and larvae), vertebrates (cells and transgenic animals), and plants (cells and transgenic plants). Each of these systems has unique advantages. Yeasts and fungi have robust growth characteristics like bacteria. They are also readily amenable to genetic manipulation. Furthermore, they can perform many posttranslational modifications found in higher eukaryotes. The purpose of this article is to highlight the strategies that have been used for optimal protein expression with the *Pichia* yeast expression system, which in the past few years has turned out to be versatile and impressive yeast for production of proteins.

## BACKGROUND

*Pichia pastoris* is a methylotrophic yeast. It is able to use methanol as sole carbon source for energy as well as for growth (1,2). It was originally developed by Phillips Petroleum Company (Bartlesville, Oklahoma) as an organism of choice for bioconversion of natural gas (methane) into food (single cell protein). This may sound ironic, because nowadays every one is talking bioconversion of food crops to fuels (biofuels). Phillips Petroleum Company in the 1970s developed an efficient *Pichia* fermentation process [cell density >130 g dry cell weight per liter and biomass productivity >10 g/(L h)] (3,4). Though impressive, it could not compete with the economics of production of soy protein. After this set back, Phillips in the early 1980s directed its future course with *Pichia* into two areas:

1. *Speciality Food or Feed Application:* A rather impressive 100,000 L fermentation plant to churn out tons of *Pichia* for potential speciality food applications was completed in 1988. However, by 1993, the idea to use *Pichia* for speciality food/feed was abandoned as *Pichia* did not have generally regarded as safe (GRAS) status and Phillips Petroleum Company decided to focus on its core business of oil exploration, production, and petrochemicals.

become available for *Saccharomyces cerevisiae*, such as yeast–*E.coli* shuttle vectors, transformation protocols, and site-specific integrative transformation (8,9) were found to be readily applicable to *Pichia*. With the advent of cloning of *AOX1* promoter (10,11), availability of auxotrophic *Pichia* strain [GS115 (*his4*) developed by George Sperl] and *Pichia* transformation protocols (12,13), successful high level expression of several proteins was readily demonstrated. They include, intracellular expression of Hepatitis B surface antigen particles (14), *E. coli* $\beta$-galactosidase (LACZ) (15), human tumor necrosis factor (TNF) (16), as well as greater than gram per liter level of secretion of *S. cervesiae* invertase (SUC2) (17), and human serum albumin (HSA) (18). *Pichia* expression technology patent was granted to Phillips Petroleum Company in 1988 (11) and thus, by the late 1980s, *Pichia* expression system was well on its way to leave the nest.

In the fall of 1988, Phillips Petroleum Company made a conscientious decision to license the system to other companies. Fortunately for Phillips licensing team (Katherine Bartosh, L.V. Benningfield, Mary Jane Hagenson, Jack Phillips, and Koti Sreekrishna), their very first commercial licensee (Glaxo–Wellcome) with just 2 days of training at Phillips produced some impressive expression results (19–23). Phillips was able to license the technology to 20 companies in 4 years. Along side, it also started distributing the *Pichia* expression kit, free of cost to any academic institution that approached the company for the kit. This was getting out of hand, and thus the company transferred distribution rights, free of cost, to Invitrogen Corporation (Carlsbad, California) in 1993. Since then, Invitrogen Corporation (http://www.invitrogen.com) has been aggressively distributing *Pichia* expression kit at reasonable cost and with some user friendly modifications. They are largely responsible for wide usage of *Pichia* technology. In the same year (1993), Phillips sold *Pichia* technology to Research Technology Corporation (Tucson,

*Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*, edited by Michael C. Flickinger

at Phillips Petroleum Company (18) and transferred to Green-Cross Corporation (Osaka, Japan) has been further developed and scaled up to produce one million dosage vials (12.5 g per vial) of authentic rHSA/year by Mitsubishi Pharma Corporation (Osaka, Japan) (26,27). This truly is a testimony to the high expression level and scale-up possible with *Pichia* (25). (Please refer to http://www.rctech.com for a more complete list of *Pichia*-produced products under various stages of commercialization.)

