold period under intermittent illumination is longer than the photo-induced damage threshold period under constant illumination.

[0018] In a related aspect, the invention also provides methods of performing an enzyme reaction, comprising providing an enzyme within a first observation region, contacting the enzyme with a fluorescent or fluorogenic substrate for the enzyme, and directing an excitation radiation at and detecting signals from the first observation region for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0019] In further aspects, the invention provides methods of monitoring a primer extension reaction, comprising providing a polymerase enzyme within a first observation region, contacting the polymerase with at least a first fluorescent or fluorogenic nucleotide analog, and monitoring a fluorescent signal emitted from the first observation region in response to illumination with excitation radiation for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0020] In addition, the invention provides methods for generating a plurality of noncontiguous sequence reads from a single nucleic acid template molecule. Such methods generally comprise preparing a reaction mixture comprising the template molecule, a polymerase enzyme, and a set of differentially labeled nucleotides or nucleotide analogs, wherein the set comprises at least one type of nucleotide or nucleotide analog for each of the natural nucleobases (A, T, C, and G). The polymerization reaction is initiated, the polymerase begins processive incorporation of the labeled nucleotides or nucleotide analogs into a nascent nucleic acid strand, and during such incorporation the reaction is monitored by optical means to detect incorporation events, thereby generating a first sequence read. In a subsequent step, the labeled nucleotides or analogs are replaced with unlabeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed without detecting incorporation events. Subsequently, the unlabeled nucleotides or analogs are replaced with labeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed once again with real time detection of incorporation events, thereby generating a second sequence read that is noncontiguous to the first sequence read. The substitution of labeled for unlabeled, and unlabeled for labeled, nucleotides and nucleotide analogs can be repeated multiple times to generate a plurality of noncontiguous sequence reads, each of the plurality generated during a period when the labeled nucleotides or nucleotide analogs are being incorporated into the nascent strand and such incorporation is being detected in real time.

[0021] In certain aspects, devices of the invention can comprise a solid support (e.g., substrate) having an observation region, a first reactant immobilized within the observation region, and a second reactant disposed within the observation region, and a means for subjecting the observation region to at least one illuminated period and at least one non-illuminated period. In certain embodiments, interaction between the first and second reactants under excitation illumination causes photo-induced damage to the first reactant, and further wherein the photo-induced damage is reduced by subjecting the observation region to intermittent illumination. In some embodiments, the first reactant is an enzyme (e.g., a polymerase), the second reactant (e.g., a nucleotide) has a detectable label (e.g., fluorescent label), and/or the observation region is within a zero-mode waveguide. The means for subjecting the observation region to one or more illuminated and non-illuminated periods may comprise, e.g., a laser, laser diode, light-emitting diode, ultra-violet light bulb, white light source, a mask, a diffraction grating, an arrayed waveguide grating, an optic fiber, an optical switch, a mirror, a lens, a collimator, an optical attenuator, a filter, a prism, a planar waveguide, a wave-plate, a delay line, a movable support coupled with the substrate, and a movable illumination source, and the like. The device may further comprise a means for collecting the data during the illuminated period(s), such as an optical train, e.g., operably coupled to a machine comprising machine-readable medium onto which such data may be written and stored.

[0022] In further aspects, the invention provides systems for performing intermittent detection of an analytical reaction comprising reagents for the analytical reaction disposed on a solid support, a mounting stage configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the solid support detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other, and a data processing system operably coupled to the optical train. In certain preferred embodiments, the analytical reaction is a sequencing reaction and/or the solid support comprises at least one zero-mode waveguide.

[0023] In still other aspects, the invention provides systems for analyzing an illuminated reaction that is susceptible to photo-induced damage when illuminated for a period longer than an photo-induced damage threshold period, comprising a solid support having reagents for the reaction disposed thereon, a mounting stage supporting the solid support and configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the

solid support to illuminate the portion of the solid support and detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, and a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other. In some embodiments, the illuminated reaction is a sequencing reaction, e.g., a nucleotide sequencing-by-synthesis reaction. In certain embodiments, the solid support comprises at least one optical confinement, e.g., a zero-mode waveguide.

[0024] The invention provides methods of performing analytical reactions, e.g., processive analytical reactions, that include preparing a reaction mixture comprising reaction components, at least one of which is a detectable component that is detectable during one or more detection periods, and at least one of which is a clocking component that is detectable during one or more non-detection periods during the analytical reaction. The methods further comprise initiation the analytical reaction and maintaining conditions that allow the analytical reaction to proceed while subjecting it to at least one detection period and at least one non-detection period, both in the presence of the clocking component and the detectable component. In certain embodiments, the detectable component emits a detectable signal is response to excitation illumination during the detection period, but not during the non-detection period when a clocking signal is emitted from the clocking component. The detectable signal is collected during the detection period and the clocking signal is detected during the non-detection period, e.g., using an optical system. Optionally, the clocking signal can also be collected during the detection period and the non-detection period. In certain preferred embodiments, detection data is collected in read time during the detection period, non-detection data is collected in real time during the non-detection period, and the detection data and non-detection data are both used to characterize the analytical reaction. In some embodiments, the transition between the detection period and the non-detection period does not involve substitution and/or addition of reaction components during progression of the analytical reaction, and in other embodiments the transition does involve substitution and/or addition of reaction components, e.g., via a reaction mixture exchange. In some preferred embodiments, a plurality of analytical reactions are disposed on a solid support, subjected to intermittent illumination, monitored to collect data, and characterized based upon the data so collected.

[0025] The detectable component and clocking component are typically linked to discrete molecules in the analytical reaction. For example, the detectable component can be linked to a first subset of nucleotide analogs and the clocking component can be linked to a second subset of

nucleotide analogs in the analytical reaction mixture. Alternatively, both the detectable component and the clocking component can be linked to a single molecule, e.g., a single nucleotide or nucleotide analog, in the analytical reaction. The detectable component and clocking component can both comprise detectable labels (e.g., luminescent, fluorescent, or fluorogenic labels, including, e.g., quantum dots), and in some embodiments, different detectable labels, e.g. having different absorption peaks.

[0026] In certain preferred embodiments, an analytical reaction performed according to the invention comprises at least one enzyme, e.g., a polymerase, ligase, ribosome, nuclease, and/or kinase. In some embodiments, pause or stop points are engineered into the analytical reaction to control activity of the enzyme. Various aspects of the analytical reaction can be changed by being subjected to at least one detection period and at least one non-detection period, such aspects including but not limited to processivity, fidelity, rate, and duration, e.g. of enzyme activity.

[0027] In certain preferred embodiments, the analytical reaction is a sequencing reaction comprising a single nucleic acid template that generates sequence reads during the detection period by detecting the detectable component, and does not generate sequence reads during the non-detection period by suspending detection of the detectable component. Such a sequencing reaction typically comprises at least two or three detection periods and generates a plurality of noncontiguous sequence reads from the single nucleic acid template. In some embodiments, the template comprises multiple repeat or complementary sequences. In some embodiments, the sequencing reaction comprises passage of the single nucleic acid or a nascent strand complementary thereto through a nanopore. In some preferred embodiments, the sequencing reaction comprises primer extension by a polymerase enzyme and the detectable component is linked to a nucleotide or nucleotide analog. In some embodiments, the clocking component is linked to the polymerase enzyme, and optionally can be a multi-component label, e.g, a FRET label.

[0028] In certain aspects, the invention provides methods of mitigating photo-induced damage during an illuminated reaction that include preparing a reaction mixture having first and second reactants, where interaction of the reactants under excitation illumination can cause photo-induced damage to the first reactant. The illuminated reaction is subjected to intermittent excitation illumination characterized by periods of maximal illumination followed by periods of modified but not absent illumination. The intermittent excitation illumination reduces the amount of photo-induced damage to the first reactant during the illuminated reaction as compared to the illuminated reaction under constant maximal excitation illumination, thereby mitigating photo-induced damage

to the first reactant. In certain preferred embodiments, the illuminated reaction is a primer extension reaction. In certain preferred embodiments, the first reactant is an enzyme, e.g., a polymerase or ligase enzyme. In certain preferred embodiments, the second reactant comprises a fluorescent or fluorogenic molecule. In certain embodiments, the modified excitation illumination is illumination with a lower intensity excitation illumination than the maximal excitation illumination. In certain embodiments, a set of illumination sources provides the maximal excitation illumination and a subset of the set of illumination sources provides the modified excitation illumination.

[0029] In other aspects, the invention provides a method of sequencing a template nucleic acid that includes subjecting the template to methylation to generate at least one methylated base, subjecting the methylated base to base excision to generate at least one abasic site in the template, annealing a primer to the template nucleic acid, contacting the template with a polymerase enzyme to promote extension of the primer in a template-dependent manner, monitoring the extension of the primer in real time to generate a nucleotide sequence read complementary to the template, extending the primer until the abasic site is encountered by the polymerase, at which time the polymerase pauses on the template, and reinitiating primer extension by facilitating abasic site bypass by the polymerase. The monitoring, extending, and reinitiating steps are repeated until a desired number of nucleotide sequence reads is generated and collected, and subsequently analyzed to determine the sequence of the template nucleic acid. In certain embodiments, the contacting step occurs during a detection period or a detection period immediately follows the contacting step. In certain embodiments, a detection period ends and a non-detection period begins prior to one or more pauses of the polymerase on the template. In certain embodiments, a non-detection period is terminated simultaneous with or immediately following one or more reinitiation steps. In some embodiments, the reinitiating step comprises introduction of a pyrene to the polymerase, where the polymerase incorporates the pyrene into the nascent strand opposite and, therefore, "pairing with" an abasic site in the template. In certain preferred embodiments, the template is circular and the polymerase pauses at the same abasic site multiple times during the primer extension reaction. In other embodiments, the method further comprises terminating the monitoring when a desired length of the nucleotide sequence read is collected, e.g., by removing or modifying excitation illumination. Optionally, the desired length can be less than a length of the template nucleic acid. Additionally, the monitoring can be reinitiated subsequent to or simultaneous with the reinitiating of primer extension.

[0030] In yet further aspects, the invention provides a method of performing an illuminated reaction that includes preparing a reaction mixture comprising multiple optically detectable components that are distinguishable from one another based upon their individual signal emissions, initiating the illuminated reaction, and maintaining conditions that allow the illuminated reaction to proceed while subjecting the reaction mixture to at least one maximal illuminated period and at least one modified illuminated period during the illuminated reaction. In preferred embodiments, at least a portion of the optically detectable components are detectable during both the maximal and modified illuminated periods. In certain embodiments, the maximal illuminated period is characterized by a first excitation radiation intensity and the modified illuminated period is characterized by a second excitation radiation intensity that is less than the first excitation radiation intensity. In certain preferred embodiments, all of the optically detectable components are detectable during both the maximal and modified illuminated periods, but are distinguishable from one another during the maximal illuminated period, but are not distinguishable during the modified illuminated period. In certain embodiments, the maximal illuminated period comprises exposing the reaction mixture to a set of excitation radiation wavelengths and the modified illuminated period comprises exposing the reaction mixture to a subset of the set of excitation radiation wavelengths. In certain preferred embodiments, all of the optically detectable components are detectable and distinguishable during the maximal illuminated period, but only a subset of the optically detectable components are detectable during the modified illuminated period.

[0031] In some embodiments, the illuminated reaction is initiated during a modified illuminated period and subsequently subjected to a maximal illuminated period, where data collected during the modified illuminated period is used in the statistical analysis of data collected during the maximal illuminated period. For example, an illuminated reaction that is a polynucleotide sequencing reaction can generate sequence read data during a modified illuminated period that is subsequently used to construct a sequence scaffold for assembly of sequence read data collected during a maximal illuminated period. Additionally or optionally, the illuminated reaction is a template-directed sequencing reaction and sequence read data collected during a modified illuminated period is used to determine a rate of translocation of a polymerase during the modified illuminated period.

[0032] Some embodiments of the invention comprise performing a plurality of illuminated reactions, each of which is exposed to the set of excitation radiation wavelengths during the maximal illuminated period, but is exposed to a different subset of the set of excitation radiation

wavelengths during the modified illuminated period, such that a distinct subset of optically detectable components are detectable during the modified illuminated period for each of the plurality of illuminated reactions. In other words, for two such illuminated reactions, although all optically detectable components are detectable during their respective maximal illuminated periods, only a subset of the optically detectable components is detectable in each reaction, and the subset detectable in the first reaction is preferably different from the subset detectable in the second reaction.

[0033] In certain aspects, the invention provides methods for performing paired-end sequencing on a single template molecule. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion. A first linker ligated to the first terminal portion of the nucleic acid molecule connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and a second linker ligated to the second terminal portion of the nucleic acid molecule connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion. A template nucleic acid molecule is thereby formed comprising the double-stranded nucleic acid molecule with both the first linker and the second linker ligated thereto. The template molecule is subjected to a sequencing process in which sequence reads are generated for the first terminal portion and the second terminal portion, but sequence reads are not generated for the intermediate portion, even if the intermediate portion is processed during the sequencing process, e.g., by a polymerase. In some embodiments, the first linker and second linker are identical, and in other embodiments they are different from one another, i.e., not identical. In certain embodiments, the first and second linkers comprise complementary regions and can be hybridized to one another prior to one or both of the ligating steps. In some cases, hybridized linkers that are ligated to the ends of a double-stranded nucleic acid molecule are separated prior to subjecting the molecule to a sequencing reaction, and in some cases the hybridized linkers remain hybridized during at least a portion of the sequencing reaction. For example, in a template-directed sequencing reaction, a polymerase capable of strand displacement separates the hybridized linkers as it sequences the template. In certain preferred embodiments, the sequencing process comprises at least one detection period (e.g., an illuminated period) and at least one non-detection period (e.g., a non-illuminated period) such that the intermediate portion of the template molecule is subjected to the sequencing process during the non-detection period. In some embodiments, the template is fragmented after ligation to remove the intermediate portion. The

sequencing process can generate redundant sequence data from one or both of the first terminal portion and the second terminal portion, and/or can generate sequence data from an additional portion of the template molecule that is noncontiguous with the first terminal portion and the second terminal portion. In preferred embodiments, the sequencing process involves circularizing the template molecule by separating the complementary strands of the template molecule and using the complementary strands in template-directed nascent strand synthesis catalyzed by a single polymerase enzyme. Optionally, the template molecule can comprise a primer binding site, a registration sequence, and/or a synthesis blocking moiety. The primer binding site, a registration sequence, or synthesis blocking moiety can be present in one or both of the linkers, or can be located elsewhere within the template molecule. In some cases, the synthesis blocking moiety is selected from the group consisting of an abasic site, a nick, a synthetic linker, a non-native nucleotide or analog thereof, a primer, a large photolabile group, a strand-binding moiety, a damaged base, and a modified base. The synthesis blocking moiety can permanently or temporarily block progression of the sequencing process, e.g., by interfering with the activity of an enzyme, e.g., a polymerase enzyme. In certain preferred embodiments, the synthesis blocking moiety is an abasic site, e.g., introduced by a DNA glycosylase.

[0034] In some aspects, the invention provides methods for generating a nucleic acid construct for analytical reactions. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion; providing a first stem-loop linker hybridized to a second stem-loop linker; ligating the first stem-loop linker to the first terminal portion of the nucleic acid molecule, wherein the first stem-loop linker connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and ligating the second stem-loop linker to the second terminal portion of the nucleic acid molecule, wherein the second stem-loop linker connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion, thereby generating the nucleic acid construct. Optionally, the nucleic acid construct can be subjected to fragmentation after the ligating of steps c and d, wherein the fragmentation removes the intermediate portion from the nucleic acid construct and introduces two double-stranded termini. The method can further include ligating the two double-stranded termini to one another. In some embodiments, one of the stem-loop linkers comprises a primer binding site, registration sequence, or a synthesis blocking moiety that is absent from the other stem-loop linker.

[0035] In further aspects, the invention includes a single template nucleic acid molecule comprising a duplex region; a first linker linking termini at a first end of the duplex region; a second linker linking termini at a second end of the duplex region, wherein a region of the first linker is complementary to a region of the second linkers. Optionally, the single template molecule comprises the first and second linkers hybridized with one another. In some embodiments, the duplex region is separated or melted apart to transform the single template nucleic acid molecule into a topologically single-stranded, circular nucleic acid molecule. Further, the invention provides a composition comprising a single, optically resolvable polymerase enzyme in association with a single-stranded circular nucleic acid molecule, wherein the single-stranded circular nucleic acid molecule comprises first, second, third, and fourth regions, and further wherein the first region is complementary to the second region, and the third region is complementary to the fourth region, and further wherein the regions are ordered on the single-stranded circular nucleic acid molecule as follows: first region, third region, second region, fourth region.

[0036] In still further aspects of the invention, machine-implemented methods for transforming nucleotide sequence read data into consensus sequence data, wherein the nucleotide sequence read data is generated by sequencing a target region of a template nucleic acid multiple times, and the consensus sequence data is representative of a most likely actual sequence of the template nucleic acid. Such machine-implemented methods can comprise various steps, such as a) mapping the nucleotide sequence data to a target sequence using a local alignment method that produces a set of local alignments comprising an optimal local alignment and sub-optimal local alignments, b) enumerating the set of local alignments, c) constructing a weighted directed graph wherein each local alignment in the set of local alignments is represented as a node, thereby generating a set of nodes in the weighted directed graph, d) drawing edges between pairs of nodes in the weighted directed graph if the pair represents a potential reconstruction of the template nucleic acid, e) assigning weights to the edges drawn in step d, wherein a given weight for a given edge represents the log-likelihood that a given pair of nodes connected by the given edge is truly a reconstruction of the template nucleic acid, f) finding the shortest path to each node in the weighted directed graph, thereby generating a set of shortest paths for the weighted directed graph, g) ranking the set of shortest paths to determine the best assignment, and h) storing the results of steps a-g on a machine-readable medium. In certain embodiments, the steps of the machine implemented methods are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Also provided are

computer program products comprising a computer usable medium having computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the machine-implemented methods of the invention, and machine-readable medium on which the results of the method steps are stored. The invention further includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods.

[0037] In certain aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme (e.g., a polymerase) during a time t, where time t occurs during a non-detection period. Such a method comprises, in certain embodiments, developing a probability model *p(v)* to describe an observed distribution of enzyme velocities during one or more detection periods; sampling velocities from *p(v)*; summing and recording the velocities sampled in step b to produce a sum that is an estimate of x/τcorr; and repeating the sampling, summing, and recording M times to generate a distribution of sums that are estimates of x/τcorr, with the distribution of sums being the distribution of the distance x travelled by an enzyme during a time t. Preferably, at least some of the steps are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Optionally, the enzyme is a polymerase enzyme. In some embodiments, multiple enzymes are observed simultaneously and the probability model *p(v)* is determined independently for each of the multiple enzymes. In certain preferred embodiments,

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}.$$

[0038] In further aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme during a time t, where time t occurs during a non-detection period. In some embodiments, the method comprises estimating a distribution of local rates p(v), making independent identically distributed draws of N = t/τcorr velocities from from p(v); summing the velocities; recording the velocities summed in c) as an estimate of x/τcorr; and repeating b-d M times, e.g., where M is preferably at least 1000. Optionally, p(v) is determined using a Hidden Markov Model or the autocorrelation function $\langle \delta v(t)\delta v(t+\Delta)\rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$. The invention further

includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods, as well as a machine-readable medium on which the results of the steps of the methods are stored.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions, whether illumination is initiated before (A) or after (B) initiation of the reaction.

[0040] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination, including depictions of multiple reactions arrayed on a solid support (A) and prophetic data (B) from certain embodiments of the invention.

[0041] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions on a solid support (A) using intermittent illumination and a mask (B). A graph (C) depicts prophetic data from certain embodiments of the invention.

[0042] Figure 4 provides additional embodiments of masks for use in the methods of the invention, including a mask that allows illumination of columns of reactions (A) and a mask that allows illumination of every other reaction in a row and column (B).

[0043] Figure 5 illustrates an aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Figure 5A illustrates a solid support comprising four quadrants, each quadrant containing a different sample. Figure 5B illustrates a mask design for selective illumination of the substrate. Figures 5C and 5D demonstrate various positions of the mask on the solid support.

[0044] Figure 6 provides an illustration of paths in a sequence alignment matrix representing sequencing data from a SMRTbell™ template.

[0045] Figure 7 illustrates a hypothetical directed graph.

[0046] Figure 8 provides data from single-molecule sequencing-by-synthesis reactions. Figure 8A provides data from a two-minute interval beginning at initiation of the reactions, i.e., from 0-120 seconds. Figure 8B provides data from a second two-minute interval from 300-420 seconds. Figure 8C provides data from a third two-minute interval from 600-720 seconds.

[0047] Figure 9 schematically illustrates one embodiment of a system for use with the methods, devices, and systems of the invention.

[0048] Figure 10 provides a graphical representation of rates of polymerase activity on different portions of a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0049] Figure 11 provides a graphical representation of the average rate of polymerase translocation over a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0050] Figure 12 provides a distribution of the physical coverage of a template nucleic acid achieved during a sequencing reaction utilizing intermittent illumination, with A showing mapping to a reference sequence with sequence reads (and portions thereof) that do not map to the reference excluded and B showing a similar mapping that further includes sequence reads corresponding to insertions in the template that are absent from the reference sequence.

[0051] Figure 13 provides a distribution of the physical coverage provided by sequence reads generated during sequencing reactions utilizing intermittent illumination across an approximately 40 kb template nucleic acid.

[0052] Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described herein and a reference sequence.

[0053] Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system.

[0054] Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters $P(S \rightarrow P) = 1/24$; $P(P \rightarrow S) = 1/11$; and $p(v) \sim \text{Gamma}(48,0.25)$. Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a non-detection period.

[0055] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period.

[0056] Figure 18 provides an exemplary strategy for selectively reducing the size of a duplex fragment within a SMRTbell™ template.

[0057] Figure 19 provides an illustrative example of nucleic acid templates having nicks.

[0058] Figures 20A and 20B illustrate two exemplary embodiments of template constructs used in the present invention.

[0059] Figures 21A and 21B schematically illustrate redundant or consensus sequencing using the constructs shown in Figures 20A and 20B.

DETAILED DESCRIPTION OF THE INVENTION

[0060] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing devices, formulations and methodologies which are described in the publication and which might be used in connection with the presently described invention.

[0061] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polymerase" refers to one agent or mixtures of such agents, and reference to "the method" includes reference to equivalent steps and methods known to those skilled in the art, and so forth. Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either both of those included limits are also included in the invention.

[0062] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention. Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

I. General

[0063] In a general sense, the methods, devices, and systems provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in

processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0064] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion.

[0065] The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions

(e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. For example, such an alteration in reaction conditions during the second stage may result in an increase in reaction rates, e.g., speeding up the progression of a template nucleic acid through a nanopore; or may reduce exposure of reaction components to harmful radiation or other reaction condition related to detection of the products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0066] "Intermittent detection," as used herein, generally refers to a means of monitoring a reaction that is carried out intermittently during the course of the reaction. Intermittent detection may refer to intermittent use of one or more monitoring methods, but does not necessarily mean that all means of monitoring a given reaction are intermittently halted. For example, monitoring of one or more nucleotide incorporations to generate nucleotide sequence reads may be intermittently halted while other aspects of a sequencing reaction are constantly monitored, e.g., temperature, reaction time, pH, etc. In certain embodiments, intermittent detection is achieved by intermittent or differential illumination of a given reaction, e.g., a reaction that uses an illumination system to detect reaction products and/or progression. Although various aspects of the invention are described herein in terms of embodiments using intermittent illumination, it should be understood that where applicable intermittent detection by other means (e.g., electrochemical, radiochemical, etc.) can be utilized in the methods of the invention. Likewise, a stage of a reaction during which an intermittent detection method is active may be referred to as a "detection period" and a stage of a reaction during which an intermittent detection method is inactive may be referred to as a "non-detection period." In illuminated reactions, such periods may also be referred to as "illuminated periods" and "non-illuminated periods," respectively, although it is to be understood that the term "non-illuminated period" included periods in which illumination may be present but altered as compared to illumination during an "illuminated period." For example, a non-illuminated period may be characterized by a complete absence of illumination, or a modification of illumination, including but not limited to changes in wavelength, frequency, intensity, and/or number of illumination sources. Alternatively or additionally, reaction components that are excited by the illumination source(s) may be modified or removed from a reaction mixture to create a non-illuminated period.

For example, a fluorescent dye detected during an illuminated period may be removed from the reaction mixture, e.g., by buffer exchange, thereby producing a non-illuminated period during which time the fluorescent dye cannot be detected even if the excitation illumination is present. In a further example, a non-illuminated period can indicate a period during an illuminated reaction during which a type of illumination-based detection that occurs during an illuminated period is not occurring, e.g., the identity of fluorescently labeled nucleotides incorporated into a nascent strand is not being detected or recorded.

[0067] In certain aspects, the present invention is generally directed to improved methods, devices, and systems for performing illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In certain preferred embodiments, illuminated reactions comprise one or more fluorogenic or fluorescent components. In accordance with certain methods of the invention, such illuminated analyses are subjected to intermittent detection (e.g., data collection) for one or more aspects of the data typically collected for a given reaction. For example, aspects of the data typically collected for nucleotide sequencing reactions include nucleotide sequence data, read quality data, signal to background ratios, reaction rates and durations, measures of the fidelity of the reaction, reaction times, and the like. In certain preferred embodiments, nucleotide sequence data is iteratively collected during an ongoing sequencing reaction to generate nucleotide sequence reads for at least two or more noncontiguous regions of a template nucleic acid molecule. Such iterative sequence data acquisition may be achieved in various ways depending on the sequencing technology in use. For example, in sequencing methods that utilize luminescent components that generate a signal indicative of the identity of a base position, iterative sequence data collection may be achieved by removing or altering an illumination source (or a reaction relative to an illumination source), substituting the luminescent components for unlabeled components that do not generate signal, or otherwise interrupting signal acquisition in the experimental system.

[0068] In certain preferred embodiments, such illuminated reactions are illuminated for an amount of time that permits the effective performance of the analysis. Traditionally, illuminated reactions are illuminated from initiation through completion, and the time during which reaction data may be reliably collected is dictated by the progression (as measured by, e.g., processivity,

rate, fidelity, duration, etc.) of the reaction under constant illumination. Some reactions are sensitive to such constant illumination, which can reduce their performance (e.g., processivity), and thereby prevent collection of data from later stages of the reaction, i.e., stages that would otherwise occur if the reaction were carried out with no illumination. The present invention provides methods for performing illuminated reactions comprising subjecting the reactions to intermittent illumination. Such intermittent illumination can increase performance (e.g., processivity, rate, fidelity, duration, etc.) of the reactions, thereby allowing generation of data that cannot be collected under constant illumination, such as data from later stages of an ongoing reaction whose progression is compromised under constant illumination. For example, in sequencing-by-incorporation reactions the use of intermittent excitation illumination can increase processivity, which has the benefit of providing sequence reads more distal from the polymerase binding/initiation site than such reactions subjected to constant exposure to excitation illumination.

[0069] Further, it is an object of the instant invention to provide sequence data from noncontiguous regions of a nucleic acid template in a single reaction. Other commercially available platforms have attempted to achieve such noncontiguous sequence data through, e.g., complex cloning and sequencing strategies. The present invention provides a clear advantage over such strategies by providing a simple and economical solution that is applicable across various platforms, and is particularly applicable to illuminated, single-molecule sequencing-by-incorporation reactions.

[0070] In preferred embodiments, illuminated reactions for use with the instant invention are nucleic acid sequencing reactions, e.g., sequencing-by-incorporation reactions. In preferred embodiments, such an illuminated reaction analyzes a single molecule to generate nucleotide sequence data pertaining to that single molecule. For example, a single nucleic acid template may be subjected to a sequencing-by-incorporation reaction to generate one or more sequence reads corresponding to the nucleotide sequence of the nucleic acid template. For a detailed discussion of such single molecule sequencing, see, *e.g.*, U.S. Patent Nos. 6,056,661, 6,917,726, 7,033,764, 7,052,847, 7,056,676, 7,170,050, 7,361,466, 7,416,844; Published U.S. Patent Application Nos. 2007-0134128 and 2003/0044781; and M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, SCIENCE 299:682-686, January 2003 Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations, all of which are incorporated herein by reference in their entireties for all purposes. In some embodiments, a plurality of single nucleic acid templates are analyzed separately and often simultaneously to generate a plurality of sequence reads

corresponding to the nucleotide sequences of the plurality of nucleic acid templates. In certain preferred embodiments, the plurality of nucleic acid templates includes at least two nucleic acid templates that comprise identical nucleotide sequences such that analysis of the two nucleic acid templates generates overlapping sequence reads. In certain preferred embodiments, at least one of the nucleic acid templates is configured to provide redundant sequence data in a single sequence read, e.g., via duplications, sense and antisense sequences, and/or circularization.

[0071] Certain aspects of the invention are directed to methods, devices, and systems for generating a sequence scaffold for a nucleic acid template, e.g., chromosome, genome, or portion thereof. A sequence scaffold as used herein refers to a set of sequence reads that extends across at least a portion of a nucleic acid template. In some embodiments, such a sequence scaffold is used to generate a consensus sequence for the nucleic acid template. In some embodiments, the nucleic acid template is very large, e.g., at least about 100, 1000, 10,000, 100,000, or more bases or base pairs in length. In some embodiments, the sequence scaffold and/or consensus sequence is based on at least 1-, 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of at least a portion of the nucleic acid template. In some preferred embodiments, the portion of the nucleic acid is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the entire length of the nucleic acid template.

[0072] In certain aspects, the invention is particularly suitable for sequencing nucleic acid templates interspersed with repetitive elements. Such repetitive elements present major logistical and computational difficulties for assembling fragments produced by sequencing strategies, especially those with read-lengths that are too short to encompass unique reads outside the repeat region. For example, the human T-cell receptor locus contains a five-fold repeat of a trypsinogen gene that is 4 kbp long and that varies 3 to 5% between copies. Therefore, a sequencing strategy that cannot provide nucleotide sequence information that spans at least 20 kb for a single molecule containing the locus will have difficulty providing consensus sequence for the locus. Further, Alu repeats (~300 bp retrotransposons) are also problematic because they cluster and can constitute up to 50-60% of the template sequence, with copies varying from 5-15% between each other. The human genome contains an estimated one million Alu repeats and 200,000 LINE elements (average length ~1000 bp), representing roughly 10% and 5% of the entire genome, respectively. In certain embodiments, the present methods facilitate efficient and accurate sequence determination for long templates comprising such repetitive sequences, in part because the present methods do not rely solely on sequence overlap to generate consensus sequences, but also include information related to

the expected location of the polymerase on the template nucleic acid, thereby linking a particular sequence read to a particular location on the template nucleic acid. This greatly facilitates accurate assembly of sequence reads to generate sequence scaffolds and/or consensus sequences.

[0073] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads in an illuminated sequencing-by-incorporation reaction that are distal from one another (i.e., noncontiguous) on a single nucleic acid template by removing the excitation illumination during the course of the reaction, and subsequently reinitiating the excitation illumination. Sequence reads are generated only during the periods of time when the excitation illumination is present, resulting in a "gap" between the sequence reads from a single template nucleic acid that corresponds to the time during which the excitation illumination was absent but the incorporation of nascent nucleotides continued "in the dark." As such, the number of sequence reads generated for a given template nucleic acid is equal to the number of periods during which the excitation illumination is present.

[0074] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads from a plurality of nucleic acid templates comprising identical nucleotide sequences. In some embodiments, the multiple sequence reads are not all from the same region of the nucleic acid templates. In some embodiments, there is overlap between the multiple sequence reads. In some embodiments, a single sequence read is generated from each of the plurality of nucleic acid templates, and in other embodiments multiple noncontiguous sequence reads are generated from each of the plurality of nucleic acid templates. In certain preferred embodiments, the multiple noncontiguous sequence reads from each of the plurality of nucleic acid templates together extend across the nucleic acid templates such that they can be combined to provide a consensus sequence for the identical nucleotide sequence in the nucleic acid templates. In some embodiments, the consensus sequence is based on at least 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of the identical nucleotide sequence. In some embodiments, the identical nucleotide sequence represents at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the nucleic acid template.

[0075] Certain aspects of the invention are directed to methods, devices, and systems for reducing or limiting the effects of photo-induced damage during illuminated reactions, particularly reactions that employ fluorescent or fluorogenic reactants. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction. Without being bound to a particular theory or

mechanism of operation, some illuminated reactions are subject to photo-induced damage that can hinder progression of the reaction, e.g., via damage to reaction components, such as enzymes, cofactors, templates, etc. As such, the illumination of the illuminated reaction can directly or indirectly negatively impact progression of the reaction, and such an impact can be measured based on various characteristics of the reaction progression, e.g., processivity, rate, fidelity, duration, etc. The present invention provides methods for subjecting an illuminated reaction to intermittent exposure to illumination, which reduces the amount of photo-induced damage at a given time during the reaction, allowing the reaction to proceed further than it does when constantly exposed to the illumination.

[0076] In some embodiments, the methods herein may further comprise the addition of one or more photo-induced damage mitigating agents (e.g., triplet-state quenchers and/or free radical quenchers) to the illuminated reaction. Such photo-damage mitigating agents are generally known to those of skill in the art. Further discussion of photo-induced damage and related compounds, compositions, methods, devices, and systems are also provided in U.S. Pub. No. 20070161017, filed December 1, 2006; and U.S.S.N. 61/116,048, filed November 19, 2008, which are incorporated by reference herein in their entireties for all purposes.

II. Intermittent Illumination of Analytical Reactions

[0077] Certain aspects of the invention are generally directed to improved methods for performing illuminated analyses. The terms "illuminated analysis" and "illuminated reaction" are used interchangeably and generally refer to an analytical reaction that is occurring while being illuminated (e.g., with excitation radiation), so as to evaluate the production, consumption, and/or conversion of luminescent (e.g., fluorescent) reactants and/or products. As used herein, the terms "reactant" and "reagent" are used interchangeably. As used herein, the terms "excitation illumination" and "excitation radiation" are used interchangeably. In certain embodiments, the illuminated reaction is a sequencing reaction, e.g., a sequencing-by-incorporation reaction. In certain embodiments, the illuminated reaction is designed to analyze a single molecule, e.g., by ensuring the molecule is optically resolvable from any other molecule being analyzed and/or in the reaction mixture. In certain embodiments, one or more components of the reaction are susceptible to photo-induced damage directly or indirectly elicited by an excitation radiation source. In certain preferred embodiments, an illuminated reaction is subjected to intermittent excitation radiation during the course of the illuminated reaction. In certain preferred embodiments, a sequencing-by-

incorporation reaction is subjected to intermittent excitation radiation during the course of a polymerization reaction to generate a plurality of noncontiguous sequence reads from a single nucleic acid template.

[0078] In certain aspects, the methods herein provide benefits over methods currently used for sequencing large template nucleic acids, such as human genomes. For example, the traditional shotgun sequencing approach entails sequencing nucleic acid fragments and analyzing the resulting sequence information for overlap and similarity to known sequences to construct the complete sequence of the template nucleic acid. One disadvantage to the shotgun approach is that assembly may be difficult if the template nucleic acid comprises numerous repeated sequences, and the inability to assemble a genomic sequence in repeat regions leads to gaps in the assembled sequence. (See, e.g., Myers, G.; "Whole-Genome DNA Sequencing" in Computing in Science and Engineering; Vol 1, Issue 3; pgs. 33-43; May/Jun 1999.) One method of resolving these gaps is to sequence fragments large enough to span the repeat regions, but sequencing large fragments can be difficult and time-consuming. Another approach to spanning a gap is to determine the sequence of two ends of a large fragment which has known spacing and orientation, and this approach is generally termed paired end sequencing (see, e.g., Smith, M. W. et al., (1994) Nature Genetics 7:40-47; and U.S. Pub. No. 2006/0292611, filed June 6, 2006, both of which are incorporated by reference herein in their entireties for all purposes). This method is limited by the requirement for information about the spacing and orientation of the ends of the long fragment, and/or complex sample preparation of the nucleic acid template. The present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.) or complex sample preparation, thereby allowing economical, efficient, and effective de novo sequencing or resequencing of long template nucleic acids.

[0079] In certain aspects, the methods herein provide various strategies for achieving intermittent illumination of illuminated reactions. Essentially, at least one type of illumination (e.g., excitation illumination) is present for at least one time period ("illuminated period") and absent during at least one other time period ("non-illuminated period") during an illuminated reaction. As described above, the term "non-illuminated" indicates a change in illumination including, but not limited to a complete absence of illumination. For example, a non-illuminated period may also be characterized by a different illumination source or intensity than an illuminated period, or by a change in reaction components, e.g., detectable labels. In general, at least one type of data collected

during an illuminated period (e.g., nucleotide sequence data) is not collected during a non-illuminated period. An absence of the illumination may be due to, e.g., inactivation of the illumination source (e.g., laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source), removal of the illuminated reaction from the illumination source (or vice versa), or may be due to blockage of the illumination from the reaction, as discussed below. Modifications to the illumination may be due to, e.g., adjustment of the intensity of an illumination source, or a substitution of one illumination wavelength and/or frequency for another. Further, components detectable during an illuminated period may be removed from the reaction mixture during a non-illuminated period, e.g., a fluorescently labeled nucleotide may be replaced with an unlabeled nucleotide. Knowledge of the rate of the reaction and the time during which the illumination is absent is used to estimate the progress of the reaction during the non-illuminated period. For example, if a reaction proceeds such that one molecule is incorporated into a macromolecule per second, and the illumination is absent for 20 seconds, it can be estimated that 20 molecules were incorporated during the non-illuminated period. This information is useful during data analysis to provide context for the reaction data collected during the illuminated period(s). For example, in a sequencing-by-incorporation reaction the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. Further the processivity of an enzyme during a non-illuminated period (or other type of non-detection period) can be manipulated or adjusted by methods known to those of skill in the art. In particular, the kinetics of replication by a polymerase enzyme can be altered by changing the chemical environment in which it operates, and such methods are further described, e.g., in U.S. Patent Application Nos. 12/414,191, filed March 30, 2009; 12/537,130, filed August 6, 2009; and U.S. Patent Application No. [unassigned], attorney docket no. 105-006301US, entitled "Engineering Polymerases and Reaction Conditions for Modified Incorporation Properties," filed September 4, 2009, the disclosures of all of which are incorporated herein by reference in their entireties for all purposes. For example, methods are provided for adjusting the enzyme activity, and these methods find particular relevance in the instant invention when used to enhance accuracy

during detection periods, and to enhance processivity during non-detection periods. Information regarding enzyme translocation rate and processivity is useful for positioning the sequence reads for a single template nucleic acid relative to one another in the construction of a sequence scaffold and/or consensus sequence for the template nucleic acid.

[0080] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions. A reaction mix is prepared at step 100. In process A shown on the left, illumination of the reaction 105 is begun prior to initiation of the reaction 110, which allows "illumination data" to be collected at initiation. (In an alternative embodiment, illumination may commence simultaneously with initiation of the reaction.) "Illumination data" as used herein refers to data collected during an illuminated period, e.g., the length of the illuminated period and luminescent signal(s) from the reaction product. At least one non-illuminated period 115 occurs during the course of the reaction, followed by at least one additional illuminated period 120. Multiple additional non-illuminated and illuminated periods may follow. During the illuminated periods (105 and 120), illumination data is collected 175. During the non-illuminated period(s), non-illumination data is collected 180. As used herein, "non-illumination data" refers to data collected during a non-illuminated period, e.g., the length of the non-illuminated period can be monitored. In process B shown on the right, the reaction is initiated 155 during a first non-illuminated period 150. At least one illuminated period 160 occurs during the course of the reaction, optionally followed by at least one additional non-illuminated period 165. Multiple additional illuminated and non-illuminated periods may follow. As for process A, illumination data is collected 175 during the illuminated period(s) 160, and non-illumination data is collected 180 during non-illuminated periods (155 and 165).

[0081] One benefit provided in certain embodiments of the invention is that the reaction need not be further manipulated after initiation (aside from the control of illumination). For example, the method can be used to analyze reaction mixtures without the need for buffer changes, addition of further reaction components, or removal of detectable components, e.g., light-activatable components such as fluorophores. For example, in a sequencing-by-incorporation reaction, labeled nucleotides may be present throughout the life of the reaction, even when the reaction is not generating nucleotide sequence data (e.g., during a non-illuminated period). This provides clear advantages over methods that require additional handling of the reaction after initiation, which tend to not only be expensive and time-consuming, but which also provide opportunities for contamination of the reaction. For example, illumination can be reinitiated at any time during the

reaction at the whim of the ordinary practitioner by simply activating the illumination. In certain preferred embodiments, the concentration of labeled nucleotides or nucleotide analogs in the reaction mixture is greater than the concentration of unlabeled nucleotides in the reaction mixture throughout the course of the reactions, and may represent at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the total nucleotides in the reaction mixture. Methods for ensuring a high ratio of labeled versus unlabeled nucleotides in a reaction mixture are known in the art and certain preferred embodiments are provided in U.S. Patent Pub. Nos. 2006/0063264, 2006/0194232, and 2007/0141598, which are incorporated herein by reference in their entireties for all purposes.

[0082] In embodiments in which a sequencing-by-incorporation reaction is subjected to intermittent illumination, the sequence reads collected during the illuminated periods are arranged in order and separated from one another by an estimated number of nucleotides incorporated into the nascent strand during the intervening non-illuminated periods. The resulting gapped read can then be used to assess certain characteristics of the template nucleic acid. When multiple identical template nucleic acids are subjected to such a sequencing-by-incorporation reaction, the resulting set of gapped reads can be combined to create a sequence scaffold and/or a consensus sequence for the template nucleic acid.

[0083] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods can be used to provide additional data during the course of the reaction, e.g., data from illuminated or non-illuminated periods. In certain preferred embodiments, such alternative labeling methods may comprise using labels that are incorporated into a product of the reaction. For example, in sequencing-by-incorporation reactions that use nucleotides comprising labeled terminal phosphates (*e.g.*, the gamma phosphate as in dNTP, or terminal phosphates on nucleotide analogs with a greater number of phosphate groups) to identify the nucleotides incorporated into a nascent polynucleotide, the reaction mixture may also include nucleotides comprising a base-linked label. During the reaction, these "base-labeled nucleotides" will be incorporated into the nascent strand, but unlike the terminal phosphate labels removed during incorporation, the base-linked labels are not cleaved from the nucleotide upon incorporation by the polymerase, resulting in a nascent strand that comprises the base-linked labels. The concentration of such base-labeled nucleotides can be adjusted in the reaction mixture to promote their incorporation into the nascent strand at a predictable rate, e.g., based on the known sequence of the template or the average frequency of a given nucleotide. The presence and/or rate of

incorporation of the base-linked labels into the nascent strand can provide a measure of the length of the nascent strand generated (and, therefore, the distance traveled by the polymerase along the template nucleic acid) during the reaction by subjecting the reaction to excitation illumination that excites the base-linked label (but preferably not the non-base-linked labels), and detecting the signal emitted. The excitation of the base-linked labels preferably occurs as a pulse during or immediately following a non-illuminated period, and is otherwise absent during the reaction. The strength of the signal is indicative of how many labels are present in the nascent strand, thereby providing a measure of the processivity of the polymerase for a given period during the ongoing reaction, e.g. during one or more illuminated or non-illuminated periods. Since the base-linked labels remain in the nascent strand, it is beneficial to minimize the amount of time those fluorophores are subjected to excitation illumination to mitigate the potential of photo-induced damage to the reaction components. As such, in preferred embodiments, the excitation illumination wavelength for the base-labeled nucleotides is different than that of other fluorescent labels in the reaction.

[0084] This method can be modified in various ways. For example, the base-labeled nucleotides may also comprise a terminal phosphate label so that their incorporation can be monitored in the same manner during an illuminated period as the non-base-labeled nucleotides. There may be a single type of base-labeled nucleotide in a reaction mixture, or multiple types may be present, e.g., each type carrying a different nucleobase. The concentration of base-labeled nucleotides in the reaction mix may be varied, although it is preferred that the ratio of base-labeled nucleotides to non-base-labeled nucleotides be relatively low. For example, in a reaction mixture comprising a single type of base-labeled nucleotide (e.g., base-labeled dATP), it is preferred that the ratio of base-labeled dATP to non-base labeled dATP be less than 1:8, and more preferably 1:10 or less. The low concentration of base-labeled nucleotides is preferred in order to minimize sterically induced polymerase stalling when incorporating multiple base-labeled nucleotides in a row. In some embodiments, the optimal ratio is pre-determined using capillary electrophoresis for any specific base-labeled nucleotide and likely homopolymer sequence prevalence. In certain preferred embodiments, at least 50, 75, 100, 125, or 150 base-labeled nucleotides are incorporated into the nascent strand during a single non-detection period. The base-labeled nucleotides may be present throughout the reaction, or may be washed in during non-illuminated periods and washed out after the pulse of excitation illumination. The reaction mixture comprising base-linked nucleotides being washed in may also include unlabeled nucleotides for incorporation during a non-detection period. During a subsequent illuminated period, a reaction mixture comprising terminal phosphate-labeled

nucleotides replaces the reaction mixture comprising base-linked nucleotides and unlabeled nucleotides. This protocol is one embodiment of the methods of the invention in which a non-detection period is not necessarily a non-illuminated period because in this case illumination may be present, but no incorporation of nucleotides is be detected.

[0085] Alternatively or in addition, a low concentration of a fifth terminal phosphate labeled nucleotide can be present in the sequencing reaction, wherein the label has a different excitation wavelength than the other labels in the reaction mixture. For example, a small proportion of one nucleotide analog, e.g., dA6P, can be labeled with the "fifth label." During non-detection periods when the sequence of incorporation of nucleotides is not being monitored, the reaction site is illuminated by excitation radiation specific for the fifth label, and this fifth label excitation radiation can be inactivated during the detection periods. Emissions detected upon incorporation of the nucleotide analog comprising the fifth label are used to "clock" the pace of the polymerase during the non-detection period, e.g., based upon the known or estimated frequency of the complementary nucleotide in the template strand. The fifth label can be chosen such that the excitation and emission radiation are less likely or unlikely to cause photo-induced damage to reaction components, e.g. by choosing a label with a long excitation wavelength (e.g., toward the red end of the visible spectrum), a label that has a low propensity for entering into a triplet state, and/or a label that has a low propensity to form a radical. Since the fifth label is being excited when other labels are not, there is no requirement for optimal spectral separation from other labels in the reaction mixture. Further, since the fifth label is not being used for sequencing, other optimizations are also not necessary, e.g., related to branching, accuracy, and the like. Various types of labels can be used as a fifth label of the invention including, but not limited to, organic and non-organic dye fluorophores. For example, latex nanoparticles or quantum dots are particularly suitable due to their lower propensity for photo-induced damage of certain analytical reaction components. In certain preferred embodiments, a quantum dot label has an emission spectrum within the same spectral window as the labels that are used to identify the sequence of base incorporations into the nascent strand ("sequencing labels") but an excitation spectrum that does not overlap those of the sequencing labels to allow detection of the fifth label emissions using the same optical system as is used to detect the sequencing label emissions.

[0086] This method can be modified in various ways. For example, more than one small subset of a nucleotide analog can be labeled with a fifth label, and in certain embodiments, a small subset of each nucleotide analog present in the reaction mixture is labeled with the fifth label.

Further, there may be a plurality of additional labels present in the reaction, each of which is present on a small subset of a single type of nucleotide analog, e.g., sixth, seventh, and eighth labels. By increasing the number of types of nucleotide analogs labeled with fifth (or sixth, seventh, eighth) labels, their frequency of incorporation is likewise increased, which improves the translocation rate calculation for the polymerase during the non-detection periods. Alternatively, each type of nucleotide analog can comprise both a sequencing label that is specific for the cognate base in the nucleotide, as well as a fifth label for clocking the polymerase. The sequencing labels are excited and detected during the detection periods and the fifth labels are excited and detected during the non-detection periods. Since every nucleotide analog is labeled with a fifth base, each incorporation event can be counted during the non-detection period and the exact rate of incorporation can be determined. Both the sequencing and fifth labels may be bound to the same or different linkers on the nucleotide analogs. In certain preferred embodiments, a linker on a nucleotide analog positions the fifth label within an illumination zone to allow excitation, but far from an enzyme (e.g., polymerase) to mitigate photo-induced damage related to excitation of and/or emission from the fifth label.

[0087] In some embodiments, the fifth label is also excited by an illumination during the detection periods. The availability of the clocking function during the detection period can be used during sequence analysis to identify positions in the resulting sequence read where a signal was not detected (resulting in an apparent "missing base" in the read) and to distinguish between true insertions and branching events in which two signals are detected for a single incorporation event.

[0088] In yet further embodiments, assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid is facilitated by using "non-illuminated periods" characterized by modified excitation illumination rather than a complete absence of excitation illumination (which can also be termed "low-illuminated periods"). For example, in some embodiments a lower intensity excitation illumination is used during the non-illuminated periods that excites one or more of the labels that are excited during the illuminated periods. As such, unlike various strategies described above, no fifth label is necessary. The lower intensity excitation illumination results in emissions that are lower intensity but still intense enough to identify an emission signal over background counts, though typically not intense enough to be used to identify the particular label generating the emission signal. For example, if label "A" and label "B" are in a reaction mixture, during an illuminated period the intensity of the signal emissions from each are high enough that the artisan can distinguish from which label a particular signal originates by the

wavelength and/or frequency of the signal. However, during a low-illuminated period the artisan can only identify that a signal emission occurs, but is unable to distinguish the originating label because its particular wavelength and/or frequency cannot be accurately determined. The decrease in excitation illumination intensity provides both a mitigation of photo-induced damage to reaction components within the observation volume while allowing the practitioner to count the emissions, and therefore the incorporations, during the non-illuminated period.

[0089] In other embodiments, multiple excitation illumination sources are used during an illuminated period, and a first subset of these illumination sources is removed during a non-illuminated period, while a second subset remains. The illumination sources that remain during the non-illuminated period may be present in the same manner as during the illuminated period, or various aspects may be altered, e.g., intensity may be reduced. For example, if labels A and B present in a reaction mixture are excited by a first illumination source and labels C and D present in the reaction mixture are excited by a second illumination source, removal of the first illumination source during the non-illuminated period results in an inability to detect labels A and B, while C and D are still detectable. Such an incomplete data set can be used to clock the progress of the reaction during the non-illuminated period(s). Further, it can also be used in various ways to facilitate the statistical analysis of data collected during the illuminated period(s). For example, for nucleotide sequencing applications (as described elsewhere herein) the incomplete data set(s) collected during non-illuminated period(s) can be used during assembly of a sequence scaffold. For example, during *de novo* sequence assembly a collection of sequences (contigs) are generated, but the order of the contigs relative to the template nucleic acid is not always apparent. The scaffolding process uses extra information to determine the correct order of the contigs. So, if only two bases are identifiable in the non-illuminated periods, the incomplete sequence reads comprising only incorporation of these two bases can be aligned to modified versions of the contigs assembled from data collected during an illuminated period, but in which the two bases not detected during the non-illuminated periods have been removed. Once the order of the contigs has been determined, the incorporation data for the two bases not detected during the non-illuminated periods is restored and the assembly of the contigs is complete. This method can be modified in various ways. For example, the practitioner may choose which illumination sources to remove during the non-illuminated periods based on various characteristics, such as their propensity to cause photo-induced damage to one or more reaction components, the propensity of the corresponding emission signal to cause photo-induced damage to one or more reaction components; their energy consumption, and

wear-and-tear on the source device. Further, as described elsewhere herein, rather then removing an illumination source, reaction components that are excited by the illumination source may be removed from the reaction mixture during the non-illuminated period, necessarily rendering them undetectable. For example, one or more fluorescently labeled nucleotide analogs may be replaced with unlabeled nucleotide analogs during the non-illuminated periods.

[0090] In certain aspects, the invention provides advantages to performing intramolecular redundant sequencing, in which a template nucleic acid is used to generate multiple copies of a sequence read of interest, whether by virtue of multiple copies of the complement being present in the template, repeated replication of the template, or a combination thereof. For example, a first stage of a template-dependent sequencing reaction on a single-stranded circular template can comprise a non-illuminated period during which the template is completely replicated at least one time to generate at least one incomplete sequence read for a sequence complementary to the template. The first stage is followed by a second stage comprising an illuminated period during which the template is replicated multiple times to generate multiple complete sequence reads for the complementary sequence. The incomplete reads generated in the first stage can be used to construct a scaffold for assembly of the complete sequence reads generated in the second stage. Further, incomplete sequence reads can also be used to clock the progress of the reaction during the non-illuminated periods by providing a count of the detectable reaction components and combining that information with known or estimated characteristics of the template, e.g., nucleotide composition or sequence.

[0091] The subset of signal emissions detectable in the non-illuminated periods as compared to the number detectable in the illuminated periods is not limiting and may be chosen based upon the non-illumination data desired by the ordinary practitioner and/or other considerations, such as mitigation of photo-induced damage to extend readlength. For example, to lower the likelihood of photo-induced damage, the ordinary practitioner may choose to remove the illumination source that is most damaging, e.g., has the highest frequency. In certain embodiments, multiple sequencing reactions may be performed for a single amplified template, each with a different combination of illumination sources and/or detectable components. Alternatively or additionally, multiple replicate reactions can also be performed for one or more of the combinations of illumination sources and/or detectable components. The combination of data from multiple different and/or replicate reactions performed on a single template provides myriad benefits during statistical analysis. As noted above, data can be combined to facilitate assembly of contigs generated during illuminated periods. Data

from non-illuminated periods can also provide value in assessing the quality of the sequence reads generated during the illuminated periods.

[0092] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods used to provide additional data during the course of the reaction can comprise using labels that are incorporated into an enzyme of the reaction. For example, FRET labels can be used to label portions of a polymerase enzyme such that the conformational change between the open and closed states of the enzyme change the FRET value. For example, a FRET-based system can be used to monitor the kinetics of opening and closing of the finger subdomain of DNA polymerase, as described in Allen, et al. (2008) Protein Science 17:401-408, incorporated herein by reference in its entirety for all purposes. In certain preferred embodiments, a closed conformation produces a FRET signal because the donor and acceptor are close to one another, and an open conformation silences the signal because there is no energy transferred between the donor and acceptor. By monitoring the emission from the FRET pair, each incorporation event can be monitored during non-detection periods, and optionally or additionally during detection periods. In certain preferred embodiments, the FRET donor is GFP (excitation at 484 nm; emission at 510 nm), and the FRET acceptor is YFP (excitation at 512 nm; emission at 529 nm). Methods for monitoring polymerase activity using FRET labels are known in the art, e.g., in WO/2007/070572 A2, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

[0093] A given reaction may experience one or a plurality of illuminated periods or non-illuminated periods, but preferably experiences at least two illuminated periods. For example, a given reaction providing nucleotide sequence information from a single template nucleic acid may have at least about 2, 3, 5, 10, 20, 50, or 100 illuminated periods with intervening non-illuminated periods. In an embodiment employing multiple periods of illumination and/or non-illuminated, the periods may be the same for both, e.g., 100 seconds "on" and 100 seconds "off." Alternatively, the illuminated periods may be longer or shorter than the non-illuminated periods. For example, in certain embodiments, a non-illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent illuminated period; or an illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent non-illuminated period. Further, each illuminated period may be the same or different from each other illuminated period, and each non-illuminated period may be the same or different from each other non-illuminated period. For

example, some embodiments generate a smaller number of long reads, and other embodiments generate a larger number of short reads. It will be understood that the number and length of the illuminated and non-illuminated periods is limited only by the experimental system in use and the data acquisition goals of the ordinary practitioner. In some embodiments, a nucleotide sequence read generated during a single illuminated period comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, a region of a nucleic acid template processed during a non-illuminated period during a single reaction comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, the set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises at least about 40, 60, 80, 100, 1000, 10,000, 25,000, 50,000, 100,000, 250,000, 500,000, or 1,000,000 nucleotide sequence positions from a single nucleic acid template. In some embodiments, a set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises multiple reads of at least a portion of the nucleotide sequence positions from a single nucleic acid template.

[0094] As noted above, the present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.). However, such information, if available, may also be useful to the ordinary practitioner in determining an optimal periodicity for illuminated and non-illuminated periods during a sequencing reaction, especially when sequencing repetitive sequences. For example, if a genomic region is known to contain five adjacent copies of a one kilobase nucleotide sequence (i.e., five "repeat regions"), it would be beneficial to keep the non-illuminated periods short enough to be able to confidently map the resulting sequence reads to the correct repeat region. If a non-illuminated period were too long, the natural variation in translocation rate of the polymerase would make it difficult to assign a sequence read to a particular repeat region, especially those farther from the binding/initiation site of the polymerase. In a further example, if the "copies" each had a few mutations that could be used to distinguish them from each other, it would be beneficial to keep the illuminated periods long enough to increase the chance one of these mutations would be included in a resulting sequence read, thereby allowing the unambiguous assignment of the read to a particular repeat region. If the illuminated period were too short the sequence reads from two different repeat regions could be identical, making mapping the sequence

read challenging. (Another way to mitigate these difficulties would be to incorporate pause or stop points into the template nucleic acid, as discussed below.)

[0095] Essentially, the practitioner may design the number of and lengths of time for each illuminated and non-illuminated period to best suit the illuminated reactions being analyzed and the invention is not limited in this regard. In certain embodiments, a practitioner may wish to increase the processivity of a polymerase thereby extending the length of the template nucleic acid processed in a sequencing reaction to be, e.g., at least 2-, 3-, 4-, 6-, 8-, 10-, or 20-fold, thereby generating sequence data much farther away from the polymerase binding/initiation site than would be achieved under constant illumination. In certain embodiments, a practitioner of the instant invention may wish to focus on data from one or more stages of an ongoing reaction, such as stages for which more data is required for analysis. In the case of sequencing-by-synthesis, one or more particular regions of a template nucleic acid may need to be resequenced. Some traditional methods require that new template nucleic acids be prepared to bring a region requiring resequencing closer to the initiation point of the sequencing reaction, or require preparation of multiple new templates if multiple regions to be resequenced. In contrast, the methods herein allow the practitioner to subject a template identical to the previously sequenced template (e.g., from a large genomic DNA sample preparation) to a sequencing reaction wherein illuminated periods are timed to illuminate the sample only when the polymerase is incorporating nucleotides into the nascent strand at the one or more particular regions requiring resequencing. This advantage substantially lowers the time and resources required for such resequencing operations, therefore providing a significant advantage over traditional methods.

[0096] The instant invention contemplates various means for providing non-illuminated periods during illuminated reactions. In some embodiments, the illumination source is turned off during the ongoing reaction to create one or more non-illuminated periods. In some embodiments, the illumination source remains on during the course of the reaction, but the illuminated reaction is removed from the system for a period of time. In some embodiments, the illumination source remains on during the course of the reaction, but the illumination is blocked to create one or more non-illuminated periods. For example, a movable mask may be manually or mechanically positioned between the illumination source and the illuminated reaction to block the illumination during non-illuminated periods and removed to allow exposure to the illumination during illuminated periods. Such a mask may also be dynamically controlled, such as a thin film transistor

display (e.g., an LCD mask). Masks for blocking illumination and manufacture thereof are well known to those of ordinary skill in the art and need no further elaboration herein.

[0097] One aspect of the present invention is multiplexing of large numbers of single-molecule analyses. For a number of approaches, e.g., single molecule methods as described above, it may be desirable to provide the reaction components in individually optically resolvable configurations, such that a single reaction component or complex can be individually monitored. Providing such individually resolvable configurations can be accomplished through a number of mechanisms. For example, by providing a dilute solution of complexes on a substrate surface suited for immobilization, one will be able to provide individually optically resolvable complexes. (See, e.g., European Patent No.1105529 to Balasubramanian, et al., the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Alternatively, one may provide a low density activated surface to which complexes are coupled. (See, e.g., Published International Patent Application No. WO 2007/041394, the full disclosure of which is incorporated herein by reference in its entirety for all purposes). Such individual complexes may be provided on planar substrates or otherwise incorporated into other structures, e.g., zero-mode waveguides or waveguide arrays, to facilitate their observation.

[0098] In some embodiments, a plurality of illuminated reactions are carried out simultaneously, e.g., on a solid support. In some preferred embodiments, a solid support comprises an array of reaction sites. In preferred embodiments, the reaction sites on a solid support are optically resolvable from each other. In further preferred embodiments, each of the reaction sites on a solid support contains no more than a single reaction to be interrogated. For example, in a sequencing-by-incorporation embodiment, each reaction site preferably has no more than one polymerase and no more than one nucleic acid template. The reaction sites may be confinements (e.g., optical and/or physical confinements), each with an effective observation volume that permits resolution of individual molecules present at a concentration that is higher than one nanomolar, or higher than 100 nanomolar, or on the order of micromolar range. In certain preferred embodiments, each of the individual confinements yields an effective observation volume that permits resolution of individual molecules present at a physiologically relevant concentration, *e.g.*, at a concentration higher than about 1 micromolar, or higher than 50 micromolar range or even higher than 100 micromolar. In addition, for purposes of discussion herein, whether a particular reagent is confined by virtue of structural barriers to its free movement, or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined."

[0099] As used herein, a solid support may comprise any of a variety of formats, from planar substrates, e.g., glass slides or planar surfaces within a larger structure, e.g., a multi-well plates such as 96 well, 384 well and 1536 well plates or regularly spaced micro- or nano-porous substrates, or such substrates may comprise more irregular porous materials, such as membranes, aerogels, fibrous mats, or the like, or they may comprise particulate substrates, e.g., beads, spheres, metal or semiconductor nanoparticles, or the like. The solid support may comprise an array of one or more zero-mode waveguides or other nanoscale optical structures.

[00100] As used herein, "zero-mode waveguide" refers to an optical guide in which the majority of incident radiation is attenuated, preferably more than 80%, more preferably more than 90%, even more preferably more than 99% of the incident radiation is attenuated. As such high level of attenuation, no significant propagating modes of electromagnetic radiation exist in the guide. Consequently, the rapid decay of incident electromagnetic radiation at the entrance of such guide provides an extremely small observation volume effective to detect single molecules, even when they are present at a concentration as high as in the micromolar range. The fabrication and application of ZMWs in biochemical analysis, and methods for calling bases in sequencing-by-incorporation methods are described, e.g., in U.S. Patent Nos. 7,315,019, 6,917,726, 7,013,054, 7,181,122, and 7,292,742, U.S. Patent Pub. No. 2003/0174992, and U.S. Patent Application No. 12/134,186, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00101] A set of reactions (e.g., contained on a solid support) may comprise identical or different components. For example, a single template nucleic acid may be analyzed in all reactions in the set, or a plurality of template nucleic acids may be analyzed, each present in only one or a subset of the set of reactions. In preferred embodiments, template nucleic acids comprising the same nucleotide sequence are analyzed in a plurality of reactions sufficient to provide adequate redundant nucleotide sequence data to determine a consensus sequence for the template nucleic acids. A number of sequence reads that will provide adequate nucleotide sequence data will vary, depending, e.g., on the quality of the template nucleic acid and other components of the reaction, but in general coverage for a template nucleic acid or portion(s) thereof is at least about 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage. Further, the numbers and lengths of illuminated and non-illuminated periods for a given reaction in the set of reactions may be the same or different than those for other reactions in the set. In some embodiments, a mixture of different periodicities are used for a set of reactions comprising the same template nucleic acid. This strategy can be

beneficial for providing nucleotide sequence reads from varying regions of the template sequence, thereby increasing the likelihood of overlapping sequence reads between individual reactions. These overlapping sequence reads can facilitate construction of a more robust sequence scaffold than could be constructed were the reactions all subjected to the same periodicity of illuminated and non-illuminated periods.

[00102] Methods of controlling polymerase progress and/or synchronizing polymerases in different reactions are also useful in analysis (e.g., mapping, validation, etc.) of nucleic acid reads farther from the initial binding site of the polymerase. During detection periods earlier in the reaction (i.e., closer to the time at which the polymerase began to process the template nucleic acid, such as during a first illuminated period), the position of a polymerase on the template can be estimated with generally good accuracy based on the known translocation rate of the polymerase under a given set of reaction conditions. As the duration of the reaction increases, however, the natural variation in polymerase translocation rate makes it more difficult to accurately determine the exact position of the polymerase on a template using estimation based on translocation rate alone; and through each subsequent illuminated period such estimations of polymerase position become less accurate, making subsequent analysis and mapping of the sequence reads to the template more difficult. Methods of regulating the position of the polymerase on the template allow more accurate determinations the polymerase's position. For example, causing the polymerase to pause or stop at a given location on the template during a non-illuminated period and reinitiating the polymerization during or immediately prior to a subsequent illuminated period provides a way to reorient the subsequently generated read with the template sequence, allowing easier consensus sequence determination and mapping analyses. Further, such pause/stop points can provide a means of controlling what regions of the template are processed during the illuminated periods by restricting where the polymerase will reinitiate on the template, thereby allowing a practioner of the instant invention to target one or more particular regions of a template for analysis during one or more detection periods during the course of an analytical reaction. Such methods are also useful to synchronize a set reactions being monitored simultaneously. For example, a plurality of reactions, each comprising a single polymerase/template complex, may be synchronized by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that show less spreading (i.e., less variation in the position on the template from which the sequence reads are generated) in the later stages of the reactions than would otherwise be observed without such regulation.

[00103] Various methods can be used to control or monitor the progress of a polymerase on a template nucleic acid. For example, as noted above, one may employ a reaction stop or pause point within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the single-stranded portion that was not used in priming. Reaction stop or pause points can be engineered into a portion of the template for which the nucleotide sequence is unknown (e.g., a genomic fragment), but is preferably located within a portion for which the nucleotide sequence is known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, certain preferred sequencing templates (e.g., SMRTbell™ templates, described elsewhere herein) are closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, and one or both of these linkers can comprise a stop or pause point to control the passage of the polymerase through them. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00104] By way of example, at a selected time following initiation of polymerization the reaction may be subjected to a non-illuminated period. The incorporation of a synthesis blocking moiety coupled to the template nucleic acid at a position encountered by the polymerase during the non-illuminated period will cause the polymerase to pause. An example of an engineered pause point is a known sequence on the template nucleic acid where a primer sits and blocks progression of a polymerase that is actively synthesizing a complementary strand. The presence of the primer by itself could introduce a pause in the polymerase sequencing or the primer could be chemically modified to force a full stop (and synchronization of multiple polymerases in multiple reactions). The chemical modification could be subsequently removed (for example, photo-chemically) and the polymerase would subsequently continue along the template nucleic acid. In some embodiments, multiple primers could be included in a reaction to introduce multiple pause or stop points along the template nucleic acid. Other methods for inducing a reversible pause (stop) in synthesis are known in the art and include, e.g., reversible sequestering of required cofactors (e.g., Mn^{2+}, one or more nucleotides, etc.). Once sufficient time has passed that the polymerase is paused at the blocking group, illumination is reintroduced and the blocking group removed. This allows control of the position on the template nucleic acid at which the polymerase will begin generating nucleotide

sequence data during the illuminated period. A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the template nucleic acid that inhibit polymerase mediated replication, strand-binding moieties that prevent processive synthesis, non-native nucleotides included within the primer and/or the template, and the like. Such reaction stops/pause points are useful in providing more certainty about the relationship of the reads to each other. For example, since the exact position on a template nucleic acid at which each sequence read begins would be known, the resulting reads could be better mapped relative to one another for construction of a sequence scaffold and/or consensus sequence. Further description of these and other methods for regulating the progress of a polymerase on a template are provided, e.g., in U.S.S.N. 61/099,696, U.S. Patent Pub. No. 2006/0160113, and U.S. Patent Pub. No. 2008/0009007, all of which are incorporated by reference herein in their entireties for all purposes.)

[00106] By way of example, a sequencing reaction may be initiated on a template comprising a non-native base in the absence of the complement to the non-native base, which would not impact the overall sequence determination of other portions of the template that are complementary to native bases. By starving the reaction for the complement to the non-native base, one can prohibit synthesis, and thus, the sequencing process, until the non-native base complement is added to the mixture. This can provide a "hot start" capability for the system and/or an internal check on the sequencing process and progress that is configurable to not interfere with sequence analysis of the regions of interest in the template, which would be complementary to only native bases. In some embodiments, the non-native base complement in the sequence mixture is provided with a detectably different label than the complements to the four native bases in the sequence, and the production of incorporation-based signals associated with such labels provides an indication that the polymerase has initiated or reinitiated. Although described as the "non-native base" it will be appreciated that this may comprise a set of non-natural bases that can provide multiple control elements within the template structure. In certain embodiments, two different non-native bases are included within the template structure, but at different points, to regulate procession of the sequencing process, e.g., allowing controlled initiation and a controlled stop/start position later in the sequence, e.g., prior to a subsequent illuminated period. For example, the complement to the first non-native base can be added to initiate sequencing immediately prior to the start of a first illuminated period. During a first non-illuminated period following the first illuminated period, the polymerase encounters the second non-native base, e.g., at a nucleotide position near but upstream of a nucleotide region desired to be sequenced in a second illuminated period. Sequencing would

stop until the complement to the second non-native base is added to the reaction mixture. Likewise, multiple such non-native bases could be incorporated into the template to effectively target the polymerase to multiple regions of interest for which sequence data is desired. Further, in applications in which multiple identical templates are being sequenced, this would allow a resynchronization of the various sequencing reactions and the data generated therefrom.

[00107] Methods of controlling polymerase progress in different stages of a sequencing reaction are also useful for not only creating "condition-dependent" non-detection periods (during which time illumination may or may not be present), but also for minimizing the amount of time required for traversing a given length of template during a non-detection period (whether or not illumination is present). In order to reliably detect incorporation events, non-natural reagent conditions are typically used to limit polymerization during detection periods to approximately 1-5, or about 3 bases per second. In certain embodiments, replacement of Mg^{2+} ions with Mn^{2+} ions serves to stabilize and slow the translocation of the polymerase. When magnesium and, optionally, native nucleotides (e.g., lacking fluorescent labels) are used, the rate of translocation and/or processivity of the polymerase may increase up to two orders of magnitude. Use of such "rapid translocation" conditions during the non-detection periods can provide myriad benefits, including but not limited to a more rapid polymerization rate, an increased processivity (e.g., due to decreased stalling and misincorporation), and an overall savings due to reduced use of expensive labeled nucleotide analogs and/or reagents that mitigate oxidative stress.

[00108] In certain embodiments, a protocol for intermittent detection comprises alternating reaction mixtures, where a first reaction mixture used during the detection periods is optimized for sequence read generation, and a second reaction mixture used during the non-detection periods is optimized for processivity and/or rapid polymerization. For example, when reagents for optimal sequence read generation are present, DNA synthesis rate is low, and there is a fluorescence signal associated with each incorporation event. After replacing the reaction mixture optimized for sequence read generation with the reaction mixture optimized for processivity and/or rapid polymerization, the polymerase rapidly advances across the template. In certain embodiments, a flow cell is used to deliver and switch between the two (or more) reaction mixtures during the course of the reaction.

[00109] In an exemplary embodiment, a first reaction mixture comprises fluorescently-labeled nucleotide analogs and manganese ions that restrict polymerization to a rate appropriate for high fidelity detection of nucleotide incorporation. The first reaction mixture can also include

additional agents for mitigation of photo-induced damage of various components of the reaction mixture. A second reaction mixture comprises natural nucleotides and an appropriate magnesium ion concentration for rapid synthesis of the nascent strand complementary to the template. A first detection period of a sequencing reaction is initiated by introduction of the first reaction mixture, and a sequence read is generated based upon synthesis of the nascent strand during the detection period. After a predetermined time interval a sufficient quantity of the second reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, thereby initiating a first non-detection period. As noted above, the lack of labeled nucleotides in the second reaction mixture alone can produce the non-detection period, since there will be no signal emitted coincident with incorporation of the native nucleotides, but in certain embodiments illumination may also be removed, e.g., to further mitigate photo-induced damage during the non-detection period. At a time appropriate to initiate a second detection period, a sufficient quantity of the first reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, and detection of incorporation event is reinitiated. The cycle of reaction mixture exchange is repeated to generate multiple detection and non-detection periods.

[00110] A flow cell for reaction mixture exchange preferably has two inputs that are gated such that only a single reaction mixture flows into a reaction site or plurality of reaction sites, e.g., on a substrate. A single out-flow line may be used to remove reaction mixtures from the reaction site(s) to a single collection vessel, or multiple collection vessels may be used, one for each type of reaction mixture used. Further, accurate estimation of the distance a polymerase translocates during a non-detection period is important for bioinformatics applications. This estimation is complicated if the time for reaction mixture exchange is slow. As such, the flow is preferably at a sufficient rate that the time for exchange is significantly less than the time spent in the presence of either reaction mixture alone.

[00111] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination. In this embodiment, sixteen sequencing-by-incorporation reactions are performed on single nucleic acid templates (each of which comprises the same nucleotide sequence) with the timing of the illuminated and non-illuminated periods the same for all sixteen reactions. In A, the sixteen reactions are shown disposed on sixteen reaction sites on a solid support and are numbered for convenience. A representation of the illumination data is shown in B, with bars extending across the graph indicative of illumination data collected during illuminated

periods for each reaction. In this illustrative example, each reaction is subjected to three illuminated periods, each followed by a non-illuminated period, resulting in three noncontiguous sequence reads for each reaction, i.e., three noncontiguous reads per template molecule sequenced. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. During the first illuminated period, the sequence reads generally overlap, but the natural variation of polymerase translocation rate over the set of reactions results in a "spreading" of the sequence reads as the reaction proceeds through the second and third illuminated periods with increasing variation in the exact position of each polymerase on the template at the beginning and end of each illuminated period. As such, the earlier illumination data provides better redundancy ("oversampling") of sequence information over a relatively narrow portion of the template nucleic acid, while the later illuminated periods provide less redundant sequencing data over a broader region of the template nucleic acid. The timing of the non-illuminated periods between the illuminated periods and the known or calculated rate of incorporation are used to determine approximate spacing between the resulting sequence reads, providing context for building a sequence scaffold or consensus sequence. It is important to note that although shown disposed on a solid support in A, the data shown in B could also have been generated from reactions not disposed on a solid support nor performed simultaneously and the methods are generally not so limited. Further, as described above, the spreading of the sequence reads from later stages of the reactions can be mitigated by synchronizing the reactions, e.g., by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that provides better redundancy (i.e., more overlap in the positions on the template from which the sequence reads are generated), especially in the later stages of the reactions.

[00112] Using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be reset to change the regions of the template for which nucleotide sequence data is generated. As described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation

during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest. Alternatively, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be kept the same to provide a greater-fold coverage of one or more regions of interest in the template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which are incorporated by reference herein in their entireties for all purposes. A specific embodiment is also provided in the Exemplary Applications section herein.

[00113] The present invention provides novel template configurations and methods for exploiting these compositions in template directed sequencing processes. While these compositions and methods have utility across all of the various template directed processes described herein, for ease of discussion, they are being primarily discussed in terms of preferred single molecule, real-time sequencing processes, in which they provide myriad benefits. In particular, the present invention is generally directed to nucleic acid sequences that employ improved template sequences to improve the accuracy of sequencing processes. For example, in at least one aspect, the template compositions of the invention are generally characterized by the presence of a double stranded segment or a pair of sub-segments that are internally complementary, i.e., complementary to each other. In particular contexts, the target nucleic acid segment that is included within a template construct will typically be substantially comprised of a double stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double stranded or otherwise internally complementary.

[00114] Examples of template configurations of the invention that are partially and completely contiguous are schematically illustrated in Figure 20A and 20B, respectively. In particular, as shown in Figure 20A, a partially contiguous template sequence 200 is shown which includes a double stranded portion, comprised of two complementary segments 202 and 204, which,

for example, represent a target sequence or portion thereof. As shown, the 3' end of segment 202 is linked to the 5' end of segment 204 by linking oligonucleotide 206, providing a single stranded portion of the template, and yielding a partially contiguous sequence. By comparison, as shown in Figure 20B, a completely contiguous template sequence 210 is shown. Sequence 210 includes a double stranded portion again comprised of two complementary segments 212 and 214. As with the partially contiguous sequence of Figure 20A, the 3' end of segment 212 is joined to the 5' end of segment 214 via oligonucleotide 216 in a first single stranded portion. In addition, the 5' end of segment 212 is joined to the 3' end of segment 214 via linking oligonucleotide 218, providing a second single stranded portion, and yielding a completely contiguous or circular template sequence.

[00115] In addition, the templates of the invention, by virtue of their inclusion of double stranded segments, provide consensus through the identification of both the sense and antisense strand of such sequences (in both the partially and completely contiguous configurations).

[00116] By way of example, and with reference to Figures 20A and 20B, with respect to a partially contiguous template shown in Figure 20A, obtaining the entire sequence, e.g., that of segments 202, 204 and 206 provides a measure of consensus by virtue of having sequenced both sense strand, e.g., segment 202, and the antisense strand, e.g., segment 204. In addition to providing sense and antisense consensus within a single template molecule that can be sequenced in one integrated process, the presence of linking segment 206 also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., 202, is completed and the other begins, e.g., 204. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. The progress of sequencing processes is schematically illustrated in Figure 21A. In particular, as shown, a sequencing process that begins, e.g., is primed, at the open end of the partially contiguous template, proceeds along the first or sense strand, providing the nucleotide sequence (A) of that strand, as represented in the schematic sequence readout provided. The process then proceeds around the linking oligonucleotide of the template, providing the nucleotide sequence (B) of that segment. The process then continues along the antisense strand to the A sequence, and provides the nucleotide sequence (A'), which provides consensus data for the sense strand as its antisense counterpart. As noted, because the B sequence may be exogenously provided, and thus known, it may also provide a registration sequence indicating a point in the sequence determination at which the data transitions from sense to antisense strands.

[00117] With respect to completely contiguous or circular template sequences configured in accordance with the invention, the consensus potential is further increased. In particular, as with the partially contiguous sequences shown in Figure 20A, the completely contiguous sequences also provide sense and antisense consensus. In addition, such templates provide for the potential for iterative sequencing of the same molecule multiple times, by virtue of the circular configuration of the template. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining consensus for each segment from the complementary sequences, as well as consensus within each segment, by repeatedly sequencing that segment. This is schematically illustrated in Figure 21B, again with a representative illustration of a sequence readout provided. As shown, a sequencing process that is primed at one end, e.g., primed within one linking oligonucleotide sequence, e.g., linking oligonucleotide 218 of Figure 20, proceeds along the first or sense strand 214, again providing the nucleotide sequence A of that strand. The sequence process then proceeds around the first linking oligonucleotide, e.g., linking oligonucleotide 216 from Figure 20, to provide the nucleotide sequence B of that segment of the template. Proceeding along the antisense strand, e.g., segment 212 of Figure 20B), provides the nucleotide sequence A', which is again, complementary to sequence A. The sequencing process then continues around the template providing the nucleotide sequence for the other linking oligonucleotide, e.g., linking oligonucleotide 218 of Figure 20B, where the illustrated sequencing process began, providing nucleotide sequence C. Because the template is circular, this process can continue to provided multiple repeated sequence reads from the one template, e.g., shown as providing a second round of the sequence data A-B-A'-C-A-B-A'. Thus, sequence redundancy comes from both the determination of complementary sequences A and A', and the repeated sequencing of each segment. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing. Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post synthesis.

[00118] Another exemplary embodiment of an analysis of a plurality of illuminated reactions using intermittent illumination comprises a first illuminated period that is initiated at different times over the plurality of reactions. For example, the illuminated period for a first reaction may start at 0 seconds, the illuminated period for a second reaction may start at 5 seconds, the illuminated period for a third reaction may start at 10 seconds, and so forth. Additionally or alternatively, a first subset

of reactions may begin at a first time, a second subset may begin at a second time, etc. The first illuminated period continues for a given length of time, followed by a non-illuminated period and a subsequent second illuminated period. Optionally, a plurality of non-illuminated periods and illuminated periods follow the first illuminated period. Staggered start times can provide staggered data sets (e.g., two or more sequence reads) for the plurality of reactions, allowing multiple different stages of the overall reaction to be interrogated in different reactions. Preferably, the staggered data sets overlap to an extent that allows further analysis and validation of the reaction data. For example, a sequencing-by-incorporation reaction subjected to such an embodiment of the invention would preferably have sufficient overlap between sequence reads from different individual reactions to allow construction of a sequence scaffold and/or consensus sequence for a template nucleic acid.

[00119] A mask for use with a solid support (e.g., an array of confinements) can be designed to allow illumination of one or more portions of the solid support while blocking illumination to other portions of the solid support. For example, a mask may comprise one or more windows that allow excitation illumination to pass through the mask. Such a mask may be physically moved over the surface of the solid support (or the solid support can be moved relative to the mask), e.g., to selectively allow excitation illumination to reach a subset of confinements in an array. For example, a mask that allows 10% of reaction sites to be illuminated could be used to increase the sequencing scaffold coverage by sliding the illumination area (the area being subjected to excitation illumination) back and forth across the solid support. The 10% of reactions would cover certain regions of the nucleic acid template for any given time period (and therefore region of sequence in the template). In certain embodiments, an automated mask that selectively controls the timing of illumination of reactions on a solid support during the course of the reaction/acquisition may be used rather than a mask that must be physically moved.

[00120] The timing of the illuminated and non-illuminated periods for a set of reactions on a solid support may be the same or may vary, and may be synchronized or random. In certain embodiments in which the excitation illumination source is turned on and off, the timing of the illuminated and non-illuminated periods for the set of reactions will be identical. In other embodiments, for example, those that comprise use of a mask, the timing of the illuminated and non-illuminated periods for the set of reactions can vary so that while a subset of the reactions are illuminated, another subset of the reactions are not illuminated. Various exemplary and nonlimiting embodiments of masks that may be used with a set of reactions on a solid substrate are provided in Figures 3-5, as described below. In certain embodiments, the illuminated/non-illuminated status of

each reaction may be random across the solid support, e.g., to remove any experimental bias potentially introduced by actively selecting which reactions to illuminate at a given time, as long as the sequence reads being generated at the illuminated reactions and the time at which these reactions are not illuminated are able to be assigned to a particular reaction. For ease of discussion, the action of both illuminating and collecting emission signals from a reaction of interest, or a particular region on a solid support in which a reaction of interest is taking place, is referred to as "interrogating" that reaction and/or that region. A region being so interrogated is termed an "observation region."

[00121] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination and a mask. As in Figure 2, an array of reactions on a solid support 310 is provided containing sixteen reaction sites, numbered for convenience (A). In B, a mask 320 is provided with a single window 330 to allow passage of illumination to a subset of reactions on the solid support. Window 330 is wide enough to allow illumination of at least two columns of reaction sites on solid support 310. As in Figure 2, a representation of the illumination data is shown in C, with bars extending across the graph indicative of illumination data collected for each reaction. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. When the sequencing reaction is initiated at all positions on solid support 310, the window 330 is positioned to allow illumination to only reactions 1, 5, 9, and 13, and these four reactions provide sequence reads 350 for the earliest stage of the reactions. The window 330 is subsequently moved to provide an illuminated period for reactions 2, 6, 10, and 14 while still continuing the illuminated period for reactions 1, 5, 9, and 13. The illumination data for reactions 2, 6, 10, and 14 provides sequence reads 360, which partially overlap sequence reads 350 for reactions 1, 5, 9, and 13. The window 330 is moved again to provide illuminated periods for reactions 3, 7, 11, and 15 while still continuing the illuminated period for reactions 2, 6, 10, and 14, but removing illumination from reactions 1, 5, 9, and 13. The illumination data for 3, 7, 11, and 15 results in sequence reads 370, which partially overlap sequence reads 360 for reactions 2, 6, 10, and 14. A fourth position of the mask 320 initiates an illuminated period for reactions 4, 8, 12, and 16 while continuing illumination of reactions 3, 7, 11, and 15, but ending the illuminated period for reactions 2, 6, 10, and 14. Sequence reads 380 correspond to sequence reads from reactions 4, 8, 12, and 16. Finally, the window is moved to end the illuminated period for reactions 3, 7, 11, and 15 while continuing the illuminated period for reactions 4, 8, 12, and 16. Repeating the above process allows

a second read to be generated from each reaction, and this second read is noncontiguous with the first read. For example, reactions 1, 5, 9, and 13 correspond to reads 350 and, later in the reaction, reads 355. The two reads generated in a single reaction do not overlap and are separated by a length of nucleotides that was incorporated during the non-illuminated period between the two illuminated periods.

[00122] The mask can optionally be passed over the substrate additional times to generate additional reads until the reactions are complete or no longer provide reliable data, such as when the total illumination time (computed by summing the times for the multiple illuminated periods) has surpassed a photo-induced damage threshold period. Further, the mask may be passed back and forth, or may pass over the solid support in only one direction, e.g., always left to right, or vice versa.

[00123] Further, unlike the data shown in Figure 2B which has gaps in the sequence coverage for the template nucleic acid, the strategy provided in this embodiment results in at least two-fold coverage across the entire template nucleic acid (Figure 3C), although at a lower-fold redundancy. The portion of the template covered by only reads 380 and reads 355 has the least-fold redundancy, and in some instances a gap in coverage may be present in this region due to the movement of the mask 320 from the far right to the far left of the solid support 310. Of course, oversampling by adding replicate reactions to the set of reactions, or using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the coverage of a nucleic acid template, thereby providing more data for construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which have been previously incorporated by reference herein. The natural variation of polymerase translocation rate over the set of reactions is also apparent in this prophetic example as the spreading of the sequence reads and decreasing overlap between reads from reactions in adjacent columns in the later stages of the reactions as compared to the earlier stages.

[00124] Figure 4A provides an embodiment of a mask similar to that provided in Figure 3 except that it comprises three windows allowing multiple nonadjacent columns of reaction sites to be illuminated simultaneously. Figure 4B provides an embodiment of a mask comprising twelve windows, each of which allows illumination of a single reaction site on a solid support. The

windows are oriented in the mask to allow illumination of every other reaction in each row and every other reaction in each column. It will be understood that these mask designs are merely exemplary and nonlimiting embodiments as it is well within the abilities of the ordinary practitioner to determine an appropriate mask design depending on the experimental design or the illuminated reactions to be interrogated.

[00125] Figure 5B illustrates yet another aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Four different samples are disposed on a solid support, one in each quadrant 510, 520, 530, and 540 (A). A mask 550 is used that comprises two windows 560 that allow multiple rows of reaction sites to be illuminated simultaneously (B). A first position of this mask over a solid support in which two reactions in each quadrant are illuminated is demonstrated in C. A second position of the mask allowing illumination of the previously non-illuminated reactions is demonstrated in D. The mask may be moved back and forth as indicated by the double-arrow to provide multiple illuminated and non-illuminated periods for each reaction containing one of the four samples.

[00126] The present invention is also useful for redundant interrogation of reactions or portions of a solid support of interest. In certain aspects, sequential interrogation of different observation regions may be repeated a number of times, e.g., more than 2, 5, 10, 50, 100, 500, 1000, or even more than 10,000 times. In general, this method of stepping the observation region to another, preferably adjacent region, and repeating the interrogation process is generally referred to as a "step and repeat" process, and may be performed by various methods, including but not limited to moving the incident light and the solid support relative to one another and moving a mask across the surface of the solid support, as described above. Although described as a "step and repeat" method, in some embodiments where the observation region is moved across a substrate, that movement is not step-wise and iterative, but instead constitutes a continuous motion, substantially continuous motion, or stepped movement, or an iterative motion whereby each iterative step interrogates a new region that overlaps with some portion of the previously interrogated region. In particular, a substrate may be moved continuously relative to an optical system, whereby the observation region moves continuously across the substrate being interrogated (in a "scan mode").

[00127] The present invention is optionally combined with an optical system that provides illumination and/or collection of emitted illumination. Preferably, the optical system is operatively coupled to the reaction sites, e.g., on a solid support. One example of a particularly preferred optical

system is described in U.S.SN. 11/201,768, filed August 11, 2005, and incorporated herein by reference in its entirety for all purposes. Optical systems are described further below.

[00128] In some embodiments, one or both of the solid support and optical system are moved during interrogation. For example, a solid support being interrogated may be held stationary while the optical system is moved, or the solid support may be moved relative to a stationary optical system. Such movement may be accomplished using any of a variety of manipulation hardware or robotic set-ups, e.g., a stepper/feeder apparatus, and are well known in high performance printing technologies and in the semiconductor industry. For example, robotic systems may be used to pick up and re-orient a given solid support in order to interrogate different regions of the solid support, or make a previously unaccessible region (e.g., blocked by clips, support structure, or the like) of the solid support accessible. Such robotic systems are generally available from, e.g., Beckman, Inc., Tecan, Inc., Caliper Life Sciences, and the like.

[00129] In addition to the foregoing, it will be appreciated that the reagents in a given reaction of interest, including those reagents for which photo-induced damage is being mitigated in accordance with the invention, may be provided in any of a variety of different configurations. For example, they may be provided free in solution, or complexed with other materials, e.g., other reagents and/or solid supports. Likewise, such reagents may be provided coupled to beads, particles, nanocrystals or other nanoparticles, or they may be tethered to larger solid supports, such as matrices or planar surfaces. These reagents may be further coupled or complexed together with other reagents, or as separate reagent populations or even as individual molecules, e.g., that are detectably resolvable from other molecules within the reaction space. As noted above, whether a particular reagent is confined by virtue of structural barriers to its free movement or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined." Further examples of such confined reagents include surface immobilized or localized reagents, e.g., surface immobilized or associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflection microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. For example, in some preferred embodiments, one or more reagents in an assay system are confined within an optical confinement. Such an optical confinement may be an internal reflection confinement (IRC) or an external reflection confinement

(ERC), a zero-mode waveguide, or an alternative optical structure, such as one comprising porous film with reflective index media or a confinement using index matching solids. More detailed descriptions of various types of optical confinements are provided, e.g., in International Application Publication No. WO/2006/083751, incorporated herein by reference in its entirety for all purposes.

[00130] The invention is generally applicable to any of a variety of optical assays that require substantial illumination and/or photoactivated conversion or excitation of chemical groups, e.g., fluorophores. For example, the compositions and methods provided herein may be used with fluorescence microscopy, optical traps and tweezers, spectrophotometry, fluorescence correlation spectroscopy, confocal microscopy, near-field optical methods, fluorescence resonance energy transfer (FRET), structured illumination microscopy, total internal reflection fluorescence microscopy (TIRF), etc. The methods provided herein may be particularly useful in assays that are negatively impacted, directly or indirectly, by prolonged exposure to illumination. Of particular interest are those assays that are impaired by the generation and/or accumulation of triplet-state forms or free radicals during illumination.

[00131] One particularly apt example of analyses that benefit from the invention are single-molecule biological analyses, including, inter alia, single molecule nucleic acid sequencing analyses, single molecule enzyme analyses, hybridization assays (e.g., antibody assays), nucleic acid hybridization assays, and the like, where the reagents of primary import are subjected to prolonged illumination with relatively concentrated light sources (e.g., lasers and other concentrated light sources, such as mercury, xenon, halogen, or other lamps) in an environment where photoconversion/excitation is occurring with its associated generation of products. In certain embodiments, the methods, compositions, and systems are used in nucleic acid sequencing processes that rely on detection of fluorescent or fluorogenic reagents. Examples of such sequencing technologies include, for example, SMRT™ nucleic acid sequencing (described in, e.g., U.S. Patent Nos. 6,399,335, 6,056,661, 7,052,847, 7,033,764, 7,056,676, 7,361,466, 7,416,844, the full disclosures of which are incorporated herein by reference in their entirety for all purposes), non-real-time, or "one base at a time" sequencing methods available from, e.g., Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), Clonal Single Molecule Array™, and SOLiD™ sequencing. (See, e.g., Harris, et al. (2008) Science 320 (5872):106-9, incorporated by reference herein in its entirety for all purposes.) Such prolonged illumination can negatively impact (e.g., by introducing photo-induced damage) these reagents and diminish their effectiveness in the desired reaction.

III. Prevention of Photo-induced Damage

[00132] The methods provided herein are particularly useful in analyses that utilize very limited concentrations of reactants, such as single molecule detection/monitoring assays. As will be appreciated, in such reagent limited analyses, any loss, degradation, or depletion of a critical reagent will dramatically impact the analysis by further limiting the reagent, which not only can adversely effect the detectable signal, but may also directly impact the reaction being monitored, e.g., by changing its rate, duration, or product(s). For example, photo-induced damage can include a photoinduced change in a given reagent that reduces the reactivity of that reagent in the reaction, e.g., photobleaching of a fluorescent molecule, which diminishes or removes its ability to act as a signaling molecule. Also included in the term photo-induced damage are other changes that reduce a reactant's usefulness in a reaction, e.g., by making the reagent less specific in its activity in the reaction. Likewise, photo-induced damage includes undesired changes in a reagent that are caused by interaction of that reagent with a product of another photoinduced reaction, e.g., the generation of singlet oxygen during a fluorescence excitation event, which singlet oxygen may damage organic or other reagents, e.g., proteins. Photo-induced damage also includes downstream effects of damage to reactants, such as irreversible interactions between damaged reactants and other critical components of the reaction, e.g., reactive proteins or enzymes. For example, damage to an enzyme that catalyzes a reaction being monitored may cause a reduction in the rate of the reaction, in some cases stopping it altogether, or may reduce the duration or fidelity of the reaction.

[00133] As suggested by the foregoing, photo-induced damage generally refers to an alteration in a given reagent, reactant, or the like, that causes such reagent to have altered functionality in a desired reaction, e.g., reduced activity, reduced specificity, or a reduced ability to be acted upon, converted, or modified, by another molecule, that results from, either directly or indirectly, a photo-induced reaction, e.g., a photo-induced reaction creates a reactant that interacts with and causes damage to one or more other reactants. Typically, such photoreaction directly impacts either the reactant of interest, e.g., direct photo-induced damage, or impacts a reactant within one, two or three reactive steps of such reactant of interest. Further, such photoreaction can directly impact the reaction of interest, e.g., causing a change in rate, duration, processivity, or fidelity of the reaction.

[00134] The amount of time an illuminated analysis may be carried out before photo-induced damage so substantially impacts the reactants to render the analysis non-useful is referred to as the

"photo-induced damage threshold period." A photo-induced damage threshold period is assay-dependent, and is affected by various factors, including but not limited to characteristics of enzymes in the assay (e.g., susceptibility to photo-induced damage and the effect of such damage on enzyme activity/processivity), characteristics of the radiation source (e.g., wavelength, intensity), characteristics of the signal-generating molecule (e.g., type of emission, susceptibility to photo-induced damage, propensity to enter triplet state, and the effect of such damage on the brightness/duration of the signal), similar characteristics of other components of the assay. It can also depend on various components of the assay system, e.g., signal transmission and detection, data collection and analysis procedures, etc. It is well within the abilities of the ordinary practitioner to determine an acceptable photo-induced damage threshold period for a given assay, e.g., by monitoring the signal decay for the assay in the presence of a photodamaging agent and identifying a period for which the signal is a reliable measure for the assay. In terms of the invention, the photo-induced damage threshold period is that period of illuminated analysis during which such photo-induced damage occurs so as to reduce the rate or processivity of the subject reaction by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over the same reaction in the absence of such illumination. It is an object of the invention to increase the photo-induced damage threshold period, thereby increasing the amount of time reactions can proceed toward completion with minimal damage to the reactants, thereby lengthening the time in which the detectable signal is an accurate measure of reaction progression.

[00135] In some contexts, a "photo-induced damaged" reaction may be subject to spurious activity, and thus be more active than desired. In such cases, it will be appreciated that the photo-induced damage threshold period of interest would be characterized by that period of illuminated analysis during which such spurious activity, e.g., as measured by an increase in reaction rate, or an increase in non-specific reaction rate, is no more than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over a non-illuminated reaction. In one non-limiting example, where a nucleic acid polymerase, by virtue of a photodamaging event, begins to incorrectly incorporate nucleotides during template directed synthesis, such activity would impact the photo-induced damage threshold period as set forth above. In this case, the methods, devices, and systems of the invention would increase the photo-induced damage threshold period, thus increasing the amount of time the reaction could proceed before the above-described spurious activity occurred.

[00136] With reference to nucleic acid analyses, it has been observed that in template-directed synthesis of nucleic acids using fluorescent nucleotide analogs as a substrate, prolonged

illumination can result in a substantial degradation in the ability of the polymerase to synthesize the nascent strand of DNA, as described previously, e.g., in U.S. Published Patent Application No. 20070161017, incorporated by reference herein in its entirety for all purposes. Damage to polymerase enzymes, template sequences, and/or primer sequences can significantly hinder the ability of the polymerase to process longer strands of nucleic acids. For example, reduction in the processivity of a polymerase leads to a reduction in read lengths for sequencing processes that identify sequence constituents based upon their incorporation into the nascent strand. As is appreciated in the art of genetic analysis, the length of contiguous reads of sequence directly impacts the ability to assemble genomic information from segments of genomic DNA. Such a reduction in the activity of an enzyme can have significant effects on many different kinds of reactions in addition to sequencing reactions, such as ligations, cleavages, digestions, phosphorylations, etc.

[00137] Without being bound to a particular theory or mechanism of operation, it is believed that at least one cause of photo-induced damage to enzyme activity, particularly in the presence of fluorescent reagents, results from the direct interaction of the enzyme with photo-induced damaged fluorescent reagents. Further, it is believed that this photo-induced damage of the fluorescent reagents (and possibly additional damage to the enzyme) is at least partially mediated by reactive intermediates (e.g., reactive oxygen species) that are generated during the relaxation of triplet-state fluorophores. One or both of the photo-induced damaged fluorescent reagents and/or reactive intermediates may be included in the overall detrimental effects of photo-induced damage.

[00138] In certain aspects, the invention is directed to methods, devices, and systems that reduce the amount of photo-induced damage to one or more reactants during an illuminated reaction, e.g., thereby improving the reaction, e.g., by increasing the processivity, rate, fidelity, processivity, or duration of the reaction. In particular, methods are provided that yield a reduction in the level of photo-induced damage and/or an increase in the photo-induced damage threshold period as compared to such reactions in the absence of such methods, devices, and systems. In particular embodiments, such methods comprise subjecting an illuminated reaction to periods of non-illuminated during the course of the reaction, as described above, or by temporarily removing components of the reaction mixture that are believed to cause such damage, as described below.

[00139] As generally referred to herein, limited quantity reagents or reactants may be present in solution, but at very limited concentrations, e.g., less than 200 nM, in some cases less than 10 nM and in still other cases, less than 10 pM. In preferred aspects, however, such limited quantity

reagents or reactants refer to reactants that are immobilized or otherwise confined within a given area or reaction site (e.g., a zero-mode waveguide), so as to provide limited quantity of reagents in that given area, and in certain cases, provide small numbers of molecules of such reagents within that given area, e.g., from 1 to 1000 individual molecules, preferably between 1 and 10 molecules. As will be appreciated, photo-induced damage of immobilized reactants in a given area will have a substantial impact on the reactivity of that area, as other, non-damaged reactants are not free to diffuse into and mask the effects of such damage. Examples of immobilized reactants include surface-immobilized or -localized reagents, e.g., surface-immobilized or -associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflectance microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. Various types of solid supports upon which one or more reactants can be immobilized are described above.

[00140] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is interrogated for one or more illuminated periods that cumulatively are less than a photo-induced damage threshold period, as set forth elsewhere herein. Such interrogation may occur coincident with or independent of interrogation of additional observation regions on a solid support containing the first observation region. In accordance with the present invention, the observation region typically includes confined reagents (e.g., enzymes, substrates, etc.) that are susceptible to photo-induced damage, and may include an area of a planar or other solid support upon which confined reagents are immobilized. Alternatively or additionally, the observation region may include a physical confinement that constrains the reagents that are susceptible to photo-induced damage, including, e.g., microwells, nanowells, planar surfaces that include hydrophobic barriers to confine reagents.

[00141] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is intermittently interrogated under constant illumination by virtue of intermittent presence of detectable components of the reaction, wherein the presence of such detectable components has the potential to directly or indirectly cause photo-induced damage to one or more other reaction components. For example, a buffer comprising detectable components of a reaction can be temporarily replaced with a buffer comprising non-detectable versions of the same

components of the reaction, thereby interrupting data acquisition for the reaction. When data acquisition is to be recommenced, the buffer comprising detectable component is substituted for the buffer comprising non-detectable components. This substitution of reaction components may be repeated multiple times to generate multiple sets of data collected at noncontiguous stages of the reaction. For example, such a substitution can occur at least about 2, 4, 6, 8, or 10 times during the course of the reaction.

[00142] In certain preferred embodiments, the detectable components are fluorescently-labeled components that can be damaged by exposure to excitation illumination, and can further cause damage to other reaction components, as described above. For example, a sequencing-by-incorporation reaction can be initiated in the presence of fluorescently-labeled nucleotides whose incorporation is indicative of the nucleotide sequence of the nascent strand synthesized by a polymerase, and by complementarity, of the template nucleic acid molecule. At a selected time point during the ongoing reaction, the labeled nucleotides can be removed and replaced with unlabeled nucleotides, for example, by buffer exchange. After a period of time during which data acquisition has been interrupted by the absence of signal from the ongoing reaction, the labeled nucleotides can be reintroduced to reinitiate data acquisition. The labeled nucleotides may be removed and reintroduced multiple times and for various lengths of time, as preferred by the ordinary practitioner. In this way, multiple noncontiguous sequence reads can be generated from a single nucleic acid molecule in real time.

[00143] The methods herein slow the accumulation of photo-induced damage to one or more reagents, and may therefore indirectly mitigate the impact of photo-induced damage in an ongoing reaction of interest. By way of example, methods that reduce exposure of a critical enzyme component to illumination radiation (e.g., by subjecting the reaction to periods of non-illumination or by temporarily removing a component of the reaction responsible for such damage) do not necessarily prevent the photo-induced damage to the enzyme component, but rather extend the photo-induced damage threshold period by slowing the accumulation of photo-induced damage in the reaction mixture. Measurements of reduction of photo-induced damage as a result of implementation of intermittent illumination may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction subjected to constant illumination. Likewise, measurements of reduction of photo-induced damage as a result of temporary removal of reaction components responsible for such damage may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction in which such components are present

throughout. Further, characterization of a reduction in photo-induced damage generally utilizes a comparison of reaction rates, durations, or fidelities, processivities, e.g., of enzyme activity, and/or a comparison of the photo-induced damage threshold period, between a reaction mixture subjected to such the methods and/or systems of the invention and a reaction mixture not so subjected.

[00144] In the case of the present invention, implementation of the methods, devices, and systems of the invention generally results in a reduction of photo-induced damage of one or more reactants in a given reaction, as measured in terms of "prevented loss of reactivity" in the system. Using methods known in the art, the amount of prevented loss of activity can at least 10%, preferably greater than 20%, 30%, or 40%, and more preferably at least 50% reduction in loss of reactivity or increase in processivity, and in many cases greater than a 90% and up to and greater than 99% reduction in loss of reactivity or increase in processivity. By way of illustration, and purely for the purpose of example, when referring to reduction in photo-induced damage as a measure of enzyme activity in the presence and absence of intermittent illumination, if a reaction included a reaction mixture having 100 units of enzyme activity that would, under constant illumination, yield a reaction mixture having only 50 units of activity, then a 10% reduction in photo-induced damage would yield a final reaction mixture of 55 units (e.g., 10% of the 50 units otherwise lost, would no longer be lost). Further, use of the invention is expected to increase the performance (e.g., processivity, duration, fidelity, rate, etc.) of a reaction whose performance is negatively impacted by constant exposure to illumination by at least about 2-, 5-, 10-, 20-, 30-, 50-, 80-, 100-, 500-, or 1000-fold over that achieved by the reaction under constant illumination. For example, it is a specific object of the instant invention to increase the processivity of a polymerase enzyme in a sequencing reaction to allow collection of data across a longer length of the template.

[00145] With regards to sequencing applications, the methods herein facilitate the scaffolding of nucleic acid sequences in reactions susceptible to photo-induced damage. For example, if the sequencing device has 1000 base pair average readlength under constant illumination, one could subject the reaction to illuminated periods timed to allow approximately 100 nucleotides to be incorporated into the nascent strand of read, followed by non-illuminated periods timed to allow approximately 1000 nucleotides to be incorporated "in the dark." The sequence reads resulting from this experimental design would comprise about ten sequence reads of about 100 nucleotides each separated by gaps of about 1000 nucleotides each. If a plurality of sequencing reactions were carried out in this manner, and the illuminated periods were staggered appropriately, the reads from the plurality of reactions could be combined to provide nucleotide sequence data for the entire

template nucleic acid. This would potentially allow sequence scaffolds to be built much more easily than can be done with short-read systems, enabling structural analysis of previously impossible-to-sequence sections of highly repetitive DNA, given the sequencing system is capable of long reads in the absence of photodamage.

IV. Software and Algorithm Implementations

[00146] The methods herein may operate with numerous methods for sequence alignment including those generated by various types of known multiple sequence alignment (MSA) algorithms. For example, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read using a reference sequence. In some embodiments in which a reference sequence is known for the region containing the target sequence, the reference sequence can be used to produce an MSA using a variant of the center-star algorithm. Alternatively, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read relative to every other read without using a reference sequence ("*de novo* assembly routines"), e.g., PHRAP, CAP, ClustalW, T-Coffee, AMOS make-consensus, or other dynamic programming MSAs. Depending on the sequence-generating methods used, the determination of sequence alignment may also involve analysis of read quality (e.g., using TraceTuner™, Phred, etc.), signal intensity, peak data (e.g., height, width, proximity to neighboring peak(s), etc.), information indicative of the orientation of the read (e.g., 5'→3' designations), clear range identifiers indicative of the usable range of calls in the sequence, and the like. Additional algorithms and systems for sequence alignment are well know to those of skill in the art, and are described further, e.g., in G. A. Churchill, M. S. Waterman (1992) "The Accuracy of DNA Sequences: Estimating Sequence Quality," *Genomics* 14: 89-98; M. Stephens, et al. (2006) "Automating sequence-based detection and genotyping of SNPs from diploid samples," *Nat. Genet.*, **38**: 375-381; J. Hein (1989) *Mol. Biol. Evol.*, **6**: 649-668; U.S.S.N. 12/134,186, filed June 5, 2008; and U.S.S.N. 61/116,439, filed November 20, 2008.

[00147] A standard sequence alignment problem in the context of DNA sequencing is to align the sequence of a relatively short fragment (<2 kilobases) to a large target sequence. The assumption is made that this fragment represents a contiguous portion of DNA to be mapped to a single location on the reference sequence. (A "contiguous portion" to be mapped to a single location may contain small insertions and/or deletions and still be considered contiguous in this context.) With the further development of nucleic acid sequencing technologies (e.g., from Illumina,

Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), and Applied Biosystems, Inc. (Foster City, CA)) and mate-pair sequencing protocols (see, e.g., U.S. Patent Pub. No. 2006/0292611 A1, which is incorporated by reference herein in its entirety for all purposes), the alignment problem has been extended to align two fragments coming from the same read to the reference sequence using some knowledge of the expected mate-pair configuration (distance and orientation).

[00148] With regards to mate-paired reads, mapping two fragments with a distance constraint and orientation constraint has been treated by various short-read mapping algorithms, e.g., SOAP (Li, et al. (2008) Bioinformatics, **24**, 713-714); SOAPdenovo; and Maq, a set of programs that map and/or assemble fixed-length Solexa/SOLiD reads (SourceForge, Inc.). While these algorithms can handle simple cases of mate-pair alignment, which generally treat the specific problem of only two reads coming from a mate-paired sequence and use the distance constraint as a hard filter (i.e., if two reads are within *x* bp of each other and in the correct orientations, report them as a mate-pair hit), the methods provided herein are more general and can handle much more complex data sets, including those with multiple reads, those for which a reference sequence is or is not present, potential non-template sequence (e.g., adapter regions or linker portions described below), and complex distance and orientation constraints. Other programs are also available that attempt to generalize on top of the mapping and aligning performed by the programs described above. These include, e.g., Breakdancer, variationhunter, GASV, etc., which can handle more complex mappings, e.g., by clustering.

[00149] Real-time single molecule sequencing presents opportunities for obtaining much more complex sequence fragments from a single DNA sequencing read. Two examples are the reading of multiple discontiguous sequence fragments from a single long stretch of DNA using a pulsed or intermittent detection system (e.g., intermittent illumination) as described herein and the contiguous reading of forward, reverse and adapter fragments from a circular templates (SMRTbell™ templates; see e.g., U.S.S.N. 61/099,696, filed September 24, 2008; U.S. Patent Application No. 12/383,855, filed March 27, 2009 and U.S. Patent Application No. 12/413,258, filed March 27, 2009, all of which are incorporated by reference herein in their entireties for all purposes). Further, methods for sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent

Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.

[00150] Certain aspects of the invention provide methods for optimally aligning such sequences to a reference sequence using knowledge of the molecular configuration and/or sequencing protocol used to generate the related sequence reads. In particular, methods are provided to address the general problem of mapping multiple fragments to a reference sequence with variable distance and orientation constraints.

[00151] Beginning with raw sequence data generated by a nucleic acid sequencing instrument (step 1), the sequence data is mapped to a target sequence (step 2) using a local alignment method which produces sub-optimal local alignments as well as the optimal alignment, for example, the Smith-Waterman algorithm. Another, more flexible example of a local alignment method is a chaining method using a method for aligning very short fragments to the target sequence (e.g., kmer-indexing, suffix trees, suffix arrays, etc.) and chaining the resulting hits back into longer chains of significant matches (see, e.g. D. Gusfield, Algorithms on Strings, Trees, and Sequences, Cambridge University Press: Cambridge, UK, 1997, which is incorporated by reference herein in its entirety for all purposes). The chains do not necessarily need to be refined by dynamic programming in order to be useful for the following algorithm, permitting a very fast algorithm. In certain embodiments, dynamic-programming refinement of the chain might improve the power (area under the ROC curve) of the algorithm.

[00152] The target sequence consists of the potential hypotheses for the molecular template in question. In the example of nucleic acid sequencing methods using iterative illumination for sequencing a shotgun fragment from a linear DNA sequence, the potential hypotheses are both orientations of the genome (since we do not know the original orientation of the fragment). In the example of sequencing of a SMRTbell™ template (e.g., see Example 1 herein), the hypotheses include both orientations of the genome and known adapter sequences. The parameters determining how many hits are reported for each local fragment can be varied to change the specificity and sensitivity of this algorithm. Figure 6 shows what these hits might look like for a SMRTbell™ template (represented as paths in the sequence alignment matrix, which is often called the dynamic-programming matrix, although it isn't necessary to use dynamic programming to find these paths).

[00153] After the potential local alignments have been enumerated, a weighted directed graph is constructed with each local alignment represented as a node in the graph (step 3). The edges are drawn between nodes if they represent a potential reconstruction of the original molecular

template using knowledge of the expected molecular configuration. The directed connection of an alignment path A to an alignment path B is interpreted as "The target sequence represented by B could follow the target sequence represented by A in the original molecule." For example, if a linear single-stranded DNA molecule is being sequenced by a method that uses iterative illumination, then fragments from opposite orientations would not be expected to be connected (unless the linear single-stranded DNA molecule also included oppositely oriented sequences, e.g., as in the case of a linearized SMRTbell™ template.) In general, fragments that represent the same stretch of the sequencing read but that align to different regions on the target sequence would not be connected. Aside from these examples, the rules for connecting nodes should be fairly loose to permit exploration of weak possibilities that gain significance when all the evidence (e.g. all the sequence reads) are considered. The assignment of edge weights handles the proper weighting of the likelihood of these edges, and the speed of the algorithm can be tuned by optimizing the pruning of highly unlikely edges. As usual this represents a tradeoff between speed and sensitivity.

[00154] Weights are assigned to connections ($A \rightarrow B$) in the graph representing the log-likelihood that target fragment A is followed by target fragment B in the original molecule.

$$w(A \rightarrow B) = -\log P(B|A)$$

The conditional probability $P(B\ A)$ encodes the knowledge of the possible molecular configurations and the alignment significance of B.

$$P(B|A) = f(B)g(A,B)$$

where f is a measure of alignment significance (either theoretical or empirically obtained) and g encodes the physical constraints representing the allowed molecular configurations.

[00155] For example, in the context of sequencing using iterative illumination the following may be known: the time between the end of one fragment and the beginning of the next fragment is 200 seconds. If the polymerase incorporates bases with an average rate of 4bp/sec with a standard deviation of 1bp/sec, it can be hypothesized that the probability of target fragment 2 following target fragment 1 is determined by the distance between these fragments on the target and a normal probability:

$$g(A,B) = \frac{1}{\sqrt{2\pi(200)}} \exp\left[-(d-800)/2(200)^2\right]$$

[00156] In a SMRTbell™ template example, knowledge of the expected insert size and the observed distance and orientation between fragments would be used to weight the likelihood that

these two fragments could come from a correctly generated SMRTbell™ template. This weight could include the expected rate of the polymerase as well and rules for the orientation of fragments with respect to each other and their distance apart in the original read. For example, while it may be expected that two forward fragments mapping to the same region in the target genome potentially come from multiple passes around a SMRTbell™ template molecule, those fragments would not be expected to be immediately adjacent in sequencing time. The weighting function would account for the proper amount of expected time between such fragments (i.e. the elapsed time would be expected to be long enough to include two adapter sequences and a reverse sequence).

[00157] In general, the weighting function could be arbitrarily complex and tuned to empirically observed relationships between sequencing fragments given the available knowledge (distance between fragments on the target sequence, sequencing time between fragments, expected length of the template, etc.). For example, the empirical probability distributions might be observed to exhibit longer tails than a Gaussian probability model might predict. The use of a conditional log-likelihood for the assignment of edge weights is motivated by the following logic. In a graph of possible local alignments it is desirable to find a highly likely path that best explains the observed data. Consider a path through three nodes *A*, *B*, and *C*, with *P(ABC)* being the probability that *ABC* is the correct assignment:

$$P(ABC) = P(C|AB)P(B|A)P(A)$$
$$\approx P(C|B)P(B|A)P(A)$$

where the last approximation is justified by the observation that the constraints between allowable assignments to the target sequence are typically local in nature. Generalizing this formula for a path $a_1, \ldots, a_N$ and taking the negative logarithm of both sides gives

$$-\log P(a_1 \ldots a_N) = -\sum_{i=1}^{N-1} \log P(a_{i+1}|a_i) - \log P(a_1)$$

[00158] It is apparent that the edge weights are additive if we use log-likelihood and we can use standard shortest-path algorithms for directed graphs to find the optimal path. A hypothetical directed graph is illustrated in Figure 7. This graph corresponds to the situation depicted by the alignments pictured above. Heavier lines correspond to more likely paths with the optimal path shown in blue. Dashed lines represent forbidden transitions. Not all paths are considered in the illustration to avoid clutter in the presentation. The general formula listed above includes a "one-body" term $P(a_1)$ for the starting node in each path that weights the probability that this initial alignment is correct. To accommodate this probability in a path-finding algorithm we add a

pseudo-source s to the graph which connects to every possible node (not shown in the graph above). The edge weight connecting the pseudo-source with a node a_i is $-\log P(a_i)$. This allows the use of a conventional single-source shortest-path algorithm starting from the pseudo-source. The desired probability $P(a_i)$ can come from a measure of alignment significance (theoretical or empirically determined) or could be set uniformly across all alignments to allow the path logic to determine the best path assignment, independent of the relative value of the starting points. It is anticipated that a threshold will be required here to only allow edges between the pseudo-source and nodes for highly likely alignments; otherwise the shortest path algorithm in the next step will not give the desired path.

[00159] After construction of the weighted directed graph, the shortest path to each node is determined (step 4). The graph is directed and acyclic (DAG) so we can use the standard shortest-path DAG algorithm (see T.H. Cormen, CE Leiserson, RL Rivest, Introduction to Algorithms, MIT Press: Cambridge, Massachusetts, 1990). This algorithm scales as O(V+E) and should be very quick for these graphs. After the shortest path to each node is determined, the paths need to be ranked to declare the best assignment. It is suggested that the best metric would be a measure which rewards paths that explain more of the sequenced read (longer paths) with high likelihood. One such metric would be the normalized negative log-likelihood: dividing the total weight of the path by the number of bases in the sequenced read explained by this path. For more complicated graphs or edge-weight assignments, Dijkstra's algorithm, the Bellman-Ford algorithm, or the A* algorithm could be applied. Other algorithms that may also be used include, but are not limited to the Floyd-Warshall algorithm.

[00160] For noisy sequence data it is likely that the local alignments found in step 2 will occasionally overlap with each other in the sequenced read even though it is physically impossible for such overlaps to occur in a perfect system (unless there has been a rearrangement relative to the reference genome). As such, some amount of slack must be allowed in the edge assignment logic in step 3 to account for not knowing the precise boundaries of each local alignment. Once the best physical model explaining the observed read is determined, the boundaries of the local alignments can be refined to reflect the physical necessity that each base in the sequenced read can only be represented in one local alignment. It is also desirable to explain all of the bases in between the local alignments that haven't been assigned in the graph. One straightforward approach to refinement would be to construct the perfect model of the sequence and to realign the sequenced read to this sequence. This refinement algorithm would preserve physical constraints (each base in

the sequenced read can only be explained by one location in the template) and would assign all bases between the extremal nodes in the optimal path.

[00161] Certain aspects of the software and algorithm implementations described herein may be varied or altered without departing from the spirit and scope of the invention. For example, with regards to algorithm seeding, many algorithms can be applied for the original determination of sub-optimal local alignments (step 2). Conventional examples include FASTA, BLAST, or Smith-Waterman. It is expected that the best benefit will be obtained from using short-sequence alignment algorithms (suffix array, suffix tree, Boyer-Moore, Rabin-Karp, kmer-indexing, and the like) followed by chaining to establish regions of significant matches. An advantage of the algorithm described here is that it does not require dynamic-programming refinement of the resulting chains and therefore can be quite fast, however it is expected that using dynamic programming to refine the chains in step 2 could increase the power of the algorithm.

[00162] With regards to graph construction, there will be advantages to tuning the logic of edge assignments to keep the size of the graph manageable. It is possible that steps 2 and 3 might be combined to in a greedy fashion to focus the potentially slow step 2 into productive areas of the graph. For example, if a particularly strong hit is found early in step 2, then it may be beneficial to search for sub-optimal hits only in this local vicinity, knowing that this strong hit should be in the final solution. Tuning of the graph construction might include thresholds, below which edges are not created. Further, there are multiple parameters (minimum chain length, minimum probability for edge assignment, relative weighting of length vs. accuracy, etc.) which can be exposed and tuned in this algorithm to maximize the sensitivity and specificity of the algorithm for a given scenario.

[00163] With regards to determination of the distance a polymerase travels between reads, various strategies are provided that are more sophisticated than estimation based upon the rate of incorporation and the time between detection (e.g., illuminated) periods. In certain embodiments, the distribution of the base pair distance travelled by a polymerase during a non-detection period is called $p(x)$. The distribution of enzyme velocities, $p(v)$, is estimated by aligning observed reads to a reference sequence, and this distribution is represented as the number of reference bases per unit time. There is a length of time, τ, over which measurement of the instantaneous rate is not independent. While this method of determining the distance the polymerase travels during a non-detection period should not be overly sensitive to non-independent estimation of the polymerase rate, it is likely to strive for independent measurements of the rate. The distance τ can be estimated

from an exponential fit to the auto-correlation function $<\delta v(t)\delta v(t-\Delta t)>$, and $v(t)$ tabulated across the aligned sequence at increments of τ.

[00164] Where multiple single polymerase enzymes are being observed simultaneously, e.g., each being optically resolvable from every other on a single array, the $p(v)$ for each is preferably determined independently for each enzyme. Further, information regarding rare but extended events, such as polymerase "stalling" on the template, can be measured across a larger data set. For example, the statistics of stalls can be determined by aggregating rate measurements across an entire array. Where a stall distribution is characterized by a "long tail" corresponding to multi-exponential behavior of IPD distribution, such a distribution of polymerization rates can be extended for stalls longer than the observed reaction by fitting the long-tail behavior to an appropriate functional form, e.g., using a single-exponential parametric model or other physically motivated model (e.g., multi-exponential, stretched exponential, power-law, etc.) In certain preferred embodiments, the following representation of a "per-enzyme" $p(v)$ is used:

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}$$

where $f(v)$ is an interpolating function designed to retain information about the zero-velocity tail of the global $p(v)$ distribution while taking the estimate of the polymerase velocity dynamics(e.g., the dominant high velocity mode) from the specific enzyme. Such an interpolating function is:

$$f(v) = \frac{1 + erf(v / v_0)}{2}$$

where v_0 is a scale parameter to be chosen based on experience (but optionally fixed). Alternatively, the average of the empirical $p_{enzyme}(v)$ and $p_{array}(v)$ can be used. This approach can be motivated by a Bayesian approach to density estimation. Other kernel density and Bayesian methods can be suggested. Alternatively or in addition, the robustness of $p(v)$ to conditions and daily phenomenon can be explored and used to estimate $p(v)$ more globally, e.g., using one or more weekly control experiments.

[00165] Given the lack of a known reference sequence for *de novo* assembly, several alternative ways to formulate $p(v)$ are provided as follows. For example, in a first embodiment a control template (essentially a proxy reference sequence) can be subjected to sequencing, e.g., in the same reaction as the *de novo* sample or in an identical reaction. The observed velocity for the sequencing reactions would be measured based upon alignments of the reads from the control template to its known sequence. Typically, a per-enzyme correction would not be available for the

p(v) and *p(v)* would default to an array-averaged *p(v)*. In a second embodiment, a previously determined *p(v)* from experiments using a known reference sequence can be used, e.g., where the previous experiments were performed under the same conditions as the *de novo* experiments. In a third embodiment, *p(v)* is estimated by using quality information/metrics to screen for the most likely "true" calls, and restricting the estimates of *v* to regions containing those calls. In a fourth embodiment, where error is low, the called base rate and reference base rate converge to the same rate, and measurements of p(v) without knowledge of the reference become substantially reliable. Further, even if they do not fully converge, they can still be used to accurately infer *p(v)*,as long as the called base rate is predictably higher/lower than the reference base rate. Yet further, the measurement of *p(v)* when a reference sequence is not available can benefit from a detailed look at the probability model which is available from an algorithm like a CRF. That is, *p(v)* can be tabulated using a weighted sum over paths through a CRF probability model.

[00166] As will be clear to the ordinary practioner based upon the teachings herein, this framework extends naturally to the measurements of other potentially systematic variations in *p(v)* across an array, e.g., even where a single reaction mixture is applied to the entire array. For example, the local temperature of the reaction environment can vary systematically across an array of reactions. The average and variation in the rates of polymerase enzymes on the array would likely have a dependence on this hidden variable. Where the functional form of the temperature dependence is known, the measurement of *p(v)* can be stabilized across the array by modeling a de novo *p(v)* as $p_{cond}(v) + p_{x,y}(v)$ where *x,y* are geometrical variables defining the location on the array. Further, *p(v)* has been found to be somewhat variable over time. As such, in certain embodiments a model of *p(v;t)* is developed using an appropriate model for the evolution of *p(v)* over time.

[00167] Once a representative distribution of velocities *p(v)* has been obtained for a given read from a given reaction, the expected travel distance in the non-detection period can be expressed as:

$$p\left(\frac{x}{\tau}\right) = IL\left[L[p(v)]^{t/\tau} \frac{1 - L[p(v)]}{s} \right]$$

where L[] and IL[] stand for the Laplace and inverse Laplace transform, respectively. A similar result is derived in Svoboda, et al. (PNAS 91:11782 (1994)) and readily follows from considering the pdf of a sum of random variables. Optionally, in certain embodiments density estimation techniques (e.g., kernel density estimation, etc.) are useful when modeling *p(v)* since they can smooth the resulting numerical calculations in the Laplace and inverse Laplace transform.

[00168] Knowledge of the complete distribution has several advantages over the commonly applied Gaussian approximation. For example, knowledge of the complete distribution of insert lengths is very desirable when using a Bayesian framework approach to detect structural variation. (See, e.g., Bashir, et al. (2008) PLoS Comput. Biol. 4:51; Hormozdiari, et al. (2009) Genome Res. 19:1270; and Lee, et al. (2008) Bioinformatics 24:59.) While Bashir, et al. does not strictly follow a Bayesian approach, the geometric approach described in the paper can be straightforwardly modified to incorporate an actual posterior instead of the boxcar posterior assumed in the paper. Further, during mapping of noncontiguous reads to a genome where they are expected to be concordant (*i.e.*, not a structural variation), it is useful to consider the known distribution when judging the significance of the resulting alignments between the observed reads and the genomic sequence. In addition, when clustering noncontiguous reads that scaffold contigs in a *de novo* assembly, a path of Bayesian significance can be followed that is very similar to that followed in the structural variation case discussed *supra*.

[00169] In further embodiments, the determination of the distance a polymerase travels between reads is performed using an algorithm based on a simulation approach rather than the exact analytical result used in the algorithm described above. This method relies on Monte Carlo sampling from a distribution, which allows a better extension to arbitrary empirical distributions. It also lacks the difficult computations of numerical Laplace and inverse Laplace transforms, and permits calculation of distances traveled during non-detection periods when the underlying kinetic processes have multi-phasic kinetics, e.g., the presence of long stalls.

[00170] This approach aims to calculate the distribution of the distance x travelled by an enzyme during a time t during which it was not being observed (e.g., during a non-detection period). In some embodiments, a distribution of local rates, $p(v)$, is estimated, where the definition of "local" is set by the correlation length of the rate autocorrelation function, e.g.:

$$\langle \delta v(t) \delta v(t+\Delta) \rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$$

Given a local rate distribution and an assumption that independent identically distributed (i.i.d.) draws can be made from this distribution, one approach to calculating the distribution is as follows. First, draw $N = t/\tau_{corr}$ velocities from $p(v)$, and subsequently sum them and record them as an estimate of x/τ_{corr}. Repeat the process M times, with the optimal choice of M dependent on the desired level of precision for estimation of the $p(x)$ distribution. In certain preferred embodiments,

M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000.

[00171] In some embodiments in which the enzyme system is not well explained by a single kinetic process or cycle (as in the case of observed stalling behavior), above-described rate autocorrelation function and the i.i.d. assumption will be violated. As such, a probability model having a richer structure is preferably used. One such probability model is a Hidden Markov Model (HMM). Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system. Where the kinetics of the pausing state can be well described by a single-exponential, this model is expected to describe the observed distribution of local velocities. The single-exponential assumption is implicit in the state structure of the model since the amount of time spent in the pause state will be a geometric distribution with mean p/(1-p) [*i.e.*, the observed stall times will have to be added to this model]. If the stall kinetics are multi-phasic, then more "dark states" will have to be added to this model. Further, the model shown in Figure 15 can actually be treated as a Markov Model and not a Hidden Markov Model without much loss of generality because the "pause" state is not actually hidden due to the fact that the data collected during the pause state is highly distinguishable from the data collected during the sequencing state. As such, the general HMM apparatus is not necessary. The model in Figure 15 can be used to simulate the distribution of local velocities when there is a long-term pause or stall phase present in the reaction data kinetics. S0 is the start state, and there is no explicit end state since this model is used as a generative model and it is assumed that it is run forward for a prescribed number of steps. The qualities P(P→S) and P(S→P) represent exit from a stalled state and entry into a stalled state, respectively. These qualities can be measured by an EM algorithm or they can be quickly estimated by physical observables. $P_{P\to S} = 1/\left(1+\frac{\tau_{stall}}{\tau_{corr}}\right)$ and P(S→P) is the frequency of stall starts per τ_{corr}. (Example parameters are τ_{stall}=80 seconds; τ_{corr} = 10 seconds; and P(S→P) = 1/24.) The simulation estimate of $p(x)$ can now be produced using the procedure outlined above in which $N = t/\tau_{corr}$ velocities are drawn from $p(v)$; and they are subsequently summed and recorded as an estimate of x/τ_{corr}. The process is repeated M times, with the optimal choice of M dependent on the desired level of precision for estimation of the $p(x)$ distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000. Figure 16 shows exemplary simulated applications of this method. Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters P(S→P) = 1/24; P(P→S) = 1/11;

and $p(v)$~Gamma(48,0.25). Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a 1300 second non-detection period, which is calculated by sampling 2000 estimates from the HMM model.

[00172] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period. The influence of pause/stall behavior can be seen in the heavy-left tailing of both distributions.

[00173] While the simulation method in which i.i.d. draw assumption is valid is more general and can treat arbitrary $p(v)$ and more complex models for non-sequencing states, the two-state model using the HMM can be treated analytically. The result of this is:

$$p(x/\tau_{corr}) = \sum_{N_S=0}^{N} \pi_{N_S}(x) p_N(N_S)$$

where $\pi_{N_S}(x)$ is the distribution of the sum of N_S variables drawn from $p(v)$. For the general case, this distribution is given by the Laplace transform approach presented above. For $p(v)$~Normal(μ,σ), this distribution is distributed as Normal($N_S\mu$,$\sqrt{N_S}\sigma$). For $p(v)$~Gamma(k,θ), this distribution is distributed as Gamma(N_Sk,θ). $P_N(N_S)$ is the number of cycles spent in the sequencing state if we observe N cycles from the Markov process in Figure 15. The expression for this is described in Pedler, et al. (1971) J. Appl. Prob. 8:381, which is incorporated herein by reference in its entirety for all purposes.

[00174] As will be clear to one of ordinary skill in the art upon review of the teachings herein, these methods can be readily extended to the non-detection period estimations of procession by other cyclical biological reactions, such as the action of reverse transcriptase or the synthesis of proteins by a ribosome complex, e.g., and certain preferred embodiments of such reactions are further described in U.S.S.N. 12/767,673, filed April 26, 2010; and U.S.S.N. 12/813,968, filed June 11, 2010, the disclosures of which are incorporated herein by reference in their entireties for all purposes. Further, the simulation model described above is not restricted to simple two-state kinetics, and the use of $p(v)$ is not restricted to analytical models. In fact, in certain embodiments, empirical estimates are preferably used.

[00175] Although useful in certain preferred embodiments of the invention, certain algorithms as presented above do not easily handle the case where the template does not match a physically-motivated expected model. A relevant example of such a case is when the template contains a genomic structural variation (SV), such as translocation, whereby two fragments which are correctly adjacent in the template are located very far apart in the reference genome. Such

structural variation cases are best handled in the context of the current algorithm by reporting the confidence of an observed path and reporting situations when no physically expected path seems to fit the observed data. In general, the detection of structural variation requires the presence of multiple highly significant local alignments which can be identified as significantly overturning the null hypothesis of matching the genomic ordering of fragments with their own individual merit. Nevertheless, with molecular redundant sequencing such as SMRTbell™ template sequencing the current algorithm can be adapted to improve the ability to identify an SV event. Such a modification could be a feedback approach which allows modification of the linking constraints in step 3 to allow very far separations on the target sequence when the individual alignments are very significant. Only one such highly-significant pair would be needed to enable the rescue of less significant partial matches that support the same SV hypothesis.

[00176] The software and algorithm implementations provided herein are particularly suited for transforming sequence read data generated from various sequencing technologies (e.g., sequencing-by-synthesis, intramolecular redundant sequencing, Sanger sequencing, capillary electrophoretic sequencing, pyrosequencing, ligase-mediated sequencing, etc.) into consensus sequence data that provides a representation of the actual nucleotide sequence of the template nucleic acid that was subjected to the sequencing reaction(s) from which the sequence read data was generated. The software and algorithm implementations provided herein are preferably machine-implemented methods. The various steps recited herein are preferably performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. The results of these methods are preferably stored on a machine-readable medium, as well. Further, the invention provides a computer program product comprising a computer usable medium having a computer readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention.

[00177] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data).

[00178] While described with reference to certain specific applications above, it will be understood that these methods are also applicable to other types of complex data sets, and the invention should not be limited to only the specific examples provided herein. Other applications of the instant methods will be clear to those of ordinary skill in the art and are considered to be additional aspects of the instant invention.

V. Devices and Systems

[00179] The invention also provides systems that are used in conjunction with the compositions and methods of the invention in order to provide for intermittent detection of analytical reactions. In particular, such systems typically include the reagent systems described herein, in conjunction with an analytical system, e.g., for detecting data from those reagent systems. For example, a sequencing reaction may be subjected to intermittent illumination, and the sequencing system may include the system components provided with or sold for use with commercially available nucleic acid sequencing systems, such as the Genome Analyzer System available from Illumina, Inc., the GS FLX System, available from 454 Life Sciences, or the ABI 3730 System available from Life Technologies, Inc.

[00180] In certain preferred embodiments, reactions subjected to intermittent illumination are monitored using an optical system capable of detecting and/or monitoring interactions between reactants at the single-molecule level. Such an optical system achieves these functions by first generating and transmitting an incident wavelength to the reactants, followed by collecting and analyzing the optical signals from the reactants. Such systems typically employ an optical train that directs signals from the reactions to a detector, and in certain embodiments in which a plurality of reactions is disposed on a solid surface, such systems typically direct signals from the solid surface (e.g., array of confinements) onto different locations of an array-based detector to simultaneously detect multiple different optical signals from each of multiple different reactions. In particular, the optical trains typically include optical gratings or wedge prisms to simultaneously direct and separate signals having differing spectral characteristics from each confinement in an array to different locations on an array based detector, *e.g.*, a CCD, and may also comprise additional optical transmission elements and optical reflection elements.

[00181] An optical system applicable for use with the present invention preferably comprises at least an excitation source and a photon detector. The excitation source generates and transmits incident light used to optically excite the reactants in the reaction. Depending on the intended

application, the source of the incident light can be a laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source. Further, the excitation light may be evanescent light, e.g., as in total internal reflection microscopy, certain types of waveguides that carry light to a reaction site (see, e.g., U.S. Application Pub. Nos. 20080128627, 20080152281, and 200801552280), or zero-mode waveguides, described below. Where desired, more than one source can be employed simultaneously. The use of multiple sources is particularly desirable in applications that employ multiple different reagent compounds having differing excitation spectra, consequently allowing detection of more than one fluorescent signal to track the interactions of more than one or one type of molecules simultaneously. A wide variety of photon detectors or detector arrays are available in the art. Representative detectors include but are not limited to optical reader, high-efficiency photon detection system, photodiode (*e.g.* avalanche photo diodes (APD)), camera, charge couple device (CCD), electron-multiplying charge-coupled device (EMCCD), intensified charge coupled device (ICCD), and confocal microscope equipped with any of the foregoing detectors. For example, in some embodiments an optical train includes a fluorescence microscope capable of resolving fluorescent signals from individual sequencing complexes. Where desired, the subject arrays of optical confinements contain various alignment aides or keys to facilitate a proper spatial placement of the optical confinement and the excitation sources, the photon detectors, or the optical train as described below.

