OPEN ACCESS Freely available online
PLoS COMPUTATIONAL BIOLOGY

# Structural Consensus among Antibodies Defines the Antigen Binding Site

**Vered Kunik[1], Bjoern Peters[2], Yanay Ofran[1]***

**1** The Goodman faculty of life sciences, Nanotechnology building, Bar Ilan University, Ramat Gan, Israel, **2** Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, California, United States of America

## Abstract

The Complementarity Determining Regions (CDRs) of antibodies are assumed to account for the antigen recognition and binding and thus to contain also the antigen binding site. CDRs are typically discerned by searching for regions that are most different, in sequence or in structure, between different antibodies. Here, we show that ~20% of the antibody residues that actually bind the antigen fall outside the CDRs. However, virtually all antigen binding residues lie in regions of structural consensus across antibodies. Furthermore, we show that these regions of structural consensus which cover the antigen binding site are identifiable from the sequence of the antibody. Analyzing the predicted contribution of antigen binding residues to the stability of the antibody-antigen complex, we show that residues that fall outside of the traditionally defined CDRs are at least as important to antigen binding as residues within the CDRs, and in some cases, they are even more important energetically. Furthermore, antigen binding residues that fall outside of the structural consensus regions but within traditionally defined CDRs show a marginal energetic contribution to antigen binding. These findings allow for systematic and comprehensive identification of antigen binding sites, which can improve the understanding of antigenic interactions and may be useful in antibody engineering and B-cell epitope identification.

**Citation:** Kunik V, Peters B, Ofran Y (2012) Structural Consensus among Antibodies Defines the Antigen Binding Site. PLoS Comput Biol 8(2): e1002388. doi:10.1371/journal.pcbi.1002388

**Editor:** Brian Baker, University of Notre Dame, United States of America

**Received** August 1, 2011; **Accepted** December 30, 2011; **Published** February 23, 2012



**Funding:** This work was supported by NIAID contract N01-AI-900048C (http://www.niaid.nih.gov/) and by the Israeli Science Foundation, grants No. 511/10(www.isf.org.il). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: Yanay@ofranlab.org

## Introduction

Antibody-Antigen (Ab-Ag) interactions are based on non-covalent binding between the antibody (Ab) and the antigen (Ag). Correct identification of the residues that mediate Ag recognition and binding would improve our understanding of antigenic interactions and may permit the modification and manipulation of Abs. For example, introducing mutations into the V-genes has been suggested as a way to improve Ab affinity [1–3]. However, mutations in the framework regions (FRs) rather than in the Ag binding residues themselves are more likely to evoke an undesired immune response [4]. Knowing which residues bind the Ag can help direct such mutations and be beneficial to Ab engineering [5–7]. It has been shown that Ag binding residues are primarily located in the so called complementarity determining regions (CDRs) [7–9]. Thus, the attempt to identify CDRs, and particularly the attempt to define their boundaries, has become the focus of extensive research over the last few decades [7,8,10]. Kabat and co-workers [9,11] attempted to systematically identify CDRs in newly sequenced Abs. Their approach was based on the assumption that CDRs include the most variable positions in Abs and therefore could be identified by aligning the fairly limited number of Abs available then. Based on this alignment they introduced a numbering scheme for the residues in the hypervariable regions and determined which positions mark the beginning and the end of each CDR. The Kabat numbering scheme was developed when no structural information was available. Chothia et al. [12,13] analyzed a small number of Ab structures and determined the relationship between the sequences of the Abs and the structures of their CDRs. The boundaries of the FRs and the CDRs were determined and the latter have been shown to adopt a restricted set of conformations based on the presence of certain residues at key positions in the CDRs and the flanking FRs. This analysis suggested that the sites of insertions and deletions in CDRs L1 and H1 are different than those suggested by Kabat. Thus, the Chothia numbering scheme is almost identical to the Kabat scheme, but based on structural considerations, places the insertions in CDRs L1 and H1 at different positions. As more experimental data became available, the analysis was performed anew, re-defining the boundaries of the CDRs. These definitions of CDRs are mostly based on manual analysis and may require adjustments as the structure of more Abs become available. Abhinandan et al. [14] aligned Ab sequences in the context of structure and found that approximately 10% of the sequences in the manually annotated Kabat database have erroneous numbering. A more recent attempt to define CDRs is that of the IMGT database [15] which curates nucleotide sequence information for immunoglobulins (IG), T-cell receptors (TcR) and Major Histocompatibility Complex (MHC) molecules. It proposes a uniform numbering system for IG and TcR sequences, based on aligning more than 5000 IG and TcR variable region sequences, taking into account and combining the Kabat definition of FRs and CDRs [16], structural data [17] and Chothia's characterization of the hypervariable loops [12]. Their