## STRATEGIES FOR OPTIMIZATION OF PROTEIN EXPRESSION

The typical *Pichia* expression vectors are all yeast–*E. coli* shuttle plasmids (see "Glossary of *P. pastoris* Vectors"). The most commonly used vectors are based on *AOX1* promoter (10,11) (see section titled "Alternative Promoters for Expression" below for other promoters available for *Pichia* expression). Numerous selectable marker genes available include *HIS4* ($His^+$ selection) (12), *ARG4* ($Arg^+$ selection) (28), *ADE1* ($Ade^+$ selection) (29), and *URA3* and *URA5* ($Ura^+$ selection) (29,30), which can be used in conjunction with the appropriate *Pichia* auxotrophic strains. Several dominant selection markers that can be used for transformation of any *P. pastoris* strain are also available. These include *SUC2* (allows growth on sucrose) (13), $Kan^R$ (G418/Geneticin selection) (23,31), $Zeo^R$ (Zeocin selection) (32), $Bsd^R$ (Blasticidin selection) (http://www.invitrogen.com), FLD1 (Formaldehyde selection) (33), and $Sor^R$ (Soraphen A selection) (34).

A wide assortment of proteins have been produced in *Pichia*, which include human insulin (35), glucagon-like peptide (GLP) (36), G-protein coupled receptors (GPCRs) (37), Aquaporin (AQP1) (38), as well as correctly assembled human collagen fibers (39) and fibrinogen chains (40). These are among >500 proteins that have expressed in this system (24). Obviously, not all proteins are expressed at multiple grams per liter range. The expression level is largely influenced by inherent properties of a protein. Some proteins are readily expressed at high levels with minimal manipulation, while some proteins barely reach milligram levels, despite tremendous effort. Encouragingly, in many instances, the initial production yield of a protein can be dramatically enhanced by addressing the

Utilization Phenotype of the Host: $Mut^+$ or $Mut^s$").

Plasmids based on autonomous replication sequence (*ARS*) such as pHIL-A1 (see "Glossary of *P. pastoris* Vectors") although can be introduced into *Pichia* cells with a high transformation frequency, they are rapidly lost from the population of dividing cells, and eventually integrate at one or more of the homologous sites on the chromosome (12,13). Owing to their ability of transform *Pichia* at high frequency ($>10^5\ \mu g^{-1}$) and ease of plasmid rescue the autonomous plasmids are useful for cloning genes in *Pichia* by functional complementation.

### Site of Integration of the Expression Cassette

The *AOX1* promoter used in *Pichia* expression vectors is active irrespective of the site of integration (*AOX1*, *HIS4* *ARG4*, *ADE1*, or *URA3* loci). However, AOX1 locus is the preferred site of integration for stable expression, because integration at the other loci can result in loss of the expression cassette due to intrachromosmal cross over between the mutant and good copy of the gene (unpublished observations). If the vector also contains a dominant selection marker (discussed earlier), then selection pressure can be applied for stable maintenance of expression cassette at those sites.

### Methanol Utilization Phenotype of the Host: $Mut^+$ or $Mut^s$

Transformation of a *P. pastoris his4* strain (GS115) using linear DNA expression cassette with the ends bearing homology to the 5′ and 3′ regions of the *AOX1* chromosomal locus results in the site-specific eviction of the *AOX1* structural gene at a high frequency (5–20% of the $His^+$ transformants) (16,28). Such clones can be readily distinguished by replica-plating the colonies from the initial $His^+$ selection plate on to a minimal methanol (MM) medium plate. The clones that have undergone eviction of *AOX1* grow slower ($Mut^S$) compared to $Mut^+$ clones with an intact AOX1 gene, because, in such clones, growth on methanol is dependent on the alcohol oxidase encoded by *AOX2*, which is expressed at much lower level (weaker promoter) (44). The $Mut^+$ clones arise due to circularization of the linear DNA expression cassette inside the yeast cell prior to integration. Thus, both $Mut^S$ and $Mut^+$ clones result in the same experiment.