[00182] The subject optical system may also include an optical train whose function can be manifold and may comprise one or more optical transmission or reflection elements. Such optical trains preferably encompass a variety of optical devices that channel light from one location to another in either an altered or unaltered state. First, the optical train collects and/or directs the incident wavelength to the reaction site (e.g., optical confinement). Second, it transmits and/or directs the optical signals emitted from the reactants to the photon detector. Third, it may select and/or modify the optical properties of the incident wavelengths or the emitted wavelengths from the reactants. In certain embodiments, the optical train controls an on/off cycle of the illumination source to provide illuminated and non-illuminated periods to one or more illuminated reaction sites. Illustrative examples of such optical transmission or reflection elements are diffraction gratings, arrayed waveguide gratings (AWG), optic fibers, optical switches, mirrors (including dichroic mirrors), lenses (including microlenses, nanolenses, objective lenses, imaging lenses, and the like), collimators, optical attenuators, filters (*e.g.*, polarization or dichroic filters), prisms, wavelength filters (low-pass, band-pass, or high-pass), planar waveguides, wave-plates, delay lines, and any

other devices that guide the transmission of light through proper refractive indices and geometries. One example of a particularly preferred optical train is described in U.S. Patent Pub. No. 20070036511, filed August 11, 2005, and incorporated by reference herein in its entirety for all purposes.

[00183] In a preferred embodiment, a reaction site (e.g., optical confinement) containing a reaction of interest is operatively coupled to a photon detector. The reaction site and the respective detector can be spatially aligned (*e.g.*, 1:1 mapping) to permit an efficient collection of optical signals from the reactants. In certain preferred embodiments, a reaction substrate is disposed upon a translation stage, which is typically coupled to appropriate robotics to provide lateral translation of the substrate in two dimensions over a fixed optical train. Alternative embodiments could couple the translation system to the optical train to move that aspect of the system relative to the substrate. For example, a translation stage provide a means of removing a reaction substrate (or a portion thereof) out of the path of illumination to create a non-illuminated period for the reaction substrate (or a portion thereof), and returning the substrate at a later time to initiate a subsequent illuminated period. An exemplary embodiment is provided in U.S. Patent Pub. No. 20070161017, filed December 1, 2006.

[00184] In particularly preferred aspects, such systems include arrays of reaction regions, e.g, zero-mode waveguide arrays, that are illuminated by the system, in order to detect signals (e.g., fluorescent signals) therefrom, that are in conjunction with analytical reactions being carried out within each reaction region. Each individual reaction region can be operatively coupled to a respective microlens or a nanolens, preferably spatially aligned to optimize the signal collection efficiency. Alternatively, a combination of an objective lens, a spectral filter set or prism for resolving signals of different wavelengths, and an imaging lens can be used in an optical train, to direct optical signals from each confinement to an array detector, *e.g.*, a CCD, and concurrently separate signals from each different confinement into multiple constituent signal elements, *e.g.*, different wavelength spectra, that correspond to different reaction events occurring within each confinement. In preferred embodiments, the setup further comprises means to control illumination of each confinement, and such means may be a feature of the optical system or may be found elsewhere is the system, e.g., as a mask positioned over an array of confinements. Detailed descriptions of such optical systems are provided, e.g, in U.S. Patent Pub. No. 20060063264, filed September 16, 2005, which is incorporated herein by reference in its entirety for all purposes.

[00185] The systems of the invention also typically include information processors or computers operably coupled to the detection portions of the systems, in order to store the signal data obtained from the detector(s) on a computer readable medium, e.g., hard disk, CD, DVD or other optical medium, flash memory device, or the like. For purposes of this aspect of the invention, such operable connection provide for the electronic transfer of data from the detection system to the processor for subsequent analysis and conversion. Operable connections may be accomplished through any of a variety of well known computer networking or connecting methods, e.g., Firewire®, USB connections, wireless connections, WAN or LAN connections, or other connections that preferably include high data transfer rates. The computers also typically include software that analyzes the raw signal data, identifies signal pulses that are likely associated with incorporation events, and identifies bases incorporated during the sequencing reaction, in order to convert or transform the raw signal data into user interpretable sequence data (See, e.g., Published U.S. Patent Application No. 2009-0024331, the full disclosure of which is incorporated herein by reference in its entirety for all purposes).

[00186] Exemplary systems are described in detail in, e.g., U.S. Patent Application No. 11/901,273, filed September 14, 2007 and U.S. Patent Application No. 12/134,186, filed June 5, 2008, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00187] Further, as noted above, the invention provides data processing systems for transforming sequence read data into consensus sequence data. In certain embodiments, the data processing systems include machines for generating sequence read data by interrogating a template nucleic acid molecule. In certain preferred embodiments, the machine generates the sequence read data using a sequencing-by-synthesis technology, as described elsewhere herein, but the machine may generate the sequence read data using other sequencing technologies known to those of ordinary skill in the art, e.g., pyrosequencing, ligation-mediated sequencing, Sanger sequencing, capillary electrophoretic sequencing, etc. Such machines and methods for using them are available to the ordinary practioner.

[00188] The sequence read data generated is representative of the nucleotide sequence of the template nucleic acid molecule only to the extent that a given sequencing technology is able to generate such data, and so may not be identical to the actual sequence of the template nucleic acid molecule. For example, it may contain a deletion or a different base at a given position as compared to the actual sequence of the template, e.g., when a base call is missed or incorrect, respectively. As

such, it is beneficial to generate redundant sequence read data, and the methods described herein provide manipulations and computations that transform redundant sequence read data into consensus sequence data that is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Redundant sequence read data comprises multiple reads, each of which includes at least a portion of sequence read that overlaps with at least a portion of at least one other of the multiple reads. As such, the multiple reads need not all overlap with one another, and a first subset may overlap for a different portion of the template nucleic acid sequence than does a second subset. Such redundant sequence read data can be generated by various methods, including repeated sequencing of a single nucleic acid template, sequencing of multiple identical nucleic acid templates, or a combination thereof.

[00189] In another aspect, the data processing systems can include software and algorithm implementations provided herein, e.g. those configured to transform redundant sequence read data into consensus sequence data, which, as noted above, is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Further, the transformation of the redundant sequence read data into consensus sequence data identifies and negates some or all of the single-read variation between the multiple reads in the redundant sequence read data. As such, the transformation provides a representation of the actual nucleotide sequence of the nucleic acid template from which redundant sequence read data is generated that is more accurate than a representation based on a single read.

[00190] The software and algorithm implementations provided herein are preferably machine-implemented methods, e.g., carried out on a machine comprising computer-readable medium configured to carry out various aspects of the methods herein. For example, the computer-readable medium preferably comprises at least one or more of the following: a) a user interface; b) memory for storing redundant sequence read data; c) memory storing software-implemented instructions for carrying out the algorithms for transforming redundant sequence read data into consensus sequence data; d) a processor for executing the instructions; e) software for recording the results of the transformation into memory; and f) memory for recordation and storage of the resulting consensus sequence read data. In preferred embodiments, the user interface is used by the practitioner to manage various aspects of the machine, e.g., to direct the machine to carry out the various steps in the transformation of redundant sequence read data into consensus sequence data,

recordation of the results of the transformation, and management of the consensus sequence data stored in memory.

[00191] As such, in preferred embodiments, the methods further comprise a transformation of the computer-readable medium by recordation of the redundant sequence read data and/or the consensus sequence data generated by the methods. Further, the computer-readable medium may comprise software for providing a graphical representation of the redundant sequence read data and/or the consensus sequence read data, and the graphical representation may be provided, e.g., in soft-copy (e.g., on an electronic display) and/or hard-copy (e.g., on a print-out) form.

[00192] The invention also provides a computer program product comprising a computer-readable medium having a computer-readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention. In certain preferred embodiments, the computer program product comprises the computer-readable medium described above.

[00193] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data), such as the computer-readable medium described above.

[00194] As shown in Figure 9, the system 900 includes a substrate 902 that includes a plurality of discrete sources of chromophore emission signals, e.g., an array of zero-mode waveguides 904. An excitation illumination source, e.g., laser 906, is provided in the system and is positioned to direct excitation radiation at the various signal sources. This is typically done by directing excitation radiation at or through appropriate optical components, e.g., dichroic 108 and objective lens 910, that direct the excitation radiation at the substrate 902, and particularly the signal sources 904. Emitted signals from the sources 904 are then collected by the optical components, e.g., objective 910, and passed through additional optical elements, e.g., dichroic 908, prism 912 and lens 914, until they are directed to and impinge upon an optical detection system, e.g., detector array 916. The signals are then detected by detector array 916, and the data from that detection is

transmitted to an appropriate data processing system, e.g., computer 918, where the data is subjected to interpretation, analysis, and ultimately presented in a user ready format, e.g., on display 920, or printout 922, from printer 924. As will be appreciated, a variety of modifications may be made to such systems, including, for example, the use of multiplexing components to direct multiple discrete beams at different locations on the substrate, the use of spatial filter components, such as confocal masks, to filter out-of focus components, beam shaping elements to modify the spot configuration incident upon the substrates, and the like (See, e.g., Published U.S. Patent Application Nos. 2007/0036511 and 2007/095119, and U.S. Patent Application No. 11/901,273, all of which are incorporated herein by reference in their entireties for all purposes.)

VI. Exemplary Applications

[00195] The methods and compositions of the invention are useful in a broad range of analytical reactions in which one or more aspects of a detection method are detrimental to one or more aspects of the analytical reaction, such as rate, duration, fidelity, processivity, and the like. In such cases, intermittent detection at least partially mitigates the detrimental effect while allowing collection of data from stages of the analytical reaction that were previously uncollectable. As noted above, illuminated reactions are one example of analytical reactions that benefit from the compositions and methods described herein, particularly those using photoluminescent or fluorescent reagents, and particularly such reactions where one or more of the reaction components that are susceptible to photo-induced damage are present at relatively low levels. One exemplary application of the methods and compositions described herein is in single molecule analytical reactions, where the reaction of a single molecule (or very limited number of molecules) is observed in the analysis, such as observation of the action of a single enzyme molecule. In another aspect, the present invention is directed to illuminated reactions for single molecule analysis, including sequencing of nucleic acids by observing incorporation of nucleotides into a nascent nucleic acid sequence during template-directed polymerase-based synthesis. Such methods, generally referred to as "sequencing-by-incorporation" or "sequencing-by-synthesis," involve the observation of the addition of nucleotides or nucleotide analogs in a template-dependent fashion in order to determine the sequence of the template strand. See, e.g., U.S. Patent Nos. 6,780,591, 7,037,687, 7,344,865, 7,302,146. Processes for performing this detection include the use of fluorescently labeled nucleotide analogs within a confined observation region, e.g., within a nanoscale well and/or tethered, either directly or indirectly to a surface. By using excitation illumination (i.e., illumination

of an appropriate wavelength to excite the fluorescent label and induce a detectable signal), the fluorescently labeled bases can be detected as they are incorporated into the nascent strand, thus identifying the nature of the incorporated base, and as a result, the complementary base in the template strand.

[00196] In particular aspects, when an analysis relies upon a small population of reagent molecules, damage to any significant fraction of that population will have a substantial impact on the analysis being performed. For example, prolonged interrogation of a limited population of reagents, e.g., fluorescent analogs and enzymes, can lead to photo-induced damage of the various reagents to the point of substantially impacting the activity or functionality of the enzyme. It has been shown that prolonged illumination of DNA polymerases involved in synthesis using fluorescent nucleotide analogs results in a dramatic decrease in the enzyme's ability to synthesize DNA, often measured as a reduction in processivity. Without being bound to any theory of operation, it is believed that in some cases a photo-induced damage event affects the catalytic region of the enzyme thus affecting either the ability of the enzyme to remain complexed with the template, or its ability to continue synthesis. In general, the methods, devices, and systems of the present invention can increase performance and/or selectively monitor one or more stages of an illuminated reaction by subjecting the reaction to intermittent illumination.

[00197] One particularly preferred aspect of the invention is in conjunction with the sequencing by incorporation of nucleic acids within an optical confinement, such as a zero-mode waveguide. Such reactions involve observation of an extremely small reaction volume in which one or only a few polymerase enzymes and their fluorescent substrates may be present. Zero-mode waveguides, and their use in sequencing applications are generally described in U.S. Patent Nos. 6,917,726 and 7,033,764, and preferred methods of sequencing by incorporation are generally described in Published U.S. Patent Application No. 2003-0044781, the full disclosures of which are incorporated herein by reference in their entireties for all purposes, and in particular for their teachings regarding such sequencing applications and methods. Briefly, arrays of zero-mode waveguides ("ZMWs"), configured in accordance with the present invention may be employed as optical confinements for single molecule DNA sequence determination. In particular, as noted above, these ZMWs provide extremely small observation volumes at or near the transparent substrate surface, also termed the "base" of the ZMW. A nucleic acid synthesis complex, e.g., template sequence, polymerase, and primer, which is immobilized at the base of the ZMW, may then be specifically observed during synthesis to monitor incorporation of nucleotides in a template

dependent fashion, and thus provide the identity and sequences of nucleotides in the template strand. This identification is typically accomplished by providing detectable label groups, such as fluorescent labeling molecules, on the nucleotides. In some instances, the labeled nucleotides terminate primer extension, allowing a "one base at a time" interrogation of the complex. If, upon exposure to a given labeled base, a base is incorporated, its representative fluorescent signal may be detected at the base of the ZMW. If no signal is detected, then the base was not incorporated and the complex is interrogated with each of the other bases, in turn. Once a base is incorporated, the labeling group is removed, e.g., through the use of a photocleavable linking group, and where the label was not the terminating group, a terminator, upon the 3' end of the incorporated nucleotide, may be removed prior to subsequent interrogation. In other more preferred embodiments, the incorporation of a labeled nucleotide does not terminate primer extension and the processive incorporation of multiple labeled nucleotides can be monitored in real time by detecting a series of fluorescent signals at the base of the ZMW. In some such embodiments, the label is naturally released upon incorporation of the labeled nucleotides by the polymerase, and so need not be released by alternative means, e.g., a photocleavage event. As such, a processive sequencing reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. (Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.)

[00198] In accordance with the present invention, sequencing reactions may be carried out by only interrogating a reaction mixture, e.g., detecting fluorescent emission for one or more illuminated periods before excessive photo-induced damage has occurred. In general, the methods described herein are implemented in a manner sufficient to provide beneficial impact, e.g., reduced photo-induced damage and/or extension of the photo-induced damage threshold period, but are not implemented in such a manner to interfere with the reaction of interest, e.g., a sequencing reaction. The present invention also contemplates alternative methods of and compositions for mitigating the impact of photo-induced damage on a reaction, as described above and in, e.g., U.S.S.N. 61/116,048, filed November 19, 2008. Such alternative methods and compounds can be used in combination with the compositions and methods provided herein to further alleviate the effects of species that can be generated during an illuminated reaction.

[00199] Another method of mitigating the impact of photo-induced damage on the results of a given reaction provides for the elimination of potentially damaging oxygen species using means other than the use of the photo-induced damage mitigating agents described above. In one example, dissolved oxygen species may be flushed out of aqueous systems by providing the reaction system under different gas environments, such as by exposing an aqueous reaction to neutral gas environments, such as argon, nitrogen, helium, xenon, or the like, to prevent dissolution of excess oxygen in the reaction mixture. By reducing the initial oxygen load of the system, it has been observed that photo-induced damage effects, e.g., on polymerase mediated DNA synthesis, is markedly reduced. In particularly preferred aspects, the system is exposed to a xenon atmosphere. In particular, since xenon can be induced to form a dipole, it operates as a triplet-state quencher in addition to supplanting oxygen in the aqueous system. (See, e.g., Vierstra and Poff, Plant Physiol. 1981 May; 67(5): 996–998) As such, xenon would also be categorized as a quencher, as set forth above.

[00200] Although described in terms of zero-mode waveguides, it will be appreciated that a variety of selective illumination strategies may be employed to selectively interrogate different regions of a solid support over time, e.g., so as to only damage molecules within certain selected regions of a substrate while not damaging molecules in other selected regions of the substrate. In certain embodiments, such methods can involve using a directed light source (e.g., a laser) to illuminate only selected regions; changing the illumination angle of the light source; or refocusing the illumination, e.g., by passing the illumination through an optical train that alters the shape of the incident light on the solid support. These and further examples of alternative methods of mitigating photo-induced damage which can be used in combination with methods and systems of the invention described herein are provided in U.S. Patent Pub. No. 20070036511, filed August 11, 2005; U.S. Patent No. 6,881,312; U.S.S.N. 61/116,048, filed November 19, 2008; and U.S. Patent Pub. No. 20070161017, filed December 1, 2006, all of which are incorporated herein by reference in their entireties for all purposes, and in particular for disclosure related to these methods of mitigating photo-induced damage.

[00201] As noted above, using templates that allow repeated sequencing (e.g., circular templates, SMRTbell™ templates, etc.) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated and/or increase the fold-coverage of the sequence reads for one or more regions of interest in the template, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus

sequences for the nucleic acid template. For example, in certain preferred embodiments, templates sequenced by the methods described herein are templates comprising a double-stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double-stranded or otherwise internally complementary. Such templates may, for example, comprise a double-stranded portion comprised of two complementary sequences and two single-stranded linking portions (e.g., oligos or "hairpins") joining the 3' end of each strand of the double-stranded region to the 5' end of the other strand (sometimes referred to as "SMRTbell™" templates). In certain embodiments, double-stranded portions for use in such templates are PCR-amplified. Optionally, restriction sites are incorporated within the PCR primers such that subsequent digestion of the amplified products with appropriate restriction enzymes generates double-stranded portions containing known overhang sequences on either end, which are then ligated to hairpin adapters containing a complementary overhang to generate the SMRTbell™ templates.

[00202] These template molecules are particularly useful as nucleotide sequence data generated therefrom comprises both sense and antisense nucleotide sequences for the double-stranded portion, and the circular conformation of the template enables repeated sequencing (e.g., using a polymerase capable of strand-displacement) provides duplicative or redundant sequence information. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining sequence data for each segment from the complementary sequences, as well as sequence data within each segment, by repeatedly sequencing that segment. Iterative illumination is useful in such sequencing applications, e.g., to focus nucleotide sequence data collection on stages of the sequencing reaction most of interest, such as the stages during which nucleotide sequence data is being generated from a strand of the (previously) double-stranded portion. Iterative illumination may also allow additional "rounds" of sequencing the template by virtue of the reduction in photo-induced damage to reaction components, as described elsewhere herein, thereby providing more complete and robust nucleotide sequence data for future analysis, e.g., sequence scaffold construction and/or consensus sequence determination. Further, as described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data

being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest, e.g., from the ends of the double-stranded portion of a SMRTbell™ template.

[00203] In addition to providing sense and antisense sequence data within a single template molecule that can be sequenced in one integrated process, the presence of the single-stranded linking portions also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., the sense strand, is completed and the other begins, e.g., the antisense strand. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. Additional aspects of and uses for registration sequences, e.g., for molecular redundant sequencing, are further described in U.S. Patent Publication No. 20090029385, which is incorporated herein by reference in its entirety for all purposes.

[00204] In certain embodiments, such a sequencing process begins by priming the template nucleic acid within one of the linking portions and allowing the polymerase to proceed along the strand of the double-stranded portion of the template that is immediately downstream of the primed linking portion when the double-stranded portion is melted or denatured. The sequence process proceeds around the second linking portion and proceeds along the complementary strand of the (now previously) double-stranded portion of the template. Because the template is circular, this process can continue to provide multiple repeated sequence reads from the one template. Thus, sequence redundancy comes from both the determination of complementary sequences (sense and antisense strands of the double-stranded portion), and the repeated sequencing of each circular template. The ongoing sequencing reaction is subjected to multiple illuminated and non-illuminated periods to generate at least two or more sequence reads per pass around the template. The illuminated periods are preferably timed to allow generation of nucleotide sequence data for selected regions of the template. For example, it may be beneficial to only generate nucleotide sequence data for the complementary strands of the double-stranded portion, or segments thereof. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing. Other approaches will

similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post-synthesis.

[00205] One may optionally employ various means for controlling initiation and/or progression of a sequencing reaction, and such means may include the addition of specific sequences or other moieties into the template nucleic acid, such as binding sites, e.g., for primers or proteins. Various methods of incorporating control elements into an analytical reaction, e.g. by integrating stop or pause points into a template, are discussed elsewhere herein and are further described in related application, U.S. Application No. 12/413,258, filed March 27, 2009, which is incorporated herein by reference in its entirety for all purposes.

[00206] In certain embodiments, a reaction stop or pause point may be included within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the linking portion that was not used in priming. By way of example, following initial sequencing from the original priming location, e.g., from the single-stranded linking portion used in priming synthesis through a first portion of the sense strand (e.g., the 3' end), the data acquisition may be switched off and the polymerase allowed to proceed around the template, e.g., through the remainder of the sense strand to the other linking portion. The incorporation of a synthesis blocking moiety coupled to this linking portion will allow control of reinitiation of the polymerase activity at the 3' end of the antisense strand. One would thereby obtain paired-end sequence data for the overall (previously) double-stranded segment, with sequence data from one end coming from the sense strand and sequence data from the other end coming from the antisense strand. This template construction and sequencing methodology is particularly useful in the case of long double-stranded segments, especially given the short read lengths generated by some sequencing technologies.

[00207] A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the nucleobase portion of one or more bases in the single-stranded portion that inhibit polymerase-mediated replication; strand-binding moieties that prevent processive synthesis; non-native nucleotides included within the primer and/or template; and the like. The use of strand-binding moieties includes, but is not limited to, reversible, specific binding of particular proteins to recognition sequences incorporated into the template (or primer bound thereto) for this purpose. In certain embodiments, such control sequences may include binding sites for transcription factors, e.g., repressor binding regions provided within the linking portion(s). For example, the lac repressor recognition sequence is bound by the lac repressor protein, and this

binding has been shown to block replication in a manner reversible by addition of appropriate initiators, such as isophenylthiogalactoside (IPTG) or allolactose.

[00208] In some embodiments, primer recognition sequences and/or additional control sequences may also be provided for control of initiation and/or progression of polymerization, e.g., through a hybridized probe or reversibly modified nucleotide, or the like. (See, e.g., U.S. Patent Application No. 2008-0009007, the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Such probes include but are not limited to probes at which a polymerase initiates polymerization, probes containing various types of detectable labels, molecular beacons, TaqMan® probes, Invader® probes (Third Wave Technologies, Inc.), or the like, that can be used for various purposes, e.g., to provide indications of the commencement and/or progress of synthesis.

[00209] An engineered pause point (reversible or irreversible) can include one or more non-native (non-natural) or fifth bases that do not pair with any of the four native nucleoside polyphosphates in the synthesis reaction, e.g., in the template and/or oligonucleotides probe(s), and/or that exhibit a distinct kinetic signature during template-dependent synthesis at such a base. Upon encountering such a base, the polymerase pauses until the complement to the non-natural base is added to the reaction mixture. Likewise, an engineered pause point could include a "damaged" base that causes a stop in replication until repair enzymes are added to the mixture. For example, a template having a pyrimidine dimer would cause the replication complex to pause, and addition of the photolyase DNA repair enzyme would repair the problem location and allow replication, and sequencing to continue. In yet further embodiments, a combination of modification enzymes could be used to engineer a set of modified bases on a template, e.g., a combination of glycosylases, methylases, nucleases, and the like. (Further information on sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.)

[00210] As noted elsewhere herein, stop or pause points can be engineered into various portions of the template, e.g., portions for which the nucleotide sequence is unknown (e.g., a genomic fragment) or known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, SMRTbell™ templates are topologically closed, single-stranded molecules having regions

of internal complementarity separated by hairpin or stem-loop linkers, such that hybridization of the regions of internal complementarity produces a double-stranded portion within the template. One or both of the linkers can comprise a stop or pause point to modulate polymerase activity. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00211] In other embodiments, an abasic site is used as a synthesis blocking moiety or pause point until addition of a non-natural "base," such as a pyrene, which has been shown to "base-pair" with an abasic site during DNA synthesis. (See, e.g., Matray, et al. (1999) Nature 399(6737):704-8, which is incorporated herein by reference in its entirety for all purposes.) Where a permanent termination of sequencing is desired, no non-natural analog is added and the polymerase is permanently blocked at the abasic site. DNA (or RNA) glycosylases create abasic sites that are quite different from the normal coding bases, A, T, G, and C (and U in RNA). A wide variety of monofunctional and bifunctional DNA glycosylases that have specificity for most common DNA or RNA adducts, including 5-methylcytosine, are known in the art, with different glycosylases capable of recognizing different types of modified DNA and/or RNA bases. The molecular structures of many glycosylases have been solved, and based on structural similarity they are grouped into four superfamilies. The UDG and AAG families contain small, compact glycosylases, whereas the MutM/Fpg and HhH-GPD families comprise larger enzymes with multiple domains. As an example, four enzymes have been identified in *Arabidopsis thaliana* in the plant pathway for cytosine demethylation. Additionally, other enzymes are also known to recognize 5-methyl cytosine and remove the methylated base to create an abasic site. Further, various enzymes are known to methylate cytosine in a sequence-specific manner. As such, a combination of a cytosine-methylase and an enzyme that creates an abasic site from a methylated cytosine nucleotide can be used to create one or more abasic sites in a template nucleic acid. The size of the recognition site of the methylase and the base composition of the template determine how frequently methylation occurs, and therefore, the number of abasic sites created in a given template nucleic acid, allowing the ordinary practioner to choose a methylase with a recognition site that produces a desired spacing between modified nucleotides. For example, if the recognition site is three bases long, then on average an abasic site is expected every 64 bases; if the recognition site is four bases long, then on

average an abasic site is expected every 256 bases; if the recognition site is six bases long, then on average an abasic site is expected every 4096 bases; and so forth. Of course, templates with a higher GC content would be expected to have more frequent abasic site formation, and templates with lower GC content would be expected to have less frequent abasic site formation.

[00212] Uracil-DNA glycosylases can also be used to introduce abasic sites into a template nucleic acid comprising deoxyuridine nucleotides. This strategy has the advantage of allowing the practitioner to choose the locations of the abasic sites within a DNA template since deoxyuridine nucleotides are not generally found in DNA. Various methods of inserting deoxyuridine nucleotides into a DNA template may be used, and different methods will be preferred for different applications. In certain embodiments, one or more site-specific deoxyuracils are incorporated during standard phosphoramidite oligonucleotide synthesis. To place uracils at indeterminate positions in a DNA, replacing a portion of the deoxythymidine triphosphate with deoxyuridine triphosphate will result in an amplimer with random U sites in place of T sites after polymerase chain reaction. In other embodiments, deoxyuridine nucleotides are engineered into the template, e.g., by ligation of a synthetic linker or adaptor comprising one or more deoxyuridine nucleotides to a nucleic acid sequence to be sequenced. In certain preferred embodiments, deoxyuridine nucleotides are incorporated into the linker portions of a SMRTbell™ template.

[00213] To subsequently introduce abasic sites prior to sequencing, the deoxyuridine nucleotide-containing template is subjected to treatment with uracil-DNA glycosylase, which removes the one or more uracil bases from the deoxyuridine nucleotides, thereby generating one or more abasic sites in the template. Alternatively, since the deoxyuridine nucleotide can be recognized as a template base and paired with deoxyadenosine during template-dependent nascent strand synthesis, the synthesis-blocking abasic site can instead be introduced after initiation of the sequencing reaction, e.g., at a time chosen by the practitioner. For example, the reaction can be initiated with a deoxyuridine-containing template, and uracil-DNA glycosylase can subsequently be added to block the polymerase and halt the reaction after the reaction has proceeded for a given time. As such, termination of the reaction is optional rather than required.

[00214] While uracil-DNA glycosylase activity is useful for introducing abasic sites into a template as described above, this activity can be problematic during the preparation of such templates. As such, strategies are typically implemented during preparation and manipulation of uracil-containing DNA, e.g., using molecular biology enzymes, to avoid uracil-DNA glycosylase activity, in particular, due to the *E. coli* UDG enzyme. Since a majority of standard molecular

biology enzymes are overexpressed and subsequently purified from an *E. coli* host, UDG activity can be a contaminating activity that is often not monitored by the enzyme manufacturer's quality control procedures. To mitigate contaminating UDG activity, a commercially available UDG inhibitor, also known as uracil glycosylase inhibitor or UGI (e.g., from New England Biolabs, Ipswich, MA) can be included in molecular biology reactions. This is a small protein inhibitor from the *B. subtilis* bacteriophase PBS1 that binds reversibly to E. coli UDG to inhibit its catalytic activity. UGI is also capable of dissociating UDG from a DNA molecule. Alternatively, UDG activity can be inhibited without exogenous protein using a chemical inhibitor of the enzyme, such as an oligonucleotide containing a 1-aza-deoxyribose base, a transition state analog for the UDG enzyme. This and other cationic nitrogenous sugars have been used for mechanistic studies of UDG activity and show potent inhibition activity. (See, e.g., Jiang et al. Biochemistry, 2002, 41 (22), pp 7116–7124.)

[00215] In certain applications, UDG activity needs to be inhibited temporarily, and subsequently enabled to remove create an abasic site as described above. In some embodiments, a DNA purification that removes proteins is employed, e.g., including a phenol-chloroform extraction with subsequent ethanol precipitation, a silica-based column approach (e.g., QiaQuick columns from Qiagen and similar products), and/or a PEG/sodium chloride precipitation (e.g., AMPure beads from Beckman Coulter). Alternatively or additionally, a commercially-available UDG enzyme that is not inhibited by UGI is added when abasic site formation is desired. For example, the *A. fulgidus* UDG is from a thermophilic organism and cannot be inhibited by the same bacteriophage protein as is the *E. coli* UDG enzyme. In certain preferred embodiments, UDG-inhibition is employed during template preparation, and inhibition-resistant UDG activity is added at a subsequent time to trigger the creation of abasic sites at deoxyuridine nucleotides, e.g., immediately prior to or during an ongoing reaction.

[00216] In some preferred embodiments, one or more abasic sites are engineered into a linker or adapter sequence within a sequencing template molecule. Abasic sugar residues serve as efficient terminators of polymerization for many polymerases, e.g., Φ29. 1',2'-dideoxyribose is the most common synthetic "abasic site". In other embodiments, a synthetic linker is incorporated into a linker or adaptor. For example, an internal spacer (e.g., Spacer 3 from Biosearch Technologies, Inc.) or other carbon-based linker can be used in lieu of a sugar-base nucleotide. Similar to an abasic nucleotide, the polymerase will be blocked upon encountering these moieties in the template nucleic acid.

[00217] In certain embodiments, synthesis blocking moieties are nicks in the template nucleic acid. Nicking enzymes (e.g., nicking endonucleases) are known in the art and can be used to specifically nick the template prior to or during a template-directed sequencing reaction. The use of site-specific nicking endonucleases allows the practitioner to incorporate a recognition sequence at a particular location within the template nucleic acid, and such nicking endonucleases are commercially available, e.g., from New England Biolabs, Inc. For example, a linker or adapter can be synthesized with a nicking endonuclease recognition sequence, ligated to a nucleic acid molecule to be sequenced, and can be specifically nicked either before or during a subsequent sequencing reaction. Nicks can also be introduced by ligating duplex segments that lack either a terminal 3'-hydroxy (e.g., have a dideoxynucleotide at the 3'-terminius) and/or 5'-phosphate group on one strand. The ligation results in covalent linkage of the phosphodiester backbone on one strand, but not on the other, which is therefore effectively "nicked." In certain embodiments, a SMRTbell™ template is constructed using a duplex (or "insert") nucleic acid molecule lacking a 5'-phosphate group at one or both termini. Upon ligation of the hairpin or stem-loop adaptors at each end, nicks are created at one or both ligation site(s), depending on whether the duplex lacked a 5'-phosphate at one or both ends, respectively. In other embodiments, a SMRTbell™ template is constructed using one or two stem-loop adaptors lacking a 3'-hydroxy group at the terminus (e.g., comprising a 2',3'-dideoxynucleotide rather than a 2'-deoxynucleotide). Upon ligation of one or two stem-loop adaptors lacking a 3'-hydroxy group, one or two nicks are created at the ligation site(s), depending on whether one or two adaptors lacked the 3'-hydroxy group, respectively. In both cases, a nick is created in the template nucleic acid, and a primer bound to one of the adaptors provides an initiation site for the polymerase, which will process the template until encountering a nick, at which point the polymerase will terminate the reaction by dissociation from the template. Regardless of how a nick is created, the position of a nick relative to the initiation site for the polymerase determines how much of the template will be sequenced. For example, Figure 19A provides an illustrative example of an embodiment in which a nick is present on a first strand of a duplex portion at a position distal to the adaptor containing the primer binding site. The first strand is processed by a polymerase, but the complementary strand is not processed because the polymerase dissociates at the nick site. An alternative embodiment is shown in Figure 19B, in which a nick is present on the strand complementary to the first strand at a position proximal to the adaptor containing the primer binding site. In this case both the first and complementary strands, as well as the adaptor not containing the primer binding site, are processed by the polymerase prior to dissociation. The

position of the primer binding site also determines how much of the template is processed by the polymerase. Figure 19C provides a template having a primer binding site at a position from which a polymerase would process a significant portion of the adaptor prior to entering the duplex portion. An additional advantage to using a 3'-dideoxynucleotide at a nick is that it prevents the use of the nick as a polymerse initiation site, since strand extension requires a 3-hydroxy group. As such, the resulting nick would not compete with a primer site for initiation of nascent strand synthesis by the polymerase. Having a single, known site of initiation on a template molecule is beneficial, e.g., for subsequent mapping of a read generated in such a reaction. In certain preferred embodiments, a nick site both lacks a 5'-phosphate group and comprises a 3'-dideoxynucleotide.

[00218] In certain preferred embodiments, modification and base excision is performed prior to introduction of a template nucleic acid to a reaction site, e.g., a zero-mode waveguide. As noted above, the choice of recognition site for the methylase depends on how far apart the practitioner wishes point of synthesis initiation to be on the template. For example, after initiating the template-dependent sequencing reaction, the sequence of nucleotide incorporations into the nascent strand is monitored for a desired sequence read, which may extend from the initiation point to the pause point, or may end before the polymerase reaches the pause point. In some preferred embodiments, as described elsewhere herein, the monitoring is suspended by modifying or removing an illumination source, e.g., by moving the illumination source or a substrate comprising the reaction site. Synthesis of the nascent strand will continue until the pause site is reached, whether or not the reaction is being actively monitored. When the reaction is to be reinitiated, reaction components are added that allow bypass, e.g., pyrene, polymerase, etc., and these can be subsequently removed (e.g., by buffer exchange) to allow additional pauses at other pause sites on the template.

[00219] In certain embodiments using pyrosequencing-based technologies (e.g., as developed by 454 Life Sciences), abasic sites can be introduced into a set of amplified template nucleic acids and synthesis initiated. Since all templates in the set are identical, they will comprise the same number of abasic sites in the same positions. During the course of the synthesis reaction, the synchronous incorporation of nucleotides into the nascent strands is monitored until either an abasic site is reached (at which point the synthesis is paused) or until the incorporation becomes asynchronous (which increases the background noise and decreases reliability of the sequence read). In the latter case, the practioner may opt to speed up the reaction, e.g., by adding all nucleotides at one time, to extend all nascent strands to the first abasic site in the templates. When synthesis is to be reinitiated, reaction components are added that allow bypass of the abasic site, e.g. one or more

pyrenes. A wash step may be performed to remove nucleotides and/or polymerases from the reaction sites prior to such addition. Further, in some cases, a different polymerase may be used for pyrene incorporation as is used for sequencing-by-synthesis reactions. In certain preferred embodiments, the reaction mixture comprising the pyrene for abasic site bypass allows readthrough of the abasic site, but no further on the template. Subsequent addition of sequencing reaction mixture allows the sequencing-by-synthesis reaction to recommence and incorporation of nucleotides into the nascent strand to be monitored. Alternatively or additionally, the practioner need not wait until an abasic site is reached to suspend detection and, optionally, speed up the reaction to bring all nascent strands to a given abasic site, but can choose to do this before a reaction has become asynchronous, e.g., after desired sequence data has been collected for a particular region of interest in a template nucleic acid.

[00220] In certain embodiments using ligation-based technologies (e.g., the SOLiD™ System developed by Life Technologies), a pause site can be engineered by using an oligonucleotide that cannot participate in the ligation reaction and that is complementary to a desired location on the set of identical template nucleic acids, e.g., on a bead. When the serial ligation reaction hits the position recognized by this polynucleotide, the reaction cannot proceed and any reactions that have become asynchronous will "catch up." The user can then unblock the oligo (e.g., using chemical treatment or photo-cleavage) and reinitiate the sequencing reaction.

[00221] In some cases, it may be desirable to provide endonuclease recognition sites within the template nucleic acid. For example, inclusion of such sites within a circular template can allow for a mechanism to release the template from a synthesis reaction, i.e., by linearizing it, and allowing the polymerase to run off the linear template, and/or to expose the template to exonuclease activity, and thus terminate synthesis through removal of the template. Such sites could additionally be exploited as control sequences by providing specific binding locations for endonucleases engineered to lack cleavage activity, but retain sequence specific binding, and could therefore be used to block progression of the polymerase enzyme on a template nucleic acid.

[00222] In some cases, nicking sites, e.g., sites recognized by nicking endonucleases, may be included within a portion of the template molecule, and particularly within a double-stranded portion of the template, e.g., in a double-stranded segment of a SMRT bell™ or in the stem portion of an exogenous hairpin structure. Such nicking sites provide one or more breaks in one strand of a double-stranded sequence and can thereby provide one or more priming locations for, e.g., a strand-displacing polymerase enzyme. A variety of nicking enzymes and their recognition sequences are

known in the art, with such enzymes being generally commercially available, e.g., from New England Biolabs.

[00223] In certain embodiments, methods for intermittent detection described herein are useful in "paired-end" sequencing applications in which sequence information is generated from two ends of a template nucleic acid but not for at least a portion of the intervening portion of the template. Typically, paired-end sequencing applications provide sequence data for only the two ends of a nucleic acid template, but the present invention also allows generation of additional sequence reads that are noncontiguous with the sequence reads from the ends of the template. In certain preferred embodiments, a duplex fragment (e.g., genomic fragment) is ligated to a single-stranded linker that connects the 3' end of the sense strand to the 5' end of the antisense strand, or that connects the 5' end of the sense strand to the 3' end of the antisense strand. In either orientation, separation of the two strands of the duplex fragment results in a single-stranded linear template nucleic acid that contains the linker in between the sense and antisense strands. Subsequent sequencing can involve intermittent detection that generates sequence reads for only the portions of the sense and antisense strands that are of interest, e.g., one or both of the ends. In certain embodiments, both sense and antisense strands may be sequenced at both ends to provide redundancy in the sequence data. Sequence reads recognized as being from the linker portion of the template (e.g., based on the known linker sequence or specific registration sequences encoded therein) can be used to orient the alignment of the sequence reads from the sense and antisense portions of the template, providing context for determining the sequences of the ends of the duplex fragment and subsequent sequence scaffold construction and/or mapping. In certain embodiments, pause or stop points may be incorporated into the linker to control the processing of the template by the polymerase, and therefore may be used to synchronize the detection periods to ensure generation of sequence reads from particular regions of template. Further, additional detection periods can be included that are timed to provide sequence reads from portions of the sense and/or antisense strand that are noncontiguous with the end regions.

[00224] In a related embodiment, paired-end sequencing may be accomplished by using a nucleic acid template that has linkers connecting the sense and antisense strands of a duplex fragment at both ends, such that separation of the strands of the duplex fragment provides a single-stranded circular template that contains a linkers in between each end of the sense and antisense strands of the original duplex fragment. Such a template molecule would allow a strand-displacing polymerase to proceed around the template multiple times, thereby potentially generating redundant

sequence data from both ends of both strands of the original duplex fragment. As noted elsewhere herein, such redundancy is useful for determination of consensus sequences and/or construction of sequence scaffolds. As the polymerase enzyme processes the template, detection periods can be timed (e.g., based on knowledge of the rate at which the polymerase processes the template, which is dependent not only on the polymerase but also on the sequence of the template itself) to generate nucleotide sequence reads from the regions of the template corresponding to one or both ends of the sense and antisense strands, and can also include detection periods to generate additional reads from other, noncontiguous regions of the duplex fragment, as well. Although such timing can be used to determine the appropriate periodicity of the detection periods, at later stages of the reaction (e.g., as the polymerase repeatedly proceeds around the template), the exact location of reinitiation of sequence read generation becomes more approximate. Incorporation of pause or stop points into one or both linkers to regulate the processing of the template by the polymerase may be used to synchronize the detection periods regardless of the total distance travelled by the polymerase around the template. This strategy more reliably ensures generation of sequence reads from selected regions of template, e.g, the ends of the sense and antisense portions and, optionally, regions in between and noncontiguous with the end regions regardless of the number of passes of the polymerase around the template nucleic acid, especially in later stages of the reaction. Further, the known sequence of one or both of the linkers can be used to orient sequence reads from the sense and antisense portions for consensus sequence determination and/or mapping.

[00225] In some such embodiments, a duplex fragment inserted between two hairpin linkers may be much larger than desired, increasing the difficulty of limiting nucleotide sequence read data to particular regions of the fragment. The size of the duplex fragment ligated to the two hairpin linkers can be selectively reduced to retain the regions attached to the linkers and to lose a central portion of the duplex fragment. One particularly preferred strategy, illustrated in Figure 18, comprises hairpin linkers (1802, 1804) having a regions of cross-complementarity (1806, 1808), such that the two linkers 1802 and 1804 can anneal to each other in a manner that does not interfere with ligation to a duplex fragment 1810. Duplex fragment 1810 comprises ends 1812 and 1814, as well as a long central region 1816, which is not shown but is understood to be between the two curvy lines. Once end 1812 is ligated to linker 1802 and end 1814 is ligated to linker 1804, the construct is subjected to fragmentation, which removes the central region 1816 of the duplex fragment 1810, producing construct 1818 having ends 1820 and 1822. After fragmentation, the ends of the portions of the duplex fragment still associated with the annealed linker pair (ends 1820 and

1822) are ligated together to produce construct 1824, which can then be treated (e.g., with heat, gentle denaturation, primer invasion, changing salt concentration, etc.) to separate cross-complementary regions 1806 and 1808 from one another, e.g., to generate a circular single-stranded nucleic acid molecule. Alternatively, the separation may occur during the course of the subsequent reaction, e.g., by polymerase-mediated strand displacement. Yet further, where the hybridized cross-complementary regions are long enough to undergo a complete DNA turn, an additional reaction component (e.g., helicase, topoisomerase, polymerase, etc.) may be needed to unwind the duplex and allow separation. As such, the resulting "mate-pair" construct has only the ends of the original duplex fragment ligated together and capped with adaptors that link the 5' end of each strand of the duplex with the 3' end of the other strand of the duplex, and denaturation of the duplex produces a closed, single-stranded circular construct.

[00226] Fragmentation of the duplex fragment can be performed by a variety of known methods. For example, fragmentation can be performed enzymatically (e.g., using restriction enzymes or other nucleases) or mechanically, by shearing or sonication. The type of fragmentation chosen will determine various characteristics of the resulting construct, e.g., how large a central region is removed and the types of ends remaining (e.g., blunt, 5' overhang, 3' overhang, random, identical on both ends, etc.). Optionally, the ends can be modified after fragmentation to facilitate the subsequent ligation step. Although not shown in Figure 18, it is expected that the ligation of the duplex fragment to the hybridized linkers will be a two-step process, with one end being ligated first and unimolecular kinetics favoring ligation of the second end to the second linker. The cross-complementary regions of the linkers can be designed to produce varying levels of complementarity, and therefore varying strengths of the hybridization. For example, a longer or higher GC content in a cross-complementary region lends a higher stability to the linker:linker interaction, but separation of the hybridized linkers requires a more severe treatment, e.g., higher temperature, more stringent conditions, etc. As such the cross-complementary regions should be engineered to produce a stable linker:linker interaction that is disruptable under conditions that are not destructive to the overall construct. Further the linkers can vary in regions apart from the cross-complementary regions For example, one linker can have a primer binding site that the other lacks, which would provide a single polymerase initiation site in the final construct. Other sequence characteristics described herein (e.g., pause sites, registrations sequences, etc.) can also be included in one or both linker regions. If topological constraints limit the subsequent processing of the resulting contruct, e.g., during template-directed nascent strand synthesis, these can be addressed by

addition of a reaction component (e.g., a helicase or topoisomerase) to resolve the topological constraint. As such, the methods can be used to add asymmetric linkers to duplex polynucleotides, whether or not the duplex is to be selectively reduced in size, or not, as long as the asymmetric linkers can cross-hybridize to one another.

[00227] Although in preferred embodiments, the two linkers to be ligated to a single duplex fragment are hybridized to one another prior to ligation, in some embodiments they are instead hybridized after the initial ligation reaction, and where topological constraints inhibit such a post-ligation hybridization a reaction component (e.g., topoisomerase) may be included to relieve such constraints. In certain embodiments, the hybridized linkers are separated prior to addition of reaction components for a subsequent reaction, and in other embodiments the hybridized linkers are not separated until after the addition of reaction components for a subsequent reaction. For example, a polymerase enzyme may bind to a primer annealed to a linker before or after separation of the linker from a second linker. In fact, it may be benefical in some embodiments to postpone separation of the linkers, e.g., where compaction of the nucleic acid construct is beneficial, such as when the construct must be loaded into a confinement of some kind, e.g., a nanowell, optical confinement, etc.

[00228] In some embodiments, the methods further include separation of single linker constructs from hybridized linker pair constructs. This can be accomplished by an exonuclease treatment after ligation of the duplex fragment to the linkers, which would degrade any constructs having an unannealed end. Alternatively, it may be desirable to remove the single linkers prior to ligation, for example using a size separation methodology or by allowing them to bind to oligonucleotides that are complementary to the cross-complementary regions and bound to a column or magnetic beads. (The cross-complementary regions of the hybridized linker pairs will not be available for binding to the oligonucleotides. Other methods known in the art can also be used to separate single linkers from hybridized linker pairs.

[00229] Interestingly, the use of the sense/antisense nucleic acid templates described above would represent a unidirectional processing of a template to provide paired-end sequence data, as opposed to the more traditional bi-directional processing of a linear template molecule. Further, unlike traditional approaches, these methods for paired-end sequencing involve processing, chemically or otherwise, of not just the regions at the ends, but also regions in between the ends, and in some embodiments comprising processing of the entire template. For example, a polymerase incorporates nucleotides into a nascent strand for each position of the template (thereby

"processing" each position of the template), yet the sequencing data generated is limited to specific regions of the template that are of particular interest to the practitioner, such as the end regions. As such, in certain embodiments the duplex fragment is not further reduced in size after ligation to a linker pair, and the entire duplex fragment is processed by the polymerase.

[00230] In certain embodiments, methods for intermittent detection described herein are useful in analysis systems that employ nanopores. A nanopore is a small pore in an electrically insulating membrane that can be used for single molecule detection. In general, a nanopore functions as a Coulter counter for much smaller particles, and can take various forms, e.g., a protein channel in a lipid bilayer or a pore in a solid-state membrane. The detection principal is based on monitoring the ionic current of an electrolyte solution passing through the nanopore as a voltage is applied across the membrane. For example, passage of a polynucleotide molecule (e.g., DNA, RNA, etc.) through a nanopore causes changes in the magnitude of the current through the nanopore, with each nucleotide obstructing the nanopore to a different, characteristic degree. As such, the pattern of variations in the current passing through the nanopore as the polynucleotide is drawn through may be monitored and analyzed to determine the nucleotide sequence of the polynucleotide. A polynucleotide may be drawn through the nanopore by various means, e.g., by electrophoresis, or using enzyme chaperones to guide the polynucleotide through the nanopore. For additional discussion of methods of fabrication and use of nanopores, see, e.g., U.S. Patent No. 5,795,782; Kasianowicz, J.J., et al. (1996) *Proc Natl Acad Sci USA* 93(24):13770-3; Ashkenas, N., et al. (2005) *Angew Chem Int Ed Engl* 44(9):1401-4; Winters-Hilt, S., et al. (2003) *Biophys J* 84:967-76; Astier, Y., et al. (2006) *J Am Chem Soc* 128(5):1705-10; Fologea, D., et al. (2005) *Nano Lett* 5(10):1905-9; Deamer, D.W., et al. (2000) *Trends Biotechnol* 18(4):147-51; and Church, G.M. (2006) *Scientific American* 294(1):52, all of which are incorporated by reference herein in their entireties for all purposes. In some embodiments, intermittent detection of nucleic acid sequence data from a nanopore may be achieved by modifying the progress of the polynucleotide through the nanopore so that progress is sped up during non-detection periods and progress is slowed to allow sequence determination during detection periods. The rate of passage of the polynucleotide through the nanopore may be modified by various methods, including but not limited to increasing an electrophoretic field carrying the polynucleotide (e.g., by increasing the voltage, changing the conductivity of the reaction mixture, and the like), or changing various reaction conditions to alter the speed at which a protein chaperone carries the polynucleotide. Further, in embodiments utilizing

a processive exonuclease to feed individual bases through the nanopore, the kinetics of the exonuclease may be modified based on the known biochemical characteristics of the exonuclease.

[00231] In diagnostic sequencing applications, it may be necessary only to provide sequence data for a small fragment of DNA, but do so in an extremely accurate sequencing process. For such applications, shorter target segments may be employed, thus permitting a higher level of redundancy by sequencing multiple times around a smaller circular template, where such redundancy provides the desired accuracy. Thus, in some cases, the double stranded target segment may be much shorter, e.g., from 10 to 200, from 20 to 100 or from 20 to 50 or from 20 to 75 bases in length. For purposes of the foregoing, the length of the target segment in terms of bases denotes the length of one strand of the double stranded segment. In such applications, various methods for intermittent detection described herein may be used to analyze the sequence of the template, thereby targeting the sequence data to the portion(s) of the template of particular interest to the diagnostician, and/or improving various aspects of the reaction performance, e.g., by virtue of the reduction of photo-induced damage to one or more reaction components.

[00232] It is to be understood that the above description is intended to be illustrative and not restrictive. It readily should be apparent to one skilled in the art that various embodiments and modifications may be made to the invention disclosed in this application, including but not limited to combinations of various aspects of the invention, without departing from the scope and spirit of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. All publications mentioned herein are cited for the purpose of describing and disclosing reagents, methodologies and concepts that may be used in connection with the present invention. Nothing herein is to be construed as an admission that these references are prior art in relation to the inventions described herein. Throughout the disclosure various patents, patent applications and publications are referenced. Unless otherwise indicated, each is incorporated by reference in its entirety for all purposes.

[00233] Although described in some detail for purposes of illustration, it will be readily appreciated that a number of variations known or appreciated by those of skill in the art may be practiced within the scope of present invention. Unless otherwise clear from the context or expressly stated, any concentration values provided herein are generally given in terms of admixture values or percentages without regard to any conversion that occurs upon or following addition of

the particular component of the mixture. To the extent not already expressly incorporated herein, all published references and patent documents referred to in this disclosure are incorporated herein by reference in their entirety for all purposes.

[00234] The following non-limiting examples are provided to further illustrate the invention.

VI. Examples of Intermittent Illumination of a Single Molecule Sequencing-by-Synthesis Reaction

Example I

[00235] A nucleic acid template was provided that comprised a double-stranded region and two single-stranded linker portions at each end. The first linker portion connected the 3' end of the sense strand with the 5' end of the antisense strand, and the second linker portion connected the 3' end of the antisense strand with the 5' end of the sense strand. This template was designed to form a single-stranded circle of approximately 500 bases when the double-stranded region was opened (e.g., by heat denaturization, helicase activity, etc.), and is sometimes referred to as a SMRTbell™ template. A plurality of this nucleic acid template was incubated with polymerases, primers, and other reaction components to allow formation of polymerase-template complexes. (See, e.g., Korlach, J., et al. (2008) Nucleosides, Nucleotides and Nucleic Acids, 27:1072-1083; and Eid, J. (2009) Science 323:133-138.) The complexes were immobilized in zero-mode waveguides in a reaction mixture containing all necessary buffer and nucleotide analog components for carrying out sequencing-by-synthesis reactions with the exception of a cognate starting base and a metal dication. A Smith-Waterman algorithm was used to perform the alignment of the known sequence of the template with the sequence reads generated in the reaction, and the positions of the sequence reads is graphically illustrated in Figure 8.

[00236] Acquisition of the data shown in Figure 8 was collected as follows. Illumination of the array of zero-mode waveguides was initiated with laser excitation (532 nm and 641 nm laser lines) at t = -5 seconds, and the missing cognate starting base and metal dication (manganese metal) were added at t = 0 seconds to simultaneously initiate the sequencing-by-synthesis reactions in all zero-mode waveguides. The reactions were monitored under illumination for 120 seconds at which time the illumination was removed; the sequencing reads generated during that stage of the reaction are shown in Figure 8A as a function of the template position to which each read maps. At 295 seconds illumination was resumed and data acquisition was reinitiated at 300 seconds and maintained for another 120 second interval; the sequencing reads during this second illuminated

period are shown in Figure 8B. At 595 seconds illumination was resumed and data acquisition was reinitiated at 600 seconds and maintained for another 120 second interval; the sequencing reads during this third illuminated period are shown in Figure 8C.

[00237] As expected, the longer the amount of time before the sequence data is collected (that is, the later the illuminated period), the further into the template the alignments shift, and this shift is a rough function of time since initiation of the reaction. Further, the distribution of sequence reads generated during each subsequent illuminated period becomes more dispersed than the previous illuminated period(s). Further, due to the circular nature of the template, Figure 8C clearly shows that some polymerases have passed completely around the substrate and are beginning to generate sequence reads from a second pass around the template, thereby generating redundant sequence information for a single template nucleic acid.

Example II

[00238] As in Example I, a SMRTbell™ template was used. For templates of defined sequence, PCR was used to generate 3 or 6 kb DNA inserts for the double-stranded region in the SMRTbell™ templates using a standard PCR methodology. For genomic and other biological samples, a DNA fragmentation protocol was used that generates DNA fragments distributed around 3 or 6 kb. Generation of fragments in these ranges was done using a HydroShear® (Genomic Solutions®) device with settings recommended by the manufacturer. The random genomic DNA fragments were enzymatically treated to generate blunt ends. Both the PCR products and randomly generated DNA fragments were phosphorylated and then immediately put into a ligation reaction with a blunt hairpin adapter. The products were purified through two size selection steps using reduced volumes of AMPure® magnetic beads (Agencourt®) to remove hairpin dimers and other short products. (Fabrication of SMRTbell™ templates is further described elsewhere herein.)

[00239] The system components used for polynucleotide sequencing using intermittent detection are comparable to single-molecule sequencing applications under constant illumination, which are described, e.g., in Eid, et al. (2009) Science 323:133-138. Specifically, the immobilization and sequencing buffer compositions, nucleotide analogs identity and concentration, polymerase, ZMWs, surface treatment and instrumentation were identical to the standard methodology. Modifications to the SMRTbell™ template DNA and polymerase binding and immobilization and data acquisition protocols are as follows.

[00240] A binding solution was prepared by incubation of 3 or 6 kb DNA SMRTbell™ templates (1-10 nM) with a 10-fold excess of DNA polymerase (10-100 nM, respectively) in 10 mM MOPS (pH 7.5), 10 mM KOAc, 100 mM DTT & 0.05% Tween-20 for 2 hours at 30°C, followed by 1 hour at 37°C and subsequent storage at 4°C prior to immobilization on the ZMWs. Immediately prior to immobilization, the binding solution was diluted in the standard immobilization solution (50 mM MOPS (pH 7.5), 75 mM KOAc, 5 mM DTT, 0.05% Tween-20) to the desired final concentration, typically 0.1 to 1 nM, and incubated for 30 to 60 minutes at 22°C. Post-immobilization chip preparation and sequencing initiation were identical to the standard methods.

[00241] The data acquisition protocol was similar to the standard application with coordinated modifications to the collection timing and ZMW positioning. In the standard acquisition procedure, a single long acquisition (~10 minutes) is performed for each ZMW. In the intermittent illumination acquisition procedure, multiple short acquisitions (~3 minutes) of sequence reads (also termed "strobe reads") were performed for each ZMW (during "detection periods") with an interval between each acquisition period during which no acquisition of sequence reads was performed ("non-detection period"). The duration of the interval between each acquisition of sequence reads was determined based upon a desired distance (i.e., number of nucleotide positions) between each sequence (or strobe) read, the polymerization rate of the polymerase, and the SMRTbell™ template insert size.

[00242] SMRTbell™ templates were generated as described above for AC223433, a fosmid clone comprising a sequence of an approximately 40 kb region of *Homo sapiens* chromosome 15. The reference sequences used to map the sequence reads generated in the sequencing reactions were the publically available sequences of *Homo sapiens* chromosome 15 (Hg18; NCBI Build 36.1) and fosmid AC223433 (NCBI GenBank accession number). Table 1 shows the number of statistically significantly mapped sequence reads for several types of intermittent illumination sequencing reactions. The number of mappable "looks" is equivalent to the number of mappable sequence reads generated during detection periods for a single template molecule. For example, a "mapped 1-look read" means, for a single template molecule, only a single detection period generated a sequence read that could be mapped to the reference sequence.

Table 1: Summary of Sequencing Results

Mapping	Mapped 1-	Mapped 2-	Mapped 3-	Mapped 4-

Reference	look reads	look reads	look reads	look reads
Human chr15 (Hg18)	13834	1289	127	4
Fosmid	15253	1571	158	5

[00243] Deviations in the expected time span for a set of sequencing reads from a single sequencing reaction are indicative of genomic events such as genomic rearrangements, e.g., insertions, deletions, etc. Figures 10 and 11 illustrate this point. Specifically, the time and distance travelled along the template (based upon the reference sequence) by the polymerase was computed within and between the sequence reads generated during the detection (illuminated) periods. These calculations were used to detect unexpected variations, indicating possible genomic events in the template as compared to the reference sequence. Figure 10 provides a plot that illustrates the normalized average time it took for the polymerase to traverse a region of the template based on the length of that region in the *Homo sapiens* chromosome 15 reference sequence. The sequence reads are fit to a diagonal having a slope equal to the average speed for sequencing reads. Deviations from the regressed diagonal indicate genomic events (for example, structural variants), and the slope of the sequence reads around such deviations indicate the relative size of the genomic event (e.g., in the case of insertions/deletions). For example, if the time for the polymerase to traverse a region was unexpectedly long, this indicated the polymerase actually traversed a longer region than was expected based on the reference sequence. The two distinct off-diagonal deviations (upper right hand corner) with higher slope indicated that an insertion had occurred in the reference sequence, and this was verified by comparison to the known fosmid sequence.

[00244] Figure 11 shows the average time it took the polymerase to traverse the template. For each mapped read, starting and ending times and positions were determined and used to compute the distance traversed by the polymerase between sequence reads. Based on these determinations, an average time across any particular region of the human reference sequence was computed. Regions that were traversed by the polymerase more slowly have peaks of higher ΔT, and were indicative of insertions in the template relative to the *Homo sapiens* chromosome 15 reference sequence. The insertions identified were the same insertions identified above.

[00245] Intermittent illumination-based sequencing reactions across fosmid sequence AC223433 showed significant sequence read coverage across the insertion events. The distribution of the physical coverage is shown in Figure 12, which illustrates examples of three-look strobes (i.e., sequencing reactions having three detection/illuminated periods) that span or intersect the

insertion events. Figure 12A shows the mapping of the strobe sequence reads to the *Homo sapiens* chromosome 15 reference sequence, where the sequence reads generated from the insert sequences in the template are excluded. Arrows indicate the locations of the insertions. Figure 12B shows a similar mapping with the sequence reads generated from the insert sequences indicated with brackets. A number of sequence reads flank the insertions, connect the two insertions, or clarify sequence within (or at the boundaries of) the insertion sequences. Such flanking and connecting sequence reads are useful for predicting and detecting genomic events, anchoring them to genomic references, and scaffolding for de novo assembly of novel sequences. In particular, there are 30 and 38 "3-look" reads that intersect the two regions of insertion of (1192 bp and 6879 bp, respectively). These sequence reads facilitated mapping of the insertions to the human reference sequence, which would have been extremely difficult, if not impossible, with commercially available short-read sequencing technologies. Further the sequence of the smaller insertion was a highly repetitive sequence, which would also have made mapping difficult with certain short-read technologies.

[00246] Figure 13 illustrates the sequence coverage obtained across the fosmid sequence, showing all two-, three-, and four-look strobe sequence reads spanning the sequence that are mappable to the known AC223433 fosmid sequence.

A consensus sequence was derived from the set of mappable sequence reads generated in these sequencing reactions. Strobe sequence reads were combined with sequence reads generated under constant illumination and assembled based on the human reference sequence (Hg18). High quality reads surrounding the (suspected) insertion sites, as well as high quality reads that did not map to the reference sequence, were extracted and assembled with a "de novo" greedy suffix tree assembler; the resulting contigs were mapped to the Hg18 reference sequence. Contigs spanning the (suspected) insertion sites were identified and fed back into the "de novo" assembler, and the resulting contigs were manually edited using standard techniques and placed back into the derived reference guided assembly. The final consensus sequence was a hybrid of a reference guided assembly and attempts at de novo assembly of novel insert sequences. Alignments to reference sequences were performed and plotted. Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described above and the fosmid reference sequence, and this plot confirmed a high degree of alignment between the two sequences. This dot plot was generated using Gepard 1.21 ("GEnome PAir – Rapid Dotter," available from the Munich Information Center for Protein Sequences (MIPS)) with a word size of 7. Nucleic acid dot plots are widely used in the art and are further described, e.g., in Krumsiek et al. (2007) Bioinformatics

23(8):1026-8; Maizel et al. (1981) Proc Natl Acad Sci USA 78:7665; Pustell, et al. (1982) Nucleic Acids Res 10:4765; and Quigley, et al. (1984) Nucleic Acids Res 12:347, all of which are incorporated herein by reference in their entireties for all purposes.

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation application of U.S. Patent Application No. 14/708,603, filed May 11, 2015, which is a continuation application of U.S. Patent Application No. 14/091,961, filed November 27, 2013, now U.S. Patent No. 9,057,102, which is a continuation application of U.S. Patent Application No. 12/982,029, filed December 30, 2010, now U.S. Patent No. 8,628,940, which (1) claims the benefit of U.S. Provisional Application No. 61/099,696, filed September 24, 2008; (2) claims the benefit of U.S. Provisional Application No. 61/139,402, filed December 19, 2008; and (3) is a continuation-in-part application of U.S. Patent Application No. 12/413,226, filed March 27, 2009, now U.S. Patent No. 8,143,030, the full disclosures of all of which are incorporated herein by reference in their entireties for all purposes.

[0002] This application is also related to U.S. Provisional Application No. 61/072,160, filed March 28, 2008, U.S. Patent Application No. 12/383,855, filed March 27, 2009, now U.S. Patent No. 8,236,499, and U.S. Patent Application No. 12/413,258, filed March 27, 2009, now U.S. Patent No. 8,153,375, all of which are incorporated herein by reference in their entireties for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0003] Not Applicable.

BACKGROUND OF THE INVENTION

[0004] The use of optically detectable labeling groups, and particularly those groups having high quantum yields, e.g., fluorescent or chemiluminescent groups, is ubiquitous throughout the fields of analytical chemistry, biochemistry, and biology. In particular, by providing a highly visible signal associated with a given reaction, one can better monitor that reaction as well as any potential effectors of that reaction. Such analyses are the basic tools of life science research in genomics, diagnostics, pharmaceutical research, and related fields.

[0005] Such analyses have generally been performed under conditions where the amounts of reactants are present far in excess of what is required for the reaction in question. The result of this excess is to provide ample detectability, as well as to compensate for any damage caused by the detection system and allow for signal detection with minimal impact on the reactants. For example, analyses based on fluorescent labeling groups generally require the use of an excitation radiation

source directed at the reaction mixture to excite the fluorescent labeling group, which is then separately detectable. However, one drawback to the use of optically detectable labeling groups is that prolonged exposure of chemical and biochemical reactants to such light sources, alone, or when in the presence of other components, e.g., the fluorescent groups, can damage such reactants. The traditional solution to this drawback is to have the reactants present so far in excess that the number of undamaged reactant molecules far outnumbers the damaged reactant molecules, thus minimizing or negating the effects of the photo-induced damage.

[0006] A variety of analytical techniques currently being explored deviate from the traditional techniques. In particular, many reactions are based on increasingly smaller amounts of reagents, e.g., in microfluidic or nanofluidic reaction vessels or channels, or in "single molecule" analyses. Such low reactant volumes are increasingly important in many high throughput applications, such as microarrays. The use of smaller reactant volumes offers challenges to the use of optical detection systems. When smaller reactant volumes are used, damage to reactants, such as from exposure to light sources for fluorescent detection, can become problematic and have a dramatic impact on the operation of a given analysis. In other cases, other reaction conditions may impact the processivity, rate, fidelity, or duration of the reaction, including salt or buffer conditions, pH, temperature, or even immobilization of reaction components within observable reaction regions. In many cases, the effects of these different reaction or environmental conditions can degrade the performance of the system over time. This can be particularly detrimental, for example, in real-time analysis of reactions that include fluorescent reagents that can expose multiple different reactions components to optical energy. In addition, smaller reactant volumes can lead to limitations in the amount of signal generated upon application of optical energy.

[0007] Further, in the case of sequencing-by-synthesis applications, an additional challenge has been to develop ways to effectively sequence noncontiguous portions of a template nucleic acid on a single molecule. This challenge is exacerbated in template nucleic acids that contain highly repetitive sequence and/or are hundreds or thousands of nucleotides in length, such as certain genomic DNA fragments. The difficulty in generating such noncontiguous reads from a single template has hampered efforts to construct consensus sequences for long templates, for example, in genome sequencing projects.

[0008] As such, methods and systems that result in enhanced reaction performance, such as an increase in processivity, rate, fidelity, or duration of a reaction of interest, would provide useful improvements to the methods and compositions currently available. For example, methods, devices,

and systems that increase reaction performance by, e.g., mitigating to some extent photo-induced damage in a reaction of interest and/or increasing various other performance metrics for the reaction would be particularly useful.

BRIEF SUMMARY OF THE INVENTION

[0009] In a general sense, the methods provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0010] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one

reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion. The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0011] In certain aspects, the present invention is generally directed to methods, devices, and systems for enhancing the performance of illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. In certain preferred embodiments, illuminated reactions comprise one or more fluorescent or fluorogenic reactants. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In some aspects, enhancing the performance of an illuminated reaction means increasing the processivity, rate, fidelity, and/or duration of the reaction. For example, enhancing the performance of an illuminated reaction can involve reducing or limiting the effects of photo-induced damage during the reaction. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction.

[0012] In certain aspects, methods of the invention useful for characterizing an analytical reaction comprise preparing a reaction mixture and initiating the analytical reaction therein, subjecting the reaction mixture to at least one detection period and at least one non-detection period during the course of the analytical reaction, collecting data during both the detection period(s) and the non-detection period(s), and combining the collected data to characterize the analytical reaction. In certain embodiments, the analytical reaction comprises an enzyme that exhibits an improvement in performance as compared to its performance in the analytical reaction under constant illumination, and such improvement may be related to various aspects of enzyme activity, e.g., processivity, fidelity, rate, duration of the analytical reaction, and the like. In certain embodiments, stop or pause points are used to control the activity of the enzyme, and such stop or pause points may comprise elements such as large photolabile groups, strand-binding moieties, non-native bases, and others well known in the art. In certain preferred embodiments, the one or more detection periods are illuminated periods and the one or more non-detection periods are non-illuminated periods. In certain preferred embodiments, a plurality of analytical reactions disposed on a solid support are characterized, preferably in a coordinated fashion as described elsewhere herein.

[0013] In certain preferred embodiments, the analytical reaction is a sequencing reaction that generates sequence reads from a single nucleic acid template during the detection period(s) but not during the non-detection period(s). For example, the analytical reaction can comprise at least two or more detection periods and can generate a plurality of noncontiguous reads from the single nucleic acid template. In some embodiments, the single nucleic acid template is at least 100 bases in length and/or comprises multiple repeat sequences. In certain embodiments, the sequencing reaction comprises passage of the single nucleic acid template through a nanopore, and in other embodiments the sequencing reaction comprises primer extension by a polymerase enzyme.

[0014] The analytical may optionally be a processive reaction monitored in real time, i.e., during the course of the processive reaction. In preferred embodiments, such a processive reaction is carried out by a processive enzyme that can repetitively execute its catalytic function, thereby completing multiple sequential steps of the reaction. For example, a processive polymerization reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.

[0015] In certain preferred embodiments, the analytical reaction comprises at least one component comprising a detectable label, e.g., a fluorescently labeled nucleotide. In certain embodiments, the labeled component is present throughout the course of the analytical reaction, i.e., during both the detection and the non-detection periods. The method may further comprise an optical system to collect the data during the detection period, but optionally not to collect the data during the non-detection period.

[0016] In certain aspects, methods of the invention comprise providing a substrate having a reaction mixture disposed thereon and illuminating the reaction mixture on the substrate with an excitation illumination for multiple, noncontiguous periods during the course of the reaction, thereby subjecting the reaction mixture to intermittent excitation illumination. In some embodiments, the reaction mixture comprises first reactant and a second reactant, wherein an amount of photo-induced damage to the first reactant occurs as a result of interaction between the first reactant and the second reactant under excitation illumination. In certain embodiments, the method further comprises monitoring a reaction between the first and second reactants during illumination and collecting the data generated therefrom. In some embodiments, the reaction is a primer extension reaction and/or the first reactant is a polymerase enzyme. In certain embodiments, the second reactant is a fluorogenic or fluorescent molecule.

[0017] In yet another aspect, the methods are useful for mitigating photo-induced damage in an illuminated reaction by subjecting the illuminated reaction to intermittent illumination rather than constant illumination. For example, certain methods of the invention monitor a reaction mixture comprising at least one enzyme and a fluorescent or fluorogenic substrate for the enzyme, wherein interaction of the enzyme and the substrate under excitation illumination can result in altered activity of the enzyme, e.g if such excitation illumination is present over an extended period of time. Such methods can comprise directing intermittent excitation illumination at a first observation region for a first period that is less than a photo-induced damage threshold period under the intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions. As such, certain aspects of the invention lengthen a photo-induced damage threshold period for an analytical reaction through intermittent inactivation of the excitation illumination source since the photo-induced damage threshold period under intermittent illumination is longer than the photo-induced damage threshold period under constant illumination.

[0018] In a related aspect, the invention also provides methods of performing an enzyme reaction, comprising providing an enzyme within a first observation region, contacting the enzyme with a fluorescent or fluorogenic substrate for the enzyme, and directing an excitation radiation at and detecting signals from the first observation region for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0019] In further aspects, the invention provides methods of monitoring a primer extension reaction, comprising providing a polymerase enzyme within a first observation region, contacting the polymerase with at least a first fluorescent or fluorogenic nucleotide analog, and monitoring a fluorescent signal emitted from the first observation region in response to illumination with excitation radiation for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0020] In addition, the invention provides methods for generating a plurality of noncontiguous sequence reads from a single nucleic acid template molecule. Such methods generally comprise preparing a reaction mixture comprising the template molecule, a polymerase enzyme, and a set of differentially labeled nucleotides or nucleotide analogs, wherein the set comprises at least one type of nucleotide or nucleotide analog for each of the natural nucleobases (A, T, C, and G). The polymerization reaction is initiated, the polymerase begins processive incorporation of the labeled nucleotides or nucleotide analogs into a nascent nucleic acid strand, and during such incorporation the reaction is monitored by optical means to detect incorporation events, thereby generating a first sequence read. In a subsequent step, the labeled nucleotides or analogs are replaced with unlabeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed without detecting incorporation events. Subsequently, the unlabeled nucleotides or analogs are replaced with labeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed once again with real time detection of incorporation events, thereby generating a second sequence read that is noncontiguous to the first sequence read. The substitution of labeled for unlabeled, and unlabeled for labeled, nucleotides and nucleotide analogs can be repeated multiple times to generate a plurality of noncontiguous sequence reads, each of the plurality generated during a period when the labeled nucleotides or nucleotide analogs are being incorporated into the nascent strand and such incorporation is being detected in real time.

[0021] In certain aspects, devices of the invention can comprise a solid support (e.g., substrate) having an observation region, a first reactant immobilized within the observation region, and a second reactant disposed within the observation region, and a means for subjecting the observation region to at least one illuminated period and at least one non-illuminated period. In certain embodiments, interaction between the first and second reactants under excitation illumination causes photo-induced damage to the first reactant, and further wherein the photo-induced damage is reduced by subjecting the observation region to intermittent illumination. In some embodiments, the first reactant is an enzyme (e.g., a polymerase), the second reactant (e.g., a nucleotide) has a detectable label (e.g., fluorescent label), and/or the observation region is within a zero-mode waveguide. The means for subjecting the observation region to one or more illuminated and non-illuminated periods may comprise, e.g., a laser, laser diode, light-emitting diode, ultra-violet light bulb, white light source, a mask, a diffraction grating, an arrayed waveguide grating, an optic fiber, an optical switch, a mirror, a lens, a collimator, an optical attenuator, a filter, a prism, a planar waveguide, a wave-plate, a delay line, a movable support coupled with the substrate, and a movable illumination source, and the like. The device may further comprise a means for collecting the data during the illuminated period(s), such as an optical train, e.g., operably coupled to a machine comprising machine-readable medium onto which such data may be written and stored.

[0022] In further aspects, the invention provides systems for performing intermittent detection of an analytical reaction comprising reagents for the analytical reaction disposed on a solid support, a mounting stage configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the solid support detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other, and a data processing system operably coupled to the optical train. In certain preferred embodiments, the analytical reaction is a sequencing reaction and/or the solid support comprises at least one zero-mode waveguide.

[0023] In still other aspects, the invention provides systems for analyzing an illuminated reaction that is susceptible to photo-induced damage when illuminated for a period longer than an photo-induced damage threshold period, comprising a solid support having reagents for the reaction disposed thereon, a mounting stage supporting the solid support and configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the

solid support to illuminate the portion of the solid support and detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, and a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other. In some embodiments, the illuminated reaction is a sequencing reaction, e.g., a nucleotide sequencing-by-synthesis reaction. In certain embodiments, the solid support comprises at least one optical confinement, e.g., a zero-mode waveguide.

[0024] The invention provides methods of performing analytical reactions, e.g., processive analytical reactions, that include preparing a reaction mixture comprising reaction components, at least one of which is a detectable component that is detectable during one or more detection periods, and at least one of which is a clocking component that is detectable during one or more non-detection periods during the analytical reaction. The methods further comprise initiation the analytical reaction and maintaining conditions that allow the analytical reaction to proceed while subjecting it to at least one detection period and at least one non-detection period, both in the presence of the clocking component and the detectable component. In certain embodiments, the detectable component emits a detectable signal is response to excitation illumination during the detection period, but not during the non-detection period when a clocking signal is emitted from the clocking component. The detectable signal is collected during the detection period and the clocking signal is detected during the non-detection period, e.g., using an optical system. Optionally, the clocking signal can also be collected during the detection period and the non-detection period. In certain preferred embodiments, detection data is collected in read time during the detection period, non-detection data is collected in real time during the non-detection period, and the detection data and non-detection data are both used to characterize the analytical reaction. In some embodiments, the transition between the detection period and the non-detection period does not involve substitution and/or addition of reaction components during progression of the analytical reaction, and in other embodiments the transition does involve substitution and/or addition of reaction components, e.g., via a reaction mixture exchange. In some preferred embodiments, a plurality of analytical reactions are disposed on a solid support, subjected to intermittent illumination, monitored to collect data, and characterized based upon the data so collected.

[0025] The detectable component and clocking component are typically linked to discrete molecules in the analytical reaction. For example, the detectable component can be linked to a first subset of nucleotide analogs and the clocking component can be linked to a second subset of

nucleotide analogs in the analytical reaction mixture. Alternatively, both the detectable component and the clocking component can be linked to a single molecule, e.g., a single nucleotide or nucleotide analog, in the analytical reaction. The detectable component and clocking component can both comprise detectable labels (e.g., luminescent, fluorescent, or fluorogenic labels, including, e.g., quantum dots), and in some embodiments, different detectable labels, e.g. having different absorption peaks.

[0026] In certain preferred embodiments, an analytical reaction performed according to the invention comprises at least one enzyme, e.g., a polymerase, ligase, ribosome, nuclease, and/or kinase. In some embodiments, pause or stop points are engineered into the analytical reaction to control activity of the enzyme. Various aspects of the analytical reaction can be changed by being subjected to at least one detection period and at least one non-detection period, such aspects including but not limited to processivity, fidelity, rate, and duration, e.g. of enzyme activity.

[0027] In certain preferred embodiments, the analytical reaction is a sequencing reaction comprising a single nucleic acid template that generates sequence reads during the detection period by detecting the detectable component, and does not generate sequence reads during the non-detection period by suspending detection of the detectable component. Such a sequencing reaction typically comprises at least two or three detection periods and generates a plurality of noncontiguous sequence reads from the single nucleic acid template. In some embodiments, the template comprises multiple repeat or complementary sequences. In some embodiments, the sequencing reaction comprises passage of the single nucleic acid or a nascent strand complementary thereto through a nanopore. In some preferred embodiments, the sequencing reaction comprises primer extension by a polymerase enzyme and the detectable component is linked to a nucleotide or nucleotide analog. In some embodiments, the clocking component is linked to the polymerase enzyme, and optionally can be a multi-component label, e.g, a FRET label.

[0028] In certain aspects, the invention provides methods of mitigating photo-induced damage during an illuminated reaction that include preparing a reaction mixture having first and second reactants, where interaction of the reactants under excitation illumination can cause photo-induced damage to the first reactant. The illuminated reaction is subjected to intermittent excitation illumination characterized by periods of maximal illumination followed by periods of modified but not absent illumination. The intermittent excitation illumination reduces the amount of photo-induced damage to the first reactant during the illuminated reaction as compared to the illuminated reaction under constant maximal excitation illumination, thereby mitigating photo-induced damage

to the first reactant. In certain preferred embodiments, the illuminated reaction is a primer extension reaction. In certain preferred embodiments, the first reactant is an enzyme, e.g., a polymerase or ligase enzyme. In certain preferred embodiments, the second reactant comprises a fluorescent or fluorogenic molecule. In certain embodiments, the modified excitation illumination is illumination with a lower intensity excitation illumination than the maximal excitation illumination. In certain embodiments, a set of illumination sources provides the maximal excitation illumination and a subset of the set of illumination sources provides the modified excitation illumination.

[0029] In other aspects, the invention provides a method of sequencing a template nucleic acid that includes subjecting the template to methylation to generate at least one methylated base, subjecting the methylated base to base excision to generate at least one abasic site in the template, annealing a primer to the template nucleic acid, contacting the template with a polymerase enzyme to promote extension of the primer in a template-dependent manner, monitoring the extension of the primer in real time to generate a nucleotide sequence read complementary to the template, extending the primer until the abasic site is encountered by the polymerase, at which time the polymerase pauses on the template, and reinitiating primer extension by facilitating abasic site bypass by the polymerase. The monitoring, extending, and reinitiating steps are repeated until a desired number of nucleotide sequence reads is generated and collected, and subsequently analyzed to determine the sequence of the template nucleic acid. In certain embodiments, the contacting step occurs during a detection period or a detection period immediately follows the contacting step. In certain embodiments, a detection period ends and a non-detection period begins prior to one or more pauses of the polymerase on the template. In certain embodiments, a non-detection period is terminated simultaneous with or immediately following one or more reinitiation steps. In some embodiments, the reinitiating step comprises introduction of a pyrene to the polymerase, where the polymerase incorporates the pyrene into the nascent strand opposite and, therefore, "pairing with" an abasic site in the template. In certain preferred embodiments, the template is circular and the polymerase pauses at the same abasic site multiple times during the primer extension reaction. In other embodiments, the method further comprises terminating the monitoring when a desired length of the nucleotide sequence read is collected, e.g., by removing or modifying excitation illumination. Optionally, the desired length can be less than a length of the template nucleic acid. Additionally, the monitoring can be reinitiated subsequent to or simultaneous with the reinitiating of primer extension.

[0030] In yet further aspects, the invention provides a method of performing an illuminated reaction that includes preparing a reaction mixture comprising multiple optically detectable components that are distinguishable from one another based upon their individual signal emissions, initiating the illuminated reaction, and maintaining conditions that allow the illuminated reaction to proceed while subjecting the reaction mixture to at least one maximal illuminated period and at least one modified illuminated period during the illuminated reaction. In preferred embodiments, at least a portion of the optically detectable components are detectable during both the maximal and modified illuminated periods. In certain embodiments, the maximal illuminated period is characterized by a first excitation radiation intensity and the modified illuminated period is characterized by a second excitation radiation intensity that is less than the first excitation radiation intensity. In certain preferred embodiments, all of the optically detectable components are detectable during both the maximal and modified illuminated periods, but are distinguishable from one another during the maximal illuminated period, but are not distinguishable during the modified illuminated period. In certain embodiments, the maximal illuminated period comprises exposing the reaction mixture to a set of excitation radiation wavelengths and the modified illuminated period comprises exposing the reaction mixture to a subset of the set of excitation radiation wavelengths. In certain preferred embodiments, all of the optically detectable components are detectable and distinguishable during the maximal illuminated period, but only a subset of the optically detectable components are detectable during the modified illuminated period.

[0031] In some embodiments, the illuminated reaction is initiated during a modified illuminated period and subsequently subjected to a maximal illuminated period, where data collected during the modified illuminated period is used in the statistical analysis of data collected during the maximal illuminated period. For example, an illuminated reaction that is a polynucleotide sequencing reaction can generate sequence read data during a modified illuminated period that is subsequently used to construct a sequence scaffold for assembly of sequence read data collected during a maximal illuminated period. Additionally or optionally, the illuminated reaction is a template-directed sequencing reaction and sequence read data collected during a modified illuminated period is used to determine a rate of translocation of a polymerase during the modified illuminated period.

[0032] Some embodiments of the invention comprise performing a plurality of illuminated reactions, each of which is exposed to the set of excitation radiation wavelengths during the maximal illuminated period, but is exposed to a different subset of the set of excitation radiation

wavelengths during the modified illuminated period, such that a distinct subset of optically detectable components are detectable during the modified illuminated period for each of the plurality of illuminated reactions. In other words, for two such illuminated reactions, although all optically detectable components are detectable during their respective maximal illuminated periods, only a subset of the optically detectable components is detectable in each reaction, and the subset detectable in the first reaction is preferably different from the subset detectable in the second reaction.

[0033] In certain aspects, the invention provides methods for performing paired-end sequencing on a single template molecule. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion. A first linker ligated to the first terminal portion of the nucleic acid molecule connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and a second linker ligated to the second terminal portion of the nucleic acid molecule connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion. A template nucleic acid molecule is thereby formed comprising the double-stranded nucleic acid molecule with both the first linker and the second linker ligated thereto. The template molecule is subjected to a sequencing process in which sequence reads are generated for the first terminal portion and the second terminal portion, but sequence reads are not generated for the intermediate portion, even if the intermediate portion is processed during the sequencing process, e.g., by a polymerase. In some embodiments, the first linker and second linker are identical, and in other embodiments they are different from one another, i.e., not identical. In certain embodiments, the first and second linkers comprise complementary regions and can be hybridized to one another prior to one or both of the ligating steps. In some cases, hybridized linkers that are ligated to the ends of a double-stranded nucleic acid molecule are separated prior to subjecting the molecule to a sequencing reaction, and in some cases the hybridized linkers remain hybridized during at least a portion of the sequencing reaction. For example, in a template-directed sequencing reaction, a polymerase capable of strand displacement separates the hybridized linkers as it sequences the template. In certain preferred embodiments, the sequencing process comprises at least one detection period (e.g., an illuminated period) and at least one non-detection period (e.g., a non-illuminated period) such that the intermediate portion of the template molecule is subjected to the sequencing process during the non-detection period. In some embodiments, the template is fragmented after ligation to remove the intermediate portion. The

sequencing process can generate redundant sequence data from one or both of the first terminal portion and the second terminal portion, and/or can generate sequence data from an additional portion of the template molecule that is noncontiguous with the first terminal portion and the second terminal portion. In preferred embodiments, the sequencing process involves circularizing the template molecule by separating the complementary strands of the template molecule and using the complementary strands in template-directed nascent strand synthesis catalyzed by a single polymerase enzyme. Optionally, the template molecule can comprise a primer binding site, a registration sequence, and/or a synthesis blocking moiety. The primer binding site, a registration sequence, or synthesis blocking moiety can be present in one or both of the linkers, or can be located elsewhere within the template molecule. In some cases, the synthesis blocking moiety is selected from the group consisting of an abasic site, a nick, a synthetic linker, a non-native nucleotide or analog thereof, a primer, a large photolabile group, a strand-binding moiety, a damaged base, and a modified base. The synthesis blocking moiety can permanently or temporarily block progression of the sequencing process, e.g., by interfering with the activity of an enzyme, e.g., a polymerase enzyme. In certain preferred embodiments, the synthesis blocking moiety is an abasic site, e.g., introduced by a DNA glycosylase.

[0034] In some aspects, the invention provides methods for generating a nucleic acid construct for analytical reactions. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion; providing a first stem-loop linker hybridized to a second stem-loop linker; ligating the first stem-loop linker to the first terminal portion of the nucleic acid molecule, wherein the first stem-loop linker connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and ligating the second stem-loop linker to the second terminal portion of the nucleic acid molecule, wherein the second stem-loop linker connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion, thereby generating the nucleic acid construct. Optionally, the nucleic acid construct can be subjected to fragmentation after the ligating of steps c and d, wherein the fragmentation removes the intermediate portion from the nucleic acid construct and introduces two double-stranded termini. The method can further include ligating the two double-stranded termini to one another. In some embodiments, one of the stem-loop linkers comprises a primer binding site, registration sequence, or a synthesis blocking moiety that is absent from the other stem-loop linker.

[0035] In further aspects, the invention includes a single template nucleic acid molecule comprising a duplex region; a first linker linking termini at a first end of the duplex region; a second linker linking termini at a second end of the duplex region, wherein a region of the first linker is complementary to a region of the second linkers. Optionally, the single template molecule comprises the first and second linkers hybridized with one another. In some embodiments, the duplex region is separated or melted apart to transform the single template nucleic acid molecule into a topologically single-stranded, circular nucleic acid molecule. Further, the invention provides a composition comprising a single, optically resolvable polymerase enzyme in association with a single-stranded circular nucleic acid molecule, wherein the single-stranded circular nucleic acid molecule comprises first, second, third, and fourth regions, and further wherein the first region is complementary to the second region, and the third region is complementary to the fourth region, and further wherein the regions are ordered on the single-stranded circular nucleic acid molecule as follows: first region, third region, second region, fourth region.

[0036] In still further aspects of the invention, machine-implemented methods for transforming nucleotide sequence read data into consensus sequence data, wherein the nucleotide sequence read data is generated by sequencing a target region of a template nucleic acid multiple times, and the consensus sequence data is representative of a most likely actual sequence of the template nucleic acid. Such machine-implemented methods can comprise various steps, such as a) mapping the nucleotide sequence data to a target sequence using a local alignment method that produces a set of local alignments comprising an optimal local alignment and sub-optimal local alignments, b) enumerating the set of local alignments, c) constructing a weighted directed graph wherein each local alignment in the set of local alignments is represented as a node, thereby generating a set of nodes in the weighted directed graph, d) drawing edges between pairs of nodes in the weighted directed graph if the pair represents a potential reconstruction of the template nucleic acid, e) assigning weights to the edges drawn in step d, wherein a given weight for a given edge represents the log-likelihood that a given pair of nodes connected by the given edge is truly a reconstruction of the template nucleic acid, f) finding the shortest path to each node in the weighted directed graph, thereby generating a set of shortest paths for the weighted directed graph, g) ranking the set of shortest paths to determine the best assignment, and h) storing the results of steps a-g on a machine-readable medium. In certain embodiments, the steps of the machine implemented methods are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Also provided are

computer program products comprising a computer usable medium having computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the machine-implemented methods of the invention, and machine-readable medium on which the results of the method steps are stored. The invention further includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods.

[0037] In certain aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme (e.g., a polymerase) during a time t, where time t occurs during a non-detection period. Such a method comprises, in certain embodiments, developing a probability model *p(v)* to describe an observed distribution of enzyme velocities during one or more detection periods; sampling velocities from *p(v)*; summing and recording the velocities sampled in step b to produce a sum that is an estimate of x/τcorr; and repeating the sampling, summing, and recording M times to generate a distribution of sums that are estimates of x/τcorr, with the distribution of sums being the distribution of the distance x travelled by an enzyme during a time t. Preferably, at least some of the steps are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Optionally, the enzyme is a polymerase enzyme. In some embodiments, multiple enzymes are observed simultaneously and the probability model *p(v)* is determined independently for each of the multiple enzymes. In certain preferred embodiments,

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}.$$

[0038] In further aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme during a time t, where time t occurs during a non-detection period. In some embodiments, the method comprises estimating a distribution of local rates p(v), making independent identically distributed draws of N = t/τcorr velocities from from p(v); summing the velocities; recording the velocities summed in c) as an estimate of x/τcorr; and repeating b-d M times, e.g., where M is preferably at least 1000. Optionally, p(v) is determined using a Hidden Markov Model or the autocorrelation function $\langle \delta v(t)\delta v(t+\Delta) \rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$. The invention further

includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods, as well as a machine-readable medium on which the results of the steps of the methods are stored.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions, whether illumination is initiated before (A) or after (B) initiation of the reaction.

[0040] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination, including depictions of multiple reactions arrayed on a solid support (A) and prophetic data (B) from certain embodiments of the invention.

[0041] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions on a solid support (A) using intermittent illumination and a mask (B). A graph (C) depicts prophetic data from certain embodiments of the invention.

[0042] Figure 4 provides additional embodiments of masks for use in the methods of the invention, including a mask that allows illumination of columns of reactions (A) and a mask that allows illumination of every other reaction in a row and column (B).

[0043] Figure 5 illustrates an aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Figure 5A illustrates a solid support comprising four quadrants, each quadrant containing a different sample. Figure 5B illustrates a mask design for selective illumination of the substrate. Figures 5C and 5D demonstrate various positions of the mask on the solid support.

[0044] Figure 6 provides an illustration of paths in a sequence alignment matrix representing sequencing data from a SMRTbell™ template.

[0045] Figure 7 illustrates a hypothetical directed graph.

[0046] Figure 8 provides data from single-molecule sequencing-by-synthesis reactions. Figure 8A provides data from a two-minute interval beginning at initiation of the reactions, i.e., from 0-120 seconds. Figure 8B provides data from a second two-minute interval from 300-420 seconds. Figure 8C provides data from a third two-minute interval from 600-720 seconds.

[0047] Figure 9 schematically illustrates one embodiment of a system for use with the methods, devices, and systems of the invention.

[0048] Figure 10 provides a graphical representation of rates of polymerase activity on different portions of a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0049] Figure 11 provides a graphical representation of the average rate of polymerase translocation over a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0050] Figure 12 provides a distribution of the physical coverage of a template nucleic acid achieved during a sequencing reaction utilizing intermittent illumination, with A showing mapping to a reference sequence with sequence reads (and portions thereof) that do not map to the reference excluded and B showing a similar mapping that further includes sequence reads corresponding to insertions in the template that are absent from the reference sequence.

[0051] Figure 13 provides a distribution of the physical coverage provided by sequence reads generated during sequencing reactions utilizing intermittent illumination across an approximately 40 kb template nucleic acid.

[0052] Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described herein and a reference sequence.

[0053] Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system.

[0054] Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters $P(S \rightarrow P) = 1/24$; $P(P \rightarrow S) = 1/11$; and $p(v) \sim \text{Gamma}(48,0.25)$. Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a non-detection period.

[0055] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period.

[0056] Figure 18 provides an exemplary strategy for selectively reducing the size of a duplex fragment within a SMRTbell™ template.

[0057] Figure 19 provides an illustrative example of nucleic acid templates having nicks.

[0058] Figures 20A and 20B illustrate two exemplary embodiments of template constructs used in the present invention.

[0059] Figures 21A and 21B schematically illustrate redundant or consensus sequencing using the constructs shown in Figures 20A and 20B.

DETAILED DESCRIPTION OF THE INVENTION

[0060] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing devices, formulations and methodologies which are described in the publication and which might be used in connection with the presently described invention.

[0061] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polymerase" refers to one agent or mixtures of such agents, and reference to "the method" includes reference to equivalent steps and methods known to those skilled in the art, and so forth. Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either both of those included limits are also included in the invention.

[0062] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention. Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

I. General

[0063] In a general sense, the methods, devices, and systems provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in

processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0064] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion.

[0065] The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions

(e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. For example, such an alteration in reaction conditions during the second stage may result in an increase in reaction rates, e.g., speeding up the progression of a template nucleic acid through a nanopore; or may reduce exposure of reaction components to harmful radiation or other reaction condition related to detection of the products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0066] "Intermittent detection," as used herein, generally refers to a means of monitoring a reaction that is carried out intermittently during the course of the reaction. Intermittent detection may refer to intermittent use of one or more monitoring methods, but does not necessarily mean that all means of monitoring a given reaction are intermittently halted. For example, monitoring of one or more nucleotide incorporations to generate nucleotide sequence reads may be intermittently halted while other aspects of a sequencing reaction are constantly monitored, e.g., temperature, reaction time, pH, etc. In certain embodiments, intermittent detection is achieved by intermittent or differential illumination of a given reaction, e.g., a reaction that uses an illumination system to detect reaction products and/or progression. Although various aspects of the invention are described herein in terms of embodiments using intermittent illumination, it should be understood that where applicable intermittent detection by other means (e.g., electrochemical, radiochemical, etc.) can be utilized in the methods of the invention. Likewise, a stage of a reaction during which an intermittent detection method is active may be referred to as a "detection period" and a stage of a reaction during which an intermittent detection method is inactive may be referred to as a "non-detection period." In illuminated reactions, such periods may also be referred to as "illuminated periods" and "non-illuminated periods," respectively, although it is to be understood that the term "non-illuminated period" included periods in which illumination may be present but altered as compared to illumination during an "illuminated period." For example, a non-illuminated period may be characterized by a complete absence of illumination, or a modification of illumination, including but not limited to changes in wavelength, frequency, intensity, and/or number of illumination sources. Alternatively or additionally, reaction components that are excited by the illumination source(s) may be modified or removed from a reaction mixture to create a non-illuminated period.

For example, a fluorescent dye detected during an illuminated period may be removed from the reaction mixture, e.g., by buffer exchange, thereby producing a non-illuminated period during which time the fluorescent dye cannot be detected even if the excitation illumination is present. In a further example, a non-illuminated period can indicate a period during an illuminated reaction during which a type of illumination-based detection that occurs during an illuminated period is not occurring, e.g., the identity of fluorescently labeled nucleotides incorporated into a nascent strand is not being detected or recorded.

[0067] In certain aspects, the present invention is generally directed to improved methods, devices, and systems for performing illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In certain preferred embodiments, illuminated reactions comprise one or more fluorogenic or fluorescent components. In accordance with certain methods of the invention, such illuminated analyses are subjected to intermittent detection (e.g., data collection) for one or more aspects of the data typically collected for a given reaction. For example, aspects of the data typically collected for nucleotide sequencing reactions include nucleotide sequence data, read quality data, signal to background ratios, reaction rates and durations, measures of the fidelity of the reaction, reaction times, and the like. In certain preferred embodiments, nucleotide sequence data is iteratively collected during an ongoing sequencing reaction to generate nucleotide sequence reads for at least two or more noncontiguous regions of a template nucleic acid molecule. Such iterative sequence data acquisition may be achieved in various ways depending on the sequencing technology in use. For example, in sequencing methods that utilize luminescent components that generate a signal indicative of the identity of a base position, iterative sequence data collection may be achieved by removing or altering an illumination source (or a reaction relative to an illumination source), substituting the luminescent components for unlabeled components that do not generate signal, or otherwise interrupting signal acquisition in the experimental system.

[0068] In certain preferred embodiments, such illuminated reactions are illuminated for an amount of time that permits the effective performance of the analysis. Traditionally, illuminated reactions are illuminated from initiation through completion, and the time during which reaction data may be reliably collected is dictated by the progression (as measured by, e.g., processivity,

rate, fidelity, duration, etc.) of the reaction under constant illumination. Some reactions are sensitive to such constant illumination, which can reduce their performance (e.g., processivity), and thereby prevent collection of data from later stages of the reaction, i.e., stages that would otherwise occur if the reaction were carried out with no illumination. The present invention provides methods for performing illuminated reactions comprising subjecting the reactions to intermittent illumination. Such intermittent illumination can increase performance (e.g., processivity, rate, fidelity, duration, etc.) of the reactions, thereby allowing generation of data that cannot be collected under constant illumination, such as data from later stages of an ongoing reaction whose progression is compromised under constant illumination. For example, in sequencing-by-incorporation reactions the use of intermittent excitation illumination can increase processivity, which has the benefit of providing sequence reads more distal from the polymerase binding/initiation site than such reactions subjected to constant exposure to excitation illumination.

[0069] Further, it is an object of the instant invention to provide sequence data from noncontiguous regions of a nucleic acid template in a single reaction. Other commercially available platforms have attempted to achieve such noncontiguous sequence data through, e.g., complex cloning and sequencing strategies. The present invention provides a clear advantage over such strategies by providing a simple and economical solution that is applicable across various platforms, and is particularly applicable to illuminated, single-molecule sequencing-by-incorporation reactions.

[0070] In preferred embodiments, illuminated reactions for use with the instant invention are nucleic acid sequencing reactions, e.g., sequencing-by-incorporation reactions. In preferred embodiments, such an illuminated reaction analyzes a single molecule to generate nucleotide sequence data pertaining to that single molecule. For example, a single nucleic acid template may be subjected to a sequencing-by-incorporation reaction to generate one or more sequence reads corresponding to the nucleotide sequence of the nucleic acid template. For a detailed discussion of such single molecule sequencing, see, *e.g.*, U.S. Patent Nos. 6,056,661, 6,917,726, 7,033,764, 7,052,847, 7,056,676, 7,170,050, 7,361,466, 7,416,844; Published U.S. Patent Application Nos. 2007-0134128 and 2003/0044781; and M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, SCIENCE 299:682-686, January 2003 Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations, all of which are incorporated herein by reference in their entireties for all purposes. In some embodiments, a plurality of single nucleic acid templates are analyzed separately and often simultaneously to generate a plurality of sequence reads

corresponding to the nucleotide sequences of the plurality of nucleic acid templates. In certain preferred embodiments, the plurality of nucleic acid templates includes at least two nucleic acid templates that comprise identical nucleotide sequences such that analysis of the two nucleic acid templates generates overlapping sequence reads. In certain preferred embodiments, at least one of the nucleic acid templates is configured to provide redundant sequence data in a single sequence read, e.g., via duplications, sense and antisense sequences, and/or circularization.

[0071] Certain aspects of the invention are directed to methods, devices, and systems for generating a sequence scaffold for a nucleic acid template, e.g., chromosome, genome, or portion thereof. A sequence scaffold as used herein refers to a set of sequence reads that extends across at least a portion of a nucleic acid template. In some embodiments, such a sequence scaffold is used to generate a consensus sequence for the nucleic acid template. In some embodiments, the nucleic acid template is very large, e.g., at least about 100, 1000, 10,000, 100,000, or more bases or base pairs in length. In some embodiments, the sequence scaffold and/or consensus sequence is based on at least 1-, 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of at least a portion of the nucleic acid template. In some preferred embodiments, the portion of the nucleic acid is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the entire length of the nucleic acid template.

[0072] In certain aspects, the invention is particularly suitable for sequencing nucleic acid templates interspersed with repetitive elements. Such repetitive elements present major logistical and computational difficulties for assembling fragments produced by sequencing strategies, especially those with read-lengths that are too short to encompass unique reads outside the repeat region. For example, the human T-cell receptor locus contains a five-fold repeat of a trypsinogen gene that is 4 kbp long and that varies 3 to 5% between copies. Therefore, a sequencing strategy that cannot provide nucleotide sequence information that spans at least 20 kb for a single molecule containing the locus will have difficulty providing consensus sequence for the locus. Further, Alu repeats (~300 bp retrotransposons) are also problematic because they cluster and can constitute up to 50-60% of the template sequence, with copies varying from 5-15% between each other. The human genome contains an estimated one million Alu repeats and 200,000 LINE elements (average length ~1000 bp), representing roughly 10% and 5% of the entire genome, respectively. In certain embodiments, the present methods facilitate efficient and accurate sequence determination for long templates comprising such repetitive sequences, in part because the present methods do not rely solely on sequence overlap to generate consensus sequences, but also include information related to

the expected location of the polymerase on the template nucleic acid, thereby linking a particular sequence read to a particular location on the template nucleic acid. This greatly facilitates accurate assembly of sequence reads to generate sequence scaffolds and/or consensus sequences.

[0073] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads in an illuminated sequencing-by-incorporation reaction that are distal from one another (i.e., noncontiguous) on a single nucleic acid template by removing the excitation illumination during the course of the reaction, and subsequently reinitiating the excitation illumination. Sequence reads are generated only during the periods of time when the excitation illumination is present, resulting in a "gap" between the sequence reads from a single template nucleic acid that corresponds to the time during which the excitation illumination was absent but the incorporation of nascent nucleotides continued "in the dark." As such, the number of sequence reads generated for a given template nucleic acid is equal to the number of periods during which the excitation illumination is present.

[0074] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads from a plurality of nucleic acid templates comprising identical nucleotide sequences. In some embodiments, the multiple sequence reads are not all from the same region of the nucleic acid templates. In some embodiments, there is overlap between the multiple sequence reads. In some embodiments, a single sequence read is generated from each of the plurality of nucleic acid templates, and in other embodiments multiple noncontiguous sequence reads are generated from each of the plurality of nucleic acid templates. In certain preferred embodiments, the multiple noncontiguous sequence reads from each of the plurality of nucleic acid templates together extend across the nucleic acid templates such that they can be combined to provide a consensus sequence for the identical nucleotide sequence in the nucleic acid templates. In some embodiments, the consensus sequence is based on at least 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of the identical nucleotide sequence. In some embodiments, the identical nucleotide sequence represents at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the nucleic acid template.

[0075] Certain aspects of the invention are directed to methods, devices, and systems for reducing or limiting the effects of photo-induced damage during illuminated reactions, particularly reactions that employ fluorescent or fluorogenic reactants. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction. Without being bound to a particular theory or

mechanism of operation, some illuminated reactions are subject to photo-induced damage that can hinder progression of the reaction, e.g., via damage to reaction components, such as enzymes, cofactors, templates, etc. As such, the illumination of the illuminated reaction can directly or indirectly negatively impact progression of the reaction, and such an impact can be measured based on various characteristics of the reaction progression, e.g., processivity, rate, fidelity, duration, etc. The present invention provides methods for subjecting an illuminated reaction to intermittent exposure to illumination, which reduces the amount of photo-induced damage at a given time during the reaction, allowing the reaction to proceed further than it does when constantly exposed to the illumination.

[0076] In some embodiments, the methods herein may further comprise the addition of one or more photo-induced damage mitigating agents (e.g., triplet-state quenchers and/or free radical quenchers) to the illuminated reaction. Such photo-damage mitigating agents are generally known to those of skill in the art. Further discussion of photo-induced damage and related compounds, compositions, methods, devices, and systems are also provided in U.S. Pub. No. 20070161017, filed December 1, 2006; and U.S.S.N. 61/116,048, filed November 19, 2008, which are incorporated by reference herein in their entireties for all purposes.

II. Intermittent Illumination of Analytical Reactions

[0077] Certain aspects of the invention are generally directed to improved methods for performing illuminated analyses. The terms "illuminated analysis" and "illuminated reaction" are used interchangeably and generally refer to an analytical reaction that is occurring while being illuminated (e.g., with excitation radiation), so as to evaluate the production, consumption, and/or conversion of luminescent (e.g., fluorescent) reactants and/or products. As used herein, the terms "reactant" and "reagent" are used interchangeably. As used herein, the terms "excitation illumination" and "excitation radiation" are used interchangeably. In certain embodiments, the illuminated reaction is a sequencing reaction, e.g., a sequencing-by-incorporation reaction. In certain embodiments, the illuminated reaction is designed to analyze a single molecule, e.g., by ensuring the molecule is optically resolvable from any other molecule being analyzed and/or in the reaction mixture. In certain embodiments, one or more components of the reaction are susceptible to photo-induced damage directly or indirectly elicited by an excitation radiation source. In certain preferred embodiments, an illuminated reaction is subjected to intermittent excitation radiation during the course of the illuminated reaction. In certain preferred embodiments, a sequencing-by-

incorporation reaction is subjected to intermittent excitation radiation during the course of a polymerization reaction to generate a plurality of noncontiguous sequence reads from a single nucleic acid template.

[0078] In certain aspects, the methods herein provide benefits over methods currently used for sequencing large template nucleic acids, such as human genomes. For example, the traditional shotgun sequencing approach entails sequencing nucleic acid fragments and analyzing the resulting sequence information for overlap and similarity to known sequences to construct the complete sequence of the template nucleic acid. One disadvantage to the shotgun approach is that assembly may be difficult if the template nucleic acid comprises numerous repeated sequences, and the inability to assemble a genomic sequence in repeat regions leads to gaps in the assembled sequence. (See, e.g., Myers, G.; "Whole-Genome DNA Sequencing" in Computing in Science and Engineering; Vol 1, Issue 3; pgs. 33-43; May/Jun 1999.) One method of resolving these gaps is to sequence fragments large enough to span the repeat regions, but sequencing large fragments can be difficult and time-consuming. Another approach to spanning a gap is to determine the sequence of two ends of a large fragment which has known spacing and orientation, and this approach is generally termed paired end sequencing (see, e.g., Smith, M. W. et al., (1994) Nature Genetics 7:40-47; and U.S. Pub. No. 2006/0292611, filed June 6, 2006, both of which are incorporated by reference herein in their entireties for all purposes). This method is limited by the requirement for information about the spacing and orientation of the ends of the long fragment, and/or complex sample preparation of the nucleic acid template. The present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.) or complex sample preparation, thereby allowing economical, efficient, and effective de novo sequencing or resequencing of long template nucleic acids.

[0079] In certain aspects, the methods herein provide various strategies for achieving intermittent illumination of illuminated reactions. Essentially, at least one type of illumination (e.g., excitation illumination) is present for at least one time period ("illuminated period") and absent during at least one other time period ("non-illuminated period") during an illuminated reaction. As described above, the term "non-illuminated" indicates a change in illumination including, but not limited to a complete absence of illumination. For example, a non-illuminated period may also be characterized by a different illumination source or intensity than an illuminated period, or by a change in reaction components, e.g., detectable labels. In general, at least one type of data collected

during an illuminated period (e.g., nucleotide sequence data) is not collected during a non-illuminated period. An absence of the illumination may be due to, e.g., inactivation of the illumination source (e.g., laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source), removal of the illuminated reaction from the illumination source (or vice versa), or may be due to blockage of the illumination from the reaction, as discussed below. Modifications to the illumination may be due to, e.g., adjustment of the intensity of an illumination source, or a substitution of one illumination wavelength and/or frequency for another. Further, components detectable during an illuminated period may be removed from the reaction mixture during a non-illuminated period, e.g., a fluorescently labeled nucleotide may be replaced with an unlabeled nucleotide. Knowledge of the rate of the reaction and the time during which the illumination is absent is used to estimate the progress of the reaction during the non-illuminated period. For example, if a reaction proceeds such that one molecule is incorporated into a macromolecule per second, and the illumination is absent for 20 seconds, it can be estimated that 20 molecules were incorporated during the non-illuminated period. This information is useful during data analysis to provide context for the reaction data collected during the illuminated period(s). For example, in a sequencing-by-incorporation reaction the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. Further the processivity of an enzyme during a non-illuminated period (or other type of non-detection period) can be manipulated or adjusted by methods known to those of skill in the art. In particular, the kinetics of replication by a polymerase enzyme can be altered by changing the chemical environment in which it operates, and such methods are further described, e.g., in U.S. Patent Application Nos. 12/414,191, filed March 30, 2009; 12/537,130, filed August 6, 2009; and U.S. Patent Application No. [unassigned], attorney docket no. 105-006301US, entitled "Engineering Polymerases and Reaction Conditions for Modified Incorporation Properties," filed September 4, 2009, the disclosures of all of which are incorporated herein by reference in their entireties for all purposes. For example, methods are provided for adjusting the enzyme activity, and these methods find particular relevance in the instant invention when used to enhance accuracy

during detection periods, and to enhance processivity during non-detection periods. Information regarding enzyme translocation rate and processivity is useful for positioning the sequence reads for a single template nucleic acid relative to one another in the construction of a sequence scaffold and/or consensus sequence for the template nucleic acid.

[0080] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions. A reaction mix is prepared at step 100. In process A shown on the left, illumination of the reaction 105 is begun prior to initiation of the reaction 110, which allows "illumination data" to be collected at initiation. (In an alternative embodiment, illumination may commence simultaneously with initiation of the reaction.) "Illumination data" as used herein refers to data collected during an illuminated period, e.g., the length of the illuminated period and luminescent signal(s) from the reaction product. At least one non-illuminated period 115 occurs during the course of the reaction, followed by at least one additional illuminated period 120. Multiple additional non-illuminated and illuminated periods may follow. During the illuminated periods (105 and 120), illumination data is collected 175. During the non-illuminated period(s), non-illumination data is collected 180. As used herein, "non-illumination data" refers to data collected during a non-illuminated period, e.g., the length of the non-illuminated period can be monitored. In process B shown on the right, the reaction is initiated 155 during a first non-illuminated period 150. At least one illuminated period 160 occurs during the course of the reaction, optionally followed by at least one additional non-illuminated period 165. Multiple additional illuminated and non-illuminated periods may follow. As for process A, illumination data is collected 175 during the illuminated period(s) 160, and non-illumination data is collected 180 during non-illuminated periods (155 and 165).

[0081] One benefit provided in certain embodiments of the invention is that the reaction need not be further manipulated after initiation (aside from the control of illumination). For example, the method can be used to analyze reaction mixtures without the need for buffer changes, addition of further reaction components, or removal of detectable components, e.g., light-activatable components such as fluorophores. For example, in a sequencing-by-incorporation reaction, labeled nucleotides may be present throughout the life of the reaction, even when the reaction is not generating nucleotide sequence data (e.g., during a non-illuminated period). This provides clear advantages over methods that require additional handling of the reaction after initiation, which tend to not only be expensive and time-consuming, but which also provide opportunities for contamination of the reaction. For example, illumination can be reinitiated at any time during the

reaction at the whim of the ordinary practitioner by simply activating the illumination. In certain preferred embodiments, the concentration of labeled nucleotides or nucleotide analogs in the reaction mixture is greater than the concentration of unlabeled nucleotides in the reaction mixture throughout the course of the reactions, and may represent at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the total nucleotides in the reaction mixture. Methods for ensuring a high ratio of labeled versus unlabeled nucleotides in a reaction mixture are known in the art and certain preferred embodiments are provided in U.S. Patent Pub. Nos. 2006/0063264, 2006/0194232, and 2007/0141598, which are incorporated herein by reference in their entireties for all purposes.

[0082] In embodiments in which a sequencing-by-incorporation reaction is subjected to intermittent illumination, the sequence reads collected during the illuminated periods are arranged in order and separated from one another by an estimated number of nucleotides incorporated into the nascent strand during the intervening non-illuminated periods. The resulting gapped read can then be used to assess certain characteristics of the template nucleic acid. When multiple identical template nucleic acids are subjected to such a sequencing-by-incorporation reaction, the resulting set of gapped reads can be combined to create a sequence scaffold and/or a consensus sequence for the template nucleic acid.

[0083] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods can be used to provide additional data during the course of the reaction, e.g., data from illuminated or non-illuminated periods. In certain preferred embodiments, such alternative labeling methods may comprise using labels that are incorporated into a product of the reaction. For example, in sequencing-by-incorporation reactions that use nucleotides comprising labeled terminal phosphates (*e.g.*, the gamma phosphate as in dNTP, or terminal phosphates on nucleotide analogs with a greater number of phosphate groups) to identify the nucleotides incorporated into a nascent polynucleotide, the reaction mixture may also include nucleotides comprising a base-linked label. During the reaction, these "base-labeled nucleotides" will be incorporated into the nascent strand, but unlike the terminal phosphate labels removed during incorporation, the base-linked labels are not cleaved from the nucleotide upon incorporation by the polymerase, resulting in a nascent strand that comprises the base-linked labels. The concentration of such base-labeled nucleotides can be adjusted in the reaction mixture to promote their incorporation into the nascent strand at a predictable rate, e.g., based on the known sequence of the template or the average frequency of a given nucleotide. The presence and/or rate of

incorporation of the base-linked labels into the nascent strand can provide a measure of the length of the nascent strand generated (and, therefore, the distance traveled by the polymerase along the template nucleic acid) during the reaction by subjecting the reaction to excitation illumination that excites the base-linked label (but preferably not the non-base-linked labels), and detecting the signal emitted. The excitation of the base-linked labels preferably occurs as a pulse during or immediately following a non-illuminated period, and is otherwise absent during the reaction. The strength of the signal is indicative of how many labels are present in the nascent strand, thereby providing a measure of the processivity of the polymerase for a given period during the ongoing reaction, e.g. during one or more illuminated or non-illuminated periods. Since the base-linked labels remain in the nascent strand, it is beneficial to minimize the amount of time those fluorophores are subjected to excitation illumination to mitigate the potential of photo-induced damage to the reaction components. As such, in preferred embodiments, the excitation illumination wavelength for the base-labeled nucleotides is different than that of other fluorescent labels in the reaction.

[0084] This method can be modified in various ways. For example, the base-labeled nucleotides may also comprise a terminal phosphate label so that their incorporation can be monitored in the same manner during an illuminated period as the non-base-labeled nucleotides. There may be a single type of base-labeled nucleotide in a reaction mixture, or multiple types may be present, e.g., each type carrying a different nucleobase. The concentration of base-labeled nucleotides in the reaction mix may be varied, although it is preferred that the ratio of base-labeled nucleotides to non-base-labeled nucleotides be relatively low. For example, in a reaction mixture comprising a single type of base-labeled nucleotide (e.g., base-labeled dATP), it is preferred that the ratio of base-labeled dATP to non-base labeled dATP be less than 1:8, and more preferably 1:10 or less. The low concentration of base-labeled nucleotides is preferred in order to minimize sterically induced polymerase stalling when incorporating multiple base-labeled nucleotides in a row. In some embodiments, the optimal ratio is pre-determined using capillary electrophoresis for any specific base-labeled nucleotide and likely homopolymer sequence prevalence. In certain preferred embodiments, at least 50, 75, 100, 125, or 150 base-labeled nucleotides are incorporated into the nascent strand during a single non-detection period. The base-labeled nucleotides may be present throughout the reaction, or may be washed in during non-illuminated periods and washed out after the pulse of excitation illumination. The reaction mixture comprising base-linked nucleotides being washed in may also include unlabeled nucleotides for incorporation during a non-detection period. During a subsequent illuminated period, a reaction mixture comprising terminal phosphate-labeled

nucleotides replaces the reaction mixture comprising base-linked nucleotides and unlabeled nucleotides. This protocol is one embodiment of the methods of the invention in which a non-detection period is not necessarily a non-illuminated period because in this case illumination may be present, but no incorporation of nucleotides is be detected.

[0085] Alternatively or in addition, a low concentration of a fifth terminal phosphate labeled nucleotide can be present in the sequencing reaction, wherein the label has a different excitation wavelength than the other labels in the reaction mixture. For example, a small proportion of one nucleotide analog, e.g., dA6P, can be labeled with the "fifth label." During non-detection periods when the sequence of incorporation of nucleotides is not being monitored, the reaction site is illuminated by excitation radiation specific for the fifth label, and this fifth label excitation radiation can be inactivated during the detection periods. Emissions detected upon incorporation of the nucleotide analog comprising the fifth label are used to "clock" the pace of the polymerase during the non-detection period, e.g., based upon the known or estimated frequency of the complementary nucleotide in the template strand. The fifth label can be chosen such that the excitation and emission radiation are less likely or unlikely to cause photo-induced damage to reaction components, e.g. by choosing a label with a long excitation wavelength (e.g., toward the red end of the visible spectrum), a label that has a low propensity for entering into a triplet state, and/or a label that has a low propensity to form a radical. Since the fifth label is being excited when other labels are not, there is no requirement for optimal spectral separation from other labels in the reaction mixture. Further, since the fifth label is not being used for sequencing, other optimizations are also not necessary, e.g., related to branching, accuracy, and the like. Various types of labels can be used as a fifth label of the invention including, but not limited to, organic and non-organic dye fluorophores. For example, latex nanoparticles or quantum dots are particularly suitable due to their lower propensity for photo-induced damage of certain analytical reaction components. In certain preferred embodiments, a quantum dot label has an emission spectrum within the same spectral window as the labels that are used to identify the sequence of base incorporations into the nascent strand ("sequencing labels") but an excitation spectrum that does not overlap those of the sequencing labels to allow detection of the fifth label emissions using the same optical system as is used to detect the sequencing label emissions.

[0086] This method can be modified in various ways. For example, more than one small subset of a nucleotide analog can be labeled with a fifth label, and in certain embodiments, a small subset of each nucleotide analog present in the reaction mixture is labeled with the fifth label.

Further, there may be a plurality of additional labels present in the reaction, each of which is present on a small subset of a single type of nucleotide analog, e.g., sixth, seventh, and eighth labels. By increasing the number of types of nucleotide analogs labeled with fifth (or sixth, seventh, eighth) labels, their frequency of incorporation is likewise increased, which improves the translocation rate calculation for the polymerase during the non-detection periods. Alternatively, each type of nucleotide analog can comprise both a sequencing label that is specific for the cognate base in the nucleotide, as well as a fifth label for clocking the polymerase. The sequencing labels are excited and detected during the detection periods and the fifth labels are excited and detected during the non-detection periods. Since every nucleotide analog is labeled with a fifth base, each incorporation event can be counted during the non-detection period and the exact rate of incorporation can be determined. Both the sequencing and fifth labels may be bound to the same or different linkers on the nucleotide analogs. In certain preferred embodiments, a linker on a nucleotide analog positions the fifth label within an illumination zone to allow excitation, but far from an enzyme (e.g., polymerase) to mitigate photo-induced damage related to excitation of and/or emission from the fifth label.

[0087] In some embodiments, the fifth label is also excited by an illumination during the detection periods. The availability of the clocking function during the detection period can be used during sequence analysis to identify positions in the resulting sequence read where a signal was not detected (resulting in an apparent "missing base" in the read) and to distinguish between true insertions and branching events in which two signals are detected for a single incorporation event.

[0088] In yet further embodiments, assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid is facilitated by using "non-illuminated periods" characterized by modified excitation illumination rather than a complete absence of excitation illumination (which can also be termed "low-illuminated periods"). For example, in some embodiments a lower intensity excitation illumination is used during the non-illuminated periods that excites one or more of the labels that are excited during the illuminated periods. As such, unlike various strategies described above, no fifth label is necessary. The lower intensity excitation illumination results in emissions that are lower intensity but still intense enough to identify an emission signal over background counts, though typically not intense enough to be used to identify the particular label generating the emission signal. For example, if label "A" and label "B" are in a reaction mixture, during an illuminated period the intensity of the signal emissions from each are high enough that the artisan can distinguish from which label a particular signal originates by the

wavelength and/or frequency of the signal. However, during a low-illuminated period the artisan can only identify that a signal emission occurs, but is unable to distinguish the originating label because its particular wavelength and/or frequency cannot be accurately determined. The decrease in excitation illumination intensity provides both a mitigation of photo-induced damage to reaction components within the observation volume while allowing the practitioner to count the emissions, and therefore the incorporations, during the non-illuminated period.

[0089] In other embodiments, multiple excitation illumination sources are used during an illuminated period, and a first subset of these illumination sources is removed during a non-illuminated period, while a second subset remains. The illumination sources that remain during the non-illuminated period may be present in the same manner as during the illuminated period, or various aspects may be altered, e.g., intensity may be reduced. For example, if labels A and B present in a reaction mixture are excited by a first illumination source and labels C and D present in the reaction mixture are excited by a second illumination source, removal of the first illumination source during the non-illuminated period results in an inability to detect labels A and B, while C and D are still detectable. Such an incomplete data set can be used to clock the progress of the reaction during the non-illuminated period(s). Further, it can also be used in various ways to facilitate the statistical analysis of data collected during the illuminated period(s). For example, for nucleotide sequencing applications (as described elsewhere herein) the incomplete data set(s) collected during non-illuminated period(s) can be used during assembly of a sequence scaffold. For example, during *de novo* sequence assembly a collection of sequences (contigs) are generated, but the order of the contigs relative to the template nucleic acid is not always apparent. The scaffolding process uses extra information to determine the correct order of the contigs. So, if only two bases are identifiable in the non-illuminated periods, the incomplete sequence reads comprising only incorporation of these two bases can be aligned to modified versions of the contigs assembled from data collected during an illuminated period, but in which the two bases not detected during the non-illuminated periods have been removed. Once the order of the contigs has been determined, the incorporation data for the two bases not detected during the non-illuminated periods is restored and the assembly of the contigs is complete. This method can be modified in various ways. For example, the practitioner may choose which illumination sources to remove during the non-illuminated periods based on various characteristics, such as their propensity to cause photo-induced damage to one or more reaction components, the propensity of the corresponding emission signal to cause photo-induced damage to one or more reaction components; their energy consumption, and

wear-and-tear on the source device. Further, as described elsewhere herein, rather then removing an illumination source, reaction components that are excited by the illumination source may be removed from the reaction mixture during the non-illuminated period, necessarily rendering them undetectable. For example, one or more fluorescently labeled nucleotide analogs may be replaced with unlabeled nucleotide analogs during the non-illuminated periods.

[0090] In certain aspects, the invention provides advantages to performing intramolecular redundant sequencing, in which a template nucleic acid is used to generate multiple copies of a sequence read of interest, whether by virtue of multiple copies of the complement being present in the template, repeated replication of the template, or a combination thereof. For example, a first stage of a template-dependent sequencing reaction on a single-stranded circular template can comprise a non-illuminated period during which the template is completely replicated at least one time to generate at least one incomplete sequence read for a sequence complementary to the template. The first stage is followed by a second stage comprising an illuminated period during which the template is replicated multiple times to generate multiple complete sequence reads for the complementary sequence. The incomplete reads generated in the first stage can be used to construct a scaffold for assembly of the complete sequence reads generated in the second stage. Further, incomplete sequence reads can also be used to clock the progress of the reaction during the non-illuminated periods by providing a count of the detectable reaction components and combining that information with known or estimated characteristics of the template, e.g., nucleotide composition or sequence.

[0091] The subset of signal emissions detectable in the non-illuminated periods as compared to the number detectable in the illuminated periods is not limiting and may be chosen based upon the non-illumination data desired by the ordinary practitioner and/or other considerations, such as mitigation of photo-induced damage to extend readlength. For example, to lower the likelihood of photo-induced damage, the ordinary practitioner may choose to remove the illumination source that is most damaging, e.g., has the highest frequency. In certain embodiments, multiple sequencing reactions may be performed for a single amplified template, each with a different combination of illumination sources and/or detectable components. Alternatively or additionally, multiple replicate reactions can also be performed for one or more of the combinations of illumination sources and/or detectable components. The combination of data from multiple different and/or replicate reactions performed on a single template provides myriad benefits during statistical analysis. As noted above, data can be combined to facilitate assembly of contigs generated during illuminated periods. Data

from non-illuminated periods can also provide value in assessing the quality of the sequence reads generated during the illuminated periods.

[0092] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods used to provide additional data during the course of the reaction can comprise using labels that are incorporated into an enzyme of the reaction. For example, FRET labels can be used to label portions of a polymerase enzyme such that the conformational change between the open and closed states of the enzyme change the FRET value. For example, a FRET-based system can be used to monitor the kinetics of opening and closing of the finger subdomain of DNA polymerase, as described in Allen, et al. (2008) Protein Science 17:401-408, incorporated herein by reference in its entirety for all purposes. In certain preferred embodiments, a closed conformation produces a FRET signal because the donor and acceptor are close to one another, and an open conformation silences the signal because there is no energy transferred between the donor and acceptor. By monitoring the emission from the FRET pair, each incorporation event can be monitored during non-detection periods, and optionally or additionally during detection periods. In certain preferred embodiments, the FRET donor is GFP (excitation at 484 nm; emission at 510 nm), and the FRET acceptor is YFP (excitation at 512 nm; emission at 529 nm). Methods for monitoring polymerase activity using FRET labels are known in the art, e.g., in WO/2007/070572 A2, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

[0093] A given reaction may experience one or a plurality of illuminated periods or non-illuminated periods, but preferably experiences at least two illuminated periods. For example, a given reaction providing nucleotide sequence information from a single template nucleic acid may have at least about 2, 3, 5, 10, 20, 50, or 100 illuminated periods with intervening non-illuminated periods. In an embodiment employing multiple periods of illumination and/or non-illuminated, the periods may be the same for both, e.g., 100 seconds "on" and 100 seconds "off." Alternatively, the illuminated periods may be longer or shorter than the non-illuminated periods. For example, in certain embodiments, a non-illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent illuminated period; or an illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent non-illuminated period. Further, each illuminated period may be the same or different from each other illuminated period, and each non-illuminated period may be the same or different from each other non-illuminated period. For

example, some embodiments generate a smaller number of long reads, and other embodiments generate a larger number of short reads. It will be understood that the number and length of the illuminated and non-illuminated periods is limited only by the experimental system in use and the data acquisition goals of the ordinary practitioner. In some embodiments, a nucleotide sequence read generated during a single illuminated period comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, a region of a nucleic acid template processed during a non-illuminated period during a single reaction comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, the set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises at least about 40, 60, 80, 100, 1000, 10,000, 25,000, 50,000, 100,000, 250,000, 500,000, or 1,000,000 nucleotide sequence positions from a single nucleic acid template. In some embodiments, a set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises multiple reads of at least a portion of the nucleotide sequence positions from a single nucleic acid template.

[0094] As noted above, the present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.). However, such information, if available, may also be useful to the ordinary practitioner in determining an optimal periodicity for illuminated and non-illuminated periods during a sequencing reaction, especially when sequencing repetitive sequences. For example, if a genomic region is known to contain five adjacent copies of a one kilobase nucleotide sequence (i.e., five "repeat regions"), it would be beneficial to keep the non-illuminated periods short enough to be able to confidently map the resulting sequence reads to the correct repeat region. If a non-illuminated period were too long, the natural variation in translocation rate of the polymerase would make it difficult to assign a sequence read to a particular repeat region, especially those farther from the binding/initiation site of the polymerase. In a further example, if the "copies" each had a few mutations that could be used to distinguish them from each other, it would be beneficial to keep the illuminated periods long enough to increase the chance one of these mutations would be included in a resulting sequence read, thereby allowing the unambiguous assignment of the read to a particular repeat region. If the illuminated period were too short the sequence reads from two different repeat regions could be identical, making mapping the sequence

read challenging. (Another way to mitigate these difficulties would be to incorporate pause or stop points into the template nucleic acid, as discussed below.)

[0095] Essentially, the practitioner may design the number of and lengths of time for each illuminated and non-illuminated period to best suit the illuminated reactions being analyzed and the invention is not limited in this regard. In certain embodiments, a practitioner may wish to increase the processivity of a polymerase thereby extending the length of the template nucleic acid processed in a sequencing reaction to be, e.g., at least 2-, 3-, 4-, 6-, 8-, 10-, or 20-fold, thereby generating sequence data much farther away from the polymerase binding/initiation site than would be achieved under constant illumination. In certain embodiments, a practitioner of the instant invention may wish to focus on data from one or more stages of an ongoing reaction, such as stages for which more data is required for analysis. In the case of sequencing-by-synthesis, one or more particular regions of a template nucleic acid may need to be resequenced. Some traditional methods require that new template nucleic acids be prepared to bring a region requiring resequencing closer to the initiation point of the sequencing reaction, or require preparation of multiple new templates if multiple regions to be resequenced. In contrast, the methods herein allow the practitioner to subject a template identical to the previously sequenced template (e.g., from a large genomic DNA sample preparation) to a sequencing reaction wherein illuminated periods are timed to illuminate the sample only when the polymerase is incorporating nucleotides into the nascent strand at the one or more particular regions requiring resequencing. This advantage substantially lowers the time and resources required for such resequencing operations, therefore providing a significant advantage over traditional methods.

[0096] The instant invention contemplates various means for providing non-illuminated periods during illuminated reactions. In some embodiments, the illumination source is turned off during the ongoing reaction to create one or more non-illuminated periods. In some embodiments, the illumination source remains on during the course of the reaction, but the illuminated reaction is removed from the system for a period of time. In some embodiments, the illumination source remains on during the course of the reaction, but the illumination is blocked to create one or more non-illuminated periods. For example, a movable mask may be manually or mechanically positioned between the illumination source and the illuminated reaction to block the illumination during non-illuminated periods and removed to allow exposure to the illumination during illuminated periods. Such a mask may also be dynamically controlled, such as a thin film transistor

display (e.g., an LCD mask). Masks for blocking illumination and manufacture thereof are well known to those of ordinary skill in the art and need no further elaboration herein.

[0097] One aspect of the present invention is multiplexing of large numbers of single-molecule analyses. For a number of approaches, e.g., single molecule methods as described above, it may be desirable to provide the reaction components in individually optically resolvable configurations, such that a single reaction component or complex can be individually monitored. Providing such individually resolvable configurations can be accomplished through a number of mechanisms. For example, by providing a dilute solution of complexes on a substrate surface suited for immobilization, one will be able to provide individually optically resolvable complexes. (See, e.g., European Patent No.1105529 to Balasubramanian, et al., the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Alternatively, one may provide a low density activated surface to which complexes are coupled. (See, e.g., Published International Patent Application No. WO 2007/041394, the full disclosure of which is incorporated herein by reference in its entirety for all purposes). Such individual complexes may be provided on planar substrates or otherwise incorporated into other structures, e.g., zero-mode waveguides or waveguide arrays, to facilitate their observation.

[0098] In some embodiments, a plurality of illuminated reactions are carried out simultaneously, e.g., on a solid support. In some preferred embodiments, a solid support comprises an array of reaction sites. In preferred embodiments, the reaction sites on a solid support are optically resolvable from each other. In further preferred embodiments, each of the reaction sites on a solid support contains no more than a single reaction to be interrogated. For example, in a sequencing-by-incorporation embodiment, each reaction site preferably has no more than one polymerase and no more than one nucleic acid template. The reaction sites may be confinements (e.g., optical and/or physical confinements), each with an effective observation volume that permits resolution of individual molecules present at a concentration that is higher than one nanomolar, or higher than 100 nanomolar, or on the order of micromolar range. In certain preferred embodiments, each of the individual confinements yields an effective observation volume that permits resolution of individual molecules present at a physiologically relevant concentration, *e.g.*, at a concentration higher than about 1 micromolar, or higher than 50 micromolar range or even higher than 100 micromolar. In addition, for purposes of discussion herein, whether a particular reagent is confined by virtue of structural barriers to its free movement, or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined."

[0099] As used herein, a solid support may comprise any of a variety of formats, from planar substrates, e.g., glass slides or planar surfaces within a larger structure, e.g., a multi-well plates such as 96 well, 384 well and 1536 well plates or regularly spaced micro- or nano-porous substrates, or such substrates may comprise more irregular porous materials, such as membranes, aerogels, fibrous mats, or the like, or they may comprise particulate substrates, e.g., beads, spheres, metal or semiconductor nanoparticles, or the like. The solid support may comprise an array of one or more zero-mode waveguides or other nanoscale optical structures.

[00100] As used herein, "zero-mode waveguide" refers to an optical guide in which the majority of incident radiation is attenuated, preferably more than 80%, more preferably more than 90%, even more preferably more than 99% of the incident radiation is attenuated. As such high level of attenuation, no significant propagating modes of electromagnetic radiation exist in the guide. Consequently, the rapid decay of incident electromagnetic radiation at the entrance of such guide provides an extremely small observation volume effective to detect single molecules, even when they are present at a concentration as high as in the micromolar range. The fabrication and application of ZMWs in biochemical analysis, and methods for calling bases in sequencing-by-incorporation methods are described, e.g., in U.S. Patent Nos. 7,315,019, 6,917,726, 7,013,054, 7,181,122, and 7,292,742, U.S. Patent Pub. No. 2003/0174992, and U.S. Patent Application No. 12/134,186, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00101] A set of reactions (e.g., contained on a solid support) may comprise identical or different components. For example, a single template nucleic acid may be analyzed in all reactions in the set, or a plurality of template nucleic acids may be analyzed, each present in only one or a subset of the set of reactions. In preferred embodiments, template nucleic acids comprising the same nucleotide sequence are analyzed in a plurality of reactions sufficient to provide adequate redundant nucleotide sequence data to determine a consensus sequence for the template nucleic acids. A number of sequence reads that will provide adequate nucleotide sequence data will vary, depending, e.g., on the quality of the template nucleic acid and other components of the reaction, but in general coverage for a template nucleic acid or portion(s) thereof is at least about 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage. Further, the numbers and lengths of illuminated and non-illuminated periods for a given reaction in the set of reactions may be the same or different than those for other reactions in the set. In some embodiments, a mixture of different periodicities are used for a set of reactions comprising the same template nucleic acid. This strategy can be

beneficial for providing nucleotide sequence reads from varying regions of the template sequence, thereby increasing the likelihood of overlapping sequence reads between individual reactions. These overlapping sequence reads can facilitate construction of a more robust sequence scaffold than could be constructed were the reactions all subjected to the same periodicity of illuminated and non-illuminated periods.

[00102] Methods of controlling polymerase progress and/or synchronizing polymerases in different reactions are also useful in analysis (e.g., mapping, validation, etc.) of nucleic acid reads farther from the initial binding site of the polymerase. During detection periods earlier in the reaction (i.e., closer to the time at which the polymerase began to process the template nucleic acid, such as during a first illuminated period), the position of a polymerase on the template can be estimated with generally good accuracy based on the known translocation rate of the polymerase under a given set of reaction conditions. As the duration of the reaction increases, however, the natural variation in polymerase translocation rate makes it more difficult to accurately determine the exact position of the polymerase on a template using estimation based on translocation rate alone; and through each subsequent illuminated period such estimations of polymerase position become less accurate, making subsequent analysis and mapping of the sequence reads to the template more difficult. Methods of regulating the position of the polymerase on the template allow more accurate determinations the polymerase's position. For example, causing the polymerase to pause or stop at a given location on the template during a non-illuminated period and reinitiating the polymerization during or immediately prior to a subsequent illuminated period provides a way to reorient the subsequently generated read with the template sequence, allowing easier consensus sequence determination and mapping analyses. Further, such pause/stop points can provide a means of controlling what regions of the template are processed during the illuminated periods by restricting where the polymerase will reinitiate on the template, thereby allowing a practioner of the instant invention to target one or more particular regions of a template for analysis during one or more detection periods during the course of an analytical reaction. Such methods are also useful to synchronize a set reactions being monitored simultaneously. For example, a plurality of reactions, each comprising a single polymerase/template complex, may be synchronized by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that show less spreading (i.e., less variation in the position on the template from which the sequence reads are generated) in the later stages of the reactions than would otherwise be observed without such regulation.