**Author Summary**

Antibodies are a primary adaptive defence mechanism against infection, and function by recognizing and binding to non-self antigens. While most of the sequence of all antibodies of a given individual is identical, relatively small variations turn each antibody into a specific binder of one antigen. It is widely assumed that antigen binding sites correspond to the so called Complementarity Determining Regions (CDRs) of the antibody, which are defined as the elements that are most different between antibodies. We analysed all known antibody-antigen complexes and found that about 20% of the residues that actually bind the antigen fall outside the CDRs. However, we also found that virtually all antigen binding residues fall within regions of structural consensus between antibodies. Moreover, we demonstrate that antigen binding residues that reside within these structural consensus regions but outside of the traditionally-defined CDRs make significant energetic contribution to antigen binding. Furthermore, we show that these regions are organized along the sequence of the antibody chains and are identifiable from the sequence of the antibody.

numbering scheme does not differentiate between the various immunoglobulins (i.e., IG or TcR), the chain type (i.e., heavy or light) or the species.

A drawback of these numbering schemes is that CDRs length variability is accommodated with either annotation of insertion (Kabat and Chothia) or by providing excess numbers (IMGT). Abs with unusually long insertions may be hard to annotate this way, and therefore their CDRs may not be identified correctly. Honegger and Pluckthun [18] suggested a structurally improved version of the IMGT scheme. Instead of introducing unidirectional insertions and deletions as in the IMGT and Chothia schemes, they were placed symmetrically around a key position. MacCallum et al. [8] have proposed focusing on the specific notion of Ag binding residues rather than the more vague concept of CDRs. They suggested that these residues could be identified based on structural analysis of the binding patterns of canonical loops. Other studies have dubbed those Ag binding residues Specificity Determining Regions (SDRs) [5,7]. Here, we analyze Ag-Ab complexes and show that virtually all Ag binding residues fall within regions of structural consensus. We refer to these regions as Ag Binding Regions (ABRs). We show that these regions can be identified from the Ab sequence as well. We used "Paratome", an implementation of a structural approach for the identification of structural consensus in Abs [19]. While residues identified by Paratome cover virtually all the Ag binding sites, the CDRs (as identified by the commonly used CDR identification tools) miss significant portions of them. We refer to the Ag binding residues which are identified by Paratome but are not identified by any of the common CDR identification methods, as Paratome-unique residues. Similarly, Ag binding residues that are identified by any of the common CDR identification methods but are not identified by Paratome are referred to as CDRs-unique residues. We show that Paratome-unique residues make crucial energetic contribution to Ab-Ag interactions, while CDRs-unique residues have a rather minor contribution. These results allow for better identification of Ag binding sites and thus for better identification of B-cell epitopes. They may also help improve vaccine and Ab design.

## Results

### Structural consensus defines ABRs

The outline of our structure-based ABRs identification method is delineated in Figure 1. Briefly, the algorithm structurally aligns all known Abs and marks the residues that contact the Ag in each of them. We have shown [19,20] that in this multiple structure alignment there is a consensus among Abs that some structurally aligned positions contact the Ag. These positions form six sequence stretches along the Ab sequence that roughly correspond to the six CDRs. Beyond the edges of these stretches there were no structurally aligned positions in which more than 10% of the Abs contact the Ag. Thus, we defined the boundaries of the ABRs based on these stretches and marked the ABRs in all the Abs in our dataset.

### Paratome: Automatic sequence based ABRs identification

Figure 2 depicts the automated ABRs identification tool we developed. Given a query sequence (Figure 2A) a BLAST search is performed against all Abs in the dataset described above. The best hit (i.e., lowest E-value) is used to infer the positions of the ABRs in the query sequence, based on its alignment to the annotated Ab from the dataset. When the query Ab has a known 3-D structure, it can be used to identify the ABRs as described in Figure 2B (see Methods).

### Content statistics

Figure 3 summarizes the number of residues identified by each method on the test set. In all regions except L1 and H2, Paratome identified a slightly larger number of residues than any other method. The largest differences were recorded in L2 and H2. In L2, Paratome had 50% more residues identified than Kabat and Chothia and four times the number of residues identified by IMGT. For H2, Kabat and Paratome identified twice the number of residues suggested by Chothia and IMGT.