Early on a dogma prevailed among the researchers of the *Pichia* system, based on limited experience (14,15), that increasing the gene dosage did not impact expression because of the remarkable strength of the *AOX1* promoter. This dogma was shattered on the Good Friday of March 28, 1986, with my observation of dramatic clonal variation (<1% to >30%) in the expression of human TNF, that was subsequently attributed to gene dosage. Furthermore, the high copy number was stably maintained even after several days of growth in high cell density fermentor. The observation with TNF expression was readily exploited for expression of tetanus toxin fragment C (21), *Bordetella pertussus* pertactin P69 (20), and mouse epidermal growth factor (EGF) (19), and has been a key strategy for successful expression of hundreds of proteins in the *Pichia* system. However, in numerous cases, a single copy of the expression cassette is sufficient and deliberately increasing the copy number had no significant effect on the production (14,15,17,18). In some rare instances, an increase in copy number has a deleterious effect on the production level as has been noted with secretion of human insulin-like growth factor (IGF) (46) and *Necator americanus* (hookworm) secretory protein (Na-ASP1) (47).

Because the effect of gene copy number on expression is unpredictable, it is prudent to examine the production level as a function of gene dosage. The spheroplast method of transformation of *Pichia* results in transformants with a wide range of copy numbers (16,48). Evaluation of as few as 100 individual clones for protein production is generally adequate to arrive at a good producer. Though bit laborious, spheroplast method of transformation yields clones with wide range of gene dosage at a high frequency (48). If other methods of transformation such as using LiCl or electroporation, which do not yield high frequency of multicopy clones, then one can use more efficient screens. These include use of colony hybridization with DNA probes or using appropriate vectors that would allow selection based on increased level of resistance to one of G418/geneticin (23,31), zeocin (32), formaldehyde (33), blasticidin (http://www.invitrogen.com), or soraphen A (34). Increased drug resistance does not automatically ensure multicopy integration and several dozen resistant colonies must be analyzed for copy number and expression.

Another approach to identify high producers may be to use a visual tag to the protein being expressed. In one

tag outside the interfering range, one may be able to reliably monitor simultaneously both cell and product yield on-line.

Vectors such as pAO815 (54,55) have also been described that would allow *in vitro* construction of expression cassette concatamers. This approach is useful to accurately correlate copy number to expression level over a narrow range of gene dosage.

### Translational Optimization: 5′ Untranslated Region

The nucleotide sequence and the length of the 5′ untranslated region (5′UTR) are detrimental to optimal protein translation. The leader length of the highly expressed *AOX1* mRNA is 114 nucleotides long, and the sequence is A + U rich (10,11). For optimal synthesis of heterologous proteins, it is essential that the 5′UTR should closely resemble that of the *AOX1* mRNA. Ideally, it is preferable to make it identical to that of *AOX1* mRNA. The expression level of HSA is increased >50 fold by optimizing the 5′UTR to mimic that of *AOX1* mRNA (18). Expression plasmids such as pHIL-D7 (42) can be used to make an exact construct. This plasmid has unique *Asu* II and *Eco* RI sites immediately following 5′*AOX1*. The second *Asu* II site that was originally present in 3′*AOX1* has been eliminated. Therefore, the sequence TTCGAAACG can be added immediately upstream of the ATG start codon of the gene of interest, and an *Eco* R I site can be engineered downstream of the stop codon for insertion at *Asu* II–*Eco* RI sites of pHIL-D7 (42).

### Transcriptional Optimization

Genes with high A + T nucleotide clusters are poorly transcribed in *Pichia* due to premature termination of the transcription. For example, ATTATTTTATAAA, present in HIV-gp120 has been identified to block transcription in *P. pastoris*, and the premature termination is overcome by altering the sequence to TTTCTTCTACAAG (22). Because we are not aware of all the problematic A + T rich clusters, a general strategy with A + T rich genes is to redesign them using *P. pastoris* preferred codons (48,54), http://www.kazusa.or.jp/codon/) so as to have an A + T content in the range of 30–55%. By using this approach, it has been possible to construct *Pichia* strains for efficient production of several proteins, which include tetanus toxin levels in the intracellular compartment (16). Interestingly, the intracellular human tetrameric catalase has successfully been expressed as a secreted protein in *Pichia* (60).