[00103] Various methods can be used to control or monitor the progress of a polymerase on a template nucleic acid. For example, as noted above, one may employ a reaction stop or pause point within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the single-stranded portion that was not used in priming. Reaction stop or pause points can be engineered into a portion of the template for which the nucleotide sequence is unknown (e.g., a genomic fragment), but is preferably located within a portion for which the nucleotide sequence is known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, certain preferred sequencing templates (e.g., SMRTbell™ templates, described elsewhere herein) are closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, and one or both of these linkers can comprise a stop or pause point to control the passage of the polymerase through them. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00104] By way of example, at a selected time following initiation of polymerization the reaction may be subjected to a non-illuminated period. The incorporation of a synthesis blocking moiety coupled to the template nucleic acid at a position encountered by the polymerase during the non-illuminated period will cause the polymerase to pause. An example of an engineered pause point is a known sequence on the template nucleic acid where a primer sits and blocks progression of a polymerase that is actively synthesizing a complementary strand. The presence of the primer by itself could introduce a pause in the polymerase sequencing or the primer could be chemically modified to force a full stop (and synchronization of multiple polymerases in multiple reactions). The chemical modification could be subsequently removed (for example, photo-chemically) and the polymerase would subsequently continue along the template nucleic acid. In some embodiments, multiple primers could be included in a reaction to introduce multiple pause or stop points along the template nucleic acid. Other methods for inducing a reversible pause (stop) in synthesis are known in the art and include, e.g., reversible sequestering of required cofactors (e.g., Mn^{2+}, one or more nucleotides, etc.). Once sufficient time has passed that the polymerase is paused at the blocking group, illumination is reintroduced and the blocking group removed. This allows control of the position on the template nucleic acid at which the polymerase will begin generating nucleotide

sequence data during the illuminated period. A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the template nucleic acid that inhibit polymerase mediated replication, strand-binding moieties that prevent processive synthesis, non-native nucleotides included within the primer and/or the template, and the like. Such reaction stops/pause points are useful in providing more certainty about the relationship of the reads to each other. For example, since the exact position on a template nucleic acid at which each sequence read begins would be known, the resulting reads could be better mapped relative to one another for construction of a sequence scaffold and/or consensus sequence. Further description of these and other methods for regulating the progress of a polymerase on a template are provided, e.g., in U.S.S.N. 61/099,696, U.S. Patent Pub. No. 2006/0160113, and U.S. Patent Pub. No. 2008/0009007, all of which are incorporated by reference herein in their entireties for all purposes.)

[00106] By way of example, a sequencing reaction may be initiated on a template comprising a non-native base in the absence of the complement to the non-native base, which would not impact the overall sequence determination of other portions of the template that are complementary to native bases. By starving the reaction for the complement to the non-native base, one can prohibit synthesis, and thus, the sequencing process, until the non-native base complement is added to the mixture. This can provide a "hot start" capability for the system and/or an internal check on the sequencing process and progress that is configurable to not interfere with sequence analysis of the regions of interest in the template, which would be complementary to only native bases. In some embodiments, the non-native base complement in the sequence mixture is provided with a detectably different label than the complements to the four native bases in the sequence, and the production of incorporation-based signals associated with such labels provides an indication that the polymerase has initiated or reinitiated. Although described as the "non-native base" it will be appreciated that this may comprise a set of non-natural bases that can provide multiple control elements within the template structure. In certain embodiments, two different non-native bases are included within the template structure, but at different points, to regulate procession of the sequencing process, e.g., allowing controlled initiation and a controlled stop/start position later in the sequence, e.g., prior to a subsequent illuminated period. For example, the complement to the first non-native base can be added to initiate sequencing immediately prior to the start of a first illuminated period. During a first non-illuminated period following the first illuminated period, the polymerase encounters the second non-native base, e.g., at a nucleotide position near but upstream of a nucleotide region desired to be sequenced in a second illuminated period. Sequencing would

stop until the complement to the second non-native base is added to the reaction mixture. Likewise, multiple such non-native bases could be incorporated into the template to effectively target the polymerase to multiple regions of interest for which sequence data is desired. Further, in applications in which multiple identical templates are being sequenced, this would allow a resynchronization of the various sequencing reactions and the data generated therefrom.

[00107] Methods of controlling polymerase progress in different stages of a sequencing reaction are also useful for not only creating "condition-dependent" non-detection periods (during which time illumination may or may not be present), but also for minimizing the amount of time required for traversing a given length of template during a non-detection period (whether or not illumination is present). In order to reliably detect incorporation events, non-natural reagent conditions are typically used to limit polymerization during detection periods to approximately 1-5, or about 3 bases per second. In certain embodiments, replacement of Mg^{2+} ions with Mn^{2+} ions serves to stabilize and slow the translocation of the polymerase. When magnesium and, optionally, native nucleotides (e.g., lacking fluorescent labels) are used, the rate of translocation and/or processivity of the polymerase may increase up to two orders of magnitude. Use of such "rapid translocation" conditions during the non-detection periods can provide myriad benefits, including but not limited to a more rapid polymerization rate, an increased processivity (e.g., due to decreased stalling and misincorporation), and an overall savings due to reduced use of expensive labeled nucleotide analogs and/or reagents that mitigate oxidative stress.

[00108] In certain embodiments, a protocol for intermittent detection comprises alternating reaction mixtures, where a first reaction mixture used during the detection periods is optimized for sequence read generation, and a second reaction mixture used during the non-detection periods is optimized for processivity and/or rapid polymerization. For example, when reagents for optimal sequence read generation are present, DNA synthesis rate is low, and there is a fluorescence signal associated with each incorporation event. After replacing the reaction mixture optimized for sequence read generation with the reaction mixture optimized for processivity and/or rapid polymerization, the polymerase rapidly advances across the template. In certain embodiments, a flow cell is used to deliver and switch between the two (or more) reaction mixtures during the course of the reaction.

[00109] In an exemplary embodiment, a first reaction mixture comprises fluorescently-labeled nucleotide analogs and manganese ions that restrict polymerization to a rate appropriate for high fidelity detection of nucleotide incorporation. The first reaction mixture can also include

additional agents for mitigation of photo-induced damage of various components of the reaction mixture. A second reaction mixture comprises natural nucleotides and an appropriate magnesium ion concentration for rapid synthesis of the nascent strand complementary to the template. A first detection period of a sequencing reaction is initiated by introduction of the first reaction mixture, and a sequence read is generated based upon synthesis of the nascent strand during the detection period. After a predetermined time interval a sufficient quantity of the second reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, thereby initiating a first non-detection period. As noted above, the lack of labeled nucleotides in the second reaction mixture alone can produce the non-detection period, since there will be no signal emitted coincident with incorporation of the native nucleotides, but in certain embodiments illumination may also be removed, e.g., to further mitigate photo-induced damage during the non-detection period. At a time appropriate to initiate a second detection period, a sufficient quantity of the first reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, and detection of incorporation event is reinitiated. The cycle of reaction mixture exchange is repeated to generate multiple detection and non-detection periods.

[00110] A flow cell for reaction mixture exchange preferably has two inputs that are gated such that only a single reaction mixture flows into a reaction site or plurality of reaction sites, e.g., on a substrate. A single out-flow line may be used to remove reaction mixtures from the reaction site(s) to a single collection vessel, or multiple collection vessels may be used, one for each type of reaction mixture used. Further, accurate estimation of the distance a polymerase translocates during a non-detection period is important for bioinformatics applications. This estimation is complicated if the time for reaction mixture exchange is slow. As such, the flow is preferably at a sufficient rate that the time for exchange is significantly less than the time spent in the presence of either reaction mixture alone.

[00111] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination. In this embodiment, sixteen sequencing-by-incorporation reactions are performed on single nucleic acid templates (each of which comprises the same nucleotide sequence) with the timing of the illuminated and non-illuminated periods the same for all sixteen reactions. In A, the sixteen reactions are shown disposed on sixteen reaction sites on a solid support and are numbered for convenience. A representation of the illumination data is shown in B, with bars extending across the graph indicative of illumination data collected during illuminated

periods for each reaction. In this illustrative example, each reaction is subjected to three illuminated periods, each followed by a non-illuminated period, resulting in three noncontiguous sequence reads for each reaction, i.e., three noncontiguous reads per template molecule sequenced. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. During the first illuminated period, the sequence reads generally overlap, but the natural variation of polymerase translocation rate over the set of reactions results in a "spreading" of the sequence reads as the reaction proceeds through the second and third illuminated periods with increasing variation in the exact position of each polymerase on the template at the beginning and end of each illuminated period. As such, the earlier illumination data provides better redundancy ("oversampling") of sequence information over a relatively narrow portion of the template nucleic acid, while the later illuminated periods provide less redundant sequencing data over a broader region of the template nucleic acid. The timing of the non-illuminated periods between the illuminated periods and the known or calculated rate of incorporation are used to determine approximate spacing between the resulting sequence reads, providing context for building a sequence scaffold or consensus sequence. It is important to note that although shown disposed on a solid support in A, the data shown in B could also have been generated from reactions not disposed on a solid support nor performed simultaneously and the methods are generally not so limited. Further, as described above, the spreading of the sequence reads from later stages of the reactions can be mitigated by synchronizing the reactions, e.g., by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that provides better redundancy (i.e., more overlap in the positions on the template from which the sequence reads are generated), especially in the later stages of the reactions.

[00112] Using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be reset to change the regions of the template for which nucleotide sequence data is generated. As described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation

during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest. Alternatively, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be kept the same to provide a greater-fold coverage of one or more regions of interest in the template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which are incorporated by reference herein in their entireties for all purposes. A specific embodiment is also provided in the Exemplary Applications section herein.

[00113] The present invention provides novel template configurations and methods for exploiting these compositions in template directed sequencing processes. While these compositions and methods have utility across all of the various template directed processes described herein, for ease of discussion, they are being primarily discussed in terms of preferred single molecule, real-time sequencing processes, in which they provide myriad benefits. In particular, the present invention is generally directed to nucleic acid sequences that employ improved template sequences to improve the accuracy of sequencing processes. For example, in at least one aspect, the template compositions of the invention are generally characterized by the presence of a double stranded segment or a pair of sub-segments that are internally complementary, i.e., complementary to each other. In particular contexts, the target nucleic acid segment that is included within a template construct will typically be substantially comprised of a double stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double stranded or otherwise internally complementary.

[00114] Examples of template configurations of the invention that are partially and completely contiguous are schematically illustrated in Figure 20A and 20B, respectively. In particular, as shown in Figure 20A, a partially contiguous template sequence 200 is shown which includes a double stranded portion, comprised of two complementary segments 202 and 204, which,

for example, represent a target sequence or portion thereof. As shown, the 3' end of segment 202 is linked to the 5' end of segment 204 by linking oligonucleotide 206, providing a single stranded portion of the template, and yielding a partially contiguous sequence. By comparison, as shown in Figure 20B, a completely contiguous template sequence 210 is shown. Sequence 210 includes a double stranded portion again comprised of two complementary segments 212 and 214. As with the partially contiguous sequence of Figure 20A, the 3' end of segment 212 is joined to the 5' end of segment 214 via oligonucleotide 216 in a first single stranded portion. In addition, the 5' end of segment 212 is joined to the 3' end of segment 214 via linking oligonucleotide 218, providing a second single stranded portion, and yielding a completely contiguous or circular template sequence.

[00115] In addition, the templates of the invention, by virtue of their inclusion of double stranded segments, provide consensus through the identification of both the sense and antisense strand of such sequences (in both the partially and completely contiguous configurations).

[00116] By way of example, and with reference to Figures 20A and 20B, with respect to a partially contiguous template shown in Figure 20A, obtaining the entire sequence, e.g., that of segments 202, 204 and 206 provides a measure of consensus by virtue of having sequenced both sense strand, e.g., segment 202, and the antisense strand, e.g., segment 204. In addition to providing sense and antisense consensus within a single template molecule that can be sequenced in one integrated process, the presence of linking segment 206 also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., 202, is completed and the other begins, e.g., 204. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. The progress of sequencing processes is schematically illustrated in Figure 21A. In particular, as shown, a sequencing process that begins, e.g., is primed, at the open end of the partially contiguous template, proceeds along the first or sense strand, providing the nucleotide sequence (A) of that strand, as represented in the schematic sequence readout provided. The process then proceeds around the linking oligonucleotide of the template, providing the nucleotide sequence (B) of that segment. The process then continues along the antisense strand to the A sequence, and provides the nucleotide sequence (A'), which provides consensus data for the sense strand as its antisense counterpart. As noted, because the B sequence may be exogenously provided, and thus known, it may also provide a registration sequence indicating a point in the sequence determination at which the data transitions from sense to antisense strands.

[00117] With respect to completely contiguous or circular template sequences configured in accordance with the invention, the consensus potential is further increased. In particular, as with the partially contiguous sequences shown in Figure 20A, the completely contiguous sequences also provide sense and antisense consensus. In addition, such templates provide for the potential for iterative sequencing of the same molecule multiple times, by virtue of the circular configuration of the template. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining consensus for each segment from the complementary sequences, as well as consensus within each segment, by repeatedly sequencing that segment. This is schematically illustrated in Figure 21B, again with a representative illustration of a sequence readout provided. As shown, a sequencing process that is primed at one end, e.g., primed within one linking oligonucleotide sequence, e.g., linking oligonucleotide 218 of Figure 20, proceeds along the first or sense strand 214, again providing the nucleotide sequence A of that strand. The sequence process then proceeds around the first linking oligonucleotide, e.g., linking oligonucleotide 216 from Figure 20, to provide the nucleotide sequence B of that segment of the template. Proceeding along the antisense strand, e.g., segment 212 of Figure 20B), provides the nucleotide sequence A', which is again, complementary to sequence A. The sequencing process then continues around the template providing the nucleotide sequence for the other linking oligonucleotide, e.g., linking oligonucleotide 218 of Figure 20B, where the illustrated sequencing process began, providing nucleotide sequence C. Because the template is circular, this process can continue to provided multiple repeated sequence reads from the one template, e.g., shown as providing a second round of the sequence data A-B-A'-C-A-B-A'. Thus, sequence redundancy comes from both the determination of complementary sequences A and A', and the repeated sequencing of each segment. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing. Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post synthesis.

[00118] Another exemplary embodiment of an analysis of a plurality of illuminated reactions using intermittent illumination comprises a first illuminated period that is initiated at different times over the plurality of reactions. For example, the illuminated period for a first reaction may start at 0 seconds, the illuminated period for a second reaction may start at 5 seconds, the illuminated period for a third reaction may start at 10 seconds, and so forth. Additionally or alternatively, a first subset

of reactions may begin at a first time, a second subset may begin at a second time, etc. The first illuminated period continues for a given length of time, followed by a non-illuminated period and a subsequent second illuminated period. Optionally, a plurality of non-illuminated periods and illuminated periods follow the first illuminated period. Staggered start times can provide staggered data sets (e.g., two or more sequence reads) for the plurality of reactions, allowing multiple different stages of the overall reaction to be interrogated in different reactions. Preferably, the staggered data sets overlap to an extent that allows further analysis and validation of the reaction data. For example, a sequencing-by-incorporation reaction subjected to such an embodiment of the invention would preferably have sufficient overlap between sequence reads from different individual reactions to allow construction of a sequence scaffold and/or consensus sequence for a template nucleic acid.

[00119] A mask for use with a solid support (e.g., an array of confinements) can be designed to allow illumination of one or more portions of the solid support while blocking illumination to other portions of the solid support. For example, a mask may comprise one or more windows that allow excitation illumination to pass through the mask. Such a mask may be physically moved over the surface of the solid support (or the solid support can be moved relative to the mask), e.g., to selectively allow excitation illumination to reach a subset of confinements in an array. For example, a mask that allows 10% of reaction sites to be illuminated could be used to increase the sequencing scaffold coverage by sliding the illumination area (the area being subjected to excitation illumination) back and forth across the solid support. The 10% of reactions would cover certain regions of the nucleic acid template for any given time period (and therefore region of sequence in the template). In certain embodiments, an automated mask that selectively controls the timing of illumination of reactions on a solid support during the course of the reaction/acquisition may be used rather than a mask that must be physically moved.

[00120] The timing of the illuminated and non-illuminated periods for a set of reactions on a solid support may be the same or may vary, and may be synchronized or random. In certain embodiments in which the excitation illumination source is turned on and off, the timing of the illuminated and non-illuminated periods for the set of reactions will be identical. In other embodiments, for example, those that comprise use of a mask, the timing of the illuminated and non-illuminated periods for the set of reactions can vary so that while a subset of the reactions are illuminated, another subset of the reactions are not illuminated. Various exemplary and nonlimiting embodiments of masks that may be used with a set of reactions on a solid substrate are provided in Figures 3-5, as described below. In certain embodiments, the illuminated/non-illuminated status of

each reaction may be random across the solid support, e.g., to remove any experimental bias potentially introduced by actively selecting which reactions to illuminate at a given time, as long as the sequence reads being generated at the illuminated reactions and the time at which these reactions are not illuminated are able to be assigned to a particular reaction. For ease of discussion, the action of both illuminating and collecting emission signals from a reaction of interest, or a particular region on a solid support in which a reaction of interest is taking place, is referred to as "interrogating" that reaction and/or that region. A region being so interrogated is termed an "observation region."

[00121] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination and a mask. As in Figure 2, an array of reactions on a solid support 310 is provided containing sixteen reaction sites, numbered for convenience (A). In B, a mask 320 is provided with a single window 330 to allow passage of illumination to a subset of reactions on the solid support. Window 330 is wide enough to allow illumination of at least two columns of reaction sites on solid support 310. As in Figure 2, a representation of the illumination data is shown in C, with bars extending across the graph indicative of illumination data collected for each reaction. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. When the sequencing reaction is initiated at all positions on solid support 310, the window 330 is positioned to allow illumination to only reactions 1, 5, 9, and 13, and these four reactions provide sequence reads 350 for the earliest stage of the reactions. The window 330 is subsequently moved to provide an illuminated period for reactions 2, 6, 10, and 14 while still continuing the illuminated period for reactions 1, 5, 9, and 13. The illumination data for reactions 2, 6, 10, and 14 provides sequence reads 360, which partially overlap sequence reads 350 for reactions 1, 5, 9, and 13. The window 330 is moved again to provide illuminated periods for reactions 3, 7, 11, and 15 while still continuing the illuminated period for reactions 2, 6, 10, and 14, but removing illumination from reactions 1, 5, 9, and 13. The illumination data for 3, 7, 11, and 15 results in sequence reads 370, which partially overlap sequence reads 360 for reactions 2, 6, 10, and 14. A fourth position of the mask 320 initiates an illuminated period for reactions 4, 8, 12, and 16 while continuing illumination of reactions 3, 7, 11, and 15, but ending the illuminated period for reactions 2, 6, 10, and 14. Sequence reads 380 correspond to sequence reads from reactions 4, 8, 12, and 16. Finally, the window is moved to end the illuminated period for reactions 3, 7, 11, and 15 while continuing the illuminated period for reactions 4, 8, 12, and 16. Repeating the above process allows

a second read to be generated from each reaction, and this second read is noncontiguous with the first read. For example, reactions 1, 5, 9, and 13 correspond to reads 350 and, later in the reaction, reads 355. The two reads generated in a single reaction do not overlap and are separated by a length of nucleotides that was incorporated during the non-illuminated period between the two illuminated periods.

[00122] The mask can optionally be passed over the substrate additional times to generate additional reads until the reactions are complete or no longer provide reliable data, such as when the total illumination time (computed by summing the times for the multiple illuminated periods) has surpassed a photo-induced damage threshold period. Further, the mask may be passed back and forth, or may pass over the solid support in only one direction, e.g., always left to right, or vice versa.

[00123] Further, unlike the data shown in Figure 2B which has gaps in the sequence coverage for the template nucleic acid, the strategy provided in this embodiment results in at least two-fold coverage across the entire template nucleic acid (Figure 3C), although at a lower-fold redundancy. The portion of the template covered by only reads 380 and reads 355 has the least-fold redundancy, and in some instances a gap in coverage may be present in this region due to the movement of the mask 320 from the far right to the far left of the solid support 310. Of course, oversampling by adding replicate reactions to the set of reactions, or using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the coverage of a nucleic acid template, thereby providing more data for construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which have been previously incorporated by reference herein. The natural variation of polymerase translocation rate over the set of reactions is also apparent in this prophetic example as the spreading of the sequence reads and decreasing overlap between reads from reactions in adjacent columns in the later stages of the reactions as compared to the earlier stages.

[00124] Figure 4A provides an embodiment of a mask similar to that provided in Figure 3 except that it comprises three windows allowing multiple nonadjacent columns of reaction sites to be illuminated simultaneously. Figure 4B provides an embodiment of a mask comprising twelve windows, each of which allows illumination of a single reaction site on a solid support. The

windows are oriented in the mask to allow illumination of every other reaction in each row and every other reaction in each column. It will be understood that these mask designs are merely exemplary and nonlimiting embodiments as it is well within the abilities of the ordinary practitioner to determine an appropriate mask design depending on the experimental design or the illuminated reactions to be interrogated.

[00125] Figure 5B illustrates yet another aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Four different samples are disposed on a solid support, one in each quadrant 510, 520, 530, and 540 (A). A mask 550 is used that comprises two windows 560 that allow multiple rows of reaction sites to be illuminated simultaneously (B). A first position of this mask over a solid support in which two reactions in each quadrant are illuminated is demonstrated in C. A second position of the mask allowing illumination of the previously non-illuminated reactions is demonstrated in D. The mask may be moved back and forth as indicated by the double-arrow to provide multiple illuminated and non-illuminated periods for each reaction containing one of the four samples.

[00126] The present invention is also useful for redundant interrogation of reactions or portions of a solid support of interest. In certain aspects, sequential interrogation of different observation regions may be repeated a number of times, e.g., more than 2, 5, 10, 50, 100, 500, 1000, or even more than 10,000 times. In general, this method of stepping the observation region to another, preferably adjacent region, and repeating the interrogation process is generally referred to as a "step and repeat" process, and may be performed by various methods, including but not limited to moving the incident light and the solid support relative to one another and moving a mask across the surface of the solid support, as described above. Although described as a "step and repeat" method, in some embodiments where the observation region is moved across a substrate, that movement is not step-wise and iterative, but instead constitutes a continuous motion, substantially continuous motion, or stepped movement, or an iterative motion whereby each iterative step interrogates a new region that overlaps with some portion of the previously interrogated region. In particular, a substrate may be moved continuously relative to an optical system, whereby the observation region moves continuously across the substrate being interrogated (in a "scan mode").

[00127] The present invention is optionally combined with an optical system that provides illumination and/or collection of emitted illumination. Preferably, the optical system is operatively coupled to the reaction sites, e.g., on a solid support. One example of a particularly preferred optical

system is described in U.S.SN. 11/201,768, filed August 11, 2005, and incorporated herein by reference in its entirety for all purposes. Optical systems are described further below.

[00128] In some embodiments, one or both of the solid support and optical system are moved during interrogation. For example, a solid support being interrogated may be held stationary while the optical system is moved, or the solid support may be moved relative to a stationary optical system. Such movement may be accomplished using any of a variety of manipulation hardware or robotic set-ups, e.g., a stepper/feeder apparatus, and are well known in high performance printing technologies and in the semiconductor industry. For example, robotic systems may be used to pick up and re-orient a given solid support in order to interrogate different regions of the solid support, or make a previously unaccessible region (e.g., blocked by clips, support structure, or the like) of the solid support accessible. Such robotic systems are generally available from, e.g., Beckman, Inc., Tecan, Inc., Caliper Life Sciences, and the like.

[00129] In addition to the foregoing, it will be appreciated that the reagents in a given reaction of interest, including those reagents for which photo-induced damage is being mitigated in accordance with the invention, may be provided in any of a variety of different configurations. For example, they may be provided free in solution, or complexed with other materials, e.g., other reagents and/or solid supports. Likewise, such reagents may be provided coupled to beads, particles, nanocrystals or other nanoparticles, or they may be tethered to larger solid supports, such as matrices or planar surfaces. These reagents may be further coupled or complexed together with other reagents, or as separate reagent populations or even as individual molecules, e.g., that are detectably resolvable from other molecules within the reaction space. As noted above, whether a particular reagent is confined by virtue of structural barriers to its free movement or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined." Further examples of such confined reagents include surface immobilized or localized reagents, e.g., surface immobilized or associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflection microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. For example, in some preferred embodiments, one or more reagents in an assay system are confined within an optical confinement. Such an optical confinement may be an internal reflection confinement (IRC) or an external reflection confinement

(ERC), a zero-mode waveguide, or an alternative optical structure, such as one comprising porous film with reflective index media or a confinement using index matching solids. More detailed descriptions of various types of optical confinements are provided, e.g., in International Application Publication No. WO/2006/083751, incorporated herein by reference in its entirety for all purposes.

[00130] The invention is generally applicable to any of a variety of optical assays that require substantial illumination and/or photoactivated conversion or excitation of chemical groups, e.g., fluorophores. For example, the compositions and methods provided herein may be used with fluorescence microscopy, optical traps and tweezers, spectrophotometry, fluorescence correlation spectroscopy, confocal microscopy, near-field optical methods, fluorescence resonance energy transfer (FRET), structured illumination microscopy, total internal reflection fluorescence microscopy (TIRF), etc. The methods provided herein may be particularly useful in assays that are negatively impacted, directly or indirectly, by prolonged exposure to illumination. Of particular interest are those assays that are impaired by the generation and/or accumulation of triplet-state forms or free radicals during illumination.

[00131] One particularly apt example of analyses that benefit from the invention are single-molecule biological analyses, including, inter alia, single molecule nucleic acid sequencing analyses, single molecule enzyme analyses, hybridization assays (e.g., antibody assays), nucleic acid hybridization assays, and the like, where the reagents of primary import are subjected to prolonged illumination with relatively concentrated light sources (e.g., lasers and other concentrated light sources, such as mercury, xenon, halogen, or other lamps) in an environment where photoconversion/excitation is occurring with its associated generation of products. In certain embodiments, the methods, compositions, and systems are used in nucleic acid sequencing processes that rely on detection of fluorescent or fluorogenic reagents. Examples of such sequencing technologies include, for example, SMRT™ nucleic acid sequencing (described in, e.g., U.S. Patent Nos. 6,399,335, 6,056,661, 7,052,847, 7,033,764, 7,056,676, 7,361,466, 7,416,844, the full disclosures of which are incorporated herein by reference in their entirety for all purposes), non-real-time, or "one base at a time" sequencing methods available from, e.g., Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), Clonal Single Molecule Array™, and SOLiD™ sequencing. (See, e.g., Harris, et al. (2008) Science 320 (5872):106-9, incorporated by reference herein in its entirety for all purposes.) Such prolonged illumination can negatively impact (e.g., by introducing photo-induced damage) these reagents and diminish their effectiveness in the desired reaction.

III. Prevention of Photo-induced Damage

[00132] The methods provided herein are particularly useful in analyses that utilize very limited concentrations of reactants, such as single molecule detection/monitoring assays. As will be appreciated, in such reagent limited analyses, any loss, degradation, or depletion of a critical reagent will dramatically impact the analysis by further limiting the reagent, which not only can adversely effect the detectable signal, but may also directly impact the reaction being monitored, e.g., by changing its rate, duration, or product(s). For example, photo-induced damage can include a photoinduced change in a given reagent that reduces the reactivity of that reagent in the reaction, e.g., photobleaching of a fluorescent molecule, which diminishes or removes its ability to act as a signaling molecule. Also included in the term photo-induced damage are other changes that reduce a reactant's usefulness in a reaction, e.g., by making the reagent less specific in its activity in the reaction. Likewise, photo-induced damage includes undesired changes in a reagent that are caused by interaction of that reagent with a product of another photoinduced reaction, e.g., the generation of singlet oxygen during a fluorescence excitation event, which singlet oxygen may damage organic or other reagents, e.g., proteins. Photo-induced damage also includes downstream effects of damage to reactants, such as irreversible interactions between damaged reactants and other critical components of the reaction, e.g., reactive proteins or enzymes. For example, damage to an enzyme that catalyzes a reaction being monitored may cause a reduction in the rate of the reaction, in some cases stopping it altogether, or may reduce the duration or fidelity of the reaction.

[00133] As suggested by the foregoing, photo-induced damage generally refers to an alteration in a given reagent, reactant, or the like, that causes such reagent to have altered functionality in a desired reaction, e.g., reduced activity, reduced specificity, or a reduced ability to be acted upon, converted, or modified, by another molecule, that results from, either directly or indirectly, a photo-induced reaction, e.g., a photo-induced reaction creates a reactant that interacts with and causes damage to one or more other reactants. Typically, such photoreaction directly impacts either the reactant of interest, e.g., direct photo-induced damage, or impacts a reactant within one, two or three reactive steps of such reactant of interest. Further, such photoreaction can directly impact the reaction of interest, e.g., causing a change in rate, duration, processivity, or fidelity of the reaction.

[00134] The amount of time an illuminated analysis may be carried out before photo-induced damage so substantially impacts the reactants to render the analysis non-useful is referred to as the

"photo-induced damage threshold period." A photo-induced damage threshold period is assay-dependent, and is affected by various factors, including but not limited to characteristics of enzymes in the assay (e.g., susceptibility to photo-induced damage and the effect of such damage on enzyme activity/processivity), characteristics of the radiation source (e.g., wavelength, intensity), characteristics of the signal-generating molecule (e.g., type of emission, susceptibility to photo-induced damage, propensity to enter triplet state, and the effect of such damage on the brightness/duration of the signal), similar characteristics of other components of the assay. It can also depend on various components of the assay system, e.g., signal transmission and detection, data collection and analysis procedures, etc. It is well within the abilities of the ordinary practitioner to determine an acceptable photo-induced damage threshold period for a given assay, e.g., by monitoring the signal decay for the assay in the presence of a photodamaging agent and identifying a period for which the signal is a reliable measure for the assay. In terms of the invention, the photo-induced damage threshold period is that period of illuminated analysis during which such photo-induced damage occurs so as to reduce the rate or processivity of the subject reaction by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over the same reaction in the absence of such illumination. It is an object of the invention to increase the photo-induced damage threshold period, thereby increasing the amount of time reactions can proceed toward completion with minimal damage to the reactants, thereby lengthening the time in which the detectable signal is an accurate measure of reaction progression.

[00135] In some contexts, a "photo-induced damaged" reaction may be subject to spurious activity, and thus be more active than desired. In such cases, it will be appreciated that the photo-induced damage threshold period of interest would be characterized by that period of illuminated analysis during which such spurious activity, e.g., as measured by an increase in reaction rate, or an increase in non-specific reaction rate, is no more than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over a non-illuminated reaction. In one non-limiting example, where a nucleic acid polymerase, by virtue of a photodamaging event, begins to incorrectly incorporate nucleotides during template directed synthesis, such activity would impact the photo-induced damage threshold period as set forth above. In this case, the methods, devices, and systems of the invention would increase the photo-induced damage threshold period, thus increasing the amount of time the reaction could proceed before the above-described spurious activity occurred.

[00136] With reference to nucleic acid analyses, it has been observed that in template-directed synthesis of nucleic acids using fluorescent nucleotide analogs as a substrate, prolonged

illumination can result in a substantial degradation in the ability of the polymerase to synthesize the nascent strand of DNA, as described previously, e.g., in U.S. Published Patent Application No. 20070161017, incorporated by reference herein in its entirety for all purposes. Damage to polymerase enzymes, template sequences, and/or primer sequences can significantly hinder the ability of the polymerase to process longer strands of nucleic acids. For example, reduction in the processivity of a polymerase leads to a reduction in read lengths for sequencing processes that identify sequence constituents based upon their incorporation into the nascent strand. As is appreciated in the art of genetic analysis, the length of contiguous reads of sequence directly impacts the ability to assemble genomic information from segments of genomic DNA. Such a reduction in the activity of an enzyme can have significant effects on many different kinds of reactions in addition to sequencing reactions, such as ligations, cleavages, digestions, phosphorylations, etc.

[00137] Without being bound to a particular theory or mechanism of operation, it is believed that at least one cause of photo-induced damage to enzyme activity, particularly in the presence of fluorescent reagents, results from the direct interaction of the enzyme with photo-induced damaged fluorescent reagents. Further, it is believed that this photo-induced damage of the fluorescent reagents (and possibly additional damage to the enzyme) is at least partially mediated by reactive intermediates (e.g., reactive oxygen species) that are generated during the relaxation of triplet-state fluorophores. One or both of the photo-induced damaged fluorescent reagents and/or reactive intermediates may be included in the overall detrimental effects of photo-induced damage.

[00138] In certain aspects, the invention is directed to methods, devices, and systems that reduce the amount of photo-induced damage to one or more reactants during an illuminated reaction, e.g., thereby improving the reaction, e.g., by increasing the processivity, rate, fidelity, processivity, or duration of the reaction. In particular, methods are provided that yield a reduction in the level of photo-induced damage and/or an increase in the photo-induced damage threshold period as compared to such reactions in the absence of such methods, devices, and systems. In particular embodiments, such methods comprise subjecting an illuminated reaction to periods of non-illuminated during the course of the reaction, as described above, or by temporarily removing components of the reaction mixture that are believed to cause such damage, as described below.

[00139] As generally referred to herein, limited quantity reagents or reactants may be present in solution, but at very limited concentrations, e.g., less than 200 nM, in some cases less than 10 nM and in still other cases, less than 10 pM. In preferred aspects, however, such limited quantity

reagents or reactants refer to reactants that are immobilized or otherwise confined within a given area or reaction site (e.g., a zero-mode waveguide), so as to provide limited quantity of reagents in that given area, and in certain cases, provide small numbers of molecules of such reagents within that given area, e.g., from 1 to 1000 individual molecules, preferably between 1 and 10 molecules. As will be appreciated, photo-induced damage of immobilized reactants in a given area will have a substantial impact on the reactivity of that area, as other, non-damaged reactants are not free to diffuse into and mask the effects of such damage. Examples of immobilized reactants include surface-immobilized or -localized reagents, e.g., surface-immobilized or -associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflectance microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. Various types of solid supports upon which one or more reactants can be immobilized are described above.

[00140] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is interrogated for one or more illuminated periods that cumulatively are less than a photo-induced damage threshold period, as set forth elsewhere herein. Such interrogation may occur coincident with or independent of interrogation of additional observation regions on a solid support containing the first observation region. In accordance with the present invention, the observation region typically includes confined reagents (e.g., enzymes, substrates, etc.) that are susceptible to photo-induced damage, and may include an area of a planar or other solid support upon which confined reagents are immobilized. Alternatively or additionally, the observation region may include a physical confinement that constrains the reagents that are susceptible to photo-induced damage, including, e.g., microwells, nanowells, planar surfaces that include hydrophobic barriers to confine reagents.

[00141] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is intermittently interrogated under constant illumination by virtue of intermittent presence of detectable components of the reaction, wherein the presence of such detectable components has the potential to directly or indirectly cause photo-induced damage to one or more other reaction components. For example, a buffer comprising detectable components of a reaction can be temporarily replaced with a buffer comprising non-detectable versions of the same

components of the reaction, thereby interrupting data acquisition for the reaction. When data acquisition is to be recommenced, the buffer comprising detectable component is substituted for the buffer comprising non-detectable components. This substitution of reaction components may be repeated multiple times to generate multiple sets of data collected at noncontiguous stages of the reaction. For example, such a substitution can occur at least about 2, 4, 6, 8, or 10 times during the course of the reaction.

[00142] In certain preferred embodiments, the detectable components are fluorescently-labeled components that can be damaged by exposure to excitation illumination, and can further cause damage to other reaction components, as described above. For example, a sequencing-by-incorporation reaction can be initiated in the presence of fluorescently-labeled nucleotides whose incorporation is indicative of the nucleotide sequence of the nascent strand synthesized by a polymerase, and by complementarity, of the template nucleic acid molecule. At a selected time point during the ongoing reaction, the labeled nucleotides can be removed and replaced with unlabeled nucleotides, for example, by buffer exchange. After a period of time during which data acquisition has been interrupted by the absence of signal from the ongoing reaction, the labeled nucleotides can be reintroduced to reinitiate data acquisition. The labeled nucleotides may be removed and reintroduced multiple times and for various lengths of time, as preferred by the ordinary practitioner. In this way, multiple noncontiguous sequence reads can be generated from a single nucleic acid molecule in real time.

[00143] The methods herein slow the accumulation of photo-induced damage to one or more reagents, and may therefore indirectly mitigate the impact of photo-induced damage in an ongoing reaction of interest. By way of example, methods that reduce exposure of a critical enzyme component to illumination radiation (e.g., by subjecting the reaction to periods of non-illumination or by temporarily removing a component of the reaction responsible for such damage) do not necessarily prevent the photo-induced damage to the enzyme component, but rather extend the photo-induced damage threshold period by slowing the accumulation of photo-induced damage in the reaction mixture. Measurements of reduction of photo-induced damage as a result of implementation of intermittent illumination may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction subjected to constant illumination. Likewise, measurements of reduction of photo-induced damage as a result of temporary removal of reaction components responsible for such damage may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction in which such components are present

throughout. Further, characterization of a reduction in photo-induced damage generally utilizes a comparison of reaction rates, durations, or fidelities, processivities, e.g., of enzyme activity, and/or a comparison of the photo-induced damage threshold period, between a reaction mixture subjected to such the methods and/or systems of the invention and a reaction mixture not so subjected.

[00144] In the case of the present invention, implementation of the methods, devices, and systems of the invention generally results in a reduction of photo-induced damage of one or more reactants in a given reaction, as measured in terms of "prevented loss of reactivity" in the system. Using methods known in the art, the amount of prevented loss of activity can at least 10%, preferably greater than 20%, 30%, or 40%, and more preferably at least 50% reduction in loss of reactivity or increase in processivity, and in many cases greater than a 90% and up to and greater than 99% reduction in loss of reactivity or increase in processivity. By way of illustration, and purely for the purpose of example, when referring to reduction in photo-induced damage as a measure of enzyme activity in the presence and absence of intermittent illumination, if a reaction included a reaction mixture having 100 units of enzyme activity that would, under constant illumination, yield a reaction mixture having only 50 units of activity, then a 10% reduction in photo-induced damage would yield a final reaction mixture of 55 units (e.g., 10% of the 50 units otherwise lost, would no longer be lost). Further, use of the invention is expected to increase the performance (e.g., processivity, duration, fidelity, rate, etc.) of a reaction whose performance is negatively impacted by constant exposure to illumination by at least about 2-, 5-, 10-, 20-, 30-, 50-, 80-, 100-, 500-, or 1000-fold over that achieved by the reaction under constant illumination. For example, it is a specific object of the instant invention to increase the processivity of a polymerase enzyme in a sequencing reaction to allow collection of data across a longer length of the template.

[00145] With regards to sequencing applications, the methods herein facilitate the scaffolding of nucleic acid sequences in reactions susceptible to photo-induced damage. For example, if the sequencing device has 1000 base pair average readlength under constant illumination, one could subject the reaction to illuminated periods timed to allow approximately 100 nucleotides to be incorporated into the nascent strand of read, followed by non-illuminated periods timed to allow approximately 1000 nucleotides to be incorporated "in the dark." The sequence reads resulting from this experimental design would comprise about ten sequence reads of about 100 nucleotides each separated by gaps of about 1000 nucleotides each. If a plurality of sequencing reactions were carried out in this manner, and the illuminated periods were staggered appropriately, the reads from the plurality of reactions could be combined to provide nucleotide sequence data for the entire

template nucleic acid. This would potentially allow sequence scaffolds to be built much more easily than can be done with short-read systems, enabling structural analysis of previously impossible-to-sequence sections of highly repetitive DNA, given the sequencing system is capable of long reads in the absence of photodamage.

IV. Software and Algorithm Implementations

[00146] The methods herein may operate with numerous methods for sequence alignment including those generated by various types of known multiple sequence alignment (MSA) algorithms. For example, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read using a reference sequence. In some embodiments in which a reference sequence is known for the region containing the target sequence, the reference sequence can be used to produce an MSA using a variant of the center-star algorithm. Alternatively, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read relative to every other read without using a reference sequence ("*de novo* assembly routines"), e.g., PHRAP, CAP, ClustalW, T-Coffee, AMOS make-consensus, or other dynamic programming MSAs. Depending on the sequence-generating methods used, the determination of sequence alignment may also involve analysis of read quality (e.g., using TraceTuner™, Phred, etc.), signal intensity, peak data (e.g., height, width, proximity to neighboring peak(s), etc.), information indicative of the orientation of the read (e.g., 5'→3' designations), clear range identifiers indicative of the usable range of calls in the sequence, and the like. Additional algorithms and systems for sequence alignment are well know to those of skill in the art, and are described further, e.g., in G. A. Churchill, M. S. Waterman (1992) "The Accuracy of DNA Sequences: Estimating Sequence Quality," *Genomics* 14: 89-98; M. Stephens, et al. (2006) "Automating sequence-based detection and genotyping of SNPs from diploid samples," *Nat. Genet.*, **38**: 375-381; J. Hein (1989) *Mol. Biol. Evol.*, **6**: 649-668; U.S.S.N. 12/134,186, filed June 5, 2008; and U.S.S.N. 61/116,439, filed November 20, 2008.

[00147] A standard sequence alignment problem in the context of DNA sequencing is to align the sequence of a relatively short fragment (<2 kilobases) to a large target sequence. The assumption is made that this fragment represents a contiguous portion of DNA to be mapped to a single location on the reference sequence. (A "contiguous portion" to be mapped to a single location may contain small insertions and/or deletions and still be considered contiguous in this context.) With the further development of nucleic acid sequencing technologies (e.g., from Illumina,

Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), and Applied Biosystems, Inc. (Foster City, CA)) and mate-pair sequencing protocols (see, e.g., U.S. Patent Pub. No. 2006/0292611 A1, which is incorporated by reference herein in its entirety for all purposes), the alignment problem has been extended to align two fragments coming from the same read to the reference sequence using some knowledge of the expected mate-pair configuration (distance and orientation).

[00148] With regards to mate-paired reads, mapping two fragments with a distance constraint and orientation constraint has been treated by various short-read mapping algorithms, e.g., SOAP (Li, et al. (2008) Bioinformatics, **24**, 713-714); SOAPdenovo; and Maq, a set of programs that map and/or assemble fixed-length Solexa/SOLiD reads (SourceForge, Inc.). While these algorithms can handle simple cases of mate-pair alignment, which generally treat the specific problem of only two reads coming from a mate-paired sequence and use the distance constraint as a hard filter (i.e., if two reads are within *x* bp of each other and in the correct orientations, report them as a mate-pair hit), the methods provided herein are more general and can handle much more complex data sets, including those with multiple reads, those for which a reference sequence is or is not present, potential non-template sequence (e.g., adapter regions or linker portions described below), and complex distance and orientation constraints. Other programs are also available that attempt to generalize on top of the mapping and aligning performed by the programs described above. These include, e.g., Breakdancer, variationhunter, GASV, etc., which can handle more complex mappings, e.g., by clustering.

[00149] Real-time single molecule sequencing presents opportunities for obtaining much more complex sequence fragments from a single DNA sequencing read. Two examples are the reading of multiple discontiguous sequence fragments from a single long stretch of DNA using a pulsed or intermittent detection system (e.g., intermittent illumination) as described herein and the contiguous reading of forward, reverse and adapter fragments from a circular templates (SMRTbell™ templates; see e.g., U.S.S.N. 61/099,696, filed September 24, 2008; U.S. Patent Application No. 12/383,855, filed March 27, 2009 and U.S. Patent Application No. 12/413,258, filed March 27, 2009, all of which are incorporated by reference herein in their entireties for all purposes). Further, methods for sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent

Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.

[00150] Certain aspects of the invention provide methods for optimally aligning such sequences to a reference sequence using knowledge of the molecular configuration and/or sequencing protocol used to generate the related sequence reads. In particular, methods are provided to address the general problem of mapping multiple fragments to a reference sequence with variable distance and orientation constraints.

[00151] Beginning with raw sequence data generated by a nucleic acid sequencing instrument (step 1), the sequence data is mapped to a target sequence (step 2) using a local alignment method which produces sub-optimal local alignments as well as the optimal alignment, for example, the Smith-Waterman algorithm. Another, more flexible example of a local alignment method is a chaining method using a method for aligning very short fragments to the target sequence (e.g., kmer-indexing, suffix trees, suffix arrays, etc.) and chaining the resulting hits back into longer chains of significant matches (see, e.g. D. Gusfield, Algorithms on Strings, Trees, and Sequences, Cambridge University Press: Cambridge, UK, 1997, which is incorporated by reference herein in its entirety for all purposes). The chains do not necessarily need to be refined by dynamic programming in order to be useful for the following algorithm, permitting a very fast algorithm. In certain embodiments, dynamic-programming refinement of the chain might improve the power (area under the ROC curve) of the algorithm.

[00152] The target sequence consists of the potential hypotheses for the molecular template in question. In the example of nucleic acid sequencing methods using iterative illumination for sequencing a shotgun fragment from a linear DNA sequence, the potential hypotheses are both orientations of the genome (since we do not know the original orientation of the fragment). In the example of sequencing of a SMRTbell™ template (e.g., see Example 1 herein), the hypotheses include both orientations of the genome and known adapter sequences. The parameters determining how many hits are reported for each local fragment can be varied to change the specificity and sensitivity of this algorithm. Figure 6 shows what these hits might look like for a SMRTbell™ template (represented as paths in the sequence alignment matrix, which is often called the dynamic-programming matrix, although it isn't necessary to use dynamic programming to find these paths).

[00153] After the potential local alignments have been enumerated, a weighted directed graph is constructed with each local alignment represented as a node in the graph (step 3). The edges are drawn between nodes if they represent a potential reconstruction of the original molecular

template using knowledge of the expected molecular configuration. The directed connection of an alignment path A to an alignment path B is interpreted as "The target sequence represented by B could follow the target sequence represented by A in the original molecule." For example, if a linear single-stranded DNA molecule is being sequenced by a method that uses iterative illumination, then fragments from opposite orientations would not be expected to be connected (unless the linear single-stranded DNA molecule also included oppositely oriented sequences, e.g., as in the case of a linearized SMRTbell™ template.) In general, fragments that represent the same stretch of the sequencing read but that align to different regions on the target sequence would not be connected. Aside from these examples, the rules for connecting nodes should be fairly loose to permit exploration of weak possibilities that gain significance when all the evidence (e.g. all the sequence reads) are considered. The assignment of edge weights handles the proper weighting of the likelihood of these edges, and the speed of the algorithm can be tuned by optimizing the pruning of highly unlikely edges. As usual this represents a tradeoff between speed and sensitivity.

[00154] Weights are assigned to connections ($A \rightarrow B$) in the graph representing the log-likelihood that target fragment A is followed by target fragment B in the original molecule.

$$w(A \rightarrow B) = -\log P(B|A)$$

The conditional probability $P(B\ A)$ encodes the knowledge of the possible molecular configurations and the alignment significance of B.

$$P(B|A) = f(B)g(A,B)$$

where f is a measure of alignment significance (either theoretical or empirically obtained) and g encodes the physical constraints representing the allowed molecular configurations.

[00155] For example, in the context of sequencing using iterative illumination the following may be known: the time between the end of one fragment and the beginning of the next fragment is 200 seconds. If the polymerase incorporates bases with an average rate of 4bp/sec with a standard deviation of 1bp/sec, it can be hypothesized that the probability of target fragment 2 following target fragment 1 is determined by the distance between these fragments on the target and a normal probability:

$$g(A,B) = \frac{1}{\sqrt{2\pi(200)}} \exp\left[-(d-800)/2(200)^2\right]$$

[00156] In a SMRTbell™ template example, knowledge of the expected insert size and the observed distance and orientation between fragments would be used to weight the likelihood that

these two fragments could come from a correctly generated SMRTbell™ template. This weight could include the expected rate of the polymerase as well and rules for the orientation of fragments with respect to each other and their distance apart in the original read. For example, while it may be expected that two forward fragments mapping to the same region in the target genome potentially come from multiple passes around a SMRTbell™ template molecule, those fragments would not be expected to be immediately adjacent in sequencing time. The weighting function would account for the proper amount of expected time between such fragments (i.e. the elapsed time would be expected to be long enough to include two adapter sequences and a reverse sequence).

[00157] In general, the weighting function could be arbitrarily complex and tuned to empirically observed relationships between sequencing fragments given the available knowledge (distance between fragments on the target sequence, sequencing time between fragments, expected length of the template, etc.). For example, the empirical probability distributions might be observed to exhibit longer tails than a Gaussian probability model might predict. The use of a conditional log-likelihood for the assignment of edge weights is motivated by the following logic. In a graph of possible local alignments it is desirable to find a highly likely path that best explains the observed data. Consider a path through three nodes *A*, *B*, and *C*, with *P(ABC)* being the probability that *ABC* is the correct assignment:

$$P(ABC) = P(C|AB)P(B|A)P(A)$$
$$\approx P(C|B)P(B|A)P(A)$$

where the last approximation is justified by the observation that the constraints between allowable assignments to the target sequence are typically local in nature. Generalizing this formula for a path $a_1, \ldots, a_N$ and taking the negative logarithm of both sides gives

$$-\log P(a_1 \ldots a_N) = -\sum_{i=1}^{N-1} \log P(a_{i+1}|a_i) - \log P(a_1)$$

[00158] It is apparent that the edge weights are additive if we use log-likelihood and we can use standard shortest-path algorithms for directed graphs to find the optimal path. A hypothetical directed graph is illustrated in Figure 7. This graph corresponds to the situation depicted by the alignments pictured above. Heavier lines correspond to more likely paths with the optimal path shown in blue. Dashed lines represent forbidden transitions. Not all paths are considered in the illustration to avoid clutter in the presentation. The general formula listed above includes a "one-body" term $P(a_1)$ for the starting node in each path that weights the probability that this initial alignment is correct. To accommodate this probability in a path-finding algorithm we add a

pseudo-source s to the graph which connects to every possible node (not shown in the graph above). The edge weight connecting the pseudo-source with a node a_i is $-\log P(a_i)$. This allows the use of a conventional single-source shortest-path algorithm starting from the pseudo-source. The desired probability $P(a_i)$ can come from a measure of alignment significance (theoretical or empirically determined) or could be set uniformly across all alignments to allow the path logic to determine the best path assignment, independent of the relative value of the starting points. It is anticipated that a threshold will be required here to only allow edges between the pseudo-source and nodes for highly likely alignments; otherwise the shortest path algorithm in the next step will not give the desired path.

[00159] After construction of the weighted directed graph, the shortest path to each node is determined (step 4). The graph is directed and acyclic (DAG) so we can use the standard shortest-path DAG algorithm (see T.H. Cormen, CE Leiserson, RL Rivest, Introduction to Algorithms, MIT Press: Cambridge, Massachusetts, 1990). This algorithm scales as O(V+E) and should be very quick for these graphs. After the shortest path to each node is determined, the paths need to be ranked to declare the best assignment. It is suggested that the best metric would be a measure which rewards paths that explain more of the sequenced read (longer paths) with high likelihood. One such metric would be the normalized negative log-likelihood: dividing the total weight of the path by the number of bases in the sequenced read explained by this path. For more complicated graphs or edge-weight assignments, Dijkstra's algorithm, the Bellman-Ford algorithm, or the A* algorithm could be applied. Other algorithms that may also be used include, but are not limited to the Floyd-Warshall algorithm.

[00160] For noisy sequence data it is likely that the local alignments found in step 2 will occasionally overlap with each other in the sequenced read even though it is physically impossible for such overlaps to occur in a perfect system (unless there has been a rearrangement relative to the reference genome). As such, some amount of slack must be allowed in the edge assignment logic in step 3 to account for not knowing the precise boundaries of each local alignment. Once the best physical model explaining the observed read is determined, the boundaries of the local alignments can be refined to reflect the physical necessity that each base in the sequenced read can only be represented in one local alignment. It is also desirable to explain all of the bases in between the local alignments that haven't been assigned in the graph. One straightforward approach to refinement would be to construct the perfect model of the sequence and to realign the sequenced read to this sequence. This refinement algorithm would preserve physical constraints (each base in

the sequenced read can only be explained by one location in the template) and would assign all bases between the extremal nodes in the optimal path.

[00161] Certain aspects of the software and algorithm implementations described herein may be varied or altered without departing from the spirit and scope of the invention. For example, with regards to algorithm seeding, many algorithms can be applied for the original determination of sub-optimal local alignments (step 2). Conventional examples include FASTA, BLAST, or Smith-Waterman. It is expected that the best benefit will be obtained from using short-sequence alignment algorithms (suffix array, suffix tree, Boyer-Moore, Rabin-Karp, kmer-indexing, and the like) followed by chaining to establish regions of significant matches. An advantage of the algorithm described here is that it does not require dynamic-programming refinement of the resulting chains and therefore can be quite fast, however it is expected that using dynamic programming to refine the chains in step 2 could increase the power of the algorithm.

[00162] With regards to graph construction, there will be advantages to tuning the logic of edge assignments to keep the size of the graph manageable. It is possible that steps 2 and 3 might be combined to in a greedy fashion to focus the potentially slow step 2 into productive areas of the graph. For example, if a particularly strong hit is found early in step 2, then it may be beneficial to search for sub-optimal hits only in this local vicinity, knowing that this strong hit should be in the final solution. Tuning of the graph construction might include thresholds, below which edges are not created. Further, there are multiple parameters (minimum chain length, minimum probability for edge assignment, relative weighting of length vs. accuracy, etc.) which can be exposed and tuned in this algorithm to maximize the sensitivity and specificity of the algorithm for a given scenario.

[00163] With regards to determination of the distance a polymerase travels between reads, various strategies are provided that are more sophisticated than estimation based upon the rate of incorporation and the time between detection (e.g., illuminated) periods. In certain embodiments, the distribution of the base pair distance travelled by a polymerase during a non-detection period is called $p(x)$. The distribution of enzyme velocities, $p(v)$, is estimated by aligning observed reads to a reference sequence, and this distribution is represented as the number of reference bases per unit time. There is a length of time, τ, over which measurement of the instantaneous rate is not independent. While this method of determining the distance the polymerase travels during a non-detection period should not be overly sensitive to non-independent estimation of the polymerase rate, it is likely to strive for independent measurements of the rate. The distance τ can be estimated

from an exponential fit to the auto-correlation function $<\delta v(t)\delta v(t-\Delta t)>$, and $v(t)$ tabulated across the aligned sequence at increments of τ.

[00164] Where multiple single polymerase enzymes are being observed simultaneously, e.g., each being optically resolvable from every other on a single array, the $p(v)$ for each is preferably determined independently for each enzyme. Further, information regarding rare but extended events, such as polymerase "stalling" on the template, can be measured across a larger data set. For example, the statistics of stalls can be determined by aggregating rate measurements across an entire array. Where a stall distribution is characterized by a "long tail" corresponding to multi-exponential behavior of IPD distribution, such a distribution of polymerization rates can be extended for stalls longer than the observed reaction by fitting the long-tail behavior to an appropriate functional form, e.g., using a single-exponential parametric model or other physically motivated model (e.g., multi-exponential, stretched exponential, power-law, etc.) In certain preferred embodiments, the following representation of a "per-enzyme" $p(v)$ is used:

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}$$

where $f(v)$ is an interpolating function designed to retain information about the zero-velocity tail of the global $p(v)$ distribution while taking the estimate of the polymerase velocity dynamics(e.g., the dominant high velocity mode) from the specific enzyme. Such an interpolating function is:

$$f(v) = \frac{1 + erf(v / v_0)}{2}$$

where v_0 is a scale parameter to be chosen based on experience (but optionally fixed). Alternatively, the average of the empirical $p_{enzyme}(v)$ and $p_{array}(v)$ can be used. This approach can be motivated by a Bayesian approach to density estimation. Other kernel density and Bayesian methods can be suggested. Alternatively or in addition, the robustness of $p(v)$ to conditions and daily phenomenon can be explored and used to estimate $p(v)$ more globally, e.g., using one or more weekly control experiments.

[00165] Given the lack of a known reference sequence for *de novo* assembly, several alternative ways to formulate $p(v)$ are provided as follows. For example, in a first embodiment a control template (essentially a proxy reference sequence) can be subjected to sequencing, e.g., in the same reaction as the *de novo* sample or in an identical reaction. The observed velocity for the sequencing reactions would be measured based upon alignments of the reads from the control template to its known sequence. Typically, a per-enzyme correction would not be available for the

p(v) and *p(v)* would default to an array-averaged *p(v)*. In a second embodiment, a previously determined *p(v)* from experiments using a known reference sequence can be used, e.g., where the previous experiments were performed under the same conditions as the *de novo* experiments. In a third embodiment, *p(v)* is estimated by using quality information/metrics to screen for the most likely "true" calls, and restricting the estimates of *v* to regions containing those calls. In a fourth embodiment, where error is low, the called base rate and reference base rate converge to the same rate, and measurements of p(v) without knowledge of the reference become substantially reliable. Further, even if they do not fully converge, they can still be used to accurately infer *p(v)*,as long as the called base rate is predictably higher/lower than the reference base rate. Yet further, the measurement of *p(v)* when a reference sequence is not available can benefit from a detailed look at the probability model which is available from an algorithm like a CRF. That is, *p(v)* can be tabulated using a weighted sum over paths through a CRF probability model.

[00166] As will be clear to the ordinary practioner based upon the teachings herein, this framework extends naturally to the measurements of other potentially systematic variations in *p(v)* across an array, e.g., even where a single reaction mixture is applied to the entire array. For example, the local temperature of the reaction environment can vary systematically across an array of reactions. The average and variation in the rates of polymerase enzymes on the array would likely have a dependence on this hidden variable. Where the functional form of the temperature dependence is known, the measurement of *p(v)* can be stabilized across the array by modeling a de novo *p(v)* as $p_{cond}(v) + p_{x,y}(v)$ where *x,y* are geometrical variables defining the location on the array. Further, *p(v)* has been found to be somewhat variable over time. As such, in certain embodiments a model of *p(v;t)* is developed using an appropriate model for the evolution of *p(v)* over time.

[00167] Once a representative distribution of velocities *p(v)* has been obtained for a given read from a given reaction, the expected travel distance in the non-detection period can be expressed as:

$$p\left(\frac{x}{\tau}\right) = IL\left[L[p(v)]^{t/\tau} \frac{1 - L[p(v)]}{s} \right]$$

where L[] and IL[] stand for the Laplace and inverse Laplace transform, respectively. A similar result is derived in Svoboda, et al. (PNAS 91:11782 (1994)) and readily follows from considering the pdf of a sum of random variables. Optionally, in certain embodiments density estimation techniques (e.g., kernel density estimation, etc.) are useful when modeling *p(v)* since they can smooth the resulting numerical calculations in the Laplace and inverse Laplace transform.

[00168] Knowledge of the complete distribution has several advantages over the commonly applied Gaussian approximation. For example, knowledge of the complete distribution of insert lengths is very desirable when using a Bayesian framework approach to detect structural variation. (See, e.g., Bashir, et al. (2008) PLoS Comput. Biol. 4:51; Hormozdiari, et al. (2009) Genome Res. 19:1270; and Lee, et al. (2008) Bioinformatics 24:59.) While Bashir, et al. does not strictly follow a Bayesian approach, the geometric approach described in the paper can be straightforwardly modified to incorporate an actual posterior instead of the boxcar posterior assumed in the paper. Further, during mapping of noncontiguous reads to a genome where they are expected to be concordant (*i.e.*, not a structural variation), it is useful to consider the known distribution when judging the significance of the resulting alignments between the observed reads and the genomic sequence. In addition, when clustering noncontiguous reads that scaffold contigs in a *de novo* assembly, a path of Bayesian significance can be followed that is very similar to that followed in the structural variation case discussed *supra*.

[00169] In further embodiments, the determination of the distance a polymerase travels between reads is performed using an algorithm based on a simulation approach rather than the exact analytical result used in the algorithm described above. This method relies on Monte Carlo sampling from a distribution, which allows a better extension to arbitrary empirical distributions. It also lacks the difficult computations of numerical Laplace and inverse Laplace transforms, and permits calculation of distances traveled during non-detection periods when the underlying kinetic processes have multi-phasic kinetics, e.g., the presence of long stalls.

[00170] This approach aims to calculate the distribution of the distance x travelled by an enzyme during a time t during which it was not being observed (e.g., during a non-detection period). In some embodiments, a distribution of local rates, $p(v)$, is estimated, where the definition of "local" is set by the correlation length of the rate autocorrelation function, e.g.:

$$\langle \delta v(t)\delta v(t+\Delta)\rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$$

Given a local rate distribution and an assumption that independent identically distributed (i.i.d.) draws can be made from this distribution, one approach to calculating the distribution is as follows. First, draw $N = t/\tau_{corr}$ velocities from $p(v)$, and subsequently sum them and record them as an estimate of x/τ_{corr}. Repeat the process M times, with the optimal choice of M dependent on the desired level of precision for estimation of the $p(x)$ distribution. In certain preferred embodiments,

M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000.

[00171] In some embodiments in which the enzyme system is not well explained by a single kinetic process or cycle (as in the case of observed stalling behavior), above-described rate autocorrelation function and the i.i.d. assumption will be violated. As such, a probability model having a richer structure is preferably used. One such probability model is a Hidden Markov Model (HMM). Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system. Where the kinetics of the pausing state can be well described by a single-exponential, this model is expected to describe the observed distribution of local velocities. The single-exponential assumption is implicit in the state structure of the model since the amount of time spent in the pause state will be a geometric distribution with mean p/(1-p) [*i.e.*, the observed stall times will have to be added to this model]. If the stall kinetics are multi-phasic, then more "dark states" will have to be added to this model. Further, the model shown in Figure 15 can actually be treated as a Markov Model and not a Hidden Markov Model without much loss of generality because the "pause" state is not actually hidden due to the fact that the data collected during the pause state is highly distinguishable from the data collected during the sequencing state. As such, the general HMM apparatus is not necessary. The model in Figure 15 can be used to simulate the distribution of local velocities when there is a long-term pause or stall phase present in the reaction data kinetics. S0 is the start state, and there is no explicit end state since this model is used as a generative model and it is assumed that it is run forward for a prescribed number of steps. The qualities P(P→S) and P(S→P) represent exit from a stalled state and entry into a stalled state, respectively. These qualities can be measured by an EM algorithm or they can be quickly estimated by physical observables. $P_{P\to S} = 1/\left(1+\frac{\tau_{stall}}{\tau_{corr}}\right)$ and P(S→P) is the frequency of stall starts per τ_{corr}. (Example parameters are τ_{stall}=80 seconds; τ_{corr} = 10 seconds; and P(S→P) = 1/24.) The simulation estimate of *p(x)* can now be produced using the procedure outlined above in which $N = t/\tau_{corr}$ velocities are drawn from *p(v)*; and they are subsequently summed and recorded as an estimate of x/τ_{corr}. The process is repeated M times, with the optimal choice of M dependent on the desired level of precision for estimation of the *p(x)* distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000. Figure 16 shows exemplary simulated applications of this method. Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters P(S→P) = 1/24; P(P→S) = 1/11;

and $p(v)$~Gamma(48,0.25). Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a 1300 second non-detection period, which is calculated by sampling 2000 estimates from the HMM model.

[00172] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period. The influence of pause/stall behavior can be seen in the heavy-left tailing of both distributions.

[00173] While the simulation method in which i.i.d. draw assumption is valid is more general and can treat arbitrary $p(v)$ and more complex models for non-sequencing states, the two-state model using the HMM can be treated analytically. The result of this is:

$$p(x/\tau_{corr}) = \sum_{N_S=0}^{N} \pi_{N_S}(x) p_N(N_S)$$

where $\pi_{N_S}(x)$ is the distribution of the sum of N_S variables drawn from $p(v)$. For the general case, this distribution is given by the Laplace transform approach presented above. For $p(v)$~Normal(μ,σ), this distribution is distributed as Normal($N_S\mu,\sqrt{N_S}\sigma$). For $p(v)$~Gamma(k,θ), this distribution is distributed as Gamma(N_Sk,θ). $P_N(N_S)$ is the number of cycles spent in the sequencing state if we observe N cycles from the Markov process in Figure 15. The expression for this is described in Pedler, et al. (1971) J. Appl. Prob. 8:381, which is incorporated herein by reference in its entirety for all purposes.

[00174] As will be clear to one of ordinary skill in the art upon review of the teachings herein, these methods can be readily extended to the non-detection period estimations of procession by other cyclical biological reactions, such as the action of reverse transcriptase or the synthesis of proteins by a ribosome complex, e.g., and certain preferred embodiments of such reactions are further described in U.S.S.N. 12/767,673, filed April 26, 2010; and U.S.S.N. 12/813,968, filed June 11, 2010, the disclosures of which are incorporated herein by reference in their entireties for all purposes. Further, the simulation model described above is not restricted to simple two-state kinetics, and the use of $p(v)$ is not restricted to analytical models. In fact, in certain embodiments, empirical estimates are preferably used.

[00175] Although useful in certain preferred embodiments of the invention, certain algorithms as presented above do not easily handle the case where the template does not match a physically-motivated expected model. A relevant example of such a case is when the template contains a genomic structural variation (SV), such as translocation, whereby two fragments which are correctly adjacent in the template are located very far apart in the reference genome. Such

structural variation cases are best handled in the context of the current algorithm by reporting the confidence of an observed path and reporting situations when no physically expected path seems to fit the observed data. In general, the detection of structural variation requires the presence of multiple highly significant local alignments which can be identified as significantly overturning the null hypothesis of matching the genomic ordering of fragments with their own individual merit. Nevertheless, with molecular redundant sequencing such as SMRTbell™ template sequencing the current algorithm can be adapted to improve the ability to identify an SV event. Such a modification could be a feedback approach which allows modification of the linking constraints in step 3 to allow very far separations on the target sequence when the individual alignments are very significant. Only one such highly-significant pair would be needed to enable the rescue of less significant partial matches that support the same SV hypothesis.

[00176] The software and algorithm implementations provided herein are particularly suited for transforming sequence read data generated from various sequencing technologies (e.g., sequencing-by-synthesis, intramolecular redundant sequencing, Sanger sequencing, capillary electrophoretic sequencing, pyrosequencing, ligase-mediated sequencing, etc.) into consensus sequence data that provides a representation of the actual nucleotide sequence of the template nucleic acid that was subjected to the sequencing reaction(s) from which the sequence read data was generated. The software and algorithm implementations provided herein are preferably machine-implemented methods. The various steps recited herein are preferably performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. The results of these methods are preferably stored on a machine-readable medium, as well. Further, the invention provides a computer program product comprising a computer usable medium having a computer readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention.

[00177] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data).

[00178] While described with reference to certain specific applications above, it will be understood that these methods are also applicable to other types of complex data sets, and the invention should not be limited to only the specific examples provided herein. Other applications of the instant methods will be clear to those of ordinary skill in the art and are considered to be additional aspects of the instant invention.

V. Devices and Systems

[00179] The invention also provides systems that are used in conjunction with the compositions and methods of the invention in order to provide for intermittent detection of analytical reactions. In particular, such systems typically include the reagent systems described herein, in conjunction with an analytical system, e.g., for detecting data from those reagent systems. For example, a sequencing reaction may be subjected to intermittent illumination, and the sequencing system may include the system components provided with or sold for use with commercially available nucleic acid sequencing systems, such as the Genome Analyzer System available from Illumina, Inc., the GS FLX System, available from 454 Life Sciences, or the ABI 3730 System available from Life Technologies, Inc.

[00180] In certain preferred embodiments, reactions subjected to intermittent illumination are monitored using an optical system capable of detecting and/or monitoring interactions between reactants at the single-molecule level. Such an optical system achieves these functions by first generating and transmitting an incident wavelength to the reactants, followed by collecting and analyzing the optical signals from the reactants. Such systems typically employ an optical train that directs signals from the reactions to a detector, and in certain embodiments in which a plurality of reactions is disposed on a solid surface, such systems typically direct signals from the solid surface (e.g., array of confinements) onto different locations of an array-based detector to simultaneously detect multiple different optical signals from each of multiple different reactions. In particular, the optical trains typically include optical gratings or wedge prisms to simultaneously direct and separate signals having differing spectral characteristics from each confinement in an array to different locations on an array based detector, *e.g.*, a CCD, and may also comprise additional optical transmission elements and optical reflection elements.

[00181] An optical system applicable for use with the present invention preferably comprises at least an excitation source and a photon detector. The excitation source generates and transmits incident light used to optically excite the reactants in the reaction. Depending on the intended

application, the source of the incident light can be a laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source. Further, the excitation light may be evanescent light, e.g., as in total internal reflection microscopy, certain types of waveguides that carry light to a reaction site (see, e.g., U.S. Application Pub. Nos. 20080128627, 20080152281, and 200801552280), or zero-mode waveguides, described below. Where desired, more than one source can be employed simultaneously. The use of multiple sources is particularly desirable in applications that employ multiple different reagent compounds having differing excitation spectra, consequently allowing detection of more than one fluorescent signal to track the interactions of more than one or one type of molecules simultaneously. A wide variety of photon detectors or detector arrays are available in the art. Representative detectors include but are not limited to optical reader, high-efficiency photon detection system, photodiode (*e.g.* avalanche photo diodes (APD)), camera, charge couple device (CCD), electron-multiplying charge-coupled device (EMCCD), intensified charge coupled device (ICCD), and confocal microscope equipped with any of the foregoing detectors. For example, in some embodiments an optical train includes a fluorescence microscope capable of resolving fluorescent signals from individual sequencing complexes. Where desired, the subject arrays of optical confinements contain various alignment aides or keys to facilitate a proper spatial placement of the optical confinement and the excitation sources, the photon detectors, or the optical train as described below.

[00182] The subject optical system may also include an optical train whose function can be manifold and may comprise one or more optical transmission or reflection elements. Such optical trains preferably encompass a variety of optical devices that channel light from one location to another in either an altered or unaltered state. First, the optical train collects and/or directs the incident wavelength to the reaction site (e.g., optical confinement). Second, it transmits and/or directs the optical signals emitted from the reactants to the photon detector. Third, it may select and/or modify the optical properties of the incident wavelengths or the emitted wavelengths from the reactants. In certain embodiments, the optical train controls an on/off cycle of the illumination source to provide illuminated and non-illuminated periods to one or more illuminated reaction sites. Illustrative examples of such optical transmission or reflection elements are diffraction gratings, arrayed waveguide gratings (AWG), optic fibers, optical switches, mirrors (including dichroic mirrors), lenses (including microlenses, nanolenses, objective lenses, imaging lenses, and the like), collimators, optical attenuators, filters (*e.g.*, polarization or dichroic filters), prisms, wavelength filters (low-pass, band-pass, or high-pass), planar waveguides, wave-plates, delay lines, and any

other devices that guide the transmission of light through proper refractive indices and geometries. One example of a particularly preferred optical train is described in U.S. Patent Pub. No. 20070036511, filed August 11, 2005, and incorporated by reference herein in its entirety for all purposes.

[00183] In a preferred embodiment, a reaction site (e.g., optical confinement) containing a reaction of interest is operatively coupled to a photon detector. The reaction site and the respective detector can be spatially aligned (*e.g.*, 1:1 mapping) to permit an efficient collection of optical signals from the reactants. In certain preferred embodiments, a reaction substrate is disposed upon a translation stage, which is typically coupled to appropriate robotics to provide lateral translation of the substrate in two dimensions over a fixed optical train. Alternative embodiments could couple the translation system to the optical train to move that aspect of the system relative to the substrate. For example, a translation stage provide a means of removing a reaction substrate (or a portion thereof) out of the path of illumination to create a non-illuminated period for the reaction substrate (or a portion thereof), and returning the substrate at a later time to initiate a subsequent illuminated period. An exemplary embodiment is provided in U.S. Patent Pub. No. 20070161017, filed December 1, 2006.

[00184] In particularly preferred aspects, such systems include arrays of reaction regions, e.g, zero-mode waveguide arrays, that are illuminated by the system, in order to detect signals (e.g., fluorescent signals) therefrom, that are in conjunction with analytical reactions being carried out within each reaction region. Each individual reaction region can be operatively coupled to a respective microlens or a nanolens, preferably spatially aligned to optimize the signal collection efficiency. Alternatively, a combination of an objective lens, a spectral filter set or prism for resolving signals of different wavelengths, and an imaging lens can be used in an optical train, to direct optical signals from each confinement to an array detector, *e.g.*, a CCD, and concurrently separate signals from each different confinement into multiple constituent signal elements, *e.g.*, different wavelength spectra, that correspond to different reaction events occurring within each confinement. In preferred embodiments, the setup further comprises means to control illumination of each confinement, and such means may be a feature of the optical system or may be found elsewhere is the system, e.g., as a mask positioned over an array of confinements. Detailed descriptions of such optical systems are provided, e.g, in U.S. Patent Pub. No. 20060063264, filed September 16, 2005, which is incorporated herein by reference in its entirety for all purposes.

[00185] The systems of the invention also typically include information processors or computers operably coupled to the detection portions of the systems, in order to store the signal data obtained from the detector(s) on a computer readable medium, e.g., hard disk, CD, DVD or other optical medium, flash memory device, or the like. For purposes of this aspect of the invention, such operable connection provide for the electronic transfer of data from the detection system to the processor for subsequent analysis and conversion. Operable connections may be accomplished through any of a variety of well known computer networking or connecting methods, e.g., Firewire®, USB connections, wireless connections, WAN or LAN connections, or other connections that preferably include high data transfer rates. The computers also typically include software that analyzes the raw signal data, identifies signal pulses that are likely associated with incorporation events, and identifies bases incorporated during the sequencing reaction, in order to convert or transform the raw signal data into user interpretable sequence data (See, e.g., Published U.S. Patent Application No. 2009-0024331, the full disclosure of which is incorporated herein by reference in its entirety for all purposes).

[00186] Exemplary systems are described in detail in, e.g., U.S. Patent Application No. 11/901,273, filed September 14, 2007 and U.S. Patent Application No. 12/134,186, filed June 5, 2008, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00187] Further, as noted above, the invention provides data processing systems for transforming sequence read data into consensus sequence data. In certain embodiments, the data processing systems include machines for generating sequence read data by interrogating a template nucleic acid molecule. In certain preferred embodiments, the machine generates the sequence read data using a sequencing-by-synthesis technology, as described elsewhere herein, but the machine may generate the sequence read data using other sequencing technologies known to those of ordinary skill in the art, e.g., pyrosequencing, ligation-mediated sequencing, Sanger sequencing, capillary electrophoretic sequencing, etc. Such machines and methods for using them are available to the ordinary practioner.

[00188] The sequence read data generated is representative of the nucleotide sequence of the template nucleic acid molecule only to the extent that a given sequencing technology is able to generate such data, and so may not be identical to the actual sequence of the template nucleic acid molecule. For example, it may contain a deletion or a different base at a given position as compared to the actual sequence of the template, e.g., when a base call is missed or incorrect, respectively. As

such, it is beneficial to generate redundant sequence read data, and the methods described herein provide manipulations and computations that transform redundant sequence read data into consensus sequence data that is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Redundant sequence read data comprises multiple reads, each of which includes at least a portion of sequence read that overlaps with at least a portion of at least one other of the multiple reads. As such, the multiple reads need not all overlap with one another, and a first subset may overlap for a different portion of the template nucleic acid sequence than does a second subset. Such redundant sequence read data can be generated by various methods, including repeated sequencing of a single nucleic acid template, sequencing of multiple identical nucleic acid templates, or a combination thereof.

[00189] In another aspect, the data processing systems can include software and algorithm implementations provided herein, e.g. those configured to transform redundant sequence read data into consensus sequence data, which, as noted above, is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Further, the transformation of the redundant sequence read data into consensus sequence data identifies and negates some or all of the single-read variation between the multiple reads in the redundant sequence read data. As such, the transformation provides a representation of the actual nucleotide sequence of the nucleic acid template from which redundant sequence read data is generated that is more accurate than a representation based on a single read.

[00190] The software and algorithm implementations provided herein are preferably machine-implemented methods, e.g., carried out on a machine comprising computer-readable medium configured to carry out various aspects of the methods herein. For example, the computer-readable medium preferably comprises at least one or more of the following: a) a user interface; b) memory for storing redundant sequence read data; c) memory storing software-implemented instructions for carrying out the algorithms for transforming redundant sequence read data into consensus sequence data; d) a processor for executing the instructions; e) software for recording the results of the transformation into memory; and f) memory for recordation and storage of the resulting consensus sequence read data. In preferred embodiments, the user interface is used by the practitioner to manage various aspects of the machine, e.g., to direct the machine to carry out the various steps in the transformation of redundant sequence read data into consensus sequence data,

recordation of the results of the transformation, and management of the consensus sequence data stored in memory.

[00191] As such, in preferred embodiments, the methods further comprise a transformation of the computer-readable medium by recordation of the redundant sequence read data and/or the consensus sequence data generated by the methods. Further, the computer-readable medium may comprise software for providing a graphical representation of the redundant sequence read data and/or the consensus sequence read data, and the graphical representation may be provided, e.g., in soft-copy (e.g., on an electronic display) and/or hard-copy (e.g., on a print-out) form.

[00192] The invention also provides a computer program product comprising a computer-readable medium having a computer-readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention. In certain preferred embodiments, the computer program product comprises the computer-readable medium described above.

[00193] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data), such as the computer-readable medium described above.

[00194] As shown in Figure 9, the system 900 includes a substrate 902 that includes a plurality of discrete sources of chromophore emission signals, e.g., an array of zero-mode waveguides 904. An excitation illumination source, e.g., laser 906, is provided in the system and is positioned to direct excitation radiation at the various signal sources. This is typically done by directing excitation radiation at or through appropriate optical components, e.g., dichroic 108 and objective lens 910, that direct the excitation radiation at the substrate 902, and particularly the signal sources 904. Emitted signals from the sources 904 are then collected by the optical components, e.g., objective 910, and passed through additional optical elements, e.g., dichroic 908, prism 912 and lens 914, until they are directed to and impinge upon an optical detection system, e.g., detector array 916. The signals are then detected by detector array 916, and the data from that detection is

transmitted to an appropriate data processing system, e.g., computer 918, where the data is subjected to interpretation, analysis, and ultimately presented in a user ready format, e.g., on display 920, or printout 922, from printer 924. As will be appreciated, a variety of modifications may be made to such systems, including, for example, the use of multiplexing components to direct multiple discrete beams at different locations on the substrate, the use of spatial filter components, such as confocal masks, to filter out-of focus components, beam shaping elements to modify the spot configuration incident upon the substrates, and the like (See, e.g., Published U.S. Patent Application Nos. 2007/0036511 and 2007/095119, and U.S. Patent Application No. 11/901,273, all of which are incorporated herein by reference in their entireties for all purposes.)

VI. Exemplary Applications

[00195] The methods and compositions of the invention are useful in a broad range of analytical reactions in which one or more aspects of a detection method are detrimental to one or more aspects of the analytical reaction, such as rate, duration, fidelity, processivity, and the like. In such cases, intermittent detection at least partially mitigates the detrimental effect while allowing collection of data from stages of the analytical reaction that were previously uncollectable. As noted above, illuminated reactions are one example of analytical reactions that benefit from the compositions and methods described herein, particularly those using photoluminescent or fluorescent reagents, and particularly such reactions where one or more of the reaction components that are susceptible to photo-induced damage are present at relatively low levels. One exemplary application of the methods and compositions described herein is in single molecule analytical reactions, where the reaction of a single molecule (or very limited number of molecules) is observed in the analysis, such as observation of the action of a single enzyme molecule. In another aspect, the present invention is directed to illuminated reactions for single molecule analysis, including sequencing of nucleic acids by observing incorporation of nucleotides into a nascent nucleic acid sequence during template-directed polymerase-based synthesis. Such methods, generally referred to as "sequencing-by-incorporation" or "sequencing-by-synthesis," involve the observation of the addition of nucleotides or nucleotide analogs in a template-dependent fashion in order to determine the sequence of the template strand. See, e.g., U.S. Patent Nos. 6,780,591, 7,037,687, 7,344,865, 7,302,146. Processes for performing this detection include the use of fluorescently labeled nucleotide analogs within a confined observation region, e.g., within a nanoscale well and/or tethered, either directly or indirectly to a surface. By using excitation illumination (i.e., illumination

of an appropriate wavelength to excite the fluorescent label and induce a detectable signal), the fluorescently labeled bases can be detected as they are incorporated into the nascent strand, thus identifying the nature of the incorporated base, and as a result, the complementary base in the template strand.

[00196] In particular aspects, when an analysis relies upon a small population of reagent molecules, damage to any significant fraction of that population will have a substantial impact on the analysis being performed. For example, prolonged interrogation of a limited population of reagents, e.g., fluorescent analogs and enzymes, can lead to photo-induced damage of the various reagents to the point of substantially impacting the activity or functionality of the enzyme. It has been shown that prolonged illumination of DNA polymerases involved in synthesis using fluorescent nucleotide analogs results in a dramatic decrease in the enzyme's ability to synthesize DNA, often measured as a reduction in processivity. Without being bound to any theory of operation, it is believed that in some cases a photo-induced damage event affects the catalytic region of the enzyme thus affecting either the ability of the enzyme to remain complexed with the template, or its ability to continue synthesis. In general, the methods, devices, and systems of the present invention can increase performance and/or selectively monitor one or more stages of an illuminated reaction by subjecting the reaction to intermittent illumination.

[00197] One particularly preferred aspect of the invention is in conjunction with the sequencing by incorporation of nucleic acids within an optical confinement, such as a zero-mode waveguide. Such reactions involve observation of an extremely small reaction volume in which one or only a few polymerase enzymes and their fluorescent substrates may be present. Zero-mode waveguides, and their use in sequencing applications are generally described in U.S. Patent Nos. 6,917,726 and 7,033,764, and preferred methods of sequencing by incorporation are generally described in Published U.S. Patent Application No. 2003-0044781, the full disclosures of which are incorporated herein by reference in their entireties for all purposes, and in particular for their teachings regarding such sequencing applications and methods. Briefly, arrays of zero-mode waveguides ("ZMWs"), configured in accordance with the present invention may be employed as optical confinements for single molecule DNA sequence determination. In particular, as noted above, these ZMWs provide extremely small observation volumes at or near the transparent substrate surface, also termed the "base" of the ZMW. A nucleic acid synthesis complex, e.g., template sequence, polymerase, and primer, which is immobilized at the base of the ZMW, may then be specifically observed during synthesis to monitor incorporation of nucleotides in a template

dependent fashion, and thus provide the identity and sequences of nucleotides in the template strand. This identification is typically accomplished by providing detectable label groups, such as fluorescent labeling molecules, on the nucleotides. In some instances, the labeled nucleotides terminate primer extension, allowing a "one base at a time" interrogation of the complex. If, upon exposure to a given labeled base, a base is incorporated, its representative fluorescent signal may be detected at the base of the ZMW. If no signal is detected, then the base was not incorporated and the complex is interrogated with each of the other bases, in turn. Once a base is incorporated, the labeling group is removed, e.g., through the use of a photocleavable linking group, and where the label was not the terminating group, a terminator, upon the 3' end of the incorporated nucleotide, may be removed prior to subsequent interrogation. In other more preferred embodiments, the incorporation of a labeled nucleotide does not terminate primer extension and the processive incorporation of multiple labeled nucleotides can be monitored in real time by detecting a series of fluorescent signals at the base of the ZMW. In some such embodiments, the label is naturally released upon incorporation of the labeled nucleotides by the polymerase, and so need not be released by alternative means, e.g., a photocleavage event. As such, a processive sequencing reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. (Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.)

[00198] In accordance with the present invention, sequencing reactions may be carried out by only interrogating a reaction mixture, e.g., detecting fluorescent emission for one or more illuminated periods before excessive photo-induced damage has occurred. In general, the methods described herein are implemented in a manner sufficient to provide beneficial impact, e.g., reduced photo-induced damage and/or extension of the photo-induced damage threshold period, but are not implemented in such a manner to interfere with the reaction of interest, e.g., a sequencing reaction. The present invention also contemplates alternative methods of and compositions for mitigating the impact of photo-induced damage on a reaction, as described above and in, e.g., U.S.S.N. 61/116,048, filed November 19, 2008. Such alternative methods and compounds can be used in combination with the compositions and methods provided herein to further alleviate the effects of species that can be generated during an illuminated reaction.

[00199] Another method of mitigating the impact of photo-induced damage on the results of a given reaction provides for the elimination of potentially damaging oxygen species using means other than the use of the photo-induced damage mitigating agents described above. In one example, dissolved oxygen species may be flushed out of aqueous systems by providing the reaction system under different gas environments, such as by exposing an aqueous reaction to neutral gas environments, such as argon, nitrogen, helium, xenon, or the like, to prevent dissolution of excess oxygen in the reaction mixture. By reducing the initial oxygen load of the system, it has been observed that photo-induced damage effects, e.g., on polymerase mediated DNA synthesis, is markedly reduced. In particularly preferred aspects, the system is exposed to a xenon atmosphere. In particular, since xenon can be induced to form a dipole, it operates as a triplet-state quencher in addition to supplanting oxygen in the aqueous system. (See, e.g., Vierstra and Poff, Plant Physiol. 1981 May; 67(5): 996–998) As such, xenon would also be categorized as a quencher, as set forth above.

[00200] Although described in terms of zero-mode waveguides, it will be appreciated that a variety of selective illumination strategies may be employed to selectively interrogate different regions of a solid support over time, e.g., so as to only damage molecules within certain selected regions of a substrate while not damaging molecules in other selected regions of the substrate. In certain embodiments, such methods can involve using a directed light source (e.g., a laser) to illuminate only selected regions; changing the illumination angle of the light source; or refocusing the illumination, e.g., by passing the illumination through an optical train that alters the shape of the incident light on the solid support. These and further examples of alternative methods of mitigating photo-induced damage which can be used in combination with methods and systems of the invention described herein are provided in U.S. Patent Pub. No. 20070036511, filed August 11, 2005; U.S. Patent No. 6,881,312; U.S.S.N. 61/116,048, filed November 19, 2008; and U.S. Patent Pub. No. 20070161017, filed December 1, 2006, all of which are incorporated herein by reference in their entireties for all purposes, and in particular for disclosure related to these methods of mitigating photo-induced damage.

[00201] As noted above, using templates that allow repeated sequencing (e.g., circular templates, SMRTbell™ templates, etc.) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated and/or increase the fold-coverage of the sequence reads for one or more regions of interest in the template, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus

sequences for the nucleic acid template. For example, in certain preferred embodiments, templates sequenced by the methods described herein are templates comprising a double-stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double-stranded or otherwise internally complementary. Such templates may, for example, comprise a double-stranded portion comprised of two complementary sequences and two single-stranded linking portions (e.g., oligos or "hairpins") joining the 3' end of each strand of the double-stranded region to the 5' end of the other strand (sometimes referred to as "SMRTbell™" templates). In certain embodiments, double-stranded portions for use in such templates are PCR-amplified. Optionally, restriction sites are incorporated within the PCR primers such that subsequent digestion of the amplified products with appropriate restriction enzymes generates double-stranded portions containing known overhang sequences on either end, which are then ligated to hairpin adapters containing a complementary overhang to generate the SMRTbell™ templates.

[00202] These template molecules are particularly useful as nucleotide sequence data generated therefrom comprises both sense and antisense nucleotide sequences for the double-stranded portion, and the circular conformation of the template enables repeated sequencing (e.g., using a polymerase capable of strand-displacement) provides duplicative or redundant sequence information. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining sequence data for each segment from the complementary sequences, as well as sequence data within each segment, by repeatedly sequencing that segment. Iterative illumination is useful in such sequencing applications, e.g., to focus nucleotide sequence data collection on stages of the sequencing reaction most of interest, such as the stages during which nucleotide sequence data is being generated from a strand of the (previously) double-stranded portion. Iterative illumination may also allow additional "rounds" of sequencing the template by virtue of the reduction in photo-induced damage to reaction components, as described elsewhere herein, thereby providing more complete and robust nucleotide sequence data for future analysis, e.g., sequence scaffold construction and/or consensus sequence determination. Further, as described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data

being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest, e.g., from the ends of the double-stranded portion of a SMRTbell™ template.

[00203] In addition to providing sense and antisense sequence data within a single template molecule that can be sequenced in one integrated process, the presence of the single-stranded linking portions also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., the sense strand, is completed and the other begins, e.g., the antisense strand. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. Additional aspects of and uses for registration sequences, e.g., for molecular redundant sequencing, are further described in U.S. Patent Publication No. 20090029385, which is incorporated herein by reference in its entirety for all purposes.

[00204] In certain embodiments, such a sequencing process begins by priming the template nucleic acid within one of the linking portions and allowing the polymerase to proceed along the strand of the double-stranded portion of the template that is immediately downstream of the primed linking portion when the double-stranded portion is melted or denatured. The sequence process proceeds around the second linking portion and proceeds along the complementary strand of the (now previously) double-stranded portion of the template. Because the template is circular, this process can continue to provide multiple repeated sequence reads from the one template. Thus, sequence redundancy comes from both the determination of complementary sequences (sense and antisense strands of the double-stranded portion), and the repeated sequencing of each circular template. The ongoing sequencing reaction is subjected to multiple illuminated and non-illuminated periods to generate at least two or more sequence reads per pass around the template. The illuminated periods are preferably timed to allow generation of nucleotide sequence data for selected regions of the template. For example, it may be beneficial to only generate nucleotide sequence data for the complementary strands of the double-stranded portion, or segments thereof. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing. Other approaches will

similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post-synthesis.

[00205] One may optionally employ various means for controlling initiation and/or progression of a sequencing reaction, and such means may include the addition of specific sequences or other moieties into the template nucleic acid, such as binding sites, e.g., for primers or proteins. Various methods of incorporating control elements into an analytical reaction, e.g. by integrating stop or pause points into a template, are discussed elsewhere herein and are further described in related application, U.S. Application No. 12/413,258, filed March 27, 2009, which is incorporated herein by reference in its entirety for all purposes.

[00206] In certain embodiments, a reaction stop or pause point may be included within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the linking portion that was not used in priming. By way of example, following initial sequencing from the original priming location, e.g., from the single-stranded linking portion used in priming synthesis through a first portion of the sense strand (e.g., the 3' end), the data acquisition may be switched off and the polymerase allowed to proceed around the template, e.g., through the remainder of the sense strand to the other linking portion. The incorporation of a synthesis blocking moiety coupled to this linking portion will allow control of reinitiation of the polymerase activity at the 3' end of the antisense strand. One would thereby obtain paired-end sequence data for the overall (previously) double-stranded segment, with sequence data from one end coming from the sense strand and sequence data from the other end coming from the antisense strand. This template construction and sequencing methodology is particularly useful in the case of long double-stranded segments, especially given the short read lengths generated by some sequencing technologies.

[00207] A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the nucleobase portion of one or more bases in the single-stranded portion that inhibit polymerase-mediated replication; strand-binding moieties that prevent processive synthesis; non-native nucleotides included within the primer and/or template; and the like. The use of strand-binding moieties includes, but is not limited to, reversible, specific binding of particular proteins to recognition sequences incorporated into the template (or primer bound thereto) for this purpose. In certain embodiments, such control sequences may include binding sites for transcription factors, e.g., repressor binding regions provided within the linking portion(s). For example, the lac repressor recognition sequence is bound by the lac repressor protein, and this

binding has been shown to block replication in a manner reversible by addition of appropriate initiators, such as isophenylthiogalactoside (IPTG) or allolactose.

[00208] In some embodiments, primer recognition sequences and/or additional control sequences may also be provided for control of initiation and/or progression of polymerization, e.g., through a hybridized probe or reversibly modified nucleotide, or the like. (See, e.g., U.S. Patent Application No. 2008-0009007, the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Such probes include but are not limited to probes at which a polymerase initiates polymerization, probes containing various types of detectable labels, molecular beacons, TaqMan® probes, Invader® probes (Third Wave Technologies, Inc.), or the like, that can be used for various purposes, e.g., to provide indications of the commencement and/or progress of synthesis.

[00209] An engineered pause point (reversible or irreversible) can include one or more non-native (non-natural) or fifth bases that do not pair with any of the four native nucleoside polyphosphates in the synthesis reaction, e.g., in the template and/or oligonucleotides probe(s), and/or that exhibit a distinct kinetic signature during template-dependent synthesis at such a base. Upon encountering such a base, the polymerase pauses until the complement to the non-natural base is added to the reaction mixture. Likewise, an engineered pause point could include a "damaged" base that causes a stop in replication until repair enzymes are added to the mixture. For example, a template having a pyrimidine dimer would cause the replication complex to pause, and addition of the photolyase DNA repair enzyme would repair the problem location and allow replication, and sequencing to continue. In yet further embodiments, a combination of modification enzymes could be used to engineer a set of modified bases on a template, e.g., a combination of glycosylases, methylases, nucleases, and the like. (Further information on sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.)

[00210] As noted elsewhere herein, stop or pause points can be engineered into various portions of the template, e.g., portions for which the nucleotide sequence is unknown (e.g., a genomic fragment) or known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, SMRTbell™ templates are topologically closed, single-stranded molecules having regions

of internal complementarity separated by hairpin or stem-loop linkers, such that hybridization of the regions of internal complementarity produces a double-stranded portion within the template. One or both of the linkers can comprise a stop or pause point to modulate polymerase activity. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00211] In other embodiments, an abasic site is used as a synthesis blocking moiety or pause point until addition of a non-natural "base," such as a pyrene, which has been shown to "base-pair" with an abasic site during DNA synthesis. (See, e.g., Matray, et al. (1999) Nature 399(6737):704-8, which is incorporated herein by reference in its entirety for all purposes.) Where a permanent termination of sequencing is desired, no non-natural analog is added and the polymerase is permanently blocked at the abasic site. DNA (or RNA) glycosylases create abasic sites that are quite different from the normal coding bases, A, T, G, and C (and U in RNA). A wide variety of monofunctional and bifunctional DNA glycosylases that have specificity for most common DNA or RNA adducts, including 5-methylcytosine, are known in the art, with different glycosylases capable of recognizing different types of modified DNA and/or RNA bases. The molecular structures of many glycosylases have been solved, and based on structural similarity they are grouped into four superfamilies. The UDG and AAG families contain small, compact glycosylases, whereas the MutM/Fpg and HhH-GPD families comprise larger enzymes with multiple domains. As an example, four enzymes have been identified in *Arabidopsis thaliana* in the plant pathway for cytosine demethylation. Additionally, other enzymes are also known to recognize 5-methyl cytosine and remove the methylated base to create an abasic site. Further, various enzymes are known to methylate cytosine in a sequence-specific manner. As such, a combination of a cytosine-methylase and an enzyme that creates an abasic site from a methylated cytosine nucleotide can be used to create one or more abasic sites in a template nucleic acid. The size of the recognition site of the methylase and the base composition of the template determine how frequently methylation occurs, and therefore, the number of abasic sites created in a given template nucleic acid, allowing the ordinary practioner to choose a methylase with a recognition site that produces a desired spacing between modified nucleotides. For example, if the recognition site is three bases long, then on average an abasic site is expected every 64 bases; if the recognition site is four bases long, then on

average an abasic site is expected every 256 bases; if the recognition site is six bases long, then on average an abasic site is expected every 4096 bases; and so forth. Of course, templates with a higher GC content would be expected to have more frequent abasic site formation, and templates with lower GC content would be expected to have less frequent abasic site formation.

[00212] Uracil-DNA glycosylases can also be used to introduce abasic sites into a template nucleic acid comprising deoxyuridine nucleotides. This strategy has the advantage of allowing the practitioner to choose the locations of the abasic sites within a DNA template since deoxyuridine nucleotides are not generally found in DNA. Various methods of inserting deoxyuridine nucleotides into a DNA template may be used, and different methods will be preferred for different applications. In certain embodiments, one or more site-specific deoxyuracils are incorporated during standard phosphoramidite oligonucleotide synthesis. To place uracils at indeterminate positions in a DNA, replacing a portion of the deoxythymidine triphosphate with deoxyuridine triphosphate will result in an amplimer with random U sites in place of T sites after polymerase chain reaction. In other embodiments, deoxyuridine nucleotides are engineered into the template, e.g., by ligation of a synthetic linker or adaptor comprising one or more deoxyuridine nucleotides to a nucleic acid sequence to be sequenced. In certain preferred embodiments, deoxyuridine nucleotides are incorporated into the linker portions of a SMRTbell™ template.

[00213] To subsequently introduce abasic sites prior to sequencing, the deoxyuridine nucleotide-containing template is subjected to treatment with uracil-DNA glycosylase, which removes the one or more uracil bases from the deoxyuridine nucleotides, thereby generating one or more abasic sites in the template. Alternatively, since the deoxyuridine nucleotide can be recognized as a template base and paired with deoxyadenosine during template-dependent nascent strand synthesis, the synthesis-blocking abasic site can instead be introduced after initiation of the sequencing reaction, e.g., at a time chosen by the practitioner. For example, the reaction can be initiated with a deoxyuridine-containing template, and uracil-DNA glycosylase can subsequently be added to block the polymerase and halt the reaction after the reaction has proceeded for a given time. As such, termination of the reaction is optional rather than required.

[00214] While uracil-DNA glycosylase activity is useful for introducing abasic sites into a template as described above, this activity can be problematic during the preparation of such templates. As such, strategies are typically implemented during preparation and manipulation of uracil-containing DNA, e.g., using molecular biology enzymes, to avoid uracil-DNA glycosylase activity, in particular, due to the *E. coli* UDG enzyme. Since a majority of standard molecular

biology enzymes are overexpressed and subsequently purified from an *E. coli* host, UDG activity can be a contaminating activity that is often not monitored by the enzyme manufacturer's quality control procedures. To mitigate contaminating UDG activity, a commercially available UDG inhibitor, also known as uracil glycosylase inhibitor or UGI (e.g., from New England Biolabs, Ipswich, MA) can be included in molecular biology reactions. This is a small protein inhibitor from the *B. subtilis* bacteriophase PBS1 that binds reversibly to E. coli UDG to inhibit its catalytic activity. UGI is also capable of dissociating UDG from a DNA molecule. Alternatively, UDG activity can be inhibited without exogenous protein using a chemical inhibitor of the enzyme, such as an oligonucleotide containing a 1-aza-deoxyribose base, a transition state analog for the UDG enzyme. This and other cationic nitrogenous sugars have been used for mechanistic studies of UDG activity and show potent inhibition activity. (See, e.g., Jiang et al. Biochemistry, 2002, 41 (22), pp 7116–7124.)

[00215] In certain applications, UDG activity needs to be inhibited temporarily, and subsequently enabled to remove create an abasic site as described above. In some embodiments, a DNA purification that removes proteins is employed, e.g., including a phenol-chloroform extraction with subsequent ethanol precipitation, a silica-based column approach (e.g., QiaQuick columns from Qiagen and similar products), and/or a PEG/sodium chloride precipitation (e.g., AMPure beads from Beckman Coulter). Alternatively or additionally, a commercially-available UDG enzyme that is not inhibited by UGI is added when abasic site formation is desired. For example, the *A. fulgidus* UDG is from a thermophilic organism and cannot be inhibited by the same bacteriophage protein as is the *E. coli* UDG enzyme. In certain preferred embodiments, UDG-inhibition is employed during template preparation, and inhibition-resistant UDG activity is added at a subsequent time to trigger the creation of abasic sites at deoxyuridine nucleotides, e.g., immediately prior to or during an ongoing reaction.

[00216] In some preferred embodiments, one or more abasic sites are engineered into a linker or adapter sequence within a sequencing template molecule. Abasic sugar residues serve as efficient terminators of polymerization for many polymerases, e.g., Φ29. 1',2'-dideoxyribose is the most common synthetic "abasic site". In other embodiments, a synthetic linker is incorporated into a linker or adaptor. For example, an internal spacer (e.g., Spacer 3 from Biosearch Technologies, Inc.) or other carbon-based linker can be used in lieu of a sugar-base nucleotide. Similar to an abasic nucleotide, the polymerase will be blocked upon encountering these moieties in the template nucleic acid.

[00217] In certain embodiments, synthesis blocking moieties are nicks in the template nucleic acid. Nicking enzymes (e.g., nicking endonucleases) are known in the art and can be used to specifically nick the template prior to or during a template-directed sequencing reaction. The use of site-specific nicking endonucleases allows the practitioner to incorporate a recognition sequence at a particular location within the template nucleic acid, and such nicking endonucleases are commercially available, e.g., from New England Biolabs, Inc. For example, a linker or adapter can be synthesized with a nicking endonuclease recognition sequence, ligated to a nucleic acid molecule to be sequenced, and can be specifically nicked either before or during a subsequent sequencing reaction. Nicks can also be introduced by ligating duplex segments that lack either a terminal 3'-hydroxy (e.g., have a dideoxynucleotide at the 3'-terminius) and/or 5'-phosphate group on one strand. The ligation results in covalent linkage of the phosphodiester backbone on one strand, but not on the other, which is therefore effectively "nicked." In certain embodiments, a SMRTbell™ template is constructed using a duplex (or "insert") nucleic acid molecule lacking a 5'-phosphate group at one or both termini. Upon ligation of the hairpin or stem-loop adaptors at each end, nicks are created at one or both ligation site(s), depending on whether the duplex lacked a 5'-phosphate at one or both ends, respectively. In other embodiments, a SMRTbell™ template is constructed using one or two stem-loop adaptors lacking a 3'-hydroxy group at the terminus (e.g., comprising a 2',3'-dideoxynucleotide rather than a 2'-deoxynucleotide). Upon ligation of one or two stem-loop adaptors lacking a 3'-hydroxy group, one or two nicks are created at the ligation site(s), depending on whether one or two adaptors lacked the 3'-hydroxy group, respectively. In both cases, a nick is created in the template nucleic acid, and a primer bound to one of the adaptors provides an initiation site for the polymerase, which will process the template until encountering a nick, at which point the polymerase will terminate the reaction by dissociation from the template. Regardless of how a nick is created, the position of a nick relative to the initiation site for the polymerase determines how much of the template will be sequenced. For example, Figure 19A provides an illustrative example of an embodiment in which a nick is present on a first strand of a duplex portion at a position distal to the adaptor containing the primer binding site. The first strand is processed by a polymerase, but the complementary strand is not processed because the polymerase dissociates at the nick site. An alternative embodiment is shown in Figure 19B, in which a nick is present on the strand complementary to the first strand at a position proximal to the adaptor containing the primer binding site. In this case both the first and complementary strands, as well as the adaptor not containing the primer binding site, are processed by the polymerase prior to dissociation. The

position of the primer binding site also determines how much of the template is processed by the polymerase. Figure 19C provides a template having a primer binding site at a position from which a polymerase would process a significant portion of the adaptor prior to entering the duplex portion. An additional advantage to using a 3'-dideoxynucleotide at a nick is that it prevents the use of the nick as a polymerse initiation site, since strand extension requires a 3-hydroxy group. As such, the resulting nick would not compete with a primer site for initiation of nascent strand synthesis by the polymerase. Having a single, known site of initiation on a template molecule is beneficial, e.g., for subsequent mapping of a read generated in such a reaction. In certain preferred embodiments, a nick site both lacks a 5'-phosphate group and comprises a 3'-dideoxynucleotide.

[00218] In certain preferred embodiments, modification and base excision is performed prior to introduction of a template nucleic acid to a reaction site, e.g., a zero-mode waveguide. As noted above, the choice of recognition site for the methylase depends on how far apart the practitioner wishes point of synthesis initiation to be on the template. For example, after initiating the template-dependent sequencing reaction, the sequence of nucleotide incorporations into the nascent strand is monitored for a desired sequence read, which may extend from the initiation point to the pause point, or may end before the polymerase reaches the pause point. In some preferred embodiments, as described elsewhere herein, the monitoring is suspended by modifying or removing an illumination source, e.g., by moving the illumination source or a substrate comprising the reaction site. Synthesis of the nascent strand will continue until the pause site is reached, whether or not the reaction is being actively monitored. When the reaction is to be reinitiated, reaction components are added that allow bypass, e.g., pyrene, polymerase, etc., and these can be subsequently removed (e.g., by buffer exchange) to allow additional pauses at other pause sites on the template.

[00219] In certain embodiments using pyrosequencing-based technologies (e.g., as developed by 454 Life Sciences), abasic sites can be introduced into a set of amplified template nucleic acids and synthesis initiated. Since all templates in the set are identical, they will comprise the same number of abasic sites in the same positions. During the course of the synthesis reaction, the synchronous incorporation of nucleotides into the nascent strands is monitored until either an abasic site is reached (at which point the synthesis is paused) or until the incorporation becomes asynchronous (which increases the background noise and decreases reliability of the sequence read). In the latter case, the practioner may opt to speed up the reaction, e.g., by adding all nucleotides at one time, to extend all nascent strands to the first abasic site in the templates. When synthesis is to be reinitiated, reaction components are added that allow bypass of the abasic site, e.g. one or more

pyrenes. A wash step may be performed to remove nucleotides and/or polymerases from the reaction sites prior to such addition. Further, in some cases, a different polymerase may be used for pyrene incorporation as is used for sequencing-by-synthesis reactions. In certain preferred embodiments, the reaction mixture comprising the pyrene for abasic site bypass allows readthrough of the abasic site, but no further on the template. Subsequent addition of sequencing reaction mixture allows the sequencing-by-synthesis reaction to recommence and incorporation of nucleotides into the nascent strand to be monitored. Alternatively or additionally, the practioner need not wait until an abasic site is reached to suspend detection and, optionally, speed up the reaction to bring all nascent strands to a given abasic site, but can choose to do this before a reaction has become asynchronous, e.g., after desired sequence data has been collected for a particular region of interest in a template nucleic acid.

[00220] In certain embodiments using ligation-based technologies (e.g., the SOLiD™ System developed by Life Technologies), a pause site can be engineered by using an oligonucleotide that cannot participate in the ligation reaction and that is complementary to a desired location on the set of identical template nucleic acids, e.g., on a bead. When the serial ligation reaction hits the position recognized by this polynucleotide, the reaction cannot proceed and any reactions that have become asynchronous will "catch up." The user can then unblock the oligo (e.g., using chemical treatment or photo-cleavage) and reinitiate the sequencing reaction.

[00221] In some cases, it may be desirable to provide endonuclease recognition sites within the template nucleic acid. For example, inclusion of such sites within a circular template can allow for a mechanism to release the template from a synthesis reaction, i.e., by linearizing it, and allowing the polymerase to run off the linear template, and/or to expose the template to exonuclease activity, and thus terminate synthesis through removal of the template. Such sites could additionally be exploited as control sequences by providing specific binding locations for endonucleases engineered to lack cleavage activity, but retain sequence specific binding, and could therefore be used to block progression of the polymerase enzyme on a template nucleic acid.

[00222] In some cases, nicking sites, e.g., sites recognized by nicking endonucleases, may be included within a portion of the template molecule, and particularly within a double-stranded portion of the template, e.g., in a double-stranded segment of a SMRT bell™ or in the stem portion of an exogenous hairpin structure. Such nicking sites provide one or more breaks in one strand of a double-stranded sequence and can thereby provide one or more priming locations for, e.g., a strand-displacing polymerase enzyme. A variety of nicking enzymes and their recognition sequences are

known in the art, with such enzymes being generally commercially available, e.g., from New England Biolabs.

[00223] In certain embodiments, methods for intermittent detection described herein are useful in "paired-end" sequencing applications in which sequence information is generated from two ends of a template nucleic acid but not for at least a portion of the intervening portion of the template. Typically, paired-end sequencing applications provide sequence data for only the two ends of a nucleic acid template, but the present invention also allows generation of additional sequence reads that are noncontiguous with the sequence reads from the ends of the template. In certain preferred embodiments, a duplex fragment (e.g., genomic fragment) is ligated to a single-stranded linker that connects the 3' end of the sense strand to the 5' end of the antisense strand, or that connects the 5' end of the sense strand to the 3' end of the antisense strand. In either orientation, separation of the two strands of the duplex fragment results in a single-stranded linear template nucleic acid that contains the linker in between the sense and antisense strands. Subsequent sequencing can involve intermittent detection that generates sequence reads for only the portions of the sense and antisense strands that are of interest, e.g., one or both of the ends. In certain embodiments, both sense and antisense strands may be sequenced at both ends to provide redundancy in the sequence data. Sequence reads recognized as being from the linker portion of the template (e.g., based on the known linker sequence or specific registration sequences encoded therein) can be used to orient the alignment of the sequence reads from the sense and antisense portions of the template, providing context for determining the sequences of the ends of the duplex fragment and subsequent sequence scaffold construction and/or mapping. In certain embodiments, pause or stop points may be incorporated into the linker to control the processing of the template by the polymerase, and therefore may be used to synchronize the detection periods to ensure generation of sequence reads from particular regions of template. Further, additional detection periods can be included that are timed to provide sequence reads from portions of the sense and/or antisense strand that are noncontiguous with the end regions.

[00224] In a related embodiment, paired-end sequencing may be accomplished by using a nucleic acid template that has linkers connecting the sense and antisense strands of a duplex fragment at both ends, such that separation of the strands of the duplex fragment provides a single-stranded circular template that contains a linkers in between each end of the sense and antisense strands of the original duplex fragment. Such a template molecule would allow a strand-displacing polymerase to proceed around the template multiple times, thereby potentially generating redundant

sequence data from both ends of both strands of the original duplex fragment. As noted elsewhere herein, such redundancy is useful for determination of consensus sequences and/or construction of sequence scaffolds. As the polymerase enzyme processes the template, detection periods can be timed (e.g., based on knowledge of the rate at which the polymerase processes the template, which is dependent not only on the polymerase but also on the sequence of the template itself) to generate nucleotide sequence reads from the regions of the template corresponding to one or both ends of the sense and antisense strands, and can also include detection periods to generate additional reads from other, noncontiguous regions of the duplex fragment, as well. Although such timing can be used to determine the appropriate periodicity of the detection periods, at later stages of the reaction (e.g., as the polymerase repeatedly proceeds around the template), the exact location of reinitiation of sequence read generation becomes more approximate. Incorporation of pause or stop points into one or both linkers to regulate the processing of the template by the polymerase may be used to synchronize the detection periods regardless of the total distance travelled by the polymerase around the template. This strategy more reliably ensures generation of sequence reads from selected regions of template, e.g, the ends of the sense and antisense portions and, optionally, regions in between and noncontiguous with the end regions regardless of the number of passes of the polymerase around the template nucleic acid, especially in later stages of the reaction. Further, the known sequence of one or both of the linkers can be used to orient sequence reads from the sense and antisense portions for consensus sequence determination and/or mapping.

[00225] In some such embodiments, a duplex fragment inserted between two hairpin linkers may be much larger than desired, increasing the difficulty of limiting nucleotide sequence read data to particular regions of the fragment. The size of the duplex fragment ligated to the two hairpin linkers can be selectively reduced to retain the regions attached to the linkers and to lose a central portion of the duplex fragment. One particularly preferred strategy, illustrated in Figure 18, comprises hairpin linkers (1802, 1804) having a regions of cross-complementarity (1806, 1808), such that the two linkers 1802 and 1804 can anneal to each other in a manner that does not interfere with ligation to a duplex fragment 1810. Duplex fragment 1810 comprises ends 1812 and 1814, as well as a long central region 1816, which is not shown but is understood to be between the two curvy lines. Once end 1812 is ligated to linker 1802 and end 1814 is ligated to linker 1804, the construct is subjected to fragmentation, which removes the central region 1816 of the duplex fragment 1810, producing construct 1818 having ends 1820 and 1822. After fragmentation, the ends of the portions of the duplex fragment still associated with the annealed linker pair (ends 1820 and

1822) are ligated together to produce construct 1824, which can then be treated (e.g., with heat, gentle denaturation, primer invasion, changing salt concentration, etc.) to separate cross-complementary regions 1806 and 1808 from one another, e.g., to generate a circular single-stranded nucleic acid molecule. Alternatively, the separation may occur during the course of the subsequent reaction, e.g., by polymerase-mediated strand displacement. Yet further, where the hybridized cross-complementary regions are long enough to undergo a complete DNA turn, an additional reaction component (e.g., helicase, topoisomerase, polymerase, etc.) may be needed to unwind the duplex and allow separation. As such, the resulting "mate-pair" construct has only the ends of the original duplex fragment ligated together and capped with adaptors that link the 5' end of each strand of the duplex with the 3' end of the other strand of the duplex, and denaturation of the duplex produces a closed, single-stranded circular construct.

[00226] Fragmentation of the duplex fragment can be performed by a variety of known methods. For example, fragmentation can be performed enzymatically (e.g., using restriction enzymes or other nucleases) or mechanically, by shearing or sonication. The type of fragmentation chosen will determine various characteristics of the resulting construct, e.g., how large a central region is removed and the types of ends remaining (e.g., blunt, 5' overhang, 3' overhang, random, identical on both ends, etc.). Optionally, the ends can be modified after fragmentation to facilitate the subsequent ligation step. Although not shown in Figure 18, it is expected that the ligation of the duplex fragment to the hybridized linkers will be a two-step process, with one end being ligated first and unimolecular kinetics favoring ligation of the second end to the second linker. The cross-complementary regions of the linkers can be designed to produce varying levels of complementarity, and therefore varying strengths of the hybridization. For example, a longer or higher GC content in a cross-complementary region lends a higher stability to the linker:linker interaction, but separation of the hybridized linkers requires a more severe treatment, e.g., higher temperature, more stringent conditions, etc. As such the cross-complementary regions should be engineered to produce a stable linker:linker interaction that is disruptable under conditions that are not destructive to the overall construct. Further the linkers can vary in regions apart from the cross-complementary regions For example, one linker can have a primer binding site that the other lacks, which would provide a single polymerase initiation site in the final construct. Other sequence characteristics described herein (e.g., pause sites, registrations sequences, etc.) can also be included in one or both linker regions. If topological constraints limit the subsequent processing of the resulting contruct, e.g., during template-directed nascent strand synthesis, these can be addressed by

addition of a reaction component (e.g., a helicase or topoisomerase) to resolve the topological constraint. As such, the methods can be used to add asymmetric linkers to duplex polynucleotides, whether or not the duplex is to be selectively reduced in size, or not, as long as the asymmetric linkers can cross-hybridize to one another.

[00227] Although in preferred embodiments, the two linkers to be ligated to a single duplex fragment are hybridized to one another prior to ligation, in some embodiments they are instead hybridized after the initial ligation reaction, and where topological constraints inhibit such a post-ligation hybridization a reaction component (e.g., topoisomerase) may be included to relieve such constraints. In certain embodiments, the hybridized linkers are separated prior to addition of reaction components for a subsequent reaction, and in other embodiments the hybridized linkers are not separated until after the addition of reaction components for a subsequent reaction. For example, a polymerase enzyme may bind to a primer annealed to a linker before or after separation of the linker from a second linker. In fact, it may be benefical in some embodiments to postpone separation of the linkers, e.g., where compaction of the nucleic acid construct is beneficial, such as when the construct must be loaded into a confinement of some kind, e.g., a nanowell, optical confinement, etc.

[00228] In some embodiments, the methods further include separation of single linker constructs from hybridized linker pair constructs. This can be accomplished by an exonuclease treatment after ligation of the duplex fragment to the linkers, which would degrade any constructs having an unannealed end. Alternatively, it may be desirable to remove the single linkers prior to ligation, for example using a size separation methodology or by allowing them to bind to oligonucleotides that are complementary to the cross-complementary regions and bound to a column or magnetic beads. (The cross-complementary regions of the hybridized linker pairs will not be available for binding to the oligonucleotides. Other methods known in the art can also be used to separate single linkers from hybridized linker pairs.

[00229] Interestingly, the use of the sense/antisense nucleic acid templates described above would represent a unidirectional processing of a template to provide paired-end sequence data, as opposed to the more traditional bi-directional processing of a linear template molecule. Further, unlike traditional approaches, these methods for paired-end sequencing involve processing, chemically or otherwise, of not just the regions at the ends, but also regions in between the ends, and in some embodiments comprising processing of the entire template. For example, a polymerase incorporates nucleotides into a nascent strand for each position of the template (thereby

"processing" each position of the template), yet the sequencing data generated is limited to specific regions of the template that are of particular interest to the practitioner, such as the end regions. As such, in certain embodiments the duplex fragment is not further reduced in size after ligation to a linker pair, and the entire duplex fragment is processed by the polymerase.

[00230] In certain embodiments, methods for intermittent detection described herein are useful in analysis systems that employ nanopores. A nanopore is a small pore in an electrically insulating membrane that can be used for single molecule detection. In general, a nanopore functions as a Coulter counter for much smaller particles, and can take various forms, e.g., a protein channel in a lipid bilayer or a pore in a solid-state membrane. The detection principal is based on monitoring the ionic current of an electrolyte solution passing through the nanopore as a voltage is applied across the membrane. For example, passage of a polynucleotide molecule (e.g., DNA, RNA, etc.) through a nanopore causes changes in the magnitude of the current through the nanopore, with each nucleotide obstructing the nanopore to a different, characteristic degree. As such, the pattern of variations in the current passing through the nanopore as the polynucleotide is drawn through may be monitored and analyzed to determine the nucleotide sequence of the polynucleotide. A polynucleotide may be drawn through the nanopore by various means, e.g., by electrophoresis, or using enzyme chaperones to guide the polynucleotide through the nanopore. For additional discussion of methods of fabrication and use of nanopores, see, e.g., U.S. Patent No. 5,795,782; Kasianowicz, J.J., et al. (1996) *Proc Natl Acad Sci USA* 93(24):13770-3; Ashkenas, N., et al. (2005) *Angew Chem Int Ed Engl* 44(9):1401-4; Winters-Hilt, S., et al. (2003) *Biophys J* 84:967-76; Astier, Y., et al. (2006) *J Am Chem Soc* 128(5):1705-10; Fologea, D., et al. (2005) *Nano Lett* 5(10):1905-9; Deamer, D.W., et al. (2000) *Trends Biotechnol* 18(4):147-51; and Church, G.M. (2006) *Scientific American* 294(1):52, all of which are incorporated by reference herein in their entireties for all purposes. In some embodiments, intermittent detection of nucleic acid sequence data from a nanopore may be achieved by modifying the progress of the polynucleotide through the nanopore so that progress is sped up during non-detection periods and progress is slowed to allow sequence determination during detection periods. The rate of passage of the polynucleotide through the nanopore may be modified by various methods, including but not limited to increasing an electrophoretic field carrying the polynucleotide (e.g., by increasing the voltage, changing the conductivity of the reaction mixture, and the like), or changing various reaction conditions to alter the speed at which a protein chaperone carries the polynucleotide. Further, in embodiments utilizing

a processive exonuclease to feed individual bases through the nanopore, the kinetics of the exonuclease may be modified based on the known biochemical characteristics of the exonuclease.

[00231] In diagnostic sequencing applications, it may be necessary only to provide sequence data for a small fragment of DNA, but do so in an extremely accurate sequencing process. For such applications, shorter target segments may be employed, thus permitting a higher level of redundancy by sequencing multiple times around a smaller circular template, where such redundancy provides the desired accuracy. Thus, in some cases, the double stranded target segment may be much shorter, e.g., from 10 to 200, from 20 to 100 or from 20 to 50 or from 20 to 75 bases in length. For purposes of the foregoing, the length of the target segment in terms of bases denotes the length of one strand of the double stranded segment. In such applications, various methods for intermittent detection described herein may be used to analyze the sequence of the template, thereby targeting the sequence data to the portion(s) of the template of particular interest to the diagnostician, and/or improving various aspects of the reaction performance, e.g., by virtue of the reduction of photo-induced damage to one or more reaction components.

[00232] It is to be understood that the above description is intended to be illustrative and not restrictive. It readily should be apparent to one skilled in the art that various embodiments and modifications may be made to the invention disclosed in this application, including but not limited to combinations of various aspects of the invention, without departing from the scope and spirit of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. All publications mentioned herein are cited for the purpose of describing and disclosing reagents, methodologies and concepts that may be used in connection with the present invention. Nothing herein is to be construed as an admission that these references are prior art in relation to the inventions described herein. Throughout the disclosure various patents, patent applications and publications are referenced. Unless otherwise indicated, each is incorporated by reference in its entirety for all purposes.

[00233] Although described in some detail for purposes of illustration, it will be readily appreciated that a number of variations known or appreciated by those of skill in the art may be practiced within the scope of present invention. Unless otherwise clear from the context or expressly stated, any concentration values provided herein are generally given in terms of admixture values or percentages without regard to any conversion that occurs upon or following addition of

the particular component of the mixture. To the extent not already expressly incorporated herein, all published references and patent documents referred to in this disclosure are incorporated herein by reference in their entirety for all purposes.

[00234] The following non-limiting examples are provided to further illustrate the invention.

VI. Examples of Intermittent Illumination of a Single Molecule Sequencing-by-Synthesis Reaction

Example I

[00235] A nucleic acid template was provided that comprised a double-stranded region and two single-stranded linker portions at each end. The first linker portion connected the 3' end of the sense strand with the 5' end of the antisense strand, and the second linker portion connected the 3' end of the antisense strand with the 5' end of the sense strand. This template was designed to form a single-stranded circle of approximately 500 bases when the double-stranded region was opened (e.g., by heat denaturization, helicase activity, etc.), and is sometimes referred to as a SMRTbell™ template. A plurality of this nucleic acid template was incubated with polymerases, primers, and other reaction components to allow formation of polymerase-template complexes. (See, e.g., Korlach, J., et al. (2008) Nucleosides, Nucleotides and Nucleic Acids, 27:1072-1083; and Eid, J. (2009) Science 323:133-138.) The complexes were immobilized in zero-mode waveguides in a reaction mixture containing all necessary buffer and nucleotide analog components for carrying out sequencing-by-synthesis reactions with the exception of a cognate starting base and a metal dication. A Smith-Waterman algorithm was used to perform the alignment of the known sequence of the template with the sequence reads generated in the reaction, and the positions of the sequence reads is graphically illustrated in Figure 8.

[00236] Acquisition of the data shown in Figure 8 was collected as follows. Illumination of the array of zero-mode waveguides was initiated with laser excitation (532 nm and 641 nm laser lines) at t = -5 seconds, and the missing cognate starting base and metal dication (manganese metal) were added at t = 0 seconds to simultaneously initiate the sequencing-by-synthesis reactions in all zero-mode waveguides. The reactions were monitored under illumination for 120 seconds at which time the illumination was removed; the sequencing reads generated during that stage of the reaction are shown in Figure 8A as a function of the template position to which each read maps. At 295 seconds illumination was resumed and data acquisition was reinitiated at 300 seconds and maintained for another 120 second interval; the sequencing reads during this second illuminated

period are shown in Figure 8B. At 595 seconds illumination was resumed and data acquisition was reinitiated at 600 seconds and maintained for another 120 second interval; the sequencing reads during this third illuminated period are shown in Figure 8C.

[00237] As expected, the longer the amount of time before the sequence data is collected (that is, the later the illuminated period), the further into the template the alignments shift, and this shift is a rough function of time since initiation of the reaction. Further, the distribution of sequence reads generated during each subsequent illuminated period becomes more dispersed than the previous illuminated period(s). Further, due to the circular nature of the template, Figure 8C clearly shows that some polymerases have passed completely around the substrate and are beginning to generate sequence reads from a second pass around the template, thereby generating redundant sequence information for a single template nucleic acid.

Example II

[00238] As in Example I, a SMRTbell™ template was used. For templates of defined sequence, PCR was used to generate 3 or 6 kb DNA inserts for the double-stranded region in the SMRTbell™ templates using a standard PCR methodology. For genomic and other biological samples, a DNA fragmentation protocol was used that generates DNA fragments distributed around 3 or 6 kb. Generation of fragments in these ranges was done using a HydroShear® (Genomic Solutions®) device with settings recommended by the manufacturer. The random genomic DNA fragments were enzymatically treated to generate blunt ends. Both the PCR products and randomly generated DNA fragments were phosphorylated and then immediately put into a ligation reaction with a blunt hairpin adapter. The products were purified through two size selection steps using reduced volumes of AMPure® magnetic beads (Agencourt®) to remove hairpin dimers and other short products. (Fabrication of SMRTbell™ templates is further described elsewhere herein.)

[00239] The system components used for polynucleotide sequencing using intermittent detection are comparable to single-molecule sequencing applications under constant illumination, which are described, e.g., in Eid, et al. (2009) Science 323:133-138. Specifically, the immobilization and sequencing buffer compositions, nucleotide analogs identity and concentration, polymerase, ZMWs, surface treatment and instrumentation were identical to the standard methodology. Modifications to the SMRTbell™ template DNA and polymerase binding and immobilization and data acquisition protocols are as follows.

[00240] A binding solution was prepared by incubation of 3 or 6 kb DNA SMRTbell™ templates (1-10 nM) with a 10-fold excess of DNA polymerase (10-100 nM, respectively) in 10 mM MOPS (pH 7.5), 10 mM KOAc, 100 mM DTT & 0.05% Tween-20 for 2 hours at 30°C, followed by 1 hour at 37°C and subsequent storage at 4°C prior to immobilization on the ZMWs. Immediately prior to immobilization, the binding solution was diluted in the standard immobilization solution (50 mM MOPS (pH 7.5), 75 mM KOAc, 5 mM DTT, 0.05% Tween-20) to the desired final concentration, typically 0.1 to 1 nM, and incubated for 30 to 60 minutes at 22°C. Post-immobilization chip preparation and sequencing initiation were identical to the standard methods.

[00241] The data acquisition protocol was similar to the standard application with coordinated modifications to the collection timing and ZMW positioning. In the standard acquisition procedure, a single long acquisition (~10 minutes) is performed for each ZMW. In the intermittent illumination acquisition procedure, multiple short acquisitions (~3 minutes) of sequence reads (also termed "strobe reads") were performed for each ZMW (during "detection periods") with an interval between each acquisition period during which no acquisition of sequence reads was performed ("non-detection period"). The duration of the interval between each acquisition of sequence reads was determined based upon a desired distance (i.e., number of nucleotide positions) between each sequence (or strobe) read, the polymerization rate of the polymerase, and the SMRTbell™ template insert size.

[00242] SMRTbell™ templates were generated as described above for AC223433, a fosmid clone comprising a sequence of an approximately 40 kb region of *Homo sapiens* chromosome 15. The reference sequences used to map the sequence reads generated in the sequencing reactions were the publically available sequences of *Homo sapiens* chromosome 15 (Hg18; NCBI Build 36.1) and fosmid AC223433 (NCBI GenBank accession number). Table 1 shows the number of statistically significantly mapped sequence reads for several types of intermittent illumination sequencing reactions. The number of mappable "looks" is equivalent to the number of mappable sequence reads generated during detection periods for a single template molecule. For example, a "mapped 1-look read" means, for a single template molecule, only a single detection period generated a sequence read that could be mapped to the reference sequence.

Table 1: Summary of Sequencing Results

Mapping	Mapped 1-	Mapped 2-	Mapped 3-	Mapped 4-

Reference	look reads	look reads	look reads	look reads
Human chr15 (Hg18)	13834	1289	127	4
Fosmid	15253	1571	158	5

[00243] Deviations in the expected time span for a set of sequencing reads from a single sequencing reaction are indicative of genomic events such as genomic rearrangements, e.g., insertions, deletions, etc. Figures 10 and 11 illustrate this point. Specifically, the time and distance travelled along the template (based upon the reference sequence) by the polymerase was computed within and between the sequence reads generated during the detection (illuminated) periods. These calculations were used to detect unexpected variations, indicating possible genomic events in the template as compared to the reference sequence. Figure 10 provides a plot that illustrates the normalized average time it took for the polymerase to traverse a region of the template based on the length of that region in the *Homo sapiens* chromosome 15 reference sequence. The sequence reads are fit to a diagonal having a slope equal to the average speed for sequencing reads. Deviations from the regressed diagonal indicate genomic events (for example, structural variants), and the slope of the sequence reads around such deviations indicate the relative size of the genomic event (e.g., in the case of insertions/deletions). For example, if the time for the polymerase to traverse a region was unexpectedly long, this indicated the polymerase actually traversed a longer region than was expected based on the reference sequence. The two distinct off-diagonal deviations (upper right hand corner) with higher slope indicated that an insertion had occurred in the reference sequence, and this was verified by comparison to the known fosmid sequence.

[00244] Figure 11 shows the average time it took the polymerase to traverse the template. For each mapped read, starting and ending times and positions were determined and used to compute the distance traversed by the polymerase between sequence reads. Based on these determinations, an average time across any particular region of the human reference sequence was computed. Regions that were traversed by the polymerase more slowly have peaks of higher ΔT, and were indicative of insertions in the template relative to the *Homo sapiens* chromosome 15 reference sequence. The insertions identified were the same insertions identified above.

[00245] Intermittent illumination-based sequencing reactions across fosmid sequence AC223433 showed significant sequence read coverage across the insertion events. The distribution of the physical coverage is shown in Figure 12, which illustrates examples of three-look strobes (i.e., sequencing reactions having three detection/illuminated periods) that span or intersect the

insertion events. Figure 12A shows the mapping of the strobe sequence reads to the *Homo sapiens* chromosome 15 reference sequence, where the sequence reads generated from the insert sequences in the template are excluded. Arrows indicate the locations of the insertions. Figure 12B shows a similar mapping with the sequence reads generated from the insert sequences indicated with brackets. A number of sequence reads flank the insertions, connect the two insertions, or clarify sequence within (or at the boundaries of) the insertion sequences. Such flanking and connecting sequence reads are useful for predicting and detecting genomic events, anchoring them to genomic references, and scaffolding for de novo assembly of novel sequences. In particular, there are 30 and 38 "3-look" reads that intersect the two regions of insertion of (1192 bp and 6879 bp, respectively). These sequence reads facilitated mapping of the insertions to the human reference sequence, which would have been extremely difficult, if not impossible, with commercially available short-read sequencing technologies. Further the sequence of the smaller insertion was a highly repetitive sequence, which would also have made mapping difficult with certain short-read technologies.

[00246] Figure 13 illustrates the sequence coverage obtained across the fosmid sequence, showing all two-, three-, and four-look strobe sequence reads spanning the sequence that are mappable to the known AC223433 fosmid sequence.

A consensus sequence was derived from the set of mappable sequence reads generated in these sequencing reactions. Strobe sequence reads were combined with sequence reads generated under constant illumination and assembled based on the human reference sequence (Hg18). High quality reads surrounding the (suspected) insertion sites, as well as high quality reads that did not map to the reference sequence, were extracted and assembled with a "de novo" greedy suffix tree assembler; the resulting contigs were mapped to the Hg18 reference sequence. Contigs spanning the (suspected) insertion sites were identified and fed back into the "de novo" assembler, and the resulting contigs were manually edited using standard techniques and placed back into the derived reference guided assembly. The final consensus sequence was a hybrid of a reference guided assembly and attempts at de novo assembly of novel insert sequences. Alignments to reference sequences were performed and plotted. Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described above and the fosmid reference sequence, and this plot confirmed a high degree of alignment between the two sequences. This dot plot was generated using Gepard 1.21 ("GEnome PAir – Rapid Dotter," available from the Munich Information Center for Protein Sequences (MIPS)) with a word size of 7. Nucleic acid dot plots are widely used in the art and are further described, e.g., in Krumsiek et al. (2007) Bioinformatics

23(8):1026-8; Maizel et al. (1981) Proc Natl Acad Sci USA 78:7665; Pustell, et al. (1982) Nucleic Acids Res 10:4765; and Quigley, et al. (1984) Nucleic Acids Res 12:347, all of which are incorporated herein by reference in their entireties for all purposes.

Electronic Acknowledgement Receipt

EFS ID:	27847720
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	20-DEC-2016
Filing Date:	
Time Stamp:	14:06:52
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	01007706_2016-12-20_Trans.pdf	764536 a36fc4800ed4fb48af0342bf33ef4a1cf231532b4	no	1

Warnings:

Information:					
2		01007706_2016-12-20_PrelimAmend.pdf	94465 07974f89b1a12dc1e5a9c7d0d8900c6ca1c31c820	yes	8
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Preliminary Amendment		1	1	
	Specification		2	5	
	Drawings-only black and white line drawings		6	6	
	Applicant Arguments/Remarks Made in an Amendment		7	8	
Warnings:					
Information:					
3	Drawings-only black and white line drawings	01007706_2016-12-20_AddedFigs.pdf	33138 d36fb5f42c5d57103e68337859b26eeac8304b0f	no	2
Warnings:					
Information:					
4	Specification	01007706_2016-12-20_SubstSpec-Redline.pdf	488245 829f6011b3cb41a1c3e5e4426b917c872d489f49	no	106
Warnings:					
Information:					
5	Specification	01007706_2016-12-20_SubstSpec-Clean.pdf	487236 e481ac1aac2426fb014ba5f06397bf4dc7dc35fc	no	106
Warnings:					
Information:					
		Total Files Size (in bytes):	1867620		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/21 (10-08)
Approved for use through 11/30/2008. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	15/383,965
Filing Date	December 20, 2016
First Named Inventor	Stephen Turner
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	01-007706US

ENCLOSURES *(Check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☑ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☑ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation
- ☐ Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- ☐ Proprietary Information
- ☐ Status Letter
- ☑ Other Enclosure(s) (please Identify below):
  Substitute Specification (both redline and clean)

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Pacific Biosciences of California, Inc.		
Signature	/David C. Scherer, Ph.D./		
Printed name	David C. Scherer, Ph.D.		
Date	December 20, 2016	Reg. No.	56,993

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below: via EFS-Web

Signature	/Jacqueline L. Lim/		
Typed or printed name	Jacqueline L. Lim	Date	December 20, 2016

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 15/383,965

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A			N/A	280
SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A			N/A	600
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	720
TOTAL CLAIMS (37 CFR 1.16(i))	18 minus 20 =	*			OR	x 80 =	0.00
INDEPENDENT CLAIMS (37 CFR 1.16(h))	1 minus 3 =	*				x 420 =	0.00
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).						0.00
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))							0.00
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL			TOTAL	1600

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)	OR	RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/383,965	12/19/2016	Stephen Turner	01-007706US

CONFIRMATION NO. 8144

57770
PACIFIC BIOSCIENCES OF CALIFORNIA, INC.
1380 Willow Road
MENLO PARK, CA 94025

FORMALITIES LETTER

OC000000088161609

Date Mailed: 12/30/2016

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Filing Date Granted

An application number and filing date have been accorded to this application. The application is informal since it does not comply with the regulations for the reason(s) indicated below. Applicant is given TWO MONTHS from the date of this Notice within which to correct the informalities indicated below. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

The required item(s) identified below must be timely submitted to avoid abandonment:

- Replacement drawings in compliance with 37 CFR 1.84 and 37 CFR 1.121(d) are required. The drawings submitted are not acceptable because:
  - The application contains drawings and the specification contains a brief description of the drawings. However, the specification does not contain a brief description of the several views of the drawings as required by 37 CFR 1.74 and 37 CFR 1.77(b)(7) and/or a drawing(s) has not been labeled in accordance with 37 CFR 1.84(u)(1). If each figure is not labeled "Fig." with a consecutive Arabic numeral (1, 2, etc.) or an Arabic numeral and capital letter in the English alphabet (A, B, etc.), then the drawing(s) must be relabeled in accordance with 37 CFR 1.84(u)(1). In addition, if the brief description of the several views of the drawings does not refer to the figure(s) as properly labeled, then the specification must be amended to correspond to the figure(s) as properly labeled and a substitute specification in compliance with 37 CFR 1.52, 1.121(b)(3), and 1.125, is required.

Applicant is cautioned that correction of the above items may cause the specification and drawings page count to exceed 100 pages. If the specification and drawings exceed 100 pages, applicant will need to submit the required application size fee.