### Structural consensus regions contain virtually all Ag binding residues

For each Ab in our test dataset we recorded the average recall of the residues that actually bind the Ag by each method. Given the typical trade-off between recall and precision in which the increase of one is at the cost of decreasing the other, we measured the average precision of each method. The results are presented in Figure 4. The ABRs identified by Paratome included 94% of Ag binding residues, followed by Kabat (85%), IMGT (81%) and Chothia (79%) CDRs. Precision rates ranged between 48% (IMGT) and 41% (Kabat), with Chothia (44%) and Paratome (42%) in between.

### ABRs-specific residues cover 10–17% of the Ag binding sites

Table 1 compares the consensus sets and the method specific sets of residues. The Paratome-Kabat consensus set is the largest (3476 residues), covering 83.54% of the Ag binding sites. Paratome-Chothia consensus set covered 77.08% of the Ag binding sites (3203 residues), and Paratome-IMGT consensus set covered 79.47% of the Ag binding sites (3077 residues). In all consensus sets, approximately 50% of the residues are Ag binding residues. ΔParatome contains a substantially larger percentage of Ag binding residues than ΔKabat, ΔChothia and ΔIMGT (20.8%, 26.23% and 20.6% respectively, compared with 5.03%, 4.88% and 6.88% respectively).

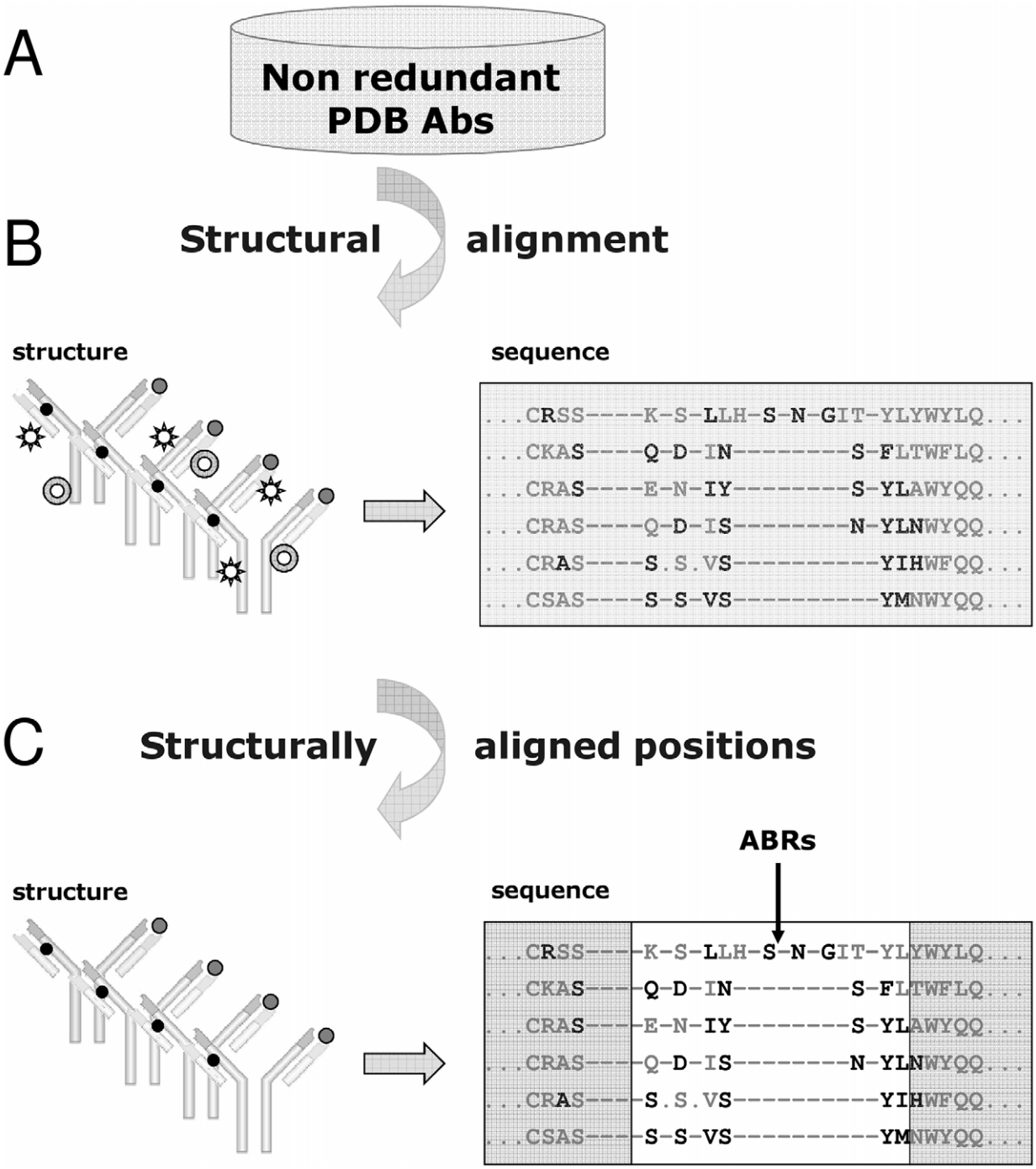


**Figure 1. Structure-based identification of ABRs.** (A) Using the non-redundant set of all Ab-Ag complexes in the PDB, (B) we created a multiple structure alignment of the Abs. Residues that are in contact with the Ag were identified by searching for structurally aligned positions that systematically create contacts with the Ag (black and grey solid circles) and disregarded positions that contact the Ag only sporadically (open shapes). (C) The contacting positions were mapped to the sequence representation of the multiple structure alignment (bold letters). The stretches of amino acids in which at least 10% of the Abs are in contact with the Ag were defined as ABRs (white rectangle).
doi:10.1371/journal.pcbi.1002388.g001

Moreover, ΔParatome residues cover a significantly larger portion of the Ag binding sites. ΔParatome residues covered 10.77% of the Ag binding sites while ΔKabat covered merely 1.78% of the Ag binding sites. The coverage of ΔParatome (14.84%) was 20 times larger than that of ΔIMGT (0.76%). When compared to Chothia, the coverage of ΔParatome (17.23%) was, again, more than an order of magnitude greater than that of ΔChothia's (0.86%). In each comparison, Paratome-specific residues covered a significantly larger portion of the Ag binding sites than the alternative method-specific residues. Thus, indicating that structural consensus regions capture more of the Ag binding portion of Abs.