A wide variety of heterologous proteins have been secreted in *Pichia*. In several instances, HSA (18), invertase (17), bovine lysozyme (61), barley alpha amylases (62), cathepsin E (63), and thaumatin (64) the native signal sequence is adequate. In the case of matrix metalloproteinases, although native signal sequences work (41), both secretion and product yield are improved while using the *S. cerevisiae* pre-proalpha mating factor (αMF) secretion signal sequence (65,66). Likewise, with *Candida rugosa* lipase 1, although native signal works, both product stability and yield are tremendously improved with the use of pre-proαMF signal sequence (67). For thaumatin secretion, native signal works, but not the pre-proαMF signal sequence (64), whereas for human interferon-alpha 2b (IFN-a2b), native signal sequence does not work, but pre-proαMF works (68).

Pre-proαMF secretion signal sequence has worked well for secretion of large variety of proteins, including the smaller-sized products such as aprotinin (69), EGF (19,69), IGF-1 (70), and ghilanten (56). The processing of pre-proαMF secretion signal involves three steps. The first is the elimination of the pre-region by signal peptidase, second, KEX2, and YPS proteinases cleave out the proregion (71,72), and finally, N-terminal Glu–Ala repeats are removed by the action of dipeptidyl-aminopeptidase (Dpap). The efficiency of each processing step depends on amino acid sequence adjacent to the processing site as well as the tertiary structure of the secreted protein. All these factors contribute to incomplete processing and/or reduced yield of mature protein (73–75).

In making αMF signal sequence constructs, it is generally preferable to retain the Glu-Ala spacers adjacent to the Kex2-like protease cleavage site ( ... Val-Ser-Ser-Leu-Glu-Lys-Arg-$^{\mathrm{Kex2p}}$Glu-Ala-$^{\mathrm{Dpap}}$-Glu-Ala-$^{\mathrm{Dpap}}$-fused protein). The presence of Glu-Ala spacers help to alleviate the steric interference imposed by the fused protein, resulting in an efficient cleavage of the prosequence by the *Pichia* Kex2 like protease (69). The Glu-Ala spacer is subsequently cleaved by diamino peptidase (Dpap coded by *STE13*) to yield the protein of interest free of additional N-terminal amino acid

Only in a limited number of cases has a thorough comparison been made on the relative efficacies of different signal sequences and variations thereof (for example (59,64,68,69)). In the case of invertase secretion, both the extent of glycosylation and secretion rate are enhanced when the native signal sequence is substituted with the pre-proαMF signal sequence (69), although both yields are greater than gram per liter level of the product. However in the case of proteins that are more susceptible to proteolysis, improvement in secretion rate can increase production level as has been noticed with MMP-1 (41,65,66) Based on the relatively high success of pre-proαMF signal sequence, it makes sense to test that signal (and variations there of) first, side by side with the native secretion signal sequence, before exploring other signal sequences.

### Production Enhancement by Manipulation of Media and Growth Conditions

More often than not, a secreted protein is rapidly degraded in the *Pichia* medium due to extracellular proteases cell-bound proteases, as well as by intracellular proteases released to medium due to cell death and cell lysis Multiple stress factors (especially in high cell density fermentor), which include starvation, pH shift, temperature change, change of carbon source, buildup of toxins, and reactive oxygen species (ROS) are considered to cause cell death and cell lysis (77).

One of the approaches to increase the stability and yield of the secreted protein in the culture medium is by manipulating the pH of the medium to arrive at the optimal pH for blocking a problem protease. Suggested pH range for experimentation is between 2.8 and 8 (see section titled “Media Compositions”). This pH range does not affect the growth significantly. HSA yield was significantly improved by raising the pH from 5.2 to 6 with adequate aeration. The yield was further enhanced by the addition of yeast extract (1%) and peptone (2%) (45,48,78,79). Production of mouse EGF was favored at pH 6 in the presence of casamino acids (19). Casamino acids is preferable to yeast extract + peptone, because the peptide components of peptone (such as collagen fragments) can interfere in product analysis and recovery (41). For both IGF-I (70) and cytokine growth-blocking peptide (80), pH 3 was found to be optimal. Greater than twofold increase

Lower cultivation temperature is shown to improve product yield in several instances, which include CBM-CABL (22°C) (81), laccase (20°C) (85), galactose oxidase (25°C) (86), Rh midkine (20°C) (87), human AFP (23°C) (88), herring proteins (23°C) (89), and human bile salt activated lipase (BSSL, 20°C) (90). *Pichia* cultivated at temperatures as low as 15°C, greatly enhanced the yield of ScFV production due to reduced protease level (91). Secretion level of bivalent T cell immunotoxin A-dmDT390-bisFy($G_4S$) was also enhanced at 15°C, which was attributed to reduced protease levels, improved folding, as well as reduced toxicity of the immunotoxin to *Pichia* (92). Thus, we may say that the low temperature expression helps to increase the yield of aggregation-prone, unstable, and toxic products in *Pichia*.