Replies must be received in the USPTO within the set time period or must include a proper Certificate of Mailing or Transmission under 37 CFR 1.8 with a mailing or transmission date within the set time period. For more information and a suggested format, see Form PTO/SB/92 and MPEP 512.

Replies should be mailed to:

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Registered users of EFS-Web may alternatively submit their reply to this notice via EFS-Web, including a copy of this Notice and selecting the document description "Applicant response to Pre-Exam Formalities Notice". https://sportal.uspto.gov/authenticate/AuthenticateUserLocalEPF.html

For more information about EFS-Web please call the USPTO Electronic Business Center at 1-866-217-9197 or visit our website at http://www.uspto.gov/ebc.

If you are not using EFS-Web to submit your reply, you must include a copy of this notice.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/cnguyen/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/383,965	12/19/2016	1634	1600	01-007706US	18	1

CONFIRMATION NO. 8144

57770
PACIFIC BIOSCIENCES OF CALIFORNIA, INC.
1380 Willow Road
MENLO PARK, CA 94025

FILING RECEIPT

OC000000088161808

Date Mailed: 12/30/2016

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Stephen Turner, Seattle, WA;
Jon Sorenson, Alameda, CA;
Kenneth Mark Maxham, Redwood City, CA;
John Eid, San Francisco, CA;

Applicant(s)

Pacific Biosciences of California, Inc., Menlo Park, CA;

Assignment For Published Patent Application

Pacific Biosciences of California, Inc.

Power of Attorney: The patent practitioners associated with Customer Number 57770

Domestic Priority data as claimed by applicant

This application is a CON of 14/708,603 05/11/2015
which is a CON of 14/091,961 11/27/2013 PAT 9057102
which is a CON of 12/982,029 12/30/2010 PAT 8628940
which claims benefit of 61/099,696 09/24/2008
and claims benefit of 61/139,402 12/19/2008
and is a CIP of 12/413,226 03/27/2009 PAT 8143030

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: No

Permission to Access Search Results: No

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 12/28/2016
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 15/383,965**
Projected Publication Date: To Be Determined - pending completion of Corrected Papers
Non-Publication Request: No
Early Publication Request: No
Title

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Preliminary Class

435

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

January 20, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
Inventor: Stephen Turner, et al.	Examiner: Not Yet Assigned
Application No.: 15/383,965	Confirmation No.: 8144
Filed: December 19, 2016	Art Unit: 1634
For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS	PRELIMINARY AMENDMENT IN RESPONSE TO NOTICE TO FILE CORRECTED APPLICATION PAPERS

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INTRODUCTORY REMARKS

This Preliminary Amendment is filed in response to the Notice to File Corrected Application Papers dated December 30, 2016 and prior to an action on the merits in this matter. Applicant respectfully requests entry of the amendments and remarks provided herein.

AMENDMENTS TO THE SPECIFICATION

Please replace paragraphs **[0039], [0040], [0041], [0042], [0043], [0046], [0050], [0057]** and **[0080]** with amended paragraphs **[0039], [0040], [0041], [0042], [0043], [0046], [0050], [0057]** and **[0080]** below. No new matter has been added.

[0039] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions, whether illumination is initiated before ([[A]]left) or after ([[B]]right) initiation of the reaction.

[0040] Figures 2A-2B provide[[s]] an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination, including depictions of multiple reactions arrayed on a solid support (Figure 2A) and prophetic data (Figure 2B) from certain embodiments of the invention.

[0041] Figures 3A–3C provide[[s]] an exemplary embodiment of analysis of a plurality of illuminated reactions on a solid support (Figure 3A) using intermittent illumination and a mask (Figure 3B). A graph (Figure 3C) depicts prophetic data from certain embodiments of the invention.

[0042] Figures 4A–4B provide[[s]] additional embodiments of masks for use in the methods of the invention, including a mask that allows illumination of columns of reactions (Figure 4A) and a mask that allows illumination of every other reaction in a row and column (Figure 4B).

[0043] Figures 5A–5D illustrate[[s]] an aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Figure 5A illustrates a solid support comprising four quadrants, each quadrant containing a different sample. Figure 5B illustrates a mask design for selective illumination of the substrate. Figures 5C and 5D demonstrate various positions of the mask on the solid support.

[0046] Figures 8A–8C provide[[s]] data from single-molecule sequencing-by-synthesis reactions. Figure 8A provides data from a two-minute interval beginning at initiation of the reactions, i.e., from 0-120 seconds. Figure 8B provides data from a

second two-minute interval from 300-420 seconds. Figure 8C provides data from a third two-minute interval from 600-720 seconds.

[0050] Figures 12A–12B provide[[s]] a distribution of the physical coverage of a template nucleic acid achieved during a sequencing reaction utilizing intermittent illumination, with Figure 12A showing mapping to a reference sequence with sequence reads (and portions thereof) that do not map to the reference excluded and Figure 12B showing a similar mapping that further includes sequence reads corresponding to insertions in the template that are absent from the reference sequence.

[0057] Figures 19A–19C provide[[s]] an illustrative example of nucleic acid templates having nicks.

[0080] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions. A reaction mix is prepared at step 100. In the process [[A]] shown on the left, illumination of the reaction 105 is begun prior to initiation of the reaction 110, which allows "illumination data" to be collected at initiation. (In an alternative embodiment, illumination may commence simultaneously with initiation of the reaction.) "Illumination data" as used herein refers to data collected during an illuminated period, e.g., the length of the illuminated period and luminescent signal(s) from the reaction product. At least one non-illuminated period 115 occurs during the course of the reaction, followed by at least one additional illuminated period 120. Multiple additional non-illuminated and illuminated periods may follow. During the illuminated periods (105 and 120), illumination data is collected 175. During the non-illuminated period(s), non-illumination data is collected 180. As used herein, "non-illumination data" refers to data collected during a non-illuminated period, e.g., the length of the non-illuminated period can be monitored. In the process [[B]] shown on the right, the reaction is initiated 155 during a first non-illuminated period 150. At least one illuminated period 160 occurs during the course of the reaction, optionally followed by at least one additional non-illuminated period 165. Multiple additional illuminated and non-illuminated periods may follow. As for process A, illumination data is collected 175 during the illuminated period(s) 160, and non-illumination data is collected 180 during non-illuminated periods (155 and 165).

AMENDMENT TO DRAWINGS

Please replace the drawings of Figures 1, 2, 3, 4, 5, 8, 12, 16 and 19 with the replacement drawings having Figures 1, 2A-2B, 3A-3C, 4A-4B, 5A-5D, 8A-8C, 12A-12B, 16A-16C and 19A-19C. No new matter has been added.

REMARKS

Amendments to the Specification

The Notice to File Corrected Application Papers dated December 30, 2016, indicates the brief description of the drawings in the specification does not refer to the figure(s) as properly labeled. Paragraphs **[0039], [0040], [0041], [0042], [0043], [0046], [0050]** and **[0057]** have been amended to ensure that the figures in the application are specifically delineated in the Brief Description of the Drawings. Paragraph **[0080]** has been amended to update reference to Figure 1 in view of the updates to this drawing. No new matter has been added.

Applicants submit herewith a marked copy of the specification showing the desired changes and a clean version of the specification with changes already incorporated.

Amendments to the Drawings

The drawings of Figures 1, 2, 3, 4, 5, 8, 12, 16 and 19 have been replaced with replacement figures to properly refer to Figures 1, 2A-2B, 3A-3C, 4A-4B, 5A-5D, 8A-8C, 12A-12B, 16A-16C and 19A-19C as recited in the specification. No new matter has been added.

CONCLUSION

Based upon the foregoing amendments and remarks, Applicant believes the instant application is in condition for examination and action toward that end is respectfully requested. If the Office believes there are additional issues that have not been addressed, the Office is encouraged to contact Applicant's undersigned representative at (650) 521-8127.

Respectfully submitted,

January 20, 2017
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF
CALIFORNIA, INC.
1305 O'Brien Drive
Menlo Park, CA 94025

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation application of U.S. Patent Application No. 14/708,603, filed May 11, 2015, which is a continuation application of U.S. Patent Application No. 14/091,961, filed November 27, 2013, now U.S. Patent No. 9,057,102, which is a continuation application of U.S. Patent Application No. 12/982,029, filed December 30, 2010, now U.S. Patent No. 8,628,940, which (1) claims the benefit of U.S. Provisional Application No. 61/099,696, filed September 24, 2008; (2) claims the benefit of U.S. Provisional Application No. 61/139,402, filed December 19, 2008; and (3) is a continuation-in-part application of U.S. Patent Application No. 12/413,226, filed March 27, 2009, now U.S. Patent No. 8,143,030, the full disclosures of all of which are incorporated herein by reference in their entireties for all purposes.

[0002] This application is also related to U.S. Provisional Application No. 61/072,160, filed March 28, 2008, U.S. Patent Application No. 12/383,855, filed March 27, 2009, now U.S. Patent No. 8,236,499, and U.S. Patent Application No. 12/413,258, filed March 27, 2009, now U.S. Patent No. 8,153,375, all of which are incorporated herein by reference in their entireties for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0003] Not Applicable.

BACKGROUND OF THE INVENTION

[0004] The use of optically detectable labeling groups, and particularly those groups having high quantum yields, e.g., fluorescent or chemiluminescent groups, is ubiquitous throughout the fields of analytical chemistry, biochemistry, and biology. In particular, by providing a highly visible signal associated with a given reaction, one can better monitor that reaction as well as any potential effectors of that reaction. Such analyses are the basic tools of life science research in genomics, diagnostics, pharmaceutical research, and related fields.

[0005] Such analyses have generally been performed under conditions where the amounts of reactants are present far in excess of what is required for the reaction in question. The result of this excess is to provide ample detectability, as well as to compensate for any damage caused by the detection system and allow for signal detection with minimal impact on the reactants. For example, analyses based on fluorescent labeling groups generally require the use of an excitation radiation

source directed at the reaction mixture to excite the fluorescent labeling group, which is then separately detectable. However, one drawback to the use of optically detectable labeling groups is that prolonged exposure of chemical and biochemical reactants to such light sources, alone, or when in the presence of other components, e.g., the fluorescent groups, can damage such reactants. The traditional solution to this drawback is to have the reactants present so far in excess that the number of undamaged reactant molecules far outnumbers the damaged reactant molecules, thus minimizing or negating the effects of the photo-induced damage.

[0006] A variety of analytical techniques currently being explored deviate from the traditional techniques. In particular, many reactions are based on increasingly smaller amounts of reagents, e.g., in microfluidic or nanofluidic reaction vessels or channels, or in "single molecule" analyses. Such low reactant volumes are increasingly important in many high throughput applications, such as microarrays. The use of smaller reactant volumes offers challenges to the use of optical detection systems. When smaller reactant volumes are used, damage to reactants, such as from exposure to light sources for fluorescent detection, can become problematic and have a dramatic impact on the operation of a given analysis. In other cases, other reaction conditions may impact the processivity, rate, fidelity, or duration of the reaction, including salt or buffer conditions, pH, temperature, or even immobilization of reaction components within observable reaction regions. In many cases, the effects of these different reaction or environmental conditions can degrade the performance of the system over time. This can be particularly detrimental, for example, in real-time analysis of reactions that include fluorescent reagents that can expose multiple different reactions components to optical energy. In addition, smaller reactant volumes can lead to limitations in the amount of signal generated upon application of optical energy.

[0007] Further, in the case of sequencing-by-synthesis applications, an additional challenge has been to develop ways to effectively sequence noncontiguous portions of a template nucleic acid on a single molecule. This challenge is exacerbated in template nucleic acids that contain highly repetitive sequence and/or are hundreds or thousands of nucleotides in length, such as certain genomic DNA fragments. The difficulty in generating such noncontiguous reads from a single template has hampered efforts to construct consensus sequences for long templates, for example, in genome sequencing projects.

[0008] As such, methods and systems that result in enhanced reaction performance, such as an increase in processivity, rate, fidelity, or duration of a reaction of interest, would provide useful improvements to the methods and compositions currently available. For example, methods, devices,

and systems that increase reaction performance by, e.g., mitigating to some extent photo-induced damage in a reaction of interest and/or increasing various other performance metrics for the reaction would be particularly useful.

BRIEF SUMMARY OF THE INVENTION

[0009] In a general sense, the methods provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0010] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one

reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion. The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0011] In certain aspects, the present invention is generally directed to methods, devices, and systems for enhancing the performance of illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. In certain preferred embodiments, illuminated reactions comprise one or more fluorescent or fluorogenic reactants. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In some aspects, enhancing the performance of an illuminated reaction means increasing the processivity, rate, fidelity, and/or duration of the reaction. For example, enhancing the performance of an illuminated reaction can involve reducing or limiting the effects of photo-induced damage during the reaction. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction.

[0012] In certain aspects, methods of the invention useful for characterizing an analytical reaction comprise preparing a reaction mixture and initiating the analytical reaction therein, subjecting the reaction mixture to at least one detection period and at least one non-detection period during the course of the analytical reaction, collecting data during both the detection period(s) and the non-detection period(s), and combining the collected data to characterize the analytical reaction. In certain embodiments, the analytical reaction comprises an enzyme that exhibits an improvement in performance as compared to its performance in the analytical reaction under constant illumination, and such improvement may be related to various aspects of enzyme activity, e.g., processivity, fidelity, rate, duration of the analytical reaction, and the like. In certain embodiments, stop or pause points are used to control the activity of the enzyme, and such stop or pause points may comprise elements such as large photolabile groups, strand-binding moieties, non-native bases, and others well known in the art. In certain preferred embodiments, the one or more detection periods are illuminated periods and the one or more non-detection periods are non-illuminated periods. In certain preferred embodiments, a plurality of analytical reactions disposed on a solid support are characterized, preferably in a coordinated fashion as described elsewhere herein.

[0013] In certain preferred embodiments, the analytical reaction is a sequencing reaction that generates sequence reads from a single nucleic acid template during the detection period(s) but not during the non-detection period(s). For example, the analytical reaction can comprise at least two or more detection periods and can generate a plurality of noncontiguous reads from the single nucleic acid template. In some embodiments, the single nucleic acid template is at least 100 bases in length and/or comprises multiple repeat sequences. In certain embodiments, the sequencing reaction comprises passage of the single nucleic acid template through a nanopore, and in other embodiments the sequencing reaction comprises primer extension by a polymerase enzyme.

[0014] The analytical may optionally be a processive reaction monitored in real time, i.e., during the course of the processive reaction. In preferred embodiments, such a processive reaction is carried out by a processive enzyme that can repetitively execute its catalytic function, thereby completing multiple sequential steps of the reaction. For example, a processive polymerization reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.

[0015] In certain preferred embodiments, the analytical reaction comprises at least one component comprising a detectable label, e.g., a fluorescently labeled nucleotide. In certain embodiments, the labeled component is present throughout the course of the analytical reaction, i.e., during both the detection and the non-detection periods. The method may further comprise an optical system to collect the data during the detection period, but optionally not to collect the data during the non-detection period.

[0016] In certain aspects, methods of the invention comprise providing a substrate having a reaction mixture disposed thereon and illuminating the reaction mixture on the substrate with an excitation illumination for multiple, noncontiguous periods during the course of the reaction, thereby subjecting the reaction mixture to intermittent excitation illumination. In some embodiments, the reaction mixture comprises first reactant and a second reactant, wherein an amount of photo-induced damage to the first reactant occurs as a result of interaction between the first reactant and the second reactant under excitation illumination. In certain embodiments, the method further comprises monitoring a reaction between the first and second reactants during illumination and collecting the data generated therefrom. In some embodiments, the reaction is a primer extension reaction and/or the first reactant is a polymerase enzyme. In certain embodiments, the second reactant is a fluorogenic or fluorescent molecule.

[0017] In yet another aspect, the methods are useful for mitigating photo-induced damage in an illuminated reaction by subjecting the illuminated reaction to intermittent illumination rather than constant illumination. For example, certain methods of the invention monitor a reaction mixture comprising at least one enzyme and a fluorescent or fluorogenic substrate for the enzyme, wherein interaction of the enzyme and the substrate under excitation illumination can result in altered activity of the enzyme, e.g if such excitation illumination is present over an extended period of time. Such methods can comprise directing intermittent excitation illumination at a first observation region for a first period that is less than a photo-induced damage threshold period under the intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions. As such, certain aspects of the invention lengthen a photo-induced damage threshold period for an analytical reaction through intermittent inactivation of the excitation illumination source since the photo-induced damage threshold period under intermittent illumination is longer than the photo-induced damage threshold period under constant illumination.

[0018] In a related aspect, the invention also provides methods of performing an enzyme reaction, comprising providing an enzyme within a first observation region, contacting the enzyme with a fluorescent or fluorogenic substrate for the enzyme, and directing an excitation radiation at and detecting signals from the first observation region for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0019] In further aspects, the invention provides methods of monitoring a primer extension reaction, comprising providing a polymerase enzyme within a first observation region, contacting the polymerase with at least a first fluorescent or fluorogenic nucleotide analog, and monitoring a fluorescent signal emitted from the first observation region in response to illumination with excitation radiation for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0020] In addition, the invention provides methods for generating a plurality of noncontiguous sequence reads from a single nucleic acid template molecule. Such methods generally comprise preparing a reaction mixture comprising the template molecule, a polymerase enzyme, and a set of differentially labeled nucleotides or nucleotide analogs, wherein the set comprises at least one type of nucleotide or nucleotide analog for each of the natural nucleobases (A, T, C, and G). The polymerization reaction is initiated, the polymerase begins processive incorporation of the labeled nucleotides or nucleotide analogs into a nascent nucleic acid strand, and during such incorporation the reaction is monitored by optical means to detect incorporation events, thereby generating a first sequence read. In a subsequent step, the labeled nucleotides or analogs are replaced with unlabeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed without detecting incorporation events. Subsequently, the unlabeled nucleotides or analogs are replaced with labeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed once again with real time detection of incorporation events, thereby generating a second sequence read that is noncontiguous to the first sequence read. The substitution of labeled for unlabeled, and unlabeled for labeled, nucleotides and nucleotide analogs can be repeated multiple times to generate a plurality of noncontiguous sequence reads, each of the plurality generated during a period when the labeled nucleotides or nucleotide analogs are being incorporated into the nascent strand and such incorporation is being detected in real time.

[0021] In certain aspects, devices of the invention can comprise a solid support (e.g., substrate) having an observation region, a first reactant immobilized within the observation region, and a second reactant disposed within the observation region, and a means for subjecting the observation region to at least one illuminated period and at least one non-illuminated period. In certain embodiments, interaction between the first and second reactants under excitation illumination causes photo-induced damage to the first reactant, and further wherein the photo-induced damage is reduced by subjecting the observation region to intermittent illumination. In some embodiments, the first reactant is an enzyme (e.g., a polymerase), the second reactant (e.g., a nucleotide) has a detectable label (e.g., fluorescent label), and/or the observation region is within a zero-mode waveguide. The means for subjecting the observation region to one or more illuminated and non-illuminated periods may comprise, e.g., a laser, laser diode, light-emitting diode, ultra-violet light bulb, white light source, a mask, a diffraction grating, an arrayed waveguide grating, an optic fiber, an optical switch, a mirror, a lens, a collimator, an optical attenuator, a filter, a prism, a planar waveguide, a wave-plate, a delay line, a movable support coupled with the substrate, and a movable illumination source, and the like. The device may further comprise a means for collecting the data during the illuminated period(s), such as an optical train, e.g., operably coupled to a machine comprising machine-readable medium onto which such data may be written and stored.

[0022] In further aspects, the invention provides systems for performing intermittent detection of an analytical reaction comprising reagents for the analytical reaction disposed on a solid support, a mounting stage configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the solid support detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other, and a data processing system operably coupled to the optical train. In certain preferred embodiments, the analytical reaction is a sequencing reaction and/or the solid support comprises at least one zero-mode waveguide.

[0023] In still other aspects, the invention provides systems for analyzing an illuminated reaction that is susceptible to photo-induced damage when illuminated for a period longer than an photo-induced damage threshold period, comprising a solid support having reagents for the reaction disposed thereon, a mounting stage supporting the solid support and configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the

solid support to illuminate the portion of the solid support and detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, and a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other. In some embodiments, the illuminated reaction is a sequencing reaction, e.g., a nucleotide sequencing-by-synthesis reaction. In certain embodiments, the solid support comprises at least one optical confinement, e.g., a zero-mode waveguide.

[0024] The invention provides methods of performing analytical reactions, e.g., processive analytical reactions, that include preparing a reaction mixture comprising reaction components, at least one of which is a detectable component that is detectable during one or more detection periods, and at least one of which is a clocking component that is detectable during one or more non-detection periods during the analytical reaction. The methods further comprise initiation the analytical reaction and maintaining conditions that allow the analytical reaction to proceed while subjecting it to at least one detection period and at least one non-detection period, both in the presence of the clocking component and the detectable component. In certain embodiments, the detectable component emits a detectable signal is response to excitation illumination during the detection period, but not during the non-detection period when a clocking signal is emitted from the clocking component. The detectable signal is collected during the detection period and the clocking signal is detected during the non-detection period, e.g., using an optical system. Optionally, the clocking signal can also be collected during the detection period and the non-detection period. In certain preferred embodiments, detection data is collected in read time during the detection period, non-detection data is collected in real time during the non-detection period, and the detection data and non-detection data are both used to characterize the analytical reaction. In some embodiments, the transition between the detection period and the non-detection period does not involve substitution and/or addition of reaction components during progression of the analytical reaction, and in other embodiments the transition does involve substitution and/or addition of reaction components, e.g., via a reaction mixture exchange. In some preferred embodiments, a plurality of analytical reactions are disposed on a solid support, subjected to intermittent illumination, monitored to collect data, and characterized based upon the data so collected.

[0025] The detectable component and clocking component are typically linked to discrete molecules in the analytical reaction. For example, the detectable component can be linked to a first subset of nucleotide analogs and the clocking component can be linked to a second subset of

nucleotide analogs in the analytical reaction mixture. Alternatively, both the detectable component and the clocking component can be linked to a single molecule, e.g., a single nucleotide or nucleotide analog, in the analytical reaction. The detectable component and clocking component can both comprise detectable labels (e.g., luminescent, fluorescent, or fluorogenic labels, including, e.g., quantum dots), and in some embodiments, different detectable labels, e.g. having different absorption peaks.

[0026] In certain preferred embodiments, an analytical reaction performed according to the invention comprises at least one enzyme, e.g., a polymerase, ligase, ribosome, nuclease, and/or kinase. In some embodiments, pause or stop points are engineered into the analytical reaction to control activity of the enzyme. Various aspects of the analytical reaction can be changed by being subjected to at least one detection period and at least one non-detection period, such aspects including but not limited to processivity, fidelity, rate, and duration, e.g. of enzyme activity.

[0027] In certain preferred embodiments, the analytical reaction is a sequencing reaction comprising a single nucleic acid template that generates sequence reads during the detection period by detecting the detectable component, and does not generate sequence reads during the non-detection period by suspending detection of the detectable component. Such a sequencing reaction typically comprises at least two or three detection periods and generates a plurality of noncontiguous sequence reads from the single nucleic acid template. In some embodiments, the template comprises multiple repeat or complementary sequences. In some embodiments, the sequencing reaction comprises passage of the single nucleic acid or a nascent strand complementary thereto through a nanopore. In some preferred embodiments, the sequencing reaction comprises primer extension by a polymerase enzyme and the detectable component is linked to a nucleotide or nucleotide analog. In some embodiments, the clocking component is linked to the polymerase enzyme, and optionally can be a multi-component label, e.g, a FRET label.

[0028] In certain aspects, the invention provides methods of mitigating photo-induced damage during an illuminated reaction that include preparing a reaction mixture having first and second reactants, where interaction of the reactants under excitation illumination can cause photo-induced damage to the first reactant. The illuminated reaction is subjected to intermittent excitation illumination characterized by periods of maximal illumination followed by periods of modified but not absent illumination. The intermittent excitation illumination reduces the amount of photo-induced damage to the first reactant during the illuminated reaction as compared to the illuminated reaction under constant maximal excitation illumination, thereby mitigating photo-induced damage

to the first reactant. In certain preferred embodiments, the illuminated reaction is a primer extension reaction. In certain preferred embodiments, the first reactant is an enzyme, e.g., a polymerase or ligase enzyme. In certain preferred embodiments, the second reactant comprises a fluorescent or fluorogenic molecule. In certain embodiments, the modified excitation illumination is illumination with a lower intensity excitation illumination than the maximal excitation illumination. In certain embodiments, a set of illumination sources provides the maximal excitation illumination and a subset of the set of illumination sources provides the modified excitation illumination.

[0029] In other aspects, the invention provides a method of sequencing a template nucleic acid that includes subjecting the template to methylation to generate at least one methylated base, subjecting the methylated base to base excision to generate at least one abasic site in the template, annealing a primer to the template nucleic acid, contacting the template with a polymerase enzyme to promote extension of the primer in a template-dependent manner, monitoring the extension of the primer in real time to generate a nucleotide sequence read complementary to the template, extending the primer until the abasic site is encountered by the polymerase, at which time the polymerase pauses on the template, and reinitiating primer extension by facilitating abasic site bypass by the polymerase. The monitoring, extending, and reinitiating steps are repeated until a desired number of nucleotide sequence reads is generated and collected, and subsequently analyzed to determine the sequence of the template nucleic acid. In certain embodiments, the contacting step occurs during a detection period or a detection period immediately follows the contacting step. In certain embodiments, a detection period ends and a non-detection period begins prior to one or more pauses of the polymerase on the template. In certain embodiments, a non-detection period is terminated simultaneous with or immediately following one or more reinitiation steps. In some embodiments, the reinitiating step comprises introduction of a pyrene to the polymerase, where the polymerase incorporates the pyrene into the nascent strand opposite and, therefore, "pairing with" an abasic site in the template. In certain preferred embodiments, the template is circular and the polymerase pauses at the same abasic site multiple times during the primer extension reaction. In other embodiments, the method further comprises terminating the monitoring when a desired length of the nucleotide sequence read is collected, e.g., by removing or modifying excitation illumination. Optionally, the desired length can be less than a length of the template nucleic acid. Additionally, the monitoring can be reinitiated subsequent to or simultaneous with the reinitiating of primer extension.

[0030] In yet further aspects, the invention provides a method of performing an illuminated reaction that includes preparing a reaction mixture comprising multiple optically detectable components that are distinguishable from one another based upon their individual signal emissions, initiating the illuminated reaction, and maintaining conditions that allow the illuminated reaction to proceed while subjecting the reaction mixture to at least one maximal illuminated period and at least one modified illuminated period during the illuminated reaction. In preferred embodiments, at least a portion of the optically detectable components are detectable during both the maximal and modified illuminated periods. In certain embodiments, the maximal illuminated period is characterized by a first excitation radiation intensity and the modified illuminated period is characterized by a second excitation radiation intensity that is less than the first excitation radiation intensity. In certain preferred embodiments, all of the optically detectable components are detectable during both the maximal and modified illuminated periods, but are distinguishable from one another during the maximal illuminated period, but are not distinguishable during the modified illuminated period. In certain embodiments, the maximal illuminated period comprises exposing the reaction mixture to a set of excitation radiation wavelengths and the modified illuminated period comprises exposing the reaction mixture to a subset of the set of excitation radiation wavelengths. In certain preferred embodiments, all of the optically detectable components are detectable and distinguishable during the maximal illuminated period, but only a subset of the optically detectable components are detectable during the modified illuminated period.

[0031] In some embodiments, the illuminated reaction is initiated during a modified illuminated period and subsequently subjected to a maximal illuminated period, where data collected during the modified illuminated period is used in the statistical analysis of data collected during the maximal illuminated period. For example, an illuminated reaction that is a polynucleotide sequencing reaction can generate sequence read data during a modified illuminated period that is subsequently used to construct a sequence scaffold for assembly of sequence read data collected during a maximal illuminated period. Additionally or optionally, the illuminated reaction is a template-directed sequencing reaction and sequence read data collected during a modified illuminated period is used to determine a rate of translocation of a polymerase during the modified illuminated period.

[0032] Some embodiments of the invention comprise performing a plurality of illuminated reactions, each of which is exposed to the set of excitation radiation wavelengths during the maximal illuminated period, but is exposed to a different subset of the set of excitation radiation

wavelengths during the modified illuminated period, such that a distinct subset of optically detectable components are detectable during the modified illuminated period for each of the plurality of illuminated reactions. In other words, for two such illuminated reactions, although all optically detectable components are detectable during their respective maximal illuminated periods, only a subset of the optically detectable components is detectable in each reaction, and the subset detectable in the first reaction is preferably different from the subset detectable in the second reaction.

[0033] In certain aspects, the invention provides methods for performing paired-end sequencing on a single template molecule. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion. A first linker ligated to the first terminal portion of the nucleic acid molecule connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and a second linker ligated to the second terminal portion of the nucleic acid molecule connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion. A template nucleic acid molecule is thereby formed comprising the double-stranded nucleic acid molecule with both the first linker and the second linker ligated thereto. The template molecule is subjected to a sequencing process in which sequence reads are generated for the first terminal portion and the second terminal portion, but sequence reads are not generated for the intermediate portion, even if the intermediate portion is processed during the sequencing process, e.g., by a polymerase. In some embodiments, the first linker and second linker are identical, and in other embodiments they are different from one another, i.e., not identical. In certain embodiments, the first and second linkers comprise complementary regions and can be hybridized to one another prior to one or both of the ligating steps. In some cases, hybridized linkers that are ligated to the ends of a double-stranded nucleic acid molecule are separated prior to subjecting the molecule to a sequencing reaction, and in some cases the hybridized linkers remain hybridized during at least a portion of the sequencing reaction. For example, in a template-directed sequencing reaction, a polymerase capable of strand displacement separates the hybridized linkers as it sequences the template. In certain preferred embodiments, the sequencing process comprises at least one detection period (e.g., an illuminated period) and at least one non-detection period (e.g., a non-illuminated period) such that the intermediate portion of the template molecule is subjected to the sequencing process during the non-detection period. In some embodiments, the template is fragmented after ligation to remove the intermediate portion. The

sequencing process can generate redundant sequence data from one or both of the first terminal portion and the second terminal portion, and/or can generate sequence data from an additional portion of the template molecule that is noncontiguous with the first terminal portion and the second terminal portion. In preferred embodiments, the sequencing process involves circularizing the template molecule by separating the complementary strands of the template molecule and using the complementary strands in template-directed nascent strand synthesis catalyzed by a single polymerase enzyme. Optionally, the template molecule can comprise a primer binding site, a registration sequence, and/or a synthesis blocking moiety. The primer binding site, a registration sequence, or synthesis blocking moiety can be present in one or both of the linkers, or can be located elsewhere within the template molecule. In some cases, the synthesis blocking moiety is selected from the group consisting of an abasic site, a nick, a synthetic linker, a non-native nucleotide or analog thereof, a primer, a large photolabile group, a strand-binding moiety, a damaged base, and a modified base. The synthesis blocking moiety can permanently or temporarily block progression of the sequencing process, e.g., by interfering with the activity of an enzyme, e.g., a polymerase enzyme. In certain preferred embodiments, the synthesis blocking moiety is an abasic site, e.g., introduced by a DNA glycosylase.

[0034] In some aspects, the invention provides methods for generating a nucleic acid construct for analytical reactions. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion; providing a first stem-loop linker hybridized to a second stem-loop linker; ligating the first stem-loop linker to the first terminal portion of the nucleic acid molecule, wherein the first stem-loop linker connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and ligating the second stem-loop linker to the second terminal portion of the nucleic acid molecule, wherein the second stem-loop linker connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion, thereby generating the nucleic acid construct. Optionally, the nucleic acid construct can be subjected to fragmentation after the ligating of steps c and d, wherein the fragmentation removes the intermediate portion from the nucleic acid construct and introduces two double-stranded termini. The method can further include ligating the two double-stranded termini to one another. In some embodiments, one of the stem-loop linkers comprises a primer binding site, registration sequence, or a synthesis blocking moiety that is absent from the other stem-loop linker.

[0035] In further aspects, the invention includes a single template nucleic acid molecule comprising a duplex region; a first linker linking termini at a first end of the duplex region; a second linker linking termini at a second end of the duplex region, wherein a region of the first linker is complementary to a region of the second linkers. Optionally, the single template molecule comprises the first and second linkers hybridized with one another. In some embodiments, the duplex region is separated or melted apart to transform the single template nucleic acid molecule into a topologically single-stranded, circular nucleic acid molecule. Further, the invention provides a composition comprising a single, optically resolvable polymerase enzyme in association with a single-stranded circular nucleic acid molecule, wherein the single-stranded circular nucleic acid molecule comprises first, second, third, and fourth regions, and further wherein the first region is complementary to the second region, and the third region is complementary to the fourth region, and further wherein the regions are ordered on the single-stranded circular nucleic acid molecule as follows: first region, third region, second region, fourth region.

[0036] In still further aspects of the invention, machine-implemented methods for transforming nucleotide sequence read data into consensus sequence data, wherein the nucleotide sequence read data is generated by sequencing a target region of a template nucleic acid multiple times, and the consensus sequence data is representative of a most likely actual sequence of the template nucleic acid. Such machine-implemented methods can comprise various steps, such as a) mapping the nucleotide sequence data to a target sequence using a local alignment method that produces a set of local alignments comprising an optimal local alignment and sub-optimal local alignments, b) enumerating the set of local alignments, c) constructing a weighted directed graph wherein each local alignment in the set of local alignments is represented as a node, thereby generating a set of nodes in the weighted directed graph, d) drawing edges between pairs of nodes in the weighted directed graph if the pair represents a potential reconstruction of the template nucleic acid, e) assigning weights to the edges drawn in step d, wherein a given weight for a given edge represents the log-likelihood that a given pair of nodes connected by the given edge is truly a reconstruction of the template nucleic acid, f) finding the shortest path to each node in the weighted directed graph, thereby generating a set of shortest paths for the weighted directed graph, g) ranking the set of shortest paths to determine the best assignment, and h) storing the results of steps a-g on a machine-readable medium. In certain embodiments, the steps of the machine implemented methods are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Also provided are

computer program products comprising a computer usable medium having computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the machine-implemented methods of the invention, and machine-readable medium on which the results of the method steps are stored. The invention further includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods.

[0037] In certain aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme (e.g., a polymerase) during a time t, where time t occurs during a non-detection period. Such a method comprises, in certain embodiments, developing a probability model *p(v)* to describe an observed distribution of enzyme velocities during one or more detection periods; sampling velocities from *p(v)*; summing and recording the velocities sampled in step b to produce a sum that is an estimate of x/τcorr; and repeating the sampling, summing, and recording M times to generate a distribution of sums that are estimates of x/τcorr, with the distribution of sums being the distribution of the distance x travelled by an enzyme during a time t. Preferably, at least some of the steps are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Optionally, the enzyme is a polymerase enzyme. In some embodiments, multiple enzymes are observed simultaneously and the probability model *p(v)* is determined independently for each of the multiple enzymes. In certain preferred embodiments,

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}.$$

[0038] In further aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme during a time t, where time t occurs during a non-detection period. In some embodiments, the method comprises estimating a distribution of local rates p(v), making independent identically distributed draws of N = t/τcorr velocities from from p(v); summing the velocities; recording the velocities summed in c) as an estimate of x/τcorr; and repeating b-d M times, e.g., where M is preferably at least 1000. Optionally, p(v) is determined using a Hidden Markov Model or the autocorrelation function $\langle \delta v(t)\delta v(t+\Delta) \rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$. The invention further

includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods, as well as a machine-readable medium on which the results of the steps of the methods are stored.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions, whether illumination is initiated before ([[A]]left) or after ([[B]]right) initiation of the reaction.

[0040] Figures 2A-2B provide[[s]] an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination, including depictions of multiple reactions arrayed on a solid support (Figure 2A) and prophetic data (Figure 2B) from certain embodiments of the invention.

[0041] Figures 3A–3C provide[[s]] an exemplary embodiment of analysis of a plurality of illuminated reactions on a solid support (Figure 3A) using intermittent illumination and a mask (Figure 3B). A graph (Figure 3C) depicts prophetic data from certain embodiments of the invention.

[0042] Figures 4A–4B provide[[s]] additional embodiments of masks for use in the methods of the invention, including a mask that allows illumination of columns of reactions (Figure 4A) and a mask that allows illumination of every other reaction in a row and column (Figure 4B).

[0043] Figures 5A–5D illustrate[[s]] an aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Figure 5A illustrates a solid support comprising four quadrants, each quadrant containing a different sample. Figure 5B illustrates a mask design for selective illumination of the substrate. Figures 5C and 5D demonstrate various positions of the mask on the solid support.

[0044] Figure 6 provides an illustration of paths in a sequence alignment matrix representing sequencing data from a SMRTbell™ template.

[0045] Figure 7 illustrates a hypothetical directed graph.

[0046] Figures 8A–8C provide[[s]] data from single-molecule sequencing-by-synthesis reactions. Figure 8A provides data from a two-minute interval beginning at initiation of the reactions, i.e., from 0-120 seconds. Figure 8B provides data from a second two-minute interval from 300-420 seconds. Figure 8C provides data from a third two-minute interval from 600-720 seconds.

[0047] Figure 9 schematically illustrates one embodiment of a system for use with the methods, devices, and systems of the invention.

[0048] Figure 10 provides a graphical representation of rates of polymerase activity on different portions of a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0049] Figure 11 provides a graphical representation of the average rate of polymerase translocation over a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0050] Figures 12A–12B provide[[s]] a distribution of the physical coverage of a template nucleic acid achieved during a sequencing reaction utilizing intermittent illumination, with Figure 12A showing mapping to a reference sequence with sequence reads (and portions thereof) that do not map to the reference excluded and Figure 12B showing a similar mapping that further includes sequence reads corresponding to insertions in the template that are absent from the reference sequence.

[0051] Figure 13 provides a distribution of the physical coverage provided by sequence reads generated during sequencing reactions utilizing intermittent illumination across an approximately 40 kb template nucleic acid.

[0052] Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described herein and a reference sequence.

[0053] Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system.

[0054] Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters $P(S \rightarrow P) = 1/24$; $P(P \rightarrow S) = 1/11$; and $p(v) \sim \text{Gamma}(48, 0.25)$. Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a non-detection period.

[0055] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period.

[0056] Figure 18 provides an exemplary strategy for selectively reducing the size of a duplex fragment within a SMRTbell™ template.

[0057] Figures 19A–19C provide[[s]] an illustrative example of nucleic acid templates having nicks.

[0058] Figures 20A and 20B illustrate two exemplary embodiments of template constructs used in the present invention.

[0059] Figures 21A and 21B schematically illustrate redundant or consensus sequencing using the constructs shown in Figures 20A and 20B.

DETAILED DESCRIPTION OF THE INVENTION

[0060] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing devices, formulations and methodologies which are described in the publication and which might be used in connection with the presently described invention.

[0061] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polymerase" refers to one agent or mixtures of such agents, and reference to "the method" includes reference to equivalent steps and methods known to those skilled in the art, and so forth. Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either both of those included limits are also included in the invention.

[0062] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention. Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

I. General

[0063] In a general sense, the methods, devices, and systems provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0064] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third

portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion.

[0065] The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. For example, such an alteration in reaction conditions during the second stage may result in an increase in reaction rates, e.g., speeding up the progression of a template nucleic acid through a nanopore; or may reduce exposure of reaction components to harmful radiation or other reaction condition related to detection of the products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0066] "Intermittent detection," as used herein, generally refers to a means of monitoring a reaction that is carried out intermittently during the course of the reaction. Intermittent detection may refer to intermittent use of one or more monitoring methods, but does not necessarily mean that all means of monitoring a given reaction are intermittently halted. For example, monitoring of one or more nucleotide incorporations to generate nucleotide sequence reads may be intermittently halted while other aspects of a sequencing reaction are constantly monitored, e.g., temperature, reaction time, pH, etc. In certain embodiments, intermittent detection is achieved by intermittent or differential illumination of a given reaction, e.g., a reaction that uses an illumination system to detect reaction products and/or progression. Although various aspects of the invention are described herein in terms of embodiments using intermittent illumination, it should be understood that where applicable intermittent detection by other means (e.g., electrochemical, radiochemical, etc.) can be utilized in the methods of the invention. Likewise, a stage of a reaction during which an intermittent detection method is active may be referred to as a "detection period" and a stage of a reaction during which an intermittent detection method is inactive may be referred to as a "non-detection period." In illuminated reactions, such periods may also be referred to as "illuminated periods" and "non-illuminated periods," respectively, although it is to be understood that the term "non-

illuminated period" included periods in which illumination may be present but altered as compared to illumination during an "illuminated period." For example, a non-illuminated period may be characterized by a complete absence of illumination, or a modification of illumination, including but not limited to changes in wavelength, frequency, intensity, and/or number of illumination sources. Alternatively or additionally, reaction components that are excited by the illumination source(s) may be modified or removed from a reaction mixture to create a non-illuminated period. For example, a fluorescent dye detected during an illuminated period may be removed from the reaction mixture, e.g., by buffer exchange, thereby producing a non-illuminated period during which time the fluorescent dye cannot be detected even if the excitation illumination is present. In a further example, a non-illuminated period can indicate a period during an illuminated reaction during which a type of illumination-based detection that occurs during an illuminated period is not occurring, e.g., the identity of fluorescently labeled nucleotides incorporated into a nascent strand is not being detected or recorded.

[0067] In certain aspects, the present invention is generally directed to improved methods, devices, and systems for performing illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In certain preferred embodiments, illuminated reactions comprise one or more fluorogenic or fluorescent components. In accordance with certain methods of the invention, such illuminated analyses are subjected to intermittent detection (e.g., data collection) for one or more aspects of the data typically collected for a given reaction. For example, aspects of the data typically collected for nucleotide sequencing reactions include nucleotide sequence data, read quality data, signal to background ratios, reaction rates and durations, measures of the fidelity of the reaction, reaction times, and the like. In certain preferred embodiments, nucleotide sequence data is iteratively collected during an ongoing sequencing reaction to generate nucleotide sequence reads for at least two or more noncontiguous regions of a template nucleic acid molecule. Such iterative sequence data acquisition may be achieved in various ways depending on the sequencing technology in use. For example, in sequencing methods that utilize luminescent components that generate a signal indicative of the identity of a base position, iterative sequence data collection may be achieved by removing or altering an illumination source (or a reaction relative to an illumination source),

substituting the luminescent components for unlabeled components that do not generate signal, or otherwise interrupting signal acquisition in the experimental system.

[0068] In certain preferred embodiments, such illuminated reactions are illuminated for an amount of time that permits the effective performance of the analysis. Traditionally, illuminated reactions are illuminated from initiation through completion, and the time during which reaction data may be reliably collected is dictated by the progression (as measured by, e.g., processivity, rate, fidelity, duration, etc.) of the reaction under constant illumination. Some reactions are sensitive to such constant illumination, which can reduce their performance (e.g., processivity), and thereby prevent collection of data from later stages of the reaction, i.e., stages that would otherwise occur if the reaction were carried out with no illumination. The present invention provides methods for performing illuminated reactions comprising subjecting the reactions to intermittent illumination. Such intermittent illumination can increase performance (e.g., processivity, rate, fidelity, duration, etc.) of the reactions, thereby allowing generation of data that cannot be collected under constant illumination, such as data from later stages of an ongoing reaction whose progression is compromised under constant illumination. For example, in sequencing-by-incorporation reactions the use of intermittent excitation illumination can increase processivity, which has the benefit of providing sequence reads more distal from the polymerase binding/initiation site than such reactions subjected to constant exposure to excitation illumination.

[0069] Further, it is an object of the instant invention to provide sequence data from noncontiguous regions of a nucleic acid template in a single reaction. Other commercially available platforms have attempted to achieve such noncontiguous sequence data through, e.g., complex cloning and sequencing strategies. The present invention provides a clear advantage over such strategies by providing a simple and economical solution that is applicable across various platforms, and is particularly applicable to illuminated, single-molecule sequencing-by-incorporation reactions.

[0070] In preferred embodiments, illuminated reactions for use with the instant invention are nucleic acid sequencing reactions, e.g., sequencing-by-incorporation reactions. In preferred embodiments, such an illuminated reaction analyzes a single molecule to generate nucleotide sequence data pertaining to that single molecule. For example, a single nucleic acid template may be subjected to a sequencing-by-incorporation reaction to generate one or more sequence reads corresponding to the nucleotide sequence of the nucleic acid template. For a detailed discussion of such single molecule sequencing, see, *e.g.*, U.S. Patent Nos. 6,056,661, 6,917,726, 7,033,764,

7,052,847, 7,056,676, 7,170,050, 7,361,466, 7,416,844; Published U.S. Patent Application Nos. 2007-0134128 and 2003/0044781; and M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, SCIENCE 299:682-686, January 2003 Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations, all of which are incorporated herein by reference in their entireties for all purposes. In some embodiments, a plurality of single nucleic acid templates are analyzed separately and often simultaneously to generate a plurality of sequence reads corresponding to the nucleotide sequences of the plurality of nucleic acid templates. In certain preferred embodiments, the plurality of nucleic acid templates includes at least two nucleic acid templates that comprise identical nucleotide sequences such that analysis of the two nucleic acid templates generates overlapping sequence reads. In certain preferred embodiments, at least one of the nucleic acid templates is configured to provide redundant sequence data in a single sequence read, e.g., via duplications, sense and antisense sequences, and/or circularization.

[0071] Certain aspects of the invention are directed to methods, devices, and systems for generating a sequence scaffold for a nucleic acid template, e.g., chromosome, genome, or portion thereof. A sequence scaffold as used herein refers to a set of sequence reads that extends across at least a portion of a nucleic acid template. In some embodiments, such a sequence scaffold is used to generate a consensus sequence for the nucleic acid template. In some embodiments, the nucleic acid template is very large, e.g., at least about 100, 1000, 10,000, 100,000, or more bases or base pairs in length. In some embodiments, the sequence scaffold and/or consensus sequence is based on at least 1-, 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of at least a portion of the nucleic acid template. In some preferred embodiments, the portion of the nucleic acid is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the entire length of the nucleic acid template.

[0072] In certain aspects, the invention is particularly suitable for sequencing nucleic acid templates interspersed with repetitive elements. Such repetitive elements present major logistical and computational difficulties for assembling fragments produced by sequencing strategies, especially those with read-lengths that are too short to encompass unique reads outside the repeat region. For example, the human T-cell receptor locus contains a five-fold repeat of a trypsinogen gene that is 4 kbp long and that varies 3 to 5% between copies. Therefore, a sequencing strategy that cannot provide nucleotide sequence information that spans at least 20 kb for a single molecule containing the locus will have difficulty providing consensus sequence for the locus. Further, Alu repeats (~300 bp retrotransposons) are also problematic because they cluster and can constitute up

to 50-60% of the template sequence, with copies varying from 5-15% between each other. The human genome contains an estimated one million Alu repeats and 200,000 LINE elements (average length ~1000 bp), representing roughly 10% and 5% of the entire genome, respectively. In certain embodiments, the present methods facilitate efficient and accurate sequence determination for long templates comprising such repetitive sequences, in part because the present methods do not rely solely on sequence overlap to generate consensus sequences, but also include information related to the expected location of the polymerase on the template nucleic acid, thereby linking a particular sequence read to a particular location on the template nucleic acid. This greatly facilitates accurate assembly of sequence reads to generate sequence scaffolds and/or consensus sequences.

[0073] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads in an illuminated sequencing-by-incorporation reaction that are distal from one another (i.e., noncontiguous) on a single nucleic acid template by removing the excitation illumination during the course of the reaction, and subsequently reinitiating the excitation illumination. Sequence reads are generated only during the periods of time when the excitation illumination is present, resulting in a "gap" between the sequence reads from a single template nucleic acid that corresponds to the time during which the excitation illumination was absent but the incorporation of nascent nucleotides continued "in the dark." As such, the number of sequence reads generated for a given template nucleic acid is equal to the number of periods during which the excitation illumination is present.

[0074] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads from a plurality of nucleic acid templates comprising identical nucleotide sequences. In some embodiments, the multiple sequence reads are not all from the same region of the nucleic acid templates. In some embodiments, there is overlap between the multiple sequence reads. In some embodiments, a single sequence read is generated from each of the plurality of nucleic acid templates, and in other embodiments multiple noncontiguous sequence reads are generated from each of the plurality of nucleic acid templates. In certain preferred embodiments, the multiple noncontiguous sequence reads from each of the plurality of nucleic acid templates together extend across the nucleic acid templates such that they can be combined to provide a consensus sequence for the identical nucleotide sequence in the nucleic acid templates. In some embodiments, the consensus sequence is based on at least 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of the identical nucleotide sequence. In some embodiments, the

identical nucleotide sequence represents at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the nucleic acid template.

[0075] Certain aspects of the invention are directed to methods, devices, and systems for reducing or limiting the effects of photo-induced damage during illuminated reactions, particularly reactions that employ fluorescent or fluorogenic reactants. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction. Without being bound to a particular theory or mechanism of operation, some illuminated reactions are subject to photo-induced damage that can hinder progression of the reaction, e.g., via damage to reaction components, such as enzymes, cofactors, templates, etc. As such, the illumination of the illuminated reaction can directly or indirectly negatively impact progression of the reaction, and such an impact can be measured based on various characteristics of the reaction progression, e.g., processivity, rate, fidelity, duration, etc. The present invention provides methods for subjecting an illuminated reaction to intermittent exposure to illumination, which reduces the amount of photo-induced damage at a given time during the reaction, allowing the reaction to proceed further than it does when constantly exposed to the illumination.

[0076] In some embodiments, the methods herein may further comprise the addition of one or more photo-induced damage mitigating agents (e.g., triplet-state quenchers and/or free radical quenchers) to the illuminated reaction. Such photo-damage mitigating agents are generally known to those of skill in the art. Further discussion of photo-induced damage and related compounds, compositions, methods, devices, and systems are also provided in U.S. Pub. No. 20070161017, filed December 1, 2006; and U.S.S.N. 61/116,048, filed November 19, 2008, which are incorporated by reference herein in their entireties for all purposes.

II. Intermittent Illumination of Analytical Reactions

[0077] Certain aspects of the invention are generally directed to improved methods for performing illuminated analyses. The terms "illuminated analysis" and "illuminated reaction" are used interchangeably and generally refer to an analytical reaction that is occurring while being illuminated (e.g., with excitation radiation), so as to evaluate the production, consumption, and/or conversion of luminescent (e.g., fluorescent) reactants and/or products. As used herein, the terms "reactant" and "reagent" are used interchangeably. As used herein, the terms "excitation illumination" and "excitation radiation" are used interchangeably. In certain embodiments, the

illuminated reaction is a sequencing reaction, e.g., a sequencing-by-incorporation reaction. In certain embodiments, the illuminated reaction is designed to analyze a single molecule, e.g., by ensuring the molecule is optically resolvable from any other molecule being analyzed and/or in the reaction mixture. In certain embodiments, one or more components of the reaction are susceptible to photo-induced damage directly or indirectly elicited by an excitation radiation source. In certain preferred embodiments, an illuminated reaction is subjected to intermittent excitation radiation during the course of the illuminated reaction. In certain preferred embodiments, a sequencing-by-incorporation reaction is subjected to intermittent excitation radiation during the course of a polymerization reaction to generate a plurality of noncontiguous sequence reads from a single nucleic acid template.

[0078] In certain aspects, the methods herein provide benefits over methods currently used for sequencing large template nucleic acids, such as human genomes. For example, the traditional shotgun sequencing approach entails sequencing nucleic acid fragments and analyzing the resulting sequence information for overlap and similarity to known sequences to construct the complete sequence of the template nucleic acid. One disadvantage to the shotgun approach is that assembly may be difficult if the template nucleic acid comprises numerous repeated sequences, and the inability to assemble a genomic sequence in repeat regions leads to gaps in the assembled sequence. (See, e.g., Myers, G.; "Whole-Genome DNA Sequencing" in Computing in Science and Engineering; Vol 1, Issue 3; pgs. 33-43; May/Jun 1999.) One method of resolving these gaps is to sequence fragments large enough to span the repeat regions, but sequencing large fragments can be difficult and time-consuming. Another approach to spanning a gap is to determine the sequence of two ends of a large fragment which has known spacing and orientation, and this approach is generally termed paired end sequencing (see, e.g., Smith, M. W. et al., (1994) Nature Genetics 7:40-47; and U.S. Pub. No. 2006/0292611, filed June 6, 2006, both of which are incorporated by reference herein in their entireties for all purposes). This method is limited by the requirement for information about the spacing and orientation of the ends of the long fragment, and/or complex sample preparation of the nucleic acid template. The present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.) or complex sample preparation, thereby allowing economical, efficient, and effective de novo sequencing or resequencing of long template nucleic acids.

[0079] In certain aspects, the methods herein provide various strategies for achieving intermittent illumination of illuminated reactions. Essentially, at least one type of illumination (e.g., excitation illumination) is present for at least one time period ("illuminated period") and absent during at least one other time period ("non-illuminated period") during an illuminated reaction. As described above, the term "non-illuminated" indicates a change in illumination including, but not limited to a complete absence of illumination. For example, a non-illuminated period may also be characterized by a different illumination source or intensity than an illuminated period, or by a change in reaction components, e.g., detectable labels. In general, at least one type of data collected during an illuminated period (e.g., nucleotide sequence data) is not collected during a non-illuminated period. An absence of the illumination may be due to, e.g., inactivation of the illumination source (e.g., laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source), removal of the illuminated reaction from the illumination source (or vice versa), or may be due to blockage of the illumination from the reaction, as discussed below. Modifications to the illumination may be due to, e.g., adjustment of the intensity of an illumination source, or a substitution of one illumination wavelength and/or frequency for another. Further, components detectable during an illuminated period may be removed from the reaction mixture during a non-illuminated period, e.g., a fluorescently labeled nucleotide may be replaced with an unlabeled nucleotide. Knowledge of the rate of the reaction and the time during which the illumination is absent is used to estimate the progress of the reaction during the non-illuminated period. For example, if a reaction proceeds such that one molecule is incorporated into a macromolecule per second, and the illumination is absent for 20 seconds, it can be estimated that 20 molecules were incorporated during the non-illuminated period. This information is useful during data analysis to provide context for the reaction data collected during the illuminated period(s). For example, in a sequencing-by-incorporation reaction the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. Further the processivity of an enzyme during a non-illuminated period (or other type of non-detection period) can be manipulated or adjusted by methods known to those of skill in the art. In

particular, the kinetics of replication by a polymerase enzyme can be altered by changing the chemical environment in which it operates, and such methods are further described, e.g., in U.S. Patent Application Nos. 12/414,191, filed March 30, 2009; 12/537,130, filed August 6, 2009; and U.S. Patent Application No. [unassigned], attorney docket no. 105-006301US, entitled "Engineering Polymerases and Reaction Conditions for Modified Incorporation Properties," filed September 4, 2009, the disclosures of all of which are incorporated herein by reference in their entireties for all purposes. For example, methods are provided for adjusting the enzyme activity, and these methods find particular relevance in the instant invention when used to enhance accuracy during detection periods, and to enhance processivity during non-detection periods. Information regarding enzyme translocation rate and processivity is useful for positioning the sequence reads for a single template nucleic acid relative to one another in the construction of a sequence scaffold and/or consensus sequence for the template nucleic acid.

[0080] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions. A reaction mix is prepared at step 100. In the process [[A]] shown on the left, illumination of the reaction 105 is begun prior to initiation of the reaction 110, which allows "illumination data" to be collected at initiation. (In an alternative embodiment, illumination may commence simultaneously with initiation of the reaction.) "Illumination data" as used herein refers to data collected during an illuminated period, e.g., the length of the illuminated period and luminescent signal(s) from the reaction product. At least one non-illuminated period 115 occurs during the course of the reaction, followed by at least one additional illuminated period 120. Multiple additional non-illuminated and illuminated periods may follow. During the illuminated periods (105 and 120), illumination data is collected 175. During the non-illuminated period(s), non-illumination data is collected 180. As used herein, "non-illumination data" refers to data collected during a non-illuminated period, e.g., the length of the non-illuminated period can be monitored. In the process [[B]] shown on the right, the reaction is initiated 155 during a first non-illuminated period 150. At least one illuminated period 160 occurs during the course of the reaction, optionally followed by at least one additional non-illuminated period 165. Multiple additional illuminated and non-illuminated periods may follow. As for process A, illumination data is collected 175 during the illuminated period(s) 160, and non-illumination data is collected 180 during non-illuminated periods (155 and 165).

[0081] One benefit provided in certain embodiments of the invention is that the reaction need not be further manipulated after initiation (aside from the control of illumination). For

example, the method can be used to analyze reaction mixtures without the need for buffer changes, addition of further reaction components, or removal of detectable components, e.g., light-activatable components such as fluorophores. For example, in a sequencing-by-incorporation reaction, labeled nucleotides may be present throughout the life of the reaction, even when the reaction is not generating nucleotide sequence data (e.g., during a non-illuminated period). This provides clear advantages over methods that require additional handling of the reaction after initiation, which tend to not only be expensive and time-consuming, but which also provide opportunities for contamination of the reaction. For example, illumination can be reinitiated at any time during the reaction at the whim of the ordinary practitioner by simply activating the illumination. In certain preferred embodiments, the concentration of labeled nucleotides or nucleotide analogs in the reaction mixture is greater than the concentration of unlabeled nucleotides in the reaction mixture throughout the course of the reactions, and may represent at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the total nucleotides in the reaction mixture. Methods for ensuring a high ratio of labeled versus unlabeled nucleotides in a reaction mixture are known in the art and certain preferred embodiments are provided in U.S. Patent Pub. Nos. 2006/0063264, 2006/0194232, and 2007/0141598, which are incorporated herein by reference in their entireties for all purposes.

[0082] In embodiments in which a sequencing-by-incorporation reaction is subjected to intermittent illumination, the sequence reads collected during the illuminated periods are arranged in order and separated from one another by an estimated number of nucleotides incorporated into the nascent strand during the intervening non-illuminated periods. The resulting gapped read can then be used to assess certain characteristics of the template nucleic acid. When multiple identical template nucleic acids are subjected to such a sequencing-by-incorporation reaction, the resulting set of gapped reads can be combined to create a sequence scaffold and/or a consensus sequence for the template nucleic acid.

[0083] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods can be used to provide additional data during the course of the reaction, e.g., data from illuminated or non-illuminated periods. In certain preferred embodiments, such alternative labeling methods may comprise using labels that are incorporated into a product of the reaction. For example, in sequencing-by-incorporation reactions that use nucleotides comprising labeled terminal phosphates (*e.g.*, the gamma phosphate as in dNTP, or terminal phosphates on nucleotide analogs with a greater number of phosphate groups) to identify

the nucleotides incorporated into a nascent polynucleotide, the reaction mixture may also include nucleotides comprising a base-linked label. During the reaction, these "base-labeled nucleotides" will be incorporated into the nascent strand, but unlike the terminal phosphate labels removed during incorporation, the base-linked labels are not cleaved from the nucleotide upon incorporation by the polymerase, resulting in a nascent strand that comprises the base-linked labels. The concentration of such base-labeled nucleotides can be adjusted in the reaction mixture to promote their incorporation into the nascent strand at a predictable rate, e.g., based on the known sequence of the template or the average frequency of a given nucleotide. The presence and/or rate of incorporation of the base-linked labels into the nascent strand can provide a measure of the length of the nascent strand generated (and, therefore, the distance traveled by the polymerase along the template nucleic acid) during the reaction by subjecting the reaction to excitation illumination that excites the base-linked label (but preferably not the non-base-linked labels), and detecting the signal emitted. The excitation of the base-linked labels preferably occurs as a pulse during or immediately following a non-illuminated period, and is otherwise absent during the reaction. The strength of the signal is indicative of how many labels are present in the nascent strand, thereby providing a measure of the processivity of the polymerase for a given period during the ongoing reaction, e.g. during one or more illuminated or non-illuminated periods. Since the base-linked labels remain in the nascent strand, it is beneficial to minimize the amount of time those fluorophores are subjected to excitation illumination to mitigate the potential of photo-induced damage to the reaction components. As such, in preferred embodiments, the excitation illumination wavelength for the base-labeled nucleotides is different than that of other fluorescent labels in the reaction.

[0084] This method can be modified in various ways. For example, the base-labeled nucleotides may also comprise a terminal phosphate label so that their incorporation can be monitored in the same manner during an illuminated period as the non-base-labeled nucleotides. There may be a single type of base-labeled nucleotide in a reaction mixture, or multiple types may be present, e.g., each type carrying a different nucleobase. The concentration of base-labeled nucleotides in the reaction mix may be varied, although it is preferred that the ratio of base-labeled nucleotides to non-base-labeled nucleotides be relatively low. For example, in a reaction mixture comprising a single type of base-labeled nucleotide (e.g., base-labeled dATP), it is preferred that the ratio of base-labeled dATP to non-base labeled dATP be less than 1:8, and more preferably 1:10 or less. The low concentration of base-labeled nucleotides is preferred in order to minimize sterically induced polymerase stalling when incorporating multiple base-labeled nucleotides in a row. In some

embodiments, the optimal ratio is pre-determined using capillary electrophoresis for any specific base-labeled nucleotide and likely homopolymer sequence prevalence. In certain preferred embodiments, at least 50, 75, 100, 125, or 150 base-labeled nucleotides are incorporated into the nascent strand during a single non-detection period. The base-labeled nucleotides may be present throughout the reaction, or may be washed in during non-illuminated periods and washed out after the pulse of excitation illumination. The reaction mixture comprising base-linked nucleotides being washed in may also include unlabeled nucleotides for incorporation during a non-detection period. During a subsequent illuminated period, a reaction mixture comprising terminal phosphate-labeled nucleotides replaces the reaction mixture comprising base-linked nucleotides and unlabeled nucleotides. This protocol is one embodiment of the methods of the invention in which a non-detection period is not necessarily a non-illuminated period because in this case illumination may be present, but no incorporation of nucleotides is be detected.

[0085] Alternatively or in addition, a low concentration of a fifth terminal phosphate labeled nucleotide can be present in the sequencing reaction, wherein the label has a different excitation wavelength than the other labels in the reaction mixture. For example, a small proportion of one nucleotide analog, e.g., dA6P, can be labeled with the "fifth label." During non-detection periods when the sequence of incorporation of nucleotides is not being monitored, the reaction site is illuminated by excitation radiation specific for the fifth label, and this fifth label excitation radiation can be inactivated during the detection periods. Emissions detected upon incorporation of the nucleotide analog comprising the fifth label are used to "clock" the pace of the polymerase during the non-detection period, e.g., based upon the known or estimated frequency of the complementary nucleotide in the template strand. The fifth label can be chosen such that the excitation and emission radiation are less likely or unlikely to cause photo-induced damage to reaction components, e.g. by choosing a label with a long excitation wavelength (e.g., toward the red end of the visible spectrum), a label that has a low propensity for entering into a triplet state, and/or a label that has a low propensity to form a radical. Since the fifth label is being excited when other labels are not, there is no requirement for optimal spectral separation from other labels in the reaction mixture. Further, since the fifth label is not being used for sequencing, other optimizations are also not necessary, e.g., related to branching, accuracy, and the like. Various types of labels can be used as a fifth label of the invention including, but not limited to, organic and non-organic dye fluorophores. For example, latex nanoparticles or quantum dots are particularly suitable due to their lower propensity for photo-induced damage of certain analytical reaction components. In certain preferred

embodiments, a quantum dot label has an emission spectrum within the same spectral window as the labels that are used to identify the sequence of base incorporations into the nascent strand ("sequencing labels") but an excitation spectrum that does not overlap those of the sequencing labels to allow detection of the fifth label emissions using the same optical system as is used to detect the sequencing label emissions.

[0086] This method can be modified in various ways. For example, more than one small subset of a nucleotide analog can be labeled with a fifth label, and in certain embodiments, a small subset of each nucleotide analog present in the reaction mixture is labeled with the fifth label. Further, there may be a plurality of additional labels present in the reaction, each of which is present on a small subset of a single type of nucleotide analog, e.g., sixth, seventh, and eighth labels. By increasing the number of types of nucleotide analogs labeled with fifth (or sixth, seventh, eighth) labels, their frequency of incorporation is likewise increased, which improves the translocation rate calculation for the polymerase during the non-detection periods. Alternatively, each type of nucleotide analog can comprise both a sequencing label that is specific for the cognate base in the nucleotide, as well as a fifth label for clocking the polymerase. The sequencing labels are excited and detected during the detection periods and the fifth labels are excited and detected during the non-detection periods. Since every nucleotide analog is labeled with a fifth base, each incorporation event can be counted during the non-detection period and the exact rate of incorporation can be determined. Both the sequencing and fifth labels may be bound to the same or different linkers on the nucleotide analogs. In certain preferred embodiments, a linker on a nucleotide analog positions the fifth label within an illumination zone to allow excitation, but far from an enzyme (e.g., polymerase) to mitigate photo-induced damage related to excitation of and/or emission from the fifth label.

[0087] In some embodiments, the fifth label is also excited by an illumination during the detection periods. The availability of the clocking function during the detection period can be used during sequence analysis to identify positions in the resulting sequence read where a signal was not detected (resulting in an apparent "missing base" in the read) and to distinguish between true insertions and branching events in which two signals are detected for a single incorporation event.

[0088] In yet further embodiments, assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid is facilitated by using "non-illuminated periods" characterized by modified excitation illumination rather than a complete absence of excitation illumination (which can also be termed "low-illuminated periods"). For example, in some

embodiments a lower intensity excitation illumination is used during the non-illuminated periods that excites one or more of the labels that are excited during the illuminated periods. As such, unlike various strategies described above, no fifth label is necessary. The lower intensity excitation illumination results in emissions that are lower intensity but still intense enough to identify an emission signal over background counts, though typically not intense enough to be used to identify the particular label generating the emission signal. For example, if label "A" and label "B" are in a reaction mixture, during an illuminated period the intensity of the signal emissions from each are high enough that the artisan can distinguish from which label a particular signal originates by the wavelength and/or frequency of the signal. However, during a low-illuminated period the artisan can only identify that a signal emission occurs, but is unable to distinguish the originating label because its particular wavelength and/or frequency cannot be accurately determined. The decrease in excitation illumination intensity provides both a mitigation of photo-induced damage to reaction components within the observation volume while allowing the practitioner to count the emissions, and therefore the incorporations, during the non-illuminated period.

[0089] In other embodiments, multiple excitation illumination sources are used during an illuminated period, and a first subset of these illumination sources is removed during a non-illuminated period, while a second subset remains. The illumination sources that remain during the non-illuminated period may be present in the same manner as during the illuminated period, or various aspects may be altered, e.g., intensity may be reduced. For example, if labels A and B present in a reaction mixture are excited by a first illumination source and labels C and D present in the reaction mixture are excited by a second illumination source, removal of the first illumination source during the non-illuminated period results in an inability to detect labels A and B, while C and D are still detectable. Such an incomplete data set can be used to clock the progress of the reaction during the non-illuminated period(s). Further, it can also be used in various ways to facilitate the statistical analysis of data collected during the illuminated period(s). For example, for nucleotide sequencing applications (as described elsewhere herein) the incomplete data set(s) collected during non-illuminated period(s) can be used during assembly of a sequence scaffold. For example, during *de novo* sequence assembly a collection of sequences (contigs) are generated, but the order of the contigs relative to the template nucleic acid is not always apparent. The scaffolding process uses extra information to determine the correct order of the contigs. So, if only two bases are identifiable in the non-illuminated periods, the incomplete sequence reads comprising only incorporation of these two bases can be aligned to modified versions of the contigs assembled from

data collected during an illuminated period, but in which the two bases not detected during the non-illuminated periods have been removed. Once the order of the contigs has been determined, the incorporation data for the two bases not detected during the non-illuminated periods is restored and the assembly of the contigs is complete. This method can be modified in various ways. For example, the practitioner may choose which illumination sources to remove during the non-illuminated periods based on various characteristics, such as their propensity to cause photo-induced damage to one or more reaction components, the propensity of the corresponding emission signal to cause photo-induced damage to one or more reaction components; their energy consumption; and wear-and-tear on the source device. Further, as described elsewhere herein, rather then removing an illumination source, reaction components that are excited by the illumination source may be removed from the reaction mixture during the non-illuminated period, necessarily rendering them undetectable. For example, one or more fluorescently labeled nucleotide analogs may be replaced with unlabeled nucleotide analogs during the non-illuminated periods.

[0090] In certain aspects, the invention provides advantages to performing intramolecular redundant sequencing, in which a template nucleic acid is used to generate multiple copies of a sequence read of interest, whether by virtue of multiple copies of the complement being present in the template, repeated replication of the template, or a combination thereof. For example, a first stage of a template-dependent sequencing reaction on a single-stranded circular template can comprise a non-illuminated period during which the template is completely replicated at least one time to generate at least one incomplete sequence read for a sequence complementary to the template. The first stage is followed by a second stage comprising an illuminated period during which the template is replicated multiple times to generate multiple complete sequence reads for the complementary sequence. The incomplete reads generated in the first stage can be used to construct a scaffold for assembly of the complete sequence reads generated in the second stage. Further, incomplete sequence reads can also be used to clock the progress of the reaction during the non-illuminated periods by providing a count of the detectable reaction components and combining that information with known or estimated characteristics of the template, e.g., nucleotide composition or sequence.

[0091] The subset of signal emissions detectable in the non-illuminated periods as compared to the number detectable in the illuminated periods is not limiting and may be chosen based upon the non-illumination data desired by the ordinary practitioner and/or other considerations, such as mitigation of photo-induced damage to extend readlength. For example, to lower the likelihood of

photo-induced damage, the ordinary practitioner may choose to remove the illumination source that is most damaging, e.g., has the highest frequency. In certain embodiments, multiple sequencing reactions may be performed for a single amplified template, each with a different combination of illumination sources and/or detectable components. Alternatively or additionally, multiple replicate reactions can also be performed for one or more of the combinations of illumination sources and/or detectable components. The combination of data from multiple different and/or replicate reactions performed on a single template provides myriad benefits during statistical analysis. As noted above, data can be combined to facilitate assembly of contigs generated during illuminated periods. Data from non-illuminated periods can also provide value in assessing the quality of the sequence reads generated during the illuminated periods.

[0092] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods used to provide additional data during the course of the reaction can comprise using labels that are incorporated into an enzyme of the reaction. For example, FRET labels can be used to label portions of a polymerase enzyme such that the conformational change between the open and closed states of the enzyme change the FRET value. For example, a FRET-based system can be used to monitor the kinetics of opening and closing of the finger subdomain of DNA polymerase, as described in Allen, et al. (2008) Protein Science 17:401-408, incorporated herein by reference in its entirety for all purposes. In certain preferred embodiments, a closed conformation produces a FRET signal because the donor and acceptor are close to one another, and an open conformation silences the signal because there is no energy transferred between the donor and acceptor. By monitoring the emission from the FRET pair, each incorporation event can be monitored during non-detection periods, and optionally or additionally during detection periods. In certain preferred embodiments, the FRET donor is GFP (excitation at 484 nm; emission at 510 nm), and the FRET acceptor is YFP (excitation at 512 nm; emission at 529 nm). Methods for monitoring polymerase activity using FRET labels are known in the art, e.g., in WO/2007/070572 A2, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

[0093] A given reaction may experience one or a plurality of illuminated periods or non-illuminated periods, but preferably experiences at least two illuminated periods. For example, a given reaction providing nucleotide sequence information from a single template nucleic acid may have at least about 2, 3, 5, 10, 20, 50, or 100 illuminated periods with intervening non-illuminated

periods. In an embodiment employing multiple periods of illumination and/or non-illuminated, the periods may be the same for both, e.g., 100 seconds "on" and 100 seconds "off." Alternatively, the illuminated periods may be longer or shorter than the non-illuminated periods. For example, in certain embodiments, a non-illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent illuminated period; or an illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent non-illuminated period. Further, each illuminated period may be the same or different from each other illuminated period, and each non-illuminated period may be the same or different from each other non-illuminated period. For example, some embodiments generate a smaller number of long reads, and other embodiments generate a larger number of short reads. It will be understood that the number and length of the illuminated and non-illuminated periods is limited only by the experimental system in use and the data acquisition goals of the ordinary practitioner. In some embodiments, a nucleotide sequence read generated during a single illuminated period comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, a region of a nucleic acid template processed during a non-illuminated period during a single reaction comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, the set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises at least about 40, 60, 80, 100, 1000, 10,000, 25,000, 50,000, 100,000, 250,000, 500,000, or 1,000,000 nucleotide sequence positions from a single nucleic acid template. In some embodiments, a set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises multiple reads of at least a portion of the nucleotide sequence positions from a single nucleic acid template.

[0094] As noted above, the present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.). However, such information, if available, may also be useful to the ordinary practitioner in determining an optimal periodicity for illuminated and non-illuminated periods during a sequencing reaction, especially when sequencing repetitive sequences. For example, if a genomic region is known to contain five adjacent copies of a one kilobase nucleotide sequence (i.e., five "repeat regions"), it would be beneficial to keep the non-illuminated periods short enough to be able to confidently map the resulting sequence reads to the correct repeat region. If a non-illuminated period were too long, the natural variation in translocation rate of the

polymerase would make it difficult to assign a sequence read to a particular repeat region, especially those farther from the binding/initiation site of the polymerase. In a further example, if the "copies" each had a few mutations that could be used to distinguish them from each other, it would be beneficial to keep the illuminated periods long enough to increase the chance one of these mutations would be included in a resulting sequence read, thereby allowing the unambiguous assignment of the read to a particular repeat region. If the illuminated period were too short the sequence reads from two different repeat regions could be identical, making mapping the sequence read challenging. (Another way to mitigate these difficulties would be to incorporate pause or stop points into the template nucleic acid, as discussed below.)

[0095] Essentially, the practitioner may design the number of and lengths of time for each illuminated and non-illuminated period to best suit the illuminated reactions being analyzed and the invention is not limited in this regard. In certain embodiments, a practitioner may wish to increase the processivity of a polymerase thereby extending the length of the template nucleic acid processed in a sequencing reaction to be, e.g., at least 2-, 3-, 4-, 6-, 8-, 10-, or 20-fold, thereby generating sequence data much farther away from the polymerase binding/initiation site than would be achieved under constant illumination. In certain embodiments, a practitioner of the instant invention may wish to focus on data from one or more stages of an ongoing reaction, such as stages for which more data is required for analysis. In the case of sequencing-by-synthesis, one or more particular regions of a template nucleic acid may need to be resequenced. Some traditional methods require that new template nucleic acids be prepared to bring a region requiring resequencing closer to the initiation point of the sequencing reaction, or require preparation of multiple new templates if multiple regions to be resequenced. In contrast, the methods herein allow the practitioner to subject a template identical to the previously sequenced template (e.g., from a large genomic DNA sample preparation) to a sequencing reaction wherein illuminated periods are timed to illuminate the sample only when the polymerase is incorporating nucleotides into the nascent strand at the one or more particular regions requiring resequencing. This advantage substantially lowers the time and resources required for such resequencing operations, therefore providing a significant advantage over traditional methods.

[0096] The instant invention contemplates various means for providing non-illuminated periods during illuminated reactions. In some embodiments, the illumination source is turned off during the ongoing reaction to create one or more non-illuminated periods. In some embodiments, the illumination source remains on during the course of the reaction, but the illuminated reaction is

removed from the system for a period of time. In some embodiments, the illumination source remains on during the course of the reaction, but the illumination is blocked to create one or more non-illuminated periods. For example, a movable mask may be manually or mechanically positioned between the illumination source and the illuminated reaction to block the illumination during non-illuminated periods and removed to allow exposure to the illumination during illuminated periods. Such a mask may also be dynamically controlled, such as a thin film transistor display (e.g., an LCD mask). Masks for blocking illumination and manufacture thereof are well known to those of ordinary skill in the art and need no further elaboration herein.

[0097] One aspect of the present invention is multiplexing of large numbers of single-molecule analyses. For a number of approaches, e.g., single molecule methods as described above, it may be desirable to provide the reaction components in individually optically resolvable configurations, such that a single reaction component or complex can be individually monitored. Providing such individually resolvable configurations can be accomplished through a number of mechanisms. For example, by providing a dilute solution of complexes on a substrate surface suited for immobilization, one will be able to provide individually optically resolvable complexes. (See, e.g., European Patent No.1105529 to Balasubramanian, et al., the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Alternatively, one may provide a low density activated surface to which complexes are coupled. (See, e.g., Published International Patent Application No. WO 2007/041394, the full disclosure of which is incorporated herein by reference in its entirety for all purposes). Such individual complexes may be provided on planar substrates or otherwise incorporated into other structures, e.g., zero-mode waveguides or waveguide arrays, to facilitate their observation.

[0098] In some embodiments, a plurality of illuminated reactions are carried out simultaneously, e.g., on a solid support. In some preferred embodiments, a solid support comprises an array of reaction sites. In preferred embodiments, the reaction sites on a solid support are optically resolvable from each other. In further preferred embodiments, each of the reaction sites on a solid support contains no more than a single reaction to be interrogated. For example, in a sequencing-by-incorporation embodiment, each reaction site preferably has no more than one polymerase and no more than one nucleic acid template. The reaction sites may be confinements (e.g., optical and/or physical confinements), each with an effective observation volume that permits resolution of individual molecules present at a concentration that is higher than one nanomolar, or higher than 100 nanomolar, or on the order of micromolar range. In certain preferred embodiments,

each of the individual confinements yields an effective observation volume that permits resolution of individual molecules present at a physiologically relevant concentration, *e.g.*, at a concentration higher than about 1 micromolar, or higher than 50 micromolar range or even higher than 100 micromolar. In addition, for purposes of discussion herein, whether a particular reagent is confined by virtue of structural barriers to its free movement, or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined."

[0099] As used herein, a solid support may comprise any of a variety of formats, from planar substrates, e.g., glass slides or planar surfaces within a larger structure, e.g., a multi-well plates such as 96 well, 384 well and 1536 well plates or regularly spaced micro- or nano-porous substrates, or such substrates may comprise more irregular porous materials, such as membranes, aerogels, fibrous mats, or the like, or they may comprise particulate substrates, e.g., beads, spheres, metal or semiconductor nanoparticles, or the like. The solid support may comprise an array of one or more zero-mode waveguides or other nanoscale optical structures.

[00100] As used herein, "zero-mode waveguide" refers to an optical guide in which the majority of incident radiation is attenuated, preferably more than 80%, more preferably more than 90%, even more preferably more than 99% of the incident radiation is attenuated. As such high level of attenuation, no significant propagating modes of electromagnetic radiation exist in the guide. Consequently, the rapid decay of incident electromagnetic radiation at the entrance of such guide provides an extremely small observation volume effective to detect single molecules, even when they are present at a concentration as high as in the micromolar range. The fabrication and application of ZMWs in biochemical analysis, and methods for calling bases in sequencing-by-incorporation methods are described, e.g., in U.S. Patent Nos. 7,315,019, 6,917,726, 7,013,054, 7,181,122, and 7,292,742, U.S. Patent Pub. No. 2003/0174992, and U.S. Patent Application No. 12/134,186, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00101] A set of reactions (e.g., contained on a solid support) may comprise identical or different components. For example, a single template nucleic acid may be analyzed in all reactions in the set, or a plurality of template nucleic acids may be analyzed, each present in only one or a subset of the set of reactions. In preferred embodiments, template nucleic acids comprising the same nucleotide sequence are analyzed in a plurality of reactions sufficient to provide adequate redundant nucleotide sequence data to determine a consensus sequence for the template nucleic acids. A number of sequence reads that will provide adequate nucleotide sequence data will vary, depending,

e.g., on the quality of the template nucleic acid and other components of the reaction, but in general coverage for a template nucleic acid or portion(s) thereof is at least about 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage. Further, the numbers and lengths of illuminated and non-illuminated periods for a given reaction in the set of reactions may be the same or different than those for other reactions in the set. In some embodiments, a mixture of different periodicities are used for a set of reactions comprising the same template nucleic acid. This strategy can be beneficial for providing nucleotide sequence reads from varying regions of the template sequence, thereby increasing the likelihood of overlapping sequence reads between individual reactions. These overlapping sequence reads can facilitate construction of a more robust sequence scaffold than could be constructed were the reactions all subjected to the same periodicity of illuminated and non-illuminated periods.

[00102] Methods of controlling polymerase progress and/or synchronizing polymerases in different reactions are also useful in analysis (e.g., mapping, validation, etc.) of nucleic acid reads farther from the initial binding site of the polymerase. During detection periods earlier in the reaction (i.e., closer to the time at which the polymerase began to process the template nucleic acid, such as during a first illuminated period), the position of a polymerase on the template can be estimated with generally good accuracy based on the known translocation rate of the polymerase under a given set of reaction conditions. As the duration of the reaction increases, however, the natural variation in polymerase translocation rate makes it more difficult to accurately determine the exact position of the polymerase on a template using estimation based on translocation rate alone; and through each subsequent illuminated period such estimations of polymerase position become less accurate, making subsequent analysis and mapping of the sequence reads to the template more difficult. Methods of regulating the position of the polymerase on the template allow more accurate determinations the polymerase's position. For example, causing the polymerase to pause or stop at a given location on the template during a non-illuminated period and reinitiating the polymerization during or immediately prior to a subsequent illuminated period provides a way to reorient the subsequently generated read with the template sequence, allowing easier consensus sequence determination and mapping analyses. Further, such pause/stop points can provide a means of controlling what regions of the template are processed during the illuminated periods by restricting where the polymerase will reinitiate on the template, thereby allowing a practioner of the instant invention to target one or more particular regions of a template for analysis during one or more detection periods during the course of an analytical reaction. Such methods are also useful to

synchronize a set reactions being monitored simultaneously. For example, a plurality of reactions, each comprising a single polymerase/template complex, may be synchronized by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that show less spreading (i.e., less variation in the position on the template from which the sequence reads are generated) in the later stages of the reactions than would otherwise be observed without such regulation.

[00103] Various methods can be used to control or monitor the progress of a polymerase on a template nucleic acid. For example, as noted above, one may employ a reaction stop or pause point within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the single-stranded portion that was not used in priming. Reaction stop or pause points can be engineered into a portion of the template for which the nucleotide sequence is unknown (e.g., a genomic fragment), but is preferably located within a portion for which the nucleotide sequence is known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, certain preferred sequencing templates (e.g., SMRTbell™ templates, described elsewhere herein) are closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, and one or both of these linkers can comprise a stop or pause point to control the passage of the polymerase through them. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00104] By way of example, at a selected time following initiation of polymerization the reaction may be subjected to a non-illuminated period. The incorporation of a synthesis blocking moiety coupled to the template nucleic acid at a position encountered by the polymerase during the non-illuminated period will cause the polymerase to pause. An example of an engineered pause point is a known sequence on the template nucleic acid where a primer sits and blocks progression of a polymerase that is actively synthesizing a complementary strand. The presence of the primer by itself could introduce a pause in the polymerase sequencing or the primer could be chemically modified to force a full stop (and synchronization of multiple polymerases in multiple reactions). The chemical modification could be subsequently removed (for example, photo-chemically) and the polymerase would subsequently continue along the template nucleic acid. In some embodiments,

multiple primers could be included in a reaction to introduce multiple pause or stop points along the template nucleic acid. Other methods for inducing a reversible pause (stop) in synthesis are known in the art and include, e.g., reversible sequestering of required cofactors (e.g., Mn^{2+}, one or more nucleotides, etc.). Once sufficient time has passed that the polymerase is paused at the blocking group, illumination is reintroduced and the blocking group removed. This allows control of the position on the template nucleic acid at which the polymerase will begin generating nucleotide sequence data during the illuminated period. A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the template nucleic acid that inhibit polymerase mediated replication, strand-binding moieties that prevent processive synthesis, non-native nucleotides included within the primer and/or the template, and the like. Such reaction stops/pause points are useful in providing more certainty about the relationship of the reads to each other. For example, since the exact position on a template nucleic acid at which each sequence read begins would be known, the resulting reads could be better mapped relative to one another for construction of a sequence scaffold and/or consensus sequence. Further description of these and other methods for regulating the progress of a polymerase on a template are provided, e.g., in U.S.S.N. 61/099,696, U.S. Patent Pub. No. 2006/0160113, and U.S. Patent Pub. No. 2008/0009007, all of which are incorporated by reference herein in their entireties for all purposes.)

[00106] By way of example, a sequencing reaction may be initiated on a template comprising a non-native base in the absence of the complement to the non-native base, which would not impact the overall sequence determination of other portions of the template that are complementary to native bases. By starving the reaction for the complement to the non-native base, one can prohibit synthesis, and thus, the sequencing process, until the non-native base complement is added to the mixture. This can provide a "hot start" capability for the system and/or an internal check on the sequencing process and progress that is configurable to not interfere with sequence analysis of the regions of interest in the template, which would be complementary to only native bases. In some embodiments, the non-native base complement in the sequence mixture is provided with a detectably different label than the complements to the four native bases in the sequence, and the production of incorporation-based signals associated with such labels provides an indication that the polymerase has initiated or reinitiated. Although described as the "non-native base" it will be appreciated that this may comprise a set of non-natural bases that can provide multiple control elements within the template structure. In certain embodiments, two different non-native bases are included within the template structure, but at different points, to regulate procession of the

sequencing process, e.g., allowing controlled initiation and a controlled stop/start position later in the sequence, e.g., prior to a subsequent illuminated period. For example, the complement to the first non-native base can be added to initiate sequencing immediately prior to the start of a first illuminated period. During a first non-illuminated period following the first illuminated period, the polymerase encounters the second non-native base, e.g., at a nucleotide position near but upstream of a nucleotide region desired to be sequenced in a second illuminated period. Sequencing would stop until the complement to the second non-native base is added to the reaction mixture. Likewise, multiple such non-native bases could be incorporated into the template to effectively target the polymerase to multiple regions of interest for which sequence data is desired. Further, in applications in which multiple identical templates are being sequenced, this would allow a resynchronization of the various sequencing reactions and the data generated therefrom.

[00107] Methods of controlling polymerase progress in different stages of a sequencing reaction are also useful for not only creating "condition-dependent" non-detection periods (during which time illumination may or may not be present), but also for minimizing the amount of time required for traversing a given length of template during a non-detection period (whether or not illumination is present). In order to reliably detect incorporation events, non-natural reagent conditions are typically used to limit polymerization during detection periods to approximately 1-5, or about 3 bases per second. In certain embodiments, replacement of Mg^{2+} ions with Mn^{2+} ions serves to stabilize and slow the translocation of the polymerase. When magnesium and, optionally, native nucleotides (e.g., lacking fluorescent labels) are used, the rate of translocation and/or processivity of the polymerase may increase up to two orders of magnitude. Use of such "rapid translocation" conditions during the non-detection periods can provide myriad benefits, including but not limited to a more rapid polymerization rate, an increased processivity (e.g., due to decreased stalling and misincorporation), and an overall savings due to reduced use of expensive labeled nucleotide analogs and/or reagents that mitigate oxidative stress.

[00108] In certain embodiments, a protocol for intermittent detection comprises alternating reaction mixtures, where a first reaction mixture used during the detection periods is optimized for sequence read generation, and a second reaction mixture used during the non-detection periods is optimized for processivity and/or rapid polymerization. For example, when reagents for optimal sequence read generation are present, DNA synthesis rate is low, and there is a fluorescence signal associated with each incorporation event. After replacing the reaction mixture optimized for sequence read generation with the reaction mixture optimized for processivity and/or rapid

polymerization, the polymerase rapidly advances across the template. In certain embodiments, a flow cell is used to deliver and switch between the two (or more) reaction mixtures during the course of the reaction.

[00109] In an exemplary embodiment, a first reaction mixture comprises fluorescently-labeled nucleotide analogs and manganese ions that restrict polymerization to a rate appropriate for high fidelity detection of nucleotide incorporation. The first reaction mixture can also include additional agents for mitigation of photo-induced damage of various components of the reaction mixture. A second reaction mixture comprises natural nucleotides and an appropriate magnesium ion concentration for rapid synthesis of the nascent strand complementary to the template. A first detection period of a sequencing reaction is initiated by introduction of the first reaction mixture, and a sequence read is generated based upon synthesis of the nascent strand during the detection period. After a predetermined time interval a sufficient quantity of the second reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, thereby initiating a first non-detection period. As noted above, the lack of labeled nucleotides in the second reaction mixture alone can produce the non-detection period, since there will be no signal emitted coincident with incorporation of the native nucleotides, but in certain embodiments illumination may also be removed, e.g., to further mitigate photo-induced damage during the non-detection period. At a time appropriate to initiate a second detection period, a sufficient quantity of the first reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, and detection of incorporation event is reinitiated. The cycle of reaction mixture exchange is repeated to generate multiple detection and non-detection periods.

[00110] A flow cell for reaction mixture exchange preferably has two inputs that are gated such that only a single reaction mixture flows into a reaction site or plurality of reaction sites, e.g., on a substrate. A single out-flow line may be used to remove reaction mixtures from the reaction site(s) to a single collection vessel, or multiple collection vessels may be used, one for each type of reaction mixture used. Further, accurate estimation of the distance a polymerase translocates during a non-detection period is important for bioinformatics applications. This estimation is complicated if the time for reaction mixture exchange is slow. As such, the flow is preferably at a sufficient rate that the time for exchange is significantly less than the time spent in the presence of either reaction mixture alone.

[00111] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination. In this embodiment, sixteen sequencing-by-incorporation reactions are performed on single nucleic acid templates (each of which comprises the same nucleotide sequence) with the timing of the illuminated and non-illuminated periods the same for all sixteen reactions. In A, the sixteen reactions are shown disposed on sixteen reaction sites on a solid support and are numbered for convenience. A representation of the illumination data is shown in B, with bars extending across the graph indicative of illumination data collected during illuminated periods for each reaction. In this illustrative example, each reaction is subjected to three illuminated periods, each followed by a non-illuminated period, resulting in three noncontiguous sequence reads for each reaction, i.e., three noncontiguous reads per template molecule sequenced. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. During the first illuminated period, the sequence reads generally overlap, but the natural variation of polymerase translocation rate over the set of reactions results in a "spreading" of the sequence reads as the reaction proceeds through the second and third illuminated periods with increasing variation in the exact position of each polymerase on the template at the beginning and end of each illuminated period. As such, the earlier illumination data provides better redundancy ("oversampling") of sequence information over a relatively narrow portion of the template nucleic acid, while the later illuminated periods provide less redundant sequencing data over a broader region of the template nucleic acid. The timing of the non-illuminated periods between the illuminated periods and the known or calculated rate of incorporation are used to determine approximate spacing between the resulting sequence reads, providing context for building a sequence scaffold or consensus sequence. It is important to note that although shown disposed on a solid support in A, the data shown in B could also have been generated from reactions not disposed on a solid support nor performed simultaneously and the methods are generally not so limited. Further, as described above, the spreading of the sequence reads from later stages of the reactions can be mitigated by synchronizing the reactions, e.g., by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that provides better redundancy (i.e., more overlap in the positions on the template from which the sequence reads are generated), especially in the later stages of the reactions.

[00112] Using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is

generated, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be reset to change the regions of the template for which nucleotide sequence data is generated. As described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest. Alternatively, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be kept the same to provide a greater-fold coverage of one or more regions of interest in the template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which are incorporated by reference herein in their entireties for all purposes. A specific embodiment is also provided in the Exemplary Applications section herein.

[00113] The present invention provides novel template configurations and methods for exploiting these compositions in template directed sequencing processes. While these compositions and methods have utility across all of the various template directed processes described herein, for ease of discussion, they are being primarily discussed in terms of preferred single molecule, real-time sequencing processes, in which they provide myriad benefits. In particular, the present invention is generally directed to nucleic acid sequences that employ improved template sequences to improve the accuracy of sequencing processes. For example, in at least one aspect, the template compositions of the invention are generally characterized by the presence of a double stranded segment or a pair of sub-segments that are internally complementary, i.e., complementary to each other. In particular contexts, the target nucleic acid segment that is included within a template

construct will typically be substantially comprised of a double stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double stranded or otherwise internally complementary.

[00114] Examples of template configurations of the invention that are partially and completely contiguous are schematically illustrated in Figure 20A and 20B, respectively. In particular, as shown in Figure 20A, a partially contiguous template sequence 200 is shown which includes a double stranded portion, comprised of two complementary segments 202 and 204, which, for example, represent a target sequence or portion thereof. As shown, the 3' end of segment 202 is linked to the 5' end of segment 204 by linking oligonucleotide 206, providing a single stranded portion of the template, and yielding a partially contiguous sequence. By comparison, as shown in Figure 20B, a completely contiguous template sequence 210 is shown. Sequence 210 includes a double stranded portion again comprised of two complementary segments 212 and 214. As with the partially contiguous sequence of Figure 20A, the 3' end of segment 212 is joined to the 5' end of segment 214 via oligonucleotide 216 in a first single stranded portion. In addition, the 5' end of segment 212 is joined to the 3' end of segment 214 via linking oligonucleotide 218, providing a second single stranded portion, and yielding a completely contiguous or circular template sequence.

[00115] In addition, the templates of the invention, by virtue of their inclusion of double stranded segments, provide consensus through the identification of both the sense and antisense strand of such sequences (in both the partially and completely contiguous configurations).

[00116] By way of example, and with reference to Figures 20A and 20B, with respect to a partially contiguous template shown in Figure 20A, obtaining the entire sequence, e.g., that of segments 202, 204 and 206 provides a measure of consensus by virtue of having sequenced both sense strand, e.g., segment 202, and the antisense strand, e.g., segment 204. In addition to providing sense and antisense consensus within a single template molecule that can be sequenced in one integrated process, the presence of linking segment 206 also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., 202, is completed and the other begins, e.g., 204. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. The progress of sequencing processes is schematically illustrated in Figure 21A. In particular, as shown, a sequencing process that begins, e.g., is primed, at the open end of the partially contiguous template, proceeds along the first or sense strand, providing the nucleotide sequence (A) of that strand, as represented in the schematic

sequence readout provided. The process then proceeds around the linking oligonucleotide of the template, providing the nucleotide sequence (B) of that segment. The process then continues along the antisense strand to the A sequence, and provides the nucleotide sequence (A'), which provides consensus data for the sense strand as its antisense counterpart. As noted, because the B sequence may be exogenously provided, and thus known, it may also provide a registration sequence indicating a point in the sequence determination at which the data transitions from sense to antisense strands.

[00117] With respect to completely contiguous or circular template sequences configured in accordance with the invention, the consensus potential is further increased. In particular, as with the partially contiguous sequences shown in Figure 20A, the completely contiguous sequences also provide sense and antisense consensus. In addition, such templates provide for the potential for iterative sequencing of the same molecule multiple times, by virtue of the circular configuration of the template. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining consensus for each segment from the complementary sequences, as well as consensus within each segment, by repeatedly sequencing that segment. This is schematically illustrated in Figure 21B, again with a representative illustration of a sequence readout provided. As shown, a sequencing process that is primed at one end, e.g., primed within one linking oligonucleotide sequence, e.g., linking oligonucleotide 218 of Figure 20, proceeds along the first or sense strand 214, again providing the nucleotide sequence A of that strand. The sequence process then proceeds around the first linking oligonucleotide, e.g., linking oligonucleotide 216 from Figure 20, to provide the nucleotide sequence B of that segment of the template. Proceeding along the antisense strand, e.g., segment 212 of Figure 20B), provides the nucleotide sequence A', which is again, complementary to sequence A. The sequencing process then continues around the template providing the nucleotide sequence for the other linking oligonucleotide, e.g., linking oligonucleotide 218 of Figure 20B, where the illustrated sequencing process began, providing nucleotide sequence C. Because the template is circular, this process can continue to provided multiple repeated sequence reads from the one template, e.g., shown as providing a second round of the sequence data A-B-A'-C-A-B-A'. Thus, sequence redundancy comes from both the determination of complementary sequences A and A', and the repeated sequencing of each segment. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing.

Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post synthesis.

[00118] Another exemplary embodiment of an analysis of a plurality of illuminated reactions using intermittent illumination comprises a first illuminated period that is initiated at different times over the plurality of reactions. For example, the illuminated period for a first reaction may start at 0 seconds, the illuminated period for a second reaction may start at 5 seconds, the illuminated period for a third reaction may start at 10 seconds, and so forth. Additionally or alternatively, a first subset of reactions may begin at a first time, a second subset may begin at a second time, etc. The first illuminated period continues for a given length of time, followed by a non-illuminated period and a subsequent second illuminated period. Optionally, a plurality of non-illuminated periods and illuminated periods follow the first illuminated period. Staggered start times can provide staggered data sets (e.g., two or more sequence reads) for the plurality of reactions, allowing multiple different stages of the overall reaction to be interrogated in different reactions. Preferably, the staggered data sets overlap to an extent that allows further analysis and validation of the reaction data. For example, a sequencing-by-incorporation reaction subjected to such an embodiment of the invention would preferably have sufficient overlap between sequence reads from different individual reactions to allow construction of a sequence scaffold and/or consensus sequence for a template nucleic acid.

[00119] A mask for use with a solid support (e.g., an array of confinements) can be designed to allow illumination of one or more portions of the solid support while blocking illumination to other portions of the solid support. For example, a mask may comprise one or more windows that allow excitation illumination to pass through the mask. Such a mask may be physically moved over the surface of the solid support (or the solid support can be moved relative to the mask), e.g., to selectively allow excitation illumination to reach a subset of confinements in an array. For example, a mask that allows 10% of reaction sites to be illuminated could be used to increase the sequencing scaffold coverage by sliding the illumination area (the area being subjected to excitation illumination) back and forth across the solid support. The 10% of reactions would cover certain regions of the nucleic acid template for any given time period (and therefore region of sequence in the template). In certain embodiments, an automated mask that selectively controls the timing of illumination of reactions on a solid support during the course of the reaction/acquisition may be used rather than a mask that must be physically moved.

[00120] The timing of the illuminated and non-illuminated periods for a set of reactions on a solid support may be the same or may vary, and may be synchronized or random. In certain

embodiments in which the excitation illumination source is turned on and off, the timing of the illuminated and non-illuminated periods for the set of reactions will be identical. In other embodiments, for example, those that comprise use of a mask, the timing of the illuminated and non-illuminated periods for the set of reactions can vary so that while a subset of the reactions are illuminated, another subset of the reactions are not illuminated. Various exemplary and nonlimiting embodiments of masks that may be used with a set of reactions on a solid substrate are provided in Figures 3-5, as described below. In certain embodiments, the illuminated/non-illuminated status of each reaction may be random across the solid support, e.g., to remove any experimental bias potentially introduced by actively selecting which reactions to illuminate at a given time, as long as the sequence reads being generated at the illuminated reactions and the time at which these reactions are not illuminated are able to be assigned to a particular reaction. For ease of discussion, the action of both illuminating and collecting emission signals from a reaction of interest, or a particular region on a solid support in which a reaction of interest is taking place, is referred to as "interrogating" that reaction and/or that region. A region being so interrogated is termed an "observation region."

[00121] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination and a mask. As in Figure 2, an array of reactions on a solid support 310 is provided containing sixteen reaction sites, numbered for convenience (A). In B, a mask 320 is provided with a single window 330 to allow passage of illumination to a subset of reactions on the solid support. Window 330 is wide enough to allow illumination of at least two columns of reaction sites on solid support 310. As in Figure 2, a representation of the illumination data is shown in C, with bars extending across the graph indicative of illumination data collected for each reaction. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. When the sequencing reaction is initiated at all positions on solid support 310, the window 330 is positioned to allow illumination to only reactions 1, 5, 9, and 13, and these four reactions provide sequence reads 350 for the earliest stage of the reactions. The window 330 is subsequently moved to provide an illuminated period for reactions 2, 6, 10, and 14 while still continuing the illuminated period for reactions 1, 5, 9, and 13. The illumination data for reactions 2, 6, 10, and 14 provides sequence reads 360, which partially overlap sequence reads 350 for reactions 1, 5, 9, and 13. The window 330 is moved again to provide illuminated periods for reactions 3, 7, 11, and 15 while still continuing the illuminated period for reactions 2, 6, 10, and 14, but removing

illumination from reactions 1, 5, 9, and 13. The illumination data for 3, 7, 11, and 15 results in sequence reads 370, which partially overlap sequence reads 360 for reactions 2, 6, 10, and 14. A fourth position of the mask 320 initiates an illuminated period for reactions 4, 8, 12, and 16 while continuing illumination of reactions 3, 7, 11, and 15, but ending the illuminated period for reactions 2, 6, 10, and 14. Sequence reads 380 correspond to sequence reads from reactions 4, 8, 12, and 16. Finally, the window is moved to end the illuminated period for reactions 3, 7, 11, and 15 while continuing the illuminated period for reactions 4, 8, 12, and 16. Repeating the above process allows a second read to be generated from each reaction, and this second read is noncontiguous with the first read. For example, reactions 1, 5, 9, and 13 correspond to reads 350 and, later in the reaction, reads 355. The two reads generated in a single reaction do not overlap and are separated by a length of nucleotides that was incorporated during the non-illuminated period between the two illuminated periods.

[00122] The mask can optionally be passed over the substrate additional times to generate additional reads until the reactions are complete or no longer provide reliable data, such as when the total illumination time (computed by summing the times for the multiple illuminated periods) has surpassed a photo-induced damage threshold period. Further, the mask may be passed back and forth, or may pass over the solid support in only one direction, e.g., always left to right, or vice versa.

[00123] Further, unlike the data shown in Figure 2B which has gaps in the sequence coverage for the template nucleic acid, the strategy provided in this embodiment results in at least two-fold coverage across the entire template nucleic acid (Figure 3C), although at a lower-fold redundancy. The portion of the template covered by only reads 380 and reads 355 has the least-fold redundancy, and in some instances a gap in coverage may be present in this region due to the movement of the mask 320 from the far right to the far left of the solid support 310. Of course, oversampling by adding replicate reactions to the set of reactions, or using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the coverage of a nucleic acid template, thereby providing more data for construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which have been previously incorporated by reference herein. The natural variation of polymerase translocation rate over the set

of reactions is also apparent in this prophetic example as the spreading of the sequence reads and decreasing overlap between reads from reactions in adjacent columns in the later stages of the reactions as compared to the earlier stages.

[00124] Figure 4A provides an embodiment of a mask similar to that provided in Figure 3 except that it comprises three windows allowing multiple nonadjacent columns of reaction sites to be illuminated simultaneously. Figure 4B provides an embodiment of a mask comprising twelve windows, each of which allows illumination of a single reaction site on a solid support. The windows are oriented in the mask to allow illumination of every other reaction in each row and every other reaction in each column. It will be understood that these mask designs are merely exemplary and nonlimiting embodiments as it is well within the abilities of the ordinary practitioner to determine an appropriate mask design depending on the experimental design or the illuminated reactions to be interrogated.

[00125] Figure 5B illustrates yet another aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Four different samples are disposed on a solid support, one in each quadrant 510, 520, 530, and 540 (A). A mask 550 is used that comprises two windows 560 that allow multiple rows of reaction sites to be illuminated simultaneously (B). A first position of this mask over a solid support in which two reactions in each quadrant are illuminated is demonstrated in C. A second position of the mask allowing illumination of the previously non-illuminated reactions is demonstrated in D. The mask may be moved back and forth as indicated by the double-arrow to provide multiple illuminated and non-illuminated periods for each reaction containing one of the four samples.

[00126] The present invention is also useful for redundant interrogation of reactions or portions of a solid support of interest. In certain aspects, sequential interrogation of different observation regions may be repeated a number of times, e.g., more than 2, 5, 10, 50, 100, 500, 1000, or even more than 10,000 times. In general, this method of stepping the observation region to another, preferably adjacent region, and repeating the interrogation process is generally referred to as a "step and repeat" process, and may be performed by various methods, including but not limited to moving the incident light and the solid support relative to one another and moving a mask across the surface of the solid support, as described above. Although described as a "step and repeat" method, in some embodiments where the observation region is moved across a substrate, that movement is not step-wise and iterative, but instead constitutes a continuous motion, substantially continuous motion, or stepped movement, or an iterative motion whereby each iterative step

interrogates a new region that overlaps with some portion of the previously interrogated region. In particular, a substrate may be moved continuously relative to an optical system, whereby the observation region moves continuously across the substrate being interrogated (in a "scan mode").

[00127] The present invention is optionally combined with an optical system that provides illumination and/or collection of emitted illumination. Preferably, the optical system is operatively coupled to the reaction sites, e.g., on a solid support. One example of a particularly preferred optical system is described in U.S.SN. 11/201,768, filed August 11, 2005, and incorporated herein by reference in its entirety for all purposes. Optical systems are described further below.

[00128] In some embodiments, one or both of the solid support and optical system are moved during interrogation. For example, a solid support being interrogated may be held stationary while the optical system is moved, or the solid support may be moved relative to a stationary optical system. Such movement may be accomplished using any of a variety of manipulation hardware or robotic set-ups, e.g., a stepper/feeder apparatus, and are well known in high performance printing technologies and in the semiconductor industry. For example, robotic systems may be used to pick up and re-orient a given solid support in order to interrogate different regions of the solid support, or make a previously unaccessible region (e.g., blocked by clips, support structure, or the like) of the solid support accessible. Such robotic systems are generally available from, e.g., Beckman, Inc., Tecan, Inc., Caliper Life Sciences, and the like.

[00129] In addition to the foregoing, it will be appreciated that the reagents in a given reaction of interest, including those reagents for which photo-induced damage is being mitigated in accordance with the invention, may be provided in any of a variety of different configurations. For example, they may be provided free in solution, or complexed with other materials, e.g., other reagents and/or solid supports. Likewise, such reagents may be provided coupled to beads, particles, nanocrystals or other nanoparticles, or they may be tethered to larger solid supports, such as matrices or planar surfaces. These reagents may be further coupled or complexed together with other reagents, or as separate reagent populations or even as individual molecules, e.g., that are detectably resolvable from other molecules within the reaction space. As noted above, whether a particular reagent is confined by virtue of structural barriers to its free movement or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined." Further examples of such confined reagents include surface immobilized or localized reagents, e.g., surface immobilized or associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal

reflection microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. For example, in some preferred embodiments, one or more reagents in an assay system are confined within an optical confinement. Such an optical confinement may be an internal reflection confinement (IRC) or an external reflection confinement (ERC), a zero-mode waveguide, or an alternative optical structure, such as one comprising porous film with reflective index media or a confinement using index matching solids. More detailed descriptions of various types of optical confinements are provided, e.g., in International Application Publication No. WO/2006/083751, incorporated herein by reference in its entirety for all purposes.

[00130] The invention is generally applicable to any of a variety of optical assays that require substantial illumination and/or photoactivated conversion or excitation of chemical groups, e.g., fluorophores. For example, the compositions and methods provided herein may be used with fluorescence microscopy, optical traps and tweezers, spectrophotometry, fluorescence correlation spectroscopy, confocal microscopy, near-field optical methods, fluorescence resonance energy transfer (FRET), structured illumination microscopy, total internal reflection fluorescence microscopy (TIRF), etc. The methods provided herein may be particularly useful in assays that are negatively impacted, directly or indirectly, by prolonged exposure to illumination. Of particular interest are those assays that are impaired by the generation and/or accumulation of triplet-state forms or free radicals during illumination.

[00131] One particularly apt example of analyses that benefit from the invention are single-molecule biological analyses, including, inter alia, single molecule nucleic acid sequencing analyses, single molecule enzyme analyses, hybridization assays (e.g., antibody assays), nucleic acid hybridization assays, and the like, where the reagents of primary import are subjected to prolonged illumination with relatively concentrated light sources (e.g., lasers and other concentrated light sources, such as mercury, xenon, halogen, or other lamps) in an environment where photoconversion/excitation is occurring with its associated generation of products. In certain embodiments, the methods, compositions, and systems are used in nucleic acid sequencing processes that rely on detection of fluorescent or fluorogenic reagents. Examples of such sequencing technologies include, for example, SMRT™ nucleic acid sequencing (described in, e.g., U.S. Patent Nos. 6,399,335, 6,056,661, 7,052,847, 7,033,764, 7,056,676, 7,361,466, 7,416,844, the full disclosures of which are incorporated herein by reference in their entirety for all purposes), non-

real-time, or "one base at a time" sequencing methods available from, e.g., Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), Clonal Single Molecule Array™, and SOLiD™ sequencing. (See, e.g., Harris, et al. (2008) Science 320 (5872):106-9, incorporated by reference herein in its entirety for all purposes.) Such prolonged illumination can negatively impact (e.g., by introducing photo-induced damage) these reagents and diminish their effectiveness in the desired reaction.

III. Prevention of Photo-induced Damage

[00132] The methods provided herein are particularly useful in analyses that utilize very limited concentrations of reactants, such as single molecule detection/monitoring assays. As will be appreciated, in such reagent limited analyses, any loss, degradation, or depletion of a critical reagent will dramatically impact the analysis by further limiting the reagent, which not only can adversely effect the detectable signal, but may also directly impact the reaction being monitored, e.g., by changing its rate, duration, or product(s). For example, photo-induced damage can include a photoinduced change in a given reagent that reduces the reactivity of that reagent in the reaction, e.g., photobleaching of a fluorescent molecule, which diminishes or removes its ability to act as a signaling molecule. Also included in the term photo-induced damage are other changes that reduce a reactant's usefulness in a reaction, e.g., by making the reagent less specific in its activity in the reaction. Likewise, photo-induced damage includes undesired changes in a reagent that are caused by interaction of that reagent with a product of another photoinduced reaction, e.g., the generation of singlet oxygen during a fluorescence excitation event, which singlet oxygen may damage organic or other reagents, e.g., proteins. Photo-induced damage also includes downstream effects of damage to reactants, such as irreversible interactions between damaged reactants and other critical components of the reaction, e.g., reactive proteins or enzymes. For example, damage to an enzyme that catalyzes a reaction being monitored may cause a reduction in the rate of the reaction, in some cases stopping it altogether, or may reduce the duration or fidelity of the reaction.

[00133] As suggested by the foregoing, photo-induced damage generally refers to an alteration in a given reagent, reactant, or the like, that causes such reagent to have altered functionality in a desired reaction, e.g., reduced activity, reduced specificity, or a reduced ability to be acted upon, converted, or modified, by another molecule, that results from, either directly or indirectly, a photo-induced reaction, e.g., a photo-induced reaction creates a reactant that interacts with and causes damage to one or more other reactants. Typically, such photoreaction directly

impacts either the reactant of interest, e.g., direct photo-induced damage, or impacts a reactant within one, two or three reactive steps of such reactant of interest. Further, such photoreaction can directly impact the reaction of interest, e.g., causing a change in rate, duration, processivity, or fidelity of the reaction.

[00134] The amount of time an illuminated analysis may be carried out before photo-induced damage so substantially impacts the reactants to render the analysis non-useful is referred to as the "photo-induced damage threshold period." A photo-induced damage threshold period is assay-dependent, and is affected by various factors, including but not limited to characteristics of enzymes in the assay (e.g., susceptibility to photo-induced damage and the effect of such damage on enzyme activity/processivity), characteristics of the radiation source (e.g., wavelength, intensity), characteristics of the signal-generating molecule (e.g., type of emission, susceptibility to photo-induced damage, propensity to enter triplet state, and the effect of such damage on the brightness/duration of the signal), similar characteristics of other components of the assay. It can also depend on various components of the assay system, e.g., signal transmission and detection, data collection and analysis procedures, etc. It is well within the abilities of the ordinary practitioner to determine an acceptable photo-induced damage threshold period for a given assay, e.g., by monitoring the signal decay for the assay in the presence of a photodamaging agent and identifying a period for which the signal is a reliable measure for the assay. In terms of the invention, the photo-induced damage threshold period is that period of illuminated analysis during which such photo-induced damage occurs so as to reduce the rate or processivity of the subject reaction by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over the same reaction in the absence of such illumination. It is an object of the invention to increase the photo-induced damage threshold period, thereby increasing the amount of time reactions can proceed toward completion with minimal damage to the reactants, thereby lengthening the time in which the detectable signal is an accurate measure of reaction progression.

[00135] In some contexts, a "photo-induced damaged" reaction may be subject to spurious activity, and thus be more active than desired. In such cases, it will be appreciated that the photo-induced damage threshold period of interest would be characterized by that period of illuminated analysis during which such spurious activity, e.g., as measured by an increase in reaction rate, or an increase in non-specific reaction rate, is no more than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over a non-illuminated reaction. In one non-limiting example, where a nucleic acid polymerase, by virtue of a photodamaging event, begins to incorrectly incorporate nucleotides

during template directed synthesis, such activity would impact the photo-induced damage threshold period as set forth above. In this case, the methods, devices, and systems of the invention would increase the photo-induced damage threshold period, thus increasing the amount of time the reaction could proceed before the above-described spurious activity occurred.

[00136] With reference to nucleic acid analyses, it has been observed that in template-directed synthesis of nucleic acids using fluorescent nucleotide analogs as a substrate, prolonged illumination can result in a substantial degradation in the ability of the polymerase to synthesize the nascent strand of DNA, as described previously, e.g., in U.S. Published Patent Application No. 20070161017, incorporated by reference herein in its entirety for all purposes. Damage to polymerase enzymes, template sequences, and/or primer sequences can significantly hinder the ability of the polymerase to process longer strands of nucleic acids. For example, reduction in the processivity of a polymerase leads to a reduction in read lengths for sequencing processes that identify sequence constituents based upon their incorporation into the nascent strand. As is appreciated in the art of genetic analysis, the length of contiguous reads of sequence directly impacts the ability to assemble genomic information from segments of genomic DNA. Such a reduction in the activity of an enzyme can have significant effects on many different kinds of reactions in addition to sequencing reactions, such as ligations, cleavages, digestions, phosphorylations, etc.

[00137] Without being bound to a particular theory or mechanism of operation, it is believed that at least one cause of photo-induced damage to enzyme activity, particularly in the presence of fluorescent reagents, results from the direct interaction of the enzyme with photo-induced damaged fluorescent reagents. Further, it is believed that this photo-induced damage of the fluorescent reagents (and possibly additional damage to the enzyme) is at least partially mediated by reactive intermediates (e.g., reactive oxygen species) that are generated during the relaxation of triplet-state fluorophores. One or both of the photo-induced damaged fluorescent reagents and/or reactive intermediates may be included in the overall detrimental effects of photo-induced damage.

[00138] In certain aspects, the invention is directed to methods, devices, and systems that reduce the amount of photo-induced damage to one or more reactants during an illuminated reaction, e.g., thereby improving the reaction, e.g., by increasing the processivity, rate, fidelity, processivity, or duration of the reaction. In particular, methods are provided that yield a reduction in the level of photo-induced damage and/or an increase in the photo-induced damage threshold period as compared to such reactions in the absence of such methods, devices, and systems. In

particular embodiments, such methods comprise subjecting an illuminated reaction to periods of non-illuminated during the course of the reaction, as described above, or by temporarily removing components of the reaction mixture that are believed to cause such damage, as described below.

[00139] As generally referred to herein, limited quantity reagents or reactants may be present in solution, but at very limited concentrations, e.g., less than 200 nM, in some cases less than 10 nM and in still other cases, less than 10 pM. In preferred aspects, however, such limited quantity reagents or reactants refer to reactants that are immobilized or otherwise confined within a given area or reaction site (e.g., a zero-mode waveguide), so as to provide limited quantity of reagents in that given area, and in certain cases, provide small numbers of molecules of such reagents within that given area, e.g., from 1 to 1000 individual molecules, preferably between 1 and 10 molecules. As will be appreciated, photo-induced damage of immobilized reactants in a given area will have a substantial impact on the reactivity of that area, as other, non-damaged reactants are not free to diffuse into and mask the effects of such damage. Examples of immobilized reactants include surface-immobilized or -localized reagents, e.g., surface-immobilized or -associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflectance microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. Various types of solid supports upon which one or more reactants can be immobilized are described above.

[00140] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is interrogated for one or more illuminated periods that cumulatively are less than a photo-induced damage threshold period, as set forth elsewhere herein. Such interrogation may occur coincident with or independent of interrogation of additional observation regions on a solid support containing the first observation region. In accordance with the present invention, the observation region typically includes confined reagents (e.g., enzymes, substrates, etc.) that are susceptible to photo-induced damage, and may include an area of a planar or other solid support upon which confined reagents are immobilized. Alternatively or additionally, the observation region may include a physical confinement that constrains the reagents that are susceptible to photo-induced damage, including, e.g., microwells, nanowells, planar surfaces that include hydrophobic barriers to confine reagents.

[00141] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is intermittently interrogated under constant illumination by virtue of intermittent presence of detectable components of the reaction, wherein the presence of such detectable components has the potential to directly or indirectly cause photo-induced damage to one or more other reaction components. For example, a buffer comprising detectable components of a reaction can be temporarily replaced with a buffer comprising non-detectable versions of the same components of the reaction, thereby interrupting data acquisition for the reaction. When data acquisition is to be recommenced, the buffer comprising detectable component is substituted for the buffer comprising non-detectable components. This substitution of reaction components may be repeated multiple times to generate multiple sets of data collected at noncontiguous stages of the reaction. For example, such a substitution can occur at least about 2, 4, 6, 8, or 10 times during the course of the reaction.

[00142] In certain preferred embodiments, the detectable components are fluorescently-labeled components that can be damaged by exposure to excitation illumination, and can further cause damage to other reaction components, as described above. For example, a sequencing-by-incorporation reaction can be initiated in the presence of fluorescently-labeled nucleotides whose incorporation is indicative of the nucleotide sequence of the nascent strand synthesized by a polymerase, and by complementarity, of the template nucleic acid molecule. At a selected time point during the ongoing reaction, the labeled nucleotides can be removed and replaced with unlabeled nucleotides, for example, by buffer exchange. After a period of time during which data acquisition has been interrupted by the absence of signal from the ongoing reaction, the labeled nucleotides can be reintroduced to reinitiate data acquisition. The labeled nucleotides may be removed and reintroduced multiple times and for various lengths of time, as preferred by the ordinary practitioner. In this way, multiple noncontiguous sequence reads can be generated from a single nucleic acid molecule in real time.

[00143] The methods herein slow the accumulation of photo-induced damage to one or more reagents, and may therefore indirectly mitigate the impact of photo-induced damage in an ongoing reaction of interest. By way of example, methods that reduce exposure of a critical enzyme component to illumination radiation (e.g., by subjecting the reaction to periods of non-illumination or by temporarily removing a component of the reaction responsible for such damage) do not necessarily prevent the photo-induced damage to the enzyme component, but rather extend the photo-induced damage threshold period by slowing the accumulation of photo-induced damage in

the reaction mixture. Measurements of reduction of photo-induced damage as a result of implementation of intermittent illumination may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction subjected to constant illumination. Likewise, measurements of reduction of photo-induced damage as a result of temporary removal of reaction components responsible for such damage may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction in which such components are present throughout. Further, characterization of a reduction in photo-induced damage generally utilizes a comparison of reaction rates, durations, or fidelities, processivities, e.g., of enzyme activity, and/or a comparison of the photo-induced damage threshold period, between a reaction mixture subjected to such the methods and/or systems of the invention and a reaction mixture not so subjected.

[00144] In the case of the present invention, implementation of the methods, devices, and systems of the invention generally results in a reduction of photo-induced damage of one or more reactants in a given reaction, as measured in terms of "prevented loss of reactivity" in the system. Using methods known in the art, the amount of prevented loss of activity can at least 10%, preferably greater than 20%, 30%, or 40%, and more preferably at least 50% reduction in loss of reactivity or increase in processivity, and in many cases greater than a 90% and up to and greater than 99% reduction in loss of reactivity or increase in processivity. By way of illustration, and purely for the purpose of example, when referring to reduction in photo-induced damage as a measure of enzyme activity in the presence and absence of intermittent illumination, if a reaction included a reaction mixture having 100 units of enzyme activity that would, under constant illumination, yield a reaction mixture having only 50 units of activity, then a 10% reduction in photo-induced damage would yield a final reaction mixture of 55 units (e.g., 10% of the 50 units otherwise lost, would no longer be lost). Further, use of the invention is expected to increase the performance (e.g., processivity, duration, fidelity, rate, etc.) of a reaction whose performance is negatively impacted by constant exposure to illumination by at least about 2-, 5-, 10-, 20-, 30-, 50-, 80-, 100-, 500-, or 1000-fold over that achieved by the reaction under constant illumination. For example, it is a specific object of the instant invention to increase the processivity of a polymerase enzyme in a sequencing reaction to allow collection of data across a longer length of the template.

[00145] With regards to sequencing applications, the methods herein facilitate the scaffolding of nucleic acid sequences in reactions susceptible to photo-induced damage. For example, if the sequencing device has 1000 base pair average readlength under constant illumination, one could subject the reaction to illuminated periods timed to allow approximately 100 nucleotides to be

incorporated into the nascent strand of read, followed by non-illuminated periods timed to allow approximately 1000 nucleotides to be incorporated "in the dark." The sequence reads resulting from this experimental design would comprise about ten sequence reads of about 100 nucleotides each separated by gaps of about 1000 nucleotides each. If a plurality of sequencing reactions were carried out in this manner, and the illuminated periods were staggered appropriately, the reads from the plurality of reactions could be combined to provide nucleotide sequence data for the entire template nucleic acid. This would potentially allow sequence scaffolds to be built much more easily than can be done with short-read systems, enabling structural analysis of previously impossible-to-sequence sections of highly repetitive DNA, given the sequencing system is capable of long reads in the absence of photodamage.

IV. Software and Algorithm Implementations

[00146] The methods herein may operate with numerous methods for sequence alignment including those generated by various types of known multiple sequence alignment (MSA) algorithms. For example, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read using a reference sequence. In some embodiments in which a reference sequence is known for the region containing the target sequence, the reference sequence can be used to produce an MSA using a variant of the center-star algorithm. Alternatively, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read relative to every other read without using a reference sequence ("*de novo* assembly routines"), e.g., PHRAP, CAP, ClustalW, T-Coffee, AMOS make-consensus, or other dynamic programming MSAs. Depending on the sequence-generating methods used, the determination of sequence alignment may also involve analysis of read quality (e.g., using TraceTuner™, Phred, etc.), signal intensity, peak data (e.g., height, width, proximity to neighboring peak(s), etc.), information indicative of the orientation of the read (e.g., 5'→3' designations), clear range identifiers indicative of the usable range of calls in the sequence, and the like. Additional algorithms and systems for sequence alignment are well know to those of skill in the art, and are described further, e.g., in G. A. Churchill, M. S. Waterman (1992) "The Accuracy of DNA Sequences: Estimating Sequence Quality," *Genomics* 14: 89-98; M. Stephens, et al. (2006) "Automating sequence-based detection and genotyping of SNPs from diploid samples," *Nat. Genet.*, **38**: 375-381; J. Hein (1989) *Mol. Biol. Evol.*, **6**: 649-668; U.S.S.N. 12/134,186, filed June 5, 2008; and U.S.S.N. 61/116,439, filed November 20, 2008.

[00147] A standard sequence alignment problem in the context of DNA sequencing is to align the sequence of a relatively short fragment (<2 kilobases) to a large target sequence. The assumption is made that this fragment represents a contiguous portion of DNA to be mapped to a single location on the reference sequence. (A "contiguous portion" to be mapped to a single location may contain small insertions and/or deletions and still be considered contiguous in this context.) With the further development of nucleic acid sequencing technologies (e.g., from Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), and Applied Biosystems, Inc. (Foster City, CA)) and mate-pair sequencing protocols (see, e.g., U.S. Patent Pub. No. 2006/0292611 A1, which is incorporated by reference herein in its entirety for all purposes), the alignment problem has been extended to align two fragments coming from the same read to the reference sequence using some knowledge of the expected mate-pair configuration (distance and orientation).

[00148] With regards to mate-paired reads, mapping two fragments with a distance constraint and orientation constraint has been treated by various short-read mapping algorithms, e.g., SOAP (Li, et al. (2008) Bioinformatics, **24**, 713-714); SOAPdenovo; and Maq, a set of programs that map and/or assemble fixed-length Solexa/SOLiD reads (SourceForge, Inc.). While these algorithms can handle simple cases of mate-pair alignment, which generally treat the specific problem of only two reads coming from a mate-paired sequence and use the distance constraint as a hard filter (i.e., if two reads are within *x* bp of each other and in the correct orientations, report them as a mate-pair hit), the methods provided herein are more general and can handle much more complex data sets, including those with multiple reads, those for which a reference sequence is or is not present, potential non-template sequence (e.g., adapter regions or linker portions described below), and complex distance and orientation constraints. Other programs are also available that attempt to generalize on top of the mapping and aligning performed by the programs described above. These include, e.g., Breakdancer, variationhunter, GASV, etc., which can handle more complex mappings, e.g., by clustering.

[00149] Real-time single molecule sequencing presents opportunities for obtaining much more complex sequence fragments from a single DNA sequencing read. Two examples are the reading of multiple discontiguous sequence fragments from a single long stretch of DNA using a pulsed or intermittent detection system (e.g., intermittent illumination) as described herein and the contiguous reading of forward, reverse and adapter fragments from a circular templates (SMRTbell™ templates; see e.g., U.S.S.N. 61/099,696, filed September 24, 2008; U.S. Patent Application No. 12/383,855, filed March 27, 2009 and U.S. Patent Application No. 12/413,258,

filed March 27, 2009, all of which are incorporated by reference herein in their entireties for all purposes). Further, methods for sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.

[00150] Certain aspects of the invention provide methods for optimally aligning such sequences to a reference sequence using knowledge of the molecular configuration and/or sequencing protocol used to generate the related sequence reads. In particular, methods are provided to address the general problem of mapping multiple fragments to a reference sequence with variable distance and orientation constraints.

[00151] Beginning with raw sequence data generated by a nucleic acid sequencing instrument (step 1), the sequence data is mapped to a target sequence (step 2) using a local alignment method which produces sub-optimal local alignments as well as the optimal alignment, for example, the Smith-Waterman algorithm. Another, more flexible example of a local alignment method is a chaining method using a method for aligning very short fragments to the target sequence (e.g., kmer-indexing, suffix trees, suffix arrays, etc.) and chaining the resulting hits back into longer chains of significant matches (see, e.g. D. Gusfield, Algorithms on Strings, Trees, and Sequences, Cambridge University Press: Cambridge, UK, 1997, which is incorporated by reference herein in its entirety for all purposes). The chains do not necessarily need to be refined by dynamic programming in order to be useful for the following algorithm, permitting a very fast algorithm. In certain embodiments, dynamic-programming refinement of the chain might improve the power (area under the ROC curve) of the algorithm.

[00152] The target sequence consists of the potential hypotheses for the molecular template in question. In the example of nucleic acid sequencing methods using iterative illumination for sequencing a shotgun fragment from a linear DNA sequence, the potential hypotheses are both orientations of the genome (since we do not know the original orientation of the fragment). In the example of sequencing of a SMRTbell™ template (e.g., see Example 1 herein), the hypotheses include both orientations of the genome and known adapter sequences. The parameters determining how many hits are reported for each local fragment can be varied to change the specificity and sensitivity of this algorithm. Figure 6 shows what these hits might look like for a SMRTbell™

template (represented as paths in the sequence alignment matrix, which is often called the dynamic-programming matrix, although it isn't necessary to use dynamic programming to find these paths).

[00153] After the potential local alignments have been enumerated, a weighted directed graph is constructed with each local alignment represented as a node in the graph (step 3). The edges are drawn between nodes if they represent a potential reconstruction of the original molecular template using knowledge of the expected molecular configuration. The directed connection of an alignment path A to an alignment path B is interpreted as "The target sequence represented by B could follow the target sequence represented by A in the original molecule." For example, if a linear single-stranded DNA molecule is being sequenced by a method that uses iterative illumination, then fragments from opposite orientations would not be expected to be connected (unless the linear single-stranded DNA molecule also included oppositely oriented sequences, e.g., as in the case of a linearized SMRTbell™ template.) In general, fragments that represent the same stretch of the sequencing read but that align to different regions on the target sequence would not be connected. Aside from these examples, the rules for connecting nodes should be fairly loose to permit exploration of weak possibilities that gain significance when all the evidence (e.g. all the sequence reads) are considered. The assignment of edge weights handles the proper weighting of the likelihood of these edges, and the speed of the algorithm can be tuned by optimizing the pruning of highly unlikely edges. As usual this represents a tradeoff between speed and sensitivity.

[00154] Weights are assigned to connections ($A \rightarrow B$) in the graph representing the log-likelihood that target fragment A is followed by target fragment B in the original molecule.

$$w(A \rightarrow B) = -\log P(B|A)$$

The conditional probability $P(B\ A)$ encodes the knowledge of the possible molecular configurations and the alignment significance of B.

$$P(B|A) = f(B)g(A,B)$$

where f is a measure of alignment significance (either theoretical or empirically obtained) and g encodes the physical constraints representing the allowed molecular configurations.

[00155] For example, in the context of sequencing using iterative illumination the following may be known: the time between the end of one fragment and the beginning of the next fragment is 200 seconds. If the polymerase incorporates bases with an average rate of 4bp/sec with a standard deviation of 1bp/sec, it can be hypothesized that the probability of target fragment 2 following target fragment 1 is determined by the distance between these fragments on the target and a normal probability:

$$g(A,B)=\frac{1}{\sqrt{2\pi(200)}}\exp\left[-(d-800)/2(200)^2\right]$$

[00156] In a SMRTbell™ template example, knowledge of the expected insert size and the observed distance and orientation between fragments would be used to weight the likelihood that these two fragments could come from a correctly generated SMRTbell™ template. This weight could include the expected rate of the polymerase as well and rules for the orientation of fragments with respect to each other and their distance apart in the original read. For example, while it may be expected that two forward fragments mapping to the same region in the target genome potentially come from multiple passes around a SMRTbell™ template molecule, those fragments would not be expected to be immediately adjacent in sequencing time. The weighting function would account for the proper amount of expected time between such fragments (i.e. the elapsed time would be expected to be long enough to include two adapter sequences and a reverse sequence).

[00157] In general, the weighting function could be arbitrarily complex and tuned to empirically observed relationships between sequencing fragments given the available knowledge (distance between fragments on the target sequence, sequencing time between fragments, expected length of the template, etc.). For example, the empirical probability distributions might be observed to exhibit longer tails than a Gaussian probability model might predict. The use of a conditional log-likelihood for the assignment of edge weights is motivated by the following logic. In a graph of possible local alignments it is desirable to find a highly likely path that best explains the observed data. Consider a path through three nodes *A*, *B*, and *C*, with *P(ABC)* being the probability that *ABC* is the correct assignment:

$$\begin{aligned}P(ABC)&=P(C|AB)P(B|A)P(A)\\&\approx P(C|B)P(B|A)P(A)\end{aligned}$$

where the last approximation is justified by the observation that the constraints between allowable assignments to the target sequence are typically local in nature. Generalizing this formula for a path $a_1,\ldots,a_N$ and taking the negative logarithm of both sides gives

$$-\log P(a_1\ldots a_N)=-\sum_{i=1}^{N-1}\log P(a_{i+1}|a_i)-\log P(a_1)$$

[00158] It is apparent that the edge weights are additive if we use log-likelihood and we can use standard shortest-path algorithms for directed graphs to find the optimal path. A hypothetical directed graph is illustrated in Figure 7. This graph corresponds to the situation depicted by the alignments pictured above. Heavier lines correspond to more likely paths with the optimal path

shown in blue. Dashed lines represent forbidden transitions. Not all paths are considered in the illustration to avoid clutter in the presentation. The general formula listed above includes a "one-body" term $P(a_1)$ for the starting node in each path that weights the probability that this initial alignment is correct. To accommodate this probability in a path-finding algorithm we add a pseudo-source s to the graph which connects to every possible node (not shown in the graph above). The edge weight connecting the pseudo-source with a node a_i is $-\log P(a_i)$. This allows the use of a conventional single-source shortest-path algorithm starting from the pseudo-source. The desired probability $P(a_i)$ can come from a measure of alignment significance (theoretical or empirically determined) or could be set uniformly across all alignments to allow the path logic to determine the best path assignment, independent of the relative value of the starting points. It is anticipated that a threshold will be required here to only allow edges between the pseudo-source and nodes for highly likely alignments; otherwise the shortest path algorithm in the next step will not give the desired path.

[00159] After construction of the weighted directed graph, the shortest path to each node is determined (step 4). The graph is directed and acyclic (DAG) so we can use the standard shortest-path DAG algorithm (see T.H. Cormen, CE Leiserson, RL Rivest, Introduction to Algorithms, MIT Press: Cambridge, Massachusetts, 1990). This algorithm scales as O(V+E) and should be very quick for these graphs. After the shortest path to each node is determined, the paths need to be ranked to declare the best assignment. It is suggested that the best metric would be a measure which rewards paths that explain more of the sequenced read (longer paths) with high likelihood. One such metric would be the normalized negative log-likelihood: dividing the total weight of the path by the number of bases in the sequenced read explained by this path. For more complicated graphs or edge-weight assignments, Dijkstra's algorithm, the Bellman-Ford algorithm, or the A* algorithm could be applied. Other algorithms that may also be used include, but are not limited to the Floyd-Warshall algorithm.

[00160] For noisy sequence data it is likely that the local alignments found in step 2 will occasionally overlap with each other in the sequenced read even though it is physically impossible for such overlaps to occur in a perfect system (unless there has been a rearrangement relative to the reference genome). As such, some amount of slack must be allowed in the edge assignment logic in step 3 to account for not knowing the precise boundaries of each local alignment. Once the best physical model explaining the observed read is determined, the boundaries of the local alignments can be refined to reflect the physical necessity that each base in the sequenced read can only be

represented in one local alignment. It is also desirable to explain all of the bases in between the local alignments that haven't been assigned in the graph. One straightforward approach to refinement would be to construct the perfect model of the sequence and to realign the sequenced read to this sequence. This refinement algorithm would preserve physical constraints (each base in the sequenced read can only be explained by one location in the template) and would assign all bases between the extremal nodes in the optimal path.

[00161] Certain aspects of the software and algorithm implementations described herein may be varied or altered without departing from the spirit and scope of the invention. For example, with regards to algorithm seeding, many algorithms can be applied for the original determination of sub-optimal local alignments (step 2). Conventional examples include FASTA, BLAST, or Smith-Waterman. It is expected that the best benefit will be obtained from using short-sequence alignment algorithms (suffix array, suffix tree, Boyer-Moore, Rabin-Karp, kmer-indexing, and the like) followed by chaining to establish regions of significant matches. An advantage of the algorithm described here is that it does not require dynamic-programming refinement of the resulting chains and therefore can be quite fast, however it is expected that using dynamic programming to refine the chains in step 2 could increase the power of the algorithm.

[00162] With regards to graph construction, there will be advantages to tuning the logic of edge assignments to keep the size of the graph manageable. It is possible that steps 2 and 3 might be combined to in a greedy fashion to focus the potentially slow step 2 into productive areas of the graph. For example, if a particularly strong hit is found early in step 2, then it may be beneficial to search for sub-optimal hits only in this local vicinity, knowing that this strong hit should be in the final solution. Tuning of the graph construction might include thresholds, below which edges are not created. Further, there are multiple parameters (minimum chain length, minimum probability for edge assignment, relative weighting of length vs. accuracy, etc.) which can be exposed and tuned in this algorithm to maximize the sensitivity and specificity of the algorithm for a given scenario.