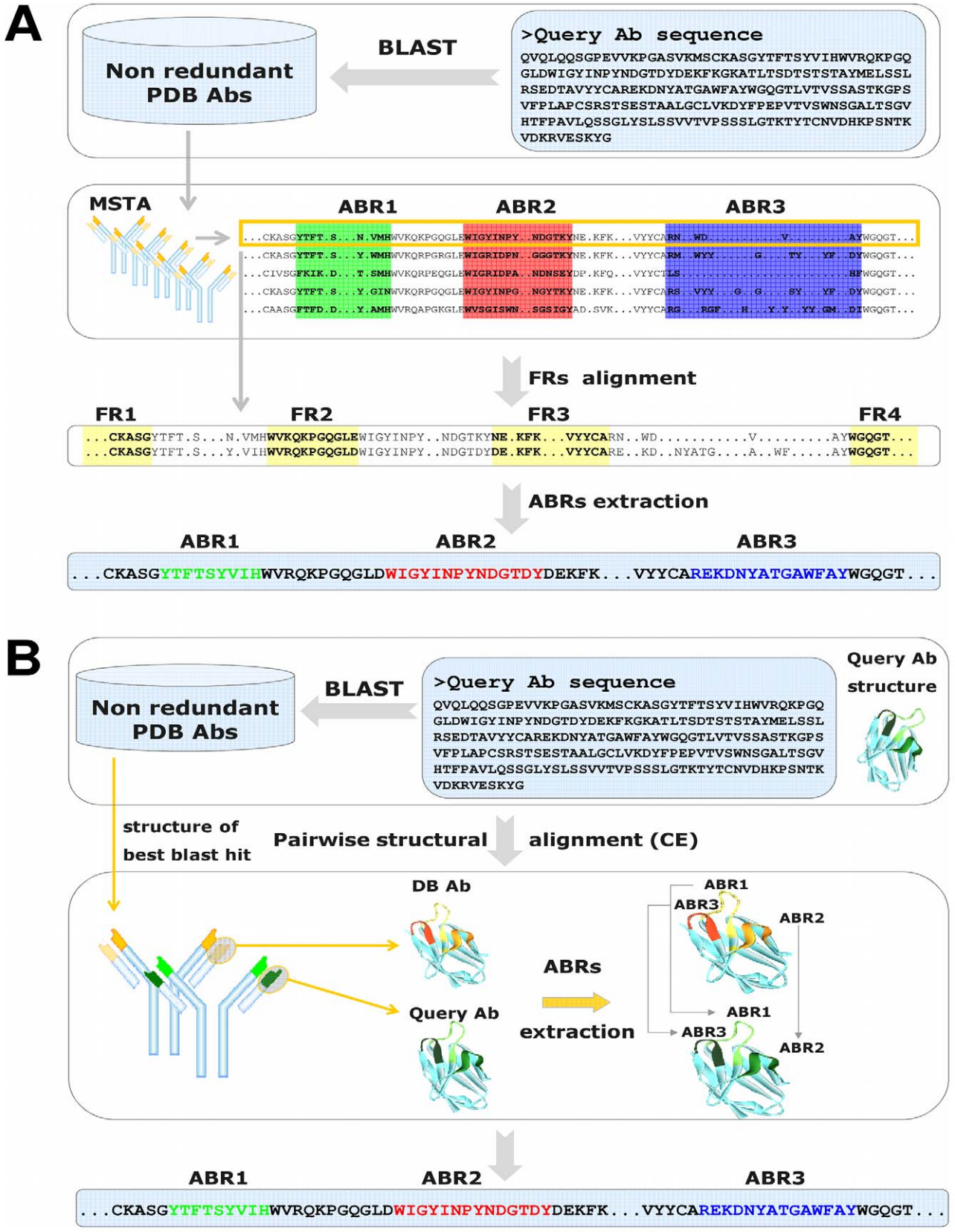


**Figure 2. Automated ABRs Identification** (**A**) **Sequence based ABRs identification.** A BLAST search is performed using the query Ab sequence versus the dataset of non-redundant PDB Abs. Using the best hit from the BLAST search, the query and annotated Abs FRs are aligned and hence the query sequence ABRs are inferred based on the location of the annotated sequence ABRs in the MSTA. (**B**) **Structure based ABRs identification.** A BLAST search is performed using the sequence of the query Ab versus our dataset of