Addition of 5 m*M* EDTA to the medium also improves accumulation of proteins expressed in *Pichia* (41). Supplementation of the induction medium with 0.4 *M* L-arginine, 5 m*M* EDTA, or 2% casamino acids in the BMMY induction medium (see section titled "Media Compositions") increased scFV expression (91). It should be noted that medium manipulation can significantly alter the profile of endogenous protein components in the culture media, such that previously unnoticed proteins accumulate at high levels. For example, we have noticed that addition of 5 m*M* EDTA causes accumulation of a protein of approximately 50 kDa in the extracellular medium (41). This protein has the aa sequence DIIWDYSSEKIMGVNLGGWL…, which matches closely with the exo-$\beta$-1,3-glucanase of *S. cerevisiae* (93) and *Candida albicans* (94). We have also noticed that overexpression of human tissue inhibitor of matrix metalloproteinase (TIMP1) in *Pichia* leads to greater gram per liter accumulation of a 18–22 kDa *Pichia* protein of sequence ADYMC$^?$GLAIYGAWEC$^?$ GPEAGPFDSEC$^?$LLATD (41).

## Production Enhancement Using Protease-Deficient Strains

In addition to the optimization of media and growth conditions, the product yield can be further improved by using a protease-deficient *Pichia* strains generated by knocking out a *Pichia* protease or by overexpressing a protease inhibitor (for example, TIMP1 strain as noted above).

Protease-deficient *Pichia* strains SMD1168 (*his4*, *pep4*), SMD1165 (*his4*, *prb1*), and SMD1163 (*his4*, *pep4*, *prb1*)

codes for a carboxy peptidase, resulted in expression of full-length endostatin in *Pichia* (97). Recently, using a gene pop-in/pop-out gene replacement approach (98), the *KEX1* was deleted from a *Pichia* hirudin production strain. This resulted in most significant improvement in intact Hir65 production, which approached 2.4 g/L for the *KEX1* deleted strain compared to 1.1 g/L seen with the strain without *KEX1* deletion (99).

## Enhancement of Protein Secretion by Overexpression of Chaperone Proteins

*Pichia* is able to perform many posttranslational modifications found in higher eukaryotes, which include correct folding, disulfide bond formation. Folding and disulfide bond formation in some cases can be the rate-limiting step in protein expression leading to protein aggregation (100). It is not clear which single chaperone is most important or which combination optimally cooperates in this process.

Overexpression of *Pichia* protein disulfide isomerase (PDI), which is important for protein folding in the endoplasmic reticulum (ER), was able to increase the secretion of Na-ASP1 protein in high copy clones (47). As noted before, high copy clones of Na-ASP1 secreted less material than single-copy constructs, perhaps due to overburdening the *Pichia* secretory/protein folding machinery, which was corrected by PDI overexpression.

A33scFV in *Pichia* is expressed at 4 g/L level, which rose to >10 g/L by overexpression of immunoglobulin binding protein (Bip) (101). The noted impressive increase is attributed to increase in folding capacity. PID overexpression did not have any effect on A33scFV expression. This was unexpected because A33scFV contains disulfide bonds. Furthermore, simultaneous overexpression of both BiP and PDF also did not have any effect in this system. It was also noted that PDI expression in the A33scFV strain caused a six-fold increase in endogenous BiP expression, suggesting that PDI was inducing an unfolded protein response due to excess chaperone and recombinant protein in the ER. In another study it was found that the chaperone combinations YDJiP/PDI, YDJiP/Sec63, and Kar2p/PDI synergistically increase secretion levels 8.7, 7.6, and 6.5 times, respectively (102).

A transcriptomics-based approach to identify novel factors enhancing heterologous protein secretion by analyzing

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you’re on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm’s cloud-native docket research platform finds what other services can’t. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

### LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

### FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

### E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

**WHAT WILL YOU BUILD?** | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.