[00163] With regards to determination of the distance a polymerase travels between reads, various strategies are provided that are more sophisticated than estimation based upon the rate of incorporation and the time between detection (e.g., illuminated) periods. In certain embodiments, the distribution of the base pair distance travelled by a polymerase during a non-detection period is called $p(x)$. The distribution of enzyme velocities, $p(v)$, is estimated by aligning observed reads to a reference sequence, and this distribution is represented as the number of reference bases per unit time. There is a length of time, τ, over which measurement of the instantaneous rate is not

independent. While this method of determining the distance the polymerase travels during a non-detection period should not be overly sensitive to non-independent estimation of the polymerase rate, it is likely to strive for independent measurements of the rate. The distance τ can be estimated from an exponential fit to the auto-correlation function $<\delta v(t)\delta v(t-\Delta t)>$, and $v(t)$ tabulated across the aligned sequence at increments of τ.

[00164] Where multiple single polymerase enzymes are being observed simultaneously, e.g., each being optically resolvable from every other on a single array, the $p(v)$ for each is preferably determined independently for each enzyme. Further, information regarding rare but extended events, such as polymerase "stalling" on the template, can be measured across a larger data set. For example, the statistics of stalls can be determined by aggregating rate measurements across an entire array. Where a stall distribution is characterized by a "long tail" corresponding to multi-exponential behavior of IPD distribution, such a distribution of polymerization rates can be extended for stalls longer than the observed reaction by fitting the long-tail behavior to an appropriate functional form, e.g., using a single-exponential parametric model or other physically motivated model (e.g., multi-exponential, stretched exponential, power-law, etc.) In certain preferred embodiments, the following representation of a "per-enzyme" $p(v)$ is used:

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}$$

where $f(v)$ is an interpolating function designed to retain information about the zero-velocity tail of the global $p(v)$ distribution while taking the estimate of the polymerase velocity dynamics(e.g., the dominant high velocity mode) from the specific enzyme. Such an interpolating function is:

$$f(v) = \frac{1 + erf(v / v_0)}{2}$$

where v_0 is a scale parameter to be chosen based on experience (but optionally fixed). Alternatively, the average of the empirical $p_{enzyme}(v)$ and $p_{array}(v)$ can be used. This approach can be motivated by a Bayesian approach to density estimation. Other kernel density and Bayesian methods can be suggested. Alternatively or in addition, the robustness of $p(v)$ to conditions and daily phenomenon can be explored and used to estimate $p(v)$ more globally, e.g., using one or more weekly control experiments.

[00165] Given the lack of a known reference sequence for *de novo* assembly, several alternative ways to formulate $p(v)$ are provided as follows. For example, in a first embodiment a control template (essentially a proxy reference sequence) can be subjected to sequencing, e.g., in the

same reaction as the *de novo* sample or in an identical reaction. The observed velocity for the sequencing reactions would be measured based upon alignments of the reads from the control template to its known sequence. Typically, a per-enzyme correction would not be available for the *p(v)* and *p(v)* would default to an array-averaged *p(v)*. In a second embodiment, a previously determined *p(v)* from experiments using a known reference sequence can be used, e.g., where the previous experiments were performed under the same conditions as the *de novo* experiments. In a third embodiment, *p(v)* is estimated by using quality information/metrics to screen for the most likely "true" calls, and restricting the estimates of *v* to regions containing those calls. In a fourth embodiment, where error is low, the called base rate and reference base rate converge to the same rate, and measurements of p(v) without knowledge of the reference become substantially reliable. Further, even if they do not fully converge, they can still be used to accurately infer *p(v)*,as long as the called base rate is predictably higher/lower than the reference base rate. Yet further, the measurement of *p(v)* when a reference sequence is not available can benefit from a detailed look at the probability model which is available from an algorithm like a CRF. That is, *p(v)* can be tabulated using a weighted sum over paths through a CRF probability model.

[00166] As will be clear to the ordinary practioner based upon the teachings herein, this framework extends naturally to the measurements of other potentially systematic variations in *p(v)* across an array, e.g., even where a single reaction mixture is applied to the entire array. For example, the local temperature of the reaction environment can vary systematically across an array of reactions. The average and variation in the rates of polymerase enzymes on the array would likely have a dependence on this hidden variable. Where the functional form of the temperature dependence is known, the measurement of *p(v)* can be stabilized across the array by modeling a de novo *p(v)* as $p_{cond}(v) + p_{x,y}(v)$ where *x,y* are geometrical variables defining the location on the array. Further, *p(v)* has been found to be somewhat variable over time. As such, in certain embodiments a model of *p(v;t)* is developed using an appropriate model for the evolution of *p(v)* over time.

[00167] Once a representative distribution of velocities *p(v)* has been obtained for a given read from a given reaction, the expected travel distance in the non-detection period can be expressed as:

$$p\left(\frac{x}{\tau}\right) = IL\left[L[p(v)]^{t/\tau} \frac{1 - L[p(v)]}{s} \right]$$

where L[] and IL[] stand for the Laplace and inverse Laplace transform, respectively. A similar result is derived in Svoboda, et al. (PNAS 91:11782 (1994)) and readily follows from considering

the pdf of a sum of random variables. Optionally, in certain embodiments density estimation techniques (e.g., kernel density estimation, etc.) are useful when modeling *p(v)* since they can smooth the resulting numerical calculations in the Laplace and inverse Laplace transform.

[00168] Knowledge of the complete distribution has several advantages over the commonly applied Gaussian approximation. For example, knowledge of the complete distribution of insert lengths is very desirable when using a Bayesian framework approach to detect structural variation. (See, e.g., Bashir, et al. (2008) PLoS Comput. Biol. 4:51; Hormozdiari, et al. (2009) Genome Res. 19:1270; and Lee, et al. (2008) Bioinformatics 24:59.) While Bashir, et al. does not strictly follow a Bayesian approach, the geometric approach described in the paper can be straightforwardly modified to incorporate an actual posterior instead of the boxcar posterior assumed in the paper. Further, during mapping of noncontiguous reads to a genome where they are expected to be concordant (*i.e.*, not a structural variation), it is useful to consider the known distribution when judging the significance of the resulting alignments between the observed reads and the genomic sequence. In addition, when clustering noncontiguous reads that scaffold contigs in a *de novo* assembly, a path of Bayesian significance can be followed that is very similar to that followed in the structural variation case discussed *supra*.

[00169] In further embodiments, the determination of the distance a polymerase travels between reads is performed using an algorithm based on a simulation approach rather than the exact analytical result used in the algorithm described above. This method relies on Monte Carlo sampling from a distribution, which allows a better extension to arbitrary empirical distributions. It also lacks the difficult computations of numerical Laplace and inverse Laplace transforms, and permits calculation of distances traveled during non-detection periods when the underlying kinetic processes have multi-phasic kinetics, e.g., the presence of long stalls.

[00170] This approach aims to calculate the distribution of the distance *x* travelled by an enzyme during a time *t* during which it was not being observed (e.g., during a non-detection period). In some embodiments, a distribution of local rates, *p(v)*, is estimated, where the definition of "local" is set by the correlation length of the rate autocorrelation function, e.g.:

$$\langle \delta v(t)\delta v(t+\Delta)\rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$$

Given a local rate distribution and an assumption that independent identically distributed (i.i.d.) draws can be made from this distribution, one approach to calculating the distribution is as follows. First, draw $N = t/\tau_{corr}$ velocities from *p(v)*; and subsequently sum them and record them as an

estimate of x/τ_{corr}. Repeat the process M times, with the optimal choice of M dependent on the desired level of precision for estimation of the *p(x)* distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000.

[00171] In some embodiments in which the enzyme system is not well explained by a single kinetic process or cycle (as in the case of observed stalling behavior), above-described rate autocorrelation function and the i.i.d. assumption will be violated. As such, a probability model having a richer structure is preferably used. One such probability model is a Hidden Markov Model (HMM). Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system. Where the kinetics of the pausing state can be well described by a single-exponential, this model is expected to describe the observed distribution of local velocities. The single-exponential assumption is implicit in the state structure of the model since the amount of time spent in the pause state will be a geometric distribution with mean p/(1-p) [*i.e.*, the observed stall times will have to be added to this model]. If the stall kinetics are multi-phasic, then more "dark states" will have to be added to this model. Further, the model shown in Figure 15 can actually be treated as a Markov Model and not a Hidden Markov Model without much loss of generality because the "pause" state is not actually hidden due to the fact that the data collected during the pause state is highly distinguishable from the data collected during the sequencing state. As such, the general HMM apparatus is not necessary. The model in Figure 15 can be used to simulate the distribution of local velocities when there is a long-term pause or stall phase present in the reaction data kinetics. S0 is the start state, and there is no explicit end state since this model is used as a generative model and it is assumed that it is run forward for a prescribed number of steps. The qualities P(P→S) and P(S→P) represent exit from a stalled state and entry into a stalled state, respectively. These qualities can be measured by an EM algorithm or they can be quickly estimated by physical observables. $P_{P\to S} = 1/\left(1+\frac{\tau_{stall}}{\tau_{corr}}\right)$ and P(S→P) is the frequency of stall starts per τ_{corr}. (Example parameters are τ_{stall}=80 seconds; τ_{corr} = 10 seconds; and P(S→P) = 1/24.) The simulation estimate of *p(x)* can now be produced using the procedure outlined above in which $N = t/\tau_{corr}$ velocities are drawn from *p(v)*; and they are subsequently summed and recorded as an estimate of x/τ_{corr}. The process is repeated M times, with the optimal choice of M dependent on the desired level of precision for estimation of the *p(x)* distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000. Figure

16 shows exemplary simulated applications of this method. Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters P(S→P) = 1/24; P(P→S) = 1/11; and $p(v)$~Gamma(48,0.25). Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a 1300 second non-detection period, which is calculated by sampling 2000 estimates from the HMM model.

[00172] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period. The influence of pause/stall behavior can be seen in the heavy-left tailing of both distributions.

[00173] While the simulation method in which i.i.d. draw assumption is valid is more general and can treat arbitrary $p(v)$ and more complex models for non-sequencing states, the two-state model using the HMM can be treated analytically. The result of this is:

$$p(x/\tau_{corr}) = \sum_{N_S=0}^{N} \pi_{N_S}(x) p_N(N_S)$$

where $\pi_{N_S}(x)$ is the distribution of the sum of N_S variables drawn from $p(v)$. For the general case, this distribution is given by the Laplace transform approach presented above. For $p(v)$~Normal(μ,σ), this distribution is distributed as Normal(Nsμ,√Nsσ). For $p(v)$~Gamma(k,θ), this distribution is distributed as Gamma(Nsk,θ). $P_N(N_S)$ is the number of cycles spent in the sequencing state if we observe N cycles from the Markov process in Figure 15. The expression for this is described in Pedler, et al. (1971) J. Appl. Prob. 8:381, which is incorporated herein by reference in its entirety for all purposes.

[00174] As will be clear to one of ordinary skill in the art upon review of the teachings herein, these methods can be readily extended to the non-detection period estimations of procession by other cyclical biological reactions, such as the action of reverse transcriptase or the synthesis of proteins by a ribosome complex, e.g., and certain preferred embodiments of such reactions are further described in U.S.S.N. 12/767,673, filed April 26, 2010; and U.S.S.N. 12/813,968, filed June 11, 2010, the disclosures of which are incorporated herein by reference in their entireties for all purposes. Further, the simulation model described above is not restricted to simple two-state kinetics, and the use of $p(v)$ is not restricted to analytical models. In fact, in certain embodiments, empirical estimates are preferably used.

[00175] Although useful in certain preferred embodiments of the invention, certain algorithms as presented above do not easily handle the case where the template does not match a physically-motivated expected model. A relevant example of such a case is when the template

contains a genomic structural variation (SV), such as translocation, whereby two fragments which are correctly adjacent in the template are located very far apart in the reference genome. Such structural variation cases are best handled in the context of the current algorithm by reporting the confidence of an observed path and reporting situations when no physically expected path seems to fit the observed data. In general, the detection of structural variation requires the presence of multiple highly significant local alignments which can be identified as significantly overturning the null hypothesis of matching the genomic ordering of fragments with their own individual merit. Nevertheless, with molecular redundant sequencing such as SMRTbell™ template sequencing the current algorithm can be adapted to improve the ability to identify an SV event. Such a modification could be a feedback approach which allows modification of the linking constraints in step 3 to allow very far separations on the target sequence when the individual alignments are very significant. Only one such highly-significant pair would be needed to enable the rescue of less significant partial matches that support the same SV hypothesis.

[00176] The software and algorithm implementations provided herein are particularly suited for transforming sequence read data generated from various sequencing technologies (e.g., sequencing-by-synthesis, intramolecular redundant sequencing, Sanger sequencing, capillary electrophoretic sequencing, pyrosequencing, ligase-mediated sequencing, etc.) into consensus sequence data that provides a representation of the actual nucleotide sequence of the template nucleic acid that was subjected to the sequencing reaction(s) from which the sequence read data was generated. The software and algorithm implementations provided herein are preferably machine-implemented methods. The various steps recited herein are preferably performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. The results of these methods are preferably stored on a machine-readable medium, as well. Further, the invention provides a computer program product comprising a computer usable medium having a computer readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention.

[00177] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and

computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data).

[00178] While described with reference to certain specific applications above, it will be understood that these methods are also applicable to other types of complex data sets, and the invention should not be limited to only the specific examples provided herein. Other applications of the instant methods will be clear to those of ordinary skill in the art and are considered to be additional aspects of the instant invention.

V. Devices and Systems

[00179] The invention also provides systems that are used in conjunction with the compositions and methods of the invention in order to provide for intermittent detection of analytical reactions. In particular, such systems typically include the reagent systems described herein, in conjunction with an analytical system, e.g., for detecting data from those reagent systems. For example, a sequencing reaction may be subjected to intermittent illumination, and the sequencing system may include the system components provided with or sold for use with commercially available nucleic acid sequencing systems, such as the Genome Analyzer System available from Illumina, Inc., the GS FLX System, available from 454 Life Sciences, or the ABI 3730 System available from Life Technologies, Inc.

[00180] In certain preferred embodiments, reactions subjected to intermittent illumination are monitored using an optical system capable of detecting and/or monitoring interactions between reactants at the single-molecule level. Such an optical system achieves these functions by first generating and transmitting an incident wavelength to the reactants, followed by collecting and analyzing the optical signals from the reactants. Such systems typically employ an optical train that directs signals from the reactions to a detector, and in certain embodiments in which a plurality of reactions is disposed on a solid surface, such systems typically direct signals from the solid surface (e.g., array of confinements) onto different locations of an array-based detector to simultaneously detect multiple different optical signals from each of multiple different reactions. In particular, the optical trains typically include optical gratings or wedge prisms to simultaneously direct and separate signals having differing spectral characteristics from each confinement in an array to different locations on an array based detector, *e.g.*, a CCD, and may also comprise additional optical transmission elements and optical reflection elements.

[00181] An optical system applicable for use with the present invention preferably comprises at least an excitation source and a photon detector. The excitation source generates and transmits incident light used to optically excite the reactants in the reaction. Depending on the intended application, the source of the incident light can be a laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source. Further, the excitation light may be evanescent light, e.g., as in total internal reflection microscopy, certain types of waveguides that carry light to a reaction site (see, e.g., U.S. Application Pub. Nos. 20080128627, 20080152281, and 200801552280), or zero-mode waveguides, described below. Where desired, more than one source can be employed simultaneously. The use of multiple sources is particularly desirable in applications that employ multiple different reagent compounds having differing excitation spectra, consequently allowing detection of more than one fluorescent signal to track the interactions of more than one or one type of molecules simultaneously. A wide variety of photon detectors or detector arrays are available in the art. Representative detectors include but are not limited to optical reader, high-efficiency photon detection system, photodiode (*e.g.* avalanche photo diodes (APD)), camera, charge couple device (CCD), electron-multiplying charge-coupled device (EMCCD), intensified charge coupled device (ICCD), and confocal microscope equipped with any of the foregoing detectors. For example, in some embodiments an optical train includes a fluorescence microscope capable of resolving fluorescent signals from individual sequencing complexes. Where desired, the subject arrays of optical confinements contain various alignment aides or keys to facilitate a proper spatial placement of the optical confinement and the excitation sources, the photon detectors, or the optical train as described below.

[00182] The subject optical system may also include an optical train whose function can be manifold and may comprise one or more optical transmission or reflection elements. Such optical trains preferably encompass a variety of optical devices that channel light from one location to another in either an altered or unaltered state. First, the optical train collects and/or directs the incident wavelength to the reaction site (e.g., optical confinement). Second, it transmits and/or directs the optical signals emitted from the reactants to the photon detector. Third, it may select and/or modify the optical properties of the incident wavelengths or the emitted wavelengths from the reactants. In certain embodiments, the optical train controls an on/off cycle of the illumination source to provide illuminated and non-illuminated periods to one or more illuminated reaction sites. Illustrative examples of such optical transmission or reflection elements are diffraction gratings, arrayed waveguide gratings (AWG), optic fibers, optical switches, mirrors (including dichroic

mirrors), lenses (including microlenses, nanolenses, objective lenses, imaging lenses, and the like), collimators, optical attenuators, filters (*e.g.*, polarization or dichroic filters), prisms, wavelength filters (low-pass, band-pass, or high-pass), planar waveguides, wave-plates, delay lines, and any other devices that guide the transmission of light through proper refractive indices and geometries. One example of a particularly preferred optical train is described in U.S. Patent Pub. No. 20070036511, filed August 11, 2005, and incorporated by reference herein in its entirety for all purposes.

[00183] In a preferred embodiment, a reaction site (e.g., optical confinement) containing a reaction of interest is operatively coupled to a photon detector. The reaction site and the respective detector can be spatially aligned (*e.g.*, 1:1 mapping) to permit an efficient collection of optical signals from the reactants. In certain preferred embodiments, a reaction substrate is disposed upon a translation stage, which is typically coupled to appropriate robotics to provide lateral translation of the substrate in two dimensions over a fixed optical train. Alternative embodiments could couple the translation system to the optical train to move that aspect of the system relative to the substrate. For example, a translation stage provide a means of removing a reaction substrate (or a portion thereof) out of the path of illumination to create a non-illuminated period for the reaction substrate (or a portion thereof), and returning the substrate at a later time to initiate a subsequent illuminated period. An exemplary embodiment is provided in U.S. Patent Pub. No. 20070161017, filed December 1, 2006.

[00184] In particularly preferred aspects, such systems include arrays of reaction regions, e.g, zero-mode waveguide arrays, that are illuminated by the system, in order to detect signals (e.g., fluorescent signals) therefrom, that are in conjunction with analytical reactions being carried out within each reaction region. Each individual reaction region can be operatively coupled to a respective microlens or a nanolens, preferably spatially aligned to optimize the signal collection efficiency. Alternatively, a combination of an objective lens, a spectral filter set or prism for resolving signals of different wavelengths, and an imaging lens can be used in an optical train, to direct optical signals from each confinement to an array detector, *e.g.*, a CCD, and concurrently separate signals from each different confinement into multiple constituent signal elements, *e.g.*, different wavelength spectra, that correspond to different reaction events occurring within each confinement. In preferred embodiments, the setup further comprises means to control illumination of each confinement, and such means may be a feature of the optical system or may be found elsewhere is the system, e.g., as a mask positioned over an array of confinements. Detailed

descriptions of such optical systems are provided, e.g, in U.S. Patent Pub. No. 20060063264, filed September 16, 2005, which is incorporated herein by reference in its entirety for all purposes.

[00185] The systems of the invention also typically include information processors or computers operably coupled to the detection portions of the systems, in order to store the signal data obtained from the detector(s) on a computer readable medium, e.g., hard disk, CD, DVD or other optical medium, flash memory device, or the like. For purposes of this aspect of the invention, such operable connection provide for the electronic transfer of data from the detection system to the processor for subsequent analysis and conversion. Operable connections may be accomplished through any of a variety of well known computer networking or connecting methods, e.g., Firewire®, USB connections, wireless connections, WAN or LAN connections, or other connections that preferably include high data transfer rates. The computers also typically include software that analyzes the raw signal data, identifies signal pulses that are likely associated with incorporation events, and identifies bases incorporated during the sequencing reaction, in order to convert or transform the raw signal data into user interpretable sequence data (See, e.g., Published U.S. Patent Application No. 2009-0024331, the full disclosure of which is incorporated herein by reference in its entirety for all purposes).

[00186] Exemplary systems are described in detail in, e.g., U.S. Patent Application No. 11/901,273, filed September 14, 2007 and U.S. Patent Application No. 12/134,186, filed June 5, 2008, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00187] Further, as noted above, the invention provides data processing systems for transforming sequence read data into consensus sequence data. In certain embodiments, the data processing systems include machines for generating sequence read data by interrogating a template nucleic acid molecule. In certain preferred embodiments, the machine generates the sequence read data using a sequencing-by-synthesis technology, as described elsewhere herein, but the machine may generate the sequence read data using other sequencing technologies known to those of ordinary skill in the art, e.g., pyrosequencing, ligation-mediated sequencing, Sanger sequencing, capillary electrophoretic sequencing, etc. Such machines and methods for using them are available to the ordinary practioner.

[00188] The sequence read data generated is representative of the nucleotide sequence of the template nucleic acid molecule only to the extent that a given sequencing technology is able to generate such data, and so may not be identical to the actual sequence of the template nucleic acid

molecule. For example, it may contain a deletion or a different base at a given position as compared to the actual sequence of the template, e.g., when a base call is missed or incorrect, respectively. As such, it is beneficial to generate redundant sequence read data, and the methods described herein provide manipulations and computations that transform redundant sequence read data into consensus sequence data that is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Redundant sequence read data comprises multiple reads, each of which includes at least a portion of sequence read that overlaps with at least a portion of at least one other of the multiple reads. As such, the multiple reads need not all overlap with one another, and a first subset may overlap for a different portion of the template nucleic acid sequence than does a second subset. Such redundant sequence read data can be generated by various methods, including repeated sequencing of a single nucleic acid template, sequencing of multiple identical nucleic acid templates, or a combination thereof.

[00189] In another aspect, the data processing systems can include software and algorithm implementations provided herein, e.g. those configured to transform redundant sequence read data into consensus sequence data, which, as noted above, is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Further, the transformation of the redundant sequence read data into consensus sequence data identifies and negates some or all of the single-read variation between the multiple reads in the redundant sequence read data. As such, the transformation provides a representation of the actual nucleotide sequence of the nucleic acid template from which redundant sequence read data is generated that is more accurate than a representation based on a single read.

[00190] The software and algorithm implementations provided herein are preferably machine-implemented methods, e.g., carried out on a machine comprising computer-readable medium configured to carry out various aspects of the methods herein. For example, the computer-readable medium preferably comprises at least one or more of the following: a) a user interface; b) memory for storing redundant sequence read data; c) memory storing software-implemented instructions for carrying out the algorithms for transforming redundant sequence read data into consensus sequence data; d) a processor for executing the instructions; e) software for recording the results of the transformation into memory; and f) memory for recordation and storage of the resulting consensus sequence read data. In preferred embodiments, the user interface is used by the

practitioner to manage various aspects of the machine, e.g., to direct the machine to carry out the various steps in the transformation of redundant sequence read data into consensus sequence data, recordation of the results of the transformation, and management of the consensus sequence data stored in memory.

[00191] As such, in preferred embodiments, the methods further comprise a transformation of the computer-readable medium by recordation of the redundant sequence read data and/or the consensus sequence data generated by the methods. Further, the computer-readable medium may comprise software for providing a graphical representation of the redundant sequence read data and/or the consensus sequence read data, and the graphical representation may be provided, e.g., in soft-copy (e.g., on an electronic display) and/or hard-copy (e.g., on a print-out) form.

[00192] The invention also provides a computer program product comprising a computer-readable medium having a computer-readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention. In certain preferred embodiments, the computer program product comprises the computer-readable medium described above.

[00193] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data), such as the computer-readable medium described above.

[00194] As shown in Figure 9, the system 900 includes a substrate 902 that includes a plurality of discrete sources of chromophore emission signals, e.g., an array of zero-mode waveguides 904. An excitation illumination source, e.g., laser 906, is provided in the system and is positioned to direct excitation radiation at the various signal sources. This is typically done by directing excitation radiation at or through appropriate optical components, e.g., dichroic 108 and objective lens 910, that direct the excitation radiation at the substrate 902, and particularly the signal sources 904. Emitted signals from the sources 904 are then collected by the optical components, e.g., objective 910, and passed through additional optical elements, e.g., dichroic 908, prism 912

and lens 914, until they are directed to and impinge upon an optical detection system, e.g., detector array 916. The signals are then detected by detector array 916, and the data from that detection is transmitted to an appropriate data processing system, e.g., computer 918, where the data is subjected to interpretation, analysis, and ultimately presented in a user ready format, e.g., on display 920, or printout 922, from printer 924. As will be appreciated, a variety of modifications may be made to such systems, including, for example, the use of multiplexing components to direct multiple discrete beams at different locations on the substrate, the use of spatial filter components, such as confocal masks, to filter out-of focus components, beam shaping elements to modify the spot configuration incident upon the substrates, and the like (See, e.g., Published U.S. Patent Application Nos. 2007/0036511 and 2007/095119, and U.S. Patent Application No. 11/901,273, all of which are incorporated herein by reference in their entireties for all purposes.)

VI. Exemplary Applications

[00195] The methods and compositions of the invention are useful in a broad range of analytical reactions in which one or more aspects of a detection method are detrimental to one or more aspects of the analytical reaction, such as rate, duration, fidelity, processivity, and the like. In such cases, intermittent detection at least partially mitigates the detrimental effect while allowing collection of data from stages of the analytical reaction that were previously uncollectable. As noted above, illuminated reactions are one example of analytical reactions that benefit from the compositions and methods described herein, particularly those using photoluminescent or fluorescent reagents, and particularly such reactions where one or more of the reaction components that are susceptible to photo-induced damage are present at relatively low levels. One exemplary application of the methods and compositions described herein is in single molecule analytical reactions, where the reaction of a single molecule (or very limited number of molecules) is observed in the analysis, such as observation of the action of a single enzyme molecule. In another aspect, the present invention is directed to illuminated reactions for single molecule analysis, including sequencing of nucleic acids by observing incorporation of nucleotides into a nascent nucleic acid sequence during template-directed polymerase-based synthesis. Such methods, generally referred to as "sequencing-by-incorporation" or "sequencing-by-synthesis," involve the observation of the addition of nucleotides or nucleotide analogs in a template-dependent fashion in order to determine the sequence of the template strand. See, e.g., U.S. Patent Nos. 6,780,591, 7,037,687, 7,344,865, 7,302,146. Processes for performing this detection include the use of fluorescently labeled

nucleotide analogs within a confined observation region, e.g., within a nanoscale well and/or tethered, either directly or indirectly to a surface. By using excitation illumination (i.e., illumination of an appropriate wavelength to excite the fluorescent label and induce a detectable signal), the fluorescently labeled bases can be detected as they are incorporated into the nascent strand, thus identifying the nature of the incorporated base, and as a result, the complementary base in the template strand.

[00196] In particular aspects, when an analysis relies upon a small population of reagent molecules, damage to any significant fraction of that population will have a substantial impact on the analysis being performed. For example, prolonged interrogation of a limited population of reagents, e.g., fluorescent analogs and enzymes, can lead to photo-induced damage of the various reagents to the point of substantially impacting the activity or functionality of the enzyme. It has been shown that prolonged illumination of DNA polymerases involved in synthesis using fluorescent nucleotide analogs results in a dramatic decrease in the enzyme's ability to synthesize DNA, often measured as a reduction in processivity. Without being bound to any theory of operation, it is believed that in some cases a photo-induced damage event affects the catalytic region of the enzyme thus affecting either the ability of the enzyme to remain complexed with the template, or its ability to continue synthesis. In general, the methods, devices, and systems of the present invention can increase performance and/or selectively monitor one or more stages of an illuminated reaction by subjecting the reaction to intermittent illumination.

[00197] One particularly preferred aspect of the invention is in conjunction with the sequencing by incorporation of nucleic acids within an optical confinement, such as a zero-mode waveguide. Such reactions involve observation of an extremely small reaction volume in which one or only a few polymerase enzymes and their fluorescent substrates may be present. Zero-mode waveguides, and their use in sequencing applications are generally described in U.S. Patent Nos. 6,917,726 and 7,033,764, and preferred methods of sequencing by incorporation are generally described in Published U.S. Patent Application No. 2003-0044781, the full disclosures of which are incorporated herein by reference in their entireties for all purposes, and in particular for their teachings regarding such sequencing applications and methods. Briefly, arrays of zero-mode waveguides ("ZMWs"), configured in accordance with the present invention may be employed as optical confinements for single molecule DNA sequence determination. In particular, as noted above, these ZMWs provide extremely small observation volumes at or near the transparent substrate surface, also termed the "base" of the ZMW. A nucleic acid synthesis complex, e.g.,

template sequence, polymerase, and primer, which is immobilized at the base of the ZMW, may then be specifically observed during synthesis to monitor incorporation of nucleotides in a template dependent fashion, and thus provide the identity and sequences of nucleotides in the template strand. This identification is typically accomplished by providing detectable label groups, such as fluorescent labeling molecules, on the nucleotides. In some instances, the labeled nucleotides terminate primer extension, allowing a "one base at a time" interrogation of the complex. If, upon exposure to a given labeled base, a base is incorporated, its representative fluorescent signal may be detected at the base of the ZMW. If no signal is detected, then the base was not incorporated and the complex is interrogated with each of the other bases, in turn. Once a base is incorporated, the labeling group is removed, e.g., through the use of a photocleavable linking group, and where the label was not the terminating group, a terminator, upon the 3' end of the incorporated nucleotide, may be removed prior to subsequent interrogation. In other more preferred embodiments, the incorporation of a labeled nucleotide does not terminate primer extension and the processive incorporation of multiple labeled nucleotides can be monitored in real time by detecting a series of fluorescent signals at the base of the ZMW. In some such embodiments, the label is naturally released upon incorporation of the labeled nucleotides by the polymerase, and so need not be released by alternative means, e.g., a photocleavage event. As such, a processive sequencing reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. (Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.)

[00198] In accordance with the present invention, sequencing reactions may be carried out by only interrogating a reaction mixture, e.g., detecting fluorescent emission for one or more illuminated periods before excessive photo-induced damage has occurred. In general, the methods described herein are implemented in a manner sufficient to provide beneficial impact, e.g., reduced photo-induced damage and/or extension of the photo-induced damage threshold period, but are not implemented in such a manner to interfere with the reaction of interest, e.g., a sequencing reaction. The present invention also contemplates alternative methods of and compositions for mitigating the impact of photo-induced damage on a reaction, as described above and in, e.g., U.S.S.N. 61/116,048, filed November 19, 2008. Such alternative methods and compounds can be used in

combination with the compositions and methods provided herein to further alleviate the effects of species that can be generated during an illuminated reaction.

[00199] Another method of mitigating the impact of photo-induced damage on the results of a given reaction provides for the elimination of potentially damaging oxygen species using means other than the use of the photo-induced damage mitigating agents described above. In one example, dissolved oxygen species may be flushed out of aqueous systems by providing the reaction system under different gas environments, such as by exposing an aqueous reaction to neutral gas environments, such as argon, nitrogen, helium, xenon, or the like, to prevent dissolution of excess oxygen in the reaction mixture. By reducing the initial oxygen load of the system, it has been observed that photo-induced damage effects, e.g., on polymerase mediated DNA synthesis, is markedly reduced. In particularly preferred aspects, the system is exposed to a xenon atmosphere. In particular, since xenon can be induced to form a dipole, it operates as a triplet-state quencher in addition to supplanting oxygen in the aqueous system. (See, e.g., Vierstra and Poff, Plant Physiol. 1981 May; 67(5): 996–998) As such, xenon would also be categorized as a quencher, as set forth above.

[00200] Although described in terms of zero-mode waveguides, it will be appreciated that a variety of selective illumination strategies may be employed to selectively interrogate different regions of a solid support over time, e.g., so as to only damage molecules within certain selected regions of a substrate while not damaging molecules in other selected regions of the substrate. In certain embodiments, such methods can involve using a directed light source (e.g., a laser) to illuminate only selected regions; changing the illumination angle of the light source; or refocusing the illumination, e.g., by passing the illumination through an optical train that alters the shape of the incident light on the solid support. These and further examples of alternative methods of mitigating photo-induced damage which can be used in combination with methods and systems of the invention described herein are provided in U.S. Patent Pub. No. 20070036511, filed August 11, 2005; U.S. Patent No. 6,881,312; U.S.S.N. 61/116,048, filed November 19, 2008; and U.S. Patent Pub. No. 20070161017, filed December 1, 2006, all of which are incorporated herein by reference in their entireties for all purposes, and in particular for disclosure related to these methods of mitigating photo-induced damage.

[00201] As noted above, using templates that allow repeated sequencing (e.g., circular templates, SMRTbell™ templates, etc.) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated and/or increase the fold-coverage of

the sequence reads for one or more regions of interest in the template, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, in certain preferred embodiments, templates sequenced by the methods described herein are templates comprising a double-stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double-stranded or otherwise internally complementary. Such templates may, for example, comprise a double-stranded portion comprised of two complementary sequences and two single-stranded linking portions (e.g., oligos or "hairpins") joining the 3' end of each strand of the double-stranded region to the 5' end of the other strand (sometimes referred to as "SMRTbell™" templates). In certain embodiments, double-stranded portions for use in such templates are PCR-amplified. Optionally, restriction sites are incorporated within the PCR primers such that subsequent digestion of the amplified products with appropriate restriction enzymes generates double-stranded portions containing known overhang sequences on either end, which are then ligated to hairpin adapters containing a complementary overhang to generate the SMRTbell™ templates.

[00202] These template molecules are particularly useful as nucleotide sequence data generated therefrom comprises both sense and antisense nucleotide sequences for the double-stranded portion, and the circular conformation of the template enables repeated sequencing (e.g., using a polymerase capable of strand-displacement) provides duplicative or redundant sequence information. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining sequence data for each segment from the complementary sequences, as well as sequence data within each segment, by repeatedly sequencing that segment. Iterative illumination is useful in such sequencing applications, e.g., to focus nucleotide sequence data collection on stages of the sequencing reaction most of interest, such as the stages during which nucleotide sequence data is being generated from a strand of the (previously) double-stranded portion. Iterative illumination may also allow additional "rounds" of sequencing the template by virtue of the reduction in photo-induced damage to reaction components, as described elsewhere herein, thereby providing more complete and robust nucleotide sequence data for future analysis, e.g., sequence scaffold construction and/or consensus sequence determination. Further, as described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including,

but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest, e.g., from the ends of the double-stranded portion of a SMRTbell™ template.

[00203] In addition to providing sense and antisense sequence data within a single template molecule that can be sequenced in one integrated process, the presence of the single-stranded linking portions also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., the sense strand, is completed and the other begins, e.g., the antisense strand. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. Additional aspects of and uses for registration sequences, e.g., for molecular redundant sequencing, are further described in U.S. Patent Publication No. 20090029385, which is incorporated herein by reference in its entirety for all purposes.

[00204] In certain embodiments, such a sequencing process begins by priming the template nucleic acid within one of the linking portions and allowing the polymerase to proceed along the strand of the double-stranded portion of the template that is immediately downstream of the primed linking portion when the double-stranded portion is melted or denatured. The sequence process proceeds around the second linking portion and proceeds along the complementary strand of the (now previously) double-stranded portion of the template. Because the template is circular, this process can continue to provide multiple repeated sequence reads from the one template. Thus, sequence redundancy comes from both the determination of complementary sequences (sense and antisense strands of the double-stranded portion), and the repeated sequencing of each circular template. The ongoing sequencing reaction is subjected to multiple illuminated and non-illuminated periods to generate at least two or more sequence reads per pass around the template. The illuminated periods are preferably timed to allow generation of nucleotide sequence data for selected regions of the template. For example, it may be beneficial to only generate nucleotide sequence data for the complementary strands of the double-stranded portion, or segments thereof. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand

with each cycle around the template, allowing continuous sequencing. Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post-synthesis.

[00205] One may optionally employ various means for controlling initiation and/or progression of a sequencing reaction, and such means may include the addition of specific sequences or other moieties into the template nucleic acid, such as binding sites, e.g., for primers or proteins. Various methods of incorporating control elements into an analytical reaction, e.g. by integrating stop or pause points into a template, are discussed elsewhere herein and are further described in related application, U.S. Application No. 12/413,258, filed March 27, 2009, which is incorporated herein by reference in its entirety for all purposes.

[00206] In certain embodiments, a reaction stop or pause point may be included within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the linking portion that was not used in priming. By way of example, following initial sequencing from the original priming location, e.g., from the single-stranded linking portion used in priming synthesis through a first portion of the sense strand (e.g., the 3' end), the data acquisition may be switched off and the polymerase allowed to proceed around the template, e.g., through the remainder of the sense strand to the other linking portion. The incorporation of a synthesis blocking moiety coupled to this linking portion will allow control of reinitiation of the polymerase activity at the 3' end of the antisense strand. One would thereby obtain paired-end sequence data for the overall (previously) double-stranded segment, with sequence data from one end coming from the sense strand and sequence data from the other end coming from the antisense strand. This template construction and sequencing methodology is particularly useful in the case of long double-stranded segments, especially given the short read lengths generated by some sequencing technologies.

[00207] A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the nucleobase portion of one or more bases in the single-stranded portion that inhibit polymerase-mediated replication; strand-binding moieties that prevent processive synthesis; non-native nucleotides included within the primer and/or template; and the like. The use of strand-binding moieties includes, but is not limited to, reversible, specific binding of particular proteins to recognition sequences incorporated into the template (or primer bound thereto) for this purpose. In certain embodiments, such control sequences may include binding sites for transcription factors, e.g., repressor binding regions provided within the linking portion(s). For example, the lac repressor recognition sequence is bound by the lac repressor protein, and this

binding has been shown to block replication in a manner reversible by addition of appropriate initiators, such as isophenylthiogalactoside (IPTG) or allolactose.

[00208] In some embodiments, primer recognition sequences and/or additional control sequences may also be provided for control of initiation and/or progression of polymerization, e.g., through a hybridized probe or reversibly modified nucleotide, or the like. (See, e.g., U.S. Patent Application No. 2008-0009007, the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Such probes include but are not limited to probes at which a polymerase initiates polymerization, probes containing various types of detectable labels, molecular beacons, TaqMan® probes, Invader® probes (Third Wave Technologies, Inc.), or the like, that can be used for various purposes, e.g., to provide indications of the commencement and/or progress of synthesis.

[00209] An engineered pause point (reversible or irreversible) can include one or more non-native (non-natural) or fifth bases that do not pair with any of the four native nucleoside polyphosphates in the synthesis reaction, e.g., in the template and/or oligonucleotides probe(s), and/or that exhibit a distinct kinetic signature during template-dependent synthesis at such a base. Upon encountering such a base, the polymerase pauses until the complement to the non-natural base is added to the reaction mixture. Likewise, an engineered pause point could include a "damaged" base that causes a stop in replication until repair enzymes are added to the mixture. For example, a template having a pyrimidine dimer would cause the replication complex to pause, and addition of the photolyase DNA repair enzyme would repair the problem location and allow replication, and sequencing to continue. In yet further embodiments, a combination of modification enzymes could be used to engineer a set of modified bases on a template, e.g., a combination of glycosylases, methylases, nucleases, and the like. (Further information on sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.)

[00210] As noted elsewhere herein, stop or pause points can be engineered into various portions of the template, e.g., portions for which the nucleotide sequence is unknown (e.g., a genomic fragment) or known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, SMRTbell™ templates are topologically closed, single-stranded molecules having regions

of internal complementarity separated by hairpin or stem-loop linkers, such that hybridization of the regions of internal complementarity produces a double-stranded portion within the template. One or both of the linkers can comprise a stop or pause point to modulate polymerase activity. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00211] In other embodiments, an abasic site is used as a synthesis blocking moiety or pause point until addition of a non-natural "base," such as a pyrene, which has been shown to "base-pair" with an abasic site during DNA synthesis. (See, e.g., Matray, et al. (1999) Nature 399(6737):704-8, which is incorporated herein by reference in its entirety for all purposes.) Where a permanent termination of sequencing is desired, no non-natural analog is added and the polymerase is permanently blocked at the abasic site. DNA (or RNA) glycosylases create abasic sites that are quite different from the normal coding bases, A, T, G, and C (and U in RNA). A wide variety of monofunctional and bifunctional DNA glycosylases that have specificity for most common DNA or RNA adducts, including 5-methylcytosine, are known in the art, with different glycosylases capable of recognizing different types of modified DNA and/or RNA bases. The molecular structures of many glycosylases have been solved, and based on structural similarity they are grouped into four superfamilies. The UDG and AAG families contain small, compact glycosylases, whereas the MutM/Fpg and HhH-GPD families comprise larger enzymes with multiple domains. As an example, four enzymes have been identified in *Arabidopsis thaliana* in the plant pathway for cytosine demethylation. Additionally, other enzymes are also known to recognize 5-methyl cytosine and remove the methylated base to create an abasic site. Further, various enzymes are known to methylate cytosine in a sequence-specific manner. As such, a combination of a cytosine-methylase and an enzyme that creates an abasic site from a methylated cytosine nucleotide can be used to create one or more abasic sites in a template nucleic acid. The size of the recognition site of the methylase and the base composition of the template determine how frequently methylation occurs, and therefore, the number of abasic sites created in a given template nucleic acid, allowing the ordinary practioner to choose a methylase with a recognition site that produces a desired spacing between modified nucleotides. For example, if the recognition site is three bases long, then on average an abasic site is expected every 64 bases; if the recognition site is four bases long, then on

average an abasic site is expected every 256 bases; if the recognition site is six bases long, then on average an abasic site is expected every 4096 bases; and so forth. Of course, templates with a higher GC content would be expected to have more frequent abasic site formation, and templates with lower GC content would be expected to have less frequent abasic site formation.

[00212] Uracil-DNA glycosylases can also be used to introduce abasic sites into a template nucleic acid comprising deoxyuridine nucleotides. This strategy has the advantage of allowing the practitioner to choose the locations of the abasic sites within a DNA template since deoxyuridine nucleotides are not generally found in DNA. Various methods of inserting deoxyuridine nucleotides into a DNA template may be used, and different methods will be preferred for different applications. In certain embodiments, one or more site-specific deoxyuracils are incorporated during standard phosphoramidite oligonucleotide synthesis. To place uracils at indeterminate positions in a DNA, replacing a portion of the deoxythymidine triphosphate with deoxyuridine triphosphate will result in an amplimer with random U sites in place of T sites after polymerase chain reaction. In other embodiments, deoxyuridine nucleotides are engineered into the template, e.g., by ligation of a synthetic linker or adaptor comprising one or more deoxyuridine nucleotides to a nucleic acid sequence to be sequenced. In certain preferred embodiments, deoxyuridine nucleotides are incorporated into the linker portions of a SMRTbell™ template.

[00213] To subsequently introduce abasic sites prior to sequencing, the deoxyuridine nucleotide-containing template is subjected to treatment with uracil-DNA glycosylase, which removes the one or more uracil bases from the deoxyuridine nucleotides, thereby generating one or more abasic sites in the template. Alternatively, since the deoxyuridine nucleotide can be recognized as a template base and paired with deoxyadenosine during template-dependent nascent strand synthesis, the synthesis-blocking abasic site can instead be introduced after initiation of the sequencing reaction, e.g., at a time chosen by the practitioner. For example, the reaction can be initiated with a deoxyuridine-containing template, and uracil-DNA glycosylase can subsequently be added to block the polymerase and halt the reaction after the reaction has proceeded for a given time. As such, termination of the reaction is optional rather than required.

[00214] While uracil-DNA glycosylase activity is useful for introducing abasic sites into a template as described above, this activity can be problematic during the preparation of such templates. As such, strategies are typically implemented during preparation and manipulation of uracil-containing DNA, e.g., using molecular biology enzymes, to avoid uracil-DNA glycosylase activity, in particular, due to the *E. coli* UDG enzyme. Since a majority of standard molecular

biology enzymes are overexpressed and subsequently purified from an *E. coli* host, UDG activity can be a contaminating activity that is often not monitored by the enzyme manufacturer's quality control procedures. To mitigate contaminating UDG activity, a commercially available UDG inhibitor, also known as uracil glycosylase inhibitor or UGI (e.g., from New England Biolabs, Ipswich, MA) can be included in molecular biology reactions. This is a small protein inhibitor from the *B. subtilis* bacteriophase PBS1 that binds reversibly to E. coli UDG to inhibit its catalytic activity. UGI is also capable of dissociating UDG from a DNA molecule. Alternatively, UDG activity can be inhibited without exogenous protein using a chemical inhibitor of the enzyme, such as an oligonucleotide containing a 1-aza-deoxyribose base, a transition state analog for the UDG enzyme. This and other cationic nitrogenous sugars have been used for mechanistic studies of UDG activity and show potent inhibition activity. (See, e.g., Jiang et al. Biochemistry, 2002, 41 (22), pp 7116–7124.)

[00215] In certain applications, UDG activity needs to be inhibited temporarily, and subsequently enabled to remove create an abasic site as described above. In some embodiments, a DNA purification that removes proteins is employed, e.g., including a phenol-chloroform extraction with subsequent ethanol precipitation, a silica-based column approach (e.g., QiaQuick columns from Qiagen and similar products), and/or a PEG/sodium chloride precipitation (e.g., AMPure beads from Beckman Coulter). Alternatively or additionally, a commercially-available UDG enzyme that is not inhibited by UGI is added when abasic site formation is desired. For example, the *A. fulgidus* UDG is from a thermophilic organism and cannot be inhibited by the same bacteriophage protein as is the *E. coli* UDG enzyme. In certain preferred embodiments, UDG-inhibition is employed during template preparation, and inhibition-resistant UDG activity is added at a subsequent time to trigger the creation of abasic sites at deoxyuridine nucleotides, e.g., immediately prior to or during an ongoing reaction.

[00216] In some preferred embodiments, one or more abasic sites are engineered into a linker or adapter sequence within a sequencing template molecule. Abasic sugar residues serve as efficient terminators of polymerization for many polymerases, e.g., Φ29. 1',2'-dideoxyribose is the most common synthetic "abasic site". In other embodiments, a synthetic linker is incorporated into a linker or adaptor. For example, an internal spacer (e.g., Spacer 3 from Biosearch Technologies, Inc.) or other carbon-based linker can be used in lieu of a sugar-base nucleotide. Similar to an abasic nucleotide, the polymerase will be blocked upon encountering these moieties in the template nucleic acid.

[00217] In certain embodiments, synthesis blocking moieties are nicks in the template nucleic acid. Nicking enzymes (e.g., nicking endonucleases) are known in the art and can be used to specifically nick the template prior to or during a template-directed sequencing reaction. The use of site-specific nicking endonucleases allows the practitioner to incorporate a recognition sequence at a particular location within the template nucleic acid, and such nicking endonucleases are commercially available, e.g., from New England Biolabs, Inc. For example, a linker or adapter can be synthesized with a nicking endonuclease recognition sequence, ligated to a nucleic acid molecule to be sequenced, and can be specifically nicked either before or during a subsequent sequencing reaction. Nicks can also be introduced by ligating duplex segments that lack either a terminal 3'-hydroxy (e.g., have a dideoxynucleotide at the 3'-terminius) and/or 5'-phosphate group on one strand. The ligation results in covalent linkage of the phosphodiester backbone on one strand, but not on the other, which is therefore effectively "nicked." In certain embodiments, a SMRTbell™ template is constructed using a duplex (or "insert") nucleic acid molecule lacking a 5'-phosphate group at one or both termini. Upon ligation of the hairpin or stem-loop adaptors at each end, nicks are created at one or both ligation site(s), depending on whether the duplex lacked a 5'-phosphate at one or both ends, respectively. In other embodiments, a SMRTbell™ template is constructed using one or two stem-loop adaptors lacking a 3'-hydroxy group at the terminus (e.g., comprising a 2',3'-dideoxynucleotide rather than a 2'-deoxynucleotide). Upon ligation of one or two stem-loop adaptors lacking a 3'-hydroxy group, one or two nicks are created at the ligation site(s), depending on whether one or two adaptors lacked the 3'-hydroxy group, respectively. In both cases, a nick is created in the template nucleic acid, and a primer bound to one of the adaptors provides an initiation site for the polymerase, which will process the template until encountering a nick, at which point the polymerase will terminate the reaction by dissociation from the template. Regardless of how a nick is created, the position of a nick relative to the initiation site for the polymerase determines how much of the template will be sequenced. For example, Figure 19A provides an illustrative example of an embodiment in which a nick is present on a first strand of a duplex portion at a position distal to the adaptor containing the primer binding site. The first strand is processed by a polymerase, but the complementary strand is not processed because the polymerase dissociates at the nick site. An alternative embodiment is shown in Figure 19B, in which a nick is present on the strand complementary to the first strand at a position proximal to the adaptor containing the primer binding site. In this case both the first and complementary strands, as well as the adaptor not containing the primer binding site, are processed by the polymerase prior to dissociation. The

position of the primer binding site also determines how much of the template is processed by the polymerase. Figure 19C provides a template having a primer binding site at a position from which a polymerase would process a significant portion of the adaptor prior to entering the duplex portion. An additional advantage to using a 3'-dideoxynucleotide at a nick is that it prevents the use of the nick as a polymerse initiation site, since strand extension requires a 3-hydroxy group. As such, the resulting nick would not compete with a primer site for initiation of nascent strand synthesis by the polymerase. Having a single, known site of initiation on a template molecule is beneficial, e.g., for subsequent mapping of a read generated in such a reaction. In certain preferred embodiments, a nick site both lacks a 5'-phosphate group and comprises a 3'-dideoxynucleotide.

[00218] In certain preferred embodiments, modification and base excision is performed prior to introduction of a template nucleic acid to a reaction site, e.g., a zero-mode waveguide. As noted above, the choice of recognition site for the methylase depends on how far apart the practitioner wishes point of synthesis initiation to be on the template. For example, after initiating the template-dependent sequencing reaction, the sequence of nucleotide incorporations into the nascent strand is monitored for a desired sequence read, which may extend from the initiation point to the pause point, or may end before the polymerase reaches the pause point. In some preferred embodiments, as described elsewhere herein, the monitoring is suspended by modifying or removing an illumination source, e.g., by moving the illumination source or a substrate comprising the reaction site. Synthesis of the nascent strand will continue until the pause site is reached, whether or not the reaction is being actively monitored. When the reaction is to be reinitiated, reaction components are added that allow bypass, e.g., pyrene, polymerase, etc., and these can be subsequently removed (e.g., by buffer exchange) to allow additional pauses at other pause sites on the template.

[00219] In certain embodiments using pyrosequencing-based technologies (e.g., as developed by 454 Life Sciences), abasic sites can be introduced into a set of amplified template nucleic acids and synthesis initiated. Since all templates in the set are identical, they will comprise the same number of abasic sites in the same positions. During the course of the synthesis reaction, the synchronous incorporation of nucleotides into the nascent strands is monitored until either an abasic site is reached (at which point the synthesis is paused) or until the incorporation becomes asynchronous (which increases the background noise and decreases reliability of the sequence read). In the latter case, the practioner may opt to speed up the reaction, e.g., by adding all nucleotides at one time, to extend all nascent strands to the first abasic site in the templates. When synthesis is to be reinitiated, reaction components are added that allow bypass of the abasic site, e.g. one or more

pyrenes. A wash step may be performed to remove nucleotides and/or polymerases from the reaction sites prior to such addition. Further, in some cases, a different polymerase may be used for pyrene incorporation as is used for sequencing-by-synthesis reactions. In certain preferred embodiments, the reaction mixture comprising the pyrene for abasic site bypass allows readthrough of the abasic site, but no further on the template. Subsequent addition of sequencing reaction mixture allows the sequencing-by-synthesis reaction to recommence and incorporation of nucleotides into the nascent strand to be monitored. Alternatively or additionally, the practioner need not wait until an abasic site is reached to suspend detection and, optionally, speed up the reaction to bring all nascent strands to a given abasic site, but can choose to do this before a reaction has become asynchronous, e.g., after desired sequence data has been collected for a particular region of interest in a template nucleic acid.

[00220] In certain embodiments using ligation-based technologies (e.g., the SOLiD™ System developed by Life Technologies), a pause site can be engineered by using an oligonucleotide that cannot participate in the ligation reaction and that is complementary to a desired location on the set of identical template nucleic acids, e.g., on a bead. When the serial ligation reaction hits the position recognized by this polynucleotide, the reaction cannot proceed and any reactions that have become asynchronous will "catch up." The user can then unblock the oligo (e.g., using chemical treatment or photo-cleavage) and reinitiate the sequencing reaction.

[00221] In some cases, it may be desirable to provide endonuclease recognition sites within the template nucleic acid. For example, inclusion of such sites within a circular template can allow for a mechanism to release the template from a synthesis reaction, i.e., by linearizing it, and allowing the polymerase to run off the linear template, and/or to expose the template to exonuclease activity, and thus terminate synthesis through removal of the template. Such sites could additionally be exploited as control sequences by providing specific binding locations for endonucleases engineered to lack cleavage activity, but retain sequence specific binding, and could therefore be used to block progression of the polymerase enzyme on a template nucleic acid.

[00222] In some cases, nicking sites, e.g., sites recognized by nicking endonucleases, may be included within a portion of the template molecule, and particularly within a double-stranded portion of the template, e.g., in a double-stranded segment of a SMRT bell™ or in the stem portion of an exogenous hairpin structure. Such nicking sites provide one or more breaks in one strand of a double-stranded sequence and can thereby provide one or more priming locations for, e.g., a strand-displacing polymerase enzyme. A variety of nicking enzymes and their recognition sequences are

known in the art, with such enzymes being generally commercially available, e.g., from New England Biolabs.

[00223] In certain embodiments, methods for intermittent detection described herein are useful in "paired-end" sequencing applications in which sequence information is generated from two ends of a template nucleic acid but not for at least a portion of the intervening portion of the template. Typically, paired-end sequencing applications provide sequence data for only the two ends of a nucleic acid template, but the present invention also allows generation of additional sequence reads that are noncontiguous with the sequence reads from the ends of the template. In certain preferred embodiments, a duplex fragment (e.g., genomic fragment) is ligated to a single-stranded linker that connects the 3' end of the sense strand to the 5' end of the antisense strand, or that connects the 5' end of the sense strand to the 3' end of the antisense strand. In either orientation, separation of the two strands of the duplex fragment results in a single-stranded linear template nucleic acid that contains the linker in between the sense and antisense strands. Subsequent sequencing can involve intermittent detection that generates sequence reads for only the portions of the sense and antisense strands that are of interest, e.g., one or both of the ends. In certain embodiments, both sense and antisense strands may be sequenced at both ends to provide redundancy in the sequence data. Sequence reads recognized as being from the linker portion of the template (e.g., based on the known linker sequence or specific registration sequences encoded therein) can be used to orient the alignment of the sequence reads from the sense and antisense portions of the template, providing context for determining the sequences of the ends of the duplex fragment and subsequent sequence scaffold construction and/or mapping. In certain embodiments, pause or stop points may be incorporated into the linker to control the processing of the template by the polymerase, and therefore may be used to synchronize the detection periods to ensure generation of sequence reads from particular regions of template. Further, additional detection periods can be included that are timed to provide sequence reads from portions of the sense and/or antisense strand that are noncontiguous with the end regions.

[00224] In a related embodiment, paired-end sequencing may be accomplished by using a nucleic acid template that has linkers connecting the sense and antisense strands of a duplex fragment at both ends, such that separation of the strands of the duplex fragment provides a single-stranded circular template that contains a linkers in between each end of the sense and antisense strands of the original duplex fragment. Such a template molecule would allow a strand-displacing polymerase to proceed around the template multiple times, thereby potentially generating redundant

sequence data from both ends of both strands of the original duplex fragment. As noted elsewhere herein, such redundancy is useful for determination of consensus sequences and/or construction of sequence scaffolds. As the polymerase enzyme processes the template, detection periods can be timed (e.g., based on knowledge of the rate at which the polymerase processes the template, which is dependent not only on the polymerase but also on the sequence of the template itself) to generate nucleotide sequence reads from the regions of the template corresponding to one or both ends of the sense and antisense strands, and can also include detection periods to generate additional reads from other, noncontiguous regions of the duplex fragment, as well. Although such timing can be used to determine the appropriate periodicity of the detection periods, at later stages of the reaction (e.g., as the polymerase repeatedly proceeds around the template), the exact location of reinitiation of sequence read generation becomes more approximate. Incorporation of pause or stop points into one or both linkers to regulate the processing of the template by the polymerase may be used to synchronize the detection periods regardless of the total distance travelled by the polymerase around the template. This strategy more reliably ensures generation of sequence reads from selected regions of template, e.g, the ends of the sense and antisense portions and, optionally, regions in between and noncontiguous with the end regions regardless of the number of passes of the polymerase around the template nucleic acid, especially in later stages of the reaction. Further, the known sequence of one or both of the linkers can be used to orient sequence reads from the sense and antisense portions for consensus sequence determination and/or mapping.

[00225] In some such embodiments, a duplex fragment inserted between two hairpin linkers may be much larger than desired, increasing the difficulty of limiting nucleotide sequence read data to particular regions of the fragment. The size of the duplex fragment ligated to the two hairpin linkers can be selectively reduced to retain the regions attached to the linkers and to lose a central portion of the duplex fragment. One particularly preferred strategy, illustrated in Figure 18, comprises hairpin linkers (1802, 1804) having a regions of cross-complementarity (1806, 1808), such that the two linkers 1802 and 1804 can anneal to each other in a manner that does not interfere with ligation to a duplex fragment 1810. Duplex fragment 1810 comprises ends 1812 and 1814, as well as a long central region 1816, which is not shown but is understood to be between the two curvy lines. Once end 1812 is ligated to linker 1802 and end 1814 is ligated to linker 1804, the construct is subjected to fragmentation, which removes the central region 1816 of the duplex fragment 1810, producing construct 1818 having ends 1820 and 1822. After fragmentation, the ends of the portions of the duplex fragment still associated with the annealed linker pair (ends 1820 and

1822) are ligated together to produce construct 1824, which can then be treated (e.g., with heat, gentle denaturation, primer invasion, changing salt concentration, etc.) to separate cross-complementary regions 1806 and 1808 from one another, e.g., to generate a circular single-stranded nucleic acid molecule. Alternatively, the separation may occur during the course of the subsequent reaction, e.g., by polymerase-mediated strand displacement. Yet further, where the hybridized cross-complementary regions are long enough to undergo a complete DNA turn, an additional reaction component (e.g., helicase, topoisomerase, polymerase, etc.) may be needed to unwind the duplex and allow separation. As such, the resulting "mate-pair" construct has only the ends of the original duplex fragment ligated together and capped with adaptors that link the 5' end of each strand of the duplex with the 3' end of the other strand of the duplex, and denaturation of the duplex produces a closed, single-stranded circular construct.

[00226] Fragmentation of the duplex fragment can be performed by a variety of known methods. For example, fragmentation can be performed enzymatically (e.g., using restriction enzymes or other nucleases) or mechanically, by shearing or sonication. The type of fragmentation chosen will determine various characteristics of the resulting construct, e.g., how large a central region is removed and the types of ends remaining (e.g., blunt, 5' overhang, 3' overhang, random, identical on both ends, etc.). Optionally, the ends can be modified after fragmentation to facilitate the subsequent ligation step. Although not shown in Figure 18, it is expected that the ligation of the duplex fragment to the hybridized linkers will be a two-step process, with one end being ligated first and unimolecular kinetics favoring ligation of the second end to the second linker. The cross-complementary regions of the linkers can be designed to produce varying levels of complementarity, and therefore varying strengths of the hybridization. For example, a longer or higher GC content in a cross-complementary region lends a higher stability to the linker:linker interaction, but separation of the hybridized linkers requires a more severe treatment, e.g., higher temperature, more stringent conditions, etc. As such the cross-complementary regions should be engineered to produce a stable linker:linker interaction that is disruptable under conditions that are not destructive to the overall construct. Further the linkers can vary in regions apart from the cross-complementary regions For example, one linker can have a primer binding site that the other lacks, which would provide a single polymerase initiation site in the final construct. Other sequence characteristics described herein (e.g., pause sites, registrations sequences, etc.) can also be included in one or both linker regions. If topological constraints limit the subsequent processing of the resulting contruct, e.g., during template-directed nascent strand synthesis, these can be addressed by

addition of a reaction component (e.g., a helicase or topoisomerase) to resolve the topological constraint. As such, the methods can be used to add asymmetric linkers to duplex polynucleotides, whether or not the duplex is to be selectively reduced in size, or not, as long as the asymmetric linkers can cross-hybridize to one another.

[00227] Although in preferred embodiments, the two linkers to be ligated to a single duplex fragment are hybridized to one another prior to ligation, in some embodiments they are instead hybridized after the initial ligation reaction, and where topological constraints inhibit such a post-ligation hybridization a reaction component (e.g., topoisomerase) may be included to relieve such constraints. In certain embodiments, the hybridized linkers are separated prior to addition of reaction components for a subsequent reaction, and in other embodiments the hybridized linkers are not separated until after the addition of reaction components for a subsequent reaction. For example, a polymerase enzyme may bind to a primer annealed to a linker before or after separation of the linker from a second linker. In fact, it may be benefical in some embodiments to postpone separation of the linkers, e.g., where compaction of the nucleic acid construct is beneficial, such as when the construct must be loaded into a confinement of some kind, e.g., a nanowell, optical confinement, etc.

[00228] In some embodiments, the methods further include separation of single linker constructs from hybridized linker pair constructs. This can be accomplished by an exonuclease treatment after ligation of the duplex fragment to the linkers, which would degrade any constructs having an unannealed end. Alternatively, it may be desirable to remove the single linkers prior to ligation, for example using a size separation methodology or by allowing them to bind to oligonucleotides that are complementary to the cross-complementary regions and bound to a column or magnetic beads. (The cross-complementary regions of the hybridized linker pairs will not be available for binding to the oligonucleotides. Other methods known in the art can also be used to separate single linkers from hybridized linker pairs.

[00229] Interestingly, the use of the sense/antisense nucleic acid templates described above would represent a unidirectional processing of a template to provide paired-end sequence data, as opposed to the more traditional bi-directional processing of a linear template molecule. Further, unlike traditional approaches, these methods for paired-end sequencing involve processing, chemically or otherwise, of not just the regions at the ends, but also regions in between the ends, and in some embodiments comprising processing of the entire template. For example, a polymerase incorporates nucleotides into a nascent strand for each position of the template (thereby

"processing" each position of the template), yet the sequencing data generated is limited to specific regions of the template that are of particular interest to the practitioner, such as the end regions. As such, in certain embodiments the duplex fragment is not further reduced in size after ligation to a linker pair, and the entire duplex fragment is processed by the polymerase.

[00230] In certain embodiments, methods for intermittent detection described herein are useful in analysis systems that employ nanopores. A nanopore is a small pore in an electrically insulating membrane that can be used for single molecule detection. In general, a nanopore functions as a Coulter counter for much smaller particles, and can take various forms, e.g., a protein channel in a lipid bilayer or a pore in a solid-state membrane. The detection principal is based on monitoring the ionic current of an electrolyte solution passing through the nanopore as a voltage is applied across the membrane. For example, passage of a polynucleotide molecule (e.g., DNA, RNA, etc.) through a nanopore causes changes in the magnitude of the current through the nanopore, with each nucleotide obstructing the nanopore to a different, characteristic degree. As such, the pattern of variations in the current passing through the nanopore as the polynucleotide is drawn through may be monitored and analyzed to determine the nucleotide sequence of the polynucleotide. A polynucleotide may be drawn through the nanopore by various means, e.g., by electrophoresis, or using enzyme chaperones to guide the polynucleotide through the nanopore. For additional discussion of methods of fabrication and use of nanopores, see, e.g., U.S. Patent No. 5,795,782; Kasianowicz, J.J., et al. (1996) *Proc Natl Acad Sci USA* 93(24):13770-3; Ashkenas, N., et al. (2005) *Angew Chem Int Ed Engl* 44(9):1401-4; Winters-Hilt, S., et al. (2003) *Biophys J* 84:967-76; Astier, Y., et al. (2006) *J Am Chem Soc* 128(5):1705-10; Fologea, D., et al. (2005) *Nano Lett* 5(10):1905-9; Deamer, D.W., et al. (2000) *Trends Biotechnol* 18(4):147-51; and Church, G.M. (2006) *Scientific American* 294(1):52, all of which are incorporated by reference herein in their entireties for all purposes. In some embodiments, intermittent detection of nucleic acid sequence data from a nanopore may be achieved by modifying the progress of the polynucleotide through the nanopore so that progress is sped up during non-detection periods and progress is slowed to allow sequence determination during detection periods. The rate of passage of the polynucleotide through the nanopore may be modified by various methods, including but not limited to increasing an electrophoretic field carrying the polynucleotide (e.g., by increasing the voltage, changing the conductivity of the reaction mixture, and the like), or changing various reaction conditions to alter the speed at which a protein chaperone carries the polynucleotide. Further, in embodiments utilizing

a processive exonuclease to feed individual bases through the nanopore, the kinetics of the exonuclease may be modified based on the known biochemical characteristics of the exonuclease.

[00231] In diagnostic sequencing applications, it may be necessary only to provide sequence data for a small fragment of DNA, but do so in an extremely accurate sequencing process. For such applications, shorter target segments may be employed, thus permitting a higher level of redundancy by sequencing multiple times around a smaller circular template, where such redundancy provides the desired accuracy. Thus, in some cases, the double stranded target segment may be much shorter, e.g., from 10 to 200, from 20 to 100 or from 20 to 50 or from 20 to 75 bases in length. For purposes of the foregoing, the length of the target segment in terms of bases denotes the length of one strand of the double stranded segment. In such applications, various methods for intermittent detection described herein may be used to analyze the sequence of the template, thereby targeting the sequence data to the portion(s) of the template of particular interest to the diagnostician, and/or improving various aspects of the reaction performance, e.g., by virtue of the reduction of photo-induced damage to one or more reaction components.

[00232] It is to be understood that the above description is intended to be illustrative and not restrictive. It readily should be apparent to one skilled in the art that various embodiments and modifications may be made to the invention disclosed in this application, including but not limited to combinations of various aspects of the invention, without departing from the scope and spirit of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. All publications mentioned herein are cited for the purpose of describing and disclosing reagents, methodologies and concepts that may be used in connection with the present invention. Nothing herein is to be construed as an admission that these references are prior art in relation to the inventions described herein. Throughout the disclosure various patents, patent applications and publications are referenced. Unless otherwise indicated, each is incorporated by reference in its entirety for all purposes.

[00233] Although described in some detail for purposes of illustration, it will be readily appreciated that a number of variations known or appreciated by those of skill in the art may be practiced within the scope of present invention. Unless otherwise clear from the context or expressly stated, any concentration values provided herein are generally given in terms of admixture values or percentages without regard to any conversion that occurs upon or following addition of

the particular component of the mixture. To the extent not already expressly incorporated herein, all published references and patent documents referred to in this disclosure are incorporated herein by reference in their entirety for all purposes.

[00234] The following non-limiting examples are provided to further illustrate the invention.

VI. Examples of Intermittent Illumination of a Single Molecule Sequencing-by-Synthesis Reaction

Example I

[00235] A nucleic acid template was provided that comprised a double-stranded region and two single-stranded linker portions at each end. The first linker portion connected the 3' end of the sense strand with the 5' end of the antisense strand, and the second linker portion connected the 3' end of the antisense strand with the 5' end of the sense strand. This template was designed to form a single-stranded circle of approximately 500 bases when the double-stranded region was opened (e.g., by heat denaturization, helicase activity, etc.), and is sometimes referred to as a SMRTbell™ template. A plurality of this nucleic acid template was incubated with polymerases, primers, and other reaction components to allow formation of polymerase-template complexes. (See, e.g., Korlach, J., et al. (2008) Nucleosides, Nucleotides and Nucleic Acids, 27:1072-1083; and Eid, J. (2009) Science 323:133-138.) The complexes were immobilized in zero-mode waveguides in a reaction mixture containing all necessary buffer and nucleotide analog components for carrying out sequencing-by-synthesis reactions with the exception of a cognate starting base and a metal dication. A Smith-Waterman algorithm was used to perform the alignment of the known sequence of the template with the sequence reads generated in the reaction, and the positions of the sequence reads is graphically illustrated in Figure 8.

[00236] Acquisition of the data shown in Figure 8 was collected as follows. Illumination of the array of zero-mode waveguides was initiated with laser excitation (532 nm and 641 nm laser lines) at t = -5 seconds, and the missing cognate starting base and metal dication (manganese metal) were added at t = 0 seconds to simultaneously initiate the sequencing-by-synthesis reactions in all zero-mode waveguides. The reactions were monitored under illumination for 120 seconds at which time the illumination was removed; the sequencing reads generated during that stage of the reaction are shown in Figure 8A as a function of the template position to which each read maps. At 295 seconds illumination was resumed and data acquisition was reinitiated at 300 seconds and maintained for another 120 second interval; the sequencing reads during this second illuminated

period are shown in Figure 8B. At 595 seconds illumination was resumed and data acquisition was reinitiated at 600 seconds and maintained for another 120 second interval; the sequencing reads during this third illuminated period are shown in Figure 8C.

[00237] As expected, the longer the amount of time before the sequence data is collected (that is, the later the illuminated period), the further into the template the alignments shift, and this shift is a rough function of time since initiation of the reaction. Further, the distribution of sequence reads generated during each subsequent illuminated period becomes more dispersed than the previous illuminated period(s). Further, due to the circular nature of the template, Figure 8C clearly shows that some polymerases have passed completely around the substrate and are beginning to generate sequence reads from a second pass around the template, thereby generating redundant sequence information for a single template nucleic acid.

Example II

[00238] As in Example I, a SMRTbell™ template was used. For templates of defined sequence, PCR was used to generate 3 or 6 kb DNA inserts for the double-stranded region in the SMRTbell™ templates using a standard PCR methodology. For genomic and other biological samples, a DNA fragmentation protocol was used that generates DNA fragments distributed around 3 or 6 kb. Generation of fragments in these ranges was done using a HydroShear® (Genomic Solutions®) device with settings recommended by the manufacturer. The random genomic DNA fragments were enzymatically treated to generate blunt ends. Both the PCR products and randomly generated DNA fragments were phosphorylated and then immediately put into a ligation reaction with a blunt hairpin adapter. The products were purified through two size selection steps using reduced volumes of AMPure® magnetic beads (Agencourt®) to remove hairpin dimers and other short products. (Fabrication of SMRTbell™ templates is further described elsewhere herein.)

[00239] The system components used for polynucleotide sequencing using intermittent detection are comparable to single-molecule sequencing applications under constant illumination, which are described, e.g., in Eid, et al. (2009) Science 323:133-138. Specifically, the immobilization and sequencing buffer compositions, nucleotide analogs identity and concentration, polymerase, ZMWs, surface treatment and instrumentation were identical to the standard methodology. Modifications to the SMRTbell™ template DNA and polymerase binding and immobilization and data acquisition protocols are as follows.

[00240] A binding solution was prepared by incubation of 3 or 6 kb DNA SMRTbell™ templates (1-10 nM) with a 10-fold excess of DNA polymerase (10-100 nM, respectively) in 10 mM MOPS (pH 7.5), 10 mM KOAc, 100 mM DTT & 0.05% Tween-20 for 2 hours at 30°C, followed by 1 hour at 37°C and subsequent storage at 4°C prior to immobilization on the ZMWs. Immediately prior to immobilization, the binding solution was diluted in the standard immobilization solution (50 mM MOPS (pH 7.5), 75 mM KOAc, 5 mM DTT, 0.05% Tween-20) to the desired final concentration, typically 0.1 to 1 nM, and incubated for 30 to 60 minutes at 22°C. Post-immobilization chip preparation and sequencing initiation were identical to the standard methods.

[00241] The data acquisition protocol was similar to the standard application with coordinated modifications to the collection timing and ZMW positioning. In the standard acquisition procedure, a single long acquisition (~10 minutes) is performed for each ZMW. In the intermittent illumination acquisition procedure, multiple short acquisitions (~3 minutes) of sequence reads (also termed "strobe reads") were performed for each ZMW (during "detection periods") with an interval between each acquisition period during which no acquisition of sequence reads was performed ("non-detection period"). The duration of the interval between each acquisition of sequence reads was determined based upon a desired distance (i.e., number of nucleotide positions) between each sequence (or strobe) read, the polymerization rate of the polymerase, and the SMRTbell™ template insert size.

[00242] SMRTbell™ templates were generated as described above for AC223433, a fosmid clone comprising a sequence of an approximately 40 kb region of *Homo sapiens* chromosome 15. The reference sequences used to map the sequence reads generated in the sequencing reactions were the publically available sequences of *Homo sapiens* chromosome 15 (Hg18; NCBI Build 36.1) and fosmid AC223433 (NCBI GenBank accession number). Table 1 shows the number of statistically significantly mapped sequence reads for several types of intermittent illumination sequencing reactions. The number of mappable "looks" is equivalent to the number of mappable sequence reads generated during detection periods for a single template molecule. For example, a "mapped 1-look read" means, for a single template molecule, only a single detection period generated a sequence read that could be mapped to the reference sequence.

Table 1: Summary of Sequencing Results

Mapping	Mapped 1-	Mapped 2-	Mapped 3-	Mapped 4-

Reference	look reads	look reads	look reads	look reads
Human chr15 (Hg18)	13834	1289	127	4
Fosmid	15253	1571	158	5

[00243] Deviations in the expected time span for a set of sequencing reads from a single sequencing reaction are indicative of genomic events such as genomic rearrangements, e.g., insertions, deletions, etc. Figures 10 and 11 illustrate this point. Specifically, the time and distance travelled along the template (based upon the reference sequence) by the polymerase was computed within and between the sequence reads generated during the detection (illuminated) periods. These calculations were used to detect unexpected variations, indicating possible genomic events in the template as compared to the reference sequence. Figure 10 provides a plot that illustrates the normalized average time it took for the polymerase to traverse a region of the template based on the length of that region in the *Homo sapiens* chromosome 15 reference sequence. The sequence reads are fit to a diagonal having a slope equal to the average speed for sequencing reads. Deviations from the regressed diagonal indicate genomic events (for example, structural variants), and the slope of the sequence reads around such deviations indicate the relative size of the genomic event (e.g., in the case of insertions/deletions). For example, if the time for the polymerase to traverse a region was unexpectedly long, this indicated the polymerase actually traversed a longer region than was expected based on the reference sequence. The two distinct off-diagonal deviations (upper right hand corner) with higher slope indicated that an insertion had occurred in the reference sequence, and this was verified by comparison to the known fosmid sequence.

[00244] Figure 11 shows the average time it took the polymerase to traverse the template. For each mapped read, starting and ending times and positions were determined and used to compute the distance traversed by the polymerase between sequence reads. Based on these determinations, an average time across any particular region of the human reference sequence was computed. Regions that were traversed by the polymerase more slowly have peaks of higher ΔT, and were indicative of insertions in the template relative to the *Homo sapiens* chromosome 15 reference sequence. The insertions identified were the same insertions identified above.

[00245] Intermittent illumination-based sequencing reactions across fosmid sequence AC223433 showed significant sequence read coverage across the insertion events. The distribution of the physical coverage is shown in Figure 12, which illustrates examples of three-look strobes (i.e., sequencing reactions having three detection/illuminated periods) that span or intersect the

insertion events. Figure 12A shows the mapping of the strobe sequence reads to the *Homo sapiens* chromosome 15 reference sequence, where the sequence reads generated from the insert sequences in the template are excluded. Arrows indicate the locations of the insertions. Figure 12B shows a similar mapping with the sequence reads generated from the insert sequences indicated with brackets. A number of sequence reads flank the insertions, connect the two insertions, or clarify sequence within (or at the boundaries of) the insertion sequences. Such flanking and connecting sequence reads are useful for predicting and detecting genomic events, anchoring them to genomic references, and scaffolding for de novo assembly of novel sequences. In particular, there are 30 and 38 "3-look" reads that intersect the two regions of insertion of (1192 bp and 6879 bp, respectively). These sequence reads facilitated mapping of the insertions to the human reference sequence, which would have been extremely difficult, if not impossible, with commercially available short-read sequencing technologies. Further the sequence of the smaller insertion was a highly repetitive sequence, which would also have made mapping difficult with certain short-read technologies.

[00246] Figure 13 illustrates the sequence coverage obtained across the fosmid sequence, showing all two-, three-, and four-look strobe sequence reads spanning the sequence that are mappable to the known AC223433 fosmid sequence.

A consensus sequence was derived from the set of mappable sequence reads generated in these sequencing reactions. Strobe sequence reads were combined with sequence reads generated under constant illumination and assembled based on the human reference sequence (Hg18). High quality reads surrounding the (suspected) insertion sites, as well as high quality reads that did not map to the reference sequence, were extracted and assembled with a "de novo" greedy suffix tree assembler; the resulting contigs were mapped to the Hg18 reference sequence. Contigs spanning the (suspected) insertion sites were identified and fed back into the "de novo" assembler, and the resulting contigs were manually edited using standard techniques and placed back into the derived reference guided assembly. The final consensus sequence was a hybrid of a reference guided assembly and attempts at de novo assembly of novel insert sequences. Alignments to reference sequences were performed and plotted. Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described above and the fosmid reference sequence, and this plot confirmed a high degree of alignment between the two sequences. This dot plot was generated using Gepard 1.21 ("GEnome PAir – Rapid Dotter," available from the Munich Information Center for Protein Sequences (MIPS)) with a word size of 7. Nucleic acid dot plots are widely used in the art and are further described, e.g., in Krumsiek et al. (2007) Bioinformatics

23(8):1026-8; Maizel et al. (1981) Proc Natl Acad Sci USA 78:7665; Pustell, et al. (1982) Nucleic Acids Res 10:4765; and Quigley, et al. (1984) Nucleic Acids Res 12:347, all of which are incorporated herein by reference in their entireties for all purposes.

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation application of U.S. Patent Application No. 14/708,603, filed May 11, 2015, which is a continuation application of U.S. Patent Application No. 14/091,961, filed November 27, 2013, now U.S. Patent No. 9,057,102, which is a continuation application of U.S. Patent Application No. 12/982,029, filed December 30, 2010, now U.S. Patent No. 8,628,940, which (1) claims the benefit of U.S. Provisional Application No. 61/099,696, filed September 24, 2008; (2) claims the benefit of U.S. Provisional Application No. 61/139,402, filed December 19, 2008; and (3) is a continuation-in-part application of U.S. Patent Application No. 12/413,226, filed March 27, 2009, now U.S. Patent No. 8,143,030, the full disclosures of all of which are incorporated herein by reference in their entireties for all purposes.

[0002] This application is also related to U.S. Provisional Application No. 61/072,160, filed March 28, 2008, U.S. Patent Application No. 12/383,855, filed March 27, 2009, now U.S. Patent No. 8,236,499, and U.S. Patent Application No. 12/413,258, filed March 27, 2009, now U.S. Patent No. 8,153,375, all of which are incorporated herein by reference in their entireties for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0003] Not Applicable.

BACKGROUND OF THE INVENTION

[0004] The use of optically detectable labeling groups, and particularly those groups having high quantum yields, e.g., fluorescent or chemiluminescent groups, is ubiquitous throughout the fields of analytical chemistry, biochemistry, and biology. In particular, by providing a highly visible signal associated with a given reaction, one can better monitor that reaction as well as any potential effectors of that reaction. Such analyses are the basic tools of life science research in genomics, diagnostics, pharmaceutical research, and related fields.

[0005] Such analyses have generally been performed under conditions where the amounts of reactants are present far in excess of what is required for the reaction in question. The result of this excess is to provide ample detectability, as well as to compensate for any damage caused by the detection system and allow for signal detection with minimal impact on the reactants. For example, analyses based on fluorescent labeling groups generally require the use of an excitation radiation

source directed at the reaction mixture to excite the fluorescent labeling group, which is then separately detectable. However, one drawback to the use of optically detectable labeling groups is that prolonged exposure of chemical and biochemical reactants to such light sources, alone, or when in the presence of other components, e.g., the fluorescent groups, can damage such reactants. The traditional solution to this drawback is to have the reactants present so far in excess that the number of undamaged reactant molecules far outnumbers the damaged reactant molecules, thus minimizing or negating the effects of the photo-induced damage.

[0006] A variety of analytical techniques currently being explored deviate from the traditional techniques. In particular, many reactions are based on increasingly smaller amounts of reagents, e.g., in microfluidic or nanofluidic reaction vessels or channels, or in "single molecule" analyses. Such low reactant volumes are increasingly important in many high throughput applications, such as microarrays. The use of smaller reactant volumes offers challenges to the use of optical detection systems. When smaller reactant volumes are used, damage to reactants, such as from exposure to light sources for fluorescent detection, can become problematic and have a dramatic impact on the operation of a given analysis. In other cases, other reaction conditions may impact the processivity, rate, fidelity, or duration of the reaction, including salt or buffer conditions, pH, temperature, or even immobilization of reaction components within observable reaction regions. In many cases, the effects of these different reaction or environmental conditions can degrade the performance of the system over time. This can be particularly detrimental, for example, in real-time analysis of reactions that include fluorescent reagents that can expose multiple different reactions components to optical energy. In addition, smaller reactant volumes can lead to limitations in the amount of signal generated upon application of optical energy.

[0007] Further, in the case of sequencing-by-synthesis applications, an additional challenge has been to develop ways to effectively sequence noncontiguous portions of a template nucleic acid on a single molecule. This challenge is exacerbated in template nucleic acids that contain highly repetitive sequence and/or are hundreds or thousands of nucleotides in length, such as certain genomic DNA fragments. The difficulty in generating such noncontiguous reads from a single template has hampered efforts to construct consensus sequences for long templates, for example, in genome sequencing projects.

[0008] As such, methods and systems that result in enhanced reaction performance, such as an increase in processivity, rate, fidelity, or duration of a reaction of interest, would provide useful improvements to the methods and compositions currently available. For example, methods, devices,

and systems that increase reaction performance by, e.g., mitigating to some extent photo-induced damage in a reaction of interest and/or increasing various other performance metrics for the reaction would be particularly useful.

BRIEF SUMMARY OF THE INVENTION

[0009] In a general sense, the methods provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0010] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one

reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion. The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0011] In certain aspects, the present invention is generally directed to methods, devices, and systems for enhancing the performance of illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. In certain preferred embodiments, illuminated reactions comprise one or more fluorescent or fluorogenic reactants. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In some aspects, enhancing the performance of an illuminated reaction means increasing the processivity, rate, fidelity, and/or duration of the reaction. For example, enhancing the performance of an illuminated reaction can involve reducing or limiting the effects of photo-induced damage during the reaction. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction.

[0012] In certain aspects, methods of the invention useful for characterizing an analytical reaction comprise preparing a reaction mixture and initiating the analytical reaction therein, subjecting the reaction mixture to at least one detection period and at least one non-detection period during the course of the analytical reaction, collecting data during both the detection period(s) and the non-detection period(s), and combining the collected data to characterize the analytical reaction. In certain embodiments, the analytical reaction comprises an enzyme that exhibits an improvement in performance as compared to its performance in the analytical reaction under constant illumination, and such improvement may be related to various aspects of enzyme activity, e.g., processivity, fidelity, rate, duration of the analytical reaction, and the like. In certain embodiments, stop or pause points are used to control the activity of the enzyme, and such stop or pause points may comprise elements such as large photolabile groups, strand-binding moieties, non-native bases, and others well known in the art. In certain preferred embodiments, the one or more detection periods are illuminated periods and the one or more non-detection periods are non-illuminated periods. In certain preferred embodiments, a plurality of analytical reactions disposed on a solid support are characterized, preferably in a coordinated fashion as described elsewhere herein.

[0013] In certain preferred embodiments, the analytical reaction is a sequencing reaction that generates sequence reads from a single nucleic acid template during the detection period(s) but not during the non-detection period(s). For example, the analytical reaction can comprise at least two or more detection periods and can generate a plurality of noncontiguous reads from the single nucleic acid template. In some embodiments, the single nucleic acid template is at least 100 bases in length and/or comprises multiple repeat sequences. In certain embodiments, the sequencing reaction comprises passage of the single nucleic acid template through a nanopore, and in other embodiments the sequencing reaction comprises primer extension by a polymerase enzyme.

[0014] The analytical may optionally be a processive reaction monitored in real time, i.e., during the course of the processive reaction. In preferred embodiments, such a processive reaction is carried out by a processive enzyme that can repetitively execute its catalytic function, thereby completing multiple sequential steps of the reaction. For example, a processive polymerization reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.

[0015] In certain preferred embodiments, the analytical reaction comprises at least one component comprising a detectable label, e.g., a fluorescently labeled nucleotide. In certain embodiments, the labeled component is present throughout the course of the analytical reaction, i.e., during both the detection and the non-detection periods. The method may further comprise an optical system to collect the data during the detection period, but optionally not to collect the data during the non-detection period.

[0016] In certain aspects, methods of the invention comprise providing a substrate having a reaction mixture disposed thereon and illuminating the reaction mixture on the substrate with an excitation illumination for multiple, noncontiguous periods during the course of the reaction, thereby subjecting the reaction mixture to intermittent excitation illumination. In some embodiments, the reaction mixture comprises first reactant and a second reactant, wherein an amount of photo-induced damage to the first reactant occurs as a result of interaction between the first reactant and the second reactant under excitation illumination. In certain embodiments, the method further comprises monitoring a reaction between the first and second reactants during illumination and collecting the data generated therefrom. In some embodiments, the reaction is a primer extension reaction and/or the first reactant is a polymerase enzyme. In certain embodiments, the second reactant is a fluorogenic or fluorescent molecule.

[0017] In yet another aspect, the methods are useful for mitigating photo-induced damage in an illuminated reaction by subjecting the illuminated reaction to intermittent illumination rather than constant illumination. For example, certain methods of the invention monitor a reaction mixture comprising at least one enzyme and a fluorescent or fluorogenic substrate for the enzyme, wherein interaction of the enzyme and the substrate under excitation illumination can result in altered activity of the enzyme, e.g if such excitation illumination is present over an extended period of time. Such methods can comprise directing intermittent excitation illumination at a first observation region for a first period that is less than a photo-induced damage threshold period under the intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions. As such, certain aspects of the invention lengthen a photo-induced damage threshold period for an analytical reaction through intermittent inactivation of the excitation illumination source since the photo-induced damage threshold period under intermittent illumination is longer than the photo-induced damage threshold period under constant illumination.

[0018] In a related aspect, the invention also provides methods of performing an enzyme reaction, comprising providing an enzyme within a first observation region, contacting the enzyme with a fluorescent or fluorogenic substrate for the enzyme, and directing an excitation radiation at and detecting signals from the first observation region for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0019] In further aspects, the invention provides methods of monitoring a primer extension reaction, comprising providing a polymerase enzyme within a first observation region, contacting the polymerase with at least a first fluorescent or fluorogenic nucleotide analog, and monitoring a fluorescent signal emitted from the first observation region in response to illumination with excitation radiation for a period that is less than a photo-induced damage threshold period under intermittent illumination conditions, but that is greater than a photo-induced damage threshold period under constant illumination conditions.

[0020] In addition, the invention provides methods for generating a plurality of noncontiguous sequence reads from a single nucleic acid template molecule. Such methods generally comprise preparing a reaction mixture comprising the template molecule, a polymerase enzyme, and a set of differentially labeled nucleotides or nucleotide analogs, wherein the set comprises at least one type of nucleotide or nucleotide analog for each of the natural nucleobases (A, T, C, and G). The polymerization reaction is initiated, the polymerase begins processive incorporation of the labeled nucleotides or nucleotide analogs into a nascent nucleic acid strand, and during such incorporation the reaction is monitored by optical means to detect incorporation events, thereby generating a first sequence read. In a subsequent step, the labeled nucleotides or analogs are replaced with unlabeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed without detecting incorporation events. Subsequently, the unlabeled nucleotides or analogs are replaced with labeled nucleotides or nucleotide analogs and the polymerization is allowed to proceed once again with real time detection of incorporation events, thereby generating a second sequence read that is noncontiguous to the first sequence read. The substitution of labeled for unlabeled, and unlabeled for labeled, nucleotides and nucleotide analogs can be repeated multiple times to generate a plurality of noncontiguous sequence reads, each of the plurality generated during a period when the labeled nucleotides or nucleotide analogs are being incorporated into the nascent strand and such incorporation is being detected in real time.

[0021] In certain aspects, devices of the invention can comprise a solid support (e.g., substrate) having an observation region, a first reactant immobilized within the observation region, and a second reactant disposed within the observation region, and a means for subjecting the observation region to at least one illuminated period and at least one non-illuminated period. In certain embodiments, interaction between the first and second reactants under excitation illumination causes photo-induced damage to the first reactant, and further wherein the photo-induced damage is reduced by subjecting the observation region to intermittent illumination. In some embodiments, the first reactant is an enzyme (e.g., a polymerase), the second reactant (e.g., a nucleotide) has a detectable label (e.g., fluorescent label), and/or the observation region is within a zero-mode waveguide. The means for subjecting the observation region to one or more illuminated and non-illuminated periods may comprise, e.g., a laser, laser diode, light-emitting diode, ultra-violet light bulb, white light source, a mask, a diffraction grating, an arrayed waveguide grating, an optic fiber, an optical switch, a mirror, a lens, a collimator, an optical attenuator, a filter, a prism, a planar waveguide, a wave-plate, a delay line, a movable support coupled with the substrate, and a movable illumination source, and the like. The device may further comprise a means for collecting the data during the illuminated period(s), such as an optical train, e.g., operably coupled to a machine comprising machine-readable medium onto which such data may be written and stored.

[0022] In further aspects, the invention provides systems for performing intermittent detection of an analytical reaction comprising reagents for the analytical reaction disposed on a solid support, a mounting stage configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the solid support detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other, and a data processing system operably coupled to the optical train. In certain preferred embodiments, the analytical reaction is a sequencing reaction and/or the solid support comprises at least one zero-mode waveguide.

[0023] In still other aspects, the invention provides systems for analyzing an illuminated reaction that is susceptible to photo-induced damage when illuminated for a period longer than an photo-induced damage threshold period, comprising a solid support having reagents for the reaction disposed thereon, a mounting stage supporting the solid support and configured to receive the solid support, an optical train positioned to be in optical communication with at least a portion of the

solid support to illuminate the portion of the solid support and detect signals emanating therefrom, a means for subjecting the portion of the solid support to at least one detection period and at least one non-detection period, and a translation system operably coupled to the mounting stage or the optical train for moving one of the optical train and the solid support relative to the other. In some embodiments, the illuminated reaction is a sequencing reaction, e.g., a nucleotide sequencing-by-synthesis reaction. In certain embodiments, the solid support comprises at least one optical confinement, e.g., a zero-mode waveguide.

[0024] The invention provides methods of performing analytical reactions, e.g., processive analytical reactions, that include preparing a reaction mixture comprising reaction components, at least one of which is a detectable component that is detectable during one or more detection periods, and at least one of which is a clocking component that is detectable during one or more non-detection periods during the analytical reaction. The methods further comprise initiation the analytical reaction and maintaining conditions that allow the analytical reaction to proceed while subjecting it to at least one detection period and at least one non-detection period, both in the presence of the clocking component and the detectable component. In certain embodiments, the detectable component emits a detectable signal is response to excitation illumination during the detection period, but not during the non-detection period when a clocking signal is emitted from the clocking component. The detectable signal is collected during the detection period and the clocking signal is detected during the non-detection period, e.g., using an optical system. Optionally, the clocking signal can also be collected during the detection period and the non-detection period. In certain preferred embodiments, detection data is collected in read time during the detection period, non-detection data is collected in real time during the non-detection period, and the detection data and non-detection data are both used to characterize the analytical reaction. In some embodiments, the transition between the detection period and the non-detection period does not involve substitution and/or addition of reaction components during progression of the analytical reaction, and in other embodiments the transition does involve substitution and/or addition of reaction components, e.g., via a reaction mixture exchange. In some preferred embodiments, a plurality of analytical reactions are disposed on a solid support, subjected to intermittent illumination, monitored to collect data, and characterized based upon the data so collected.

[0025] The detectable component and clocking component are typically linked to discrete molecules in the analytical reaction. For example, the detectable component can be linked to a first subset of nucleotide analogs and the clocking component can be linked to a second subset of

nucleotide analogs in the analytical reaction mixture. Alternatively, both the detectable component and the clocking component can be linked to a single molecule, e.g., a single nucleotide or nucleotide analog, in the analytical reaction. The detectable component and clocking component can both comprise detectable labels (e.g., luminescent, fluorescent, or fluorogenic labels, including, e.g., quantum dots), and in some embodiments, different detectable labels, e.g. having different absorption peaks.

[0026] In certain preferred embodiments, an analytical reaction performed according to the invention comprises at least one enzyme, e.g., a polymerase, ligase, ribosome, nuclease, and/or kinase. In some embodiments, pause or stop points are engineered into the analytical reaction to control activity of the enzyme. Various aspects of the analytical reaction can be changed by being subjected to at least one detection period and at least one non-detection period, such aspects including but not limited to processivity, fidelity, rate, and duration, e.g. of enzyme activity.

[0027] In certain preferred embodiments, the analytical reaction is a sequencing reaction comprising a single nucleic acid template that generates sequence reads during the detection period by detecting the detectable component, and does not generate sequence reads during the non-detection period by suspending detection of the detectable component. Such a sequencing reaction typically comprises at least two or three detection periods and generates a plurality of noncontiguous sequence reads from the single nucleic acid template. In some embodiments, the template comprises multiple repeat or complementary sequences. In some embodiments, the sequencing reaction comprises passage of the single nucleic acid or a nascent strand complementary thereto through a nanopore. In some preferred embodiments, the sequencing reaction comprises primer extension by a polymerase enzyme and the detectable component is linked to a nucleotide or nucleotide analog. In some embodiments, the clocking component is linked to the polymerase enzyme, and optionally can be a multi-component label, e.g, a FRET label.

[0028] In certain aspects, the invention provides methods of mitigating photo-induced damage during an illuminated reaction that include preparing a reaction mixture having first and second reactants, where interaction of the reactants under excitation illumination can cause photo-induced damage to the first reactant. The illuminated reaction is subjected to intermittent excitation illumination characterized by periods of maximal illumination followed by periods of modified but not absent illumination. The intermittent excitation illumination reduces the amount of photo-induced damage to the first reactant during the illuminated reaction as compared to the illuminated reaction under constant maximal excitation illumination, thereby mitigating photo-induced damage

to the first reactant. In certain preferred embodiments, the illuminated reaction is a primer extension reaction. In certain preferred embodiments, the first reactant is an enzyme, e.g., a polymerase or ligase enzyme. In certain preferred embodiments, the second reactant comprises a fluorescent or fluorogenic molecule. In certain embodiments, the modified excitation illumination is illumination with a lower intensity excitation illumination than the maximal excitation illumination. In certain embodiments, a set of illumination sources provides the maximal excitation illumination and a subset of the set of illumination sources provides the modified excitation illumination.

[0029] In other aspects, the invention provides a method of sequencing a template nucleic acid that includes subjecting the template to methylation to generate at least one methylated base, subjecting the methylated base to base excision to generate at least one abasic site in the template, annealing a primer to the template nucleic acid, contacting the template with a polymerase enzyme to promote extension of the primer in a template-dependent manner, monitoring the extension of the primer in real time to generate a nucleotide sequence read complementary to the template, extending the primer until the abasic site is encountered by the polymerase, at which time the polymerase pauses on the template, and reinitiating primer extension by facilitating abasic site bypass by the polymerase. The monitoring, extending, and reinitiating steps are repeated until a desired number of nucleotide sequence reads is generated and collected, and subsequently analyzed to determine the sequence of the template nucleic acid. In certain embodiments, the contacting step occurs during a detection period or a detection period immediately follows the contacting step. In certain embodiments, a detection period ends and a non-detection period begins prior to one or more pauses of the polymerase on the template. In certain embodiments, a non-detection period is terminated simultaneous with or immediately following one or more reinitiation steps. In some embodiments, the reinitiating step comprises introduction of a pyrene to the polymerase, where the polymerase incorporates the pyrene into the nascent strand opposite and, therefore, "pairing with" an abasic site in the template. In certain preferred embodiments, the template is circular and the polymerase pauses at the same abasic site multiple times during the primer extension reaction. In other embodiments, the method further comprises terminating the monitoring when a desired length of the nucleotide sequence read is collected, e.g., by removing or modifying excitation illumination. Optionally, the desired length can be less than a length of the template nucleic acid. Additionally, the monitoring can be reinitiated subsequent to or simultaneous with the reinitiating of primer extension.

[0030] In yet further aspects, the invention provides a method of performing an illuminated reaction that includes preparing a reaction mixture comprising multiple optically detectable components that are distinguishable from one another based upon their individual signal emissions, initiating the illuminated reaction, and maintaining conditions that allow the illuminated reaction to proceed while subjecting the reaction mixture to at least one maximal illuminated period and at least one modified illuminated period during the illuminated reaction. In preferred embodiments, at least a portion of the optically detectable components are detectable during both the maximal and modified illuminated periods. In certain embodiments, the maximal illuminated period is characterized by a first excitation radiation intensity and the modified illuminated period is characterized by a second excitation radiation intensity that is less than the first excitation radiation intensity. In certain preferred embodiments, all of the optically detectable components are detectable during both the maximal and modified illuminated periods, but are distinguishable from one another during the maximal illuminated period, but are not distinguishable during the modified illuminated period. In certain embodiments, the maximal illuminated period comprises exposing the reaction mixture to a set of excitation radiation wavelengths and the modified illuminated period comprises exposing the reaction mixture to a subset of the set of excitation radiation wavelengths. In certain preferred embodiments, all of the optically detectable components are detectable and distinguishable during the maximal illuminated period, but only a subset of the optically detectable components are detectable during the modified illuminated period.

[0031] In some embodiments, the illuminated reaction is initiated during a modified illuminated period and subsequently subjected to a maximal illuminated period, where data collected during the modified illuminated period is used in the statistical analysis of data collected during the maximal illuminated period. For example, an illuminated reaction that is a polynucleotide sequencing reaction can generate sequence read data during a modified illuminated period that is subsequently used to construct a sequence scaffold for assembly of sequence read data collected during a maximal illuminated period. Additionally or optionally, the illuminated reaction is a template-directed sequencing reaction and sequence read data collected during a modified illuminated period is used to determine a rate of translocation of a polymerase during the modified illuminated period.

[0032] Some embodiments of the invention comprise performing a plurality of illuminated reactions, each of which is exposed to the set of excitation radiation wavelengths during the maximal illuminated period, but is exposed to a different subset of the set of excitation radiation

wavelengths during the modified illuminated period, such that a distinct subset of optically detectable components are detectable during the modified illuminated period for each of the plurality of illuminated reactions. In other words, for two such illuminated reactions, although all optically detectable components are detectable during their respective maximal illuminated periods, only a subset of the optically detectable components is detectable in each reaction, and the subset detectable in the first reaction is preferably different from the subset detectable in the second reaction.

[0033] In certain aspects, the invention provides methods for performing paired-end sequencing on a single template molecule. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion. A first linker ligated to the first terminal portion of the nucleic acid molecule connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and a second linker ligated to the second terminal portion of the nucleic acid molecule connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion. A template nucleic acid molecule is thereby formed comprising the double-stranded nucleic acid molecule with both the first linker and the second linker ligated thereto. The template molecule is subjected to a sequencing process in which sequence reads are generated for the first terminal portion and the second terminal portion, but sequence reads are not generated for the intermediate portion, even if the intermediate portion is processed during the sequencing process, e.g., by a polymerase. In some embodiments, the first linker and second linker are identical, and in other embodiments they are different from one another, i.e., not identical. In certain embodiments, the first and second linkers comprise complementary regions and can be hybridized to one another prior to one or both of the ligating steps. In some cases, hybridized linkers that are ligated to the ends of a double-stranded nucleic acid molecule are separated prior to subjecting the molecule to a sequencing reaction, and in some cases the hybridized linkers remain hybridized during at least a portion of the sequencing reaction. For example, in a template-directed sequencing reaction, a polymerase capable of strand displacement separates the hybridized linkers as it sequences the template. In certain preferred embodiments, the sequencing process comprises at least one detection period (e.g., an illuminated period) and at least one non-detection period (e.g., a non-illuminated period) such that the intermediate portion of the template molecule is subjected to the sequencing process during the non-detection period. In some embodiments, the template is fragmented after ligation to remove the intermediate portion. The

sequencing process can generate redundant sequence data from one or both of the first terminal portion and the second terminal portion, and/or can generate sequence data from an additional portion of the template molecule that is noncontiguous with the first terminal portion and the second terminal portion. In preferred embodiments, the sequencing process involves circularizing the template molecule by separating the complementary strands of the template molecule and using the complementary strands in template-directed nascent strand synthesis catalyzed by a single polymerase enzyme. Optionally, the template molecule can comprise a primer binding site, a registration sequence, and/or a synthesis blocking moiety. The primer binding site, a registration sequence, or synthesis blocking moiety can be present in one or both of the linkers, or can be located elsewhere within the template molecule. In some cases, the synthesis blocking moiety is selected from the group consisting of an abasic site, a nick, a synthetic linker, a non-native nucleotide or analog thereof, a primer, a large photolabile group, a strand-binding moiety, a damaged base, and a modified base. The synthesis blocking moiety can permanently or temporarily block progression of the sequencing process, e.g., by interfering with the activity of an enzyme, e.g., a polymerase enzyme. In certain preferred embodiments, the synthesis blocking moiety is an abasic site, e.g., introduced by a DNA glycosylase.

[0034] In some aspects, the invention provides methods for generating a nucleic acid construct for analytical reactions. In certain embodiments, such a method comprises providing a double-stranded nucleic acid molecule comprising a first terminal portion, an intermediate portion, and a second terminal portion; providing a first stem-loop linker hybridized to a second stem-loop linker; ligating the first stem-loop linker to the first terminal portion of the nucleic acid molecule, wherein the first stem-loop linker connects the 3' terminus at the first terminal portion with the 5' terminus at the first terminal portion; and ligating the second stem-loop linker to the second terminal portion of the nucleic acid molecule, wherein the second stem-loop linker connects the 3' terminus at the second terminal portion with the 5' terminus at the second terminal portion, thereby generating the nucleic acid construct. Optionally, the nucleic acid construct can be subjected to fragmentation after the ligating of steps c and d, wherein the fragmentation removes the intermediate portion from the nucleic acid construct and introduces two double-stranded termini. The method can further include ligating the two double-stranded termini to one another. In some embodiments, one of the stem-loop linkers comprises a primer binding site, registration sequence, or a synthesis blocking moiety that is absent from the other stem-loop linker.

[0035] In further aspects, the invention includes a single template nucleic acid molecule comprising a duplex region; a first linker linking termini at a first end of the duplex region; a second linker linking termini at a second end of the duplex region, wherein a region of the first linker is complementary to a region of the second linkers. Optionally, the single template molecule comprises the first and second linkers hybridized with one another. In some embodiments, the duplex region is separated or melted apart to transform the single template nucleic acid molecule into a topologically single-stranded, circular nucleic acid molecule. Further, the invention provides a composition comprising a single, optically resolvable polymerase enzyme in association with a single-stranded circular nucleic acid molecule, wherein the single-stranded circular nucleic acid molecule comprises first, second, third, and fourth regions, and further wherein the first region is complementary to the second region, and the third region is complementary to the fourth region, and further wherein the regions are ordered on the single-stranded circular nucleic acid molecule as follows: first region, third region, second region, fourth region.

[0036] In still further aspects of the invention, machine-implemented methods for transforming nucleotide sequence read data into consensus sequence data, wherein the nucleotide sequence read data is generated by sequencing a target region of a template nucleic acid multiple times, and the consensus sequence data is representative of a most likely actual sequence of the template nucleic acid. Such machine-implemented methods can comprise various steps, such as a) mapping the nucleotide sequence data to a target sequence using a local alignment method that produces a set of local alignments comprising an optimal local alignment and sub-optimal local alignments, b) enumerating the set of local alignments, c) constructing a weighted directed graph wherein each local alignment in the set of local alignments is represented as a node, thereby generating a set of nodes in the weighted directed graph, d) drawing edges between pairs of nodes in the weighted directed graph if the pair represents a potential reconstruction of the template nucleic acid, e) assigning weights to the edges drawn in step d, wherein a given weight for a given edge represents the log-likelihood that a given pair of nodes connected by the given edge is truly a reconstruction of the template nucleic acid, f) finding the shortest path to each node in the weighted directed graph, thereby generating a set of shortest paths for the weighted directed graph, g) ranking the set of shortest paths to determine the best assignment, and h) storing the results of steps a-g on a machine-readable medium. In certain embodiments, the steps of the machine implemented methods are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Also provided are

computer program products comprising a computer usable medium having computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the machine-implemented methods of the invention, and machine-readable medium on which the results of the method steps are stored. The invention further includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods.

[0037] In certain aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme (e.g., a polymerase) during a time t, where time t occurs during a non-detection period. Such a method comprises, in certain embodiments, developing a probability model *p(v)* to describe an observed distribution of enzyme velocities during one or more detection periods; sampling velocities from *p(v)*; summing and recording the velocities sampled in step b to produce a sum that is an estimate of x/τcorr; and repeating the sampling, summing, and recording M times to generate a distribution of sums that are estimates of x/τcorr, with the distribution of sums being the distribution of the distance x travelled by an enzyme during a time t. Preferably, at least some of the steps are performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. Optionally, the enzyme is a polymerase enzyme. In some embodiments, multiple enzymes are observed simultaneously and the probability model *p(v)* is determined independently for each of the multiple enzymes. In certain preferred embodiments,

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}.$$

[0038] In further aspects, the invention provides machine-implemented methods for transforming enzyme velocity data from one or more detection periods into a distribution of the distance x travelled by an enzyme during a time t, where time t occurs during a non-detection period. In some embodiments, the method comprises estimating a distribution of local rates p(v), making independent identically distributed draws of N = t/τcorr velocities from from p(v); summing the velocities; recording the velocities summed in c) as an estimate of x/τcorr; and repeating b-d M times, e.g., where M is preferably at least 1000. Optionally, p(v) is determined using a Hidden Markov Model or the autocorrelation function $\langle \delta v(t)\delta v(t+\Delta) \rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$. The invention further

includes a computer program product comprising a computer usable medium having a computer readable program code embodied therein, said computer readable program code adapted to be executed to implement the above methods, as well as a machine-readable medium on which the results of the steps of the methods are stored.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions, whether illumination is initiated before (left) or after (right) initiation of the reaction.

[0040] Figures 2A-2B provide an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination, including depictions of multiple reactions arrayed on a solid support (Figure 2A) and prophetic data (Figure 2B) from certain embodiments of the invention.

[0041] Figures 3A–3C provide an exemplary embodiment of analysis of a plurality of illuminated reactions on a solid support (Figure 3A) using intermittent illumination and a mask (Figure 3B). A graph (Figure 3C) depicts prophetic data from certain embodiments of the invention.

[0042] Figures 4A–4B provide additional embodiments of masks for use in the methods of the invention, including a mask that allows illumination of columns of reactions (Figure 4A) and a mask that allows illumination of every other reaction in a row and column (Figure 4B).

[0043] Figures 5A–5D illustrate an aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Figure 5A illustrates a solid support comprising four quadrants, each quadrant containing a different sample. Figure 5B illustrates a mask design for selective illumination of the substrate. Figures 5C and 5D demonstrate various positions of the mask on the solid support.

[0044] Figure 6 provides an illustration of paths in a sequence alignment matrix representing sequencing data from a SMRTbell™ template.

[0045] Figure 7 illustrates a hypothetical directed graph.

[0046] Figures 8A–8C provide data from single-molecule sequencing-by-synthesis reactions. Figure 8A provides data from a two-minute interval beginning at initiation of the reactions, i.e., from 0-120 seconds. Figure 8B provides data from a second two-minute interval from 300-420 seconds. Figure 8C provides data from a third two-minute interval from 600-720 seconds.

[0047] Figure 9 schematically illustrates one embodiment of a system for use with the methods, devices, and systems of the invention.

[0048] Figure 10 provides a graphical representation of rates of polymerase activity on different portions of a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0049] Figure 11 provides a graphical representation of the average rate of polymerase translocation over a template nucleic acid during a sequencing reaction utilizing intermittent illumination.

[0050] Figures 12A–12B provide a distribution of the physical coverage of a template nucleic acid achieved during a sequencing reaction utilizing intermittent illumination, with Figure 12A showing mapping to a reference sequence with sequence reads (and portions thereof) that do not map to the reference excluded and Figure 12B showing a similar mapping that further includes sequence reads corresponding to insertions in the template that are absent from the reference sequence.

[0051] Figure 13 provides a distribution of the physical coverage provided by sequence reads generated during sequencing reactions utilizing intermittent illumination across an approximately 40 kb template nucleic acid.

[0052] Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described herein and a reference sequence.

[0053] Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system.

[0054] Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters $P(S \rightarrow P) = 1/24$; $P(P \rightarrow S) = 1/11$; and $p(v) \sim \text{Gamma}(48, 0.25)$. Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a non-detection period.

[0055] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period.

[0056] Figure 18 provides an exemplary strategy for selectively reducing the size of a duplex fragment within a SMRTbell™ template.

[0057] Figures 19A–19C provide an illustrative example of nucleic acid templates having nicks.

[0058] Figures 20A and 20B illustrate two exemplary embodiments of template constructs used in the present invention.

[0059] Figures 21A and 21B schematically illustrate redundant or consensus sequencing using the constructs shown in Figures 20A and 20B.

DETAILED DESCRIPTION OF THE INVENTION

[0060] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing devices, formulations and methodologies which are described in the publication and which might be used in connection with the presently described invention.

[0061] Note that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polymerase" refers to one agent or mixtures of such agents, and reference to "the method" includes reference to equivalent steps and methods known to those skilled in the art, and so forth. Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either both of those included limits are also included in the invention.

[0062] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features and procedures well known to those skilled in the art have not been described in order to avoid obscuring the invention. Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

I. General

[0063] In a general sense, the methods, devices, and systems provided herein implement intermittent detection of analytical reactions as a means to collect reliable data from times during the reaction that are less or not able to be analyzed if detection is constant throughout the reaction. In particular, certain detection methods can cause damage to reaction components, and such intermittent detection allows the damage to be avoided or at least delayed, thereby facilitating detection of the reaction at later stages. For example, if a detection method causes a reduction in processivity of a polymerase enzyme, then intermittent detection would allow data collection at noncontiguous regions of a template nucleic acid that extend farther from the initial binding site of the polymerase on the template than would be achievable under constant detection. Further, some detection methods have limits on how much data or for how long a time data may be generated in a single reaction, and intermittent detection of such a reaction can allow this data to be collected from various stages of a reaction, thereby increasing the flexibility of the investigator to spread out the data collection over multiple stages of a reaction. In certain aspects, the present invention is particularly suitable to characterization of analytical reactions in real time, that is, during the course of the reaction. In certain aspects, the present invention is particularly suitable to characterization of single molecules or molecular complexes monitored in analytical reactions, for example, single enzymes, nucleotides, polynucleotides, and complexes thereof.

[0064] In certain aspects, the present invention is directed to methods, devices, and systems for obtaining sequence data from discontiguous portions of single nucleic acid templates. The methods generally comprise providing a monitorable sequencing reaction comprising a polymerase, template, and primer sequence, as well as the various types of nucleotides or nucleotide analogs that are to be incorporated by the polymerase enzyme in the template-directed primer extension reaction. Typically, at least one or more or all of the nucleotides or nucleotide analogs are embodied with a detectable property that permits their identification upon or following incorporation. In the context of the present invention, the sequence data for a first portion of a template nucleic acid is acquired during a first stage of the reaction under a first set of reaction conditions that includes at least one reaction condition that results in degraded performance of the reaction, but that may contribute to the detectability of the nucleotides being incorporated. During a second stage of the reaction, the degradative influence is eliminated or reduced, which may result in an inability or a reduced ability to obtain sequence data from a second portion of the template nucleic acid, but where the second portion of the template nucleic acid is contiguous with the first portion. Subsequently, the reaction condition resulting in degraded performance is reinstated and sequence data is obtained for a third

portion of the template nucleic acid during a third stage of the reaction, but where the third portion of the sequence is not contiguous with the first portion of the sequence, but is contiguous with the second portion.

[0065] The elimination or reduction of the degradative influence during the second stage of the reaction may be accomplished by changing or shortening one or more reaction conditions underlying degradative reaction performance, e.g., by changing one or more reaction conditions (e.g., temperature, pH, exposure to radiation, physical manipulation, etc.), and in particular may involve altering a reaction condition related to detection of one or more aspects or products of the reaction. For example, such an alteration in reaction conditions during the second stage may result in an increase in reaction rates, e.g., speeding up the progression of a template nucleic acid through a nanopore; or may reduce exposure of reaction components to harmful radiation or other reaction condition related to detection of the products of the reaction. However, in preferred embodiments, nucleotides or nucleotide analogs having the detectable property are present in the reaction mixture during all stages of the reaction, including stages in which the degradative influence is eliminated or reduced; as such, the reaction condition changed in stage two of such an embodiment would not comprise removal or dilution of such detectable nucleotides or nucleotide analogs.

[0066] "Intermittent detection," as used herein, generally refers to a means of monitoring a reaction that is carried out intermittently during the course of the reaction. Intermittent detection may refer to intermittent use of one or more monitoring methods, but does not necessarily mean that all means of monitoring a given reaction are intermittently halted. For example, monitoring of one or more nucleotide incorporations to generate nucleotide sequence reads may be intermittently halted while other aspects of a sequencing reaction are constantly monitored, e.g., temperature, reaction time, pH, etc. In certain embodiments, intermittent detection is achieved by intermittent or differential illumination of a given reaction, e.g., a reaction that uses an illumination system to detect reaction products and/or progression. Although various aspects of the invention are described herein in terms of embodiments using intermittent illumination, it should be understood that where applicable intermittent detection by other means (e.g., electrochemical, radiochemical, etc.) can be utilized in the methods of the invention. Likewise, a stage of a reaction during which an intermittent detection method is active may be referred to as a "detection period" and a stage of a reaction during which an intermittent detection method is inactive may be referred to as a "non-detection period." In illuminated reactions, such periods may also be referred to as "illuminated periods" and "non-illuminated periods," respectively, although it is to be understood that the term "non-

illuminated period" included periods in which illumination may be present but altered as compared to illumination during an "illuminated period." For example, a non-illuminated period may be characterized by a complete absence of illumination, or a modification of illumination, including but not limited to changes in wavelength, frequency, intensity, and/or number of illumination sources. Alternatively or additionally, reaction components that are excited by the illumination source(s) may be modified or removed from a reaction mixture to create a non-illuminated period. For example, a fluorescent dye detected during an illuminated period may be removed from the reaction mixture, e.g., by buffer exchange, thereby producing a non-illuminated period during which time the fluorescent dye cannot be detected even if the excitation illumination is present. In a further example, a non-illuminated period can indicate a period during an illuminated reaction during which a type of illumination-based detection that occurs during an illuminated period is not occurring, e.g., the identity of fluorescently labeled nucleotides incorporated into a nascent strand is not being detected or recorded.

[0067] In certain aspects, the present invention is generally directed to improved methods, devices, and systems for performing illuminated reactions. The term "illuminated reactions" as used herein refers to reactions which are exposed to an optical energy source. Typically, such illumination is provided in order to observe the generation and/or consumption of reactants or products that possess a particular optical characteristic indicative of their presence, such as a shift in the absorbance spectrum and/or emission spectrum of the reaction mixture or its components. In certain preferred embodiments, illuminated reactions comprise one or more fluorogenic or fluorescent components. In accordance with certain methods of the invention, such illuminated analyses are subjected to intermittent detection (e.g., data collection) for one or more aspects of the data typically collected for a given reaction. For example, aspects of the data typically collected for nucleotide sequencing reactions include nucleotide sequence data, read quality data, signal to background ratios, reaction rates and durations, measures of the fidelity of the reaction, reaction times, and the like. In certain preferred embodiments, nucleotide sequence data is iteratively collected during an ongoing sequencing reaction to generate nucleotide sequence reads for at least two or more noncontiguous regions of a template nucleic acid molecule. Such iterative sequence data acquisition may be achieved in various ways depending on the sequencing technology in use. For example, in sequencing methods that utilize luminescent components that generate a signal indicative of the identity of a base position, iterative sequence data collection may be achieved by removing or altering an illumination source (or a reaction relative to an illumination source),

substituting the luminescent components for unlabeled components that do not generate signal, or otherwise interrupting signal acquisition in the experimental system.

[0068] In certain preferred embodiments, such illuminated reactions are illuminated for an amount of time that permits the effective performance of the analysis. Traditionally, illuminated reactions are illuminated from initiation through completion, and the time during which reaction data may be reliably collected is dictated by the progression (as measured by, e.g., processivity, rate, fidelity, duration, etc.) of the reaction under constant illumination. Some reactions are sensitive to such constant illumination, which can reduce their performance (e.g., processivity), and thereby prevent collection of data from later stages of the reaction, i.e., stages that would otherwise occur if the reaction were carried out with no illumination. The present invention provides methods for performing illuminated reactions comprising subjecting the reactions to intermittent illumination. Such intermittent illumination can increase performance (e.g., processivity, rate, fidelity, duration, etc.) of the reactions, thereby allowing generation of data that cannot be collected under constant illumination, such as data from later stages of an ongoing reaction whose progression is compromised under constant illumination. For example, in sequencing-by-incorporation reactions the use of intermittent excitation illumination can increase processivity, which has the benefit of providing sequence reads more distal from the polymerase binding/initiation site than such reactions subjected to constant exposure to excitation illumination.

[0069] Further, it is an object of the instant invention to provide sequence data from noncontiguous regions of a nucleic acid template in a single reaction. Other commercially available platforms have attempted to achieve such noncontiguous sequence data through, e.g., complex cloning and sequencing strategies. The present invention provides a clear advantage over such strategies by providing a simple and economical solution that is applicable across various platforms, and is particularly applicable to illuminated, single-molecule sequencing-by-incorporation reactions.

[0070] In preferred embodiments, illuminated reactions for use with the instant invention are nucleic acid sequencing reactions, e.g., sequencing-by-incorporation reactions. In preferred embodiments, such an illuminated reaction analyzes a single molecule to generate nucleotide sequence data pertaining to that single molecule. For example, a single nucleic acid template may be subjected to a sequencing-by-incorporation reaction to generate one or more sequence reads corresponding to the nucleotide sequence of the nucleic acid template. For a detailed discussion of such single molecule sequencing, see, *e.g.*, U.S. Patent Nos. 6,056,661, 6,917,726, 7,033,764,

7,052,847, 7,056,676, 7,170,050, 7,361,466, 7,416,844; Published U.S. Patent Application Nos. 2007-0134128 and 2003/0044781; and M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, SCIENCE 299:682-686, January 2003 Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations, all of which are incorporated herein by reference in their entireties for all purposes. In some embodiments, a plurality of single nucleic acid templates are analyzed separately and often simultaneously to generate a plurality of sequence reads corresponding to the nucleotide sequences of the plurality of nucleic acid templates. In certain preferred embodiments, the plurality of nucleic acid templates includes at least two nucleic acid templates that comprise identical nucleotide sequences such that analysis of the two nucleic acid templates generates overlapping sequence reads. In certain preferred embodiments, at least one of the nucleic acid templates is configured to provide redundant sequence data in a single sequence read, e.g., via duplications, sense and antisense sequences, and/or circularization.

[0071] Certain aspects of the invention are directed to methods, devices, and systems for generating a sequence scaffold for a nucleic acid template, e.g., chromosome, genome, or portion thereof. A sequence scaffold as used herein refers to a set of sequence reads that extends across at least a portion of a nucleic acid template. In some embodiments, such a sequence scaffold is used to generate a consensus sequence for the nucleic acid template. In some embodiments, the nucleic acid template is very large, e.g., at least about 100, 1000, 10,000, 100,000, or more bases or base pairs in length. In some embodiments, the sequence scaffold and/or consensus sequence is based on at least 1-, 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of at least a portion of the nucleic acid template. In some preferred embodiments, the portion of the nucleic acid is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the entire length of the nucleic acid template.

[0072] In certain aspects, the invention is particularly suitable for sequencing nucleic acid templates interspersed with repetitive elements. Such repetitive elements present major logistical and computational difficulties for assembling fragments produced by sequencing strategies, especially those with read-lengths that are too short to encompass unique reads outside the repeat region. For example, the human T-cell receptor locus contains a five-fold repeat of a trypsinogen gene that is 4 kbp long and that varies 3 to 5% between copies. Therefore, a sequencing strategy that cannot provide nucleotide sequence information that spans at least 20 kb for a single molecule containing the locus will have difficulty providing consensus sequence for the locus. Further, Alu repeats (~300 bp retrotransposons) are also problematic because they cluster and can constitute up

to 50-60% of the template sequence, with copies varying from 5-15% between each other. The human genome contains an estimated one million Alu repeats and 200,000 LINE elements (average length ~1000 bp), representing roughly 10% and 5% of the entire genome, respectively. In certain embodiments, the present methods facilitate efficient and accurate sequence determination for long templates comprising such repetitive sequences, in part because the present methods do not rely solely on sequence overlap to generate consensus sequences, but also include information related to the expected location of the polymerase on the template nucleic acid, thereby linking a particular sequence read to a particular location on the template nucleic acid. This greatly facilitates accurate assembly of sequence reads to generate sequence scaffolds and/or consensus sequences.

[0073] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads in an illuminated sequencing-by-incorporation reaction that are distal from one another (i.e., noncontiguous) on a single nucleic acid template by removing the excitation illumination during the course of the reaction, and subsequently reinitiating the excitation illumination. Sequence reads are generated only during the periods of time when the excitation illumination is present, resulting in a "gap" between the sequence reads from a single template nucleic acid that corresponds to the time during which the excitation illumination was absent but the incorporation of nascent nucleotides continued "in the dark." As such, the number of sequence reads generated for a given template nucleic acid is equal to the number of periods during which the excitation illumination is present.

[0074] Certain aspects of the invention are directed to methods, devices, and systems for generating multiple sequence reads from a plurality of nucleic acid templates comprising identical nucleotide sequences. In some embodiments, the multiple sequence reads are not all from the same region of the nucleic acid templates. In some embodiments, there is overlap between the multiple sequence reads. In some embodiments, a single sequence read is generated from each of the plurality of nucleic acid templates, and in other embodiments multiple noncontiguous sequence reads are generated from each of the plurality of nucleic acid templates. In certain preferred embodiments, the multiple noncontiguous sequence reads from each of the plurality of nucleic acid templates together extend across the nucleic acid templates such that they can be combined to provide a consensus sequence for the identical nucleotide sequence in the nucleic acid templates. In some embodiments, the consensus sequence is based on at least 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage of the identical nucleotide sequence. In some embodiments, the

identical nucleotide sequence represents at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the nucleic acid template.

[0075] Certain aspects of the invention are directed to methods, devices, and systems for reducing or limiting the effects of photo-induced damage during illuminated reactions, particularly reactions that employ fluorescent or fluorogenic reactants. The term "photo-induced damage" refers generally to any direct or indirect impact of illumination on one or more reagents in a reaction resulting in a negative impact upon that reaction. Without being bound to a particular theory or mechanism of operation, some illuminated reactions are subject to photo-induced damage that can hinder progression of the reaction, e.g., via damage to reaction components, such as enzymes, cofactors, templates, etc. As such, the illumination of the illuminated reaction can directly or indirectly negatively impact progression of the reaction, and such an impact can be measured based on various characteristics of the reaction progression, e.g., processivity, rate, fidelity, duration, etc. The present invention provides methods for subjecting an illuminated reaction to intermittent exposure to illumination, which reduces the amount of photo-induced damage at a given time during the reaction, allowing the reaction to proceed further than it does when constantly exposed to the illumination.

[0076] In some embodiments, the methods herein may further comprise the addition of one or more photo-induced damage mitigating agents (e.g., triplet-state quenchers and/or free radical quenchers) to the illuminated reaction. Such photo-damage mitigating agents are generally known to those of skill in the art. Further discussion of photo-induced damage and related compounds, compositions, methods, devices, and systems are also provided in U.S. Pub. No. 20070161017, filed December 1, 2006; and U.S.S.N. 61/116,048, filed November 19, 2008, which are incorporated by reference herein in their entireties for all purposes.

II. Intermittent Illumination of Analytical Reactions

[0077] Certain aspects of the invention are generally directed to improved methods for performing illuminated analyses. The terms "illuminated analysis" and "illuminated reaction" are used interchangeably and generally refer to an analytical reaction that is occurring while being illuminated (e.g., with excitation radiation), so as to evaluate the production, consumption, and/or conversion of luminescent (e.g., fluorescent) reactants and/or products. As used herein, the terms "reactant" and "reagent" are used interchangeably. As used herein, the terms "excitation illumination" and "excitation radiation" are used interchangeably. In certain embodiments, the

illuminated reaction is a sequencing reaction, e.g., a sequencing-by-incorporation reaction. In certain embodiments, the illuminated reaction is designed to analyze a single molecule, e.g., by ensuring the molecule is optically resolvable from any other molecule being analyzed and/or in the reaction mixture. In certain embodiments, one or more components of the reaction are susceptible to photo-induced damage directly or indirectly elicited by an excitation radiation source. In certain preferred embodiments, an illuminated reaction is subjected to intermittent excitation radiation during the course of the illuminated reaction. In certain preferred embodiments, a sequencing-by-incorporation reaction is subjected to intermittent excitation radiation during the course of a polymerization reaction to generate a plurality of noncontiguous sequence reads from a single nucleic acid template.

[0078] In certain aspects, the methods herein provide benefits over methods currently used for sequencing large template nucleic acids, such as human genomes. For example, the traditional shotgun sequencing approach entails sequencing nucleic acid fragments and analyzing the resulting sequence information for overlap and similarity to known sequences to construct the complete sequence of the template nucleic acid. One disadvantage to the shotgun approach is that assembly may be difficult if the template nucleic acid comprises numerous repeated sequences, and the inability to assemble a genomic sequence in repeat regions leads to gaps in the assembled sequence. (See, e.g., Myers, G.; "Whole-Genome DNA Sequencing" in Computing in Science and Engineering; Vol 1, Issue 3; pgs. 33-43; May/Jun 1999.) One method of resolving these gaps is to sequence fragments large enough to span the repeat regions, but sequencing large fragments can be difficult and time-consuming. Another approach to spanning a gap is to determine the sequence of two ends of a large fragment which has known spacing and orientation, and this approach is generally termed paired end sequencing (see, e.g., Smith, M. W. et al., (1994) Nature Genetics 7:40-47; and U.S. Pub. No. 2006/0292611, filed June 6, 2006, both of which are incorporated by reference herein in their entireties for all purposes). This method is limited by the requirement for information about the spacing and orientation of the ends of the long fragment, and/or complex sample preparation of the nucleic acid template. The present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.) or complex sample preparation, thereby allowing economical, efficient, and effective de novo sequencing or resequencing of long template nucleic acids.

[0079] In certain aspects, the methods herein provide various strategies for achieving intermittent illumination of illuminated reactions. Essentially, at least one type of illumination (e.g., excitation illumination) is present for at least one time period ("illuminated period") and absent during at least one other time period ("non-illuminated period") during an illuminated reaction. As described above, the term "non-illuminated" indicates a change in illumination including, but not limited to a complete absence of illumination. For example, a non-illuminated period may also be characterized by a different illumination source or intensity than an illuminated period, or by a change in reaction components, e.g., detectable labels. In general, at least one type of data collected during an illuminated period (e.g., nucleotide sequence data) is not collected during a non-illuminated period. An absence of the illumination may be due to, e.g., inactivation of the illumination source (e.g., laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source), removal of the illuminated reaction from the illumination source (or vice versa), or may be due to blockage of the illumination from the reaction, as discussed below. Modifications to the illumination may be due to, e.g., adjustment of the intensity of an illumination source, or a substitution of one illumination wavelength and/or frequency for another. Further, components detectable during an illuminated period may be removed from the reaction mixture during a non-illuminated period, e.g., a fluorescently labeled nucleotide may be replaced with an unlabeled nucleotide. Knowledge of the rate of the reaction and the time during which the illumination is absent is used to estimate the progress of the reaction during the non-illuminated period. For example, if a reaction proceeds such that one molecule is incorporated into a macromolecule per second, and the illumination is absent for 20 seconds, it can be estimated that 20 molecules were incorporated during the non-illuminated period. This information is useful during data analysis to provide context for the reaction data collected during the illuminated period(s). For example, in a sequencing-by-incorporation reaction the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. Further the processivity of an enzyme during a non-illuminated period (or other type of non-detection period) can be manipulated or adjusted by methods known to those of skill in the art. In

particular, the kinetics of replication by a polymerase enzyme can be altered by changing the chemical environment in which it operates, and such methods are further described, e.g., in U.S. Patent Application Nos. 12/414,191, filed March 30, 2009; 12/537,130, filed August 6, 2009; and U.S. Patent Application No. [unassigned], attorney docket no. 105-006301US, entitled "Engineering Polymerases and Reaction Conditions for Modified Incorporation Properties," filed September 4, 2009, the disclosures of all of which are incorporated herein by reference in their entireties for all purposes. For example, methods are provided for adjusting the enzyme activity, and these methods find particular relevance in the instant invention when used to enhance accuracy during detection periods, and to enhance processivity during non-detection periods. Information regarding enzyme translocation rate and processivity is useful for positioning the sequence reads for a single template nucleic acid relative to one another in the construction of a sequence scaffold and/or consensus sequence for the template nucleic acid.

[0080] Figure 1 provides exemplary embodiments of methods for intermittent illumination of analytical reactions. A reaction mix is prepared at step 100. In the process shown on the left, illumination of the reaction 105 is begun prior to initiation of the reaction 110, which allows "illumination data" to be collected at initiation. (In an alternative embodiment, illumination may commence simultaneously with initiation of the reaction.) "Illumination data" as used herein refers to data collected during an illuminated period, e.g., the length of the illuminated period and luminescent signal(s) from the reaction product. At least one non-illuminated period 115 occurs during the course of the reaction, followed by at least one additional illuminated period 120. Multiple additional non-illuminated and illuminated periods may follow. During the illuminated periods (105 and 120), illumination data is collected 175. During the non-illuminated period(s), non-illumination data is collected 180. As used herein, "non-illumination data" refers to data collected during a non-illuminated period, e.g., the length of the non-illuminated period can be monitored. In the process shown on the right, the reaction is initiated 155 during a first non-illuminated period 150. At least one illuminated period 160 occurs during the course of the reaction, optionally followed by at least one additional non-illuminated period 165. Multiple additional illuminated and non-illuminated periods may follow. As for process A, illumination data is collected 175 during the illuminated period(s) 160, and non-illumination data is collected 180 during non-illuminated periods (155 and 165).

[0081] One benefit provided in certain embodiments of the invention is that the reaction need not be further manipulated after initiation (aside from the control of illumination). For

example, the method can be used to analyze reaction mixtures without the need for buffer changes, addition of further reaction components, or removal of detectable components, e.g., light-activatable components such as fluorophores. For example, in a sequencing-by-incorporation reaction, labeled nucleotides may be present throughout the life of the reaction, even when the reaction is not generating nucleotide sequence data (e.g., during a non-illuminated period). This provides clear advantages over methods that require additional handling of the reaction after initiation, which tend to not only be expensive and time-consuming, but which also provide opportunities for contamination of the reaction. For example, illumination can be reinitiated at any time during the reaction at the whim of the ordinary practitioner by simply activating the illumination. In certain preferred embodiments, the concentration of labeled nucleotides or nucleotide analogs in the reaction mixture is greater than the concentration of unlabeled nucleotides in the reaction mixture throughout the course of the reactions, and may represent at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% of the total nucleotides in the reaction mixture. Methods for ensuring a high ratio of labeled versus unlabeled nucleotides in a reaction mixture are known in the art and certain preferred embodiments are provided in U.S. Patent Pub. Nos. 2006/0063264, 2006/0194232, and 2007/0141598, which are incorporated herein by reference in their entireties for all purposes.

[0082] In embodiments in which a sequencing-by-incorporation reaction is subjected to intermittent illumination, the sequence reads collected during the illuminated periods are arranged in order and separated from one another by an estimated number of nucleotides incorporated into the nascent strand during the intervening non-illuminated periods. The resulting gapped read can then be used to assess certain characteristics of the template nucleic acid. When multiple identical template nucleic acids are subjected to such a sequencing-by-incorporation reaction, the resulting set of gapped reads can be combined to create a sequence scaffold and/or a consensus sequence for the template nucleic acid.

[0083] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods can be used to provide additional data during the course of the reaction, e.g., data from illuminated or non-illuminated periods. In certain preferred embodiments, such alternative labeling methods may comprise using labels that are incorporated into a product of the reaction. For example, in sequencing-by-incorporation reactions that use nucleotides comprising labeled terminal phosphates (*e.g.*, the gamma phosphate as in dNTP, or terminal phosphates on nucleotide analogs with a greater number of phosphate groups) to identify

the nucleotides incorporated into a nascent polynucleotide, the reaction mixture may also include nucleotides comprising a base-linked label. During the reaction, these "base-labeled nucleotides" will be incorporated into the nascent strand, but unlike the terminal phosphate labels removed during incorporation, the base-linked labels are not cleaved from the nucleotide upon incorporation by the polymerase, resulting in a nascent strand that comprises the base-linked labels. The concentration of such base-labeled nucleotides can be adjusted in the reaction mixture to promote their incorporation into the nascent strand at a predictable rate, e.g., based on the known sequence of the template or the average frequency of a given nucleotide. The presence and/or rate of incorporation of the base-linked labels into the nascent strand can provide a measure of the length of the nascent strand generated (and, therefore, the distance traveled by the polymerase along the template nucleic acid) during the reaction by subjecting the reaction to excitation illumination that excites the base-linked label (but preferably not the non-base-linked labels), and detecting the signal emitted. The excitation of the base-linked labels preferably occurs as a pulse during or immediately following a non-illuminated period, and is otherwise absent during the reaction. The strength of the signal is indicative of how many labels are present in the nascent strand, thereby providing a measure of the processivity of the polymerase for a given period during the ongoing reaction, e.g. during one or more illuminated or non-illuminated periods. Since the base-linked labels remain in the nascent strand, it is beneficial to minimize the amount of time those fluorophores are subjected to excitation illumination to mitigate the potential of photo-induced damage to the reaction components. As such, in preferred embodiments, the excitation illumination wavelength for the base-labeled nucleotides is different than that of other fluorescent labels in the reaction.