Abs. Using the best hit from the BLAST search, the query and annotated Abs are structurally aligned. The ABRs of the query Ab are inferred based on the location of the annotated Ab ABRs in the MSTA.

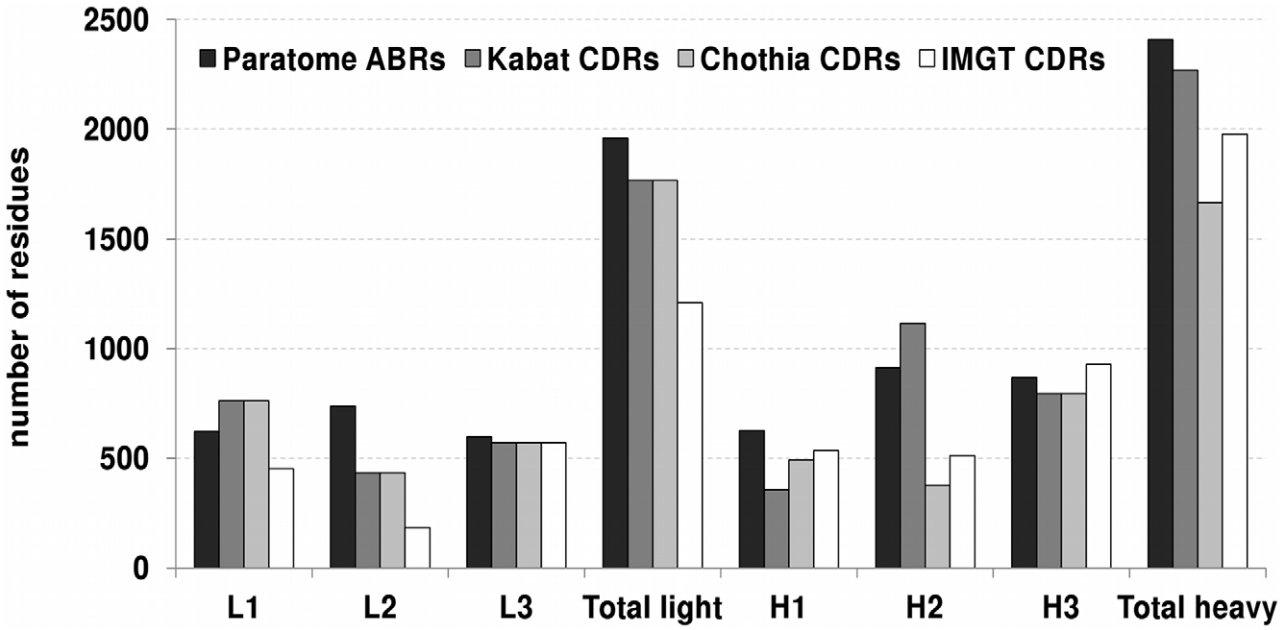


**Figure 3. Total number of residues identified by each method for all Ab-Ag complexes in the test set.** L1–L3 are ABR/CDR1-3 of the light chain. H1–H3 are ABR/CDR1-3 of the heavy chain. Total light and heavy are the sum of all identified residues in the light and heavy chains respectively.
doi:10.1371/journal.pcbi.1002388.g003