[0084] This method can be modified in various ways. For example, the base-labeled nucleotides may also comprise a terminal phosphate label so that their incorporation can be monitored in the same manner during an illuminated period as the non-base-labeled nucleotides. There may be a single type of base-labeled nucleotide in a reaction mixture, or multiple types may be present, e.g., each type carrying a different nucleobase. The concentration of base-labeled nucleotides in the reaction mix may be varied, although it is preferred that the ratio of base-labeled nucleotides to non-base-labeled nucleotides be relatively low. For example, in a reaction mixture comprising a single type of base-labeled nucleotide (e.g., base-labeled dATP), it is preferred that the ratio of base-labeled dATP to non-base labeled dATP be less than 1:8, and more preferably 1:10 or less. The low concentration of base-labeled nucleotides is preferred in order to minimize sterically induced polymerase stalling when incorporating multiple base-labeled nucleotides in a row. In some

embodiments, the optimal ratio is pre-determined using capillary electrophoresis for any specific base-labeled nucleotide and likely homopolymer sequence prevalence. In certain preferred embodiments, at least 50, 75, 100, 125, or 150 base-labeled nucleotides are incorporated into the nascent strand during a single non-detection period. The base-labeled nucleotides may be present throughout the reaction, or may be washed in during non-illuminated periods and washed out after the pulse of excitation illumination. The reaction mixture comprising base-linked nucleotides being washed in may also include unlabeled nucleotides for incorporation during a non-detection period. During a subsequent illuminated period, a reaction mixture comprising terminal phosphate-labeled nucleotides replaces the reaction mixture comprising base-linked nucleotides and unlabeled nucleotides. This protocol is one embodiment of the methods of the invention in which a non-detection period is not necessarily a non-illuminated period because in this case illumination may be present, but no incorporation of nucleotides is be detected.

[0085] Alternatively or in addition, a low concentration of a fifth terminal phosphate labeled nucleotide can be present in the sequencing reaction, wherein the label has a different excitation wavelength than the other labels in the reaction mixture. For example, a small proportion of one nucleotide analog, e.g., dA6P, can be labeled with the "fifth label." During non-detection periods when the sequence of incorporation of nucleotides is not being monitored, the reaction site is illuminated by excitation radiation specific for the fifth label, and this fifth label excitation radiation can be inactivated during the detection periods. Emissions detected upon incorporation of the nucleotide analog comprising the fifth label are used to "clock" the pace of the polymerase during the non-detection period, e.g., based upon the known or estimated frequency of the complementary nucleotide in the template strand. The fifth label can be chosen such that the excitation and emission radiation are less likely or unlikely to cause photo-induced damage to reaction components, e.g. by choosing a label with a long excitation wavelength (e.g., toward the red end of the visible spectrum), a label that has a low propensity for entering into a triplet state, and/or a label that has a low propensity to form a radical. Since the fifth label is being excited when other labels are not, there is no requirement for optimal spectral separation from other labels in the reaction mixture. Further, since the fifth label is not being used for sequencing, other optimizations are also not necessary, e.g., related to branching, accuracy, and the like. Various types of labels can be used as a fifth label of the invention including, but not limited to, organic and non-organic dye fluorophores. For example, latex nanoparticles or quantum dots are particularly suitable due to their lower propensity for photo-induced damage of certain analytical reaction components. In certain preferred

embodiments, a quantum dot label has an emission spectrum within the same spectral window as the labels that are used to identify the sequence of base incorporations into the nascent strand ("sequencing labels") but an excitation spectrum that does not overlap those of the sequencing labels to allow detection of the fifth label emissions using the same optical system as is used to detect the sequencing label emissions.

[0086] This method can be modified in various ways. For example, more than one small subset of a nucleotide analog can be labeled with a fifth label, and in certain embodiments, a small subset of each nucleotide analog present in the reaction mixture is labeled with the fifth label. Further, there may be a plurality of additional labels present in the reaction, each of which is present on a small subset of a single type of nucleotide analog, e.g., sixth, seventh, and eighth labels. By increasing the number of types of nucleotide analogs labeled with fifth (or sixth, seventh, eighth) labels, their frequency of incorporation is likewise increased, which improves the translocation rate calculation for the polymerase during the non-detection periods. Alternatively, each type of nucleotide analog can comprise both a sequencing label that is specific for the cognate base in the nucleotide, as well as a fifth label for clocking the polymerase. The sequencing labels are excited and detected during the detection periods and the fifth labels are excited and detected during the non-detection periods. Since every nucleotide analog is labeled with a fifth base, each incorporation event can be counted during the non-detection period and the exact rate of incorporation can be determined. Both the sequencing and fifth labels may be bound to the same or different linkers on the nucleotide analogs. In certain preferred embodiments, a linker on a nucleotide analog positions the fifth label within an illumination zone to allow excitation, but far from an enzyme (e.g., polymerase) to mitigate photo-induced damage related to excitation of and/or emission from the fifth label.

[0087] In some embodiments, the fifth label is also excited by an illumination during the detection periods. The availability of the clocking function during the detection period can be used during sequence analysis to identify positions in the resulting sequence read where a signal was not detected (resulting in an apparent "missing base" in the read) and to distinguish between true insertions and branching events in which two signals are detected for a single incorporation event.

[0088] In yet further embodiments, assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid is facilitated by using "non-illuminated periods" characterized by modified excitation illumination rather than a complete absence of excitation illumination (which can also be termed "low-illuminated periods"). For example, in some

embodiments a lower intensity excitation illumination is used during the non-illuminated periods that excites one or more of the labels that are excited during the illuminated periods. As such, unlike various strategies described above, no fifth label is necessary. The lower intensity excitation illumination results in emissions that are lower intensity but still intense enough to identify an emission signal over background counts, though typically not intense enough to be used to identify the particular label generating the emission signal. For example, if label "A" and label "B" are in a reaction mixture, during an illuminated period the intensity of the signal emissions from each are high enough that the artisan can distinguish from which label a particular signal originates by the wavelength and/or frequency of the signal. However, during a low-illuminated period the artisan can only identify that a signal emission occurs, but is unable to distinguish the originating label because its particular wavelength and/or frequency cannot be accurately determined. The decrease in excitation illumination intensity provides both a mitigation of photo-induced damage to reaction components within the observation volume while allowing the practitioner to count the emissions, and therefore the incorporations, during the non-illuminated period.

[0089] In other embodiments, multiple excitation illumination sources are used during an illuminated period, and a first subset of these illumination sources is removed during a non-illuminated period, while a second subset remains. The illumination sources that remain during the non-illuminated period may be present in the same manner as during the illuminated period, or various aspects may be altered, e.g., intensity may be reduced. For example, if labels A and B present in a reaction mixture are excited by a first illumination source and labels C and D present in the reaction mixture are excited by a second illumination source, removal of the first illumination source during the non-illuminated period results in an inability to detect labels A and B, while C and D are still detectable. Such an incomplete data set can be used to clock the progress of the reaction during the non-illuminated period(s). Further, it can also be used in various ways to facilitate the statistical analysis of data collected during the illuminated period(s). For example, for nucleotide sequencing applications (as described elsewhere herein) the incomplete data set(s) collected during non-illuminated period(s) can be used during assembly of a sequence scaffold. For example, during *de novo* sequence assembly a collection of sequences (contigs) are generated, but the order of the contigs relative to the template nucleic acid is not always apparent. The scaffolding process uses extra information to determine the correct order of the contigs. So, if only two bases are identifiable in the non-illuminated periods, the incomplete sequence reads comprising only incorporation of these two bases can be aligned to modified versions of the contigs assembled from

data collected during an illuminated period, but in which the two bases not detected during the non-illuminated periods have been removed. Once the order of the contigs has been determined, the incorporation data for the two bases not detected during the non-illuminated periods is restored and the assembly of the contigs is complete. This method can be modified in various ways. For example, the practitioner may choose which illumination sources to remove during the non-illuminated periods based on various characteristics, such as their propensity to cause photo-induced damage to one or more reaction components, the propensity of the corresponding emission signal to cause photo-induced damage to one or more reaction components; their energy consumption; and wear-and-tear on the source device. Further, as described elsewhere herein, rather then removing an illumination source, reaction components that are excited by the illumination source may be removed from the reaction mixture during the non-illuminated period, necessarily rendering them undetectable. For example, one or more fluorescently labeled nucleotide analogs may be replaced with unlabeled nucleotide analogs during the non-illuminated periods.

[0090] In certain aspects, the invention provides advantages to performing intramolecular redundant sequencing, in which a template nucleic acid is used to generate multiple copies of a sequence read of interest, whether by virtue of multiple copies of the complement being present in the template, repeated replication of the template, or a combination thereof. For example, a first stage of a template-dependent sequencing reaction on a single-stranded circular template can comprise a non-illuminated period during which the template is completely replicated at least one time to generate at least one incomplete sequence read for a sequence complementary to the template. The first stage is followed by a second stage comprising an illuminated period during which the template is replicated multiple times to generate multiple complete sequence reads for the complementary sequence. The incomplete reads generated in the first stage can be used to construct a scaffold for assembly of the complete sequence reads generated in the second stage. Further, incomplete sequence reads can also be used to clock the progress of the reaction during the non-illuminated periods by providing a count of the detectable reaction components and combining that information with known or estimated characteristics of the template, e.g., nucleotide composition or sequence.

[0091] The subset of signal emissions detectable in the non-illuminated periods as compared to the number detectable in the illuminated periods is not limiting and may be chosen based upon the non-illumination data desired by the ordinary practitioner and/or other considerations, such as mitigation of photo-induced damage to extend readlength. For example, to lower the likelihood of

photo-induced damage, the ordinary practitioner may choose to remove the illumination source that is most damaging, e.g., has the highest frequency. In certain embodiments, multiple sequencing reactions may be performed for a single amplified template, each with a different combination of illumination sources and/or detectable components. Alternatively or additionally, multiple replicate reactions can also be performed for one or more of the combinations of illumination sources and/or detectable components. The combination of data from multiple different and/or replicate reactions performed on a single template provides myriad benefits during statistical analysis. As noted above, data can be combined to facilitate assembly of contigs generated during illuminated periods. Data from non-illuminated periods can also provide value in assessing the quality of the sequence reads generated during the illuminated periods.

[0092] Additional methods may also be used to aid in assembly of gapped reads into a sequence scaffold and/or a consensus sequence for a template nucleic acid. For example, in some embodiments, alternative labeling methods used to provide additional data during the course of the reaction can comprise using labels that are incorporated into an enzyme of the reaction. For example, FRET labels can be used to label portions of a polymerase enzyme such that the conformational change between the open and closed states of the enzyme change the FRET value. For example, a FRET-based system can be used to monitor the kinetics of opening and closing of the finger subdomain of DNA polymerase, as described in Allen, et al. (2008) Protein Science 17:401-408, incorporated herein by reference in its entirety for all purposes. In certain preferred embodiments, a closed conformation produces a FRET signal because the donor and acceptor are close to one another, and an open conformation silences the signal because there is no energy transferred between the donor and acceptor. By monitoring the emission from the FRET pair, each incorporation event can be monitored during non-detection periods, and optionally or additionally during detection periods. In certain preferred embodiments, the FRET donor is GFP (excitation at 484 nm; emission at 510 nm), and the FRET acceptor is YFP (excitation at 512 nm; emission at 529 nm). Methods for monitoring polymerase activity using FRET labels are known in the art, e.g., in WO/2007/070572 A2, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

[0093] A given reaction may experience one or a plurality of illuminated periods or non-illuminated periods, but preferably experiences at least two illuminated periods. For example, a given reaction providing nucleotide sequence information from a single template nucleic acid may have at least about 2, 3, 5, 10, 20, 50, or 100 illuminated periods with intervening non-illuminated

periods. In an embodiment employing multiple periods of illumination and/or non-illuminated, the periods may be the same for both, e.g., 100 seconds "on" and 100 seconds "off." Alternatively, the illuminated periods may be longer or shorter than the non-illuminated periods. For example, in certain embodiments, a non-illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent illuminated period; or an illuminated period may be at least about 2-, 3-, 4-, 6-, 8-, 10-, 20-, or 50-fold longer than an adjacent non-illuminated period. Further, each illuminated period may be the same or different from each other illuminated period, and each non-illuminated period may be the same or different from each other non-illuminated period. For example, some embodiments generate a smaller number of long reads, and other embodiments generate a larger number of short reads. It will be understood that the number and length of the illuminated and non-illuminated periods is limited only by the experimental system in use and the data acquisition goals of the ordinary practitioner. In some embodiments, a nucleotide sequence read generated during a single illuminated period comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, a region of a nucleic acid template processed during a non-illuminated period during a single reaction comprises at least about 20, 30, 40, 50, 75, 100, 1000, 10,000, 25,000, 50,000, or 100,000 adjacent nucleotide positions. In some embodiments, the set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises at least about 40, 60, 80, 100, 1000, 10,000, 25,000, 50,000, 100,000, 250,000, 500,000, or 1,000,000 nucleotide sequence positions from a single nucleic acid template. In some embodiments, a set of nucleotide sequence reads generated during a single sequencing reaction comprising a plurality of illuminated periods comprises multiple reads of at least a portion of the nucleotide sequence positions from a single nucleic acid template.

[0094] As noted above, the present invention provides methods that are tolerant of large repetitive regions and do not require prior knowledge of nucleotide sequences (e.g., base sequences, spacing, orientation, etc.). However, such information, if available, may also be useful to the ordinary practitioner in determining an optimal periodicity for illuminated and non-illuminated periods during a sequencing reaction, especially when sequencing repetitive sequences. For example, if a genomic region is known to contain five adjacent copies of a one kilobase nucleotide sequence (i.e., five "repeat regions"), it would be beneficial to keep the non-illuminated periods short enough to be able to confidently map the resulting sequence reads to the correct repeat region. If a non-illuminated period were too long, the natural variation in translocation rate of the

polymerase would make it difficult to assign a sequence read to a particular repeat region, especially those farther from the binding/initiation site of the polymerase. In a further example, if the "copies" each had a few mutations that could be used to distinguish them from each other, it would be beneficial to keep the illuminated periods long enough to increase the chance one of these mutations would be included in a resulting sequence read, thereby allowing the unambiguous assignment of the read to a particular repeat region. If the illuminated period were too short the sequence reads from two different repeat regions could be identical, making mapping the sequence read challenging. (Another way to mitigate these difficulties would be to incorporate pause or stop points into the template nucleic acid, as discussed below.)

[0095] Essentially, the practitioner may design the number of and lengths of time for each illuminated and non-illuminated period to best suit the illuminated reactions being analyzed and the invention is not limited in this regard. In certain embodiments, a practitioner may wish to increase the processivity of a polymerase thereby extending the length of the template nucleic acid processed in a sequencing reaction to be, e.g., at least 2-, 3-, 4-, 6-, 8-, 10-, or 20-fold, thereby generating sequence data much farther away from the polymerase binding/initiation site than would be achieved under constant illumination. In certain embodiments, a practitioner of the instant invention may wish to focus on data from one or more stages of an ongoing reaction, such as stages for which more data is required for analysis. In the case of sequencing-by-synthesis, one or more particular regions of a template nucleic acid may need to be resequenced. Some traditional methods require that new template nucleic acids be prepared to bring a region requiring resequencing closer to the initiation point of the sequencing reaction, or require preparation of multiple new templates if multiple regions to be resequenced. In contrast, the methods herein allow the practitioner to subject a template identical to the previously sequenced template (e.g., from a large genomic DNA sample preparation) to a sequencing reaction wherein illuminated periods are timed to illuminate the sample only when the polymerase is incorporating nucleotides into the nascent strand at the one or more particular regions requiring resequencing. This advantage substantially lowers the time and resources required for such resequencing operations, therefore providing a significant advantage over traditional methods.

[0096] The instant invention contemplates various means for providing non-illuminated periods during illuminated reactions. In some embodiments, the illumination source is turned off during the ongoing reaction to create one or more non-illuminated periods. In some embodiments, the illumination source remains on during the course of the reaction, but the illuminated reaction is

removed from the system for a period of time. In some embodiments, the illumination source remains on during the course of the reaction, but the illumination is blocked to create one or more non-illuminated periods. For example, a movable mask may be manually or mechanically positioned between the illumination source and the illuminated reaction to block the illumination during non-illuminated periods and removed to allow exposure to the illumination during illuminated periods. Such a mask may also be dynamically controlled, such as a thin film transistor display (e.g., an LCD mask). Masks for blocking illumination and manufacture thereof are well known to those of ordinary skill in the art and need no further elaboration herein.

[0097] One aspect of the present invention is multiplexing of large numbers of single-molecule analyses. For a number of approaches, e.g., single molecule methods as described above, it may be desirable to provide the reaction components in individually optically resolvable configurations, such that a single reaction component or complex can be individually monitored. Providing such individually resolvable configurations can be accomplished through a number of mechanisms. For example, by providing a dilute solution of complexes on a substrate surface suited for immobilization, one will be able to provide individually optically resolvable complexes. (See, e.g., European Patent No.1105529 to Balasubramanian, et al., the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Alternatively, one may provide a low density activated surface to which complexes are coupled. (See, e.g., Published International Patent Application No. WO 2007/041394, the full disclosure of which is incorporated herein by reference in its entirety for all purposes). Such individual complexes may be provided on planar substrates or otherwise incorporated into other structures, e.g., zero-mode waveguides or waveguide arrays, to facilitate their observation.

[0098] In some embodiments, a plurality of illuminated reactions are carried out simultaneously, e.g., on a solid support. In some preferred embodiments, a solid support comprises an array of reaction sites. In preferred embodiments, the reaction sites on a solid support are optically resolvable from each other. In further preferred embodiments, each of the reaction sites on a solid support contains no more than a single reaction to be interrogated. For example, in a sequencing-by-incorporation embodiment, each reaction site preferably has no more than one polymerase and no more than one nucleic acid template. The reaction sites may be confinements (e.g., optical and/or physical confinements), each with an effective observation volume that permits resolution of individual molecules present at a concentration that is higher than one nanomolar, or higher than 100 nanomolar, or on the order of micromolar range. In certain preferred embodiments,

each of the individual confinements yields an effective observation volume that permits resolution of individual molecules present at a physiologically relevant concentration, *e.g.*, at a concentration higher than about 1 micromolar, or higher than 50 micromolar range or even higher than 100 micromolar. In addition, for purposes of discussion herein, whether a particular reagent is confined by virtue of structural barriers to its free movement, or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined."

[0099] As used herein, a solid support may comprise any of a variety of formats, from planar substrates, e.g., glass slides or planar surfaces within a larger structure, e.g., a multi-well plates such as 96 well, 384 well and 1536 well plates or regularly spaced micro- or nano-porous substrates, or such substrates may comprise more irregular porous materials, such as membranes, aerogels, fibrous mats, or the like, or they may comprise particulate substrates, e.g., beads, spheres, metal or semiconductor nanoparticles, or the like. The solid support may comprise an array of one or more zero-mode waveguides or other nanoscale optical structures.

[00100] As used herein, "zero-mode waveguide" refers to an optical guide in which the majority of incident radiation is attenuated, preferably more than 80%, more preferably more than 90%, even more preferably more than 99% of the incident radiation is attenuated. As such high level of attenuation, no significant propagating modes of electromagnetic radiation exist in the guide. Consequently, the rapid decay of incident electromagnetic radiation at the entrance of such guide provides an extremely small observation volume effective to detect single molecules, even when they are present at a concentration as high as in the micromolar range. The fabrication and application of ZMWs in biochemical analysis, and methods for calling bases in sequencing-by-incorporation methods are described, e.g., in U.S. Patent Nos. 7,315,019, 6,917,726, 7,013,054, 7,181,122, and 7,292,742, U.S. Patent Pub. No. 2003/0174992, and U.S. Patent Application No. 12/134,186, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00101] A set of reactions (e.g., contained on a solid support) may comprise identical or different components. For example, a single template nucleic acid may be analyzed in all reactions in the set, or a plurality of template nucleic acids may be analyzed, each present in only one or a subset of the set of reactions. In preferred embodiments, template nucleic acids comprising the same nucleotide sequence are analyzed in a plurality of reactions sufficient to provide adequate redundant nucleotide sequence data to determine a consensus sequence for the template nucleic acids. A number of sequence reads that will provide adequate nucleotide sequence data will vary, depending,

e.g., on the quality of the template nucleic acid and other components of the reaction, but in general coverage for a template nucleic acid or portion(s) thereof is at least about 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, or 1000-fold coverage. Further, the numbers and lengths of illuminated and non-illuminated periods for a given reaction in the set of reactions may be the same or different than those for other reactions in the set. In some embodiments, a mixture of different periodicities are used for a set of reactions comprising the same template nucleic acid. This strategy can be beneficial for providing nucleotide sequence reads from varying regions of the template sequence, thereby increasing the likelihood of overlapping sequence reads between individual reactions. These overlapping sequence reads can facilitate construction of a more robust sequence scaffold than could be constructed were the reactions all subjected to the same periodicity of illuminated and non-illuminated periods.

[00102] Methods of controlling polymerase progress and/or synchronizing polymerases in different reactions are also useful in analysis (e.g., mapping, validation, etc.) of nucleic acid reads farther from the initial binding site of the polymerase. During detection periods earlier in the reaction (i.e., closer to the time at which the polymerase began to process the template nucleic acid, such as during a first illuminated period), the position of a polymerase on the template can be estimated with generally good accuracy based on the known translocation rate of the polymerase under a given set of reaction conditions. As the duration of the reaction increases, however, the natural variation in polymerase translocation rate makes it more difficult to accurately determine the exact position of the polymerase on a template using estimation based on translocation rate alone; and through each subsequent illuminated period such estimations of polymerase position become less accurate, making subsequent analysis and mapping of the sequence reads to the template more difficult. Methods of regulating the position of the polymerase on the template allow more accurate determinations the polymerase's position. For example, causing the polymerase to pause or stop at a given location on the template during a non-illuminated period and reinitiating the polymerization during or immediately prior to a subsequent illuminated period provides a way to reorient the subsequently generated read with the template sequence, allowing easier consensus sequence determination and mapping analyses. Further, such pause/stop points can provide a means of controlling what regions of the template are processed during the illuminated periods by restricting where the polymerase will reinitiate on the template, thereby allowing a practioner of the instant invention to target one or more particular regions of a template for analysis during one or more detection periods during the course of an analytical reaction. Such methods are also useful to

synchronize a set reactions being monitored simultaneously. For example, a plurality of reactions, each comprising a single polymerase/template complex, may be synchronized by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that show less spreading (i.e., less variation in the position on the template from which the sequence reads are generated) in the later stages of the reactions than would otherwise be observed without such regulation.

[00103] Various methods can be used to control or monitor the progress of a polymerase on a template nucleic acid. For example, as noted above, one may employ a reaction stop or pause point within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the single-stranded portion that was not used in priming. Reaction stop or pause points can be engineered into a portion of the template for which the nucleotide sequence is unknown (e.g., a genomic fragment), but is preferably located within a portion for which the nucleotide sequence is known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, certain preferred sequencing templates (e.g., SMRTbell™ templates, described elsewhere herein) are closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, and one or both of these linkers can comprise a stop or pause point to control the passage of the polymerase through them. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00104] By way of example, at a selected time following initiation of polymerization the reaction may be subjected to a non-illuminated period. The incorporation of a synthesis blocking moiety coupled to the template nucleic acid at a position encountered by the polymerase during the non-illuminated period will cause the polymerase to pause. An example of an engineered pause point is a known sequence on the template nucleic acid where a primer sits and blocks progression of a polymerase that is actively synthesizing a complementary strand. The presence of the primer by itself could introduce a pause in the polymerase sequencing or the primer could be chemically modified to force a full stop (and synchronization of multiple polymerases in multiple reactions). The chemical modification could be subsequently removed (for example, photo-chemically) and the polymerase would subsequently continue along the template nucleic acid. In some embodiments,

multiple primers could be included in a reaction to introduce multiple pause or stop points along the template nucleic acid. Other methods for inducing a reversible pause (stop) in synthesis are known in the art and include, e.g., reversible sequestering of required cofactors (e.g., Mn^{2+}, one or more nucleotides, etc.). Once sufficient time has passed that the polymerase is paused at the blocking group, illumination is reintroduced and the blocking group removed. This allows control of the position on the template nucleic acid at which the polymerase will begin generating nucleotide sequence data during the illuminated period. A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the template nucleic acid that inhibit polymerase mediated replication, strand-binding moieties that prevent processive synthesis, non-native nucleotides included within the primer and/or the template, and the like. Such reaction stops/pause points are useful in providing more certainty about the relationship of the reads to each other. For example, since the exact position on a template nucleic acid at which each sequence read begins would be known, the resulting reads could be better mapped relative to one another for construction of a sequence scaffold and/or consensus sequence. Further description of these and other methods for regulating the progress of a polymerase on a template are provided, e.g., in U.S.S.N. 61/099,696, U.S. Patent Pub. No. 2006/0160113, and U.S. Patent Pub. No. 2008/0009007, all of which are incorporated by reference herein in their entireties for all purposes.)

[00106] By way of example, a sequencing reaction may be initiated on a template comprising a non-native base in the absence of the complement to the non-native base, which would not impact the overall sequence determination of other portions of the template that are complementary to native bases. By starving the reaction for the complement to the non-native base, one can prohibit synthesis, and thus, the sequencing process, until the non-native base complement is added to the mixture. This can provide a "hot start" capability for the system and/or an internal check on the sequencing process and progress that is configurable to not interfere with sequence analysis of the regions of interest in the template, which would be complementary to only native bases. In some embodiments, the non-native base complement in the sequence mixture is provided with a detectably different label than the complements to the four native bases in the sequence, and the production of incorporation-based signals associated with such labels provides an indication that the polymerase has initiated or reinitiated. Although described as the "non-native base" it will be appreciated that this may comprise a set of non-natural bases that can provide multiple control elements within the template structure. In certain embodiments, two different non-native bases are included within the template structure, but at different points, to regulate procession of the

sequencing process, e.g., allowing controlled initiation and a controlled stop/start position later in the sequence, e.g., prior to a subsequent illuminated period. For example, the complement to the first non-native base can be added to initiate sequencing immediately prior to the start of a first illuminated period. During a first non-illuminated period following the first illuminated period, the polymerase encounters the second non-native base, e.g., at a nucleotide position near but upstream of a nucleotide region desired to be sequenced in a second illuminated period. Sequencing would stop until the complement to the second non-native base is added to the reaction mixture. Likewise, multiple such non-native bases could be incorporated into the template to effectively target the polymerase to multiple regions of interest for which sequence data is desired. Further, in applications in which multiple identical templates are being sequenced, this would allow a resynchronization of the various sequencing reactions and the data generated therefrom.

[00107] Methods of controlling polymerase progress in different stages of a sequencing reaction are also useful for not only creating "condition-dependent" non-detection periods (during which time illumination may or may not be present), but also for minimizing the amount of time required for traversing a given length of template during a non-detection period (whether or not illumination is present). In order to reliably detect incorporation events, non-natural reagent conditions are typically used to limit polymerization during detection periods to approximately 1-5, or about 3 bases per second. In certain embodiments, replacement of Mg^{2+} ions with Mn^{2+} ions serves to stabilize and slow the translocation of the polymerase. When magnesium and, optionally, native nucleotides (e.g., lacking fluorescent labels) are used, the rate of translocation and/or processivity of the polymerase may increase up to two orders of magnitude. Use of such "rapid translocation" conditions during the non-detection periods can provide myriad benefits, including but not limited to a more rapid polymerization rate, an increased processivity (e.g., due to decreased stalling and misincorporation), and an overall savings due to reduced use of expensive labeled nucleotide analogs and/or reagents that mitigate oxidative stress.

[00108] In certain embodiments, a protocol for intermittent detection comprises alternating reaction mixtures, where a first reaction mixture used during the detection periods is optimized for sequence read generation, and a second reaction mixture used during the non-detection periods is optimized for processivity and/or rapid polymerization. For example, when reagents for optimal sequence read generation are present, DNA synthesis rate is low, and there is a fluorescence signal associated with each incorporation event. After replacing the reaction mixture optimized for sequence read generation with the reaction mixture optimized for processivity and/or rapid

polymerization, the polymerase rapidly advances across the template. In certain embodiments, a flow cell is used to deliver and switch between the two (or more) reaction mixtures during the course of the reaction.

[00109] In an exemplary embodiment, a first reaction mixture comprises fluorescently-labeled nucleotide analogs and manganese ions that restrict polymerization to a rate appropriate for high fidelity detection of nucleotide incorporation. The first reaction mixture can also include additional agents for mitigation of photo-induced damage of various components of the reaction mixture. A second reaction mixture comprises natural nucleotides and an appropriate magnesium ion concentration for rapid synthesis of the nascent strand complementary to the template. A first detection period of a sequencing reaction is initiated by introduction of the first reaction mixture, and a sequence read is generated based upon synthesis of the nascent strand during the detection period. After a predetermined time interval a sufficient quantity of the second reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, thereby initiating a first non-detection period. As noted above, the lack of labeled nucleotides in the second reaction mixture alone can produce the non-detection period, since there will be no signal emitted coincident with incorporation of the native nucleotides, but in certain embodiments illumination may also be removed, e.g., to further mitigate photo-induced damage during the non-detection period. At a time appropriate to initiate a second detection period, a sufficient quantity of the first reaction mixture is flowed onto the reaction site(s) until effectively all the first reaction mixture has been replaced with the second, and detection of incorporation event is reinitiated. The cycle of reaction mixture exchange is repeated to generate multiple detection and non-detection periods.

[00110] A flow cell for reaction mixture exchange preferably has two inputs that are gated such that only a single reaction mixture flows into a reaction site or plurality of reaction sites, e.g., on a substrate. A single out-flow line may be used to remove reaction mixtures from the reaction site(s) to a single collection vessel, or multiple collection vessels may be used, one for each type of reaction mixture used. Further, accurate estimation of the distance a polymerase translocates during a non-detection period is important for bioinformatics applications. This estimation is complicated if the time for reaction mixture exchange is slow. As such, the flow is preferably at a sufficient rate that the time for exchange is significantly less than the time spent in the presence of either reaction mixture alone.

[00111] Figure 2 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination. In this embodiment, sixteen sequencing-by-incorporation reactions are performed on single nucleic acid templates (each of which comprises the same nucleotide sequence) with the timing of the illuminated and non-illuminated periods the same for all sixteen reactions. In A, the sixteen reactions are shown disposed on sixteen reaction sites on a solid support and are numbered for convenience. A representation of the illumination data is shown in B, with bars extending across the graph indicative of illumination data collected during illuminated periods for each reaction. In this illustrative example, each reaction is subjected to three illuminated periods, each followed by a non-illuminated period, resulting in three noncontiguous sequence reads for each reaction, i.e., three noncontiguous reads per template molecule sequenced. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. During the first illuminated period, the sequence reads generally overlap, but the natural variation of polymerase translocation rate over the set of reactions results in a "spreading" of the sequence reads as the reaction proceeds through the second and third illuminated periods with increasing variation in the exact position of each polymerase on the template at the beginning and end of each illuminated period. As such, the earlier illumination data provides better redundancy ("oversampling") of sequence information over a relatively narrow portion of the template nucleic acid, while the later illuminated periods provide less redundant sequencing data over a broader region of the template nucleic acid. The timing of the non-illuminated periods between the illuminated periods and the known or calculated rate of incorporation are used to determine approximate spacing between the resulting sequence reads, providing context for building a sequence scaffold or consensus sequence. It is important to note that although shown disposed on a solid support in A, the data shown in B could also have been generated from reactions not disposed on a solid support nor performed simultaneously and the methods are generally not so limited. Further, as described above, the spreading of the sequence reads from later stages of the reactions can be mitigated by synchronizing the reactions, e.g., by regulating the initiation points of the polymerase on the template for each detection period, thereby creating a set of sequence reads that provides better redundancy (i.e., more overlap in the positions on the template from which the sequence reads are generated), especially in the later stages of the reactions.

[00112] Using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is

generated, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be reset to change the regions of the template for which nucleotide sequence data is generated. As described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest. Alternatively, each time a circular template is sequenced the timing of the illuminated and non-illuminated periods can be kept the same to provide a greater-fold coverage of one or more regions of interest in the template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which are incorporated by reference herein in their entireties for all purposes. A specific embodiment is also provided in the Exemplary Applications section herein.

[00113] The present invention provides novel template configurations and methods for exploiting these compositions in template directed sequencing processes. While these compositions and methods have utility across all of the various template directed processes described herein, for ease of discussion, they are being primarily discussed in terms of preferred single molecule, real-time sequencing processes, in which they provide myriad benefits. In particular, the present invention is generally directed to nucleic acid sequences that employ improved template sequences to improve the accuracy of sequencing processes. For example, in at least one aspect, the template compositions of the invention are generally characterized by the presence of a double stranded segment or a pair of sub-segments that are internally complementary, i.e., complementary to each other. In particular contexts, the target nucleic acid segment that is included within a template

construct will typically be substantially comprised of a double stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double stranded or otherwise internally complementary.

[00114] Examples of template configurations of the invention that are partially and completely contiguous are schematically illustrated in Figure 20A and 20B, respectively. In particular, as shown in Figure 20A, a partially contiguous template sequence 200 is shown which includes a double stranded portion, comprised of two complementary segments 202 and 204, which, for example, represent a target sequence or portion thereof. As shown, the 3' end of segment 202 is linked to the 5' end of segment 204 by linking oligonucleotide 206, providing a single stranded portion of the template, and yielding a partially contiguous sequence. By comparison, as shown in Figure 20B, a completely contiguous template sequence 210 is shown. Sequence 210 includes a double stranded portion again comprised of two complementary segments 212 and 214. As with the partially contiguous sequence of Figure 20A, the 3' end of segment 212 is joined to the 5' end of segment 214 via oligonucleotide 216 in a first single stranded portion. In addition, the 5' end of segment 212 is joined to the 3' end of segment 214 via linking oligonucleotide 218, providing a second single stranded portion, and yielding a completely contiguous or circular template sequence.

[00115] In addition, the templates of the invention, by virtue of their inclusion of double stranded segments, provide consensus through the identification of both the sense and antisense strand of such sequences (in both the partially and completely contiguous configurations).

[00116] By way of example, and with reference to Figures 20A and 20B, with respect to a partially contiguous template shown in Figure 20A, obtaining the entire sequence, e.g., that of segments 202, 204 and 206 provides a measure of consensus by virtue of having sequenced both sense strand, e.g., segment 202, and the antisense strand, e.g., segment 204. In addition to providing sense and antisense consensus within a single template molecule that can be sequenced in one integrated process, the presence of linking segment 206 also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., 202, is completed and the other begins, e.g., 204. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. The progress of sequencing processes is schematically illustrated in Figure 21A. In particular, as shown, a sequencing process that begins, e.g., is primed, at the open end of the partially contiguous template, proceeds along the first or sense strand, providing the nucleotide sequence (A) of that strand, as represented in the schematic

sequence readout provided. The process then proceeds around the linking oligonucleotide of the template, providing the nucleotide sequence (B) of that segment. The process then continues along the antisense strand to the A sequence, and provides the nucleotide sequence (A'), which provides consensus data for the sense strand as its antisense counterpart. As noted, because the B sequence may be exogenously provided, and thus known, it may also provide a registration sequence indicating a point in the sequence determination at which the data transitions from sense to antisense strands.

[00117] With respect to completely contiguous or circular template sequences configured in accordance with the invention, the consensus potential is further increased. In particular, as with the partially contiguous sequences shown in Figure 20A, the completely contiguous sequences also provide sense and antisense consensus. In addition, such templates provide for the potential for iterative sequencing of the same molecule multiple times, by virtue of the circular configuration of the template. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining consensus for each segment from the complementary sequences, as well as consensus within each segment, by repeatedly sequencing that segment. This is schematically illustrated in Figure 21B, again with a representative illustration of a sequence readout provided. As shown, a sequencing process that is primed at one end, e.g., primed within one linking oligonucleotide sequence, e.g., linking oligonucleotide 218 of Figure 20, proceeds along the first or sense strand 214, again providing the nucleotide sequence A of that strand. The sequence process then proceeds around the first linking oligonucleotide, e.g., linking oligonucleotide 216 from Figure 20, to provide the nucleotide sequence B of that segment of the template. Proceeding along the antisense strand, e.g., segment 212 of Figure 20B), provides the nucleotide sequence A', which is again, complementary to sequence A. The sequencing process then continues around the template providing the nucleotide sequence for the other linking oligonucleotide, e.g., linking oligonucleotide 218 of Figure 20B, where the illustrated sequencing process began, providing nucleotide sequence C. Because the template is circular, this process can continue to provided multiple repeated sequence reads from the one template, e.g., shown as providing a second round of the sequence data A-B-A'-C-A-B-A'. Thus, sequence redundancy comes from both the determination of complementary sequences A and A', and the repeated sequencing of each segment. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand with each cycle around the template, allowing continuous sequencing.

Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post synthesis.

[00118] Another exemplary embodiment of an analysis of a plurality of illuminated reactions using intermittent illumination comprises a first illuminated period that is initiated at different times over the plurality of reactions. For example, the illuminated period for a first reaction may start at 0 seconds, the illuminated period for a second reaction may start at 5 seconds, the illuminated period for a third reaction may start at 10 seconds, and so forth. Additionally or alternatively, a first subset of reactions may begin at a first time, a second subset may begin at a second time, etc. The first illuminated period continues for a given length of time, followed by a non-illuminated period and a subsequent second illuminated period. Optionally, a plurality of non-illuminated periods and illuminated periods follow the first illuminated period. Staggered start times can provide staggered data sets (e.g., two or more sequence reads) for the plurality of reactions, allowing multiple different stages of the overall reaction to be interrogated in different reactions. Preferably, the staggered data sets overlap to an extent that allows further analysis and validation of the reaction data. For example, a sequencing-by-incorporation reaction subjected to such an embodiment of the invention would preferably have sufficient overlap between sequence reads from different individual reactions to allow construction of a sequence scaffold and/or consensus sequence for a template nucleic acid.

[00119] A mask for use with a solid support (e.g., an array of confinements) can be designed to allow illumination of one or more portions of the solid support while blocking illumination to other portions of the solid support. For example, a mask may comprise one or more windows that allow excitation illumination to pass through the mask. Such a mask may be physically moved over the surface of the solid support (or the solid support can be moved relative to the mask), e.g., to selectively allow excitation illumination to reach a subset of confinements in an array. For example, a mask that allows 10% of reaction sites to be illuminated could be used to increase the sequencing scaffold coverage by sliding the illumination area (the area being subjected to excitation illumination) back and forth across the solid support. The 10% of reactions would cover certain regions of the nucleic acid template for any given time period (and therefore region of sequence in the template). In certain embodiments, an automated mask that selectively controls the timing of illumination of reactions on a solid support during the course of the reaction/acquisition may be used rather than a mask that must be physically moved.

[00120] The timing of the illuminated and non-illuminated periods for a set of reactions on a solid support may be the same or may vary, and may be synchronized or random. In certain

embodiments in which the excitation illumination source is turned on and off, the timing of the illuminated and non-illuminated periods for the set of reactions will be identical. In other embodiments, for example, those that comprise use of a mask, the timing of the illuminated and non-illuminated periods for the set of reactions can vary so that while a subset of the reactions are illuminated, another subset of the reactions are not illuminated. Various exemplary and nonlimiting embodiments of masks that may be used with a set of reactions on a solid substrate are provided in Figures 3-5, as described below. In certain embodiments, the illuminated/non-illuminated status of each reaction may be random across the solid support, e.g., to remove any experimental bias potentially introduced by actively selecting which reactions to illuminate at a given time, as long as the sequence reads being generated at the illuminated reactions and the time at which these reactions are not illuminated are able to be assigned to a particular reaction. For ease of discussion, the action of both illuminating and collecting emission signals from a reaction of interest, or a particular region on a solid support in which a reaction of interest is taking place, is referred to as "interrogating" that reaction and/or that region. A region being so interrogated is termed an "observation region."

[00121] Figure 3 provides an exemplary embodiment of analysis of a plurality of illuminated reactions using intermittent illumination and a mask. As in Figure 2, an array of reactions on a solid support 310 is provided containing sixteen reaction sites, numbered for convenience (A). In B, a mask 320 is provided with a single window 330 to allow passage of illumination to a subset of reactions on the solid support. Window 330 is wide enough to allow illumination of at least two columns of reaction sites on solid support 310. As in Figure 2, a representation of the illumination data is shown in C, with bars extending across the graph indicative of illumination data collected for each reaction. The position of the bars relative to the x-axis provides the position of the sequence read relative to the template nucleic acid sequence, which extends from position 0 (initiation of sequencing reaction) to n. When the sequencing reaction is initiated at all positions on solid support 310, the window 330 is positioned to allow illumination to only reactions 1, 5, 9, and 13, and these four reactions provide sequence reads 350 for the earliest stage of the reactions. The window 330 is subsequently moved to provide an illuminated period for reactions 2, 6, 10, and 14 while still continuing the illuminated period for reactions 1, 5, 9, and 13. The illumination data for reactions 2, 6, 10, and 14 provides sequence reads 360, which partially overlap sequence reads 350 for reactions 1, 5, 9, and 13. The window 330 is moved again to provide illuminated periods for reactions 3, 7, 11, and 15 while still continuing the illuminated period for reactions 2, 6, 10, and 14, but removing

illumination from reactions 1, 5, 9, and 13. The illumination data for 3, 7, 11, and 15 results in sequence reads 370, which partially overlap sequence reads 360 for reactions 2, 6, 10, and 14. A fourth position of the mask 320 initiates an illuminated period for reactions 4, 8, 12, and 16 while continuing illumination of reactions 3, 7, 11, and 15, but ending the illuminated period for reactions 2, 6, 10, and 14. Sequence reads 380 correspond to sequence reads from reactions 4, 8, 12, and 16. Finally, the window is moved to end the illuminated period for reactions 3, 7, 11, and 15 while continuing the illuminated period for reactions 4, 8, 12, and 16. Repeating the above process allows a second read to be generated from each reaction, and this second read is noncontiguous with the first read. For example, reactions 1, 5, 9, and 13 correspond to reads 350 and, later in the reaction, reads 355. The two reads generated in a single reaction do not overlap and are separated by a length of nucleotides that was incorporated during the non-illuminated period between the two illuminated periods.

[00122] The mask can optionally be passed over the substrate additional times to generate additional reads until the reactions are complete or no longer provide reliable data, such as when the total illumination time (computed by summing the times for the multiple illuminated periods) has surpassed a photo-induced damage threshold period. Further, the mask may be passed back and forth, or may pass over the solid support in only one direction, e.g., always left to right, or vice versa.

[00123] Further, unlike the data shown in Figure 2B which has gaps in the sequence coverage for the template nucleic acid, the strategy provided in this embodiment results in at least two-fold coverage across the entire template nucleic acid (Figure 3C), although at a lower-fold redundancy. The portion of the template covered by only reads 380 and reads 355 has the least-fold redundancy, and in some instances a gap in coverage may be present in this region due to the movement of the mask 320 from the far right to the far left of the solid support 310. Of course, oversampling by adding replicate reactions to the set of reactions, or using templates that allow repeated sequencing (e.g., circular templates) in a single reaction can increase the coverage of a nucleic acid template, thereby providing more data for construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. Various methods for generating redundant sequence reads are known in the art, and certain specific methods are provided in U.S. Patent No. 7,302,146; U.S. Patent No. 7,476,503; U.S.S.N. 61/094,837, filed September 5, 2008; U.S.S.N. 61/099,696, filed September 24, 2008; and U.S.S.N. 61/072,160, filed March 28, 2008, all of which have been previously incorporated by reference herein. The natural variation of polymerase translocation rate over the set

of reactions is also apparent in this prophetic example as the spreading of the sequence reads and decreasing overlap between reads from reactions in adjacent columns in the later stages of the reactions as compared to the earlier stages.

[00124] Figure 4A provides an embodiment of a mask similar to that provided in Figure 3 except that it comprises three windows allowing multiple nonadjacent columns of reaction sites to be illuminated simultaneously. Figure 4B provides an embodiment of a mask comprising twelve windows, each of which allows illumination of a single reaction site on a solid support. The windows are oriented in the mask to allow illumination of every other reaction in each row and every other reaction in each column. It will be understood that these mask designs are merely exemplary and nonlimiting embodiments as it is well within the abilities of the ordinary practitioner to determine an appropriate mask design depending on the experimental design or the illuminated reactions to be interrogated.

[00125] Figure 5B illustrates yet another aspect of the instant invention in which multiple samples are analyzed on a single solid support using intermittent illumination. Four different samples are disposed on a solid support, one in each quadrant 510, 520, 530, and 540 (A). A mask 550 is used that comprises two windows 560 that allow multiple rows of reaction sites to be illuminated simultaneously (B). A first position of this mask over a solid support in which two reactions in each quadrant are illuminated is demonstrated in C. A second position of the mask allowing illumination of the previously non-illuminated reactions is demonstrated in D. The mask may be moved back and forth as indicated by the double-arrow to provide multiple illuminated and non-illuminated periods for each reaction containing one of the four samples.

[00126] The present invention is also useful for redundant interrogation of reactions or portions of a solid support of interest. In certain aspects, sequential interrogation of different observation regions may be repeated a number of times, e.g., more than 2, 5, 10, 50, 100, 500, 1000, or even more than 10,000 times. In general, this method of stepping the observation region to another, preferably adjacent region, and repeating the interrogation process is generally referred to as a "step and repeat" process, and may be performed by various methods, including but not limited to moving the incident light and the solid support relative to one another and moving a mask across the surface of the solid support, as described above. Although described as a "step and repeat" method, in some embodiments where the observation region is moved across a substrate, that movement is not step-wise and iterative, but instead constitutes a continuous motion, substantially continuous motion, or stepped movement, or an iterative motion whereby each iterative step

interrogates a new region that overlaps with some portion of the previously interrogated region. In particular, a substrate may be moved continuously relative to an optical system, whereby the observation region moves continuously across the substrate being interrogated (in a "scan mode").

[00127] The present invention is optionally combined with an optical system that provides illumination and/or collection of emitted illumination. Preferably, the optical system is operatively coupled to the reaction sites, e.g., on a solid support. One example of a particularly preferred optical system is described in U.S.SN. 11/201,768, filed August 11, 2005, and incorporated herein by reference in its entirety for all purposes. Optical systems are described further below.

[00128] In some embodiments, one or both of the solid support and optical system are moved during interrogation. For example, a solid support being interrogated may be held stationary while the optical system is moved, or the solid support may be moved relative to a stationary optical system. Such movement may be accomplished using any of a variety of manipulation hardware or robotic set-ups, e.g., a stepper/feeder apparatus, and are well known in high performance printing technologies and in the semiconductor industry. For example, robotic systems may be used to pick up and re-orient a given solid support in order to interrogate different regions of the solid support, or make a previously unaccessible region (e.g., blocked by clips, support structure, or the like) of the solid support accessible. Such robotic systems are generally available from, e.g., Beckman, Inc., Tecan, Inc., Caliper Life Sciences, and the like.

[00129] In addition to the foregoing, it will be appreciated that the reagents in a given reaction of interest, including those reagents for which photo-induced damage is being mitigated in accordance with the invention, may be provided in any of a variety of different configurations. For example, they may be provided free in solution, or complexed with other materials, e.g., other reagents and/or solid supports. Likewise, such reagents may be provided coupled to beads, particles, nanocrystals or other nanoparticles, or they may be tethered to larger solid supports, such as matrices or planar surfaces. These reagents may be further coupled or complexed together with other reagents, or as separate reagent populations or even as individual molecules, e.g., that are detectably resolvable from other molecules within the reaction space. As noted above, whether a particular reagent is confined by virtue of structural barriers to its free movement or is chemically tethered or immobilized to a surface of a substrate, it will be described as being "confined." Further examples of such confined reagents include surface immobilized or localized reagents, e.g., surface immobilized or associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal

reflection microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. For example, in some preferred embodiments, one or more reagents in an assay system are confined within an optical confinement. Such an optical confinement may be an internal reflection confinement (IRC) or an external reflection confinement (ERC), a zero-mode waveguide, or an alternative optical structure, such as one comprising porous film with reflective index media or a confinement using index matching solids. More detailed descriptions of various types of optical confinements are provided, e.g., in International Application Publication No. WO/2006/083751, incorporated herein by reference in its entirety for all purposes.

[00130] The invention is generally applicable to any of a variety of optical assays that require substantial illumination and/or photoactivated conversion or excitation of chemical groups, e.g., fluorophores. For example, the compositions and methods provided herein may be used with fluorescence microscopy, optical traps and tweezers, spectrophotometry, fluorescence correlation spectroscopy, confocal microscopy, near-field optical methods, fluorescence resonance energy transfer (FRET), structured illumination microscopy, total internal reflection fluorescence microscopy (TIRF), etc. The methods provided herein may be particularly useful in assays that are negatively impacted, directly or indirectly, by prolonged exposure to illumination. Of particular interest are those assays that are impaired by the generation and/or accumulation of triplet-state forms or free radicals during illumination.

[00131] One particularly apt example of analyses that benefit from the invention are single-molecule biological analyses, including, inter alia, single molecule nucleic acid sequencing analyses, single molecule enzyme analyses, hybridization assays (e.g., antibody assays), nucleic acid hybridization assays, and the like, where the reagents of primary import are subjected to prolonged illumination with relatively concentrated light sources (e.g., lasers and other concentrated light sources, such as mercury, xenon, halogen, or other lamps) in an environment where photoconversion/excitation is occurring with its associated generation of products. In certain embodiments, the methods, compositions, and systems are used in nucleic acid sequencing processes that rely on detection of fluorescent or fluorogenic reagents. Examples of such sequencing technologies include, for example, SMRT™ nucleic acid sequencing (described in, e.g., U.S. Patent Nos. 6,399,335, 6,056,661, 7,052,847, 7,033,764, 7,056,676, 7,361,466, 7,416,844, the full disclosures of which are incorporated herein by reference in their entirety for all purposes), non-

real-time, or "one base at a time" sequencing methods available from, e.g., Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), Clonal Single Molecule Array™, and SOLiD™ sequencing. (See, e.g., Harris, et al. (2008) Science 320 (5872):106-9, incorporated by reference herein in its entirety for all purposes.) Such prolonged illumination can negatively impact (e.g., by introducing photo-induced damage) these reagents and diminish their effectiveness in the desired reaction.

III. Prevention of Photo-induced Damage

[00132] The methods provided herein are particularly useful in analyses that utilize very limited concentrations of reactants, such as single molecule detection/monitoring assays. As will be appreciated, in such reagent limited analyses, any loss, degradation, or depletion of a critical reagent will dramatically impact the analysis by further limiting the reagent, which not only can adversely effect the detectable signal, but may also directly impact the reaction being monitored, e.g., by changing its rate, duration, or product(s). For example, photo-induced damage can include a photoinduced change in a given reagent that reduces the reactivity of that reagent in the reaction, e.g., photobleaching of a fluorescent molecule, which diminishes or removes its ability to act as a signaling molecule. Also included in the term photo-induced damage are other changes that reduce a reactant's usefulness in a reaction, e.g., by making the reagent less specific in its activity in the reaction. Likewise, photo-induced damage includes undesired changes in a reagent that are caused by interaction of that reagent with a product of another photoinduced reaction, e.g., the generation of singlet oxygen during a fluorescence excitation event, which singlet oxygen may damage organic or other reagents, e.g., proteins. Photo-induced damage also includes downstream effects of damage to reactants, such as irreversible interactions between damaged reactants and other critical components of the reaction, e.g., reactive proteins or enzymes. For example, damage to an enzyme that catalyzes a reaction being monitored may cause a reduction in the rate of the reaction, in some cases stopping it altogether, or may reduce the duration or fidelity of the reaction.

[00133] As suggested by the foregoing, photo-induced damage generally refers to an alteration in a given reagent, reactant, or the like, that causes such reagent to have altered functionality in a desired reaction, e.g., reduced activity, reduced specificity, or a reduced ability to be acted upon, converted, or modified, by another molecule, that results from, either directly or indirectly, a photo-induced reaction, e.g., a photo-induced reaction creates a reactant that interacts with and causes damage to one or more other reactants. Typically, such photoreaction directly

impacts either the reactant of interest, e.g., direct photo-induced damage, or impacts a reactant within one, two or three reactive steps of such reactant of interest. Further, such photoreaction can directly impact the reaction of interest, e.g., causing a change in rate, duration, processivity, or fidelity of the reaction.

[00134] The amount of time an illuminated analysis may be carried out before photo-induced damage so substantially impacts the reactants to render the analysis non-useful is referred to as the "photo-induced damage threshold period." A photo-induced damage threshold period is assay-dependent, and is affected by various factors, including but not limited to characteristics of enzymes in the assay (e.g., susceptibility to photo-induced damage and the effect of such damage on enzyme activity/processivity), characteristics of the radiation source (e.g., wavelength, intensity), characteristics of the signal-generating molecule (e.g., type of emission, susceptibility to photo-induced damage, propensity to enter triplet state, and the effect of such damage on the brightness/duration of the signal), similar characteristics of other components of the assay. It can also depend on various components of the assay system, e.g., signal transmission and detection, data collection and analysis procedures, etc. It is well within the abilities of the ordinary practitioner to determine an acceptable photo-induced damage threshold period for a given assay, e.g., by monitoring the signal decay for the assay in the presence of a photodamaging agent and identifying a period for which the signal is a reliable measure for the assay. In terms of the invention, the photo-induced damage threshold period is that period of illuminated analysis during which such photo-induced damage occurs so as to reduce the rate or processivity of the subject reaction by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over the same reaction in the absence of such illumination. It is an object of the invention to increase the photo-induced damage threshold period, thereby increasing the amount of time reactions can proceed toward completion with minimal damage to the reactants, thereby lengthening the time in which the detectable signal is an accurate measure of reaction progression.

[00135] In some contexts, a "photo-induced damaged" reaction may be subject to spurious activity, and thus be more active than desired. In such cases, it will be appreciated that the photo-induced damage threshold period of interest would be characterized by that period of illuminated analysis during which such spurious activity, e.g., as measured by an increase in reaction rate, or an increase in non-specific reaction rate, is no more than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% over a non-illuminated reaction. In one non-limiting example, where a nucleic acid polymerase, by virtue of a photodamaging event, begins to incorrectly incorporate nucleotides

during template directed synthesis, such activity would impact the photo-induced damage threshold period as set forth above. In this case, the methods, devices, and systems of the invention would increase the photo-induced damage threshold period, thus increasing the amount of time the reaction could proceed before the above-described spurious activity occurred.

[00136] With reference to nucleic acid analyses, it has been observed that in template-directed synthesis of nucleic acids using fluorescent nucleotide analogs as a substrate, prolonged illumination can result in a substantial degradation in the ability of the polymerase to synthesize the nascent strand of DNA, as described previously, e.g., in U.S. Published Patent Application No. 20070161017, incorporated by reference herein in its entirety for all purposes. Damage to polymerase enzymes, template sequences, and/or primer sequences can significantly hinder the ability of the polymerase to process longer strands of nucleic acids. For example, reduction in the processivity of a polymerase leads to a reduction in read lengths for sequencing processes that identify sequence constituents based upon their incorporation into the nascent strand. As is appreciated in the art of genetic analysis, the length of contiguous reads of sequence directly impacts the ability to assemble genomic information from segments of genomic DNA. Such a reduction in the activity of an enzyme can have significant effects on many different kinds of reactions in addition to sequencing reactions, such as ligations, cleavages, digestions, phosphorylations, etc.

[00137] Without being bound to a particular theory or mechanism of operation, it is believed that at least one cause of photo-induced damage to enzyme activity, particularly in the presence of fluorescent reagents, results from the direct interaction of the enzyme with photo-induced damaged fluorescent reagents. Further, it is believed that this photo-induced damage of the fluorescent reagents (and possibly additional damage to the enzyme) is at least partially mediated by reactive intermediates (e.g., reactive oxygen species) that are generated during the relaxation of triplet-state fluorophores. One or both of the photo-induced damaged fluorescent reagents and/or reactive intermediates may be included in the overall detrimental effects of photo-induced damage.

[00138] In certain aspects, the invention is directed to methods, devices, and systems that reduce the amount of photo-induced damage to one or more reactants during an illuminated reaction, e.g., thereby improving the reaction, e.g., by increasing the processivity, rate, fidelity, processivity, or duration of the reaction. In particular, methods are provided that yield a reduction in the level of photo-induced damage and/or an increase in the photo-induced damage threshold period as compared to such reactions in the absence of such methods, devices, and systems. In

particular embodiments, such methods comprise subjecting an illuminated reaction to periods of non-illuminated during the course of the reaction, as described above, or by temporarily removing components of the reaction mixture that are believed to cause such damage, as described below.

[00139] As generally referred to herein, limited quantity reagents or reactants may be present in solution, but at very limited concentrations, e.g., less than 200 nM, in some cases less than 10 nM and in still other cases, less than 10 pM. In preferred aspects, however, such limited quantity reagents or reactants refer to reactants that are immobilized or otherwise confined within a given area or reaction site (e.g., a zero-mode waveguide), so as to provide limited quantity of reagents in that given area, and in certain cases, provide small numbers of molecules of such reagents within that given area, e.g., from 1 to 1000 individual molecules, preferably between 1 and 10 molecules. As will be appreciated, photo-induced damage of immobilized reactants in a given area will have a substantial impact on the reactivity of that area, as other, non-damaged reactants are not free to diffuse into and mask the effects of such damage. Examples of immobilized reactants include surface-immobilized or -localized reagents, e.g., surface-immobilized or -associated enzymes, antibodies, etc. that are interrogated upon the surface, e.g., through fluorescence scanning microscopy or scanning confocal microscopy, total internal reflectance microscopy or fluorometry, microscopy utilizing evanescent waves (see, e.g., U.S. Patent Publication Nos. 20080128627, filed August 31, 2007; 20080152281, filed October 31, 2007; and 200801552280, filed October 31, 2007, all of which are incorporated by reference in their entireties for all purposes), surface imaging, or the like. Various types of solid supports upon which one or more reactants can be immobilized are described above.

[00140] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is interrogated for one or more illuminated periods that cumulatively are less than a photo-induced damage threshold period, as set forth elsewhere herein. Such interrogation may occur coincident with or independent of interrogation of additional observation regions on a solid support containing the first observation region. In accordance with the present invention, the observation region typically includes confined reagents (e.g., enzymes, substrates, etc.) that are susceptible to photo-induced damage, and may include an area of a planar or other solid support upon which confined reagents are immobilized. Alternatively or additionally, the observation region may include a physical confinement that constrains the reagents that are susceptible to photo-induced damage, including, e.g., microwells, nanowells, planar surfaces that include hydrophobic barriers to confine reagents.

[00141] In accordance with certain aspects of the invention, a reaction of interest within a first observation region is intermittently interrogated under constant illumination by virtue of intermittent presence of detectable components of the reaction, wherein the presence of such detectable components has the potential to directly or indirectly cause photo-induced damage to one or more other reaction components. For example, a buffer comprising detectable components of a reaction can be temporarily replaced with a buffer comprising non-detectable versions of the same components of the reaction, thereby interrupting data acquisition for the reaction. When data acquisition is to be recommenced, the buffer comprising detectable component is substituted for the buffer comprising non-detectable components. This substitution of reaction components may be repeated multiple times to generate multiple sets of data collected at noncontiguous stages of the reaction. For example, such a substitution can occur at least about 2, 4, 6, 8, or 10 times during the course of the reaction.

[00142] In certain preferred embodiments, the detectable components are fluorescently-labeled components that can be damaged by exposure to excitation illumination, and can further cause damage to other reaction components, as described above. For example, a sequencing-by-incorporation reaction can be initiated in the presence of fluorescently-labeled nucleotides whose incorporation is indicative of the nucleotide sequence of the nascent strand synthesized by a polymerase, and by complementarity, of the template nucleic acid molecule. At a selected time point during the ongoing reaction, the labeled nucleotides can be removed and replaced with unlabeled nucleotides, for example, by buffer exchange. After a period of time during which data acquisition has been interrupted by the absence of signal from the ongoing reaction, the labeled nucleotides can be reintroduced to reinitiate data acquisition. The labeled nucleotides may be removed and reintroduced multiple times and for various lengths of time, as preferred by the ordinary practitioner. In this way, multiple noncontiguous sequence reads can be generated from a single nucleic acid molecule in real time.

[00143] The methods herein slow the accumulation of photo-induced damage to one or more reagents, and may therefore indirectly mitigate the impact of photo-induced damage in an ongoing reaction of interest. By way of example, methods that reduce exposure of a critical enzyme component to illumination radiation (e.g., by subjecting the reaction to periods of non-illumination or by temporarily removing a component of the reaction responsible for such damage) do not necessarily prevent the photo-induced damage to the enzyme component, but rather extend the photo-induced damage threshold period by slowing the accumulation of photo-induced damage in

the reaction mixture. Measurements of reduction of photo-induced damage as a result of implementation of intermittent illumination may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction subjected to constant illumination. Likewise, measurements of reduction of photo-induced damage as a result of temporary removal of reaction components responsible for such damage may be characterized as providing a reduction in the level of photo-induced damage as compared to a reaction in which such components are present throughout. Further, characterization of a reduction in photo-induced damage generally utilizes a comparison of reaction rates, durations, or fidelities, processivities, e.g., of enzyme activity, and/or a comparison of the photo-induced damage threshold period, between a reaction mixture subjected to such the methods and/or systems of the invention and a reaction mixture not so subjected.

[00144] In the case of the present invention, implementation of the methods, devices, and systems of the invention generally results in a reduction of photo-induced damage of one or more reactants in a given reaction, as measured in terms of "prevented loss of reactivity" in the system. Using methods known in the art, the amount of prevented loss of activity can at least 10%, preferably greater than 20%, 30%, or 40%, and more preferably at least 50% reduction in loss of reactivity or increase in processivity, and in many cases greater than a 90% and up to and greater than 99% reduction in loss of reactivity or increase in processivity. By way of illustration, and purely for the purpose of example, when referring to reduction in photo-induced damage as a measure of enzyme activity in the presence and absence of intermittent illumination, if a reaction included a reaction mixture having 100 units of enzyme activity that would, under constant illumination, yield a reaction mixture having only 50 units of activity, then a 10% reduction in photo-induced damage would yield a final reaction mixture of 55 units (e.g., 10% of the 50 units otherwise lost, would no longer be lost). Further, use of the invention is expected to increase the performance (e.g., processivity, duration, fidelity, rate, etc.) of a reaction whose performance is negatively impacted by constant exposure to illumination by at least about 2-, 5-, 10-, 20-, 30-, 50-, 80-, 100-, 500-, or 1000-fold over that achieved by the reaction under constant illumination. For example, it is a specific object of the instant invention to increase the processivity of a polymerase enzyme in a sequencing reaction to allow collection of data across a longer length of the template.

[00145] With regards to sequencing applications, the methods herein facilitate the scaffolding of nucleic acid sequences in reactions susceptible to photo-induced damage. For example, if the sequencing device has 1000 base pair average readlength under constant illumination, one could subject the reaction to illuminated periods timed to allow approximately 100 nucleotides to be

incorporated into the nascent strand of read, followed by non-illuminated periods timed to allow approximately 1000 nucleotides to be incorporated "in the dark." The sequence reads resulting from this experimental design would comprise about ten sequence reads of about 100 nucleotides each separated by gaps of about 1000 nucleotides each. If a plurality of sequencing reactions were carried out in this manner, and the illuminated periods were staggered appropriately, the reads from the plurality of reactions could be combined to provide nucleotide sequence data for the entire template nucleic acid. This would potentially allow sequence scaffolds to be built much more easily than can be done with short-read systems, enabling structural analysis of previously impossible-to-sequence sections of highly repetitive DNA, given the sequencing system is capable of long reads in the absence of photodamage.

IV. Software and Algorithm Implementations

[00146] The methods herein may operate with numerous methods for sequence alignment including those generated by various types of known multiple sequence alignment (MSA) algorithms. For example, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read using a reference sequence. In some embodiments in which a reference sequence is known for the region containing the target sequence, the reference sequence can be used to produce an MSA using a variant of the center-star algorithm. Alternatively, the sequence alignment may comprise one or more MSA algorithm-derived alignments that align each read relative to every other read without using a reference sequence ("*de novo* assembly routines"), e.g., PHRAP, CAP, ClustalW, T-Coffee, AMOS make-consensus, or other dynamic programming MSAs. Depending on the sequence-generating methods used, the determination of sequence alignment may also involve analysis of read quality (e.g., using TraceTuner™, Phred, etc.), signal intensity, peak data (e.g., height, width, proximity to neighboring peak(s), etc.), information indicative of the orientation of the read (e.g., 5'→3' designations), clear range identifiers indicative of the usable range of calls in the sequence, and the like. Additional algorithms and systems for sequence alignment are well know to those of skill in the art, and are described further, e.g., in G. A. Churchill, M. S. Waterman (1992) "The Accuracy of DNA Sequences: Estimating Sequence Quality," *Genomics* 14: 89-98; M. Stephens, et al. (2006) "Automating sequence-based detection and genotyping of SNPs from diploid samples," *Nat. Genet.*, **38**: 375-381; J. Hein (1989) *Mol. Biol. Evol.*, **6**: 649-668; U.S.S.N. 12/134,186, filed June 5, 2008; and U.S.S.N. 61/116,439, filed November 20, 2008.

[00147] A standard sequence alignment problem in the context of DNA sequencing is to align the sequence of a relatively short fragment (<2 kilobases) to a large target sequence. The assumption is made that this fragment represents a contiguous portion of DNA to be mapped to a single location on the reference sequence. (A "contiguous portion" to be mapped to a single location may contain small insertions and/or deletions and still be considered contiguous in this context.) With the further development of nucleic acid sequencing technologies (e.g., from Illumina, Inc. (San Diego, CA), Helicos BioSciences (Cambridge, MA), and Applied Biosystems, Inc. (Foster City, CA)) and mate-pair sequencing protocols (see, e.g., U.S. Patent Pub. No. 2006/0292611 A1, which is incorporated by reference herein in its entirety for all purposes), the alignment problem has been extended to align two fragments coming from the same read to the reference sequence using some knowledge of the expected mate-pair configuration (distance and orientation).

[00148] With regards to mate-paired reads, mapping two fragments with a distance constraint and orientation constraint has been treated by various short-read mapping algorithms, e.g., SOAP (Li, et al. (2008) Bioinformatics, **24**, 713-714); SOAPdenovo; and Maq, a set of programs that map and/or assemble fixed-length Solexa/SOLiD reads (SourceForge, Inc.). While these algorithms can handle simple cases of mate-pair alignment, which generally treat the specific problem of only two reads coming from a mate-paired sequence and use the distance constraint as a hard filter (i.e., if two reads are within *x* bp of each other and in the correct orientations, report them as a mate-pair hit), the methods provided herein are more general and can handle much more complex data sets, including those with multiple reads, those for which a reference sequence is or is not present, potential non-template sequence (e.g., adapter regions or linker portions described below), and complex distance and orientation constraints. Other programs are also available that attempt to generalize on top of the mapping and aligning performed by the programs described above. These include, e.g., Breakdancer, variationhunter, GASV, etc., which can handle more complex mappings, e.g., by clustering.

[00149] Real-time single molecule sequencing presents opportunities for obtaining much more complex sequence fragments from a single DNA sequencing read. Two examples are the reading of multiple discontiguous sequence fragments from a single long stretch of DNA using a pulsed or intermittent detection system (e.g., intermittent illumination) as described herein and the contiguous reading of forward, reverse and adapter fragments from a circular templates (SMRTbell™ templates; see e.g., U.S.S.N. 61/099,696, filed September 24, 2008; U.S. Patent Application No. 12/383,855, filed March 27, 2009 and U.S. Patent Application No. 12/413,258,

filed March 27, 2009, all of which are incorporated by reference herein in their entireties for all purposes). Further, methods for sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.

[00150] Certain aspects of the invention provide methods for optimally aligning such sequences to a reference sequence using knowledge of the molecular configuration and/or sequencing protocol used to generate the related sequence reads. In particular, methods are provided to address the general problem of mapping multiple fragments to a reference sequence with variable distance and orientation constraints.

[00151] Beginning with raw sequence data generated by a nucleic acid sequencing instrument (step 1), the sequence data is mapped to a target sequence (step 2) using a local alignment method which produces sub-optimal local alignments as well as the optimal alignment, for example, the Smith-Waterman algorithm. Another, more flexible example of a local alignment method is a chaining method using a method for aligning very short fragments to the target sequence (e.g., kmer-indexing, suffix trees, suffix arrays, etc.) and chaining the resulting hits back into longer chains of significant matches (see, e.g. D. Gusfield, Algorithms on Strings, Trees, and Sequences, Cambridge University Press: Cambridge, UK, 1997, which is incorporated by reference herein in its entirety for all purposes). The chains do not necessarily need to be refined by dynamic programming in order to be useful for the following algorithm, permitting a very fast algorithm. In certain embodiments, dynamic-programming refinement of the chain might improve the power (area under the ROC curve) of the algorithm.

[00152] The target sequence consists of the potential hypotheses for the molecular template in question. In the example of nucleic acid sequencing methods using iterative illumination for sequencing a shotgun fragment from a linear DNA sequence, the potential hypotheses are both orientations of the genome (since we do not know the original orientation of the fragment). In the example of sequencing of a SMRTbell™ template (e.g., see Example 1 herein), the hypotheses include both orientations of the genome and known adapter sequences. The parameters determining how many hits are reported for each local fragment can be varied to change the specificity and sensitivity of this algorithm. Figure 6 shows what these hits might look like for a SMRTbell™

template (represented as paths in the sequence alignment matrix, which is often called the dynamic-programming matrix, although it isn't necessary to use dynamic programming to find these paths).

[00153] After the potential local alignments have been enumerated, a weighted directed graph is constructed with each local alignment represented as a node in the graph (step 3). The edges are drawn between nodes if they represent a potential reconstruction of the original molecular template using knowledge of the expected molecular configuration. The directed connection of an alignment path A to an alignment path B is interpreted as "The target sequence represented by B could follow the target sequence represented by A in the original molecule." For example, if a linear single-stranded DNA molecule is being sequenced by a method that uses iterative illumination, then fragments from opposite orientations would not be expected to be connected (unless the linear single-stranded DNA molecule also included oppositely oriented sequences, e.g., as in the case of a linearized SMRTbell™ template.) In general, fragments that represent the same stretch of the sequencing read but that align to different regions on the target sequence would not be connected. Aside from these examples, the rules for connecting nodes should be fairly loose to permit exploration of weak possibilities that gain significance when all the evidence (e.g. all the sequence reads) are considered. The assignment of edge weights handles the proper weighting of the likelihood of these edges, and the speed of the algorithm can be tuned by optimizing the pruning of highly unlikely edges. As usual this represents a tradeoff between speed and sensitivity.

[00154] Weights are assigned to connections ($A \rightarrow B$) in the graph representing the log-likelihood that target fragment A is followed by target fragment B in the original molecule.

$$w(A \rightarrow B) = -\log P(B|A)$$

The conditional probability $P(B\ A)$ encodes the knowledge of the possible molecular configurations and the alignment significance of B.

$$P(B|A) = f(B)g(A,B)$$

where f is a measure of alignment significance (either theoretical or empirically obtained) and g encodes the physical constraints representing the allowed molecular configurations.

[00155] For example, in the context of sequencing using iterative illumination the following may be known: the time between the end of one fragment and the beginning of the next fragment is 200 seconds. If the polymerase incorporates bases with an average rate of 4bp/sec with a standard deviation of 1bp/sec, it can be hypothesized that the probability of target fragment 2 following target fragment 1 is determined by the distance between these fragments on the target and a normal probability:

$$g(A,B)=\frac{1}{\sqrt{2\pi(200)}}\exp\left[-(d-800)/2(200)^2\right]$$

[00156] In a SMRTbell™ template example, knowledge of the expected insert size and the observed distance and orientation between fragments would be used to weight the likelihood that these two fragments could come from a correctly generated SMRTbell™ template. This weight could include the expected rate of the polymerase as well and rules for the orientation of fragments with respect to each other and their distance apart in the original read. For example, while it may be expected that two forward fragments mapping to the same region in the target genome potentially come from multiple passes around a SMRTbell™ template molecule, those fragments would not be expected to be immediately adjacent in sequencing time. The weighting function would account for the proper amount of expected time between such fragments (i.e. the elapsed time would be expected to be long enough to include two adapter sequences and a reverse sequence).

[00157] In general, the weighting function could be arbitrarily complex and tuned to empirically observed relationships between sequencing fragments given the available knowledge (distance between fragments on the target sequence, sequencing time between fragments, expected length of the template, etc.). For example, the empirical probability distributions might be observed to exhibit longer tails than a Gaussian probability model might predict. The use of a conditional log-likelihood for the assignment of edge weights is motivated by the following logic. In a graph of possible local alignments it is desirable to find a highly likely path that best explains the observed data. Consider a path through three nodes *A*, *B*, and *C*, with *P(ABC)* being the probability that *ABC* is the correct assignment:

$$\begin{aligned}P(ABC)&=P(C|AB)P(B|A)P(A)\\&\approx P(C|B)P(B|A)P(A)\end{aligned}$$

where the last approximation is justified by the observation that the constraints between allowable assignments to the target sequence are typically local in nature. Generalizing this formula for a path $a_1,\ldots,a_N$ and taking the negative logarithm of both sides gives

$$-\log P(a_1\ldots a_N)=-\sum_{i=1}^{N-1}\log P(a_{i+1}|a_i)-\log P(a_1)$$

[00158] It is apparent that the edge weights are additive if we use log-likelihood and we can use standard shortest-path algorithms for directed graphs to find the optimal path. A hypothetical directed graph is illustrated in Figure 7. This graph corresponds to the situation depicted by the alignments pictured above. Heavier lines correspond to more likely paths with the optimal path

shown in blue. Dashed lines represent forbidden transitions. Not all paths are considered in the illustration to avoid clutter in the presentation. The general formula listed above includes a "one-body" term $P(a_i)$ for the starting node in each path that weights the probability that this initial alignment is correct. To accommodate this probability in a path-finding algorithm we add a pseudo-source s to the graph which connects to every possible node (not shown in the graph above). The edge weight connecting the pseudo-source with a node a_i is $-\log P(a_i)$. This allows the use of a conventional single-source shortest-path algorithm starting from the pseudo-source. The desired probability $P(a_i)$ can come from a measure of alignment significance (theoretical or empirically determined) or could be set uniformly across all alignments to allow the path logic to determine the best path assignment, independent of the relative value of the starting points. It is anticipated that a threshold will be required here to only allow edges between the pseudo-source and nodes for highly likely alignments; otherwise the shortest path algorithm in the next step will not give the desired path.

[00159] After construction of the weighted directed graph, the shortest path to each node is determined (step 4). The graph is directed and acyclic (DAG) so we can use the standard shortest-path DAG algorithm (see T.H. Cormen, CE Leiserson, RL Rivest, Introduction to Algorithms, MIT Press: Cambridge, Massachusetts, 1990). This algorithm scales as O(V+E) and should be very quick for these graphs. After the shortest path to each node is determined, the paths need to be ranked to declare the best assignment. It is suggested that the best metric would be a measure which rewards paths that explain more of the sequenced read (longer paths) with high likelihood. One such metric would be the normalized negative log-likelihood: dividing the total weight of the path by the number of bases in the sequenced read explained by this path. For more complicated graphs or edge-weight assignments, Dijkstra's algorithm, the Bellman-Ford algorithm, or the A* algorithm could be applied. Other algorithms that may also be used include, but are not limited to the Floyd-Warshall algorithm.

[00160] For noisy sequence data it is likely that the local alignments found in step 2 will occasionally overlap with each other in the sequenced read even though it is physically impossible for such overlaps to occur in a perfect system (unless there has been a rearrangement relative to the reference genome). As such, some amount of slack must be allowed in the edge assignment logic in step 3 to account for not knowing the precise boundaries of each local alignment. Once the best physical model explaining the observed read is determined, the boundaries of the local alignments can be refined to reflect the physical necessity that each base in the sequenced read can only be

represented in one local alignment. It is also desirable to explain all of the bases in between the local alignments that haven't been assigned in the graph. One straightforward approach to refinement would be to construct the perfect model of the sequence and to realign the sequenced read to this sequence. This refinement algorithm would preserve physical constraints (each base in the sequenced read can only be explained by one location in the template) and would assign all bases between the extremal nodes in the optimal path.

[00161] Certain aspects of the software and algorithm implementations described herein may be varied or altered without departing from the spirit and scope of the invention. For example, with regards to algorithm seeding, many algorithms can be applied for the original determination of sub-optimal local alignments (step 2). Conventional examples include FASTA, BLAST, or Smith-Waterman. It is expected that the best benefit will be obtained from using short-sequence alignment algorithms (suffix array, suffix tree, Boyer-Moore, Rabin-Karp, kmer-indexing, and the like) followed by chaining to establish regions of significant matches. An advantage of the algorithm described here is that it does not require dynamic-programming refinement of the resulting chains and therefore can be quite fast, however it is expected that using dynamic programming to refine the chains in step 2 could increase the power of the algorithm.

[00162] With regards to graph construction, there will be advantages to tuning the logic of edge assignments to keep the size of the graph manageable. It is possible that steps 2 and 3 might be combined to in a greedy fashion to focus the potentially slow step 2 into productive areas of the graph. For example, if a particularly strong hit is found early in step 2, then it may be beneficial to search for sub-optimal hits only in this local vicinity, knowing that this strong hit should be in the final solution. Tuning of the graph construction might include thresholds, below which edges are not created. Further, there are multiple parameters (minimum chain length, minimum probability for edge assignment, relative weighting of length vs. accuracy, etc.) which can be exposed and tuned in this algorithm to maximize the sensitivity and specificity of the algorithm for a given scenario.

[00163] With regards to determination of the distance a polymerase travels between reads, various strategies are provided that are more sophisticated than estimation based upon the rate of incorporation and the time between detection (e.g., illuminated) periods. In certain embodiments, the distribution of the base pair distance travelled by a polymerase during a non-detection period is called $p(x)$. The distribution of enzyme velocities, $p(v)$, is estimated by aligning observed reads to a reference sequence, and this distribution is represented as the number of reference bases per unit time. There is a length of time, τ, over which measurement of the instantaneous rate is not

independent. While this method of determining the distance the polymerase travels during a non-detection period should not be overly sensitive to non-independent estimation of the polymerase rate, it is likely to strive for independent measurements of the rate. The distance τ can be estimated from an exponential fit to the auto-correlation function $\langle \delta v(t) \delta v(t-\Delta t) \rangle$, and $v(t)$ tabulated across the aligned sequence at increments of τ.

[00164] Where multiple single polymerase enzymes are being observed simultaneously, e.g., each being optically resolvable from every other on a single array, the $p(v)$ for each is preferably determined independently for each enzyme. Further, information regarding rare but extended events, such as polymerase "stalling" on the template, can be measured across a larger data set. For example, the statistics of stalls can be determined by aggregating rate measurements across an entire array. Where a stall distribution is characterized by a "long tail" corresponding to multi-exponential behavior of IPD distribution, such a distribution of polymerization rates can be extended for stalls longer than the observed reaction by fitting the long-tail behavior to an appropriate functional form, e.g., using a single-exponential parametric model or other physically motivated model (e.g., multi-exponential, stretched exponential, power-law, etc.) In certain preferred embodiments, the following representation of a "per-enzyme" $p(v)$ is used:

$$p(v) = \frac{f(v)p_{enzyme}(v) + [1 - f(v)]p_{array}(v)}{\int f(v')p_{enzyme}(v') + [1 - f(v')]p_{array}(v')dv'}$$

where $f(v)$ is an interpolating function designed to retain information about the zero-velocity tail of the global $p(v)$ distribution while taking the estimate of the polymerase velocity dynamics(e.g., the dominant high velocity mode) from the specific enzyme. Such an interpolating function is:

$$f(v) = \frac{1 + erf(v / v_0)}{2}$$

where v_0 is a scale parameter to be chosen based on experience (but optionally fixed). Alternatively, the average of the empirical $p_{enzyme}(v)$ and $p_{array}(v)$ can be used. This approach can be motivated by a Bayesian approach to density estimation. Other kernel density and Bayesian methods can be suggested. Alternatively or in addition, the robustness of $p(v)$ to conditions and daily phenomenon can be explored and used to estimate $p(v)$ more globally, e.g., using one or more weekly control experiments.

[00165] Given the lack of a known reference sequence for *de novo* assembly, several alternative ways to formulate $p(v)$ are provided as follows. For example, in a first embodiment a control template (essentially a proxy reference sequence) can be subjected to sequencing, e.g., in the

same reaction as the *de novo* sample or in an identical reaction. The observed velocity for the sequencing reactions would be measured based upon alignments of the reads from the control template to its known sequence. Typically, a per-enzyme correction would not be available for the *p(v)* and *p(v)* would default to an array-averaged *p(v)*. In a second embodiment, a previously determined *p(v)* from experiments using a known reference sequence can be used, e.g., where the previous experiments were performed under the same conditions as the *de novo* experiments. In a third embodiment, *p(v)* is estimated by using quality information/metrics to screen for the most likely "true" calls, and restricting the estimates of *v* to regions containing those calls. In a fourth embodiment, where error is low, the called base rate and reference base rate converge to the same rate, and measurements of p(v) without knowledge of the reference become substantially reliable. Further, even if they do not fully converge, they can still be used to accurately infer *p(v)*,as long as the called base rate is predictably higher/lower than the reference base rate. Yet further, the measurement of *p(v)* when a reference sequence is not available can benefit from a detailed look at the probability model which is available from an algorithm like a CRF. That is, *p(v)* can be tabulated using a weighted sum over paths through a CRF probability model.

[00166] As will be clear to the ordinary practioner based upon the teachings herein, this framework extends naturally to the measurements of other potentially systematic variations in *p(v)* across an array, e.g., even where a single reaction mixture is applied to the entire array. For example, the local temperature of the reaction environment can vary systematically across an array of reactions. The average and variation in the rates of polymerase enzymes on the array would likely have a dependence on this hidden variable. Where the functional form of the temperature dependence is known, the measurement of *p(v)* can be stabilized across the array by modeling a de novo *p(v)* as $p_{cond}(v) + p_{x,y}(v)$ where *x,y* are geometrical variables defining the location on the array. Further, *p(v)* has been found to be somewhat variable over time. As such, in certain embodiments a model of *p(v;t)* is developed using an appropriate model for the evolution of *p(v)* over time.

[00167] Once a representative distribution of velocities *p(v)* has been obtained for a given read from a given reaction, the expected travel distance in the non-detection period can be expressed as:

$$p\left(\frac{x}{\tau}\right) = IL\left[L[p(v)]^{t/\tau} \frac{1 - L[p(v)]}{s} \right]$$

where L[] and IL[] stand for the Laplace and inverse Laplace transform, respectively. A similar result is derived in Svoboda, et al. (PNAS 91:11782 (1994)) and readily follows from considering

the pdf of a sum of random variables. Optionally, in certain embodiments density estimation techniques (e.g., kernel density estimation, etc.) are useful when modeling *p(v)* since they can smooth the resulting numerical calculations in the Laplace and inverse Laplace transform.

[00168] Knowledge of the complete distribution has several advantages over the commonly applied Gaussian approximation. For example, knowledge of the complete distribution of insert lengths is very desirable when using a Bayesian framework approach to detect structural variation. (See, e.g., Bashir, et al. (2008) PLoS Comput. Biol. 4:51; Hormozdiari, et al. (2009) Genome Res. 19:1270; and Lee, et al. (2008) Bioinformatics 24:59.) While Bashir, et al. does not strictly follow a Bayesian approach, the geometric approach described in the paper can be straightforwardly modified to incorporate an actual posterior instead of the boxcar posterior assumed in the paper. Further, during mapping of noncontiguous reads to a genome where they are expected to be concordant (*i.e.*, not a structural variation), it is useful to consider the known distribution when judging the significance of the resulting alignments between the observed reads and the genomic sequence. In addition, when clustering noncontiguous reads that scaffold contigs in a *de novo* assembly, a path of Bayesian significance can be followed that is very similar to that followed in the structural variation case discussed *supra*.

[00169] In further embodiments, the determination of the distance a polymerase travels between reads is performed using an algorithm based on a simulation approach rather than the exact analytical result used in the algorithm described above. This method relies on Monte Carlo sampling from a distribution, which allows a better extension to arbitrary empirical distributions. It also lacks the difficult computations of numerical Laplace and inverse Laplace transforms, and permits calculation of distances traveled during non-detection periods when the underlying kinetic processes have multi-phasic kinetics, e.g., the presence of long stalls.

[00170] This approach aims to calculate the distribution of the distance *x* travelled by an enzyme during a time *t* during which it was not being observed (e.g., during a non-detection period). In some embodiments, a distribution of local rates, *p(v)*, is estimated, where the definition of "local" is set by the correlation length of the rate autocorrelation function, e.g.:

$$\langle \delta v(t)\delta v(t+\Delta)\rangle \sim \exp\left(\frac{-\Delta}{\tau_{corr}}\right)$$

Given a local rate distribution and an assumption that independent identically distributed (i.i.d.) draws can be made from this distribution, one approach to calculating the distribution is as follows. First, draw $N = t/\tau_{corr}$ velocities from *p(v)*; and subsequently sum them and record them as an

estimate of x/τ_{corr}. Repeat the process M times, with the optimal choice of M dependent on the desired level of precision for estimation of the *p(x)* distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000.

[00171] In some embodiments in which the enzyme system is not well explained by a single kinetic process or cycle (as in the case of observed stalling behavior), above-described rate autocorrelation function and the i.i.d. assumption will be violated. As such, a probability model having a richer structure is preferably used. One such probability model is a Hidden Markov Model (HMM). Figure 15 provides an exemplary illustration of an HMM for modeling a simple "pausing" vs. "sequencing" system. Where the kinetics of the pausing state can be well described by a single-exponential, this model is expected to describe the observed distribution of local velocities. The single-exponential assumption is implicit in the state structure of the model since the amount of time spent in the pause state will be a geometric distribution with mean p/(1-p) [*i.e.*, the observed stall times will have to be added to this model]. If the stall kinetics are multi-phasic, then more "dark states" will have to be added to this model. Further, the model shown in Figure 15 can actually be treated as a Markov Model and not a Hidden Markov Model without much loss of generality because the "pause" state is not actually hidden due to the fact that the data collected during the pause state is highly distinguishable from the data collected during the sequencing state. As such, the general HMM apparatus is not necessary. The model in Figure 15 can be used to simulate the distribution of local velocities when there is a long-term pause or stall phase present in the reaction data kinetics. S0 is the start state, and there is no explicit end state since this model is used as a generative model and it is assumed that it is run forward for a prescribed number of steps. The qualities P(P→S) and P(S→P) represent exit from a stalled state and entry into a stalled state, respectively. These qualities can be measured by an EM algorithm or they can be quickly estimated by physical observables. $P_{P\to S} = 1/\left(1+\frac{\tau_{stall}}{\tau_{corr}}\right)$ and P(S→P) is the frequency of stall starts per τ_{corr}. (Example parameters are τ_{stall}=80 seconds; τ_{corr} = 10 seconds; and P(S→P) = 1/24.) The simulation estimate of *p(x)* can now be produced using the procedure outlined above in which $N = t/\tau_{corr}$ velocities are drawn from *p(v)*; and they are subsequently summed and recorded as an estimate of x/τ_{corr}. The process is repeated M times, with the optimal choice of M dependent on the desired level of precision for estimation of the *p(x)* distribution. In certain preferred embodiments, M is between about 1000 and about 5000, e.g. at least about 1000, 2000, 3000, or 4000, or is about 5000. Figure

16 shows exemplary simulated applications of this method. Figure 16A shows a sample of velocities drawn from the HMM in Figure 15 with the parameters P(S→P) = 1/24; P(P→S) = 1/11; and $p(v)$~Gamma(48,0.25). Figure 16B illustrates a resulting histogram of local velocities. Figure 16C provides an estimated distance traveled during a 1300 second non-detection period, which is calculated by sampling 2000 estimates from the HMM model.

[00172] Figure 17 provides an illustrative example of two observed histograms of distances traveled during a non-detection period. The influence of pause/stall behavior can be seen in the heavy-left tailing of both distributions.

[00173] While the simulation method in which i.i.d. draw assumption is valid is more general and can treat arbitrary $p(v)$ and more complex models for non-sequencing states, the two-state model using the HMM can be treated analytically. The result of this is:

$$p(x/\tau_{corr}) = \sum_{N_S=0}^{N} \pi_{N_S}(x) p_N(N_S)$$

where $\pi_{N_S}(x)$ is the distribution of the sum of N_S variables drawn from $p(v)$. For the general case, this distribution is given by the Laplace transform approach presented above. For $p(v)$~Normal(μ,σ), this distribution is distributed as Normal(Nsμ,√Nsσ). For $p(v)$~Gamma(k,θ), this distribution is distributed as Gamma(Nsk,θ). $P_N(N_S)$ is the number of cycles spent in the sequencing state if we observe N cycles from the Markov process in Figure 15. The expression for this is described in Pedler, et al. (1971) J. Appl. Prob. 8:381, which is incorporated herein by reference in its entirety for all purposes.

[00174] As will be clear to one of ordinary skill in the art upon review of the teachings herein, these methods can be readily extended to the non-detection period estimations of procession by other cyclical biological reactions, such as the action of reverse transcriptase or the synthesis of proteins by a ribosome complex, e.g., and certain preferred embodiments of such reactions are further described in U.S.S.N. 12/767,673, filed April 26, 2010; and U.S.S.N. 12/813,968, filed June 11, 2010, the disclosures of which are incorporated herein by reference in their entireties for all purposes. Further, the simulation model described above is not restricted to simple two-state kinetics, and the use of $p(v)$ is not restricted to analytical models. In fact, in certain embodiments, empirical estimates are preferably used.

[00175] Although useful in certain preferred embodiments of the invention, certain algorithms as presented above do not easily handle the case where the template does not match a physically-motivated expected model. A relevant example of such a case is when the template

contains a genomic structural variation (SV), such as translocation, whereby two fragments which are correctly adjacent in the template are located very far apart in the reference genome. Such structural variation cases are best handled in the context of the current algorithm by reporting the confidence of an observed path and reporting situations when no physically expected path seems to fit the observed data. In general, the detection of structural variation requires the presence of multiple highly significant local alignments which can be identified as significantly overturning the null hypothesis of matching the genomic ordering of fragments with their own individual merit. Nevertheless, with molecular redundant sequencing such as SMRTbell™ template sequencing the current algorithm can be adapted to improve the ability to identify an SV event. Such a modification could be a feedback approach which allows modification of the linking constraints in step 3 to allow very far separations on the target sequence when the individual alignments are very significant. Only one such highly-significant pair would be needed to enable the rescue of less significant partial matches that support the same SV hypothesis.

[00176] The software and algorithm implementations provided herein are particularly suited for transforming sequence read data generated from various sequencing technologies (e.g., sequencing-by-synthesis, intramolecular redundant sequencing, Sanger sequencing, capillary electrophoretic sequencing, pyrosequencing, ligase-mediated sequencing, etc.) into consensus sequence data that provides a representation of the actual nucleotide sequence of the template nucleic acid that was subjected to the sequencing reaction(s) from which the sequence read data was generated. The software and algorithm implementations provided herein are preferably machine-implemented methods. The various steps recited herein are preferably performed via a user interface implemented in a machine that comprises instructions stored in machine-readable medium and a processor that executes the instructions. The results of these methods are preferably stored on a machine-readable medium, as well. Further, the invention provides a computer program product comprising a computer usable medium having a computer readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention.

[00177] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and

computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data).

[00178] While described with reference to certain specific applications above, it will be understood that these methods are also applicable to other types of complex data sets, and the invention should not be limited to only the specific examples provided herein. Other applications of the instant methods will be clear to those of ordinary skill in the art and are considered to be additional aspects of the instant invention.

V. Devices and Systems

[00179] The invention also provides systems that are used in conjunction with the compositions and methods of the invention in order to provide for intermittent detection of analytical reactions. In particular, such systems typically include the reagent systems described herein, in conjunction with an analytical system, e.g., for detecting data from those reagent systems. For example, a sequencing reaction may be subjected to intermittent illumination, and the sequencing system may include the system components provided with or sold for use with commercially available nucleic acid sequencing systems, such as the Genome Analyzer System available from Illumina, Inc., the GS FLX System, available from 454 Life Sciences, or the ABI 3730 System available from Life Technologies, Inc.

[00180] In certain preferred embodiments, reactions subjected to intermittent illumination are monitored using an optical system capable of detecting and/or monitoring interactions between reactants at the single-molecule level. Such an optical system achieves these functions by first generating and transmitting an incident wavelength to the reactants, followed by collecting and analyzing the optical signals from the reactants. Such systems typically employ an optical train that directs signals from the reactions to a detector, and in certain embodiments in which a plurality of reactions is disposed on a solid surface, such systems typically direct signals from the solid surface (e.g., array of confinements) onto different locations of an array-based detector to simultaneously detect multiple different optical signals from each of multiple different reactions. In particular, the optical trains typically include optical gratings or wedge prisms to simultaneously direct and separate signals having differing spectral characteristics from each confinement in an array to different locations on an array based detector, *e.g.*, a CCD, and may also comprise additional optical transmission elements and optical reflection elements.

[00181] An optical system applicable for use with the present invention preferably comprises at least an excitation source and a photon detector. The excitation source generates and transmits incident light used to optically excite the reactants in the reaction. Depending on the intended application, the source of the incident light can be a laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source. Further, the excitation light may be evanescent light, e.g., as in total internal reflection microscopy, certain types of waveguides that carry light to a reaction site (see, e.g., U.S. Application Pub. Nos. 20080128627, 20080152281, and 200801552280), or zero-mode waveguides, described below. Where desired, more than one source can be employed simultaneously. The use of multiple sources is particularly desirable in applications that employ multiple different reagent compounds having differing excitation spectra, consequently allowing detection of more than one fluorescent signal to track the interactions of more than one or one type of molecules simultaneously. A wide variety of photon detectors or detector arrays are available in the art. Representative detectors include but are not limited to optical reader, high-efficiency photon detection system, photodiode (*e.g.* avalanche photo diodes (APD)), camera, charge couple device (CCD), electron-multiplying charge-coupled device (EMCCD), intensified charge coupled device (ICCD), and confocal microscope equipped with any of the foregoing detectors. For example, in some embodiments an optical train includes a fluorescence microscope capable of resolving fluorescent signals from individual sequencing complexes. Where desired, the subject arrays of optical confinements contain various alignment aides or keys to facilitate a proper spatial placement of the optical confinement and the excitation sources, the photon detectors, or the optical train as described below.

[00182] The subject optical system may also include an optical train whose function can be manifold and may comprise one or more optical transmission or reflection elements. Such optical trains preferably encompass a variety of optical devices that channel light from one location to another in either an altered or unaltered state. First, the optical train collects and/or directs the incident wavelength to the reaction site (e.g., optical confinement). Second, it transmits and/or directs the optical signals emitted from the reactants to the photon detector. Third, it may select and/or modify the optical properties of the incident wavelengths or the emitted wavelengths from the reactants. In certain embodiments, the optical train controls an on/off cycle of the illumination source to provide illuminated and non-illuminated periods to one or more illuminated reaction sites. Illustrative examples of such optical transmission or reflection elements are diffraction gratings, arrayed waveguide gratings (AWG), optic fibers, optical switches, mirrors (including dichroic

mirrors), lenses (including microlenses, nanolenses, objective lenses, imaging lenses, and the like), collimators, optical attenuators, filters (*e.g.*, polarization or dichroic filters), prisms, wavelength filters (low-pass, band-pass, or high-pass), planar waveguides, wave-plates, delay lines, and any other devices that guide the transmission of light through proper refractive indices and geometries. One example of a particularly preferred optical train is described in U.S. Patent Pub. No. 20070036511, filed August 11, 2005, and incorporated by reference herein in its entirety for all purposes.

[00183] In a preferred embodiment, a reaction site (e.g., optical confinement) containing a reaction of interest is operatively coupled to a photon detector. The reaction site and the respective detector can be spatially aligned (*e.g.*, 1:1 mapping) to permit an efficient collection of optical signals from the reactants. In certain preferred embodiments, a reaction substrate is disposed upon a translation stage, which is typically coupled to appropriate robotics to provide lateral translation of the substrate in two dimensions over a fixed optical train. Alternative embodiments could couple the translation system to the optical train to move that aspect of the system relative to the substrate. For example, a translation stage provide a means of removing a reaction substrate (or a portion thereof) out of the path of illumination to create a non-illuminated period for the reaction substrate (or a portion thereof), and returning the substrate at a later time to initiate a subsequent illuminated period. An exemplary embodiment is provided in U.S. Patent Pub. No. 20070161017, filed December 1, 2006.

[00184] In particularly preferred aspects, such systems include arrays of reaction regions, e.g, zero-mode waveguide arrays, that are illuminated by the system, in order to detect signals (e.g., fluorescent signals) therefrom, that are in conjunction with analytical reactions being carried out within each reaction region. Each individual reaction region can be operatively coupled to a respective microlens or a nanolens, preferably spatially aligned to optimize the signal collection efficiency. Alternatively, a combination of an objective lens, a spectral filter set or prism for resolving signals of different wavelengths, and an imaging lens can be used in an optical train, to direct optical signals from each confinement to an array detector, *e.g.*, a CCD, and concurrently separate signals from each different confinement into multiple constituent signal elements, *e.g.*, different wavelength spectra, that correspond to different reaction events occurring within each confinement. In preferred embodiments, the setup further comprises means to control illumination of each confinement, and such means may be a feature of the optical system or may be found elsewhere is the system, e.g., as a mask positioned over an array of confinements. Detailed

descriptions of such optical systems are provided, e.g, in U.S. Patent Pub. No. 20060063264, filed September 16, 2005, which is incorporated herein by reference in its entirety for all purposes.

[00185] The systems of the invention also typically include information processors or computers operably coupled to the detection portions of the systems, in order to store the signal data obtained from the detector(s) on a computer readable medium, e.g., hard disk, CD, DVD or other optical medium, flash memory device, or the like. For purposes of this aspect of the invention, such operable connection provide for the electronic transfer of data from the detection system to the processor for subsequent analysis and conversion. Operable connections may be accomplished through any of a variety of well known computer networking or connecting methods, e.g., Firewire®, USB connections, wireless connections, WAN or LAN connections, or other connections that preferably include high data transfer rates. The computers also typically include software that analyzes the raw signal data, identifies signal pulses that are likely associated with incorporation events, and identifies bases incorporated during the sequencing reaction, in order to convert or transform the raw signal data into user interpretable sequence data (See, e.g., Published U.S. Patent Application No. 2009-0024331, the full disclosure of which is incorporated herein by reference in its entirety for all purposes).

[00186] Exemplary systems are described in detail in, e.g., U.S. Patent Application No. 11/901,273, filed September 14, 2007 and U.S. Patent Application No. 12/134,186, filed June 5, 2008, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

[00187] Further, as noted above, the invention provides data processing systems for transforming sequence read data into consensus sequence data. In certain embodiments, the data processing systems include machines for generating sequence read data by interrogating a template nucleic acid molecule. In certain preferred embodiments, the machine generates the sequence read data using a sequencing-by-synthesis technology, as described elsewhere herein, but the machine may generate the sequence read data using other sequencing technologies known to those of ordinary skill in the art, e.g., pyrosequencing, ligation-mediated sequencing, Sanger sequencing, capillary electrophoretic sequencing, etc. Such machines and methods for using them are available to the ordinary practioner.

[00188] The sequence read data generated is representative of the nucleotide sequence of the template nucleic acid molecule only to the extent that a given sequencing technology is able to generate such data, and so may not be identical to the actual sequence of the template nucleic acid

molecule. For example, it may contain a deletion or a different base at a given position as compared to the actual sequence of the template, e.g., when a base call is missed or incorrect, respectively. As such, it is beneficial to generate redundant sequence read data, and the methods described herein provide manipulations and computations that transform redundant sequence read data into consensus sequence data that is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Redundant sequence read data comprises multiple reads, each of which includes at least a portion of sequence read that overlaps with at least a portion of at least one other of the multiple reads. As such, the multiple reads need not all overlap with one another, and a first subset may overlap for a different portion of the template nucleic acid sequence than does a second subset. Such redundant sequence read data can be generated by various methods, including repeated sequencing of a single nucleic acid template, sequencing of multiple identical nucleic acid templates, or a combination thereof.

[00189] In another aspect, the data processing systems can include software and algorithm implementations provided herein, e.g. those configured to transform redundant sequence read data into consensus sequence data, which, as noted above, is generally more representative of the actual sequence of the template nucleic acid molecule than sequence read data from a single read of a single template nucleic acid molecule. Further, the transformation of the redundant sequence read data into consensus sequence data identifies and negates some or all of the single-read variation between the multiple reads in the redundant sequence read data. As such, the transformation provides a representation of the actual nucleotide sequence of the nucleic acid template from which redundant sequence read data is generated that is more accurate than a representation based on a single read.

[00190] The software and algorithm implementations provided herein are preferably machine-implemented methods, e.g., carried out on a machine comprising computer-readable medium configured to carry out various aspects of the methods herein. For example, the computer-readable medium preferably comprises at least one or more of the following: a) a user interface; b) memory for storing redundant sequence read data; c) memory storing software-implemented instructions for carrying out the algorithms for transforming redundant sequence read data into consensus sequence data; d) a processor for executing the instructions; e) software for recording the results of the transformation into memory; and f) memory for recordation and storage of the resulting consensus sequence read data. In preferred embodiments, the user interface is used by the

practitioner to manage various aspects of the machine, e.g., to direct the machine to carry out the various steps in the transformation of redundant sequence read data into consensus sequence data, recordation of the results of the transformation, and management of the consensus sequence data stored in memory.

[00191] As such, in preferred embodiments, the methods further comprise a transformation of the computer-readable medium by recordation of the redundant sequence read data and/or the consensus sequence data generated by the methods. Further, the computer-readable medium may comprise software for providing a graphical representation of the redundant sequence read data and/or the consensus sequence read data, and the graphical representation may be provided, e.g., in soft-copy (e.g., on an electronic display) and/or hard-copy (e.g., on a print-out) form.

[00192] The invention also provides a computer program product comprising a computer-readable medium having a computer-readable program code embodied therein, the computer readable program code adapted to implement one or more of the methods described herein, and optionally also providing storage for the results of the methods of the invention. In certain preferred embodiments, the computer program product comprises the computer-readable medium described above.

[00193] In another aspect, the invention provides data processing systems for transforming sequence read data from one or more sequencing reactions into consensus sequence data representative of an actual sequence of one or more template nucleic acids analyzed in the one or more sequencing reactions. Such data processing systems typically comprise a computer processor for processing the sequence read data according to the steps and methods described herein, and computer usable medium for storage of the initial sequence read data and/or the results of one or more steps of the transformation (e.g., the consensus sequence data), such as the computer-readable medium described above.

[00194] As shown in Figure 9, the system 900 includes a substrate 902 that includes a plurality of discrete sources of chromophore emission signals, e.g., an array of zero-mode waveguides 904. An excitation illumination source, e.g., laser 906, is provided in the system and is positioned to direct excitation radiation at the various signal sources. This is typically done by directing excitation radiation at or through appropriate optical components, e.g., dichroic 108 and objective lens 910, that direct the excitation radiation at the substrate 902, and particularly the signal sources 904. Emitted signals from the sources 904 are then collected by the optical components, e.g., objective 910, and passed through additional optical elements, e.g., dichroic 908, prism 912

and lens 914, until they are directed to and impinge upon an optical detection system, e.g., detector array 916. The signals are then detected by detector array 916, and the data from that detection is transmitted to an appropriate data processing system, e.g., computer 918, where the data is subjected to interpretation, analysis, and ultimately presented in a user ready format, e.g., on display 920, or printout 922, from printer 924. As will be appreciated, a variety of modifications may be made to such systems, including, for example, the use of multiplexing components to direct multiple discrete beams at different locations on the substrate, the use of spatial filter components, such as confocal masks, to filter out-of focus components, beam shaping elements to modify the spot configuration incident upon the substrates, and the like (See, e.g., Published U.S. Patent Application Nos. 2007/0036511 and 2007/095119, and U.S. Patent Application No. 11/901,273, all of which are incorporated herein by reference in their entireties for all purposes.)

VI. Exemplary Applications

[00195] The methods and compositions of the invention are useful in a broad range of analytical reactions in which one or more aspects of a detection method are detrimental to one or more aspects of the analytical reaction, such as rate, duration, fidelity, processivity, and the like. In such cases, intermittent detection at least partially mitigates the detrimental effect while allowing collection of data from stages of the analytical reaction that were previously uncollectable. As noted above, illuminated reactions are one example of analytical reactions that benefit from the compositions and methods described herein, particularly those using photoluminescent or fluorescent reagents, and particularly such reactions where one or more of the reaction components that are susceptible to photo-induced damage are present at relatively low levels. One exemplary application of the methods and compositions described herein is in single molecule analytical reactions, where the reaction of a single molecule (or very limited number of molecules) is observed in the analysis, such as observation of the action of a single enzyme molecule. In another aspect, the present invention is directed to illuminated reactions for single molecule analysis, including sequencing of nucleic acids by observing incorporation of nucleotides into a nascent nucleic acid sequence during template-directed polymerase-based synthesis. Such methods, generally referred to as "sequencing-by-incorporation" or "sequencing-by-synthesis," involve the observation of the addition of nucleotides or nucleotide analogs in a template-dependent fashion in order to determine the sequence of the template strand. See, e.g., U.S. Patent Nos. 6,780,591, 7,037,687, 7,344,865, 7,302,146. Processes for performing this detection include the use of fluorescently labeled

nucleotide analogs within a confined observation region, e.g., within a nanoscale well and/or tethered, either directly or indirectly to a surface. By using excitation illumination (i.e., illumination of an appropriate wavelength to excite the fluorescent label and induce a detectable signal), the fluorescently labeled bases can be detected as they are incorporated into the nascent strand, thus identifying the nature of the incorporated base, and as a result, the complementary base in the template strand.

[00196] In particular aspects, when an analysis relies upon a small population of reagent molecules, damage to any significant fraction of that population will have a substantial impact on the analysis being performed. For example, prolonged interrogation of a limited population of reagents, e.g., fluorescent analogs and enzymes, can lead to photo-induced damage of the various reagents to the point of substantially impacting the activity or functionality of the enzyme. It has been shown that prolonged illumination of DNA polymerases involved in synthesis using fluorescent nucleotide analogs results in a dramatic decrease in the enzyme's ability to synthesize DNA, often measured as a reduction in processivity. Without being bound to any theory of operation, it is believed that in some cases a photo-induced damage event affects the catalytic region of the enzyme thus affecting either the ability of the enzyme to remain complexed with the template, or its ability to continue synthesis. In general, the methods, devices, and systems of the present invention can increase performance and/or selectively monitor one or more stages of an illuminated reaction by subjecting the reaction to intermittent illumination.

[00197] One particularly preferred aspect of the invention is in conjunction with the sequencing by incorporation of nucleic acids within an optical confinement, such as a zero-mode waveguide. Such reactions involve observation of an extremely small reaction volume in which one or only a few polymerase enzymes and their fluorescent substrates may be present. Zero-mode waveguides, and their use in sequencing applications are generally described in U.S. Patent Nos. 6,917,726 and 7,033,764, and preferred methods of sequencing by incorporation are generally described in Published U.S. Patent Application No. 2003-0044781, the full disclosures of which are incorporated herein by reference in their entireties for all purposes, and in particular for their teachings regarding such sequencing applications and methods. Briefly, arrays of zero-mode waveguides ("ZMWs"), configured in accordance with the present invention may be employed as optical confinements for single molecule DNA sequence determination. In particular, as noted above, these ZMWs provide extremely small observation volumes at or near the transparent substrate surface, also termed the "base" of the ZMW. A nucleic acid synthesis complex, e.g.,

template sequence, polymerase, and primer, which is immobilized at the base of the ZMW, may then be specifically observed during synthesis to monitor incorporation of nucleotides in a template dependent fashion, and thus provide the identity and sequences of nucleotides in the template strand. This identification is typically accomplished by providing detectable label groups, such as fluorescent labeling molecules, on the nucleotides. In some instances, the labeled nucleotides terminate primer extension, allowing a "one base at a time" interrogation of the complex. If, upon exposure to a given labeled base, a base is incorporated, its representative fluorescent signal may be detected at the base of the ZMW. If no signal is detected, then the base was not incorporated and the complex is interrogated with each of the other bases, in turn. Once a base is incorporated, the labeling group is removed, e.g., through the use of a photocleavable linking group, and where the label was not the terminating group, a terminator, upon the 3' end of the incorporated nucleotide, may be removed prior to subsequent interrogation. In other more preferred embodiments, the incorporation of a labeled nucleotide does not terminate primer extension and the processive incorporation of multiple labeled nucleotides can be monitored in real time by detecting a series of fluorescent signals at the base of the ZMW. In some such embodiments, the label is naturally released upon incorporation of the labeled nucleotides by the polymerase, and so need not be released by alternative means, e.g., a photocleavage event. As such, a processive sequencing reaction can comprise a polymerase enzyme repetitively incorporating multiple nucleotides or nucleotide analogs, as long as such are available to the polymerase within the reaction mixture, e.g., without stalling on the template nucleic acid. (Such a processive polymerization reaction can be prevented by incorporation of nucleotides or nucleotide analogs that contain groups that block additional incorporation events, e.g., certain labeling groups or other chemical modifications.)

[00198] In accordance with the present invention, sequencing reactions may be carried out by only interrogating a reaction mixture, e.g., detecting fluorescent emission for one or more illuminated periods before excessive photo-induced damage has occurred. In general, the methods described herein are implemented in a manner sufficient to provide beneficial impact, e.g., reduced photo-induced damage and/or extension of the photo-induced damage threshold period, but are not implemented in such a manner to interfere with the reaction of interest, e.g., a sequencing reaction. The present invention also contemplates alternative methods of and compositions for mitigating the impact of photo-induced damage on a reaction, as described above and in, e.g., U.S.S.N. 61/116,048, filed November 19, 2008. Such alternative methods and compounds can be used in

combination with the compositions and methods provided herein to further alleviate the effects of species that can be generated during an illuminated reaction.

[00199] Another method of mitigating the impact of photo-induced damage on the results of a given reaction provides for the elimination of potentially damaging oxygen species using means other than the use of the photo-induced damage mitigating agents described above. In one example, dissolved oxygen species may be flushed out of aqueous systems by providing the reaction system under different gas environments, such as by exposing an aqueous reaction to neutral gas environments, such as argon, nitrogen, helium, xenon, or the like, to prevent dissolution of excess oxygen in the reaction mixture. By reducing the initial oxygen load of the system, it has been observed that photo-induced damage effects, e.g., on polymerase mediated DNA synthesis, is markedly reduced. In particularly preferred aspects, the system is exposed to a xenon atmosphere. In particular, since xenon can be induced to form a dipole, it operates as a triplet-state quencher in addition to supplanting oxygen in the aqueous system. (See, e.g., Vierstra and Poff, Plant Physiol. 1981 May; 67(5): 996–998) As such, xenon would also be categorized as a quencher, as set forth above.

[00200] Although described in terms of zero-mode waveguides, it will be appreciated that a variety of selective illumination strategies may be employed to selectively interrogate different regions of a solid support over time, e.g., so as to only damage molecules within certain selected regions of a substrate while not damaging molecules in other selected regions of the substrate. In certain embodiments, such methods can involve using a directed light source (e.g., a laser) to illuminate only selected regions; changing the illumination angle of the light source; or refocusing the illumination, e.g., by passing the illumination through an optical train that alters the shape of the incident light on the solid support. These and further examples of alternative methods of mitigating photo-induced damage which can be used in combination with methods and systems of the invention described herein are provided in U.S. Patent Pub. No. 20070036511, filed August 11, 2005; U.S. Patent No. 6,881,312; U.S.S.N. 61/116,048, filed November 19, 2008; and U.S. Patent Pub. No. 20070161017, filed December 1, 2006, all of which are incorporated herein by reference in their entireties for all purposes, and in particular for disclosure related to these methods of mitigating photo-induced damage.

[00201] As noted above, using templates that allow repeated sequencing (e.g., circular templates, SMRTbell™ templates, etc.) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated and/or increase the fold-coverage of

the sequence reads for one or more regions of interest in the template, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, in certain preferred embodiments, templates sequenced by the methods described herein are templates comprising a double-stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double-stranded or otherwise internally complementary. Such templates may, for example, comprise a double-stranded portion comprised of two complementary sequences and two single-stranded linking portions (e.g., oligos or "hairpins") joining the 3' end of each strand of the double-stranded region to the 5' end of the other strand (sometimes referred to as "SMRTbell™" templates). In certain embodiments, double-stranded portions for use in such templates are PCR-amplified. Optionally, restriction sites are incorporated within the PCR primers such that subsequent digestion of the amplified products with appropriate restriction enzymes generates double-stranded portions containing known overhang sequences on either end, which are then ligated to hairpin adapters containing a complementary overhang to generate the SMRTbell™ templates.

[00202] These template molecules are particularly useful as nucleotide sequence data generated therefrom comprises both sense and antisense nucleotide sequences for the double-stranded portion, and the circular conformation of the template enables repeated sequencing (e.g., using a polymerase capable of strand-displacement) provides duplicative or redundant sequence information. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining sequence data for each segment from the complementary sequences, as well as sequence data within each segment, by repeatedly sequencing that segment. Iterative illumination is useful in such sequencing applications, e.g., to focus nucleotide sequence data collection on stages of the sequencing reaction most of interest, such as the stages during which nucleotide sequence data is being generated from a strand of the (previously) double-stranded portion. Iterative illumination may also allow additional "rounds" of sequencing the template by virtue of the reduction in photo-induced damage to reaction components, as described elsewhere herein, thereby providing more complete and robust nucleotide sequence data for future analysis, e.g., sequence scaffold construction and/or consensus sequence determination. Further, as described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including,

but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest, e.g., from the ends of the double-stranded portion of a SMRTbell™ template.

[00203] In addition to providing sense and antisense sequence data within a single template molecule that can be sequenced in one integrated process, the presence of the single-stranded linking portions also provides an opportunity to provide a registration sequence that permits the identification of when one segment, e.g., the sense strand, is completed and the other begins, e.g., the antisense strand. Such registration sequences provide a basis for alignment sequence data from multiple sequence reads from the same template sequences, e.g., the same molecule, or identical molecules in a template population. Additional aspects of and uses for registration sequences, e.g., for molecular redundant sequencing, are further described in U.S. Patent Publication No. 20090029385, which is incorporated herein by reference in its entirety for all purposes.

[00204] In certain embodiments, such a sequencing process begins by priming the template nucleic acid within one of the linking portions and allowing the polymerase to proceed along the strand of the double-stranded portion of the template that is immediately downstream of the primed linking portion when the double-stranded portion is melted or denatured. The sequence process proceeds around the second linking portion and proceeds along the complementary strand of the (now previously) double-stranded portion of the template. Because the template is circular, this process can continue to provide multiple repeated sequence reads from the one template. Thus, sequence redundancy comes from both the determination of complementary sequences (sense and antisense strands of the double-stranded portion), and the repeated sequencing of each circular template. The ongoing sequencing reaction is subjected to multiple illuminated and non-illuminated periods to generate at least two or more sequence reads per pass around the template. The illuminated periods are preferably timed to allow generation of nucleotide sequence data for selected regions of the template. For example, it may be beneficial to only generate nucleotide sequence data for the complementary strands of the double-stranded portion, or segments thereof. As will be appreciated, in iteratively sequencing circular templates, strand displacing polymerases, as discussed elsewhere herein, are particularly preferred, as they will displace the nascent strand

with each cycle around the template, allowing continuous sequencing. Other approaches will similarly allow such iterative sequencing including, e.g., use of an enzyme having 5'-3' exonuclease activity in the reaction mixture to digest the nascent strand post-synthesis.

[00205] One may optionally employ various means for controlling initiation and/or progression of a sequencing reaction, and such means may include the addition of specific sequences or other moieties into the template nucleic acid, such as binding sites, e.g., for primers or proteins. Various methods of incorporating control elements into an analytical reaction, e.g. by integrating stop or pause points into a template, are discussed elsewhere herein and are further described in related application, U.S. Application No. 12/413,258, filed March 27, 2009, which is incorporated herein by reference in its entirety for all purposes.

[00206] In certain embodiments, a reaction stop or pause point may be included within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the linking portion that was not used in priming. By way of example, following initial sequencing from the original priming location, e.g., from the single-stranded linking portion used in priming synthesis through a first portion of the sense strand (e.g., the 3' end), the data acquisition may be switched off and the polymerase allowed to proceed around the template, e.g., through the remainder of the sense strand to the other linking portion. The incorporation of a synthesis blocking moiety coupled to this linking portion will allow control of reinitiation of the polymerase activity at the 3' end of the antisense strand. One would thereby obtain paired-end sequence data for the overall (previously) double-stranded segment, with sequence data from one end coming from the sense strand and sequence data from the other end coming from the antisense strand. This template construction and sequencing methodology is particularly useful in the case of long double-stranded segments, especially given the short read lengths generated by some sequencing technologies.

[00207] A variety of synthesis controlling groups may be employed, including, e.g., large photolabile groups coupled to the nucleobase portion of one or more bases in the single-stranded portion that inhibit polymerase-mediated replication; strand-binding moieties that prevent processive synthesis; non-native nucleotides included within the primer and/or template; and the like. The use of strand-binding moieties includes, but is not limited to, reversible, specific binding of particular proteins to recognition sequences incorporated into the template (or primer bound thereto) for this purpose. In certain embodiments, such control sequences may include binding sites for transcription factors, e.g., repressor binding regions provided within the linking portion(s). For example, the lac repressor recognition sequence is bound by the lac repressor protein, and this

binding has been shown to block replication in a manner reversible by addition of appropriate initiators, such as isophenylthiogalactoside (IPTG) or allolactose.

[00208] In some embodiments, primer recognition sequences and/or additional control sequences may also be provided for control of initiation and/or progression of polymerization, e.g., through a hybridized probe or reversibly modified nucleotide, or the like. (See, e.g., U.S. Patent Application No. 2008-0009007, the full disclosure of which is incorporated herein by reference in its entirety for all purposes.) Such probes include but are not limited to probes at which a polymerase initiates polymerization, probes containing various types of detectable labels, molecular beacons, TaqMan® probes, Invader® probes (Third Wave Technologies, Inc.), or the like, that can be used for various purposes, e.g., to provide indications of the commencement and/or progress of synthesis.

[00209] An engineered pause point (reversible or irreversible) can include one or more non-native (non-natural) or fifth bases that do not pair with any of the four native nucleoside polyphosphates in the synthesis reaction, e.g., in the template and/or oligonucleotides probe(s), and/or that exhibit a distinct kinetic signature during template-dependent synthesis at such a base. Upon encountering such a base, the polymerase pauses until the complement to the non-natural base is added to the reaction mixture. Likewise, an engineered pause point could include a "damaged" base that causes a stop in replication until repair enzymes are added to the mixture. For example, a template having a pyrimidine dimer would cause the replication complex to pause, and addition of the photolyase DNA repair enzyme would repair the problem location and allow replication, and sequencing to continue. In yet further embodiments, a combination of modification enzymes could be used to engineer a set of modified bases on a template, e.g., a combination of glycosylases, methylases, nucleases, and the like. (Further information on sequencing template nucleic acids comprising modifications, including detecting kinetic signatures of such modifications during single-molecule sequencing reactions, are provided in U.S. Patent Application Nos. 61/201,551, filed December 11, 2008; 61/180,350, filed May 21, 2009; and 12/945,767, filed November 12, 2010; and U.S. Patent Publication No. 2010/0221716, the disclosures of which are incorporated herein by reference in their entireties for all purposes.)

[00210] As noted elsewhere herein, stop or pause points can be engineered into various portions of the template, e.g., portions for which the nucleotide sequence is unknown (e.g., a genomic fragment) or known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, SMRTbell™ templates are topologically closed, single-stranded molecules having regions

of internal complementarity separated by hairpin or stem-loop linkers, such that hybridization of the regions of internal complementarity produces a double-stranded portion within the template. One or both of the linkers can comprise a stop or pause point to modulate polymerase activity. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

[00211] In other embodiments, an abasic site is used as a synthesis blocking moiety or pause point until addition of a non-natural "base," such as a pyrene, which has been shown to "base-pair" with an abasic site during DNA synthesis. (See, e.g., Matray, et al. (1999) Nature 399(6737):704-8, which is incorporated herein by reference in its entirety for all purposes.) Where a permanent termination of sequencing is desired, no non-natural analog is added and the polymerase is permanently blocked at the abasic site. DNA (or RNA) glycosylases create abasic sites that are quite different from the normal coding bases, A, T, G, and C (and U in RNA). A wide variety of monofunctional and bifunctional DNA glycosylases that have specificity for most common DNA or RNA adducts, including 5-methylcytosine, are known in the art, with different glycosylases capable of recognizing different types of modified DNA and/or RNA bases. The molecular structures of many glycosylases have been solved, and based on structural similarity they are grouped into four superfamilies. The UDG and AAG families contain small, compact glycosylases, whereas the MutM/Fpg and HhH-GPD families comprise larger enzymes with multiple domains. As an example, four enzymes have been identified in *Arabidopsis thaliana* in the plant pathway for cytosine demethylation. Additionally, other enzymes are also known to recognize 5-methyl cytosine and remove the methylated base to create an abasic site. Further, various enzymes are known to methylate cytosine in a sequence-specific manner. As such, a combination of a cytosine-methylase and an enzyme that creates an abasic site from a methylated cytosine nucleotide can be used to create one or more abasic sites in a template nucleic acid. The size of the recognition site of the methylase and the base composition of the template determine how frequently methylation occurs, and therefore, the number of abasic sites created in a given template nucleic acid, allowing the ordinary practioner to choose a methylase with a recognition site that produces a desired spacing between modified nucleotides. For example, if the recognition site is three bases long, then on average an abasic site is expected every 64 bases; if the recognition site is four bases long, then on

average an abasic site is expected every 256 bases; if the recognition site is six bases long, then on average an abasic site is expected every 4096 bases; and so forth. Of course, templates with a higher GC content would be expected to have more frequent abasic site formation, and templates with lower GC content would be expected to have less frequent abasic site formation.

[00212] Uracil-DNA glycosylases can also be used to introduce abasic sites into a template nucleic acid comprising deoxyuridine nucleotides. This strategy has the advantage of allowing the practitioner to choose the locations of the abasic sites within a DNA template since deoxyuridine nucleotides are not generally found in DNA. Various methods of inserting deoxyuridine nucleotides into a DNA template may be used, and different methods will be preferred for different applications. In certain embodiments, one or more site-specific deoxyuracils are incorporated during standard phosphoramidite oligonucleotide synthesis. To place uracils at indeterminate positions in a DNA, replacing a portion of the deoxythymidine triphosphate with deoxyuridine triphosphate will result in an amplimer with random U sites in place of T sites after polymerase chain reaction. In other embodiments, deoxyuridine nucleotides are engineered into the template, e.g., by ligation of a synthetic linker or adaptor comprising one or more deoxyuridine nucleotides to a nucleic acid sequence to be sequenced. In certain preferred embodiments, deoxyuridine nucleotides are incorporated into the linker portions of a SMRTbell™ template.

[00213] To subsequently introduce abasic sites prior to sequencing, the deoxyuridine nucleotide-containing template is subjected to treatment with uracil-DNA glycosylase, which removes the one or more uracil bases from the deoxyuridine nucleotides, thereby generating one or more abasic sites in the template. Alternatively, since the deoxyuridine nucleotide can be recognized as a template base and paired with deoxyadenosine during template-dependent nascent strand synthesis, the synthesis-blocking abasic site can instead be introduced after initiation of the sequencing reaction, e.g., at a time chosen by the practitioner. For example, the reaction can be initiated with a deoxyuridine-containing template, and uracil-DNA glycosylase can subsequently be added to block the polymerase and halt the reaction after the reaction has proceeded for a given time. As such, termination of the reaction is optional rather than required.

[00214] While uracil-DNA glycosylase activity is useful for introducing abasic sites into a template as described above, this activity can be problematic during the preparation of such templates. As such, strategies are typically implemented during preparation and manipulation of uracil-containing DNA, e.g., using molecular biology enzymes, to avoid uracil-DNA glycosylase activity, in particular, due to the *E. coli* UDG enzyme. Since a majority of standard molecular

biology enzymes are overexpressed and subsequently purified from an *E. coli* host, UDG activity can be a contaminating activity that is often not monitored by the enzyme manufacturer's quality control procedures. To mitigate contaminating UDG activity, a commercially available UDG inhibitor, also known as uracil glycosylase inhibitor or UGI (e.g., from New England Biolabs, Ipswich, MA) can be included in molecular biology reactions. This is a small protein inhibitor from the *B. subtilis* bacteriophase PBS1 that binds reversibly to E. coli UDG to inhibit its catalytic activity. UGI is also capable of dissociating UDG from a DNA molecule. Alternatively, UDG activity can be inhibited without exogenous protein using a chemical inhibitor of the enzyme, such as an oligonucleotide containing a 1-aza-deoxyribose base, a transition state analog for the UDG enzyme. This and other cationic nitrogenous sugars have been used for mechanistic studies of UDG activity and show potent inhibition activity. (See, e.g., Jiang et al. Biochemistry, 2002, 41 (22), pp 7116–7124.)

[00215] In certain applications, UDG activity needs to be inhibited temporarily, and subsequently enabled to remove create an abasic site as described above. In some embodiments, a DNA purification that removes proteins is employed, e.g., including a phenol-chloroform extraction with subsequent ethanol precipitation, a silica-based column approach (e.g., QiaQuick columns from Qiagen and similar products), and/or a PEG/sodium chloride precipitation (e.g., AMPure beads from Beckman Coulter). Alternatively or additionally, a commercially-available UDG enzyme that is not inhibited by UGI is added when abasic site formation is desired. For example, the *A. fulgidus* UDG is from a thermophilic organism and cannot be inhibited by the same bacteriophage protein as is the *E. coli* UDG enzyme. In certain preferred embodiments, UDG-inhibition is employed during template preparation, and inhibition-resistant UDG activity is added at a subsequent time to trigger the creation of abasic sites at deoxyuridine nucleotides, e.g., immediately prior to or during an ongoing reaction.

[00216] In some preferred embodiments, one or more abasic sites are engineered into a linker or adapter sequence within a sequencing template molecule. Abasic sugar residues serve as efficient terminators of polymerization for many polymerases, e.g., Φ29. 1',2'-dideoxyribose is the most common synthetic "abasic site". In other embodiments, a synthetic linker is incorporated into a linker or adaptor. For example, an internal spacer (e.g., Spacer 3 from Biosearch Technologies, Inc.) or other carbon-based linker can be used in lieu of a sugar-base nucleotide. Similar to an abasic nucleotide, the polymerase will be blocked upon encountering these moieties in the template nucleic acid.

[00217] In certain embodiments, synthesis blocking moieties are nicks in the template nucleic acid. Nicking enzymes (e.g., nicking endonucleases) are known in the art and can be used to specifically nick the template prior to or during a template-directed sequencing reaction. The use of site-specific nicking endonucleases allows the practitioner to incorporate a recognition sequence at a particular location within the template nucleic acid, and such nicking endonucleases are commercially available, e.g., from New England Biolabs, Inc. For example, a linker or adapter can be synthesized with a nicking endonuclease recognition sequence, ligated to a nucleic acid molecule to be sequenced, and can be specifically nicked either before or during a subsequent sequencing reaction. Nicks can also be introduced by ligating duplex segments that lack either a terminal 3'-hydroxy (e.g., have a dideoxynucleotide at the 3'-terminius) and/or 5'-phosphate group on one strand. The ligation results in covalent linkage of the phosphodiester backbone on one strand, but not on the other, which is therefore effectively "nicked." In certain embodiments, a SMRTbell™ template is constructed using a duplex (or "insert") nucleic acid molecule lacking a 5'-phosphate group at one or both termini. Upon ligation of the hairpin or stem-loop adaptors at each end, nicks are created at one or both ligation site(s), depending on whether the duplex lacked a 5'-phosphate at one or both ends, respectively. In other embodiments, a SMRTbell™ template is constructed using one or two stem-loop adaptors lacking a 3'-hydroxy group at the terminus (e.g., comprising a 2',3'-dideoxynucleotide rather than a 2'-deoxynucleotide). Upon ligation of one or two stem-loop adaptors lacking a 3'-hydroxy group, one or two nicks are created at the ligation site(s), depending on whether one or two adaptors lacked the 3'-hydroxy group, respectively. In both cases, a nick is created in the template nucleic acid, and a primer bound to one of the adaptors provides an initiation site for the polymerase, which will process the template until encountering a nick, at which point the polymerase will terminate the reaction by dissociation from the template. Regardless of how a nick is created, the position of a nick relative to the initiation site for the polymerase determines how much of the template will be sequenced. For example, Figure 19A provides an illustrative example of an embodiment in which a nick is present on a first strand of a duplex portion at a position distal to the adaptor containing the primer binding site. The first strand is processed by a polymerase, but the complementary strand is not processed because the polymerase dissociates at the nick site. An alternative embodiment is shown in Figure 19B, in which a nick is present on the strand complementary to the first strand at a position proximal to the adaptor containing the primer binding site. In this case both the first and complementary strands, as well as the adaptor not containing the primer binding site, are processed by the polymerase prior to dissociation. The

position of the primer binding site also determines how much of the template is processed by the polymerase. Figure 19C provides a template having a primer binding site at a position from which a polymerase would process a significant portion of the adaptor prior to entering the duplex portion. An additional advantage to using a 3'-dideoxynucleotide at a nick is that it prevents the use of the nick as a polymerse initiation site, since strand extension requires a 3-hydroxy group. As such, the resulting nick would not compete with a primer site for initiation of nascent strand synthesis by the polymerase. Having a single, known site of initiation on a template molecule is beneficial, e.g., for subsequent mapping of a read generated in such a reaction. In certain preferred embodiments, a nick site both lacks a 5'-phosphate group and comprises a 3'-dideoxynucleotide.

[00218] In certain preferred embodiments, modification and base excision is performed prior to introduction of a template nucleic acid to a reaction site, e.g., a zero-mode waveguide. As noted above, the choice of recognition site for the methylase depends on how far apart the practitioner wishes point of synthesis initiation to be on the template. For example, after initiating the template-dependent sequencing reaction, the sequence of nucleotide incorporations into the nascent strand is monitored for a desired sequence read, which may extend from the initiation point to the pause point, or may end before the polymerase reaches the pause point. In some preferred embodiments, as described elsewhere herein, the monitoring is suspended by modifying or removing an illumination source, e.g., by moving the illumination source or a substrate comprising the reaction site. Synthesis of the nascent strand will continue until the pause site is reached, whether or not the reaction is being actively monitored. When the reaction is to be reinitiated, reaction components are added that allow bypass, e.g., pyrene, polymerase, etc., and these can be subsequently removed (e.g., by buffer exchange) to allow additional pauses at other pause sites on the template.

[00219] In certain embodiments using pyrosequencing-based technologies (e.g., as developed by 454 Life Sciences), abasic sites can be introduced into a set of amplified template nucleic acids and synthesis initiated. Since all templates in the set are identical, they will comprise the same number of abasic sites in the same positions. During the course of the synthesis reaction, the synchronous incorporation of nucleotides into the nascent strands is monitored until either an abasic site is reached (at which point the synthesis is paused) or until the incorporation becomes asynchronous (which increases the background noise and decreases reliability of the sequence read). In the latter case, the practioner may opt to speed up the reaction, e.g., by adding all nucleotides at one time, to extend all nascent strands to the first abasic site in the templates. When synthesis is to be reinitiated, reaction components are added that allow bypass of the abasic site, e.g. one or more

pyrenes. A wash step may be performed to remove nucleotides and/or polymerases from the reaction sites prior to such addition. Further, in some cases, a different polymerase may be used for pyrene incorporation as is used for sequencing-by-synthesis reactions. In certain preferred embodiments, the reaction mixture comprising the pyrene for abasic site bypass allows readthrough of the abasic site, but no further on the template. Subsequent addition of sequencing reaction mixture allows the sequencing-by-synthesis reaction to recommence and incorporation of nucleotides into the nascent strand to be monitored. Alternatively or additionally, the practioner need not wait until an abasic site is reached to suspend detection and, optionally, speed up the reaction to bring all nascent strands to a given abasic site, but can choose to do this before a reaction has become asynchronous, e.g., after desired sequence data has been collected for a particular region of interest in a template nucleic acid.

[00220] In certain embodiments using ligation-based technologies (e.g., the SOLiD™ System developed by Life Technologies), a pause site can be engineered by using an oligonucleotide that cannot participate in the ligation reaction and that is complementary to a desired location on the set of identical template nucleic acids, e.g., on a bead. When the serial ligation reaction hits the position recognized by this polynucleotide, the reaction cannot proceed and any reactions that have become asynchronous will "catch up." The user can then unblock the oligo (e.g., using chemical treatment or photo-cleavage) and reinitiate the sequencing reaction.

[00221] In some cases, it may be desirable to provide endonuclease recognition sites within the template nucleic acid. For example, inclusion of such sites within a circular template can allow for a mechanism to release the template from a synthesis reaction, i.e., by linearizing it, and allowing the polymerase to run off the linear template, and/or to expose the template to exonuclease activity, and thus terminate synthesis through removal of the template. Such sites could additionally be exploited as control sequences by providing specific binding locations for endonucleases engineered to lack cleavage activity, but retain sequence specific binding, and could therefore be used to block progression of the polymerase enzyme on a template nucleic acid.

[00222] In some cases, nicking sites, e.g., sites recognized by nicking endonucleases, may be included within a portion of the template molecule, and particularly within a double-stranded portion of the template, e.g., in a double-stranded segment of a SMRT bell™ or in the stem portion of an exogenous hairpin structure. Such nicking sites provide one or more breaks in one strand of a double-stranded sequence and can thereby provide one or more priming locations for, e.g., a strand-displacing polymerase enzyme. A variety of nicking enzymes and their recognition sequences are

known in the art, with such enzymes being generally commercially available, e.g., from New England Biolabs.

[00223] In certain embodiments, methods for intermittent detection described herein are useful in "paired-end" sequencing applications in which sequence information is generated from two ends of a template nucleic acid but not for at least a portion of the intervening portion of the template. Typically, paired-end sequencing applications provide sequence data for only the two ends of a nucleic acid template, but the present invention also allows generation of additional sequence reads that are noncontiguous with the sequence reads from the ends of the template. In certain preferred embodiments, a duplex fragment (e.g., genomic fragment) is ligated to a single-stranded linker that connects the 3' end of the sense strand to the 5' end of the antisense strand, or that connects the 5' end of the sense strand to the 3' end of the antisense strand. In either orientation, separation of the two strands of the duplex fragment results in a single-stranded linear template nucleic acid that contains the linker in between the sense and antisense strands. Subsequent sequencing can involve intermittent detection that generates sequence reads for only the portions of the sense and antisense strands that are of interest, e.g., one or both of the ends. In certain embodiments, both sense and antisense strands may be sequenced at both ends to provide redundancy in the sequence data. Sequence reads recognized as being from the linker portion of the template (e.g., based on the known linker sequence or specific registration sequences encoded therein) can be used to orient the alignment of the sequence reads from the sense and antisense portions of the template, providing context for determining the sequences of the ends of the duplex fragment and subsequent sequence scaffold construction and/or mapping. In certain embodiments, pause or stop points may be incorporated into the linker to control the processing of the template by the polymerase, and therefore may be used to synchronize the detection periods to ensure generation of sequence reads from particular regions of template. Further, additional detection periods can be included that are timed to provide sequence reads from portions of the sense and/or antisense strand that are noncontiguous with the end regions.

[00224] In a related embodiment, paired-end sequencing may be accomplished by using a nucleic acid template that has linkers connecting the sense and antisense strands of a duplex fragment at both ends, such that separation of the strands of the duplex fragment provides a single-stranded circular template that contains a linkers in between each end of the sense and antisense strands of the original duplex fragment. Such a template molecule would allow a strand-displacing polymerase to proceed around the template multiple times, thereby potentially generating redundant

sequence data from both ends of both strands of the original duplex fragment. As noted elsewhere herein, such redundancy is useful for determination of consensus sequences and/or construction of sequence scaffolds. As the polymerase enzyme processes the template, detection periods can be timed (e.g., based on knowledge of the rate at which the polymerase processes the template, which is dependent not only on the polymerase but also on the sequence of the template itself) to generate nucleotide sequence reads from the regions of the template corresponding to one or both ends of the sense and antisense strands, and can also include detection periods to generate additional reads from other, noncontiguous regions of the duplex fragment, as well. Although such timing can be used to determine the appropriate periodicity of the detection periods, at later stages of the reaction (e.g., as the polymerase repeatedly proceeds around the template), the exact location of reinitiation of sequence read generation becomes more approximate. Incorporation of pause or stop points into one or both linkers to regulate the processing of the template by the polymerase may be used to synchronize the detection periods regardless of the total distance travelled by the polymerase around the template. This strategy more reliably ensures generation of sequence reads from selected regions of template, e.g, the ends of the sense and antisense portions and, optionally, regions in between and noncontiguous with the end regions regardless of the number of passes of the polymerase around the template nucleic acid, especially in later stages of the reaction. Further, the known sequence of one or both of the linkers can be used to orient sequence reads from the sense and antisense portions for consensus sequence determination and/or mapping.

[00225] In some such embodiments, a duplex fragment inserted between two hairpin linkers may be much larger than desired, increasing the difficulty of limiting nucleotide sequence read data to particular regions of the fragment. The size of the duplex fragment ligated to the two hairpin linkers can be selectively reduced to retain the regions attached to the linkers and to lose a central portion of the duplex fragment. One particularly preferred strategy, illustrated in Figure 18, comprises hairpin linkers (1802, 1804) having a regions of cross-complementarity (1806, 1808), such that the two linkers 1802 and 1804 can anneal to each other in a manner that does not interfere with ligation to a duplex fragment 1810. Duplex fragment 1810 comprises ends 1812 and 1814, as well as a long central region 1816, which is not shown but is understood to be between the two curvy lines. Once end 1812 is ligated to linker 1802 and end 1814 is ligated to linker 1804, the construct is subjected to fragmentation, which removes the central region 1816 of the duplex fragment 1810, producing construct 1818 having ends 1820 and 1822. After fragmentation, the ends of the portions of the duplex fragment still associated with the annealed linker pair (ends 1820 and

1822) are ligated together to produce construct 1824, which can then be treated (e.g., with heat, gentle denaturation, primer invasion, changing salt concentration, etc.) to separate cross-complementary regions 1806 and 1808 from one another, e.g., to generate a circular single-stranded nucleic acid molecule. Alternatively, the separation may occur during the course of the subsequent reaction, e.g., by polymerase-mediated strand displacement. Yet further, where the hybridized cross-complementary regions are long enough to undergo a complete DNA turn, an additional reaction component (e.g., helicase, topoisomerase, polymerase, etc.) may be needed to unwind the duplex and allow separation. As such, the resulting "mate-pair" construct has only the ends of the original duplex fragment ligated together and capped with adaptors that link the 5' end of each strand of the duplex with the 3' end of the other strand of the duplex, and denaturation of the duplex produces a closed, single-stranded circular construct.

[00226] Fragmentation of the duplex fragment can be performed by a variety of known methods. For example, fragmentation can be performed enzymatically (e.g., using restriction enzymes or other nucleases) or mechanically, by shearing or sonication. The type of fragmentation chosen will determine various characteristics of the resulting construct, e.g., how large a central region is removed and the types of ends remaining (e.g., blunt, 5' overhang, 3' overhang, random, identical on both ends, etc.). Optionally, the ends can be modified after fragmentation to facilitate the subsequent ligation step. Although not shown in Figure 18, it is expected that the ligation of the duplex fragment to the hybridized linkers will be a two-step process, with one end being ligated first and unimolecular kinetics favoring ligation of the second end to the second linker. The cross-complementary regions of the linkers can be designed to produce varying levels of complementarity, and therefore varying strengths of the hybridization. For example, a longer or higher GC content in a cross-complementary region lends a higher stability to the linker:linker interaction, but separation of the hybridized linkers requires a more severe treatment, e.g., higher temperature, more stringent conditions, etc. As such the cross-complementary regions should be engineered to produce a stable linker:linker interaction that is disruptable under conditions that are not destructive to the overall construct. Further the linkers can vary in regions apart from the cross-complementary regions For example, one linker can have a primer binding site that the other lacks, which would provide a single polymerase initiation site in the final construct. Other sequence characteristics described herein (e.g., pause sites, registrations sequences, etc.) can also be included in one or both linker regions. If topological constraints limit the subsequent processing of the resulting contruct, e.g., during template-directed nascent strand synthesis, these can be addressed by

addition of a reaction component (e.g., a helicase or topoisomerase) to resolve the topological constraint. As such, the methods can be used to add asymmetric linkers to duplex polynucleotides, whether or not the duplex is to be selectively reduced in size, or not, as long as the asymmetric linkers can cross-hybridize to one another.

[00227] Although in preferred embodiments, the two linkers to be ligated to a single duplex fragment are hybridized to one another prior to ligation, in some embodiments they are instead hybridized after the initial ligation reaction, and where topological constraints inhibit such a post-ligation hybridization a reaction component (e.g., topoisomerase) may be included to relieve such constraints. In certain embodiments, the hybridized linkers are separated prior to addition of reaction components for a subsequent reaction, and in other embodiments the hybridized linkers are not separated until after the addition of reaction components for a subsequent reaction. For example, a polymerase enzyme may bind to a primer annealed to a linker before or after separation of the linker from a second linker. In fact, it may be benefical in some embodiments to postpone separation of the linkers, e.g., where compaction of the nucleic acid construct is beneficial, such as when the construct must be loaded into a confinement of some kind, e.g., a nanowell, optical confinement, etc.

[00228] In some embodiments, the methods further include separation of single linker constructs from hybridized linker pair constructs. This can be accomplished by an exonuclease treatment after ligation of the duplex fragment to the linkers, which would degrade any constructs having an unannealed end. Alternatively, it may be desirable to remove the single linkers prior to ligation, for example using a size separation methodology or by allowing them to bind to oligonucleotides that are complementary to the cross-complementary regions and bound to a column or magnetic beads. (The cross-complementary regions of the hybridized linker pairs will not be available for binding to the oligonucleotides. Other methods known in the art can also be used to separate single linkers from hybridized linker pairs.

[00229] Interestingly, the use of the sense/antisense nucleic acid templates described above would represent a unidirectional processing of a template to provide paired-end sequence data, as opposed to the more traditional bi-directional processing of a linear template molecule. Further, unlike traditional approaches, these methods for paired-end sequencing involve processing, chemically or otherwise, of not just the regions at the ends, but also regions in between the ends, and in some embodiments comprising processing of the entire template. For example, a polymerase incorporates nucleotides into a nascent strand for each position of the template (thereby

"processing" each position of the template), yet the sequencing data generated is limited to specific regions of the template that are of particular interest to the practitioner, such as the end regions. As such, in certain embodiments the duplex fragment is not further reduced in size after ligation to a linker pair, and the entire duplex fragment is processed by the polymerase.

[00230] In certain embodiments, methods for intermittent detection described herein are useful in analysis systems that employ nanopores. A nanopore is a small pore in an electrically insulating membrane that can be used for single molecule detection. In general, a nanopore functions as a Coulter counter for much smaller particles, and can take various forms, e.g., a protein channel in a lipid bilayer or a pore in a solid-state membrane. The detection principal is based on monitoring the ionic current of an electrolyte solution passing through the nanopore as a voltage is applied across the membrane. For example, passage of a polynucleotide molecule (e.g., DNA, RNA, etc.) through a nanopore causes changes in the magnitude of the current through the nanopore, with each nucleotide obstructing the nanopore to a different, characteristic degree. As such, the pattern of variations in the current passing through the nanopore as the polynucleotide is drawn through may be monitored and analyzed to determine the nucleotide sequence of the polynucleotide. A polynucleotide may be drawn through the nanopore by various means, e.g., by electrophoresis, or using enzyme chaperones to guide the polynucleotide through the nanopore. For additional discussion of methods of fabrication and use of nanopores, see, e.g., U.S. Patent No. 5,795,782; Kasianowicz, J.J., et al. (1996) *Proc Natl Acad Sci USA* 93(24):13770-3; Ashkenas, N., et al. (2005) *Angew Chem Int Ed Engl* 44(9):1401-4; Winters-Hilt, S., et al. (2003) *Biophys J* 84:967-76; Astier, Y., et al. (2006) *J Am Chem Soc* 128(5):1705-10; Fologea, D., et al. (2005) *Nano Lett* 5(10):1905-9; Deamer, D.W., et al. (2000) *Trends Biotechnol* 18(4):147-51; and Church, G.M. (2006) *Scientific American* 294(1):52, all of which are incorporated by reference herein in their entireties for all purposes. In some embodiments, intermittent detection of nucleic acid sequence data from a nanopore may be achieved by modifying the progress of the polynucleotide through the nanopore so that progress is sped up during non-detection periods and progress is slowed to allow sequence determination during detection periods. The rate of passage of the polynucleotide through the nanopore may be modified by various methods, including but not limited to increasing an electrophoretic field carrying the polynucleotide (e.g., by increasing the voltage, changing the conductivity of the reaction mixture, and the like), or changing various reaction conditions to alter the speed at which a protein chaperone carries the polynucleotide. Further, in embodiments utilizing

a processive exonuclease to feed individual bases through the nanopore, the kinetics of the exonuclease may be modified based on the known biochemical characteristics of the exonuclease.

[00231] In diagnostic sequencing applications, it may be necessary only to provide sequence data for a small fragment of DNA, but do so in an extremely accurate sequencing process. For such applications, shorter target segments may be employed, thus permitting a higher level of redundancy by sequencing multiple times around a smaller circular template, where such redundancy provides the desired accuracy. Thus, in some cases, the double stranded target segment may be much shorter, e.g., from 10 to 200, from 20 to 100 or from 20 to 50 or from 20 to 75 bases in length. For purposes of the foregoing, the length of the target segment in terms of bases denotes the length of one strand of the double stranded segment. In such applications, various methods for intermittent detection described herein may be used to analyze the sequence of the template, thereby targeting the sequence data to the portion(s) of the template of particular interest to the diagnostician, and/or improving various aspects of the reaction performance, e.g., by virtue of the reduction of photo-induced damage to one or more reaction components.

[00232] It is to be understood that the above description is intended to be illustrative and not restrictive. It readily should be apparent to one skilled in the art that various embodiments and modifications may be made to the invention disclosed in this application, including but not limited to combinations of various aspects of the invention, without departing from the scope and spirit of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. All publications mentioned herein are cited for the purpose of describing and disclosing reagents, methodologies and concepts that may be used in connection with the present invention. Nothing herein is to be construed as an admission that these references are prior art in relation to the inventions described herein. Throughout the disclosure various patents, patent applications and publications are referenced. Unless otherwise indicated, each is incorporated by reference in its entirety for all purposes.

[00233] Although described in some detail for purposes of illustration, it will be readily appreciated that a number of variations known or appreciated by those of skill in the art may be practiced within the scope of present invention. Unless otherwise clear from the context or expressly stated, any concentration values provided herein are generally given in terms of admixture values or percentages without regard to any conversion that occurs upon or following addition of

the particular component of the mixture. To the extent not already expressly incorporated herein, all published references and patent documents referred to in this disclosure are incorporated herein by reference in their entirety for all purposes.

[00234] The following non-limiting examples are provided to further illustrate the invention.

VI. Examples of Intermittent Illumination of a Single Molecule Sequencing-by-Synthesis Reaction

Example I

[00235] A nucleic acid template was provided that comprised a double-stranded region and two single-stranded linker portions at each end. The first linker portion connected the 3' end of the sense strand with the 5' end of the antisense strand, and the second linker portion connected the 3' end of the antisense strand with the 5' end of the sense strand. This template was designed to form a single-stranded circle of approximately 500 bases when the double-stranded region was opened (e.g., by heat denaturization, helicase activity, etc.), and is sometimes referred to as a SMRTbell™ template. A plurality of this nucleic acid template was incubated with polymerases, primers, and other reaction components to allow formation of polymerase-template complexes. (See, e.g., Korlach, J., et al. (2008) Nucleosides, Nucleotides and Nucleic Acids, 27:1072-1083; and Eid, J. (2009) Science 323:133-138.) The complexes were immobilized in zero-mode waveguides in a reaction mixture containing all necessary buffer and nucleotide analog components for carrying out sequencing-by-synthesis reactions with the exception of a cognate starting base and a metal dication. A Smith-Waterman algorithm was used to perform the alignment of the known sequence of the template with the sequence reads generated in the reaction, and the positions of the sequence reads is graphically illustrated in Figure 8.

[00236] Acquisition of the data shown in Figure 8 was collected as follows. Illumination of the array of zero-mode waveguides was initiated with laser excitation (532 nm and 641 nm laser lines) at t = -5 seconds, and the missing cognate starting base and metal dication (manganese metal) were added at t = 0 seconds to simultaneously initiate the sequencing-by-synthesis reactions in all zero-mode waveguides. The reactions were monitored under illumination for 120 seconds at which time the illumination was removed; the sequencing reads generated during that stage of the reaction are shown in Figure 8A as a function of the template position to which each read maps. At 295 seconds illumination was resumed and data acquisition was reinitiated at 300 seconds and maintained for another 120 second interval; the sequencing reads during this second illuminated

period are shown in Figure 8B. At 595 seconds illumination was resumed and data acquisition was reinitiated at 600 seconds and maintained for another 120 second interval; the sequencing reads during this third illuminated period are shown in Figure 8C.

[00237] As expected, the longer the amount of time before the sequence data is collected (that is, the later the illuminated period), the further into the template the alignments shift, and this shift is a rough function of time since initiation of the reaction. Further, the distribution of sequence reads generated during each subsequent illuminated period becomes more dispersed than the previous illuminated period(s). Further, due to the circular nature of the template, Figure 8C clearly shows that some polymerases have passed completely around the substrate and are beginning to generate sequence reads from a second pass around the template, thereby generating redundant sequence information for a single template nucleic acid.

Example II

[00238] As in Example I, a SMRTbell™ template was used. For templates of defined sequence, PCR was used to generate 3 or 6 kb DNA inserts for the double-stranded region in the SMRTbell™ templates using a standard PCR methodology. For genomic and other biological samples, a DNA fragmentation protocol was used that generates DNA fragments distributed around 3 or 6 kb. Generation of fragments in these ranges was done using a HydroShear® (Genomic Solutions®) device with settings recommended by the manufacturer. The random genomic DNA fragments were enzymatically treated to generate blunt ends. Both the PCR products and randomly generated DNA fragments were phosphorylated and then immediately put into a ligation reaction with a blunt hairpin adapter. The products were purified through two size selection steps using reduced volumes of AMPure® magnetic beads (Agencourt®) to remove hairpin dimers and other short products. (Fabrication of SMRTbell™ templates is further described elsewhere herein.)

[00239] The system components used for polynucleotide sequencing using intermittent detection are comparable to single-molecule sequencing applications under constant illumination, which are described, e.g., in Eid, et al. (2009) Science 323:133-138. Specifically, the immobilization and sequencing buffer compositions, nucleotide analogs identity and concentration, polymerase, ZMWs, surface treatment and instrumentation were identical to the standard methodology. Modifications to the SMRTbell™ template DNA and polymerase binding and immobilization and data acquisition protocols are as follows.

[00240] A binding solution was prepared by incubation of 3 or 6 kb DNA SMRTbell™ templates (1-10 nM) with a 10-fold excess of DNA polymerase (10-100 nM, respectively) in 10 mM MOPS (pH 7.5), 10 mM KOAc, 100 mM DTT & 0.05% Tween-20 for 2 hours at 30°C, followed by 1 hour at 37°C and subsequent storage at 4°C prior to immobilization on the ZMWs. Immediately prior to immobilization, the binding solution was diluted in the standard immobilization solution (50 mM MOPS (pH 7.5), 75 mM KOAc, 5 mM DTT, 0.05% Tween-20) to the desired final concentration, typically 0.1 to 1 nM, and incubated for 30 to 60 minutes at 22°C. Post-immobilization chip preparation and sequencing initiation were identical to the standard methods.

[00241] The data acquisition protocol was similar to the standard application with coordinated modifications to the collection timing and ZMW positioning. In the standard acquisition procedure, a single long acquisition (~10 minutes) is performed for each ZMW. In the intermittent illumination acquisition procedure, multiple short acquisitions (~3 minutes) of sequence reads (also termed "strobe reads") were performed for each ZMW (during "detection periods") with an interval between each acquisition period during which no acquisition of sequence reads was performed ("non-detection period"). The duration of the interval between each acquisition of sequence reads was determined based upon a desired distance (i.e., number of nucleotide positions) between each sequence (or strobe) read, the polymerization rate of the polymerase, and the SMRTbell™ template insert size.

[00242] SMRTbell™ templates were generated as described above for AC223433, a fosmid clone comprising a sequence of an approximately 40 kb region of *Homo sapiens* chromosome 15. The reference sequences used to map the sequence reads generated in the sequencing reactions were the publically available sequences of *Homo sapiens* chromosome 15 (Hg18; NCBI Build 36.1) and fosmid AC223433 (NCBI GenBank accession number). Table 1 shows the number of statistically significantly mapped sequence reads for several types of intermittent illumination sequencing reactions. The number of mappable "looks" is equivalent to the number of mappable sequence reads generated during detection periods for a single template molecule. For example, a "mapped 1-look read" means, for a single template molecule, only a single detection period generated a sequence read that could be mapped to the reference sequence.

Table 1: Summary of Sequencing Results

Mapping	Mapped 1-	Mapped 2-	Mapped 3-	Mapped 4-

Reference	look reads	look reads	look reads	look reads
Human chr15 (Hg18)	13834	1289	127	4
Fosmid	15253	1571	158	5

[00243] Deviations in the expected time span for a set of sequencing reads from a single sequencing reaction are indicative of genomic events such as genomic rearrangements, e.g., insertions, deletions, etc. Figures 10 and 11 illustrate this point. Specifically, the time and distance travelled along the template (based upon the reference sequence) by the polymerase was computed within and between the sequence reads generated during the detection (illuminated) periods. These calculations were used to detect unexpected variations, indicating possible genomic events in the template as compared to the reference sequence. Figure 10 provides a plot that illustrates the normalized average time it took for the polymerase to traverse a region of the template based on the length of that region in the *Homo sapiens* chromosome 15 reference sequence. The sequence reads are fit to a diagonal having a slope equal to the average speed for sequencing reads. Deviations from the regressed diagonal indicate genomic events (for example, structural variants), and the slope of the sequence reads around such deviations indicate the relative size of the genomic event (e.g., in the case of insertions/deletions). For example, if the time for the polymerase to traverse a region was unexpectedly long, this indicated the polymerase actually traversed a longer region than was expected based on the reference sequence. The two distinct off-diagonal deviations (upper right hand corner) with higher slope indicated that an insertion had occurred in the reference sequence, and this was verified by comparison to the known fosmid sequence.

[00244] Figure 11 shows the average time it took the polymerase to traverse the template. For each mapped read, starting and ending times and positions were determined and used to compute the distance traversed by the polymerase between sequence reads. Based on these determinations, an average time across any particular region of the human reference sequence was computed. Regions that were traversed by the polymerase more slowly have peaks of higher ΔT, and were indicative of insertions in the template relative to the *Homo sapiens* chromosome 15 reference sequence. The insertions identified were the same insertions identified above.

[00245] Intermittent illumination-based sequencing reactions across fosmid sequence AC223433 showed significant sequence read coverage across the insertion events. The distribution of the physical coverage is shown in Figure 12, which illustrates examples of three-look strobes (i.e., sequencing reactions having three detection/illuminated periods) that span or intersect the

insertion events. Figure 12A shows the mapping of the strobe sequence reads to the *Homo sapiens* chromosome 15 reference sequence, where the sequence reads generated from the insert sequences in the template are excluded. Arrows indicate the locations of the insertions. Figure 12B shows a similar mapping with the sequence reads generated from the insert sequences indicated with brackets. A number of sequence reads flank the insertions, connect the two insertions, or clarify sequence within (or at the boundaries of) the insertion sequences. Such flanking and connecting sequence reads are useful for predicting and detecting genomic events, anchoring them to genomic references, and scaffolding for de novo assembly of novel sequences. In particular, there are 30 and 38 "3-look" reads that intersect the two regions of insertion of (1192 bp and 6879 bp, respectively). These sequence reads facilitated mapping of the insertions to the human reference sequence, which would have been extremely difficult, if not impossible, with commercially available short-read sequencing technologies. Further the sequence of the smaller insertion was a highly repetitive sequence, which would also have made mapping difficult with certain short-read technologies.

[00246] Figure 13 illustrates the sequence coverage obtained across the fosmid sequence, showing all two-, three-, and four-look strobe sequence reads spanning the sequence that are mappable to the known AC223433 fosmid sequence.

A consensus sequence was derived from the set of mappable sequence reads generated in these sequencing reactions. Strobe sequence reads were combined with sequence reads generated under constant illumination and assembled based on the human reference sequence (Hg18). High quality reads surrounding the (suspected) insertion sites, as well as high quality reads that did not map to the reference sequence, were extracted and assembled with a "de novo" greedy suffix tree assembler; the resulting contigs were mapped to the Hg18 reference sequence. Contigs spanning the (suspected) insertion sites were identified and fed back into the "de novo" assembler, and the resulting contigs were manually edited using standard techniques and placed back into the derived reference guided assembly. The final consensus sequence was a hybrid of a reference guided assembly and attempts at de novo assembly of novel insert sequences. Alignments to reference sequences were performed and plotted. Figure 14 provides a sequence dot plot for an alignment between a sequence assembly produced as described above and the fosmid reference sequence, and this plot confirmed a high degree of alignment between the two sequences. This dot plot was generated using Gepard 1.21 ("GEnome PAir – Rapid Dotter," available from the Munich Information Center for Protein Sequences (MIPS)) with a word size of 7. Nucleic acid dot plots are widely used in the art and are further described, e.g., in Krumsiek et al. (2007) Bioinformatics

23(8):1026-8; Maizel et al. (1981) Proc Natl Acad Sci USA 78:7665; Pustell, et al. (1982) Nucleic Acids Res 10:4765; and Quigley, et al. (1984) Nucleic Acids Res 12:347, all of which are incorporated herein by reference in their entireties for all purposes.

Electronic Acknowledgement Receipt

EFS ID:	28126474
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	20-JAN-2017
Filing Date:	19-DEC-2016
Time Stamp:	19:01:00
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	01007706_2017-01-20_Trans.pdf	760091 56dbc8cad67ef75fe13099ad910d6d76d008c56d	no	1

Warnings:

Information:					
2		01007706_2017-01-20_PrelimAmend.pdf	88805 4f71957404da9d9bca28aa28e3f09ed0875 0f1c73	yes	5
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Preliminary Amendment		1	1	
	Specification		2	3	
	Drawings-only black and white line drawings		4	4	
	Applicant Arguments/Remarks Made in an Amendment		5	5	
Warnings:					
Information:					
3	Specification	01007706_2017-01-20_SubstSpec-Redline.pdf	487927 76cec2cb2d92ad41e63854c03d4ef1b2c7d0 6fc4	no	106
Warnings:					
Information:					
4	Specification	01007706_2017-01-20_SubstSpec-Clean.pdf	487436 7231b21a297599fdcededa3588188814040ee 3a216	no	106
Warnings:					
Information:					
5	Drawings-only black and white line drawings	01007706_2017-01-20_ReplacementFigs-1.pdf	2040416 e36c9378dc0358fdfe3325339b6ff9576675b cb82	no	14
Warnings:					
Information:					
		Total Files Size (in bytes):	3864675		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/21 (10-08)
Approved for use through 11/30/2008. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	15/383,965
Filing Date	December 19, 2016
First Named Inventor	Stephen Turner
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	01-007706US

ENCLOSURES *(Check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☑ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☑ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☑ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation
- ☐ Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- ☐ Proprietary Information
- ☐ Status Letter
- ☑ Other Enclosure(s) (please identify below):
  Substitute Specification (both redline and clean)

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Pacific Biosciences of California, Inc.		
Signature	/David C. Scherer, Ph.D./		
Printed name	David C. Scherer, Ph.D.		
Date	01-20-2017	Reg. No.	56,993

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below: via EFS-Web

Signature	/Jacqueline L. Lim/		
Typed or printed name	Jacqueline L. Lim	Date	January 20, 2017

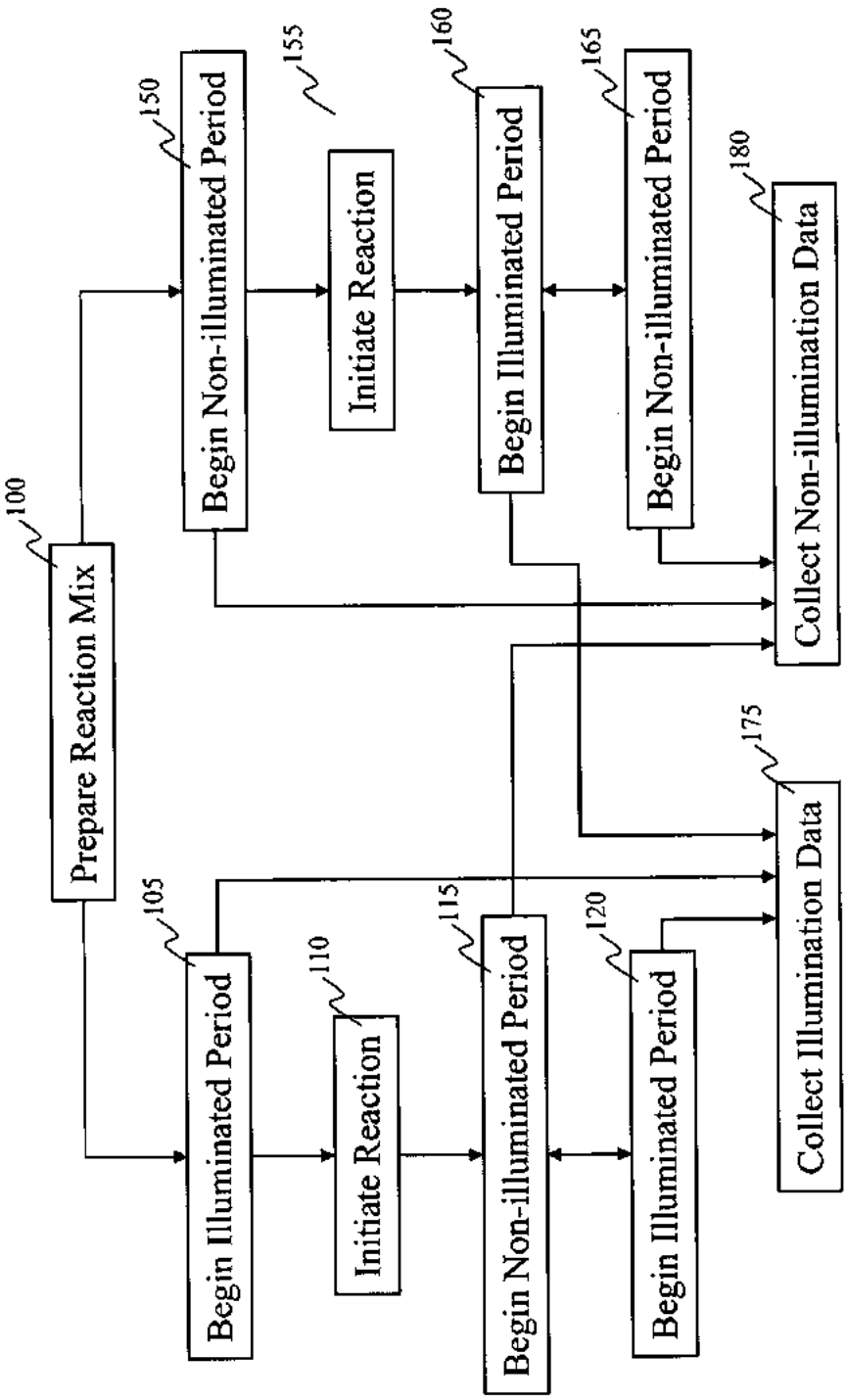

Prepare Reaction Mix
100
Begin Illuminated Period
105
Initiate Reaction
110
Begin Non-illuminated Period
115
Begin Illuminated Period
120
Collect Illumination Data
175
Begin Non-illuminated Period
150
Initiate Reaction
155
Begin Illuminated Period
160
Begin Non-illuminated Period
165
Collect Non-illumination Data
180

Figure 1

Figure 2A

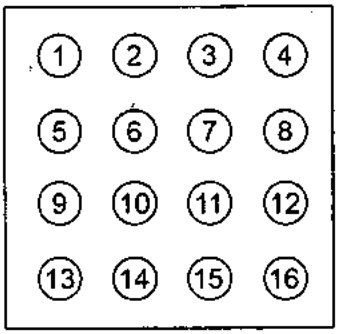


Figure 2B

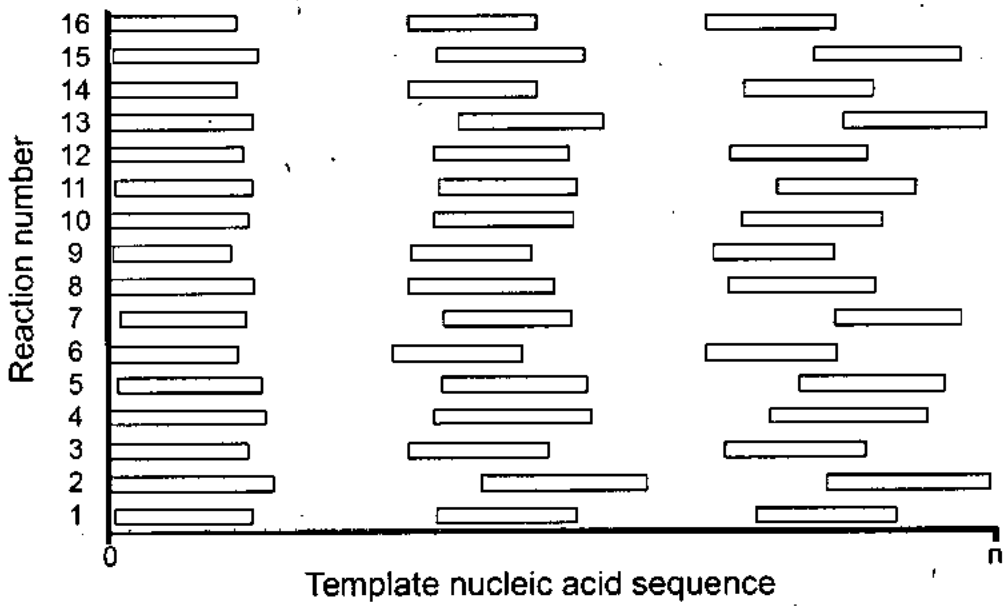


Figure 3A

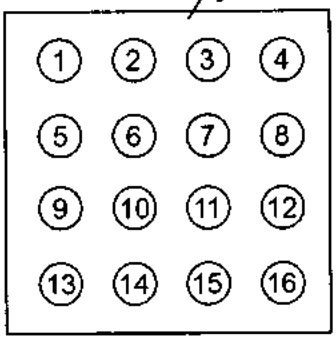


Figure 3B

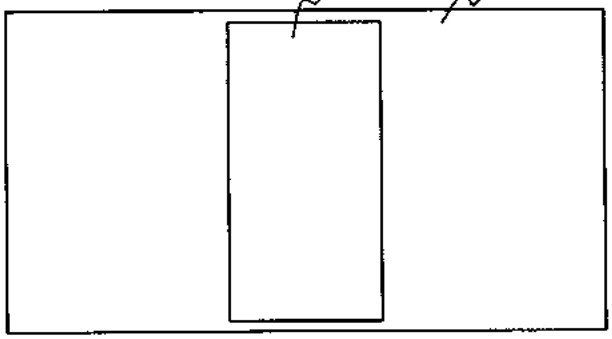


Figure 3C

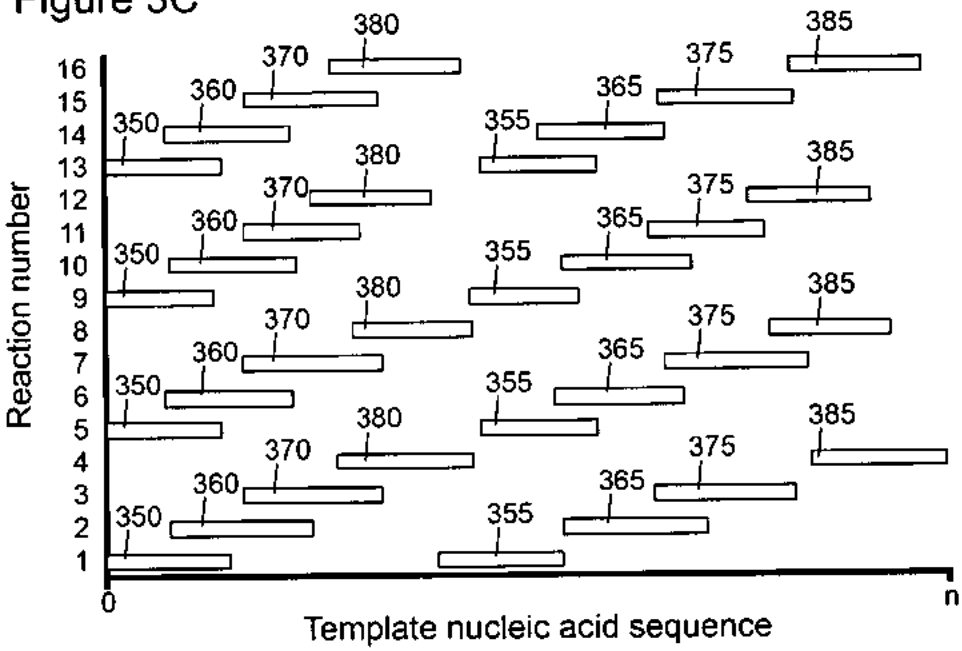


Figure 4A

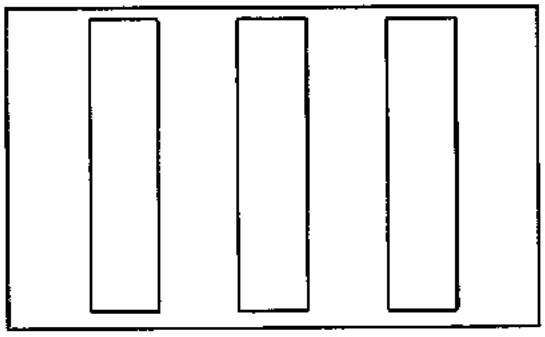


Figure 4B

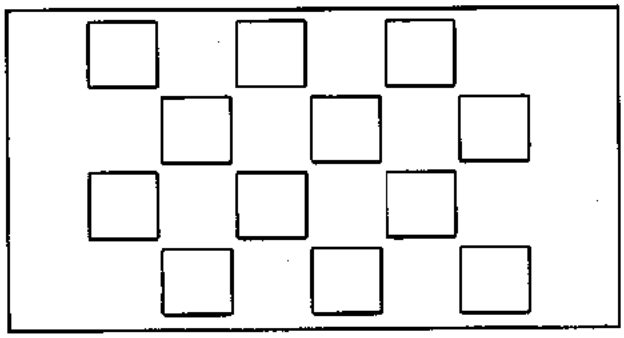


Figure 5A

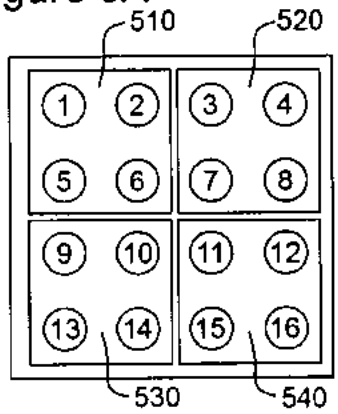


Figure 5B

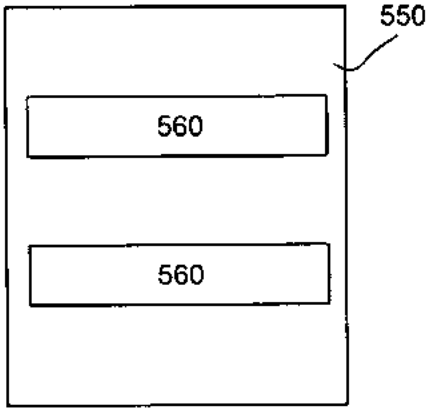


Figure 5C

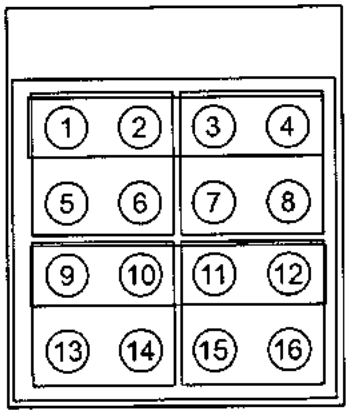
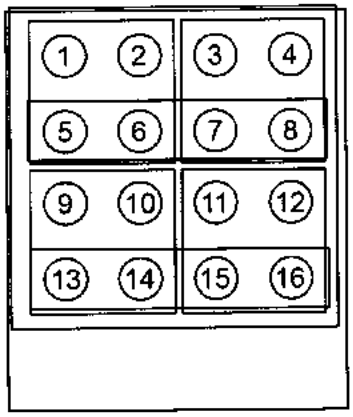


Figure 5D



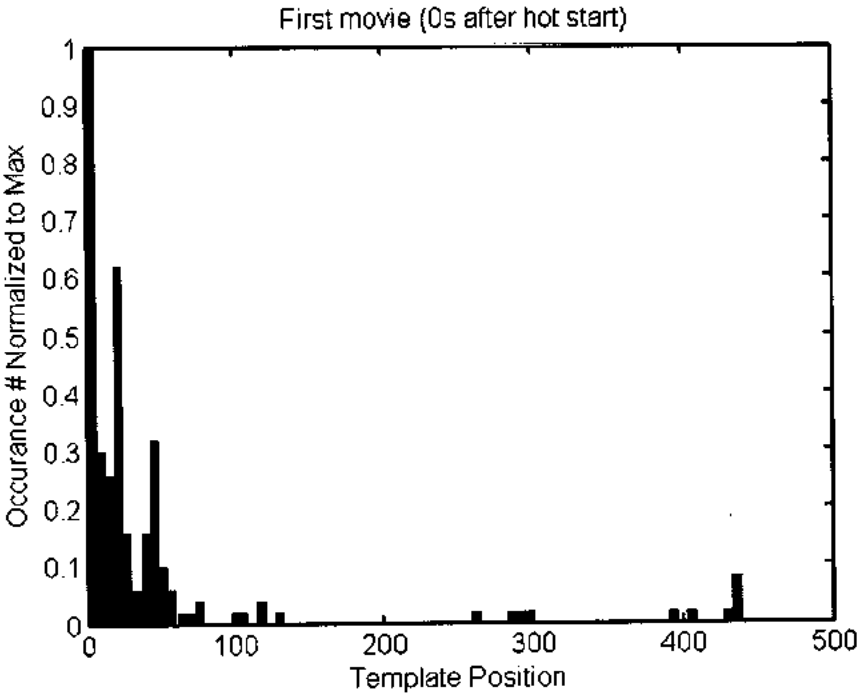

First movie (0s after hot start)
Occurance # Normalized to Max
Template Position
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500

Figure 8A

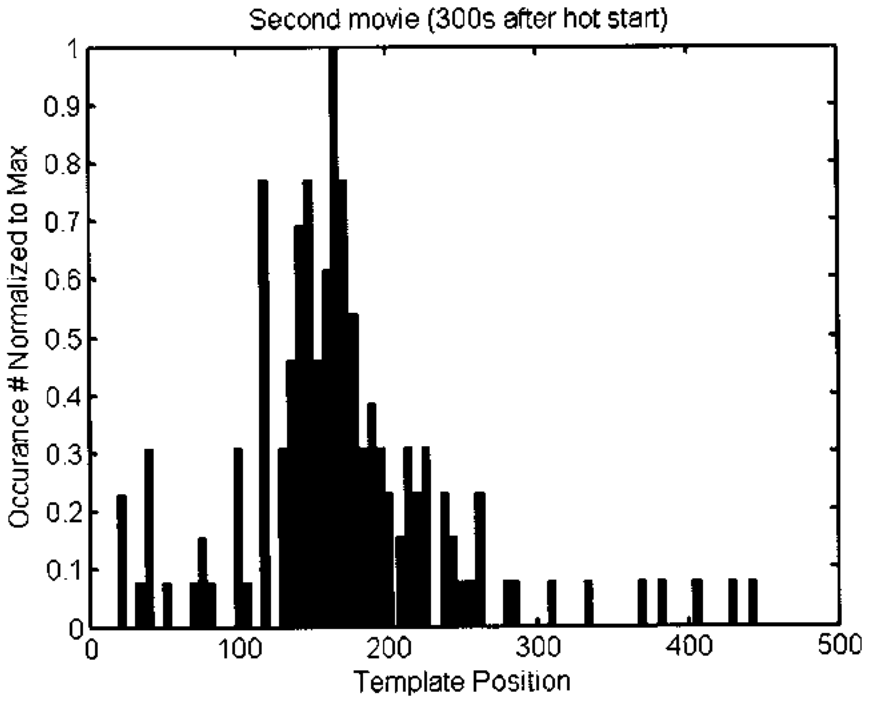

Second movie (300s after hot start)
Occurance # Normalized to Max
Template Position
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500

Figure 8B

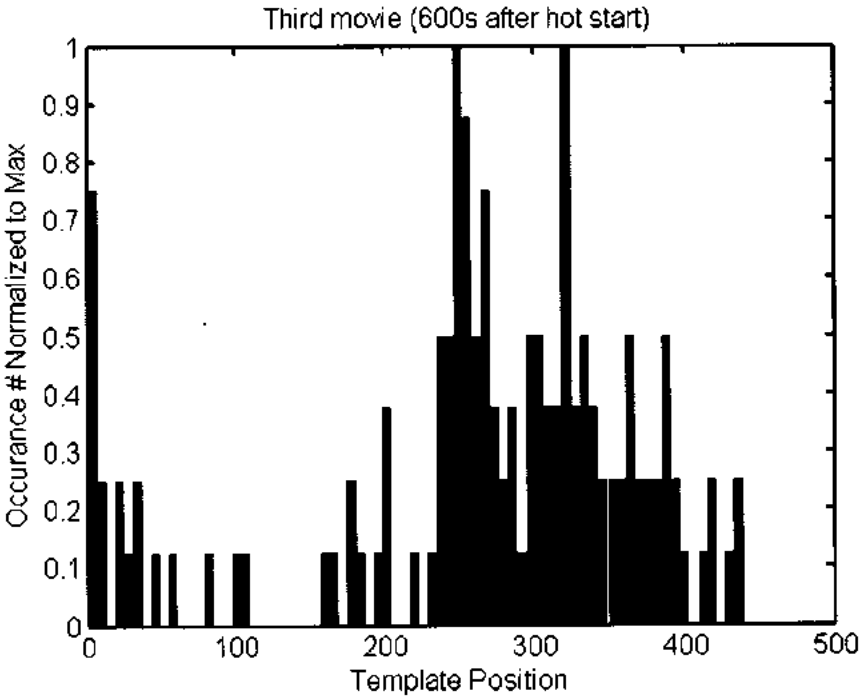

Third movie (600s after hot start)
Occurance # Normalized to Max
Template Position
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500

Figure 8C

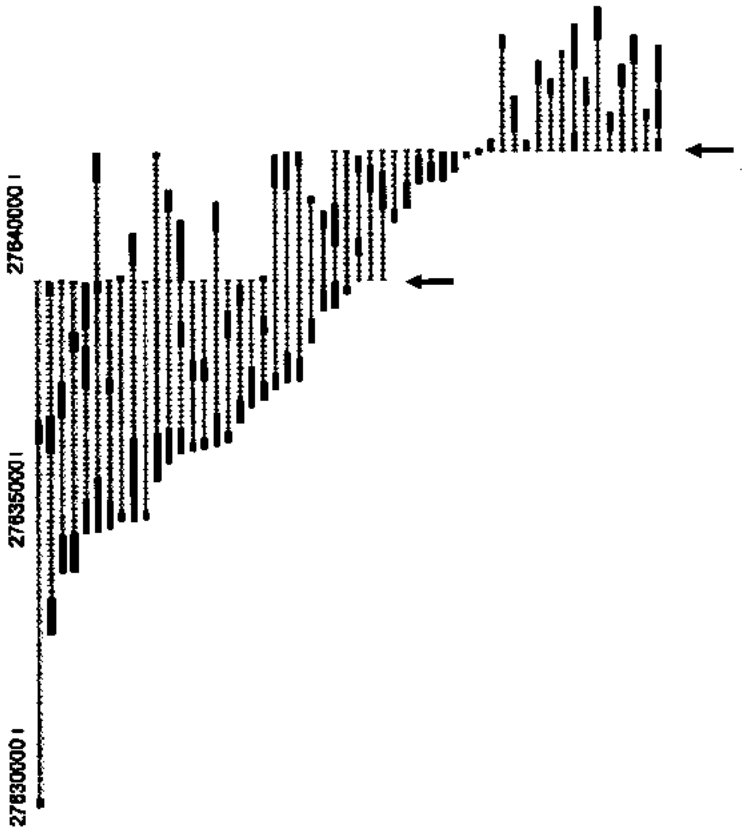


Figure 12A

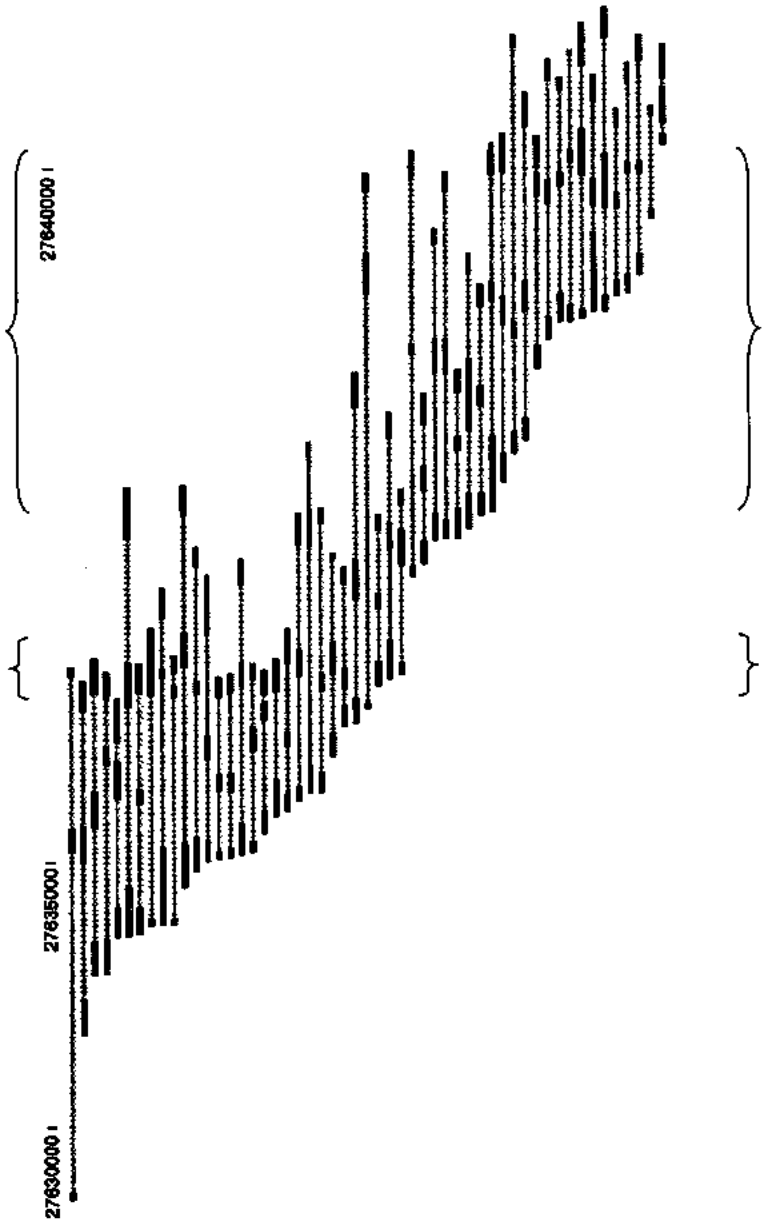


Figure 12B

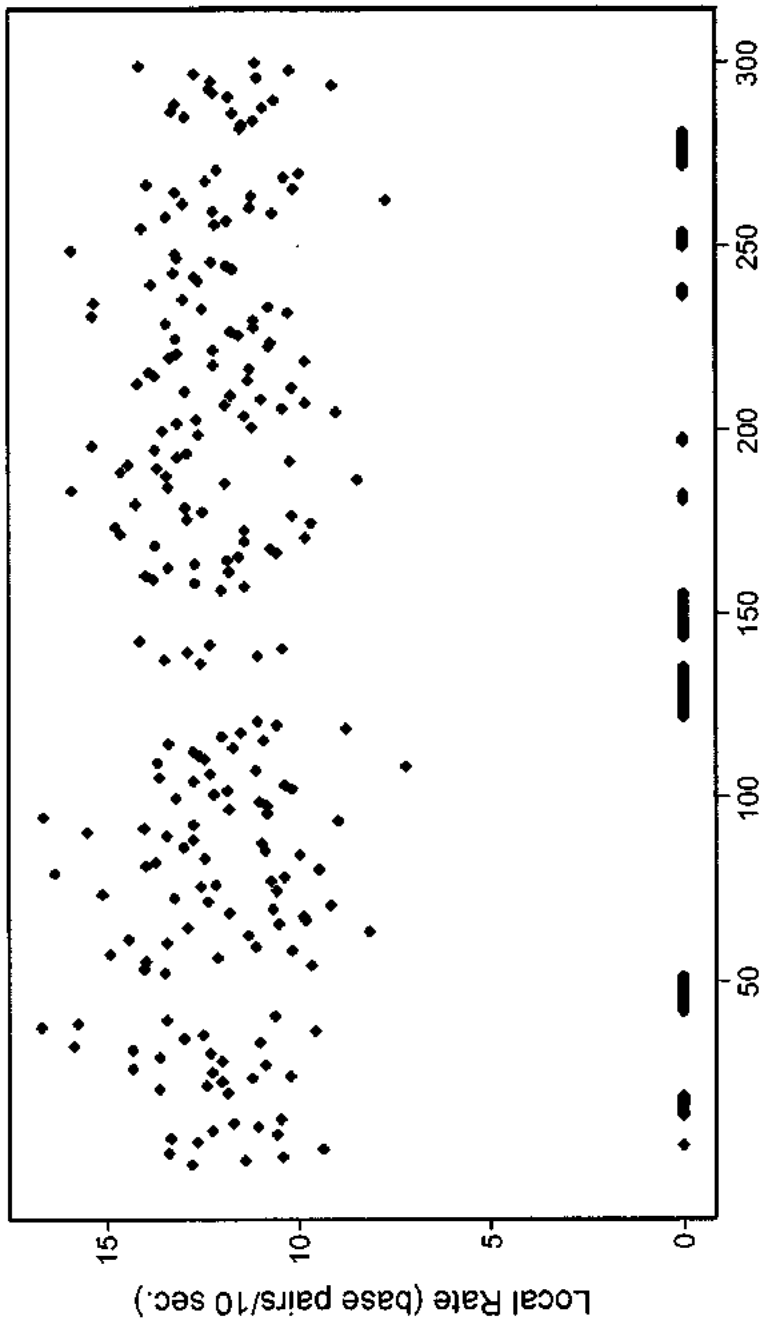

Local Rate (base pairs/10 sec.)
0
5
10
15
Time (10 sec.)
50
100
150
200
250
300

Figure 16A

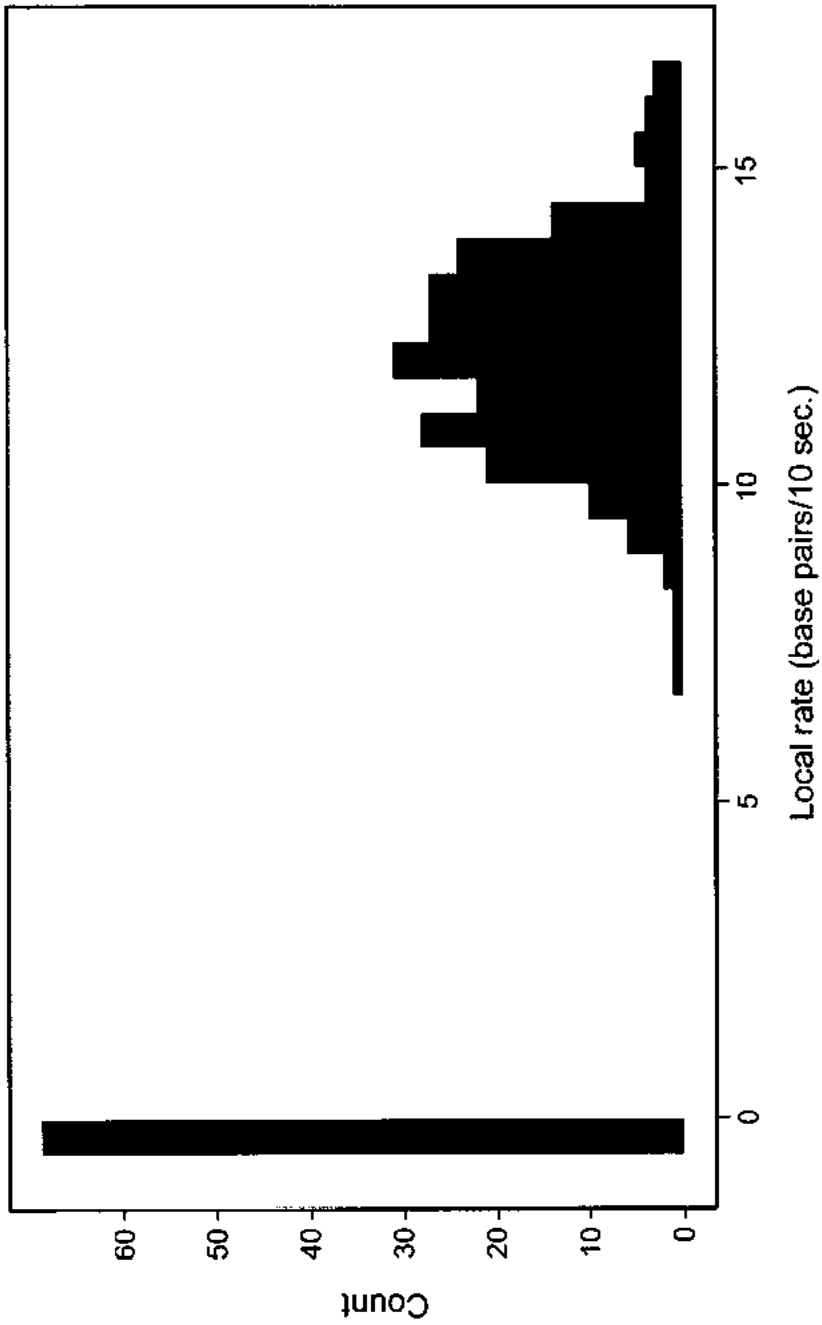


Figure 16B

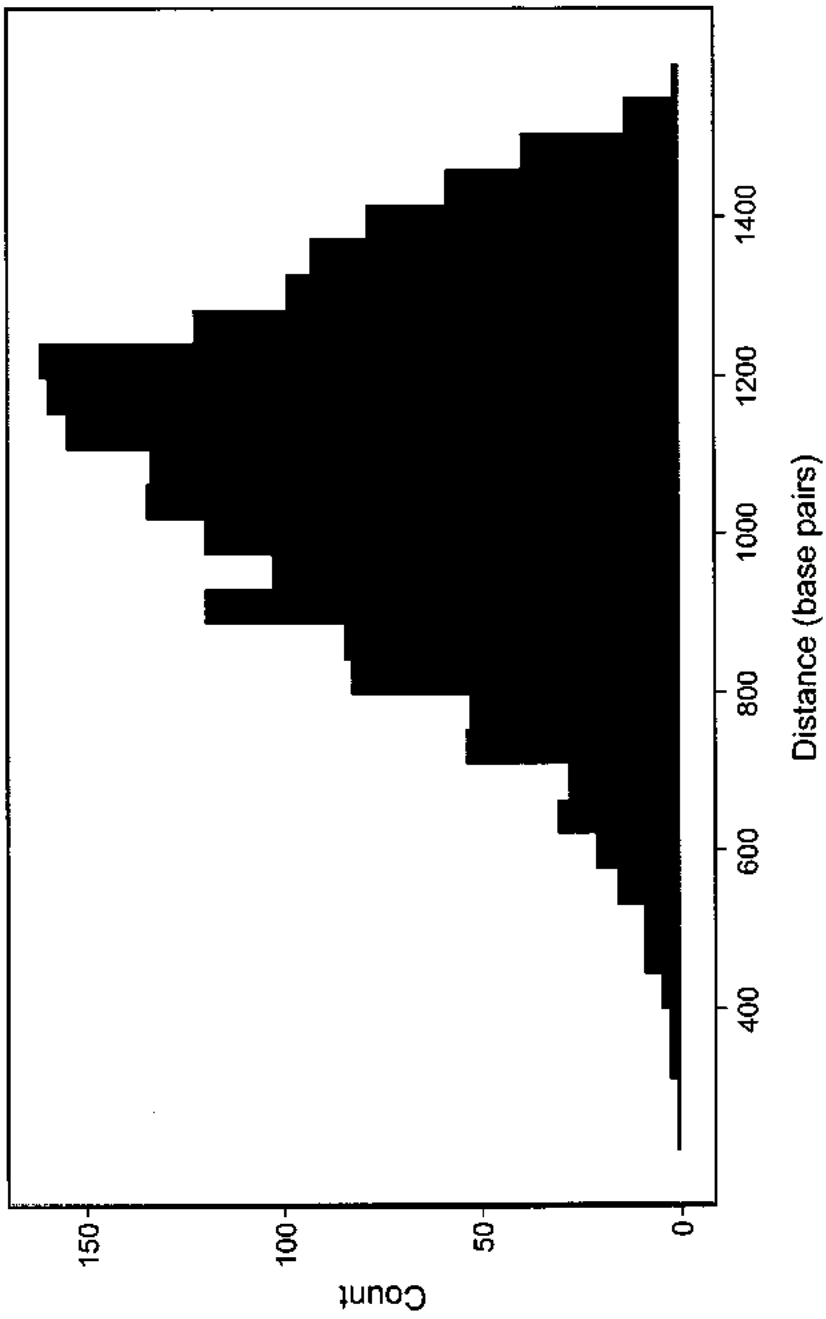

Count
150
100
50
0
400
600
800
1000
1200
1400
Distance (base pairs)

Figure 16C

Figure 19A

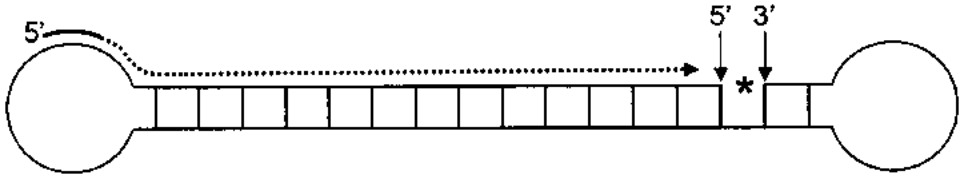


Figure 19B

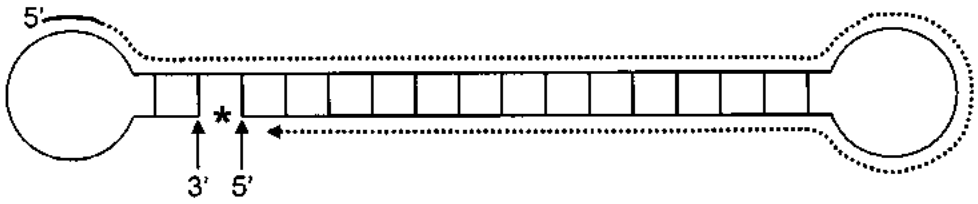
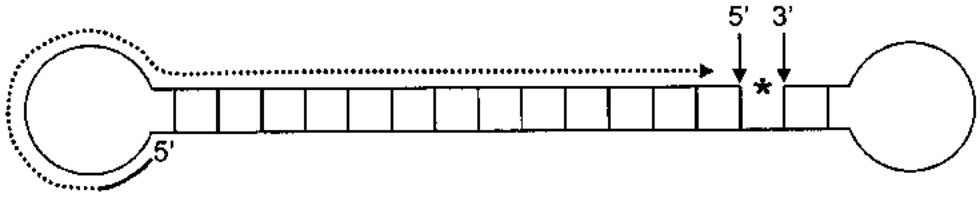


Figure 19C



SCORE Placeholder Sheet for IFW Content

Application Number: 15383965 Document Date: 01/20/2017

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

- Drawings – Other than Black and White Line Drawings

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

To access the documents in the SCORE database, refer to instructions below.

At the time of document entry (noted above):

- Examiners may access SCORE content via the eDAN interface.
- Other USPTO employees can bookmark the current SCORE URL (http://Score.uspto.gov/ScoreAccessWeb/).
- External customers may access SCORE content via the Public and Private PAIR interfaces.

Form Revision Date: September 30, 2013

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/383,965	12/19/2016	1634	1600	01-007706US	18	1

CONFIRMATION NO. 8144

UPDATED FILING RECEIPT

57770
PACIFIC BIOSCIENCES OF CALIFORNIA, INC.
1380 Willow Road
MENLO PARK, CA 94025

Date Mailed: 01/25/2017

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Stephen Turner, Seattle, WA;
Jon Sorenson, Alameda, CA;
Kenneth Mark Maxham, Redwood City, CA;
John Eid, San Francisco, CA;

Applicant(s)

Pacific Biosciences of California, Inc., Menlo Park, CA;

Assignment For Published Patent Application

Pacific Biosciences of California, Inc.

Power of Attorney: The patent practitioners associated with Customer Number 57770

Domestic Priority data as claimed by applicant

This application is a CON of 14/708,603 05/11/2015 PAT 9556480
which is a CON of 14/091,961 11/27/2013 PAT 9057102
which is a CON of 12/982,029 12/30/2010 PAT 8628940
which claims benefit of 61/099,696 09/24/2008
and claims benefit of 61/139,402 12/19/2008
and is a CIP of 12/413,226 03/27/2009 PAT 8143030

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: No

Permission to Access Search Results: No

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 12/28/2016
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 15/383,965**
Projected Publication Date: 05/04/2017
Non-Publication Request: No
Early Publication Request: No
Title

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Preliminary Class

435

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 15/383,965

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A			N/A	280
SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A			N/A	600
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	720
TOTAL CLAIMS (37 CFR 1.16(i))	18 minus 20 =	*			OR	x 80 =	0.00
INDEPENDENT CLAIMS (37 CFR 1.16(h))	1 minus 3 =	*				x 420 =	0.00
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).						0.00
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))							0.00
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL			TOTAL	1600

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE($)	ADDITIONAL FEE($)	OR	RATE($)	ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

PACIFIC BIOSCIENCES OF CALIFORNIA, INC.
1380 Willow Road
MENLO PARK CA 94025

MAILED
JAN 30 2017
OFFICE OF PETITIONS

Doc Code: TRACK1.GRANT

Decision Granting Request for Prioritized Examination (Track I or After RCE)	Application No.: 15/383,965

1. THE REQUEST FILED December 19, 2016 IS **GRANTED.**

The above-identified application has met the requirements for prioritized examination

A. ☒ for an original nonprovisional application (Track I).
B. ☐ for an application undergoing continued examination (RCE).

2. **The above-identified application will undergo prioritized examination.** The application will be accorded special status throughout its entire course of prosecution until one of the following occurs:

A. filing a **petition for extension of time** to extend the time period for filing a reply;
B. filing an **amendment to amend the application to contain more than four independent claims, more than thirty total claims**, or a multiple dependent claim;
C. filing a **request for continued examination**;
D. filing a notice of appeal;
E. filing a request for suspension of action;
F. mailing of a notice of allowance;
G. mailing of a final Office action;
H. completion of examination as defined in 37 CFR 41.102; or
I. abandonment of the application.

Telephone inquiries with regard to this decision should be directed to Brian W. Brown at 571-272-5338.

/Brian W. Brown/
[*Signature*]

Petitions Examiner, Office of Petitions
(Title)

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

57770 7590 03/14/2017
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

EXAMINER
WILDER, CYNTHIA B

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
03/14/2017	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

PATENTS@PACB.COM

Office Action Summary	Application No. 15/383,965	Applicant(s) TURNER ET AL.	
	Examiner CYNTHIA B. WILDER	Art Unit 1637	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *1/30/2017*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-18* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-8* is/are rejected.

8)☒ Claim(s) *9-18* is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☒ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *1/20/2017* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date *12/19/2016*.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions.

Drawings

2. The drawings were received on 1/20/2017. These drawings are found acceptable by the Examiner.

Information Disclosure Statement

3. The information disclosure statement (IDS) submitted on 12/19/2016 is acknowledged. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Specification

4. The use of the term "SMRTbell" at paragraph [0044], which is a trade name or a mark used in commerce, has been noted in this application. It should be capitalized wherever it appears and be accompanied by the generic terminology.

Although the use of trade names and marks used in commerce (i.e., trademarks, service marks, certification marks, and collective marks) are permissible in patent applications, the proprietary nature of the marks should be respected and every effort made to prevent their use in any manner which might adversely affect their validity as commercial marks.

Claim Rejections - 35 USC § 101

4. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

5. Claim 1 is rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. The claimed invention is not directed to patent eligible subject matter. Based upon consideration of all of the relevant factors with respect to the claim as a whole, claim(s) 1 is/are determined to be directed to an abstract idea. The rationale for this determination is explained below:

The claims are directed to a method of determining a nucleotide sequence of a region of interest in a polynucleotide, the method comprising: introducing a polynucleotide comprising a region of interest to a sequence analysis system comprising a nanopore and an enzyme chaperone; monitoring variations in a parameter of the nanopore of the sequence analysis system during passage of the polynucleotide through the nanopore, wherein the enzyme chaperone regulates the rate of passage of the polynucleotide through the nanopore; analyzing the monitored variations of the nanopore of the sequence analysis system to determine a nucleotide sequence of the polynucleotide, wherein the nucleotide sequence comprises redundant sequence information for the region of interest; and determining a consensus sequence for the region of interest based on the redundant sequence information.

According to the Updates to the interim Eligibility Guidance an initial two step analysis is required for determining statutory eligibility. Step 1 requires a determination of whether the claims are directed to a process, machine, manufacture, or a composition of matter. In the instant case the Step 1 requirement is satisfied as the

claims are directed towards a process/machine. The Step 2 analysis is a two-part analysis, Step 2A and Step 2B, with the first part Step 2A requiring a determination of whether the claims are directed towards a judicial exception, i.e. a law of nature, natural phenomenon, or an abstract idea. In the instant case, the claims are drawn to the method(s) recited above comprising steps of "monitoring variations", "analyzing monitored variations" and "determining consensus sequence" which are considered inactive language and could be interpreted as being akin to reading data on a paper or the gathering of data via the use of a computer, computer program or software and mathematical calculations/algorithms to obtain desired sequence information. The results of Step 2A is that the claims are directed to an Abstract idea. The Court has made clear if a claim is directed essentially to a method of calculating, using a mathematical formula, even if the solution is for a specific purpose, the claimed method is non-statutory. In other words, patenting abstract ideas cannot be circumvented by attempting to limit the use [the idea] to a particular technological environment..

The second part, Step 2B of the two-step analysis is to determine whether any element or combination of elements, in the claim is sufficient to ensure that the claims as a whole amount to significantly more than the judicial exception. In this case, while the claims recite an active step of introducing a polynucleotide comprising a region of interest to a sequence analysis system comprising a nanopore and an enzyme chaperone, this step does not make clear how the steps of "monitoring", "analyzing" and "determining" are intended to take place and/or what active processes are involved in these steps. This step alone do not transform the abstract idea recited in the

"monitoring", "analyzing" and "determining" steps into a patent eligible application of the abstract idea such that the claims amount to significantly more than the abstract idea. The Court has made clear that to transform an unpatentable law of nature into patent-eligible application of such a law, one must do more than simply state the law of nature while adding the words "apply it". Essentially, appending conventional steps or elements, specified at a high level of generality, to laws of nature, natural phenomena, and abstract ideas cannot make those laws, phenomena, and ideas patent-eligible. In view of the foregoing, the claims are not drawn to patent eligible subject matter under 35 USC 101.

Double Patenting

6. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory double patenting rejection is appropriate where the claims at issue are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the reference application or patent either is shown to be commonly owned with this application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. A terminal disclaimer must be signed in compliance with 37 CFR 1.321(b).

The USPTO internet Web site contains terminal disclaimer forms which may be used. Please visit http://www.uspto.gov/forms/. The filing date of the application will determine what form should be used. A web-based eTerminal Disclaimer may be filled out completely online using web-screens. An eTerminal Disclaimer that meets all requirements is auto-processed and approved immediately upon submission. For more information about eTerminal Disclaimers, refer to http://www.uspto.gov/patents/process/file/efs/guidance/eTD-info-I.jsp.

7. Claims 1-8 are rejected on the ground of nonstatutory double patenting as being unpatentable over claims 1, 2, 5, 6, 7, 9, 10, 11, 12 of U.S. Patent No. 9057102. An obviousness-type double patenting rejection is appropriate where the conflicting claims are not identical, but an examined application claim is not patentably distinct from the reference claim(s) because the examined claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F. 2d 887, 225 USPQ 645 (fed. Cir. 1985). Although the claims at issue are not identical, they are not patentably distinct from each other.

Specifically, the claims 1-8 of the instant invention are drawn to the following:

> 1. A method of determining a nucleotide sequence of a region of interest in a polynucleotide, the method comprising: introducing a polynucleotide comprising a region of interest to a sequence analysis system comprising a nanopore and an enzyme chaperone; monitoring variations in a parameter of the nanopore of the sequence analysis system during passage of the polynucleotide through the nanopore, wherein the enzyme chaperone regulates the rate of passage of the polynucleotide through the nanopore; analyzing the monitored variations of the nanopore of the sequence analysis system to determine a nucleotide sequence of the polynucleotide, wherein the nucleotide sequence comprises redundant sequence information for the region of interest; and determining a consensus sequence for the region of interest based on the redundant sequence information.
>
> 2. The method of claim 1, wherein the nanopore comprises a protein channel.
>
> 3. The method of claim 1, wherein the nanopore is in a membrane.
>
> 4. The method of claim 3, wherein the membrane is a lipid bilayer.
>
> 5. The method of claim 3, wherein the membrane is a solid-state membrane.
>
> 6. The method of claim 3, wherein a voltage is applied across the membrane.
>
> 7. The method of claim 6, wherein the parameter in the monitoring step is the ionic current of an electrolyte solution passing through the nanopore.
>
> 8. The method of claim 1, further comprising changing reaction conditions to alter the speed at which the enzyme chaperone guides the polynucleotide through the nanopore.

The claims 1, 2, 5, 6, 7, 9, 10, 11, and 12 of US Patent 9057102 are drawn to the following:

> 1. A method of performing a nucleic acid sequencing reaction, comprising: a) preparing a reaction mixture comprising components of the nucleic acid sequencing reaction including a polynucleotide to be sequenced; b) introducing the reaction mixture to a membrane comprising a nanopore; c) initiating the nucleic acid sequencing reaction in the reaction mixture to begin progression of the nucleic acid sequencing reaction, wherein the nucleic acid sequencing reaction comprises passage of the polynucleotide or a nascent strand complementary thereto through the nanopore and monitoring the nanopore as

voltage is applied across the membrane; and d) maintaining conditions that allow the nucleic acid sequencing reaction to proceed while subjecting the reaction mixture to at least two detection periods and at least one non-detection period during the progression of the nucleic acid sequencing reaction, wherein the progress of the polynucleotide or the nascent strand through the nanopore is slowed during the detection periods to allow sequence determination, and further wherein the progress of the polynucleotide or the nascent strand through the nanopore is sped up during the non-detection period, thereby performing the nucleic acid sequencing reaction.

2. The method of claim 1, wherein the polynucleotide or the nascent strand is drawn through the nanopore by an electrophoretic method.

5. The method of claim 4, wherein the enzymatic method comprises a protein chaperone that directs the polynucleotide or the nascent strand through the nanopore.

6. The method of claim 5, wherein the progress of the polynucleotide or the nascent strand through the nanopore is sped up by increasing the speed at which the protein chaperone directs the polynucleotide or the nascent strand through the nanopore.

7. The method of claim 5, wherein the protein chaperone is a polymerase enzyme.

9. The method of claim 1, wherein the monitoring comprises detecting changes in ionic current through the nanopore.

10. The method of claim 1, wherein the membrane is a lipid bilayer.

11. The method of claim 1, wherein the membrane is a solid-state membrane.

12. The method of claim 1, wherein the nanopore comprises a protein channel.

The claims of US 9057102 differs from the instant invention in that the claims 1-8 of the instant invention are broader in scope. Additionally, the claims 1, 2, 5, 6, 7, 9, 10, 11, and 12 of US Patent 9057102 do not expressly recite wherein the polynucleotide to be sequence comprises redundant sequence information for a consensus sequence.

However, the specification of US 9057102 defines that the sequences as used therein refers to in one embodiment scaffold or consensus sequences (col. 18, first full paragraph). Further definition of the term(s) comprise of overlapping, redundant sequences. Thus, The claims 1-8 of the instant invention falls within the teachings of the claims 1, 2, 5, 6, 7, 9, 10, 11, and 12 of US Patent 9057102 and are therefore obvious over the teaching of the claims 1, 2, 5, 6, 7, 9, 10, 11, and 12 of US Patent 9057102.

As the court stated in *In re Goodman*, 29 USPQ2d 2010 (CAFC 1993), "a second application-- "containing a broader claim, more generical in its character than the specific claim in the prior patent"--typically cannot support an independent valid patent. Miller, 151, U.S. at 198; See Stanley, 214 F.2d at 153. Thus, the generic invention, as noted above is "anticipated" by the species of the patented invention. Cf., Titanium metal corp. v. Banner, 778 F.2d 775, 227 USPQ 773 (Fed. Cir. 1985) (holding that an earlier species disclosure in the prior art defeats any generic claims). This court's predecessor has held that, without a terminal disclaimer, the species claims preclude issuance of the generical application. "*In re Van Ornum*, 686 F.2d 937, 944, 214 USPQ 761, 767 (CCPA 1982); *Schneller*, 397 F.2d at 354".

Prior art

6. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Eid et al (20090087850, citation made of record on IDS) teach a sequencing method that uses an exonuclease that comprises template dependent

nucleobase binding activity. Eid et al teach wherein the method is configured to detect one or more luminescent signals during the sequencing reaction.

White et al (US 20080318184, cited on IDS) teach a nanopore device for single molecule detection, DNA sequencing or other suitable applications. White teaches wherein a glass nanopore membrane with a supported lipid bilayer is used. White teaches wherein one or more protein ion channels may be inserted into the bilayer and used form DNA sequencing (see page 5).

The references cited above differ from the instant invention in that they do not expressly teach the combination of method steps recited in the claims. While the teachings of White recognizes the suitability of using a membrane comprising nanopore in a DNA sequencing reaction, the prior art alone, or in combination, does not provide any reasonable motivation for performing the recited method steps comprising maintaining conditions that allow the nucleic acid sequencing reaction to proceed while subjecting the reaction mixture to at least two detection periods and at least one non-detection period during the progression of the nucleic acid sequencing reaction, wherein the progress of the polynucleotide or the nascent strand through the nanopore is slowed during the detection periods to allow sequence determination, and further wherein the progress of the polynucleotide or the nascent strand through the nanopore is sped up during the non-detection period. The prior art does not teach the collection of data during the non-detection period because the art is concerned with enabling a detection period during analytical process.

Conclusion

7. Claims 1-8 are rejected. Claims 9-18 are objected because they depend from rejected claims. Any inquiry concerning this communication or earlier communications from the examiner should be directed to CYNTHIA B. WILDER whose telephone number is (571)272-0791. The examiner can normally be reached on a flexible schedule.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on (571) 272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/CYNTHIA B WILDER/
Primary Examiner, Art Unit 1637

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15383965	TURNER ET AL.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

CPC- SEARCHED		
Symbol	Date	Examiner
C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333	3/6/2017	CW

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES		
Search Notes	Date	Examiner
Inventor search on PALM, review of related applictions as cited in PALM, review of IDS filed in instant invention and cited documents in copending applications, text search on WEST	3/6/2017	CW

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner

INTERFERENCE SEARCH			
US Class/ CPC Symbol	**US Subclass / CPC Group**	**Date**	**Examiner**
C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333		3/7/2017	CW

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99**)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

U.S.PATENTS						
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5001050		1991-03-19	Blanco et al.	
	2	5198543		1993-03-30	Blanco et al.	
	3	5350686		1994-09-27	Jhingah	
	4	5470724		1995-11-28	Ahern	
	5	5547839		1996-08-20	Dower et al.	
	6	5576204		1996-11-19	Blanco et al.	
	7	5648245		1997-07-15	Fire et al.	
	8	5674683		1997-10-07	Kool	

/Cynthia Wilder/ 03/09/2017

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	9	5674716		1997-10-07	Tabor et al.	
	10	5714320		1998-02-03	Kool	
	11	5854033		1998-12-29	Lizardi	
	12	6210891	B1	2001-04-03	Nyren et al.	
	13	6210896		2001-04-03	Chan	
	14	6255083		2001-07-03	Williams	
	15	6261808		2001-07-17	Auerbach	
	16	6369038		2002-04-09	Blumenfeld et al.	
	17	6451563		2002-09-17	Wittig et al.	
	18	6498023	B1	2002-12-24	Abarzua	
	19	6787308		2004-09-07	Balasubramanian	

/CYNTHIA B WILDER/ (03/06/2017)

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	
Examiner Name	Not Yet Assigned
Attorney Docket Number	01-007706US

	20	6917726		2005-07-12	Levene et al.	
	21	7013054		2006-03-14	Levene et al.	
	22	7033764		2006-04-25	Korlach et al.	
	23	7045362	B2	2006-05-16	Hartwich et al.	
	24	7052847		2006-05-30	Korlach et al.	
	25	7056661		2006-06-06	Korlach et al.	
	26	7056676		2006-06-06	Korlach et al.	
	27	7170050		2006-06-06	Korlach et al.	
	28	7181122		2007-02-20	Levene et al.	
	29	7229799		2007-06-12	Williams et al.	
	30	7282337		2007-10-16	Harris et al.	

EFS Web 2.1.17

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	
Examiner Name	Not Yet Assigned
Attorney Docket Number	01-007706US

	31	7292742		2007-11-06	Levene et al.	
	32	7361466		2008-04-22	Korlach et al.	
	33	7368265		2008-05-06	Brenner et al.	
	34	7416844		2008-08-26	Korlach et al.	
	35	7476503		2007-01-30	Turner et al.	
	36	7485424		2009-02-03	Korlach et al.	
	37	7601495		2009-10-13	Chen et al.	
	38	7601499		2009-10-13	Berka et al.	
	39	7700287	B2	2010-04-20	Chen et al.	
	40	7754429	B2	2010-07-13	Rigatti et al.	
	41	7767400	B2	2010-08-03	Harris et al.	

/CYNTHIA B WILDER/ (03/06/2017)

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	42	6849404	B2	2005-02-01	Park et al.	
	43	8143030	B2	2012-03-27	Maxham et al.	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20010030290	A1	2001-10-18	Stern	
	2	20030044781	A1	2003-03-06	Korlach et al.	
	3	20030096253		2003-05-22	Nelson et al.	
	4	20030143550	A1	2003-07-31	Green et al.	
	5	20030190647		2003-10-09	Odera	
	6	20030207279	A1	2003-11-06	Crothers et al.	
	7	20030213771	A1	2003-11-20	Ohshita et al.	

/CYNTHIA B WILDER/ (03/06/2017)

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	
Examiner Name	Not Yet Assigned
Attorney Docket Number	01-007706US

	8	20030215862		2003-11-20	Parce et al.	
	9	20040048300		2004-03-11	Sood et al.	
	10	20040152119		2004-08-05	Sood et al.	
	11	20040203008	A1	2004-10-14	Uemori et al.	
	12	20040224319		2004-11-11	Sood et al.	
	13	20040259082		2004-12-23	Williams	
	14	20050176035	A1	2005-08-11	Crothers et al.	
	15	20060061754	A1	2006-03-23	Turner et al.	
	16	20060292611	A1	2006-12-28	Berka et al.	
	17	20070062934	A1	2007-03-22	King	
	18	20070161017	A1	2007-07-12	Eid et al.	

/CYNTHIA B WILDER/ (03/06/2017)

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	19	20070178482	A1	2007-08-02	Lezhava et al.	
	20	20070269825	A1	2007-11-22	Wang et al.	
	21	20080026393	A1	2008-01-31	Mindrinos et al.	
	22	20080233575		2008-09-25	Harris et al.	
	23	20090197257	A1	2009-08-06	Harris	
	24	20090269771		2009-10-29	Schroeder	
	25	20090305248		2009-12-10	Lander et al.	
	26	20020197618	A1	2002-12-26	Sampson	
	27	20080009007		2008-01-10	Lyle et al.	
	28	20080176241		2008-07-24	Eid et al.	
	29	20090005252	A1	2009-01-01	Drmanac et al.	

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	30	20090087850		2009-04-02	Eid et al.	
	31	20090233291	A1	2009-09-17	Chen et al.	
	32	20081218184	A1	2008-09-11	White et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	199106678	WO	A1	1991-05-16	SRI International		☐
	2	199416090	WO	A1	1994-07-21	Molecular Tool, Inc.		☐
	3	199627025	WO	A1	1996-09-06	Rabani		☐
	4	199905315	WO	A2	1999-02-04	Medical Biosystems Ltd.		☐
	5	2007003017	WO	A1	2007-01-11	Biochip Innovations PTY		☐
	6	2007070572	WO		2007-06-21	US Govt as represented by DHHS		☐

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	7	2008058282	WO		2008-05-15	Complete Genomics Inc.		☐
	8	2009124255	WO	A2	2009-10-08	Helicos Biosciences Corp.		☐
	9	1225234	EP	B1	2007-11-07	Agilent Technologies		☐
	10	1907573	EP	B1	2010-01-27	Braslavsky		☐
	11	2007010263	WO	A2	2007-01-25	Swerdlow et al.		☐
	12	2008064905	WO	A2	2007-06-07	Eid et al.		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	BASHIR, A. et al., "Evaluation of paired-end sequencing strategies for detection of genome rearrangements in cancer" Plos CompBiol (2008) 4(4):1-14	☐
	2	EID, et al., "Real-time DNA sequencing from single polymerase molecules" Science (2009) 323(5910):133-138	☐
	3	HARRIS, T.D. et al., "Single-molecule DNA sequencing of a viral genome" Science (2008) 320:106-109	☐

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	4	HONG, Y.S. et al., "Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing" Mol Genet Genomics (2003) 268:720-728	☐
	5	~~KEANE, T. et al., "Assessing assemblability of reads from new sequencing platforms" Wellcome Trust Roster, Page 1~~	☐
	6	KOONIN et al. "Computer-assisted dissection of rolling circle DNA replication" Biosystems (1993) 30(1-3):241-268	☐
	7	KORBEL, J.O. et al. "Paired-end mapping reveals extensive structural variation in the human genome" Science (2007) 318:420-426	☐
	8	KUHN et al., "Rolling-circle amplificaiton under topological constraints" Nucl Acids Res (2002) 30(2):574-580	☐
	9	LEVENE et al., "Zero-mode waveguides for single-molecule analysis at high concentrations" Science (2003) 299 (5607):682-686	☐
	10	MATRAY, T.J. et al. "A specific partner for abasic damage in DNA" Nature (1999) 399:704-708	☐
	11	MYERS, G. "Whole-genome DNA sequencing" IEEE (May-June 1999) pp. 33-43	☐
	12	NOVICK "Contrasting lifestyles of rolling-circle phages and plasmids" Trends Biochem Sci (1998) 23(11):434-438	☐
	13	REIFENBERGER, J. et al., Advances in Genome Biol and Tech (AGBT) (2009) Abstract February 4-7, 2009	☐
	14	REIFENBERGER, J. et al., Biophys Soc 53rd Ann Meeting (2009) Abstract, February 28, 2009	☐

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	15	SMITH, M. et al., "Genomic sequence sampling: a strategy for high resolution sequence-based physical mapping of complex genomes" Nature Genetics (1994) 7:40-47	☐
	16	SPINELLA et al., "Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles" Nucl Acids Res (1999) 27(18):e22-e22	☐
	17	VELCULESCU et al. "Serial analysis of gene expression" Science (1995) 270(5235): 484-487	☐
	18	VOLIK, S. et al., "End-sequence profiling: sequence-based analysis of aberrant genomes" PNAS (2003) 100 (13):7696-7701	☐
	19	WILEY, G. et al., "Methods for generating shotgun and mixed shotgun/paired-end libraries for the 454 DNA sequencer" Current Protocols in Human Genomics (2009) Chapter 18; Unit 18.1, pp. 1-21	☐
	20	Technology Spotlight: Illumina Sequencing Technology, current of 10/08/08, pp 1-4	☐
	21	HORMOZDIARI, et al. "Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes, " Genome Research (2009) 19:1270-1278	☐
	22	LEE, et al., "A robust framework for detecting structural variations in a genome," Bioinformatics (2008) 24:i59-i67.	☐
	23	MARGULIES, et al., "Genome sequencing in microfabricated high-density picolitre reactors," Nature (2005), 437:376-382.	☐
	24	PEDLER, "Occupation Times for Two State Markov Chains," Journ Appl Probability (1971), 8(2):381-90.	☐
	25	SVOBODA, et al., "Fluctuation analysis of motor protein movement and single enzyme kinetics," PNAS (1994), 91:11782-86.	☐

/CYNTHIA B WILDER/ (03/06/2017)

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	26	International Search Report and Written Opinion dated 4/29/2010 for related case PCT/US2009/005169.	☐
	27	International Search Report and Written Opinion dated 10/27/2009 for related case PCT/US2009/001930	☐
	28	International Preliminary Report on Patentability dated 10/7/2010 for related case PCT/US2009/001930	☐
	29	International Search Report and Written Opinion dated 11/17/2009 for related case PCT/US2009/001926	☐
	30	METZKER, M.I., "Emerging Technologies in DNA Sequencing," Genome Research (2005) 15:1767-1776	☐
	31	International Preliminary Report on Patentability dated 4/7/2011 for related case PCT/US2009/005169	☐
	32	International Search Report and Written Opinion dated 11/3/2009 for related case PCT/US2009/001927	☐
	33	International Preliminary Report on Patentability dated 4/7/2011 for related case PCT/US2009/001927	☐
	34	Supplementary European Search Report dated March 20, 2012 for related case EP 09816557.4	☐
	35	First Exam Report dated 7/18/13 of related EP 09816557.4	☐
	36	Second Exam Report dated 4/9/2015 of related EP 09816557.4	☐

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	37	Third Exam Report dated 3/30/16 of related EP 09816557.4	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature	/CYNTHIA B WILDER/ (03/06/2017)	Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

WEST Search History for Application 15383965

Creation Date: 2017030620:18

Prior Art Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
(polynucleotide same nanopore same enzyme chaperone)	PGPB, USPT, USOC, EPAB, DWPI	10	ADJ	YES		03-06-2017
(nanopore same enzyme chaperone)	PGPB, USPT, USOC, EPAB, DWPI	10	ADJ	YES		03-06-2017
9057102.pn. or 8143030.pn.	PGPB, USPT, USOC, EPAB, DWPI	5	ADJ	YES		03-06-2017
15/383965	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		03-06-2017
(nanopore same enzyme).clm.	PGPB, USPT, USOC, EPAB, DWPI	54	ADJ	YES		03-06-2017
((nanopore same enzyme).clm.) same polynucleotide	PGPB, USPT, USOC, EPAB, DWPI	8	ADJ	YES		03-06-2017
(9057102.pn. or 8143030.pn.) and ((C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333).CPC.)	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		03-06-2017

20080318184	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		03-06-2017
(20080318184) and (consensus sequence or consensus or redundant sequence)	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		03-06-2017
nanopore same (consensus sequence)	PGPB, USPT, USOC, EPAB, DWPI	31	ADJ	YES		03-06-2017

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 8144

SERIAL NUMBER	FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
15/383,965	12/19/2016 RULE	435	1637	01-007706US

APPLICANTS

Pacific Biosciences of California, Inc., Menlo Park, CA;

INVENTORS

Stephen Turner, Seattle, WA;
Jon Sorenson, Alameda, CA;
Kenneth Mark Maxham, Redwood City, CA;
John Eid, San Francisco, CA;

**** CONTINUING DATA ***************************

This application is a CON of 14/708,603 05/11/2015 PAT 9556480
which is a CON of 14/091,961 11/27/2013 PAT 9057102
which is a CON of 12/982,029 12/30/2010 PAT 8628940
which claims benefit of 61/099,696 09/24/2008
and claims benefit of 61/139,402 12/19/2008
and is a CIP of 12/413,226 03/27/2009 PAT 8143030

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED ****

12/28/2016

Foreign Priority claimed ☐ Yes ☑ No 35 USC 119(a-d) conditions met ☐ Yes ☐ No Verified and Acknowledged /CYNTHIA B WILDER/ Examiner's Signature	☐ Met after Allowance /CBW/ Initials	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
		WA	24	18	1

ADDRESS

Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025
UNITED STATES

TITLE

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

FILING FEE RECEIVED		
1600	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

March 31, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
Inventor: Stephen Turner, et al.	Examiner: Wilder, Cynthia B.
Application No.: 15/383,965	Confirmation No.: 8144
Filed: December 19, 2016	Art Unit: 1637
For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS	RESPONSE TO NON-FINAL OFFICE ACTION

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INTRODUCTORY REMARKS

This amendment is filed in response to the Non-Final Office Action dated March 14, 2017 for which a 3-month period for response was given, and as such is timely filed. Applicant respectfully requests entry of the amendments and remarks provided herein.

Amendments to the Specification begin on page **2** of this paper.

Amendments to the Claims begin on page **13** of this paper.

Remarks/Arguments begin at page **18** of this paper.

AMENDMENTS TO THE SPECIFICATION

1. Please delete the title on page 1 and replace it with the following title:

NUCLEIC ACID SEQUENCE ANALYSIS

2. Please replace paragraph [0044] with amended paragraph [0044] below:

[0044] Figure 6 provides an illustration of paths in a sequence alignment matrix representing sequencing data from a ~~SMRTbell~~ SMRTBELL™ template (a template comprising a double-stranded portion comprised of two complementary sequences and two single-stranded linking portions (e.g., oligos or "hairpins") joining the 3' end of each strand of the double-stranded region to the 5' end of the other strand).

3. Please replace paragraph [0056] with amended paragraph [0056] below:

[0056] Figure 18 provides an exemplary strategy for selectively reducing the size of a duplex fragment within a ~~SMRTbell~~ SMRTBELL™ template.

4. Please replace paragraph [00103] with amended paragraph [00103] below:

[00103] Various methods can be used to control or monitor the progress of a polymerase on a template nucleic acid. For example, as noted above, one may employ a reaction stop or pause point within the template sequence, such as a reversibly bound blocking group at one location on the template, e.g., on the single-stranded portion that was not used in priming. Reaction stop or pause points can be engineered into a portion of the template for which the nucleotide sequence is unknown (e.g., a genomic fragment), but is preferably located within a portion for which the nucleotide sequence is known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, certain preferred sequencing templates (e.g., ~~SMRTbell~~ SMRTBELL™ templates, described elsewhere herein) are closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, and one or

both of these linkers can comprise a stop or pause point to control the passage of the polymerase through them. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

5. Please replace paragraph [00152] with amended paragraph [00152] below:

[00152] The target sequence consists of the potential hypotheses for the molecular template in question. In the example of nucleic acid sequencing methods using iterative illumination for sequencing a shotgun fragment from a linear DNA sequence, the potential hypotheses are both orientations of the genome (since we do not know the original orientation of the fragment). In the example of sequencing of a ~~SMRTbell~~ SMRTBELL™ template (e.g., see Example 1 herein), the hypotheses include both orientations of the genome and known adapter sequences. The parameters determining how many hits are reported for each local fragment can be varied to change the specificity and sensitivity of this algorithm. Figure 6 shows what these hits might look like for a ~~SMRTbell~~ SMRTBELL™ template (represented as paths in the sequence alignment matrix, which is often called the dynamic-programming matrix, although it isn't necessary to use dynamic programming to find these paths).

6. Please replace paragraph [00153] with amended paragraph [00153] below:

[00153] After the potential local alignments have been enumerated, a weighted directed graph is constructed with each local alignment represented as a node in the graph (step 3). The edges are drawn between nodes if they represent a potential reconstruction of the original molecular template using knowledge of the expected molecular configuration. The directed connection of an alignment path A to an alignment path B is interpreted as "The target sequence represented by B could follow the target sequence represented by A in the original molecule." For example, if a linear single-stranded DNA molecule is being sequenced by a method that uses iterative illumination, then fragments from opposite orientations would not be expected to be

connected (unless the linear single-stranded DNA molecule also included oppositely oriented sequences, e.g., as in the case of a linearized ~~SMRTbell~~ SMRTBELL™ template.) In general, fragments that represent the same stretch of the sequencing read but that align to different regions on the target sequence would not be connected. Aside from these examples, the rules for connecting nodes should be fairly loose to permit exploration of weak possibilities that gain significance when all the evidence (e.g. all the sequence reads) are considered. The assignment of edge weights handles the proper weighting of the likelihood of these edges, and the speed of the algorithm can be tuned by optimizing the pruning of highly unlikely edges. As usual this represents a tradeoff between speed and sensitivity.

7. Please replace paragraph [00156] with amended paragraph [00156] below:

[00156] In a ~~SMRTbell~~ SMRTBELL™ template example, knowledge of the expected insert size and the observed distance and orientation between fragments would be used to weight the likelihood that these two fragments could come from a correctly generated ~~SMRTbell~~ SMRTBELL™ template. This weight could include the expected rate of the polymerase as well and rules for the orientation of fragments with respect to each other and their distance apart in the original read. For example, while it may be expected that two forward fragments mapping to the same region in the target genome potentially come from multiple passes around a ~~SMRTbell~~ SMRTBELL™ template molecule, those fragments would not be expected to be immediately adjacent in sequencing time. The weighting function would account for the proper amount of expected time between such fragments (i.e. the elapsed time would be expected to be long enough to include two adapter sequences and a reverse sequence).

8. Please replace paragraph [00175] with amended paragraph [00175] below:

[00175] Although useful in certain preferred embodiments of the invention, certain algorithms as presented above do not easily handle the case where the template does not match a physically-motivated expected model. A relevant example of such a case is when the template contains a genomic structural variation (SV), such as translocation, whereby two fragments which are correctly adjacent in the template are located very far apart in the reference genome.

Such structural variation cases are best handled in the context of the current algorithm by reporting the confidence of an observed path and reporting situations when no physically expected path seems to fit the observed data. In general, the detection of structural variation requires the presence of multiple highly significant local alignments which can be identified as significantly overturning the null hypothesis of matching the genomic ordering of fragments with their own individual merit. Nevertheless, with molecular redundant sequencing such as ~~SMRTbell~~ SMRTBELL™ template sequencing the current algorithm can be adapted to improve the ability to identify an SV event. Such a modification could be a feedback approach which allows modification of the linking constraints in step 3 to allow very far separations on the target sequence when the individual alignments are very significant. Only one such highly-significant pair would be needed to enable the rescue of less significant partial matches that support the same SV hypothesis.

9. Please replace paragraph [00201] with amended paragraph [00201] below:

[00201] As noted above, using templates that allow repeated sequencing (e.g., circular templates, ~~SMRTbell~~ SMRTBELL™ templates, etc.) in a single reaction can increase the percent of a nucleic acid template for which nucleotide sequence data is generated and/or increase the fold-coverage of the sequence reads for one or more regions of interest in the template, thereby providing more complete data for further analysis, e.g., construction of sequence scaffolds and/or consensus sequences for the nucleic acid template. For example, in certain preferred embodiments, templates sequenced by the methods described herein are templates comprising a double-stranded segment, e.g., greater than 75%, or even greater than 90% of the target segment will be double-stranded or otherwise internally complementary. Such templates may, for example, comprise a double-stranded portion comprised of two complementary sequences and two single-stranded linking portions (e.g., oligos or "hairpins") joining the 3' end of each strand of the double-stranded region to the 5' end of the other strand (sometimes referred to as "~~SMRTbell~~ SMRTBELL™" templates). In certain embodiments, double-stranded portions for use in such templates are PCR-amplified. Optionally, restriction sites are incorporated within the PCR primers such that subsequent digestion of the amplified products with appropriate restriction enzymes generates double-stranded portions containing known overhang sequences

on either end, which are then ligated to hairpin adapters containing a complementary overhang to generate the ~~SMRTbell~~ SMRTBELL™ templates.

10. Please replace paragraph [00202] with amended paragraph [00202] below:

[00202] These template molecules are particularly useful as nucleotide sequence data generated therefrom comprises both sense and antisense nucleotide sequences for the double-stranded portion, and the circular conformation of the template enables repeated sequencing (e.g., using a polymerase capable of strand-displacement) provides duplicative or redundant sequence information. Restated, a sequence process may progress around the completely contiguous sequence repeatedly obtaining sequence data for each segment from the complementary sequences, as well as sequence data within each segment, by repeatedly sequencing that segment. Iterative illumination is useful in such sequencing applications, e.g., to focus nucleotide sequence data collection on stages of the sequencing reaction most of interest, such as the stages during which nucleotide sequence data is being generated from a strand of the (previously) double-stranded portion. Iterative illumination may also allow additional "rounds" of sequencing the template by virtue of the reduction in photo-induced damage to reaction components, as described elsewhere herein, thereby providing more complete and robust nucleotide sequence data for future analysis, e.g., sequence scaffold construction and/or consensus sequence determination. Further, as described above, the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. These factors can be used to determine the appropriate timing for the illuminated and non-illuminated periods depending on the experimental objectives of the practitioner, whether it be maximizing length or depth of sequence coverage on a given template nucleic acid, or optimizing sequence data collection from particular regions of interest, e.g., from the ends of the double-stranded portion of a ~~SMRTbell~~ SMRTBELL™ template.

11. Please replace paragraph [00210] with amended paragraph [00210] below:

[00210] As noted elsewhere herein, stop or pause points can be engineered into various portions of the template, e.g., portions for which the nucleotide sequence is unknown (e.g., a genomic fragment) or known (e.g., an adaptor or linker ligated to the genomic fragment.) For example, ~~SMRTbell~~ SMRTBELL™ templates are topologically closed, single-stranded molecules having regions of internal complementarity separated by hairpin or stem-loop linkers, such that hybridization of the regions of internal complementarity produces a double-stranded portion within the template. One or both of the linkers can comprise a stop or pause point to modulate polymerase activity. In some embodiments, these regulatory sequences or sites cause a permanent cessation of nascent strand synthesis, and in other embodiments the reaction can be reinitiated, e.g., by removing a blocking moiety or adding a missing reaction component. Various types of pause and stop points are described below and elsewhere herein, and it will be understood that these can be used independently or in combination, e.g., in the same template molecule.

12. Please replace paragraph [00212] with amended paragraph [00212] below:

[00212] Uracil-DNA glycosylases can also be used to introduce abasic sites into a template nucleic acid comprising deoxyuridine nucleotides. This strategy has the advantage of allowing the practitioner to choose the locations of the abasic sites within a DNA template since deoxyuridine nucleotides are not generally found in DNA. Various methods of inserting deoxyuridine nucleotides into a DNA template may be used, and different methods will be preferred for different applications. In certain embodiments, one or more site-specific deoxyuracils are incorporated during standard phosphoramidite oligonucleotide synthesis. To place uracils at indeterminate positions in a DNA, replacing a portion of the deoxythymidine triphosphate with deoxyuridine triphosphate will result in an amplimer with random U sites in place of T sites after polymerase chain reaction. In other embodiments, deoxyuridine nucleotides are engineered into the template, e.g., by ligation of a synthetic linker or adaptor comprising one or more deoxyuridine nucleotides to a nucleic acid sequence to be sequenced. In certain

preferred embodiments, deoxyuridine nucleotides are incorporated into the linker portions of a ~~SMRTbell~~ SMRTBELL™ template.

13. Please replace paragraph [00217] with amended paragraph [00217] below:

[00217] In certain embodiments, synthesis blocking moieties are nicks in the template nucleic acid. Nicking enzymes (e.g., nicking endonucleases) are known in the art and can be used to specifically nick the template prior to or during a template-directed sequencing reaction. The use of site-specific nicking endonucleases allows the practitioner to incorporate a recognition sequence at a particular location within the template nucleic acid, and such nicking endonucleases are commercially available, e.g., from New England Biolabs, Inc. For example, a linker or adapter can be synthesized with a nicking endonuclease recognition sequence, ligated to a nucleic acid molecule to be sequenced, and can be specifically nicked either before or during a subsequent sequencing reaction. Nicks can also be introduced by ligating duplex segments that lack either a terminal 3'-hydroxy (e.g., have a dideoxynucleotide at the 3'-terminius) and/or 5'-phosphate group on one strand. The ligation results in covalent linkage of the phosphodiester backbone on one strand, but not on the other, which is therefore effectively "nicked." In certain embodiments, a ~~SMRTbell~~ SMRTBELL™ template is constructed using a duplex (or "insert") nucleic acid molecule lacking a 5'-phosphate group at one or both termini. Upon ligation of the hairpin or stem-loop adaptors at each end, nicks are created at one or both ligation site(s), depending on whether the duplex lacked a 5'-phosphate at one or both ends, respectively. In other embodiments, a ~~SMRTbell~~ SMRTBELL™ template is constructed using one or two stem-loop adaptors lacking a 3'-hydroxy group at the terminus (e.g., comprising a 2',3'-dideoxynucleotide rather than a 2'-deoxynucleotide). Upon ligation of one or two stem-loop adaptors lacking a 3'-hydroxy group, one or two nicks are created at the ligation site(s), depending on whether one or two adaptors lacked the 3'-hydroxy group, respectively. In both cases, a nick is created in the template nucleic acid, and a primer bound to one of the adaptors provides an initiation site for the polymerase, which will process the template until encountering a nick, at which point the polymerase will terminate the reaction by dissociation from the template. Regardless of how a nick is created, the position of a nick relative to the initiation site for the polymerase determines how much of the template will be sequenced. For example, Figure

19A provides an illustrative example of an embodiment in which a nick is present on a first strand of a duplex portion at a position distal to the adaptor containing the primer binding site. The first strand is processed by a polymerase, but the complementary strand is not processed because the polymerase dissociates at the nick site. An alternative embodiment is shown in Figure 19B, in which a nick is present on the strand complementary to the first strand at a position proximal to the adaptor containing the primer binding site. In this case both the first and complementary strands, as well as the adaptor not containing the primer binding site, are processed by the polymerase prior to dissociation. The position of the primer binding site also determines how much of the template is processed by the polymerase. Figure 19C provides a template having a primer binding site at a position from which a polymerase would process a significant portion of the adaptor prior to entering the duplex portion. An additional advantage to using a 3'-dideoxynucleotide at a nick is that it prevents the use of the nick as a polymerse initiation site, since strand extension requires a 3-hydroxy group. As such, the resulting nick would not compete with a primer site for initiation of nascent strand synthesis by the polymerase. Having a single, known site of initiation on a template molecule is beneficial, e.g., for subsequent mapping of a read generated in such a reaction. In certain preferred embodiments, a nick site both lacks a 5'-phosphate group and comprises a 3'-dideoxynucleotide.

14. Please replace paragraph [00235] with amended paragraph [00235] below:

[00235] A nucleic acid template was provided that comprised a double-stranded region and two single-stranded linker portions at each end. The first linker portion connected the 3' end of the sense strand with the 5' end of the antisense strand, and the second linker portion connected the 3' end of the antisense strand with the 5' end of the sense strand. This template was designed to form a single-stranded circle of approximately 500 bases when the double-stranded region was opened (e.g., by heat denaturization, helicase activity, etc.), and is sometimes referred to as a ~~SMRTbell~~ SMRTBELL™ template. A plurality of this nucleic acid template was incubated with polymerases, primers, and other reaction components to allow formation of polymerase-template complexes. (See, e.g., Korlach, J., et al. (2008) Nucleosides, Nucleotides and Nucleic Acids, 27:1072-1083; and Eid, J. (2009) Science 323:133-138.) The complexes were immobilized in zero-mode waveguides in a reaction mixture containing all

necessary buffer and nucleotide analog components for carrying out sequencing-by-synthesis reactions with the exception of a cognate starting base and a metal dication. A Smith-Waterman algorithm was used to perform the alignment of the known sequence of the template with the sequence reads generated in the reaction, and the positions of the sequence reads is graphically illustrated in Figure 8.

15. Please replace paragraph [00238] with amended paragraph [00238] below:

[00238] As in Example I, a ~~SMRTbell~~ SMRTBELL™ template was used. For templates of defined sequence, PCR was used to generate 3 or 6 kb DNA inserts for the double-stranded region in the ~~SMRTbell~~ SMRTBELL™ templates using a standard PCR methodology. For genomic and other biological samples, a DNA fragmentation protocol was used that generates DNA fragments distributed around 3 or 6 kb. Generation of fragments in these ranges was done using a ~~HydroShear~~ HYDROSHEAR® (Genomic Solutions[[®]]) device with settings recommended by the manufacturer. The random genomic DNA fragments were enzymatically treated to generate blunt ends. Both the PCR products and randomly generated DNA fragments were phosphorylated and then immediately put into a ligation reaction with a blunt hairpin adapter. The products were purified through two size selection steps using reduced volumes of ~~AMPure~~ AMPURE® magnetic beads (Agencourt[[®]]) to remove hairpin dimers and other short products. (Fabrication of ~~SMRTbell~~ SMRTBELL™ templates is further described elsewhere herein.)

16. Please replace paragraph [00239] with amended paragraph [00239] below:

[00239] The system components used for polynucleotide sequencing using intermittent detection are comparable to single-molecule sequencing applications under constant illumination, which are described, e.g., in Eid, et al. (2009) Science 323:133-138. Specifically, the immobilization and sequencing buffer compositions, nucleotide analogs identity and concentration, polymerase, ZMWs, surface treatment and instrumentation were identical to the standard methodology. Modifications to the ~~SMRTbell~~ SMRTBELL™ template DNA and polymerase binding and immobilization and data acquisition protocols are as follows.

17. Please replace paragraph [00240] with amended paragraph [00240] below:

[00240] A binding solution was prepared by incubation of 3 or 6 kb DNA ~~SMRTbell~~ SMRTBELL™ templates (1-10 nM) with a 10-fold excess of DNA polymerase (10-100 nM, respectively) in 10 mM MOPS (pH 7.5), 10 mM KOAc, 100 mM DTT & 0.05% Tween-20 for 2 hours at 30°C, followed by 1 hour at 37°C and subsequent storage at 4°C prior to immobilization on the ZMWs. Immediately prior to immobilization, the binding solution was diluted in the standard immobilization solution (50 mM MOPS (pH 7.5), 75 mM KOAc, 5 mM DTT, 0.05% Tween-20) to the desired final concentration, typically 0.1 to 1 nM, and incubated for 30 to 60 minutes at 22°C. Post-immobilization chip preparation and sequencing initiation were identical to the standard methods.

18. Please replace paragraph [00241] with amended paragraph [00241] below:

[00241] The data acquisition protocol was similar to the standard application with coordinated modifications to the collection timing and ZMW positioning. In the standard acquisition procedure, a single long acquisition (~10 minutes) is performed for each ZMW. In the intermittent illumination acquisition procedure, multiple short acquisitions (~3 minutes) of sequence reads (also termed "strobe reads") were performed for each ZMW (during "detection periods") with an interval between each acquisition period during which no acquisition of sequence reads was performed ("non-detection period"). The duration of the interval between each acquisition of sequence reads was determined based upon a desired distance (i.e., number of nucleotide positions) between each sequence (or strobe) read, the polymerization rate of the polymerase, and the ~~SMRTbell~~ SMRTBELL™ template insert size.

19. Please replace paragraph [00242] with amended paragraph [00242] below:

[00242] ~~SMRTbell~~ SMRTBELL™ templates were generated as described above for AC223433, a fosmid clone comprising a sequence of an approximately 40 kb region of Homo sapiens chromosome 15. The reference sequences used to map the sequence reads generated in

the sequencing reactions were the publically available sequences of Homo sapiens chromosome 15 (Hg18; NCBI Build 36.1) and fosmid AC223433 (NCBI GenBank accession number). Table 1 shows the number of statistically significantly mapped sequence reads for several types of intermittent illumination sequencing reactions. The number of mappable "looks" is equivalent to the number of mappable sequence reads generated during detection periods for a single template molecule. For example, a "mapped 1-look read" means, for a single template molecule, only a single detection period generated a sequence read that could be mapped to the reference sequence.

AMENDMENTS TO THE CLAIMS

Please incorporate the following amendments to the claims of the subject application.

1. (Currently Amended) A method of determining a nucleotide sequence of a region of interest in a polynucleotide, the method comprising:

introducing a polynucleotide comprising a region of interest to a sequence analysis system comprising a nanopore **in a membrane** ~~and an enzyme chaperone~~;

applying a voltage across the membrane;

monitoring variations in ~~a parameter of~~ **ionic current through** the nanopore of the sequence analysis system during **enzyme chaperone-regulated** passage of the polynucleotide through the nanopore~~, wherein the enzyme chaperone regulates the rate of passage of the polynucleotide through the nanopore~~;

analyzing the monitored variations ~~of the nanopore of the sequence analysis system~~ **in ionic current** to ~~determine a~~ **obtain** nucleotide sequence [[of]] **information for** the polynucleotide, wherein the nucleotide sequence **information** comprises redundant sequence information for the region of interest; and

determining a consensus sequence for the region of interest based on the redundant sequence information.

2. (Original) The method of claim 1, wherein the nanopore comprises a protein channel.

3. (Canceled)

4. (Currently Amended) The method of claim [[3]] 1, wherein the membrane is a lipid bilayer.

5. (Currently Amended) The method of claim [[3]] 1, wherein the membrane is a solid-state membrane.

6. (Canceled)

7. (Canceled)

8. (Currently Amended) The method of claim 1, further comprising changing reaction conditions to alter the ~~speed at which the enzyme chaperone guides the polynucleotide~~ rate of enzyme chaperone regulated passage of the polynucleotide through the nanopore.

9. (Currently Amended) The method of claim 1, wherein the polynucleotide comprises complementary sequences of the region of interest~~, wherein the redundant sequence information comprises the nucleotide sequence of the complementary sequences~~.

10. (Currently Amended) The method of claim 9, wherein the complementary sequences are linked ~~by an oligonucleotide~~.

11. (Canceled)

12. (Original) The method of claim 1, wherein the polynucleotide comprises double-stranded DNA.

13. (Original) The method of claim 12, wherein the polynucleotide is greater than 75% double-stranded DNA.

14. (Original) The method of claim 12, wherein the polynucleotide is greater than 90% double-stranded DNA.

15. (Currently Amended) The method of claim 1, wherein the polynucleotide comprises a double-stranded portion comprising complementary ~~sequences~~ strands of the region of interest~~, wherein the redundant sequence information comprises the nucleotide sequence of the complementary sequences~~.

16. (Currently Amended) The method of claim 15, wherein the complementary ~~sequences~~ strands are linked ~~by an oligonucleotide~~.

17. (Canceled)

18. (Original) The method of claim 1, wherein the polynucleotide comprises multiple repeats of the region of interest, wherein the redundant sequence information comprises the nucleotide sequence of the multiple repeats.

19. (New) The method of claim 9, wherein the redundant sequence information comprises the nucleotide sequence of the complementary sequences.

20. (New) The method of claim 10, wherein the complementary sequences are linked by a linker comprising a nucleotide.

21. (New) The method of claim 20, wherein the linker comprises an oligonucleotide.

22. (New) The method of claim 21, wherein the oligonucleotide comprises a registration sequence.

23. (New) The method of claim 21, wherein the linker comprises a nick.

24. (New) The method of claim 10, wherein the complementary sequences are linked by a synthetic linker.

25. (New) The method of claim 24, wherein the synthetic linker is a carbon-based linker.

26. (New) The method of claim 15, wherein the redundant sequence information comprises the nucleotide sequence of the complementary strands.

27. (New) The method of claim 16, wherein the complementary strands are linked by a linker comprising a nucleotide.

28. (New) The method of claim 27, wherein the linker comprises an oligonucleotide.

29. (New) The method of claim 28, wherein the oligonucleotide comprises a registration sequence.

30. (New) The method of claim 28, wherein the linker comprises a nick.

31. (New) The method of claim 16, wherein the complementary strands are linked by a synthetic linker.

32. (New) The method of claim 31, wherein the synthetic linker is a carbon-based linker.

33. (New) The method of claim 8, wherein the monitoring comprises a detection period and a non-detection period, wherein the rate of passage of the polynucleotide through the nanopore by the enzyme chaperone is sped up during the non-detection period and slowed during the detection period, wherein the monitored variations in ionic current through the nanopore in the detection period is subjected to the analyzing and determining steps.

34. (New) The method of claim 33, wherein the monitoring comprises multiple detection periods.

35. (New) A method of determining nucleotide sequences of regions of interest in a plurality of polynucleotides, the method comprising:

introducing a plurality of polynucleotides to a sequence analysis system comprising nanopores in a membrane, wherein each of the plurality of polynucleotides comprises a double-stranded portion comprising complementary strands of a corresponding region of interest;

applying a voltage across the membrane;

monitoring variations in ionic current through each of the nanopores of the sequence analysis system during enzyme chaperone-regulated passage of the plurality of polynucleotides through the nanopores;

analyzing the monitored variations in ionic current to obtain nucleotide sequence information for the complementary strands of the double-stranded portion of each of the plurality of polynucleotides; and

determining a consensus sequence for the corresponding region of interest of each of the plurality of polynucleotides based on the nucleotide sequence information.

REMARKS

Formal Matters

Claims 3, 6, 7, 11 and 17 are canceled.

Claim 1 is amended to include subject matter from Claims 3, 6 and 7 as well as to clarify claim language.

Claims 4 and 5 are amended to changed their dependency due to the cancellation of Claim 3.

Claim 8 is amended to clarify claim language.

Claims 9 and 15 are amended to delete the "wherein" clauses, which are now present in New dependent Claims 19 and 26.

Claims 10 and 16 are amended to delete the phrase "by an oligonucleotide", which are present in New dependent Claims 21 and 28.

Dependent Claims 19 to 34 and Independent Claim 35 are added. Independent Claim 35 includes the subject matter of Claim 1 with certain additional elements related to determining the sequence of each region of interest in a plurality of polynucleotides, where each region of interest is in a double stranded portion of the polynucleotide.

Support for these claims can be found throughout the application and claims as filed, see, e.g., paragraphs [0013], [0027], [0033], [0065], [00116], [00203], [00216], [00217], and [00230].

No new matter has been added.

Information Disclosure Statement

Applicants bring the Examiner's attention to the information disclosure statement (IDS) filed in conjunction with this amendment. This IDS is in compliance with 37 CFR 1.97(c)(2).

Interview Summary

Applicants thank the Examiner for extending the courtesy of a telephonic interview on March 28, 2017 with the undersigned to discuss the Office Action dated March 14, 2017. In the interview, Applicants proposed specific actions to address each rejection, including claim amendments to overcome the 35 USC §101 rejection. While no final agreement was reached, this discussion was very helpful and significantly moved the application towards resolution.

Specification

The Examiner stated that the term "SMRTbell" (e.g., at paragraph [0044]) is a trade name or a mark used in commerce and thus should be capitalized wherever it appears and be accompanied by the generic terminology.

In response, Applicants have amended the specification to capitalize this term (SMRTBELL) and to provide a generic description at its first usage (paragraph [0044]).

Claim Rejection – 25 USC § 101

Claim 1 is rejected under 35 U.S.C. § 101 because the claimed invention is directed to non-statutory subject matter.

In making this rejection, the Examiner analyzed Claim 1 using the Steps set forth in the Updates to the Interim Eligibility Guidance to assessing compliance with § 101. In brief, the Examiner acknowledged that Claim 1 is a process/machine, and thus meets the Step 1 requirement. However, the Examiner asserted that Claim 1 recites steps that are directed to an Abstract idea that do not amount to significantly more than the judicial exception and thus not in compliance with Step 2A and 2B.

While not acquiescing to the correctness of this analysis, Applicants have amended Claim 1 to make clear that it is not directed to an Abstract idea. Specifically, Claim 1 has been amended to specify: the nanopore is "in a membrane" (previous Claim 3); "applying a voltage across the membrane" (previous Claim 6); "monitoring variations in ionic current through the nanopore" (previous Claim 7); and "analyzing the monitored variations in ionic current to obtain nucleotide sequence information for the polynucleotide".

As discussed in the interview, Applicants consider that these amendments to Claim 1 clearly place it in compliance with § 101, meeting the standards set forth in the Updates to the Interim Eligibility Guidance employed by the Office. Specifically, Claim 1 clearly sets forth specific non-abstract steps for determining a nucleotide sequence of a region of interest using a specific sequence analysis platform.

Applicants further note that new independent Claim 35 includes elements similar or identical to those specified above for independent Claim 1, and thus is also in compliance with § 101, meeting the standards set forth in the Updates to the Interim Eligibility Guidance employed

by the Office.

In view of the amendments to the Claims and arguments above, withdrawal of this rejection is respectfully requested.

Double Patenting

The Examiner has rejected Claims 1 to 8 on the ground of nonstatutory double patenting as being unpatentable over claims 1, 2, 5, 6, 7, 9, 10, 11, 12 of U.S. Patent No. 9,057,102.

While not acquiescing to the Examiner's arguments, Applicants provide herewith a Terminal Disclaimer over U.S. Patent No. 9,057,102 in compliance with 37 CFR 1.321(c) or 1.321(d) to overcome this rejection.

In view of this Terminal Disclaimer, this rejection may be withdrawn.

CONCLUSION

Based upon the foregoing amendments and remarks, Applicant believes the instant application is in condition for allowance and action toward that end is respectfully requested. The Commissioner is authorized to charge any additional required fees or credit any overpayment to Deposit Account No. 50-4427 (referencing Attorney Docket No. 01-007706US). If the Office believes there are additional issues that have not been addressed, the Office is encouraged to contact Applicant's undersigned representative at (650) 521-8127.

Respectfully submitted,

March 31, 2017
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF
CALIFORNIA, INC.
1305 O'Brien Drive
Menlo Park, CA 94025

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5308751		1994-05-03	Ohkawa et al.	
	2	6673615		2004-01-06	Denison et al.	
	3	8168380		2012-05-01	Chan	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20020028458	A1	2002-03-07	Lexow	
	2	20020168645	A1	2002-11-14	Taylor	
	3	20040248161	A1	2004-12-09	Rotherberg et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	4	20060147935	A1	2006-07-06	Linnarsson	
	5	20070190556	A1	2007-08-16	Brenner et al.	
	6	20070231808	A1	2007-10-04	Gouda et al.	
	7	20080051294	A1	2008-02-28	Gormely et al.	
	8	20060024711		2006-02-02	Lapidus et al.	
	9	20020094526		2002-07-18	Bayley et al.	
	10	20100216151		2010-08-26	Lapidus et al.	
	11	20040002077		2004-01-01	Taira et al.	
	12	20050069939		2005-03-31	Wang et a.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	1							☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	Response and Exhibit List of Oxford Nanopore Technologies Ltd., Oxford Nanopore Technologies, Inc., and Metrichor, Ltd. Dated 1/6/2017 To the Complaint of Pacific Biosciences of California, Inc., and Notice of Investigation in the Matter of "Certain Single-Molecule Nucleic Acid Sequencing Systems and Reagents, Consumables, and Software for Use with Same (Investigation No. 337-TA-1032)	☐
	2	Ezzevaz-Roulet et al., "Mechanical Separation of the Complementary Strands of DNA," Proc. Natl. Acad. Sci. USA, Vol. 94, pp 11935-11940 (Oct. 1997)	☐
	3	Greenleaf et al., "Single-Molecule, Motion-Based DNA Sequencing Using RNAPolymerase," Science, 313(5788): 801 (Aug. 2006)	☐
	4	Hattori et al, "Dideoxy Sequening Method Using Denatured Plasmid Templates," Analytical Biochemistry, Vol. 152, pp 232-238 (1986)	☐
	5	Hayashizaki et al., "A new method for constructing NotI linking and boundary libraries using a restriction trapper," Genomics, Vol. 14, pp 733-739 (1992)	☐
	6	Jarvie et al., "3K Long-Tag Paired End sequencing with the Genome Sequencer FLX System," BioTechniques, Vol. 44, No. 6, pp 829-831 (2008)"	☐
	7	Kalisch et al., "Covalently linked sequencing primer linkers (splinkers) for sequence analysis of restriction fragments," Gene, Vol. 44, pp 263-270 (1986)	☐
	8	Kambara et al., "Real Time Automated Simultaneous Double-Stranded DNA Sequencing Using Two-Color Fluorophore Labeling," Biotechnology, Vol 9, pp 648-651 (July 1991)	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	9	Kaur et al., "Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA," Nucleic Acids Research, Vol. 31, No. 6 e26 (2003)	☐
	10	"Kim et al., "DARFA: a novel technique for studying differential gene expression and bacterial comparative genomics," Biochemical and Biophysical Research Communications, Vol. 336, pp 168–174 (2005)"	☐
	11	Kuhn et al., "High-Purity Preparation of a Large DNA Dumbbell," Antisense & Nucleic Acid Drug Development, Vol. 11, pp 149-153 (2001)	☐
	12	Liu et al., "Rolling Circle DNA Synthesis: Small Circular Oligonucleotides as Efficient Templates for DNA Polymerases," J. Am. Chem. Soc., Vol. 118, pp 1587-1594 (1996)"	☐
	13	Luo et al., "Small interfering RNA production by enzymatic engineering of DNA (SPEED)," PNAS, Vol. 101 No. 15, pp 5494-5499 (Apr. 2004)	☐
	14	Miller et al., "Chain Terminator Sequencing of Double-stranded DNA with Built-In Error Detection," J. theor. Biol. Vol. 161, pp 407-429 (1993)	☐
	15	Taki et al., "Small-interfering-RNA expression in cells based on an efficiently constructed dumbbell-shaped DNA," Angew. Chem. Int. Ed. Vol. 43, pp 3160-3163 (2004)	☐
	16	Thelwell et al., "Mode of action and application of Scorpion primers to mutation detection," Nucleic Acids Research, Vol. 28, No. 19, pp 3752-3761 (2000)	☐
	17	Vercoutere, et al., "Discrimination among individual Watson-Crick base pairs at the termini of single DNA hairpin molecules," Nucleic Acids Research, Vol. 31, No. 4, pp 1311-1318 (2003)	☐
	18	Vercoutere, et al., "Rapid discrimination among individual DNA hairpin molecules at singlenucleotide resolution using an ion channel," Nature Biotechnology, Vol. 19, pp 248-252 (Mar. 2001)	☐
	19	Wiemann et al., "Simultaneous On-Line DNA Sequencing on Both Strands with Two Fluourescent Dyes," Analytical Biochemistry, Vol. 224, pp 117-121 (1995)"	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	20	Winters-Hilt et al., "Highly Accurate Classification of Watson-Crick Base pairs on Termini of Single DNA Molecules," Biophysical Journal, Vol. 84, pp 967–976 (Feb. 2003)	☐
	21	Woodside et al.. "Nanomechanical Measurements of the Sequence-Dependent Folding Landscapes of Single Nucleic Acid Hairpins," Proc. Natl. Acad. Sci. USA, Vol. 103, pp6190-6195 (Apr. 2006)"	☐
	22	Zanta et al., "Gene delivery: a single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus," Proc. Natl. Acad. Sci. USA, Vol. 96, pp 91-96 (Jan. 1999)	☐
	23	Definition of "Consensus Sequence" from Medical dictionary, Printed on 2/6/2017	☐
	24	KEANE, T. et al., "Assessing Assemblability of Reads from New sequencing Platforms" Wellcome Trust Poster, Page 1, 15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) & 6th European Conference on Computational Biology (ECCB), Vienna, Austria July 21-25, 2007	☐
	25	Verified Complaint of Pacific Biosciences of California, Inc. Under Section 337 of the Tariff Act of 1930 filed with the United States International Trade Commission on November 2, 2016	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David C. Scherer, Ph.D./	Date (YYYY-MM-DD)	2017-03-31
Name/Print	David C. Scherer, Ph.D.	Registration Number	56,993

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Electronic Patent Application Fee Transmittal

Application Number:	15383965
Filing Date:	19-Dec-2016
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Filer:	David Christopher Scherer/Jacqueline Lim
Attorney Docket Number:	01-007706US

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
CLAIMS IN EXCESS OF 20	1202	10	80	800
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
Submission- Information Disclosure Stmt	1806	1	180	180
STATUTORY OR TERMINAL DISCLAIMER	1814	1	160	160
		Total in USD ($)		**1140**

Electronic Acknowledgement Receipt	
EFS ID:	28799617
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	31-MAR-2017
Filing Date:	19-DEC-2016
Time Stamp:	13:42:31
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1140
RAM confirmation Number	033117INTEFSW00011112504427
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	01007706_2017-03-31_Trans.pdf	47267 a4a75912945f7fe0b21cadea5219a4967bb3f31d	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	01007706_2017-03-31_FeeTrans.pdf	61131 244b2931ee0235506451bd8748bd9782e4441d9e	no	1
Warnings:					
Information:					
3		01007706_2017-03-31_OA1Resp.pdf	99195 19337b5563d650760f7ca2ca88f9bd2038d7a396	yes	20

Multipart Description/PDF files in .zip description

Document Description	Start	End
Amendment/Req. Reconsideration-After Non-Final Reject	1	1
Specification	2	12
Claims	13	17
Applicant Arguments/Remarks Made in an Amendment	18	20

Warnings:

Information:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
4	Terminal Disclaimer Filed	01007706_2017-03-31_TermDiscl-9057102.pdf	53259 afa1c40ce760d334c33c3df49ae1ecb23e0f3cfd	no	1
Warnings:					
Information:					

5	Transmittal Letter	01007706_2017-03-31_IDS.pdf	24767 661ce7fd6ef55afc96395b503c7af95abf8d9944	no	2
Warnings:					
Information:					
6	Information Disclosure Statement (IDS) Form (SB08)	01007706_2017-03-31_SB08A.pdf	49568 8d80c41bacb1e9eeb4771593eede1c27cf290258	no	6
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
7	Other Reference-Patent/App/Search documents	2017-01-06_ONTPublicRespToComplaint-withExhibitList.pdf	17400133 dfb13c28b41367c8d9f642da7f41d6102da6d4c8f8e	no	43
Warnings:					
The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing					
Information:					
8	Other Reference-Patent/App/Search documents	DefinitionOfConsensusSequence_2017-02-06.pdf	251713 0acccde84df629bcbea772f663e11c3328f8bb5c	no	3
Warnings:					
Information:					
9	Non Patent Literature	Ezzevaz-Roulet_PNAS_p11935-11940.pdf	278437 8c311fa564159d6fc2e7080c800869ace82714c2	no	6
Warnings:					
Information:					
10	Non Patent Literature	Greenleaf_Science_p801-R.pdf	1889184 d57283417bcffb912fe9785d5851bbb9ab6h4Rna	no	3
Warnings:					
Information:					
11	Non Patent Literature	Hattori_AnalBiochem_p232-238-R.pdf	3021510 71e95e94a420c216f51c69e6373f8e982357c066	no	7
Warnings:					

Information:					
12	Non Patent Literature	Hayashizaki_Genomics_p733-739.pdf	453998 464c0018d9659fa23e3d063e0d7ac5165536f419	no	7
Warnings:					
Information:					
13	Other Reference-Patent/App/Search documents	ITCcomplaint_2016-11-02.pdf	4465987 43d59001fe9d3258e7a8f0b012cc3ebee90ce66	no	46
Warnings:					
Information:					
14	Non Patent Literature	Jarvie_BioTechniques_p829-831.pdf	250741 66a24da884025a13e082eb21f7e2e759b3e02614	no	3
Warnings:					
Information:					
15	Non Patent Literature	Kalisch_Gene_p263-270-R.pdf	1280415 1373aa9958e711fee2b6fdc4211f2fb1ea6345c5	no	8
Warnings:					
Information:					
16	Non Patent Literature	Kambara_Biotechnology_p648-651-R.pdf	3362104 d1e3148148d25c427d7c29bdf1084be550ada30e	no	4
Warnings:					
Information:					
17	Non Patent Literature	Kaur_NuclAcidsRes_e26.pdf	210688 3893a3ddda3f5dcf51e0dba6faaeffe18e217d60	no	7
Warnings:					
Information:					
18	Non Patent Literature	Kim_BiochemBiophysResComm_p168-174-R.pdf	6659764 7d70c26ac04168d9af6f356f0adc7de70b5cbaf6e	no	7
Warnings:					
Information:					

19	Non Patent Literature	Kuhn_AntisenseNuclAcidDrug Dev_p149-153-R.pdf	5219838 276ab643c90d3f1d20f38ca006e0e1f43f873 64d0	no	5
Warnings:					
Information:					
20	Non Patent Literature	Liu_JAmChemSoc_p1587-1594 -R.pdf	5961852 dfdc535e373e7e9024390237faZ40c8b196f d731	no	8
Warnings:					
Information:					
21	Non Patent Literature	Luo_PNAS_p5494-5499.pdf	1820609 21e2baf7f400dc6ca267c797e58a599d3a155 eb2d	no	7
Warnings:					
Information:					
22	Non Patent Literature	Miller_JTheorBiol_p407-429-R. pdf	8015063 4d2ea8057f6fde25c7c3afc92e3311a76c8a 649c	no	23
Warnings:					
Information:					
23	Non Patent Literature	Taki_AngewChemIntEd_p3160 -3163.pdf	214649 db93b8da98296379cbaabaca78ddab68d de070e	no	4
Warnings:					
Information:					
24	Non Patent Literature	Thelwell_NuclAcidsRes_p3752- 3761-R.pdf	10302785 b5570bbf35f2154c371c1a42f50f9929a531f 3e5	no	10
Warnings:					
Information:					
25	Non Patent Literature	Vercoutere_NatureBio_p248-2 52-R.pdf	5160898 134e9557ba2f8914f941ab7eb5fd369365d0 302	no	6
Warnings:					
Information:					

26	Non Patent Literature	Vercoutere_NuclAcidsRes_p1311-1318.pdf	301200 172625f166978c4 0f9b71c9d88a79d0ee59778d	no	8
Warnings:					
Information:					
27	Non Patent Literature	Wiemann_AnalBiochem_p117-121-R.pdf	1057351 96c9721ba31a8d6fc4f227bbb4c19d761a54e2ff	no	5
Warnings:					
Information:					
28	Non Patent Literature	Winters-Hilt_BiophysJourn_p967-976.pdf	6450961 5cb167e3a27608aace648b5d7146af2a93ff2d5e	no	10
Warnings:					
Information:					
29	Non Patent Literature	Woodside_PNAS_6190-6195.pdf	1979039 a70b9311b0afb176c5f834c4a3c15a260fae1903	no	7
Warnings:					
Information:					
30	Non Patent Literature	Zanta_PNAS_p91-96.pdf	2566555 90f616870f916876aaac49633f3ed8c279782d5	no	7
Warnings:					
Information:					
31	Non Patent Literature	Keane_WelcomeTrustPoster-R.pdf	2018582 8168a7714e97adccb8ea7d2ef3ea005cad48d10b3	no	1
Warnings:					
Information:					
32	Fee Worksheet (SB06)	fee-info.pdf	33805 e18eef1b19b910d2e631f1a532f90be48c5e0c2c	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	90963048		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/21 (10-08)
Approved for use through 11/30/2008. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	15/383,965
Filing Date	December 19, 2016
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

ENCLOSURES *(Check all that apply)*

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [x] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [x] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation
- [ ] Change of Correspondence Address
- [x] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) (please Identify below):

PTO Form SB06A with 25 refs.

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Pacific Biosciences of California, Inc.		
Signature	/David C. Scherer, Ph.D./		
Printed name	David C. Scherer, Ph.D.		
Date	March 31, 2017	Reg. No.	56,993

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below: via EFS-Web

Signature	/Jacqueline L. Lim/		
Typed or printed name	Jacqueline L. Lim	Date	March 31, 2017

PTO/SB/17 (03-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995 no persons are required to respond to a collection of information unless it displays a valid OMB control number

FEE TRANSMITTAL

Complete if known	
Application Number	15/383,965
Filing Date	December 19, 2016
First Named Inventor	Stephen Turner
Examiner Name	Wilder, Cynthia B.
Art Unit	1637
Practitioner Docket No.	01-007706US

☐ Applicant asserts small entity status. See 37 CFR 1.27.

☐ Applicant certifies micro entity status. See 37 CFR 1.29. Form PTO/SB/15A or B or equivalent must either be enclosed or have been submitted previously.

TOTAL AMOUNT OF PAYMENT ($) 1,140.00

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☑ Deposit Account Deposit Account Number: 57770 Deposit Account Name: Pacific Biosciences

For the above-identified deposit account, the Director is hereby authorized to (check all that apply):

☑ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, **except for the filing fee**

☑ Charge any additional fee(s) or underpayment of fee(s) under 37 CFR 1.16 and 1.17 ☑ Credit any overpayment of fee(s)

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES (U = undiscounted fee; S = small entity fee; M = micro entity fee)

	FILING FEES			SEARCH FEES			EXAMINATION FEES			
Application Type	U ($)	S ($)	M ($)	U ($)	S ($)	M ($)	U ($)	S ($)	M ($)	Fees Paid ($)
Utility	280	140*	70	600	300	150	720	360	180	
Design	180	90	45	120	60	30	460	230	115	
Plant	180	90	45	380	190	95	580	290	145	
Reissue	280	140	70	600	300	150	2,160	1,080	540	
Provisional	260	130	65	0	0	0	0	0	0	

* The $140 small entity status filing fee for a utility application is further reduced to $70 for a small entity status applicant who files the application via EFS-Web.