## Differences in ratios of Ag binding residues

Figure 5A shows the average precision for each ABR/CDR on the light and heavy chains as defined by each of the methods. L2 has the lowest precision in all methods. For L3, all the methods have a similar precision, with a slightly higher rate for Paratome (0.55). IMGT has the highest precision for L1 (0.46), followed by Paratome (0.38) and Chothia and Kabat has the lowest precision (0.27). The largest difference between the methods is in H2 where Chothia has the highest precision (0.69), followed by IMGT (0.57), then Paratome (0.43) and Kabat (0.37).

Figure 5B summarizes the average recall of each method for each of the six regions. For all methods, L2 has the lowest recall (2–7%). This is expected considering L2 has the lowest precision (see Figure 5A). For L1, all methods show similar recall (11–12%). The same holds for H3, which covers the largest fraction of the Ag binding sites (24–25%). H2 shows the highest diversity; For Paratome and Kabat it covers 21% of the Ag binding sites while for Chothia and IMGT recall ranged between 13–15%, respectively. In all cases, Paratome shows the highest recall. Note that while the overall recall ranges between 0.7–1 (see Figure 4), the recall of each of the six regions ranges between 0–0.3. This is due to the fact that the total recall is the accumulation of the recall obtained by each of the six regions.

## Paratome-unique residues are important for Ag binding

To gain insight into the extent to which Paratome-unique residues contribute to Ag binding, we searched the non-redundant set of Abs for Ag binding residues residing within structural consensus regions that are not identified by any of the CDR identification methods. We obtained 153 Paratome-unique residues, originating from 104 Abs (Table S3). Using the FoldX algorithm [21,22], we performed an in-silico alanine scan in which each Paratome-unique residue and each Ag binding residue identified by the CDR identification methods (2707 residues) within the 104 Abs were mutated to Alanine. Additionally, we searched the non-redundant set of Abs for Ag binding residues residing within CDRs that are not identified by Paratome (i.e. CDRs-unique residues). We found 59 CDRs-unique residues, stemming from 41 Abs (Table S4). To each CDRs-unique residue we performed an in-silico alanine scan in which it was mutated to Alanine. The distribution of the predicted interaction energy ($\Delta\Delta G$) of these mutants is presented in figure 6A. Destabilizing residues in this analysis ($\Delta\Delta G>0.25$) are residues whose mutation to alanine is predicted to destabilize the Ab-Ag complex. These residues, therefore, are likely to be important for Ag binding. Paratome-unique residues have a slightly higher percentage of destabilizing residues (49%) than Ag binding residues that fall within the CDRs according to Kabat, Chothia or IMGT (44.15%). While it is not clear whether the differences between Paratome-unique and Ag binding residues within the CDRs are significant, it is obvious that the former are at least as important to stability as the latter. In contrast, CDRs-unique residues have substantially lower contribution to binding: only 27% of them are destabilizing and the vast majority of them (70%) are neutral. To demonstrate the importance of Paratome-unique residues we show a more detailed analysis of the complex of IL-15 with an anti-IL-15 Ab (PDB ID 2xqb). Two Ag binding residues, LEU46 and TYR49, which were identified by Paratome to be part of ABR L2, were not identified by any of the CDR identification methods (Table S1). Figure 6B shows these residues relative to the surface of the Ag. It can be seen that TYR49 protrudes into the surface of the Ag, while LEU46 is located opposite to the antigenic LEU52, forming a hydrophobic interaction. As shown is Figure 6C, only seven residues from L2 interact with the Ag, and two of them are Paratome-unique residues. TYR49 forms one of the two hydrogen bonds between the Ag and ABR L2. The results of the FoldX in-silico single-point mutations analysis indicate that mutating ARG50, ARG53 and TYR49 to Alanine have the most significant destabilizing effect (Table S2). Not surprisingly, due to the salt bridge it forms with antigenic GLU46, mutating ARG50 had the

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.