2. EXCESS CLAIM FEES

Fee Description	Undiscounted Fee ($)	Small Entity Fee ($)	Micro Entity Fee ($)
Each claim over 20 (including Reissues)	80	40	20
Each independent claim over 3 (including Reissues)	420	210	105
Multiple dependent claims	780	390	195

Total Claims		Extra Claims		Fee ($)		Fee Paid ($)
30	-20 or HP =	10	x	80	=	800.00

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims		Extra Claims		Fee ($)		Fee Paid ($)
	-3 or HP =		x		=	

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims	
Fee ($)	Fee Paid ($)

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $400 ($200 for small entity) ($100 for micro entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets		Extra Sheets		Number of each additional 50 or fraction thereof		Fee ($)		Fee Paid ($)
_____	- 100 =	_____	/ 50 =	_____ (round up to a whole number)	x	_____	=	_____

4. OTHER FEE(S) Fees Paid ($)

Non-English specification, $130 fee (no small or micro entity discount) _____

Non-electronic filing fee under 37 CFR 1.16(t) for a utility application, $400 fee ($200 small or micro entity) _____

Other (e.g., late filing surcharge): Terminal Disclaimer Fee; Information Disclosure Statement Fee 340.00

SUBMITTED BY

Signature	/David C. Scherer, Ph.D./	Registration No. (Attorney/Agent) 56,993	Telephone 650-521-8127
Name (Print/Type)	David C. Scherer, Ph.D.		Date March 31, 2017

PTO/AIA/26 (04-14)
Approved for use through 07/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TERMINAL DISCLAIMER TO OBVIATE A DOUBLE PATENTING REJECTION OVER A "PRIOR" PATENT	Docket Number (Optional) 01-007706US

In re Application of: Pacific Biosciences of California, Inc.

Application No.: 15/383,965

Filed: December 19, 2016

For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

The applicant, Pacific Biosciences of California, Inc., owner of 100 percent interest in the instant application hereby disclaims, except as provided below, the terminal part of the statutory term of any patent granted on the instant application which would extend beyond the expiration date of the full statutory term of **prior patent** No. 9,057,102 as the term of said **prior patent** is presently shortened by any terminal disclaimer. The applicant hereby agrees that any patent so granted on the instant application shall be enforceable only for and during such period that it and the **prior patent** are commonly owned. This agreement runs with any patent granted on the instant application and is binding upon the grantee, its successors or assigns.

In making the above disclaimer, the applicant does not disclaim the terminal part of the term of any patent granted on the instant application that would extend to the expiration date of the full statutory term of the **prior patent**, "as the term of said **prior patent** is presently shortened by any terminal disclaimer," in the event that said **prior patent** later:

expires for failure to pay a maintenance fee;
is held unenforceable;
is found invalid by a court of competent jurisdiction;
is statutorily disclaimed in whole or terminally disclaimed under 37 CFR 1.321;
has all claims canceled by a reexamination certificate;
is reissued; or
is in any manner terminated prior to the expiration of its full statutory term as presently shortened by any terminal disclaimer.

Check either box 1 or 2 below, if appropriate.

1. ☐ The undersigned is the applicant. If the applicant is an assignee, the undersigned is authorized to act on behalf of the assignee.

I hereby acknowledge that any willful false statements made are punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

2. ☑ The undersigned is an attorney or agent of record. Reg. No. 56,993

/David C. Scherer, Ph.D./ — Signature
March 31, 2017 — Date

David C. Scherer, Ph.D. — Typed or printed name

Senior Patent Agent — Title
650-521-8127 — Telephone Number

☑ Terminal disclaimer fee under 37 CFR 1.20(d) included

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

March 31, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
Attorney Docket No. 01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Stephen Turner et al.

Application No.: 15/383,965

Filed: December 19, 2016

For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Examiner: Wilder, Cynthia B.

Art Unit: 1637

Confirmation No: 8144

INFORMATION DISCLOSURE STATEMENT UNDER 37 CFR § 1.97 AND 1.98

COMMISSIONER FOR PATENTS
P. O. BOX 1450
Alexandria, CA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Sir:

This information Disclosure Statement is submitted:

(a) [ ] **Under 37 CFR 1.97(b)**
(Within three months of filing national application or date of entry of national application or before mailing date of First Office Action on the merits whichever occurs last).

(b) [X] **Under 37 CFR 1.97(c) together with either a:**
[ ] Statement under CFR 1.97(e) or
[X] $180.00 fee under 37 CFR 1.17(p)
(After the CFR 1.97(b) time period, but before a final action, notice of allowance, whichever occurs first).

(c) [ ] **Under 37 CFR 1.97(d) together with: a**
Statement under 37 CFR 1.97(e), and
$180.00 fee as set forth in 37 CFR 1.17(p).
(After a final action or notice of allowance, whichever occurs first, but before payment of the issue fee).

[ ] **STATEMENT UNDER 37 CFR 1.97(e)**

The undersigned certifies that:

[ ] Each item of information contained in the Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the statement, or

[ ] No item of information contained in the Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart foreign application , and to the knowledge of the undersigned after making reasonable inquiry, was known to any individual designated in 37 CFR §1.56 more than three months prior to the filing of the Information Disclosure Statement.

[ ] PRIOR APPLICATIONS

All references in the enclosed PTO Form SB08A which are required to be included in this submission, were disclosed in prior Patent Application No.: *, filed *, and, as such, copies thereof are not included pursuant to the provisions of 37 CFR 1.98(d).

[ ] FOREIGN LANGUAGE DOCUMENTS

A concise explanation of the relevance of foreign language patents, foreign language publications and other foreign language information listed on PTO form 1449, as presently understood by the individual(s) designated in 37 CFR 1.56 most knowledgeable about the content is given on the attached sheet, or where a foreign language patent is cited in a search report or other action by a foreign patent office in a counterpart foreign application, an English language version of the search or action which indicates the degree of relevance found by the foreign office is listed on form PTO 1449 and is enclosed herewith.

[X] FEE AUTHORIZATION

Please charge to Deposit Account No. 50-4427 the sum of $ 180.00 at anytime during the pendency of this application, please charge any fees required or credit any overpayment to Deposit Account No.50-4427.

Respectfully Submitted,

March 31, 2017
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF CALIFORNIA
1305 O'Brien Drive
Menlo Park, CA 94025
Phone: (650) 521-8127
Fax: (650) 323-9420
Email: dscherer@pacb.com

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PATENT APPLICATION FEE DETERMINATION RECORD
Substitute for Form PTO-875

Application or Docket Number	Filing Date	☐ To be Mailed
15/383,965	12/19/2016	

ENTITY: ☒ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*	X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*	X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT 03/31/2017

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	* 30	Minus	** 20	= 10	x $80 =	800
Independent (37 CFR 1.16(h))	* 2	Minus	***3	= 0	x $420 =	0
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	800

AMENDMENT

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1

LIE
CAROL BARNES

Application Number	Application/Control No.	Applicant(s)/Patent under Reexamination	
	15/383,965	TURNER ET AL.	

Document Code - DISQ	Internal Document – DO NOT MAIL

TERMINAL DISCLAIMER	☒ APPROVED	☐ DISAPPROVED
Date Filed : 3/31/17	This patent is subject to a Terminal Disclaimer	

Approved/Disapproved by:
Felicia D. Roberts 9,057,102

U.S. Patent and Trademark Office

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

April 28, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Inventor: Stephen Turner, et al.

Application No.: 15/383,965

Filed: December 19, 2016

For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Examiner: Wilder, Cynthia B.

Confirmation No.: 8144

Art Unit: 1637

REQUEST TO CHANGE INVENTORSHIP PURSUANT TO 37 C.F.R. 1.48 (a)

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Pursuant to 37 C.F.R. 1.48(a), please add Cheryl Heiner and Kevin Travers as inventors. A replacement application data sheet annotated to reflect this change is provided herewith.

Please charge i) the required $600.00 fee under 37 C.F.R. 1.48(c) and 1.17(d); and ii) the required $140.00 fee under 37 C.F.R. 1.17(i) for entry and consideration of this request, to deposit account number 50-4427. Please charge any other additional fee that might be required for entry and consideration of this request to the above noted deposit account, or credit any overpayment.

Respectfully submitted,

April 28, 2017
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF CALIFORNIA, INC.
1305 O'Brien Drive
Menlo Park, CA 94025

SUPPLEMENTAL APPLICATION DATA SHEET

APPLICATION INFORMATION

Application Type::	Regular
Title::	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
Attorney Docket Number::	01-007706US
Request for Early Publication?::	No
Request for Non-Publication?::	No
Total Drawing Sheets ::	24
Small Entity?::	No
Petition included?::	No
Licensed US Govt. Agency::	No
Contract or Grant Numbers::	Not Applicable
Secrecy Order in Parent Application?::	No

INVENTOR INFORMATION

Inventor 1

Primary Citizenship Country::	United States
Given Name::	Stephen
Middle Name::	
Family Name::	Turner
City of Residence::	Seattle
State or Province of Residence::	Washington
Country of Residence::	United States
Street of mailing address::	4216 NE 113th Street
City of mailing address::	Seattle
State or Province of mailing address::	Washington
Postal or Zip Code of mailing address::	98125

Inventor 2

Primary Citizenship Country::	United States
Given Name::	Jon
Middle Name::	
Family Name::	Sorenson
City of Residence::	Alameda
State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	1725 Nason Street
City of mailing address::	Alameda
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94501

Inventor 3

Primary Citizenship Country::	United States
Given Name::	Kenneth
Middle Name::	Mark
Family Name::	Maxham
City of Residence::	Redwood City
State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	241 Harrison Avenue
City of mailing address::	Redwood City
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94062

Inventor 4

Primary Citizenship Country::	United States
Given Name::	John
Middle Name::	
Family Name::	Eid
City of Residence::	San Francisco
State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	52 Sheridan Street, #2
City of mailing address::	San Francisco
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94103

Inventor 5

Primary Citizenship Country::	United States
Given Name::	Cheryl
Middle Name::	
Family Name::	Heiner
City of Residence::	La Honda
State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	112 Sueno Camino
City of mailing address::	La Honda
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94020

Inventor 6

Primary Citizenship Country::	United States
Given Name::	Kevin
Middle Name::	
Family Name::	Travers
City of Residence::	Menlo Park

State or Province of Residence::	California
Country of Residence::	United States
Street of mailing address::	1008 Sevier Avenue
City of mailing address::	Menlo Park
State or Province of mailing address::	California
Postal or Zip Code of mailing address::	94025

DOMESTIC PRIORITY INFORMATION

Application::	Continuity Type::	Parent Application::	Parent Filing Date
This Application	Continuation of	14/708,603	05/11/15
14/708,603	Continuation of	14/091,961	11/27/13
14/091,961	Continuation of	12/982,029	12/30/10
12/982,029	An application claiming the benefit under 35 USC 119(e)	61/099,696	09/24/08
12/982,029	An application claiming the benefit under 35 USC 119(e)	61/139,402	12/19/08
12/982,029	Continuation-in-part of	12/413,226	03/27/09

FOREIGN PRIORITY INFORMATION

Country::	Application number::	Filing Date::	Priority Claimed::

APPLICANT INFORMATION

Applicant is assignee?::	Yes
Applicant Name::	Pacific Biosciences of California, Inc.
Street of Mailing Address::	~~1380 Willow Road~~1305 O'Brien Drive
City of mailing address::	Menlo Park
State or Province of mailing address::	California
Postal or Zip code of mailing address::	94025

CORRESPONDENCE INFORMATION

Correspondence Address::	Pacific Biosciences of California, Inc. ~~1380 Willow Road~~1305 O'Brien Drive Menlo Park, CA 94025

Phone number::	(650) 521-8127
Fax number::	(650) 323-9420
E-Mail address::	dscherer@pacb.com
Customer Number::	57770

REPRESENTATIVE INFORMATION

Representative Name::	David C. Scherer, Ph.D.
Registration Number::	56,993

Signature:: /David C. Scherer, Ph.D./

David C. Scherer, Ph.D.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 15/383,965 filed on December 19, 2016.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in 37 CFR 1.56.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR(s)

Inventor: Cheryl Heiner Date (Optional):

Signature:

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form. Use an additional PTO/AIA/01 form for each additional inventor.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 15/383,965

filed on December 19, 2016.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in 37 CFR 1.56.

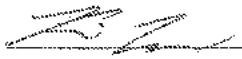
I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2036 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR(s)

Inventor: Kevin Travers Date (Optional): 4/27/2017

Signature: 	
Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form. Use an additional PTO/AIA/01 form for each additional inventor.	

Electronic Patent Application Fee Transmittal

Application Number:	15383965
Filing Date:	19-Dec-2016
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Filer:	David Christopher Scherer/Jacqueline Lim
Attorney Docket Number:	01-007706US

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
PROCESSING FEE, EXCEPT PROV. APPLS.	1830	1	140	140
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
CORRECTION OF INVENTORSHIP ON MERITS	1819	1	600	600
		Total in USD ($)		**740**

Electronic Acknowledgement Receipt	
EFS ID:	29064823
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	28-APR-2017
Filing Date:	19-DEC-2016
Time Stamp:	15:27:02
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$740
RAM confirmation Number	050117INTEFSW00002188504427
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	01007706_2017-04-28_Trans.pdf	47242 35d93a703e26ef4ee6f6763099599e1729b9146585c	no	1
Warnings:					
Information:					
2	Request under Rule 48 correcting inventorship	01007706_2017-04-28_ReqChgInventorship.pdf	14351 9095f4af90f693f142891 4a231f48a58e332De55d	no	1
Warnings:					
Information:					
3	Application Data Sheet	01007706_2017-04-28_SupplADS.pdf	19492 fac05dcfa2f828acd7b347f09847fbaa250d199e4	no	4
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
4	Oath or Declaration filed	01007706_2017-04-28_Declaration-CHeinerAndKTravers.pdf	1583339 428f6411df31b82f9156b801ddb3b618a7e30b51	no	3
Warnings:					
The page size in the PDF is too large. The pages should be 8.5 x 11 or A4. If this PDF is submitted, the pages will be resized upon entry into the Image File Wrapper and may affect subsequent processing					
Information:					
5	Fee Worksheet (SB06)	fee-info.pdf	32476 e1d346306e35fdf81173d9f8b34c123b2935e0e16	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	1696900		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/21 (10-08)
Approved for use through 11/30/2008. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	15/383,965
Filing Date	December 19, 2016
First Named Inventor	Stephen Turner
Art Unit	1634
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

ENCLOSURES *(Check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation
- ☐ Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- ☐ Proprietary Information
- ☐ Status Letter
- ☑ Other Enclosure(s) (please Identify below):
  - -Request to Change Inventorship;
  - -Declaration Signed by Inventors Being Added;
  - -Supplemental Application Data Sheet

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Pacific Biosciences of California, Inc.		
Signature	/David C. Scherer, Ph.D./		
Printed name	David C. Scherer, Ph.D.		
Date	April 28, 2017	Reg. No.	56,993

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below: via EFS-Web

Signature	/Jacqueline L. Lim/		
Typed or printed name	Jacqueline L. Lim	Date	April 28, 2017

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/383,965	12/19/2016	Stephen Turner	01-007706US

CONFIRMATION NO. 8144

57770
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

37 CFR 1.48 ACKNOWLEDGEMENT LETTER

OC000000091058038

Date Mailed: 05/03/2017

NOTICE OF ACCEPTANCE OF REQUEST UNDER 37 CFR 1.48(a)

This is in response to the applicant's request under 37 CFR 1.48(a) submitted on 04/28/2017.

The request under 37 CFR 1.48(a) to correct the inventorship, to correct or update the name of an inventor, or to correct the order of names of joint inventors is accepted.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/lqchau/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/383,965	12/19/2016	1637	2400	01-007706US	18	1

CONFIRMATION NO. 8144

UPDATED FILING RECEIPT

57770
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

OC000000091038037

Date Mailed: 05/03/2017

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Stephen Turner, Seattle, WA;
Jon Sorenson, Alameda, CA;
Kenneth Mark Maxham, Redwood City, CA;
John Eid, San Francisco, CA;
Cheryl Heiner, La Honda, CA;
Kevin Travers, Menlo Park, CA;

Applicant(s)

Pacific Biosciences of California, Inc., Menlo Park, CA;

Assignment For Published Patent Application

Pacific Biosciences of California, Inc.

Power of Attorney: The patent practitioners associated with Customer Number 57770

Domestic Priority data as claimed by applicant

This application is a CON of 14/708,603 05/11/2015 PAT 9556480
which is a CON of 14/091,961 11/27/2013 PAT 9057102
which is a CON of 12/982,029 12/30/2010 PAT 8628940
which claims benefit of 61/099,696 09/24/2008
and claims benefit of 61/139,402 12/19/2008
and is a CIP of 12/413,226 03/27/2009 PAT 8143030

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: No

Permission to Access Search Results: No

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

Projected Publication Date: 05/04/2017

Non-Publication Request: No

Early Publication Request: No

Title

INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Preliminary Class

435

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific

countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop

technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
15/383,965	12/19/2016	Stephen Turner	01-007706US

CONFIRMATION NO. 8144

57770
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

PUBLICATION NOTICE

OC000000091110502

Title:INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Publication No.US-2017-0121764-A1
Publication Date:05/04/2017

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Office of Public Records. The Office of Public Records can be reached by telephone at (571) 272-3150 or (800) 972-6382, by facsimile at (571) 273-3250, by mail addressed to the United States Patent and Trademark Office, Office of Public Records, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently http://pair.uspto.gov/. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Management, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	6404907	B1	2002-06-11	Gilchrist et al.	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20050266456	A1	2005-12-01	Williams et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2007025124	WO	A1	2007-03-01	Lapidus		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	1	Expert Report of Floyd Romesberg, Ph.D. Regarding U.S. Patent Nos. 9,404,146 and 9,542,527 as they relate to the Complaint filed by Pacific Biosciences Under Section 337 of the Tariff Act of 1930 with the United States International Trade Commission on November 2, 2016	☐
	2	MINER et al., "Molecular Barcodes Detect Redundancy and Contamination in Hairpin-Bisulfite PCR," Nucl. Acids Res. (2004) 32(17):e135.	☐
	3	Pacific Biosciences Presentation for Cold Spring Harbor Personal Genomics Meeting on October 12, 2008	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David C. Scherer, Ph.D./	Date (YYYY-MM-DD)	2017-05-09
Name/Print	David C. Scherer, Ph.D.	Registration Number	56,993

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 March 2007 (01.03.2007)

PCT

(10) International Publication Number
WO 2007/025124 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2006/033214

(22) International Filing Date: 25 August 2006 (25.08.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/213,621 26 August 2005 (26.08.2005) US

(71) Applicant *(for all designated States except US)*: **HELICOS BIOSCIENCES CORPORATION** [US/US]; One Kendall Square, Building 200, Cambridge, MA 02139 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **LAPIDUS, Stanley,** N. [US/US]; 12 Old Evergreen Road, Bedford, NH 03110 (US).

(74) Agents: **KAVANAUGH, Theresa, C.** et al.; Goodwin Procter LLP, Exchange Place, Boston, MA 02109 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SINGLE MOLECULE SEQUENCING OF CAPTURED NUCLEIC ACIDS

(57) Abstract: The invention provides methods for detecting, enumerating or identifying target nucleic acid molecules using immobilized capture probes and single molecule sequencing techniques.

WO 2007/025124 A1

SINGLE MOLECULE SEQUENCING OF CAPTURED NUCLEIC ACIDS

Cross-Reference to Related Application

[0001] This application claims priority to U.S.S.N 11/213,621 filed August 26, 2005, hereby incorporated by reference in its entirety.

Technical Field of the Invention

[0002] This invention relates to methods and compositions for detecting, enumerating, and identifying nucleic acids using capture probes to capture and purify target nucleic acid in combination with single molecule sequencing of the captured or purified nucleic acid molecules.

Background of the Invention

[0003] Knowledge of the human genome has given rise to inquiry into individual differences, as well as differences within an individual, as the basis for differences in biological function and dysfunction. Differences as small as single nucleotide polymorphisms (SNPs) or combinations of SNPs can lead to phenotypic differences, and detection of combinations of SNPs can predict the likelihood that an individual will get a specific disease or how an individual will respond to treatment.

[0004] For example, most cancers develop from a series of genomic changes, some subtle and some significant, that occur in a small subpopulation of cells. Knowledge of the sequence variations that lead to cancer will lead to an understanding of the etiology of the disease, as well as ways to treat and/or prevent it. An essential first step in understanding genomic complexity is the ability to perform high-resolution sequencing. Therefore, a true understanding of the complexities in either normal or abnormal function will require specific sequence information from large numbers of target nucleic acid molecules.

[0005] Bulk sequencing techniques are often not useful for the identification of subtle or rare nucleotide changes due to the many cloning, amplification and electrophoresis steps that complicate the process of gaining useful information regarding individual nucleotides. The ability to sequence and gain information from single molecules obtained from an individual patient is the next milestone for genomic sequencing. However, effective diagnosis and management of important diseases through single molecule sequencing is impeded by lack of cost-effective tools and methods for screening individual molecules.

[0006] There have been many proposals to develop new sequencing technologies based on single-molecule measurements, generally either by observing the interaction of particular proteins with DNA or by using ultra high resolution scanned probe microscopy. See, *e.g.*, Rigler, *et al.*, Biotech., 86(3):161 (2001); Goodwin, P.M., *et al.*, Nucleosides & Nucleotides, 16(5-6):543-550 (1997); Howorka, S., *et al.*, Nature Biotech., 19(7):636-639 (2001); Meller, A., *et al.*, Proc. Natl. Acad., 97(3):1079-1084 (2000); Driscoll, R.J., *et al.*, Nature, 346(6281):294-296 (1990). A recent technique employs optical detection in a sequencing-by-synthesis reaction at the single molecule level. Braslavsky, et al., PNAS, 100: 3960-3964 (2003). The present invention provides improvements in sequencing, especially single molecule sequencing.

Summary of the Invention

[0007] The invention provides methods for sequencing nucleic acids. According to the invention, sequencing is facilitated by capture of target nucleic acids to be sequenced prior to sequencing. In a preferred embodiment, target nucleic acids are sequenced at the single molecule level, resulting in sample-specific, high-throughput sequence information. Thus, according to one aspect of the invention, sequence-specific capture probes are used to isolate target nucleic acids of interest. The target population may be composed of same-sequence nucleic acids or may be a population of mixed-sequence nucleic acids. The isolated target nucleic acids are then sequenced. In one embodiment, sequencing is performed directly on captured nucleic acids, using the capture probes as primers. Alternatively, target/probe duplex may be melted and the target nucleic acids may then be hybridized to primers for template-dependent sequencing.

[0008] In a preferred embodiment, single molecule sequencing is conducted in order to provide high-resolution, high-throughput sequence information. Template-dependent single-molecule sequencing-by-synthesis is conducted using optically-labeled nucleotides for addition to the primer or probe. Either the target (template) or the primer/probe, or both are attached to a surface that is designed to enhance optical signal detection. A particularly-preferred surface is an epoxide surface coated onto glass or fused silica. Nucleic acids are easily attached to epoxide or epoxide derivatives. A preferred method is direct amine attachment. Nucleic acids can be purchased with a 5' or 3' amine, or terminal transferase can be used to introduce a terminal amine for attachment to the epoxide ring. Alternatively, epoxide surfaces can be derivatized for nucleic acid attachment. For example, the surface can incorporate streptavidin, which binds to

biotinylated nucleic acids. Either the target, primer/probe, or both can be biotinylated using known methods. Alternative surfaces include polyelectrolyte multilayers as described in Braslavasky, et al., PNAS 100: 3960-64 (2003), incorporated by reference herein. Essentially, any surface that has reduced native fluorescence and is amenable to attachment of oligonucleotides is useful for the invention.

[0009] Single molecule sequence is advantageously performed using optically-detectable labels. Especially preferred are fluorescent labels, including fluorescein, rhodamine, derivatized rhodamine dyes, such as TAMRA, phosphor, polymethadine dye, fluorescent phosphoramidite, texas red, green fluorescent protein, acridine, cyanine, cyanine 5 dye, cyanine 3 dye, 5-(2'-aminoethyl)-aminonaphthalene-1-sulfonic acid (EDANS), BODIPY, 120 ALEXA, or a derivative or modification of any of the foregoing.

[0010] The capture step prior to sequencing may be any suitable hybrid capture method. For example, capture can occur in solution, on beads (polystyrene beads), in a column (such as a chromatography column), in a gel (such as a polyacrylamide gel), or directly on the surface to be used for sequencing. An array of support-bound capture oligos can be used to hybridize specifically to a target sequence. Additionally, chromatography-based capture techniques are useful. For example, ion exchange chromatography, HPLC, gas chromatography, and gel-based chromatography all are useful. In one embodiment, gel-based capture is used in order to achieve sequence-specific capture. Using this method, multiple different sequences are captured simultaneously using immobilized probes in the gel. The target sequences are isolated by removing portions of the gel containing them and eluting target from the gel portions for sequencing.

[0011] In an alternative embodiment, the target nucleic acid molecule either includes, or is modified to include, an adaptor sequence (such as a polyadenylation region) that is complementary to a portion of a capture probe in order to aid in the capture of the target. A preferred embodiment comprises an immobilized capture probe having a sequence that hybridizes (*e.g.*, is complementary to) with the adaptor sequence. Methods of the invention are conducted by contacting capture probes with a sample comprising target nucleic acid molecules under conditions suitable for specific hybridization between the target nucleic acid molecule and immobilized capture probe, thereby forming target/capture probe duplex. A wash step removes debris and unhybridized nucleic acid in the sample. In one embodiment, target nucleic acid is sequenced using the capture probe as a primer for template-dependent sequencing-by-synthesis.

In another embodiment, the target/capture probe duplex is melted to release target nucleic acid. The resulting purified target population is sequenced as described below.

[0012] If target nucleic acid is melted off the capture probe, the targets are either attached to a surface for sequencing or hybridized to a primer that has been attached to the surface. Surface attachment of oligonucleotides for sequencing can be direct or indirect. For example, nucleic acids are attached to an epoxide surface via a direct amine linkage as described below. Alternatively, the surface is prepared with a binding partner, the opposite of which is attached to the nucleic acid. For example, the surface can be streptavidinated and biotinylated nucleic acids can be used to form an attachment at the surface. Other binding pairs (*e.g.*, antibody/antigen, such as digoxigenen/anti-digoxigenen and dinitorphenol/anti-dinitrophenol) can also be used.

[0013] In a preferred embodiment, the invention contemplates optical sequencing. Preferably, template-dependent sequencing-by-synthesis is conducted using optically-detectable labels. Optimal labels include fluorescent labels as described in detail below. The surface is prepared to minimize background for optical detection of incorporated nucleotides. Primer/template duplex attached to the surface is exposed to labeled nucleoside triphosphates in the presence of a suitable nucleotide polymerizing enzyme, under conditions suitable for the enzyme to add at least one nucleotide to the primer in template-dependent manner, and the added nucleotide is detected, thereby identifying at least one nucleotide in at least one attached target nucleic acid molecule.

[0014] As will be appreciated by one skilled in the art, individual features of the invention may be used separately or in any combination. A detailed description of embodiments of the invention is provided below. Other embodiments of the invention are apparent upon review of the detailed description that follows.

Detailed Description of the Invention

[0015] The present invention provides methods for sequencing nucleic acids. In a preferred embodiment, the invention comprises the use of single molecule sequencing of hybrid-captured nucleic acids. Methods of the invention are highly parallel and are amenable to multiplexing. As a result, biological samples can be rapidly analyzed over a broad dynamic range.

[0016] Methods of the invention comprise hybrid-capturing target nucleic acid and sequencing the captured targets. Captured target nucleic acids can be sequenced directly using

the capture probe as a primer for template-dependent synthesis. Thus, capture probes are bound to a surface and exposed to sample nucleic acids. Complementary binding of target nucleic acid and probes occurs and the remaining nucleic acid and other sample contents are washed from the surface. Sequencing then takes place essentially as described below. In one alternative, target nucleic acids are isolated by contacting a sample containing target nucleic acid molecules with immobilized capture probes capable of hybridizing to target nucleic acid. Target nucleic acid The surface is washed to remove non-target nucleic acids and other debris, and the targets are then melted off the target/capture probe duplex. After an optional wash, the targets are bound to a surface for sequencing essentially as described below. Target nucleic acid molecules preferably are analyzed using single molecule sequencing techniques described below. In a preferred embodiment, target nucleic acid is modified by addition of a polynucleotide sequence that specifically hybridizes to a complementary portion of a primer for nucleic acid synthesis. For example, targets can be polyadenylated and captured using poly-dT probes/primers for sequencing.

Target Nucleic Acid Molecules

[0017] Target nucleic acids include deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA). Target nucleic acids can be isolated or can be captured in situ. In a preferred alternative, target nucleic acid is isolated from a biological sample containing a variety of other components, such as proteins, lipids and non-target nucleic acids. Target nucleic acid molecules can be obtained from any cellular material, obtained from an animal, plant, bacterium, virus, fungus, or any other cellular organism. Target nucleic acids may be obtained directly from an organism or from a biological sample obtained from an organism, *e.g.*, from blood, urine, cerebrospinal fluid, seminal fluid, saliva, sputum, stool and tissue. Any tissue or body fluid specimen may be used as a source for nucleic acid for use in the invention. Nucleic acid molecules may also be isolated from cultured cells, such as a primary cell culture or a cell line. The cells from which target nucleic acids are obtained can be infected with a virus or other intracellular pathogen.

[0018] A sample can also be total RNA extracted from a biological specimen, a cDNA library, or genomic DNA. Nucleic acid typically is fragmented to produce suitable fragments for capture and/or purification and analysis. In one embodiment, nucleic acid from a biological sample is fragmented by sonication. Test samples can be obtained as described in U.S. Patent

Application 2002/0190663 A1, published October 9, 2003, the teachings of which are incorporated herein in their entirety. Generally, nucleic acid can be extracted from a biological sample by a variety of techniques such as those described by Maniatis, *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp. 280-281 (1982). Generally, target nucleic acid molecules can be from about 5 bases to about 20 kb. Nucleic acid molecules may be single-stranded, double-stranded, or double-stranded with single-stranded regions (for example, stem- and loop-structures).

[0019] A biological sample as described herein may be homogenized or fractionated in the presence of a detergent or surfactant. The concentration of the detergent in the buffer may be about 0.05% to about 10.0%. The concentration of the detergent can be up to an amount where the detergent remains soluble in the solution. In a preferred embodiment, the concentration of the detergent is between 0.1% to about 2%. The detergent, particularly a mild one that is nondenaturing, can act to solubilize the sample. Detergents may be ionic or nonionic. Examples of nonionic detergents include triton, such as the Triton® X series (Triton® X-100 t-Oct-C_6H_4-$(OCH_2-CH_2)_xOH$, x=9-10, Triton® X-100R, Triton® X-114 x=7-8), octyl glucoside, polyoxyethylene(9)dodecyl ether, digitonin, IGEPAL® CA630 octylphenyl polyethylene glycol, n-octyl-beta-D-glucopyranoside (betaOG), n-dodecyl-beta, Tween® 20 polyethylene glycol sorbitan monolaurate, Tween® 80 polyethylene glycol sorbitan monooleate, polidocanol, n-dodecyl beta-D-maltoside (DDM), NP-40 nonylphenyl polyethylene glycol, C12E8 (octaethylene glycol n-dodecyl monoether), hexaethyleneglycol mono-n-tetradecyl ether (C14EO6), octyl-beta-thioglucopyranoside (octyl thioglucoside, OTG), Emulgen, and polyoxyethylene 10 lauryl ether (C12E10). Examples of ionic detergents (anionic or cationic) include deoxycholate, sodium dodecyl sulfate (SDS), N-lauroylsarcosine, and cetyltrimethylammoniumbromide (CTAB). A zwitterionic reagent may also be used in the purification schemes of the present invention, such as Chaps, zwitterion 3-14, and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate. It is contemplated also that urea may be added with or without another detergent or surfactant.

[0020] Lysis or homogenization solutions may further contain other agents, such as reducing agents. Examples of such reducing agents include dithiothreitol (DTT), β-mercaptoethanol, DTE, GSH, cysteine, cysteamine, tricarboxyethyl phosphine (TCEP), or salts of sulfurous acid.

[0021] Preferred buffer concentration is from about 5 mM to about 500 mM in solution or in solution with the sample. The buffer concentration in the lysing solution can be between about 10 mM and 300 mM.

[0022] Subsequent to exposure to a homogenization solution, samples may be further homogenized by mechanical means. Mechanical blenders, rotor-stator homogenizers, or shear-type homogenizers may be employed. Alternatively, the tissue can be homogenized in the lysis solution, and the tissue remains separated by settling, centrifugation, or filtration. These remains could then be treated with homogenization solution and extraction conditions as described above.

[0023] Target nucleic acid molecules can include or be modified to include an adaptor sequence that is complementary to the immobilized capture probe. The adaptor sequence and immobilized capture probe are chosen such that the target nucleic acid molecule comprising the adaptor sequence can hybridize to the immobilized capture probe. Typically, the adaptor sequence is a homopolymer, such as oligo(dA), and the corresponding immobilized capture probe includes an oligo(dT) sequence. The adaptor sequence may be endogenously contained within the target nucleic acid molecule, for example, the adaptor sequence can be a sequence of interest in the target nucleic acid molecule. Alternatively, the adaptor sequence can comprise vector sequence.

[0024] In one embodiment, a universal probe is used as the immobilized capture probe and the target nucleic acid molecules are modified with an adaptor comprising a sequence complementary to the universal probe, thereby allowing the modified nucleic acid molecules to hybridize to the immobilized probe.

[0025] The adaptor sequence and complementary sequence within the immobilized capture probe are of a length suitable for hybridizing the target nucleic acid molecule to the immobilized capture probe to thereby capture target nucleic acid molecules to form target probe/duplexes. The target probe/duplexes are also sufficiently stable to permit optional washing of the duplexes to remove any remaining components of the biological sample, including non-target nucleic acids. The sequence of the adaptor and the complementary sequence of the immobilized capture probe can be about 10 to about 100, and preferably 50, nucleotides in length. The adaptor sequence and complementary immobilized capture sequence can be of the same length or of different lengths. It is routine in the art to adjust probe length and/or oligonucleotide length to optimize hybridization.

[0026] The adaptor sequence can be attached to the nucleic acid molecules with an enzyme. The enzyme can be a ligase or a polymerase. The ligase can be an enzyme capable of ligating an oligonucleotide (RNA or DNA) to the nucleic acid molecules. Suitable ligases include, for example, T4 DNA ligase and T4 RNA ligase (such ligases are available commercially, from New England BioLabs (on the World Wide Web at NEB.com). In a preferred embodiment, the target nucleic acid molecules are dephoyshorylated before ligating the adaptors. Methods for using ligases are well known in the art.

[0027] The polymerase can be any enzyme capable of adding nucleotides to the target nucleic acid molecules. The polymerase can be, for example, poly(A) polymerase, including yeast poly(A) polymerase, commercially available from USB (on the World Wide Web at USBweb.com), terminal deoxyribonucleotidyl transferase (TdT), and the like. The polymerases can be used according to the manufacturer's instructions.

Purifying Target Nucleic Acid Molecules

[0028] As described herein, target nucleic acid molecules can be purified by contacting a sample containing the target molecules with one or more classes of immobilized capture probes. Where more than one target molecule is to be purified, the different target molecules can have the same or different adaptor sequences. A given class of immobilized capture probe is selected to hybridize to target nucleic acid molecules containing the corresponding adaptor sequence, under conditions suitable for the target nucleic acid molecule to hybridize to the immobilized capture probe, thereby forming target/probe duplexes. A variety of capture probes can be used to purify the target nucleic acid molecules. The capture probes typically comprise a nucleic acid with a nucleotide sequence with substantial complementarity to a region of the target nucleic acid molecule, so that the target nucleic acid molecule can hybridize to the capture probe. Complementarity between target nucleic acid molecules and the capture probes need only be sufficient to specifically bind the target nucleic acid molecule, and thus, to effectuate purification of the target nucleic acid molecule from a test sample. Probes suitable for use in the present invention include those formed from nucleic acids, such as RNA and/or DNA, nucleic acid analogues, modified nucleic acids, and chimeric probes of a mixed class comprising a nucleic acid with another organic component such as peptide nucleic acids. Capture probes can be single stranded or double stranded. Preferably, the length of the capture probe is at least 5 nucleotides, more preferably between about 5 and about 100 nucleotides, but the length can be

up to several thousand nucleotides. Additional probes are described in U.S. Patent application 2002/0119480 A1, published August 29, 2002, the teachings of which are incorporated herein in their entirety.

[0029] Capture probes can be coupled to agarose, dextrans, cellulose, beads, microparticles, and starch polymers using cyanogen bromide or cyanuric chloride activation. Particularly useful beads and microparticles are described in U.S. Patent Application 2003/0190663 A1, published October 9, 2003, the teachings of which are incorporated herein in their entirety. Polymers containing carboxyl groups can be coupled to synthetic capture probes having primary amine groups using carbodiimide coupling. Polymers containing primary amines can be coupled to amine-containing probes with glutaraldehyde or cyanuric chloride. Polymers can also be modified with thiol-reactive groups that can be coupled to thiol-containing probes. Many other suitable methods can be found in the literature (Wong, Chemistry of Protein Conjugation and Cross-Linking, CRC Press, Boca Raton, Florida 1993). Methods for covalently attaching the capture probes to polymerizable chemical groups are also described in U.S. 2002/0119480 A1, the teachings of which are incorporated herein. In addition, as described in U.S. 2002/0119480 A1, nucleic acids can be attached to particles which themselves can be incorporated into electrophoretic media.

[0030] To purify target nucleic acid molecules, the test sample can be contacted, for example, with an electrophoretic medium that comprises at least one immobilized capture probe. An electric field is applied across the electrophoretic medium so that negatively charged molecules can migrate through the medium. The non-target components of the test sample pass into the electrophoresis buffer, while the target nucleic acid molecules are captured by hybridizing to the capture probes within the electrophoretic medium (forming target /probe duplexes). The buffer used for the electrophoresis step (containing the non-target components of the test sample) can be replaced with fresh buffer and current applied across the electrophoretic medium to denature the complex, thereby releasing the target nucleic acid molecule from the capture probe. The released target nucleic acid molecule can be eluted by applying a reversed electric field.

[0031] The method of the present invention is suitable for multiplexing. As described above, a sample can include more than one target nucleic acid molecule. If adapters are used, the different target nucleic acid molecules can have the same or different adaptor sequences. In a multiplexing reaction, where the target nucleic acid molecules contain different adaptor

sequences, the test sample is contacted with two or more classes of immobilized capture probes as described above. Target nucleic acid molecules are captured by their respective immobilized capture probes. The captured target nucleic acid molecules can be released as described above. Other methods for capturing and purifying target nucleic acid molecules are described in U.S. 2002/0119480 A1, the teachings of which are incorporated herein.

Attaching Target Nucleic Acid Molecules To A Surface

[0032] In a preferred embodiment, target nucleic acid molecules are attached to a surface and subjected to analysis by single molecules sequencing. Target nucleic acid molecules are attached to the surface such that they are individually optically resolvable. Substrates for use in the invention can be two- or three-dimensional and can comprise a planar surface (*e.g.*, a glass slide) or can be shaped. A substrate can include glass (*e.g.*, controlled pore glass (CPG)), quartz, plastic (such as polystyrene (low cross-linked and high cross-linked polystyrene), polycarbonate, polypropylene and poly(methymethacrylate)), acrylic copolymer, polyamide, silicon, metal (*e.g.*, alkanethiolate-derivatized gold), cellulose, nylon, latex, dextran, gel matrix (*e.g.*, silica gel), polyacrolein, or composites.

[0033] Suitable three-dimensional substrates include, for example, spheres, microparticles, beads, membranes, slides, plates, micromachined chips, tubes (*e.g.*, capillary tubes), microwells, microfluidic devices, filters, or any other structure suitable for anchoring a nucleic acid. Substrates can include planar arrays or matrices capable of having regions that include populations of target nucleic acids or primers. Examples include nucleoside-derivatized CPG and polystyrene slides; derivatized magnetic slides; polystyrene grafted with polyethylene glycol, and the like.

[0034] In one embodiment, a substrate is coated to allow optimum optical processing and nucleic acid attachment. Surfaces for use in the invention are treated to reduce background. Exemplary coatings include epoxides, and derivatized epoxides (*e.g.*, with a binding molecule, such as streptavidin). The surface can also be treated to improve the positioning of attached nucleic acids (*e.g.*, target nucleic acid molecules, primers, or target molecule/primer complexes) for analysis. As such, a surface according to the invention can be treated with one or more charge layers (*e.g.*, a negative charge) to repel a charged molecule (*e.g.*, a negatively charged labeled nucleotide). For example, a substrate according to the invention can be treated with polyallylamine followed by polyacrylic acid to form a polyelectrolyte multilayer. The carboxyl

groups of the polyacrylic acid layer are negatively charged and thus repel negatively charged labeled nucleotides, improving the positioning of the label for detection. Coatings or films applied to the substrate should be able to withstand subsequent treatment steps (*e.g.*, photoexposure, boiling, baking, soaking in warm detergent-containing liquids, and the like) without substantial degradation or disassociation from the substrate.

[0035] Examples of substrate coatings include, vapor phase coatings of 3-aminopropyltrimethoxysilane, as applied to glass slide products, for example, from Molecular Dynamics, Sunnyvale, California. In addition, generally, hydrophobic substrate coatings and films aid in the uniform distribution of hydrophilic molecules on the substrate surfaces. Importantly, in those embodiments of the invention that employ substrate coatings or films, the coatings or films that are substantially non-interfering with primer extension and detection steps are preferred. Additionally, it is preferable that any coatings or films applied to the substrates either increase target molecule binding to the substrate or, at least, do not substantially impair target binding.

[0036] Various methods can be used to anchor or immobilize the target nucleic acid molecule to the surface of the substrate. The immobilization can be achieved through direct or indirect bonding to the surface. The bonding can be by covalent linkage. See, Joos *et al.*, Analytical Biochemistry 247:96-101, 1997; Oroskar *et al.*, Clin. Chem. 42:1547-1555, 1996; and Khandjian, Mole. Bio. Rep. 11:107-115, 1986. A preferred attachment is direct amine bonding of a terminal nucleotide to an epoxide integrated on the surface. The bonding also can be through non-covalent linkage. For example, biotin-streptavidin (Taylor *et al.*, J. Phys. D. Appl. Phys. 24:1443, 1991) and digoxigenin with anti-digoxigenin (Smith *et al.*, Science 253:1122, 1992) are common tools for anchoring nucleic acids to surfaces and parallels. Alternatively, the attachment can be achieved by anchoring a hydrophobic chain into a lipid monolayer or bilayer. Other methods for known in the art for attaching nucleic acid molecules to supports also can be used.

[0037] In a preferred embodiment, surfaces for oligonucleotide attachment are coated with an epoxide. An epoxide may be deposited by many methods known in the art. An epoxy silane surface is preferred. Different molecules or combinations of molecules may serve to link the epoxide to a surface. Ideally, a surface will be coated with an even distribution of epoxides prior to introduction of target nucleic acid molecules. Target nucleic acid molecules can be

directly or indirectly linked to an epoxide. In a direct attachment embodiment, the epoxide is introduced to a nucleic acid bearing an amine group. In a preferred embodiment, terminal transferase is used to add an amine-terminated nucleotide to a nucleic acid to be attached to the surface.

[0038] When biotin-streptavidin linkage is used to anchor the nucleic acids, the nucleic acids can be biotinylated, while one surface of the substrates can be coated with streptavidin.

[0039] Surface density of the nucleic acid molecules can be controlled by adjusting the concentration of the streptavidin applied to the surface. Reagents for biotinylating a surface can be obtained, for example, from Vector Laboratories. Alternatively, biotinylation can be performed with BLCPA: EZ-Link Biotin LC-PEO-Amine (Pierce, Cat. 21347), or any other known or convenient method. In some embodiments, labeled streptavidin of very low concentration (*e.g.*, in the μM, nM or pM range) is used to coat the substrate surface prior to anchoring. This can facilitate immobilization of the nucleic acid with single molecule resolution.

[0040] Other examples of linkers include antigen/antibody, digoxigenin/anti-digoxigenin, dinitrophenol, fluorescein, and other haptens known in the art. Alternatively, the nucleic acid may contain other binding moieties that result in a conformational change of the epoxide ring and result in a direct attachment of the target nucleic acid molecules to the opened epoxide ring.

[0041] Alternatively, primers may be immobilized on the surface. A terminus of one or more primers may be modified to carry a linker moiety for tethering the primers to the surface or may be directly attached to the surface. Target nucleic acid molecules containing primer complementary sequence are then hybridized to the attached primer. Methods for attaching nucleic acid such as primer to the surface of a substrate are described in detail above.

[0042] In order to inhibit non-specific binding of molecules to a surface, the surface can be treated with a passivating (blocking) agent. Preferred blocking strategies include exposing the surface to a non-detectable molecule that adheres to the surface or changes the chemical properties of the surface such that non-specific binding is reduced. In methods in which optically-detectable labels are used, one way to block or passivate the surface is to expose the surface to unlabeled molecules of the same type as those that are labeled. The unlabeled molecules will out-compete labeled molecules for non-specific binding on the surface, thus reducing background due to non-specific label. Other strategies involve treating the surface with phosphate, Tris, a sulfate, or an amine that interacts with the surface to prevent non-specific

binding. Non-reactive proteins are also appropriate. In a preferred embodiment, a matrix of blocking reagents is provided on the surface in order to provide a highly washable, low non-specific background surface. In some embodiments, blocking reagents are chosen to provide electrostatic repulsion of highly anionic nucleoside triphosphates.

[0043] Blocking agents may be introduced or reintroduced at any time during the analysis. Also, in some embodiments, blocking agents may be used to pre-treat the surface prior to exposing the substrate to target nucleic acid molecules or primers. In addition, blocking agents, such as a detergent (*e.g.*, Tris) may be included in some or all wash steps in order to passivate the surface during incubation periods and/or washes.

[0044] Surface charge can be manipulated to achieve ideal conditions during both nucleic acid attachment and primer extension. For example, during the loading phase where the nucleic acid (target nucleic acid molecule or primer) is bound to the surface, the salt concentration of the solution may be increased in order to create a more positive surface charge on the substrate to facilitate reaction between the amine portion of the nucleic acid and the epoxide ring. Conversely, after the nucleic acid has been secured to the surface, the salt concentration of the solution can lowered in order to repel the nucleic acid from the surface of the substrate thereby sterically conforming the nucleic acid for annealing and sequence analysis.

[0045] In another embodiment, the substrate includes a layer of polyanions and nucleic acid molecules anchored on the layer of polyanions. Accordingly, nucleic acids are positioned to avoid being substantially parallel (*e.g.*, is hindered from lying down on the layer of polyanions.) In some embodiments, the surface of a substrate is pretreated to create a surface chemistry that facilitates nucleic acid attachment and subsequent annealing and sequence analysis. In some of these embodiments, the substrate surface is coated with a polyelectrolyte multilayer (PEM). In some cases, biotin can be applied to the PEM, followed by application of streptavidin. The substrate can then be used to attach biotinylated target nucleic acid molecules.

[0046] In some embodiments, multiple layers of alternating positive and negative charges are used. In the case of incompletely-charged surfaces, multiple-layer deposition tends to increase surface charge to a well-defined and stable level. For example, surfaces can be coated with a PEM for attachment of primers via light-directed spatial attachment. Alternatively, target nucleic acid molecules can be attached to a PEM-coated surface chemically. PEM formation has been described in Decher *et al.* (Thin Solid Films, 210:831-835, 1992), the teachings of which

are incorporated herein. PEM formation proceeds by the sequential addition of polycations and polyanions, which are polymers with many positive or negative charges, respectively. Upon addition of a polycation to a negatively-charged surface, the polycation deposits on the surface, forming a thin polymer layer and reversing the surface charge. Similarly, a polyanion deposited on a positively charged surface forms a thin layer of polymer and leaves a negatively charged surface. Alternating exposure to poly(+) and poly(-) generates a polyelectrolyte multilayer structure with a surface charge determined by the last polyelectrolyte added. This can produce a strongly-negatively-charged surface, repelling the negatively-charged nucleotides. Detailed procedures for coating a substrate with PEM for immobilizing nucleic acid are described in U.S. Patent Application No. 11/137,928, field May 25, 2005.

[0047] In another aspect of the invention, the substrate may be prepared by, for example, coating with a chemical that increases or decreases hydrophobicity or coating with a chemical that allows covalent linkage of the primers. Some chemical coatings may both alter the hydrophobicity and allow covalent linkage. Hydrophobicity on a solid substrate may readily be increased by silane treatment or other treatments known in the art. Linker molecules adhere to the surface and comprise a functional moiety that reacts with biomolecules. Many such linkers are readily available and known in the art. For example, substrates or supports are modified with photolabile-protected hydroxyl groups, alkoxy or aliphatic derivatized hydroxyl groups, or other chemicals. A preferred coating that both decreases hydrophobicity and provides linkers is poly(ethyleneimine).

[0048] Methods of the invention also optionally include a surface drying step. In some embodiments, the surface is exposed to a drying agent prior to, during and/or after a chemical reaction, such as a nucleotide incorporation step. Examples of preferred drying agents include, without limitation, phosphate buffer, an alcohol (such as, for example, EtOH), air and/or N_2.

Analyzing Attached Target Nucleic Acid Molecules

[0049] As described herein, attached target nucleic acid molecules are analyzed by single molecule sequencing. At least one nucleotide is identified in at least one attached target nucleic acid molecule. Target molecules are hybridized to a primer to form nucleic acid target molecule/primer duplex on a surface. As described above, either the target nucleic acid molecule or the primer, or both, is/are attached to the surface. Thereafter, template-dependent primer extension is conducted to identify at least one nucleotide of the hybridized nucleic acid

molecule using a nucleotide polymerizing enzyme and a nucleotide (*e.g.*, dATP, dTTP, dUTP, dCTP and/or a dGTP) or a nucleotide analog. Incorporation of a nucleotide or a nucleotide analog is detected at discrete locations on the surface. Nucleic acid target molecule/primer duplexes, as well as the incorporated nucleotides, are individually resolvable in single molecule embodiments. Alternatively, bulk signal from mixed nucleic acid populations or clonal populations of nucleic acids, are obtained.

[0050] Fast reagent application and removal is another advantage of the invention. For example, concentrations of nucleotides and/or other reaction reagents can be alternated at different time points. This is a particularly useful feature in an embodiment comprising introducing one or more single species of nucleotide individually. This could lead to increased incorporation rates and sensitivity. For example, when all four types of nucleotides are simultaneously present in the reaction to monitor dynamic incorporation of nucleotides, concentrations of the each of the respective nucleotides can be alternated between a first and a second range. This leads to both better visualization of the signal when low concentrations of nucleotides are present, and increased polymerization rate when higher concentrations of nucleotides are present.

[0051] The target nucleic acid can comprise or can be modified to comprise a sequence that is sufficiently complementary to a primer to hybridize to the primer to allow template dependent addition of nucleotides to the hybridized primer. The sequence complementary to the primer can be the same as the previously described adaptor sequence. Therefore, the immobilized capture probe and the primer can comprise the same sequence. The primer sequence can be about 10 to about 1000 nucleotides in length. The primer sequence and complementary target nucleic acid molecule sequence can be of the same length or of different lengths. Conditions for hybridizing primers to nucleic acid targets are well known. The annealing reaction is performed under conditions which are stringent enough to guarantee sequence specificity, yet sufficiently permissive to allow formation of stable hybrids at an acceptable rate. The temperature and length of time required for primer annealing depend upon several factors including the base composition, length and concentration of the primer, and the nature of the solvent used, *e.g.*, the concentration of cosolvents such as DMSO (dimethylsulfoxide), formamide, or glycerol, and counterions such as magnesium. Typically, hybridization (annealing) between primers and target nucleic acids is carried out at a temperature that is approximately 5 to 10° C below the melting temperature of the target-primer hybrid in the

annealing solvent. Typically, the annealing temperature is in the range of 55 to 75° C and the primer concentration is approximately 0.2 μM. Under such conditions, the annealing reaction is usually complete within a few seconds.

[0052] Methods according to the invention include conducting a primer extension reaction, such as exposing the nucleic acid/primer complexes to polymerase and one or more nucleotides under conditions sufficient to extend the primer by at least one base. Sequencing, as used herein can be performed such that one or more nucleotides are identified in one or more nucleic molecules. Methods according to the invention also include the step of compiling a sequence of the molecule (nucleic acid) based upon sequential incorporation of the extension bases into the primer.

[0053] In the analyzing step, the hybridized nucleic acid molecules can be sequenced using single molecule sequencing as described, for example, in U.S. Patent Application No. 11/137,928, filed May 25, 2005 and/or and described in U.S. Patent 6,780,591, the teachings of which are incorporated herein in their entirety. Polymerases useful in the invention include any nucleic acid polymerase capable of catalyzing a template-dependent addition of a nucleotide or nucleotide analog to a primer. Depending on the characteristics of the target nucleic acid, a DNA polymerase, an RNA polymerase, a reverse transcriptase, or a mutant or altered form of any of the foregoing can be used. According to one aspect of the invention, a thermophilic polymerase is used, such as ThermoSequenase®, 9°N™, Therminator™, Taq, Tne, Tma, Pfu, Tfl, Tth, Tli, Stoffel fragment, Vent™ and Deep Vent™ DNA polymerase.

[0054] The target nucleic acid molecule/primer complexes are contacted with dNTPs in the presence of the polymerase under conditions such that the polymerase catalyzes template-dependent addition of a dNTP to the 3′ terminus of the primer. The dNTP can be detectably labeled, as described herein, and the nucleotide is identified by detecting the presence of the incorporated labeled nucleotide. As described above, unincorporated labeled dNTPs can be removed from the surface prior to detecting the incorporated labeled dNTP. The process can be repeated one or more times, wherein the template/primer complex(s) are provided with additional dNTPs, in the presence of a polymerase, followed by removing the unincorporated labeled dNTPs and detecting the incorporated labeled dNTP. The sequence of the template is determined by compiling the detected (identified) dNTPs. In this manner, the entire sequence of one or more templates can be determined. In addition, by using single molecule sequencing techniques, determining the sequence for each nucleic acid molecule attached to the surface

provides the number of different or unique nucleic acid molecules in the sample. Furthermore, the number of copies of each nucleic acid sequences in a biological sample is also provided.

[0055] In order to allow for further extension and detection of subsequently added fluorophore-labeled nucleotides, the fluorophore of the incorporated nucleotide can be removed from the nucleotide incorporated into the primer. For example, the optical label (*e.g.*, fluorescent label) can be destroyed by photochemical destruction as described in U.S. Patent 6,780,591, the teachings of which are incorporated herein in their entirety. This cycle can be repeated a large number of times if sample losses are avoided. In one embodiment, such losses will be avoided by attaching the target nucleic acid molecules or primers to a surface of an array device, for example a microscope slide, and transferring the entire array device between a reaction vessel and the fluorescent reader.

[0056] The extension reactions are carried out in buffer solutions which contain the appropriate concentrations of salts, dNTP(s) and nucleotide polymerizing enzyme required for the enzyme mediated extension to proceed. For additional guidance regarding such conditions see, for example, Sambrook *et al.*, (1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, NY); and Ausubel *et al.* (1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, NY).

[0057] Nucleotides particularly useful in the invention comprise detectable labels. Labeled nucleotides include any nucleotide that has been modified to include a label that is directly or indirectly detectable. Preferred labels include optically-detectable labels, including fluorescent labels or fluorophores, such as fluorescein, rhodamine, cyanine, cyanine-5 dye, cyanine-3 dye, or a derivative or modification of any of the foregoing, and also include such labeling systems as hapten labeling. Accordingly, methods of the invention further provide for exposing the primer/target nucleic acid complex to a digoxigenin, a fluorescein, an alkaline phosphatase or a peroxidase.

[0058] The sequencing can be optimized to achieve rapid and complete addition of the correct nucleotide to primers in primer/template complexes, while limiting the misincorporation of incorrect nucleotides. For example, dNTP concentrations may be lowered to reduce misincorporation of incorrect nucleotides into the primer. K_m values for incorrect dNTPs can be as much as 1000-fold higher than for correct nucleotides, indicating that a reduction in dNTP concentrations can reduce the rate of misincorporation of nucleotides. Thus, in a preferred embodiment of the invention the concentration of dNTPs in the sequencing reactions are

approximately 5-20 μM.

[0059] In addition, relatively short reaction times can be used to reduce the probability of misincorporation. For an incorporation rate approaching the maximum rate of about 400 nucleotides per second, a reaction time of approximately 25 milliseconds will be sufficient to ensure extension of 99.99% of primer strands.

Detection

[0060] Any detection method may be used that is suitable for the type of label employed. Thus, exemplary detection methods include radioactive detection, optical absorbance detection, *e.g.*, UV-visible absorbance detection, optical emission detection, *e.g.*, fluorescence or chemiluminescence. For example, extended primers can be detected on a substrate by scanning all or portions of each substrate simultaneously or serially, depending on the scanning method used. For fluorescence labeling, selected regions on a substrate may be serially scanned one-by-one or row-by-row using a fluorescence microscope apparatus, such as described in Fodor (U.S. Patent No. 5,445,934) and Mathies *et al.* (U.S. Patent No. 5,091,652). Devices capable of sensing fluorescence from a single molecule include scanning tunneling microscope (siM) and the atomic force microscope (AFM). Hybridization patterns may also be scanned using a CCD camera (*e.g.*, Model TE/CCD512SF, Princeton Instruments, Trenton, N.J.) with suitable optics (Ploem, in Fluorescent and Luminescent Probes for Biological Activity Mason, T.G. Ed., Academic Press, Landon, pp. 1-11 (1993), such as described in Yershov *et al.*, Proc. Natl. Aca. Sci. 93:4913 (1996), or may be imaged by TV monitoring. For radioactive signals, a phosphorimager device can be used (Johnston *et al.*, Electrophoresis, 13:566, 1990; Drmanac *et al.*, Electrophoresis, 13:566, 1992; 1993). Other commercial suppliers of imaging instruments include General Scanning Inc., (Watertown, Mass. on the World Wide Web at genscan.com), Genix Technologies (Waterloo, Ontario, Canada; on the World Wide Web at confocal.com), and Applied Precision Inc. Such detection methods are particularly useful to achieve simultaneous scanning of multiple attached target nucleic acids.

[0061] The present invention provides for detection of molecules from a single nucleotide to a single target nucleic acid molecule. A number of methods are available for this purpose. Methods for visualizing single molecules within nucleic acids labeled with an intercalating dye include, for example, fluorescence microscopy. For example, the fluorescent spectrum and lifetime of a single molecule excited-state can be measured. Standard detectors

such as a photomultiplier tube or avalanche photodiode can be used. Full field imaging with a two-stage image intensified COD camera also can be used. Additionally, low noise cooled CCD can also be used to detect single fluorescent molecules.

[0062] The detection system for the signal may depend upon the labeling moiety used, which can be defined by the chemistry available. For optical signals, a combination of an optical fiber or charged couple device (CCD) can be used in the detection step. In those circumstances where the substrate is itself transparent to the radiation used, it is possible to have an incident light beam pass through the substrate with the detector located opposite the substrate from the target nucleic acid. For electromagnetic labeling moieties, various forms of spectroscopy systems can be used. Various physical orientations for the detection system are available and discussion of important design parameters is provided in the art.

[0063] A number of approaches can be used to detect incorporation of fluorescently-labeled nucleotides into a single nucleic acid molecule. Optical setups include near-field scanning microscopy, far-field confocal microscopy, wide-field epi-illumination, light scattering, dark field microscopy, photoconversion, single and/or multiphoton excitation, spectral wavelength discrimination, fluorophore identification, evanescent wave illumination, and total internal reflection fluorescence (TIRF) microscopy. In general, certain methods involve detection of laser-activated fluorescence using a microscope equipped with a camera. . Suitable photon detection systems include, but are not limited to, photodiodes and intensified CCD cameras. For example, an intensified charge couple device (ICCD) camera can be used. The use of an ICCD camera to image individual fluorescent dye molecules in a fluid near a surface provides numerous advantages. For example, with an ICCD optical setup, it is possible to acquire a sequence of images (movies) of fluorophores.

[0064] Some embodiments of the present invention use TIRF microscopy for two-dimensional imaging. TIRF microscopy uses totally internally reflected excitation light and is well known in the art. See, *e g.*, the World Wide Web at nikon-instruments.jp/eng/page/products/tirf.aspx. In certain embodiments, detection is carried out using evanescent wave illumination and total internal reflection fluorescence microscopy. An evanescent light field can be set up at the surface, for example, to image fluorescently-labeled nucleic acid molecules. When a laser beam is totally reflected at the interface between a liquid and a solid substrate (*e.g.*, a glass), the excitation light beam penetrates only a short distance into the liquid. The optical field does not end abruptly at the reflective interface, but its intensity falls

off exponentially with distance. This surface electromagnetic field, called the "evanescent wave", can selectively excite fluorescent molecules in the liquid near the interface. The thin evanescent optical field at the interface provides low background and facilitates the detection of single molecules with high signal-to-noise ratio at visible wavelengths.

[0065] The evanescent field also can image fluorescently-labeled nucleotides upon their incorporation into the attached target nucleic acid target molecule/primer complex in the presence of a polymerase. Total internal reflectance fluorescence microscopy is then used to visualize the attached target nucleic acid target molecule/primer complex and/or the incorporated nucleotides with single molecule resolution.

[0066] Measured signals can be analyzed manually or by appropriate computer methods to tabulate results. The substrates and reaction conditions can include appropriate controls for verifying the integrity of hybridization and extension conditions, and for providing standard curves for quantification, if desired. For example, a control nucleic acid can be added to the sample. The absence of the expected extension product is an indication that there is a defect with the sample or assay components requiring correction.

[0067] Fluorescence resonance energy transfer (FRET) can be used as a detection scheme. FRET in the context of sequencing is described generally in Braslavasky, *et al.*, Proc. Nat'l Acad. Sci., 100: 3960-3964 (2003), incorporated by reference herein. Essentially, in one embodiment, a donor fluorophore is attached to the primer, polymerase, or template. Nucleotides added for incorporation into the primer comprise an acceptor fluorophore that is activated by the donor when the two are in proximity.

[0068] As described herein, combination articles of manufacture are provided that are suitable for performing the method of the invention. Suitable surfaces are described above. The enzyme can be a ligase or a polymerase, as described above. The adaptor sequence can optionally comprise a linker moiety at the 5′ terminus of the adaptor sequence, the linker moiety being suitable for attaching a target molecule to the surface. Other suitable linker moieties are described above.

Example

[0069] The 7249 nucleotide genome of the bacteriophage M13mp18 was sequenced using a single molecule system of the invention. Purified, single-stranded viral M13mp18 genomic DNA was obtained from New England Biolabs. Approximately 25ug of M13 DNA

was digested to an average fragment size of 40 bp with 0.1 U Dnase I (New England Biolabs) for 10 minutes at 37°C. Digested DNA fragment sizes were estimated by running an aliquot of the digestion mixture on a precast denaturing (TBE-Urea) 10% polyacrylamide gel (Novagen) and staining with SYBR Gold (Invitrogen/Molecular Probes). The DNase I-digested genomic DNA was filtered through a YM10 ultrafiltration spin column (Millipore) to remove small digestion products less than about 30 nt. Approximately 20 pmol of the filtered DNase I digest was then polyadenylated with terminal transferase according to known methods (Roychoudhury, R and Wu, R.1980, Terminal transferase-catalyzed addition of nucleotides to the 3' termini of DNA. Methods Enzymol. 65(1):43-62.). The average dA tail length was 50+/-5 nucleotides. Terminal transferase was then used to label the fragments with Cy3-dUTP. Fragments were then terminated with dideoxyTTP (also added using terminal transferase). The resulting fragments were again filtered with a YM10 ultrafiltration spin column to remove free nucleotides and stored in ddH2O at –20°C.

[0070] Epoxide-coated glass slides were prepared for oligo attachment. Epoxide-functionalized 40mm diameter #1.5 glass cover slips (slides) were obtained from Erie Scientific (Salem, NH). The slides were preconditioned by soaking in 3xSSC for 15 minutes at 37°C. Next, a 500pM aliquot of 5' aminated polydT(50) (polythymidine of 50bp in length with a 5' terminal amine) was incubated with each slide for 30 minutes at room temperature in a volume of 80ml. The resulting slides had poly(dT50) primer attached by direct amine linkage to the epoxide. The slides were then treated with phosphate (1M) for 4 hours at room temperature in order to passivate the surface. Slides were then stored in polymerase rinse buffer (20mM Tris, 100mM NaCl, 0.001% Triton X-100, pH 8.0) until they were used for sequencing.

[0071] For sequencing, the slides were placed in a modified FCS2 flow cell (Bioptechs, Butler, PA) using a 50um thick gasket. The flow cell was placed on a movable stage that is part of a high-efficiency fluorescence imaging system built around a Nikon TE-2000 inverted microscope equipped with a total internal reflection (TIR) objective. The slide was then rinsed with HEPES buffer with 100mM NaCl and equilibrated to a temperature of 50°C. An aliquot of the M13 template fragments described above was diluted in 3xSSC to a final concentration of 1.2nM. A 100ul aliquot was placed in the flow cell and incubated on the slide for 15 minutes. After incubation, the flow cell was rinsed with 1xSSC/HEPES/0.1%SDS followed by HEPES/NaCl. A passive vacuum apparatus was used to pull fluid across the flow cell. The resulting slide contained M13 template/oligo(dT) primer duplex. The temperature of the flow

cell was then reduced to 37°C for sequencing and the objective was brought into contact with the flow cell.

[0072] For sequencing, cytosine triphosphate, guanidine triphosphate, adenine triphosphate, and uracil triphosphate, each having a cyanine-5 label (at the 7-deaza position for ATP and GTP and at the C5 position for CTP and UTP (PerkinElmer)) were stored separately in buffer containing 20mM Tris-HCl, pH 8.8, 10 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 10mM HCl, and 0.1% Triton X-100, and 100U Klenow exo$^-$ polymerase (NEN). Sequencing proceeded as follows.

[0073] First, initial imaging was used to determine the positions of duplex on the epoxide surface. The Cy3 label attached to the M13 templates was imaged by excitation using a laser tuned to 532 nm radiation (Verdi V-2 Laser, Coherent, Inc., Santa Clara, CA) in order to establish duplex position. For each slide only single fluorescent molecules were imaged in this step were counted. Imaging of incorporated nucleotides as described below was accomplished by excitation of a cyanine-5 dye using a 635 nm radiation laser (Coherent). 5uM Cy5CTP was placed into the flow cell and exposed to the slide for 2 minutes. After incubation, the slide was rinsed in 1xSSC/15 mM HEPES/0.1% SDS/pH 7.0 ("SSC/HEPES/SDS") (15 times in 60ul volumes each, followed by 150 mM HEPES/150 mM NaCl/pH 7.0 ("HEPES/NaCl") (10 times at 60ul volumes). An oxygen scavenger containing 30% acetonitrile and scavenger buffer (134ul HEPES/NaCl, 24ul 100mM Trolox in MES, pH 6.1, 10ul DABCO in MES, pH6.1, 8ul 2M glucose, 20ul NaI (50mM stock in water), and 4ul glucose oxidase) was next added. The slide was then imaged (500 frames) for 0.2 seconds using an Inova301K laser (Coherent) at 647nm, followed by green imaging with a Verdi V-2 laser (Coherent) at 532nm for 2 seconds to confirm duplex position. The positions having detectable fluorescence were recorded. After imaging, the flow cell was rinsed 5 times each with SSC/HEPES/SDS (60ul) and HEPES/NaCl (60ul). Next, the cyanine-5 label was cleaved off incorporated CTP by introduction into the flow cell of 50mM TCEP for 5 minutes, after which the flow cell was rinsed 5 times each with SSC/HEPES/SDS (60ul) and HEPES/NaCl (60ul). The remaining nucleotide was capped with 50mM iodoacetamide for 5 minutes followed by rinsing 5 times each with SSC/HEPES/SDS (60ul) and HEPES/NaCl (60ul). The scavenger was applied again in the manner described above, and the slide was again imaged to determine the effectiveness of the cleave/cap steps and to identify non-incorporated fluorescent objects.

[0074] The procedure described above was then conducted 100 nM Cy5dATP, followed by 100nM Cy5dGTP, and finally 500nM Cy5dUTP. The procedure (expose to nucleotide, polymerase, rinse, scavenger, image, rinse, cleave, rinse, cap, rinse, scavenger, final image) was repeated exactly as described for ATP, GTP, and UTP except that Cy5dUTP was incubated for 5 minutes instead of 2 minutes. Uridine was used instead of thymidine due to the fact that the Cy5 label was incorporated at the position normally occupied by the methyl group in Thymidine triphosphate, thus turning the dTTP into dUTP. In all 64 cycles (C, A, G, U) were conducted as described in this and the preceding paragraph.

[0075] Once 64 cycles were completed, the image stack data (i.e., the single molecule sequences obtained from the various surface-bound duplex) were aligned to the M13 reference sequence. The image data obtained was compressed to collapse homopolymeric regions. Thus, the sequence "TCAAAGC" would be represented as "TCAGC" in the data tags used for alignment. Similarly, homopolymeric regions in the reference sequence were collapsed for alignment. The sequencing protocol described above resulted in an aligned M13 sequence with an accuracy of between 98.8% and 99.96% (depending on depth of coverage). The individual single molecule sequence read lengths obtained ranged from 2 to 33 consecutive nucleotides with about 12.6 consecutive nucleotides being the average length.

[0076] The alignment algorithm matched sequences obtained as described above with the actual M13 linear sequence. Placement of obtained sequence on M13 was based upon the best match between the obtained sequence and a portion of M13 of the same length, taking into consideration 0, 1, or 2 possible errors. All obtained 9-mers with 0 errors (meaning that they exactly matched a 9-mer in the M13 reference sequence) were first aligned with M13. Then 10-, 11-, and 12-mers with 0 or 1 error were aligned. Finally, all 13-mers or greater with 0, 1, or 2 errors were aligned. At a coverage depth of greater than or equal to 1, 5,001 bases of the 5,066 base M13 collapsed genome were covered at an accuracy of 98.8%. Similarly, at a coverage depth of greater than or equal to 5, 83.6% of the genome was covered at an accuracy of 99.3%, and at a depth of greater than or equal to 10, 51.9% of the genome was covered at an accuracy of 99.96%. The average coverage depth was 12.6 nucleotides.

CLAIMS

We claim:

1. A method for obtaining sequence information from a nucleic acid, the method comprising the steps of:
capturing target nucleic acids with a sequence-specific capture probe to produce a target/probe duplex;
isolating said target nucleic acids from said target/probe duplexes;
attaching said target nucleic acids to a surface such that at least a portion of said target nucleic acids are individually optically resolvable;
exposing said target nucleic acids to a nucleic acid primer capable of forming a target/primer duplex;
introducing a polymerase and at least one nucleotide species comprising an optically-detectable label under conditions sufficient for template-dependent nucleotide addition to said primer;
removing unincorporated nucleotide;
identifying nucleotide species incorporated into said primer, thereby obtaining sequence information from said target.

2. The method of claim 1, further comprising the step of removing said optically-detectable label from nucleotide incorporated into said primer.

3. The method of claim 1, wherein said capturing step comprises hybridizing said target nucleic acids to capture probes immobilized in a polyacrylamide gel.

4. The method of claim 1, wherein said capture probes are attached to a bead.

5. The method of claim 1, wherein said capture probes are immobilized on a chromatography column.

6. The method of claim 1, wherein said optically-detectable label is a fluorescent label.

7. The method of claim 1, wherein said introducing, removing, and identifying steps are repeated at least once.

8. The method of claim 1, wherein said surface comprises an epoxide.

9. The method of claim 1, wherein said attaching step comprises modifying said target nucleic acid to introduce a 3' or 5' amine.

10. The method of claim 1, wherein said primer is attached to said surface.

11. The method of claim 1, wherein said surface is streptavidnated.

12. The method of claim 11, wherein said target and/or said primer comprise a biotin end label.

13. A method for obtaining sequence information from a selected nucleic acid, the method comprising the steps of:
capturing target nucleic acids with a sequence-specific capture probes to produce a plurality of target/probe duplexes;

attaching members of said plurality of duplexes to a surface such that at least some members of said plurality of duplexes are individually optically resolvable;
introducing a polymerase and at least one nucleotide species comprising an optically-detectable label under conditions sufficient for template-dependent nucleotide addition to said probe;
removing unincorporated nucleotide;
identifying nucleotide species incorporated into said probe, thereby obtaining sequence information from said target.

14. The method of claim 13, wherein said introducing, removing, and identifying steps are repeated at least once.

15. The method of claim 14, wherein said surface is an epoxide surface.

16. The method of claim 15, wherein said epoxide surface is derivatized for attachment of said duplex.

17. The method of claim 16, wherein said surface is derivatized with streptavidin.

18. The method of claim 13, wherein said target and said probe are attached to said surface.

19. The method of claim 13, wherein said label is a fluorescent label.

20. A method for obtaining sequence information from a selected nucleic acid, the method comprising the steps of:
capturing target nucleic acids with sequence-specific capture probes attached to a surface to form a target/probe duplex;
introducing a polymerase and at least one nucleotide species comprising an optically-detectable label under conditions sufficient for template-dependent nucleotide addition to said probe;
removing unincorporated nucleotide;
identifying nucleotide species incorporated into said probe, thereby obtaining sequence information from said target.

21. The method of claim 20, wherein said surface is an epoxide surface.

22. The method of claim 20, wherein said introducing, removing, and identifying steps are repeated at least once.

23. The method of claim 20, further comprising the step of washing said surface to remove uncaptured target nucleic acid.

24. The method of claim 20, wherein said label is a fluorescent label.

25. The method of claim 21, wherein said epoxide surface is derivatized with streptavidin.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2006/033214

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 03/016565 A2 (SOLEXA LTD [GB]; BALASUBRAMANIAN SHANKAR [GB]; KLENERMAN DAVID [GB]; B) 27 February 2003 (2003-02-27) the whole document	1-25
X	BRASLAVSKY IDO ET AL: "Sequence information can be obtained from single DNA molecules" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, DC, US, vol. 100, no. 7, 1 April 2003 (2003-04-01), pages 3960-3964, XP002341053 ISSN: 0027-8424 the whole document	1-19

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 January 2007	02/02/2007

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bradbrook, Derek

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2006/033214

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99/57321 A1 (UNIV ARIZONA [US]; UNIV ALBERTA [CA]; WILLIAMS PETER [US]; HAYES MARK) 11 November 1999 (1999-11-11) page 8, line 9 - page 12, line 24 page 18, line 15 - page 20, line 2	20-25
X	US 5 302 509 A (CHEESEMAN PETER C [US]) 12 April 1994 (1994-04-12) column 2, line 15 - line 68; figure 1	20-25
X	WO 2005/040425 A2 (ISIS INNOVATION [GB]; SHCHEPINOV MIKHAIL S [GB]; MIR KALIM [GB]) 6 May 2005 (2005-05-06) the whole document	1-25

1

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2006/033214

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 03016565 A2	27-02-2003	EP 1417341 A2	12-05-2004
		JP 2005500067 T	06-01-2005
		US 2004175716 A1	09-09-2004
WO 9957321 A1	11-11-1999	CA 2330673 A1	11-11-1999
		EP 1082458 A1	14-03-2001
		JP 3813818 B2	23-08-2006
		JP 2002513594 T	14-05-2002
		JP 2005000174 A	06-01-2005
US 5302509 A	12-04-1994	NONE	
WO 2005040425 A2	06-05-2005	EP 1689881 A2	16-08-2006

Electronic Patent Application Fee Transmittal

Application Number:	15383965
Filing Date:	19-Dec-2016
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Filer:	David Christopher Scherer/Jacqueline Lim
Attorney Docket Number:	01-007706US

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
Submission- Information Disclosure Stmt	1806	1	180	180
		Total in USD ($)		**180**

Electronic Acknowledgement Receipt	
EFS ID:	29160379
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	09-MAY-2017
Filing Date:	19-DEC-2016
Time Stamp:	16:09:29
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$180
RAM confirmation Number	051017INTEFSW00003297504427
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	01007706_2017-05-09_Trans.pdf	47175 0f5ca2ef4f763e4862a7837e9d31000c14b068bc7	no	1
Warnings:					
Information:					
2	Transmittal Letter	01007706_2017-05-09_IDS.pdf	24751 c057aff25109dcac83d839e1cb8c463c337a882d	no	2
Warnings:					
Information:					
3	Information Disclosure Statement (IDS) Form (SB08)	01007706_2017-05-09_SB08A.pdf	31926 d25fd5a8285a29768ca09aa05602f8df34b5a9fa	no	3
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
4	Foreign Reference	2017-05-02_ONTinvalidityReportByFloydRomesberg-LapidusWO2007025124A1.pdf	874250 1acfceef732176fae99a91658fe1dda91c2d7bd5	no	30
Warnings:					
Information:					
5	Non Patent Literature	2017-05-02_ONTinvalidityReportByFloydRomesberg-Miner_NucleicAcidsRes.pdf	197354 7d8f6e1ca0670c27229a8d519a5709e1be1f57a2	no	4
Warnings:					
Information:					
6	Other Reference-Patent/App/Search documents	2017-05-02_ONTinvalidityReportByFloydRomesberg.pdf	8097390 1ce6d220134cc1270d1aef4b62e6fed6dc3ef610d	no	309
Warnings:					

Information:					
7	Other Reference-Patent/App/Search documents	2017-05-02_ONTinvalidityReportByFloydRomesberg-PacbioColdSpringHarbor.pdf	25853604 94551c6238cc4a4a6eccdef05ac74c504d0488bc5	no	22
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	30677 d1ff83d823bf4fa1c56ca89a84cae7c21ccb9270	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	35157127		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/21 (10-08)
Approved for use through 11/30/2008. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	15/383,965
Filing Date	December 19, 2016
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

ENCLOSURES *(Check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☑ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation
- ☐ Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- ☐ Proprietary Information
- ☐ Status Letter
- ☑ Other Enclosure(s) (please Identify below):
  PTO Form SB08A with 4 refs

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Pacific Biosciences of California, Inc.		
Signature	/David C. Scherer, Ph.D./		
Printed name	David C. Scherer, Ph.D.		
Date	May 9, 2017	Reg. No.	56,993

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below: via EFS-Web

Signature	/Jacqueline L. Lim/		
Typed or printed name	Jacqueline L. Lim	Date	May 9, 2017

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

May 9, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
Attorney Docket No. 01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Stephen Turner et al.

Application No.: 15/383,965

Filed: December 19, 2016

For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Examiner: Wilder, Cynthia B.

Art Unit: 1637

Confirmation No: 8144

INFORMATION DISCLOSURE STATEMENT UNDER 37 CFR § 1.97 AND 1.98

COMMISSIONER FOR PATENTS
P. O. BOX 1450
Alexandria, CA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Sir:

This information Disclosure Statement is submitted:

(a) [ ] **Under 37 CFR 1.97(b)**
(Within three months of filing national application or date of entry of national application or before mailing date of First Office Action on the merits whichever occurs last).

(b) [X] **Under 37 CFR 1.97(c) together with either a:**
[ ] Statement under CFR 1.97(e) or
[X] $180.00 fee under 37 CFR 1.17(p)
(After the CFR 1.97(b) time period, but before a final action, notice of allowance, whichever occurs first).

(c) [ ] **Under 37 CFR 1.97(d) together with: a**
Statement under 37 CFR 1.97(e), and
$180.00 fee as set forth in 37 CFR 1.17(p).
(After a final action or notice of allowance, whichever occurs first, but before payment of the issue fee).

[ ] **STATEMENT UNDER 37 CFR 1.97(e)**

The undersigned certifies that:

[ ] Each item of information contained in the Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the statement, or

[ ] No item of information contained in the Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart foreign application , and to the knowledge of the undersigned after making reasonable inquiry, was known to any individual designated in 37 CFR §1.56 more than three months prior to the filing of the Information Disclosure Statement.

[ ] PRIOR APPLICATIONS

All references in the enclosed PTO Form SB08A which are required to be included in this submission, were disclosed in prior Patent Application No.: *, filed *, and, as such, copies thereof are not included pursuant to the provisions of 37 CFR 1.98(d).

[ ] FOREIGN LANGUAGE DOCUMENTS

A concise explanation of the relevance of foreign language patents, foreign language publications and other foreign language information listed on PTO form 1449, as presently understood by the individual(s) designated in 37 CFR 1.56 most knowledgeable about the content is given on the attached sheet, or where a foreign language patent is cited in a search report or other action by a foreign patent office in a counterpart foreign application, an English language version of the search or action which indicates the degree of relevance found by the foreign office is listed on form PTO 1449 and is enclosed herewith.

[X] FEE AUTHORIZATION

Please charge to Deposit Account No. 50-4427 the sum of $ 180.00 at anytime during the pendency of this application, please charge any fees required or credit any overpayment to Deposit Account No.50-4427.

Respectfully Submitted,

May 9, 2017
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF CALIFORNIA
1305 O'Brien Drive
Menlo Park, CA 94025
Phone: (650) 521-8127
Fax: (650) 323-9420
Email: dscherer@pacb.com

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

57770 7590 06/05/2017

Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

EXAMINER
WILDER, CYNTHIA B

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
06/05/2017	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

PATENTS@PACB.COM

Office Action Summary	Application No. 15/383,965	Applicant(s) TURNER ET AL.	
	Examiner CYNTHIA B. WILDER	Art Unit 1637	AIA (First Inventor to File) Status No

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☐ Claim(s) *1,2,4,5,8-10,12-15 and 18-35* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1,2,4,5,8-10,12-15 and 35* is/are rejected.

8)☒ Claim(s) *18-34* is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *1-20-2017* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date *5/9/2017 and 3/31/2017*.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____

4) ☐ Other: _____.

DETAILED ACTION

1. The present application is being examined under the pre-AIA first to invent provisions. Applicant's amendment filed 3/31/2017 is acknowledged. Claims 1, 4, 5, 8, 9, 10, 15, 16 have been amended. Claims 3, 6, 7, 11 and 17 have been canceled. Claims 19-35 have been added. All of the arguments have been thoroughly reviewed and considered.

Any rejection not reiterated in this action has been withdrawn as being obviated by the amendment of the claims.

This action is made FINAL.

The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

Previous Rejections

2. The objection to the specification for the trade name is withdrawn in view of Applicant's amendment to the specification. The claim rejection under 35 USC 101 directed to the claims as encompassing an abstract idea is withdrawn in view of Applicant's amendment to the claims. The examiner agrees that the combination of method steps recited in the claims as amended recite significantly more than a judicial exception under 35 USC 101. The double patenting rejection is withdrawn in view of Applicant's filing of a proper terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d).

New Ground(s) of Rejections

THE NEW GROUND(S) OF REJECTIONS WERE NECESSITATED BY APPLICANT'S AMENDMENT OF THE CLAIMS:

Claim Rejections - 35 USC § 103

3. The following is a quotation of pre-AIA 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. This application currently names joint inventors. In considering patentability of the claims under pre-AIA 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of pre-AIA 35 U.S.C. 103(c) and potential pre-AIA 35 U.S.C. 102(e), (f) or (g) prior art under pre-AIA 35 U.S.C. 103(a).

5. Claim 1, 2, 4, 5, 8, 9, 10, 12, 13, 14, 15, and 35 is/are rejected under pre-AIA 35 U.S.C. 103(a) as being unpatentable over Dennison et al (20030044816, March 2003). Regarding claims 1, 8 and 35, Dennison et al teach a method of determining a nucleotide sequence of a region of interest in a polynucleotide, the method comprising: introducing a polynucleotide comprising a region of interest to a sequence analysis system comprising a nanopore in a membrane; applying a voltage across the

membrane; monitoring variations in ionic current through the nanopore of the sequence analysis system during enzyme chaperone-regulated passage of the polynucleotide through the nanopore; analyzing the monitored variations in ionic current to obtain nucleotide sequence information for the polynucleotide (para. [0013], [0018], [0022]-[0023], [0029], [0043], [0063]-[0067], [0074], [0090], [0091] and [0095]).

With regards to the limitation "wherein the nucleotide sequence information comprises redundant sequence information for the region of interest; and determining a consensus sequence for the region of interest based on the redundant sequence information", Dennison et al no expressly teach determining the consensus sequence for the region of interest based on redundant sequence information. However, Dennison provides the capability of obtaining redundant sequence information in order to obtain consensus information. Dennison et al teach that the method allows long stretches to be read so that errors associated with assembly and repetitive sequences can be minimized (para. [0067]). Dennison et al teach to resolve stretches of repeating identical bases accurately, and to minimize reading errors in general, it may be useful for the pore to register a distinct (probably higher) level of conductance in between the bases. Dennison et al teach that if the rate of movement is constant, then punctuation between bases may not be required to resolve stretches of repeating identical bases ([0080]). At paragraph [0095], Dennison et al teach that the length of continuous DNA sequence obtainable from the methods described in the document will only be limited in certain embodiments (e.g., by the packaging limit of phage lambda heads (.about.50 kb) or by the size of the template containing polymerase promoter sequences). Dennison et

al teach that other embodiments (e.g., voltage gradients) have no such limitation and should even make it possible to sequence DNA directly from tissue samples, since the technique is not limited to cloned DNA. Dennison states that having large contiguous sequence as primary input data will substantially reduce the complexity of sequence assembly, particularly in the case of repetitive DNA.

It would have been *prima facie* obvious to one of ordinary skill in the art at the time of the claimed invention to carry out the invention with a reasonable expectation of success since such steps of detecting desired consensus sequences and redundant sequences using the sequencing methodology of Dennison is within the ordinary artisan capabilities and would not negatively alter or modify the results of sequencing a desired target sequence.

Regarding claim 2, Dennison et al teach wherein the nanopore comprises a protein channel (e.g., [0069], [0136]).

Regarding claim 4, Dennison et al teach wherein the membrane is a lipid bilayer ([0136]).

Regarding claim 5, Dennison et al teach wherein the membrane is a solid-state membrane ([0066], [0072]).

Regarding claims 9-10, Dennison et al teach wherein the polynucleotide comprises complementary sequences of the region of interest and wherein the complementary sequences are linked ([0024], [0042]).

Regarding claims 12 and 15-16, Dennison et al teach wherein the polynucleotide comprises double stranded DNA portions that may comprise of complementary sequence portions ([0023] and [0024], [0042]).

Regarding claims 13-14, Dennison teaches that the invention features detection of double stranded regions of nucleic acid (0023). Dennison teaches that long stretches of DNA may be detected using the sequencing methodology [(0067)]. Thus, the claimed limitations are deemed to be inherent in the teachings of Dennison.

Conclusion

6. Claims 1, 2, 4, 5, 8, 9, 10, 12, 13, 14, 15, and 35 are rejected. Claims 18-34 are objected because they depend from rejected claims. The claims 18-34 have not been rejected under prior art.

7. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to CYNTHIA B. WILDER whose telephone number is (571)272-0791. The examiner can normally be reached on a flexible schedule.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Gary Benzion can be reached on (571) 272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/CYNTHIA B WILDER/
Primary Examiner, Art Unit 1637

Notice of References Cited	Application/Control No. 15/383,965	Applicant(s)/Patent Under Reexamination TURNER ET AL.	
	Examiner CYNTHIA B. WILDER	Art Unit 1637	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-2003/0044816 A1	03-2003	Denison, Timothy J.	B82Y15/00	435/6.12
	B	US-				
	C	US-				
	D	US-				
	E	US-				
	F	US-				
	G	US-				
	H	US-				
	I	US-				
	J	US-				
	K	US-				
	L	US-				
	M	US-				

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

WEST Search History for Application 15383965

Creation Date: 2017052418:09

Prior Art Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
(polynucleotide same nanopore same enzyme chaperone)	PGPB, USPT, USOC, EPAB, DWPI	10	ADJ	YES		03-06-2017
(nanopore same enzyme chaperone)	PGPB, USPT, USOC, EPAB, DWPI	10	ADJ	YES		03-06-2017
9057102.pn. or 8143030.pn.	PGPB, USPT, USOC, EPAB, DWPI	5	ADJ	YES		03-06-2017
15/383965	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		03-06-2017
(nanopore same enzyme).clm.	PGPB, USPT, USOC, EPAB, DWPI	54	ADJ	YES		03-06-2017
((nanopore same enzyme).clm.) same polynucleotide	PGPB, USPT, USOC, EPAB, DWPI	8	ADJ	YES		03-06-2017
(9057102.pn. or 8143030.pn.) and ((C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333).CPC.)	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		03-06-2017

20080318184	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		03-06-2017
(20080318184) and (consensus sequence or consensus or redundant sequence)	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		03-06-2017
nanopore same (consensus sequence)	PGPB, USPT, USOC, EPAB, DWPI	31	ADJ	YES		03-06-2017
13/866603 or 12/413258 or 13/403789	PGPB, USPT, USOC, EPAB, DWPI	104	ADJ	YES		05-24-2017
8455193.pn. or 8153375.pn.	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
20050266456	PGPB, USPT, USOC, EPAB, DWPI	7	ADJ	YES		05-24-2017
(20050266456) and nanopore	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		05-24-2017
WO 2007025124	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		05-24-2017
15/383965	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017

(15/383965) and nanopore	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965 and nanopore) and Clarke	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		05-24-2017
(8455193.pn. or 8153375.pn.) and clarke	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		05-24-2017
9057102.pn. or 8143030.pn.	PGPB, USPT, USOC, EPAB, DWPI	5	ADJ	YES		05-24-2017
15/383965	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(nanopore same enzyme).clm.	PGPB, USPT, USOC, EPAB, DWPI	57	ADJ	YES		05-24-2017
((nanopore same enzyme).clm.) same polynucleotide	PGPB, USPT, USOC, EPAB, DWPI	9	ADJ	YES		05-24-2017
(9057102.pn. or 8143030.pn.) and (((C12Q1/6869) | (C12Q2533/101) | (C12Q2537/149) | (C12Q2565/631) | (C12Q2525/301) | (G01N2021/7786) | (G01N2021/6439) | (G01N21/6428) | (G01N21/6486) | (C12N9/1252) | (G06F19/22) | (Y10T436/143333)).CPC.)	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		05-24-2017
20060063171 or 20030059778 or 20030044816	PGPB, USPT, USOC,	7	ADJ	YES		05-24-2017

	EPAB, DWPI					
(20060063171 or 20030059778 or 20030044816) and nanopore	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore) and (appl$ or voltage$ or membrane)	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane)) and (ionic current or current)	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current)) and (enzyme or chaperone or pass$)	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current) and (enzyme or chaperone or pass$)) and variation	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current) and (enzyme or chaperone or pass$) and variation) and (consensus or redundant or information)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current) and (enzyme or chaperone or pass$) and variation) and (protein channel)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
20030044816	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		05-24-2017
(20030044816) and nanopore	PGPB, USPT,	1	ADJ	YES		05-24-2017

	USOC, EPAB, DWPI					
(20030044816 and nanopore) and sequencing	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(20030044816) and (enzyme or chaperone)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(20030044816 and (enzyme or chaperone)) and linker	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965) and (registration sequence)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965 and (registration sequence)) and synthetic linker	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965 and (registration sequence) and synthetic linker) and carbon-based linker	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-10)
Approved for use through 07/31/2012 OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5308751		1994-05-03	Ohkawa et al.	
	2	6673615		2004-01-06	Denison et al.	
	3	8168380		2012-05-01	Chan	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20020028458	A1	2002-03-07	Lexow	
	2	20020168645	A1	2002-11-14	Taylor	
	3	20040248161	A1	2004-12-09	Rotherberg et al.	

/CYNTHIA B WILDER/ (05/30/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

	4	20060147935	A1	2006-07-06	Linnarsson	
	5	20070190556	A1	2007-08-16	Brenner et al.	
	6	20070231808	A1	2007-10-04	Gouda et al.	
	7	20080051294	A1	2008-02-28	Gormely et al.	
	8	20060024711		2006-02-02	Lapidus et al.	
	9	20020094526		2002-07-18	Bayley et al.	
	10	20100216151		2010-08-26	Lapidus et al.	
	11	20040002077		2004-01-01	Taira et al.	
	12	20050069939		2005-03-31	Wang et a.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]

/CYNTHIA B WILDER/ (05/30/2017)

EFS Web 2.1 17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	1							☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	Response and Exhibit List of Oxford Nanopore Technologies Ltd., Oxford Nanopore Technologies, Inc., and Metrichor, Ltd. Dated 1/6/2017 To the Complaint of Pacific Biosciences of California, Inc., and Notice of Investigation in the Matter of "Certain Single-Molecule Nucleic Acid Sequencing Systems and Reagents, Consumables, and Software for Use with Same (Investigation No. 337-TA-1032)	☐
	2	Ezzevaz-Roulet et al., "Mechanical Separation of the Complementary Strands of DNA," Proc. Natl. Acad. Sci. USA, Vol. 94, pp 11935-11940 (Oct. 1997)	☐
	3	Greenleaf et al., "Single-Molecule, Motion-Based DNA Sequencing Using RNAPolymerase," Science, 313(5788): 801 (Aug. 2006)	☐
	4	Hattori et al, "Dideoxy Sequening Method Using Denatured Plasmid Templates," Analytical Biochemistry, Vol. 152, pp 232-238 (1986)	☐
	5	Hayashizaki et al., "A new method for constructing NotI linking and boundary libraries using a restriction trapper," Genomics, Vol. 14, pp 733-739 (1992)	☐
	6	Jarvie et al., "3K Long-Tag Paired End sequencing with the Genome Sequencer FLX System," BioTechniques, Vol. 44, No. 6, pp 829-831 (2008)"	☐
	7	Kalisch et al., "Covalently linked sequencing primer linkers (splinkers) for sequence analysis of restriction fragments," Gene, Vol. 44, pp 263-270 (1986)	☐
	8	Kambara et al., "Real Time Automated Simultaneous Double-Stranded DNA Sequencing Using Two-Color Fluorophore Labeling," Biotechnology, Vol 9, pp 648-651 (July 1991)	☐

/CYNTHIA B WILDER/ (05/30/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(**Not for submission under 37 CFR 1.99**)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

	9	Kaur et al., "Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA," Nucleic Acids Research, Vol. 31, No. 6 e26 (2003)	☐
	10	"Kim et al., "DARFA: a novel technique for studying differential gene expression and bacterial comparative genomics," Biochemical and Biophysical Research Communications, Vol. 336, pp 168–174 (2005)"	☐
	11	Kuhn et al., "High-Purity Preparation of a Large DNA Dumbbell," Antisense & Nucleic Acid Drug Development, Vol. 11, pp 149-153 (2001)	☐
	12	Liu et al., "Rolling Circle DNA Synthesis: Small Circular Oligonucleotides as Efficient Templates for DNA Polymerases," J. Am. Chem. Soc., Vol. 118, pp 1587-1594 (1996)"	☐
	13	Luo et al., "Small interfering RNA production by enzymatic engineering of DNA (SPEED)," PNAS, Vol. 101 No. 15, pp 5494-5499 (Apr. 2004)	☐
	14	Miller et al., "Chain Terminator Sequencing of Double-stranded DNA with Built-In Error Detection," J. theor. Biol. Vol. 161, pp 407-429 (1993)	☐
	15	Taki et al., "Small-interfering-RNA expression in cells based on an efficiently constructed dumbbell-shaped DNA," Angew. Chem. Int. Ed. Vol. 43, pp 3160-3163 (2004)	☐
	16	Thelwell et al., "Mode of action and application of Scorpion primers to mutation detection," Nucleic Acids Research, Vol. 28, No. 19, pp 3752-3761 (2000)	☐
	17	Vercoutere, et al., "Discrimination among individual Watson-Crick base pairs at the termini of single DNA hairpin molecules," Nucleic Acids Research, Vol. 31, No. 4, pp 1311-1318 (2003)	☐
	18	Vercoutere, et al., "Rapid discrimination among individual DNA hairpin molecules at singlenucleotide resolution using an ion channel," Nature Biotechnology, Vol. 19, pp 248-252 (Mar. 2001)	☐
	19	Wiemann et al., "Simultaneous On-Line DNA Sequencing on Both Strands with Two Fluourescent Dyes," Analytical Biochemistry, Vol. 224, pp 117-121 (1995)"	☐

/CYNTHIA B WILDER/ (05/30/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	20	Winters-Hilt et al., "Highly Accurate Classification of Watson-Crick Base pairs on Termini of Single DNA Molecules," Biophysical Journal, Vol. 84, pp 967–976 (Feb. 2003)	☐
	21	Woodside et al., "Nanomechanical Measurements of the Sequence-Dependent Folding Landscapes of Single Nucleic Acid Hairpins," Proc. Natl. Acad. Sci. USA, Vol. 103, pp6190-6195 (Apr. 2006)"	☐
	22	Zanta et al., "Gene delivery: a single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus," Proc. Natl. Acad. Sci. USA, Vol. 96, pp 91-96 (Jan. 1999)	☐
	23	Definition of "Consensus Sequence" from Medical dictionary, Printed on 2/6/2017	☐
	24	KEANE, T. et al., "Assessing Assemblability of Reads from New sequencing Platforms" Wellcome Trust Poster, Page 1, 15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) & 6th European Conference on Computational Biology (ECCB), Vienna, Austria July 21-25, 2007	☐
	25	Verified Complaint of Pacific Biosciences of California, Inc. Under Section 337 of the Tariff Act of 1930 filed with the United States International Trade Commission on November 2, 2016	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature	/CYNTHIA B WILDER/ (05/30/2017)	Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David C. Scherer, Ph.D./	Date (YYYY-MM-DD)	2017-03-31
Name/Print	David C. Scherer, Ph.D.	Registration Number	56,993

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15383965	TURNER ET AL.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

CPC- SEARCHED		
Symbol	Date	Examiner
C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333	3/6/2017	CW
updated	5/30/2017	CW

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES		
Search Notes	Date	Examiner
Inventor search on PALM, review of related applictions as cited in PALM, review of IDS filed in instant invention and cited documents in copending applications, text search on WEST	3/6/2017	CW
Updated	5/30/2017	CW

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner

INTERFERENCE SEARCH			
US Class/ CPC Symbol	**US Subclass / CPC Group**	**Date**	**Examiner**
C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333		3/7/2017	CW
updated		5/30/2017	CW

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-10)
Approved for use through 07/31/2012 OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	6404907	B1	2002-06-11	Gilchrist et al.	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20050266456	A1	2005-12-01	Williams et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2007025124	WO	A1	2007-03-01	Lapidus		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

/CYNTHIA B WILDER/ (05/30/2017)

EFS Web 2.1 17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	15383965
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	1637
	Examiner Name	Wilder, Cynthia B.
	Attorney Docket Number	01-007706US

	1	Expert Report of Floyd Romesberg, Ph.D. Regarding U.S. Patent Nos. 9,404,146 and 9,542,527 as they relate to the Complaint filed by Pacific Biosciences Under Section 337 of the Tariff Act of 1930 with the United States International Trade Commission on November 2, 2016	☐
	2	MINER et al., "Molecular Barcodes Detect Redundancy and Contamination in Hairpin-Bisulfite PCR," Nucl. Acids Res. (2004) 32(17):e135.	☐
	3	Pacific Biosciences Presentation for Cold Spring Harbor Personal Genomics Meeting on October 12, 2008	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature	/CYNTHIA B WILDER/ (05/30/2017)	Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

/CYNTHIA B WILDER/ (05/30/2017)

/CYNTHIA B WILDER/ (05/30/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(**Not for submission under 37 CFR 1.99**)

Application Number	15383965
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	1637
Examiner Name	Wilder, Cynthia B.
Attorney Docket Number	01-007706US

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/David C. Scherer, Ph.D./	Date (YYYY-MM-DD)	2017-05-09
Name/Print	David C. Scherer, Ph.D.	Registration Number	56,993

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

June 5, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Inventor: Stephen Turner, et al.

Application No.: 15/383,965

Filed: December 19, 2016

For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Examiner: Wilder, Cynthia B.

Confirmation No.: 8144

Art Unit: 1637

RESPONSE TO FINAL OFFICE ACTION

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INTRODUCTORY REMARKS

This amendment is filed in response to the Final Office Action dated June 5, 2017 for which a 3-month period for response was given, and as such is timely filed. Applicant respectfully requests entry of the amendments and remarks provided herein.

Amendments to the Claims begin on page **2** of this paper.

Remarks begin at page **5** of this paper.

AMENDMENTS TO THE CLAIMS

Please incorporate the following amendments to the claims of the subject application.

1. (Currently Amended) A method of determining a nucleotide sequence of a region of interest in a polynucleotide, the method comprising:

introducing a polynucleotide comprising a region of interest to a sequence analysis system comprising a nanopore in a membrane, wherein the polynucleotide comprises a double-stranded portion comprising complementary strands of the region of interest;

applying a voltage across the membrane;

monitoring variations in ionic current through the nanopore of the sequence analysis system during enzyme chaperone-regulated passage of the polynucleotide through the nanopore;

analyzing the monitored variations in ionic current to obtain nucleotide sequence information for the polynucleotide, wherein the nucleotide sequence information comprises redundant sequence information for the region of interest, wherein the redundant sequence information comprises the nucleotide sequence of the complementary strands; and

determining a consensus sequence for the region of interest based on the redundant sequence information.

2. (Original) The method of claim 1, wherein the nanopore comprises a protein channel.

3. (Canceled)

4. (Previously Presented) The method of claim 1, wherein the membrane is a lipid bilayer.

5. (Previously Presented) The method of claim 1, wherein the membrane is a solid-state membrane.

6-7. (Canceled)

8. (Previously Presented) The method of claim 1, further comprising changing reaction conditions to alter the rate of enzyme chaperone regulated passage of the polynucleotide through the nanopore.

9-12. (Canceled)

13. (Currently Amended) The method of claim [[12]]1, wherein the polynucleotide is greater than 75% double-stranded DNA.

14. (Currently Amended) The method of claim [[12]]1, wherein the polynucleotide is greater than 90% double-stranded DNA.

15. (Canceled)

16. (Currently Amended) The method of claim [[15]]1, wherein the complementary strands are linked.

17. (Canceled)

18. (Currently Amended) The method of claim 1, wherein the polynucleotide comprises multiple repeats of the region of interest, wherein the redundant sequence information further comprises the nucleotide sequence of the multiple repeats.

19-26. (Canceled)

27. (Previously Presented) The method of claim 16, wherein the complementary strands are linked by a linker comprising a nucleotide.

28. (Previously Presented) The method of claim 27, wherein the linker comprises an oligonucleotide.

29. (Previously Presented) The method of claim 28, wherein the oligonucleotide comprises a registration sequence.

30. (Previously Presented) The method of claim 28, wherein the linker comprises a nick.

31. (Previously Presented) The method of claim 16, wherein the complementary strands are linked by a synthetic linker.

32. (Previously Presented) The method of claim 31, wherein the synthetic linker is a carbon-based linker.

33. (Previously Presented) The method of claim 8, wherein the monitoring comprises a detection period and a non-detection period, wherein the rate of passage of the polynucleotide through the nanopore by the enzyme chaperone is sped up during the non-detection period and slowed during the detection period, wherein the monitored variations in ionic current through the nanopore in the detection period is subjected to the analyzing and determining steps.

34. (Previously Presented) The method of claim 33, wherein the monitoring comprises multiple detection periods.

35. (Canceled)

REMARKS

Formal Matters

Claims 3, 6, 7, 11 and 17 were previously canceled. Claims 9, 10, 12, 15, 19 to 26, and 35 are canceled in the present amendment. Claim 1 is amended to include subject matter from Claims 15 and 26. Claims 13, 14 and 16 are amended to update dependency. Claim 18 is amended for clarity.

No new matter has been added.

Claim Rejection – 25 USC § 103(a)

Claims 1, 2, 4, 5, 8, 9, 10, 12, 13, 14, 15, and 35 are rejected under 35 U.S.C. § 103(a) for assertedly being unpatentable over Dennison et al. (US20030044816, March 2003).

While not acquiescing to the correctness of this rejection, Applicants have amended the claims to overcome it.

Specifically, Applicants have amended independent Claim 1 to incorporate the subject matter Claim 26 and intervening dependent Claim 15. As noted in the Final Office Action on page 6, Claim 26 was not rejected under prior art but rather because it was dependent on a rejected base claim (Claim 15).

In view of the amendments to Claim 1, Claims 9, 10, 12, 15, 19 to 26, and 35 have been canceled. Claims 13, 14, and 16 have been amended to update their dependency, and Claim 18 has been amended for clarity with respect to the redundant sequence information.

In view of these amendments, amended Claim 1 is patentable over Dennison et al. As all remaining pending claims depend directly or indirectly from amended Claim 1, they are likewise patentable over Dennison et al.

Withdrawal of this rejection is therefore respectfully requested.

CONCLUSION

Based upon the foregoing amendments and remarks, Applicant believes the instant application is in condition for allowance and action toward that end is respectfully requested. The Commissioner is authorized to charge any additional required fees or credit any overpayment to Deposit Account No. 50-4427 (referencing Attorney Docket No. 01-007706US). If the Office believes there are additional issues that have not been addressed, the Office is encouraged to contact Applicant's undersigned representative at (650) 521-8127.

Respectfully submitted,

June 5, 2017
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF
CALIFORNIA, INC.
1305 O'Brien Drive
Menlo Park, CA 94025

Electronic Acknowledgement Receipt

EFS ID:	29397830
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	05-JUN-2017
Filing Date:	19-DEC-2016
Time Stamp:	15:48:33
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		01007706_2017-06-05_FOARes p.pdf	32433 6d920f493e69ca0fa5804b5aa0362374e1b96706	yes	6

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Response After Final Action	1	1
	Claims	2	4
	Applicant Arguments/Remarks Made in an Amendment	5	6
Warnings:			
Information:			
	Total Files Size (in bytes):	32433	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PATENT APPLICATION FEE DETERMINATION RECORD
Substitute for Form PTO-875

Application or Docket Number	Filing Date	☐ To be Mailed
15/383,965	12/19/2016	

ENTITY: ☒ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*	X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*	X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT 06/05/2017

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	* 17	Minus	** 30	= 0	X $80 =	0
Independent (37 CFR 1.16(h))	* 1	Minus	***3	= 0	X $420 =	0
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	0

AMENDMENT

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1

LIE
Tina J. Barden

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

57770 7590 06/23/2017

Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

EXAMINER	
WILDER, CYNTHIA B	
ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 06/23/2017

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-00770601JS	8144

TITLE OF INVENTION: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$960	$0	$0	$960	09/25/2017

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.

PTOL-85 (Rev. 02/11)

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail **Mail Stop ISSUE FEE**
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

57770 7590 06/23/2017
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission
I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

TITLE OF INVENTION: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$960	$0	$0	$960	09/25/2017

EXAMINER	ART UNIT	CLASS-SUBCLASS
WILDER, CYNTHIA B	1637	435-006120

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).
- ❑ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
- ❑ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively, 1
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed. 2 3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent): ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. The following fee(s) are submitted:
- ❑ Issue Fee
- ❑ Publication Fee (No small entity discount permitted)
- ❑ Advance Order - # of Copies _____

4b. Payment of Fee(s): **(Please first reapply any previously paid issue fee shown above)**
- ❑ A check is enclosed.
- ❑ Payment by credit card. Form PTO-2038 is attached.
- ❑ The director is hereby authorized to charge the required fee(s), any deficiency, or credits any overpayment, to Deposit Account Number _____ (enclose an extra copy of this form).

5. **Change in Entity Status** (from status indicated above)
- ❑ Applicant certifying micro entity status. See 37 CFR 1.29 — NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
- ❑ Applicant asserting small entity status. See 37 CFR 1.27 — NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
- ❑ Applicant changing to regular undiscounted fee status. — NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature Date

Typed or printed name Registration No.

PTOL-85 Part B (10-13) Approved for use through 10/31/2013. OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

57770 7590 06/23/2017
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

EXAMINER	
WILDER, CYNTHIA B	

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 06/23/2017

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

PTOL-85 (Rev. 02/11)

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 15/383,965	Applicant(s) TURNER ET AL.	
	Examiner CYNTHIA B. WILDER	Art Unit 1637	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to *6/5/2017*.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on_____.
2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.
3. ☒ The allowed claim(s) is/are *1,2, 4, 5, 8, 13, 14, 16, 18, 27-34*. As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.
4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:
      1. ☐ Certified copies of the priority documents have been received.
      2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
      3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____.

Applicant has **THREE MONTHS FROM THE "MAILING DATE"** of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☐ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date _____ .
5. ☐ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/CYNTHIA B WILDER/
Primary Examiner, Art Unit 1637

WEST Search History for Application 15383965

Creation Date: 2017061617:30

Prior Art Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
(polynucleotide same nanopore same enzyme chaperone)	PGPB, USPT, USOC, EPAB, DWPI	10	ADJ	YES		03-06-2017
(nanopore same enzyme chaperone)	PGPB, USPT, USOC, EPAB, DWPI	10	ADJ	YES		03-06-2017
9057102.pn. or 8143030.pn.	PGPB, USPT, USOC, EPAB, DWPI	5	ADJ	YES		03-06-2017
15/383965	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		03-06-2017
(nanopore same enzyme).clm.	PGPB, USPT, USOC, EPAB, DWPI	54	ADJ	YES		03-06-2017
((nanopore same enzyme).clm.) same polynucleotide	PGPB, USPT, USOC, EPAB, DWPI	8	ADJ	YES		03-06-2017
(9057102.pn. or 8143030.pn.) and ((C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333).CPC.)	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		03-06-2017

20080318184	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		03-06-2017
(20080318184) and (consensus sequence or consensus or redundant sequence)	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		03-06-2017
nanopore same (consensus sequence)	PGPB, USPT, USOC, EPAB, DWPI	31	ADJ	YES		03-06-2017
13/866603 or 12/413258 or 13/403789	PGPB, USPT, USOC, EPAB, DWPI	104	ADJ	YES		05-24-2017
8455193.pn. or 8153375.pn.	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
20050266456	PGPB, USPT, USOC, EPAB, DWPI	7	ADJ	YES		05-24-2017
(20050266456) and nanopore	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		05-24-2017
WO 2007025124	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		05-24-2017
15/383965	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017

(15/383965) and nanopore	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965 and nanopore) and Clarke	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		05-24-2017
(8455193.pn. or 8153375.pn.) and clarke	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		05-24-2017
9057102.pn. or 8143030.pn.	PGPB, USPT, USOC, EPAB, DWPI	5	ADJ	YES		05-24-2017
15/383965	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(nanopore same enzyme).clm.	PGPB, USPT, USOC, EPAB, DWPI	57	ADJ	YES		05-24-2017
((nanopore same enzyme).clm.) same polynucleotide	PGPB, USPT, USOC, EPAB, DWPI	9	ADJ	YES		05-24-2017
(9057102.pn. or 8143030.pn.) and (((C12Q1/6869) ! (C12Q2533/101) ! (C12Q2537/149) ! (C12Q2565/631) ! (C12Q2525/301) ! (G01N2021/7786) ! (G01N2021/6439) ! (G01N21/6428) ! (G01N21/6486) ! (C12N9/1252) ! (G06F19/22) ! (Y10T436/143333)).CPC.)	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		05-24-2017
20060063171 or 20030059778 or 20030044816	PGPB, USPT, USOC,	7	ADJ	YES		05-24-2017

	EPAB, DWPI					
(20060063171 or 20030059778 or 20030044816) and nanopore	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore) and (appl$ or voltage$ or membrane)	PGPB, USPT, USOC, EPAB, DWPI	4	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane)) and (ionic current or current)	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current)) and (enzyme or chaperone or pass$)	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current) and (enzyme or chaperone or pass$)) and variation	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current) and (enzyme or chaperone or pass$) and variation) and (consensus or redundant or information)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(20060063171 or 20030059778 or 20030044816 and nanopore and (appl$ or voltage$ or membrane) and (ionic current or current) and (enzyme or chaperone or pass$) and variation) and (protein channel)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
20030044816	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		05-24-2017
(20030044816) and nanopore	PGPB, USPT,	1	ADJ	YES		05-24-2017

	USOC, EPAB, DWPI					
(20030044816 and nanopore) and sequencing	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(20030044816) and (enzyme or chaperone)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(20030044816 and (enzyme or chaperone)) and linker	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965) and (registration sequence)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965 and (registration sequence)) and synthetic linker	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
(15/383965 and (registration sequence) and synthetic linker) and carbon-based linker	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		05-24-2017
20030044816	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		06-16-2017
(20030044816) and (redundant)	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		06-16-2017

OK TO ENTER: /C.B.W/

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

June 5, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Inventor: Stephen Turner, et al.

Application No.: 15/383,965

Filed: December 19, 2016

For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

Examiner: Wilder, Cynthia B.

Confirmation No.: 8144

Art Unit: 1637

RESPONSE TO FINAL OFFICE ACTION

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INTRODUCTORY REMARKS

This amendment is filed in response to the Final Office Action dated June 5, 2017 for which a 3-month period for response was given, and as such is timely filed. Applicant respectfully requests entry of the amendments and remarks provided herein.

Amendments to the Claims begin on page **2** of this paper.

Remarks begin at page **5** of this paper.

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15383965	TURNER ET AL.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

CPC						
Symbol					Type	Version
C12Q		1	/	6869	F	2013-01-01
G01N		33	/	48721	I	2013-01-01
			/			
			/			
			/			
			/			
			/			
			/			
			/			
			/			
			/			
			/			
			/			
			/			
			/			

CPC Combination Sets							
Symbol					Type	Set	Ranking	Version
			/					
			/					

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	17	
/CYNTHIA B WILDER/ Primary Examiner.Art Unit 1637	6/16/2017	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	NONE

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15383965	TURNER ET AL.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

US ORIGINAL CLASSIFICATION		INTERNATIONAL CLASSIFICATION										
CLASS	SUBCLASS	CLAIMED					NON-CLAIMED					
435	6.1	C	1	2	P	19 / 34 (2006.01.01)						
CROSS REFERENCE(S)												
CLASS	SUBCLASS (ONE SUBCLASS PER BLOCK)											

NONE (Assistant Examiner)	(Date)	Total Claims Allowed: 17	
/CYNTHIA B WILDER/ Primary Examiner.Art Unit 1637 (Primary Examiner)	6/16/2017 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure NONE

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15383965	TURNER ET AL.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☒ T.D. ☐ R.1.47

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
1	1	-	17	16	33										
2	2	9	18	17	34										
-	3	-	19	-	35										
3	4	-	20												
4	5	-	21												
-	6	-	22												
-	7	-	23												
5	8	-	24												
-	9	-	25												
-	10	-	26												
-	11	10	27												
-	12	11	28												
6	13	12	29												
7	14	13	30												
-	15	14	31												
8	16	15	32												

NONE (Assistant Examiner)	(Date)	Total Claims Allowed: 17	
/CYNTHIA B WILDER/ Primary Examiner.Art Unit 1637 (Primary Examiner)	6/16/2017 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure NONE

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15383965	TURNER ET AL.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

CPC- SEARCHED		
Symbol	Date	Examiner
C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333	3/6/2017	CW
updated	5/30/2017	CW
updated	6/16/2017	CW

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner
435	6.1	6/16/2017	CW

SEARCH NOTES		
Search Notes	Date	Examiner
Inventor search on PALM, review of related applictions as cited in PALM, review of IDS filed in instant invention and cited documents in copending applications, text search on WEST	3/6/2017	CW
Updated	5/30/2017	CW
updated	6/16/2017	CW

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner

INTERFERENCE SEARCH			
US Class/ CPC Symbol	**US Subclass / CPC Group**	**Date**	**Examiner**
C12Q1/6869 | C12Q2533/101 | C12Q2537/149 | C12Q2565/631 | C12Q2525/301 | G01N2021/7786 | G01N2021/6439 | G01N21/6428 | G01N21/6486 | C12N9/1252 | G06F19/22 | Y10T436/143333		3/7/2017	CW
updated		5/30/2017	CW
updated		6/16/2017	CW

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	15383965	TURNER ET AL.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☒ T.D. ☐ R.1.47

CLAIM		DATE									
Final	Original	06/16/2017									
1	1	=									
2	2	=									
-	3	-									
3	4	=									
4	5	=									
-	6	-									
-	7	-									
5	8	=									
-	9	-									
-	10	-									
-	11	-									
-	12	-									
6	13	=									
7	14	=									
-	15	-									
8	16	=									
-	17	-									
9	18	=									
-	19	-									
-	20	-									
-	21	-									
-	22	-									
-	23	-									
-	24	-									
-	25	-									
-	26	-									
10	27	=									
11	28	=									
12	29	=									
13	30	=									
14	31	=									
15	32	=									
16	33	=									
17	34	=									
-	35	-									

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail **Mail Stop ISSUE FEE**
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

57720 7590 06/23/2017
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission
I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the "Mail Stop ISSUE FEE" address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

Jacqueline L. Lim	(Depositor's name)
/Jacqueline L. Lim/	(Signature)
June 23, 2017	(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

TITLE OF INVENTION: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	UNDISCOUNTED	$960	$0	$0	$960	09/25/2017

EXAMINER	ART UNIT	CLASS-SUBCLASS
WILDER, CYNTHIA B	1637	435-006120

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).
- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
- ☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 David C. Scherer, Ph.D.
2
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: Pacific Biosciences of California, Inc.

(B) RESIDENCE: (CITY and STATE OR COUNTRY): Menlo Park, California

Please check the appropriate assignee category or categories (will not be printed on the patent): ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. The following fee(s) are submitted:
- ☒ Issue Fee
- ☐ Publication Fee (No small entity discount permitted)
- ☐ Advance Order - # of Copies _____

4b. Payment of Fee(s): **(Please first reapply any previously paid issue fee shown above)**
- ☐ A check is enclosed.
- ☐ Payment by credit card. Form PTO-2038 is attached.
- ☒ The director is hereby authorized to charge the required fee(s), any deficiency, or credits any overpayment, to Deposit Account Number 50-4427 (enclose an extra copy of this form).

5. **Change in Entity Status** (from status indicated above)
- ☐ Applicant certifying micro entity status. See 37 CFR 1.29
  NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
- ☐ Applicant asserting small entity status. See 37 CFR 1.27
  NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
- ☐ Applicant changing to regular undiscounted fee status.
  NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /David C. Scherer, Ph.D./ Date June 23, 2017

Typed or printed name David C. Scherer, Ph.D. Registration No. 56,993

PTOL-85 Part B (10-13) Approved for use through 10/31/2013. OMB 0651-0033 U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Electronic Patent Application Fee Transmittal

Application Number:	15383965
Filing Date:	19-Dec-2016
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Filer:	David Christopher Scherer/Jacqueline Lim
Attorney Docket Number:	01-007706US

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
UTILITY APPL ISSUE FEE	1501	1	960	960

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**960**

Electronic Acknowledgement Receipt	
EFS ID:	29593502
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	23-JUN-2017
Filing Date:	19-DEC-2016
Time Stamp:	15:10:35
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$960
RAM confirmation Number	062617INTEFSW00001753504427
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	01007706_2017-06-23_IssueFeeTrans.pdf	185612 d89f05ab0f01b6ad0fafe0d9fc2dc4966cef64d06	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30692 2942a43fdd5221c71288cde91f5c7338b14f31e6	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	216304		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
15/383,965	12/19/2016	1637	2400	01-007706US	18	1

CONFIRMATION NO. 8144

CORRECTED FILING RECEIPT

57770
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

OC000000092376543

Date Mailed: 06/27/2017

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Stephen Turner, Seattle, WA;
Jon Sorenson, Alameda, CA;
Kenneth Mark Maxham, Redwood City, CA;
John Eid, San Francisco, CA;
Cheryl Heiner, La Honda, CA;
Kevin Travers, Menlo Park, CA;

Applicant(s)

Pacific Biosciences of California, Inc., Menlo Park, CA;

Assignment For Published Patent Application

Pacific Biosciences of California, Inc.

Power of Attorney: The patent practitioners associated with Customer Number 57770

Domestic Priority data as claimed by applicant

This application is a CON of 14/708,603 05/11/2015 PAT 9556480
which is a CON of 14/091,961 11/27/2013 PAT 9057102
which is a CON of 12/982,029 12/30/2010 PAT 8628940
which claims benefit of 61/099,696 09/24/2008
and claims benefit of 61/139,402 12/19/2008
and is a CIP of 12/413,226 03/27/2009 PAT 8143030

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: No

Permission to Access Search Results: No

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 06/26/2017

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 15/383,965**

Projected Publication Date: Not Applicable

Non-Publication Request: No

Early Publication Request: No
Title

NUCLEIC ACID SEQUENCE ANALYSIS

Preliminary Class

435

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor

community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

57770 7590 07/05/2017
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

EXAMINER	
WILDER, CYNTHIA B	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
07/05/2017	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

PATENTS@PACB.COM

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Application No. : 15383965
Applicant : Turner
Filing Date : 12/19/2016
Date Mailed : 07/05/2017

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Notice of Allowance Mailed

This application has been accorded an Allowance Date and is being prepared for issuance. The application, however, is incomplete for the reasons below.

Applicant is given two (2) months from the mail date of this Notice within which to respond. This time period for reply is extendable under 37 CFR 1.136(a) for only TWO additional MONTHS.

The informalities requiring correction are indicated in the attachment(s). If the informality pertains to the abstract, specification (including claims) or drawings, the informality must be corrected with an amendment in compliance with 37 CFR 1.121 (or, if the application is a reissue application, 37 CFR 1.173). Such an amendment may be filed after payment of the issue fee if limited to correction of informalities noted herein. See Waiver of 37 CFR 1.312 for Documents Required by the Office of Patent Publication, 1280 Off. Gaz. Patent Office 918 (March 23, 2004). In addition, if the informality is not corrected until after payment of the issue fee, for purposes of 35 U.S.C. 154(b)(1)(iv), "all outstanding requirements" will be considered to have been satisfied when the informality has been corrected. A failure to respond within the above-identified time period will result in the application being ABANDONED.

See attachment(s).

A copy of this notice MUST be returned with the reply. Please address response to "Mail Stop Issue Fee, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450".

/Joanna Black/
Publication Branch
Office of Data Management
(571) 272-4200

Application No. 15383965

IDENTIFICATION OF APPLICATION DEFICIENCIES IN APPLICATION FILED ON OR AFTER SEPTEMBER 16, 2012

☐ Applicant must provide legible text for the following item(s).

- ☐ Specification filed , page(s) .
- ☐ Claims filed , claim(s) .
- ☐ Other: .

☐ Applicant must provide missing information on the following page(s) of the specification by amending the specification to add the missing text. No new matter may be added.
Page/line no(s).

☒ The specification refers to one or more applications by attorney docket number and does not show the U.S. application number(s). Applicant must supply the U.S. application number in place of each attorney docket number.
Page/line no(s). Paragraph 0079

☐ Applicant must provide an Abstract of the Disclosure.

☐ The Application Data Sheet (ADS dated) does not supply the inventor`s city and/or does not supply the inventor's U.S. state and/or does not supply the inventor's country. Applicant must submit a signed, in accordance with 37 CFR 1.76(e) and 1.33(b), application data sheet that corrects this deficiency. To be in compliance with 37 CFR 1.76, the corrected application data sheet must identify the information being changed by using underlining for additions and strikethroughs or brackets for deletions.

☐ Other:

I hereby certify that this correspondence is being electronically transmitted to the USPTO or deposited with the United States Postal Service as first class mail addressed to:
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

July 5, 2017

By /Jacqueline L. Lim/
Jacqueline L. Lim

PATENT
01-007706US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
Inventor: Stephen Turner, et al.	Examiner: Wilder, Cynthia
Application No.: 15/383,965	Confirmation No.: 8144
Filed: December 19, 2016	Art Unit: Not 1637
For: INTERMITTENT DETECTION DURING ANALYTICAL REACTIONS	RESPONSE TO NOTICE TO FILE CORRECTED APPLICATION PAPERS – *Notice of Allowance Mailed*

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INTRODUCTORY REMARKS

This communication is filed in response to the Notice to File Corrected Application Papers-*Notice of Allowance Mailed*, dated July 5, 2017. Applicant respectfully requests entry of the amendment and remarks provided herein.

Amendment to the Specification begins on page **2** of this paper.

Remarks/Arguments begin at page **4** of this paper.

AMENDMENTS TO THE SPECIFICATION

Please replace paragraph **[0079]** with the corresponding amended paragraph below. No new matter has been added.

[0079] In certain aspects, the methods herein provide various strategies for achieving intermittent illumination of illuminated reactions. Essentially, at least one type of illumination (e.g., excitation illumination) is present for at least one time period ("illuminated period") and absent during at least one other time period ("non-illuminated period") during an illuminated reaction. As described above, the term "non-illuminated" indicates a change in illumination including, but not limited to a complete absence of illumination. For example, a non-illuminated period may also be characterized by a different illumination source or intensity than an illuminated period, or by a change in reaction components, e.g., detectable labels. In general, at least one type of data collected during an illuminated period (e.g., nucleotide sequence data) is not collected during a non-illuminated period. An absence of the illumination may be due to, e.g., inactivation of the illumination source (e.g., laser, laser diode, a light-emitting diode (LED), a ultra-violet light bulb, and/or a white light source), removal of the illuminated reaction from the illumination source (or vice versa), or may be due to blockage of the illumination from the reaction, as discussed below. Modifications to the illumination may be due to, e.g., adjustment of the intensity of an illumination source, or a substitution of one illumination wavelength and/or frequency for another. Further, components detectable during an illuminated period may be removed from the reaction mixture during a non-illuminated period, e.g., a fluorescently labeled nucleotide may be replaced with an unlabeled nucleotide. Knowledge of the rate of the reaction and the time during which the illumination is absent is used to estimate the progress of the reaction during the non-illuminated period. For example, if a reaction proceeds such that one molecule is incorporated into a macromolecule per second, and the illumination is absent for 20 seconds, it can be estimated that 20 molecules were incorporated during the non-

illuminated period. This information is useful during data analysis to provide context for the reaction data collected during the illuminated period(s). For example, in a sequencing-by-incorporation reaction the number of base positions separating sequence reads generated in illuminated periods can be estimated based on the temporal length of intervening non-illuminated periods and the known rate of incorporation during the reaction and/or by the measured rate of incorporation during the illuminated period(s). The known rate of incorporation can be based on various factors including, but not limited to, sequence context effects due to the nucleotide sequence of the template nucleic acid, kinetics of the polymerase used, buffer effects (salt concentration, pH, etc.), and even data being collected from an ongoing reaction. Further the processivity of an enzyme during a non-illuminated period (or other type of non-detection period) can be manipulated or adjusted by methods known to those of skill in the art. In particular, the kinetics of replication by a polymerase enzyme can be altered by changing the chemical environment in which it operates, and such methods are further described, e.g., in U.S. Patent Application Nos. 12/414,191, filed March 30, 2009; 12/537,130, filed August 6, 2009; and U.S. Patent Application No. ~~[unassigned], attorney docket no. 105-006301US~~ 12/584,481, entitled "Engineering Polymerases and Reaction Conditions for Modified Incorporation Properties," filed September 4, 2009, the disclosures of all of which are incorporated herein by reference in their entireties for all purposes. For example, methods are provided for adjusting the enzyme activity, and these methods find particular relevance in the instant invention when used to enhance accuracy during detection periods, and to enhance processivity during non-detection periods. Information regarding enzyme translocation rate and processivity is useful for positioning the sequence reads for a single template nucleic acid relative to one another in the construction of a sequence scaffold and/or consensus sequence for the template nucleic acid.

REMARKS

In response to the Notice to File Corrected Application Papers dated July 5, 2017, Applicants have amended paragraph [0079] of the specification to add the missing application number. Applicants believe this amendment addresses all issues cited in the Notice.

Based upon the foregoing remarks, Applicants believe the instant application is in condition for allowance and action toward that end is respectfully requested. If the Office believes there are additional issues, the Office is encouraged to contact Applicant's undersigned representative at (650) 521-8127.

Respectfully submitted,

July 5, 2017
Date

/David C. Scherer, Ph.D./
David C. Scherer, Ph.D.
Reg. No.: 56,993

PACIFIC BIOSCIENCES OF
CALIFORNIA, INC.
1305 O'Brien Drive
Menlo Park, CA 9402

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

57770 7590 07/05/2017
Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

EXAMINER	
WILDER, CYNTHIA B	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
07/05/2017	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

PATENTS@PACB.COM

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Application No. : 15383965
Applicant : Turner
Filing Date : 12/19/2016
Date Mailed : 07/05/2017

NOTICE TO FILE CORRECTED APPLICATION PAPERS

Notice of Allowance Mailed

This application has been accorded an Allowance Date and is being prepared for issuance. The application, however, is incomplete for the reasons below.

Applicant is given two (2) months from the mail date of this Notice within which to respond. This time period for reply is extendable under 37 CFR 1.136(a) for only TWO additional MONTHS.

The informalities requiring correction are indicated in the attachment(s). If the informality pertains to the abstract, specification (including claims) or drawings, the informality must be corrected with an amendment in compliance with 37 CFR 1.121 (or, if the application is a reissue application, 37 CFR 1.173). Such an amendment may be filed after payment of the issue fee if limited to correction of informalities noted herein. See Waiver of 37 CFR 1.312 for Documents Required by the Office of Patent Publication, 1280 Off. Gaz. Patent Office 918 (March 23, 2004). In addition, if the informality is not corrected until after payment of the issue fee, for purposes of 35 U.S.C. 154(b)(1)(iv), "all outstanding requirements" will be considered to have been satisfied when the informality has been corrected. A failure to respond within the above-identified time period will result in the application being ABANDONED.

See attachment(s).

A copy of this notice MUST be returned with the reply. Please address response to "Mail Stop Issue Fee, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450".

/Joanna Black/
Publication Branch
Office of Data Management
(571) 272-4200

Application No. 15383965

IDENTIFICATION OF APPLICATION DEFICIENCIES IN APPLICATION FILED ON OR AFTER SEPTEMBER 16, 2012

☐ Applicant must provide legible text for the following item(s).

- ☐ Specification filed , page(s) .
- ☐ Claims filed , claim(s) .
- ☐ Other: .

☐ Applicant must provide missing information on the following page(s) of the specification by amending the specification to add the missing text. No new matter may be added.
Page/line no(s).

☒ The specification refers to one or more applications by attorney docket number and does not show the U.S. application number(s). Applicant must supply the U.S. application number in place of each attorney docket number.
Page/line no(s). Paragraph 0079

☐ Applicant must provide an Abstract of the Disclosure.

☐ The Application Data Sheet (ADS dated) does not supply the inventor`s city and/or does not supply the inventor's U.S. state and/or does not supply the inventor's country. Applicant must submit a signed, in accordance with 37 CFR 1.76(e) and 1.33(b), application data sheet that corrects this deficiency. To be in compliance with 37 CFR 1.76, the corrected application data sheet must identify the information being changed by using underlining for additions and strikethroughs or brackets for deletions.

☐ Other:

Electronic Acknowledgement Receipt

EFS ID:	29690459
Application Number:	15383965
International Application Number:	
Confirmation Number:	8144
Title of Invention:	NUCLEIC ACID SEQUENCE ANALYSIS
First Named Inventor/Applicant Name:	Stephen Turner
Customer Number:	57770
Filer:	David Christopher Scherer/Jacqueline Lim
Filer Authorized By:	David Christopher Scherer
Attorney Docket Number:	01-007706US
Receipt Date:	05-JUL-2017
Filing Date:	19-DEC-2016
Time Stamp:	15:12:05
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		01007706_2017-07-05_RespNoticeFileCorrAppPapers.pdf	130563 4366fadb3ea217a3e3ede38a6da19fa6f571f733	yes	7

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Amendment after Notice of Allowance (Rule 312)	1	1
	Specification	2	3
	Applicant Arguments/Remarks Made in an Amendment	4	7
Warnings:			
Information:			
	Total Files Size (in bytes):	130563	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	12/19/2016	Stephen Turner	01-007706US	8144

57770 7590 07/10/2017

Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

EXAMINER	
WILDER, CYNTHIA B	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
07/10/2017	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

PATENTS@PACB.COM

Response to Rule 312 Communication	Application No. 15/383,965	Applicant(s)
	Examiner	Art Unit

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

1. ☒ The amendment filed on *05 July 2017* under 37 CFR 1.312 has been considered, and has been:

a) ☒ entered.

b) ☐ entered as directed to matters of form not affecting the scope of the invention.

c) ☐ disapproved because the amendment was filed after the payment of the issue fee.
Any amendment filed after the date the issue fee is paid must be accompanied by a petition under 37 CFR 1.313(c)(1) and the required fee to withdraw the application from issue.

d) ☐ disapproved. See explanation below.

e) ☐ entered in part. See explanation below.

Charles Bowen

Publishing Division

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	2016-12-19
First Named Inventor	Stephen Turner
Art Unit	
Examiner Name	Not Yet Assigned
Attorney Docket Number	01-007706US

	20	6917726		2005-07-12	Levene et al.	
	21	7013054		2006-03-14	Levene et al.	
	22	7033764		2006-04-25	Korlach et al.	
	23	7045362	B2	2006-05-16	Hartwich et al.	
	24	7052847		2006-05-30	Korlach et al.	
	25	7056661		2006-06-06	Korlach et al.	
	26	7056676		2006-06-06	Korlach et al.	
Change(s) applied to document. /R.F./ 6/28/2017	27	7170050		01-2007 ~~2006-06-06~~	Korlach et al.	
	28	7181122		2007-02-20	Levene et al.	
	29	7229799		2007-06-12	Williams et al.	
	30	7282337		2007-10-16	Harris et al.	

EFS Web 2.1.17

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	2016-12-19
	First Named Inventor	Stephen Turner
	Art Unit	
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	01-007706US

	30	20090087850		2009-04-02	Eid et al.	
	31	20090233291	A1	2009-09-17	Chen et al.	
	32	20080218184 ~~20091218184~~	A1	2008-09-11	White et al.	

Change(s) applied to document.
/RKC/
7/22/2017

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	199106678	WO	A1	1991-05-16	SRI International		☐
	2	199416090	WO	A1	1994-07-21	Molecular Tool, Inc.		☐
	3	199627025	WO	A1	1996-09-06	Rabani		☐
	4	199905315	WO	A2	1999-02-04	Medical Biosystems Ltd.		☐
	5	2007003017	WO	A1	2007-01-11	Biochip Innovations PTY		☐
	6	2007070572	WO		2007-06-21	US Govt as represented by DHHS		☐

/CYNTHIA B WILDER/ (03/06/2017)

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
15/383,965	08/22/2017	9738929	01-007706US	8144

57770 7590 08/02/2017

Pacific Biosciences of California, Inc.
1305 O' Brien Drive
MENLO PARK, CA 94025

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 0 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Data Management (ODM) at (571)-272-4200.

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

Stephen Turner, Seattle, WA;
Pacific Biosciences of California, Inc., Menlo Park, CA;
Jon Sorenson, Alameda, CA;
Kenneth Mark Maxham, Redwood City, CA;
John Eid, San Francisco, CA;
Cheryl Heiner, La Honda, CA;
Kevin Travers, Menlo Park, CA;

AO 120 (Rev. 08/10)

TO:	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court District of Delaware on the following

☐ Trademarks or ☑ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
	9/25/2017	District of Delaware

PLAINTIFF	DEFENDANT
Pacific Biosciences of California, Inc.	Oxford Nanopore Technologies, Inc.

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	US 9,678,056 B2	6/13/2017	Pacific Biosciences of California, Inc.
2	US 9,738,929 B2	8/22/2017	Pacific Biosciences of California, Inc.
3			
4			
5			

In the above entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy