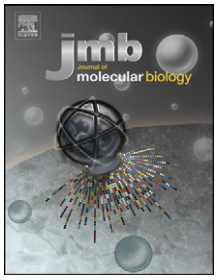

Restricted Diversity of Antigen Binding Residues of Antibodies Revealed by Computational Alanine Scanning of 227 Antibody–Antigen Complexes

Gautier Robin [1,2,3,4], Yoshiteru Sato [5], Dominique Desplancq [6], Natacha Rochel [5], Etienne Weiss [6] and Pierre Martineau [1,2,3,4]

*1 - **IRCM,** Institut de Recherche en Cancérologie de Montpellier, Montpellier, F-34298, France*
*2 - **INSERM U896,** Montpellier, F-34298, France*
*3 - **Université Montpellier 1,** Montpellier, F-34298, France*
*4 - **Institut Régional du Cancer Montpellier,** Montpellier, F-34298, France*
*5 - **Integrated Structural Biology Department,** Institute of Genetics and Molecular and Cellular Biology, INSERM U964/CNRS UMR 7104/Université de Strasbourg, Illkirch F-67404, France*
*6 - **Ecole Supérieure de Biotechnologie de Strasbourg,** UMR 7242, CNRS/Université de Strasbourg, Illkirch F-67412, France*

***Correspondence to Pierre Martineau:** Institut de Recherche en Cancérologie de Montpellier, Montpellier F-34298, France.*
pierre.martineau@inserm.fr
http://dx.doi.org/10.1016/j.jmb.2014.08.013
Edited by A. Skerra

Abstract

Antibody molecules are able to recognize any antigen with high affinity and specificity. To get insight into the molecular diversity at the source of this functional diversity, we compiled and analyzed a non-redundant aligned collection of 227 structures of antibody–antigen complexes. Free energy of binding of all the residue side chains was quantified by computational alanine scanning, allowing the first large-scale quantitative description of antibody paratopes. This demonstrated that as few as 8 residues among 30 key positions are sufficient to explain 80% of the binding free energy in most complexes. At these positions, the residue distribution is not only different from that of other surface residues but also dependent on the role played by the side chain in the interaction, residues participating in the binding energy being mainly aromatic residues, and Gly or Ser otherwise. To question the generality of these binding characteristics, we isolated an antibody fragment by phage display using a biased synthetic repertoire with only two diversified complementarity-determining regions and solved its structure in complex with its antigen. Despite this restricted diversity, the structure demonstrated that all complementarity-determining regions were involved in the interaction with the antigen and that the rules derived from the natural antibody repertoire apply to this synthetic binder, thus demonstrating the robustness and universality of our results.

Introduction

Antibody molecules are able to recognize almost any part of any antigen molecule [1]. This unique property makes the antibody one of the most used reagents in research, diagnosis, and therapy. As the number of applications increases, along with the demand for antibody improvements in affinity, specificity, stability, and solubility, so does the need to refine the knowledge of antibody–antigen recognition in greater details for enhanced rational designs.

The paratope is defined as the part of the antibody that interacts with its target antigen. It is thus only defined for a given antigen–antibody pair and is not a *per se* entity [2]. As an example, some antibodies have been shown to interact with different and unrelated antigens and thus contain two different paratopes [3,4]. However, there are some common features in the mode of binding that allow the paratope to be partially defined independently of the antigen. Antibody variable domains (Fv) can be divided into hypervariable regions and more conserved segments called framework regions (FRs) [5,6]. Because most of the diversity is located in the hypervariable regions that pack together at the surface of the antibody and make most of the contacts with the antigen, these

hypervariable regions are usually considered to be identical with the complementarity-determining regions (CDR) that form the paratope. This is indeed generally true since binding activity can be transferred between antibody molecules by just exchanging their hypervariable regions [7,8]. However, many residues of the hypervariable regions never interact directly with the antigen and only play a structural role [8,9], and several residues located in FRs may also contact the antigen [8,10]. Definition of the paratope is further complicated by the importance of water molecules in the stability of the binding complex [11].

An antibody and its antigen must possess a complementary surface in terms of both shape and chemical nature in order to achieve high affinity. The shape diversity is mainly due not only to variations in loop lengths and conformations [9,12–14] but also to modifications in the relative orientation of heavy chain variable domain (VH) and light chain variable domain (VL) [15], as well as solvation of the binding interface [11]. This shape complementarity involves many aromatic residues that pull together the two surfaces using mainly van der Waals and hydrophobic interactions [16]. Specificity and strengthening is obtained by electrostatic interactions between charged side chains, as well as hydrogen bonds bridging oxygen and/or nitrogen atoms.

Several studies have analyzed large sets of antibody–antigen complexes to precisely delineate paratopes and epitopes. However, most of these studies defined interacting residues by comparing accessibility of the antibody residues in the presence and in the absence of the bound antigen [13,17]. If a residue is buried within the interacting interface, it is considered as a contact residue and as participating in the interaction. However, the change in solvent-accessible surface area does not always correlate with the free energy of binding [18]. In other studies, authors used a distance cutoff to define paratope residues. For instance, in a series of papers, Dr. Ofran's group identified interacting residues in 200 antibody–antigen complexes by considering that two residues were in contact if they have at least one pair of atoms within a distance of 6 Å from each other [10,19,20]. Both approaches have the advantage of giving an exhaustive list of potential paratope residues at the expense of a large number of false positive positions, since residues in direct contact within the antibody–antigen interface do not necessarily participate to the binding free energy [21].

In this work, we focused on a quantitative and statistical analysis of side-chain contributions in a large set of antibody–antigen complexes. We first established a database of 227 non-redundant structures and we calculated for each residue in the binding interface its contribution to the antibody–antigen interaction free energy using a computational alanine scanning procedure. To compare the results between structures, we numbered all antibodies using a previously established numbering scheme that minimizes the average deviation from the averaged structure of the aligned domains [17]. This allowed us to define quantitatively residue average contributions, revisit paratope definition, and address questions regarding the number of significant contacts, the CDR relative free energy contributions, and side-chain usage and roles. The study demonstrates that paratopes are mainly discontinuous, comprise between 4 and 13 residues located within a set of 30 fixed positions, and contain a subset of amino acids different from the one found at the rest of the antibody surface. Finally, we solved the structure of an antibody–antigen complex whose antibody had been previously isolated from a biased synthetic antibody repertoire with only two diversified CDRs, and we demonstrated that this unnatural antibody molecule fulfilled the rules obtained in our analysis based on natural antibodies. Taken together, our results reveal conserved structural constraints that shape a restricted diversity of the antibody paratope, being of natural or synthetic source, and should help in the understanding of the mechanisms of immune recognition and the rational design of new and improved antibodies.

Results

Aligned database of antibody side-chain binding free energies to antigens

For the purpose of this study, we focused on protein and peptide antigens since haptens are known to have a different mode of binding that would require an independent analysis to derive general characteristic rules. We first recovered 506 structures of antibody–antigen complexes from the IMGT database (Supplementary Table 1), removed mis--annotated structures, and then filtered for sequence redundancy. We independently extracted VH and VL domains from the structures together with the interacting molecules of the antigen. Finally, we renamed antibody and antigen chains using a common naming scheme and renumbered variable domains using AHo numbering [17]. The choice of AHo numbering was dictated by its design, based on the spatial alignment of known three-dimensional structures of immunoglobulin domains. The final database contains 227 VH and 206 VL non-redundant renumbered structures and their corresponding interacting antigen molecules (Supplementary Table 2), representing 206 Fv and 21 VHH. To evaluate the contribution of each antibody side chain present in the binding interface to the free energy of binding, we performed computational alanine scanning experiments using the FoldX program [22]. The contribution of each residue side chain to the antibody–antigen

binding free energy for the 433 complexes is given in Supplementary File complexes.xlsm.

Few residues contribute significantly to the antibody–antigen binding free energy

For each structure, we ranked the residues in decreasing order of their binding free energy and computed the percentage of the binding free energy obtained as a function of the number of residues for the heavy chain (a), the light chain (b), and the complete Fv domain (c) (Fig. 1). The side chain with the highest antibody–antigen binding free energy contributed on average to 22%, 33%, and 16% of the $\Delta\Delta G$ for VH, VL, and Fv, respectively. As a general trend, the free energy contribution decreased exponentially with the ranking so that only the first few residues contributed significantly to the binding free energy. For instance, 80% of the binding free energy was obtained with only 6 (IC95%: 2–10), 4 (IC95%: 1–6), and 8 (IC95%: 4–13) residues for the VH, VL, and Fv, respectively. In other words, 80% of the binding free energy is due to only 4–13 residues in 95% of the Fv–antigen complexes. Thus, this represents the length of the functional paratope formed by the residues that make energetically favorable contacts with the antigen [23]. The result was independent of the light chain class and the number of contacts was identical for both λ and κ variable domains (Supplementary Fig. 1). This demonstrated that antibody paratopes require only few residues of the CDRs to raise high affinity and specificity.

Binding free energy per position

In the previous results, the considered residues were different and optimally chosen for each antibody–antigen complex. To determine if some common residues were involved in the interaction,

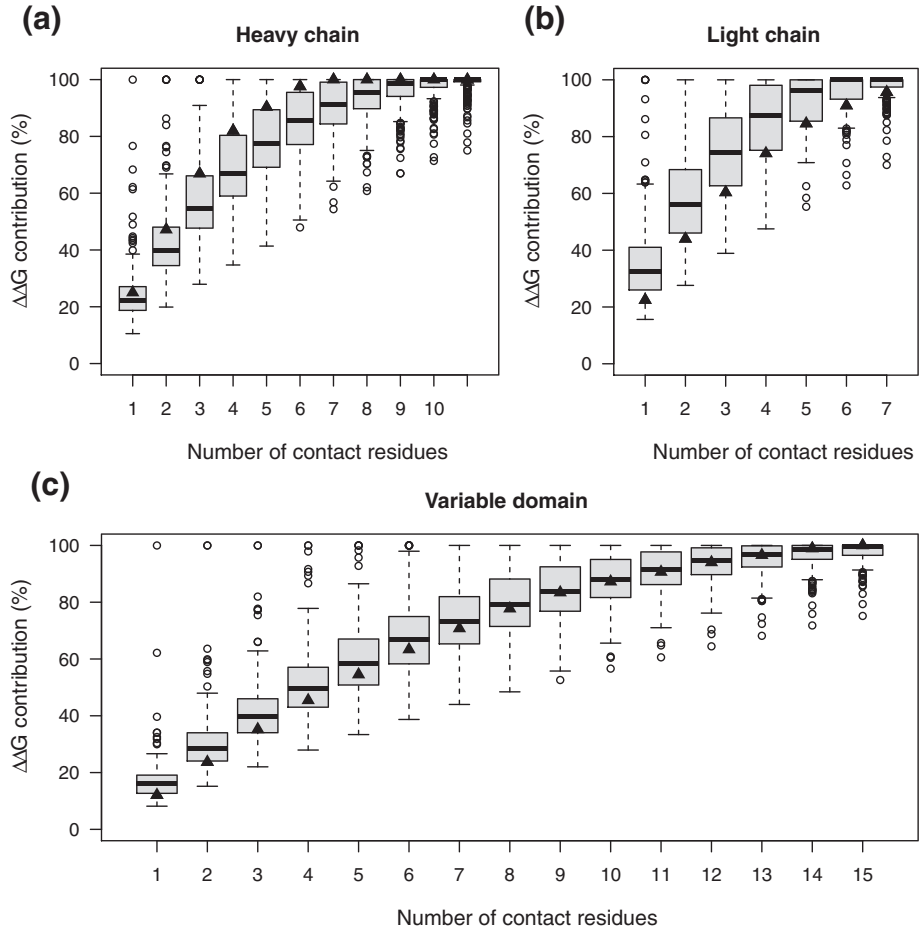

Fig. 1. Number of residues participating in binding free energy. Distribution of the percentage of the binding free energy obtained for a given number of contact residues, for VH (a), VL (b), and Fv (c) domains. Black-filled triangles are the values obtained for the GkF5 complex. Boxplot whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range.

we computed the mean binding free energy at each position in the series of complexes. This was possible thanks to the AHo numbering scheme that insured that equally numbered residues were also structural equivalents. Detailed results are presented in Supplementary File complexes.xlsm. A plot of the mean binding free energy per position is shown in Fig. 2 for the heavy and light chains. Most of the interacting sites were located in the CDR regions (Kabat's definition) but some frequently used positions such as H2, H29-32, H44, H54, H108, H139, L2, L44, L54, and L57 were located in the FRs (average $\Delta\Delta G > 0.1$ kcal/mol). Within the CDRs, there was an uneven distribution of the antibody–antigen binding free energy limited to a few specific positions (hot spots). Indeed only 2, 4, 1, 2, and 2 positions contributed to more than 60% of the total CDR binding free energy for H1, H2, L1, L2, and L3, respectively. We noticed that, only in H3, the binding free energy was more evenly distributed. This is due to the highly variable length of this CDR that smoothes the energy per position. If the energy distribution in H3 loops had been studied for any given length, the distribution would have become more heterogeneous, as for the other CDRs (Supplementary File complexes.xlsm). However, because of the very high diversity of H3 loop structures, the structural equivalence of equally numbered residues is highly questionable, except for the N- and C-terminal parts of the loops [24]. Positions were ranked according to their average $\Delta\Delta G$, and the percentage of the total $\Delta\Delta G$ as a function of the number of these fixed residues is shown in Fig. 3. This showed, for example, that 80% of the total $\Delta\Delta G$ localized among 30 positions, 27 in the CDRs and 3 in FRs, for more than half of the structures, representing only a fraction of the total CDR positions (less than half of the CDR positions according to Kabat's definition). Furthermore, several of these positions were not always occupied and their number could be easily reduced for most antibodies. The results presented in Figs. 2 and 3 were obtained for the whole database that contains mainly κ light chains. If analyzed independently, the best positions were slightly dependent on the light chain class, λ light chains having a stronger propensity than κ light chains to bind the antigen using residue L109 and κ light chains having frequently longer L1 loops making contacts with residues L32–L36 (Supplementary Fig. 2).

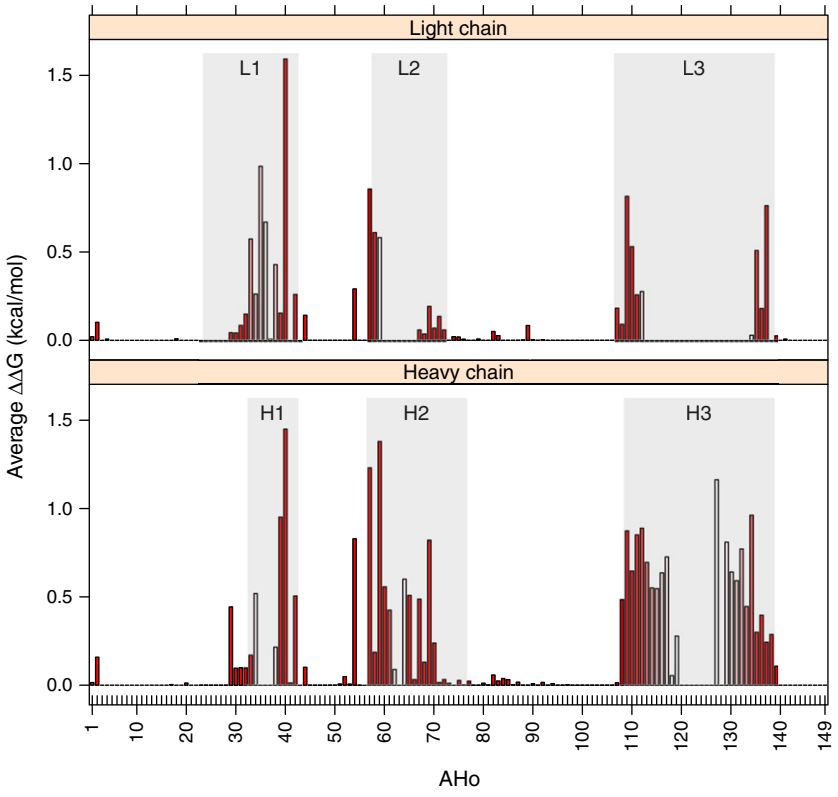

Fig. 2. Average binding free energy. Bars represent the average antibody–antigen binding free energy at each position for VL (top) and VH (bottom) domains. The bar color represents the frequency of occupancy of this position in the database of aligned antibody, from 0% (white) to 100% (deeper red).

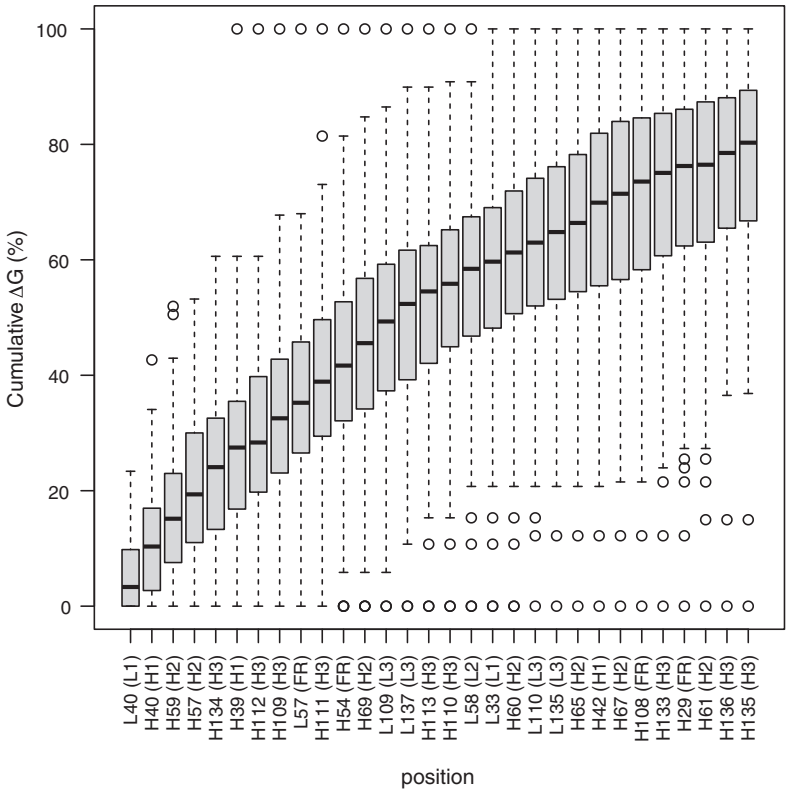

Fig. 3. Cumulative binding free energy. Distribution of the binding free energy for the 30 most contributing positions. The energy is cumulative from the left. Only positions occupied in more than 40% of the structures are shown. Boxplot whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range.

CDR contributions to antigen binding

The binding free energy contribution of each CDR was computed for the 206 Fv–antigen complexes (Fig. 4a). H2 and H3 were the two CDRs that contributed the most to the antibody–antigen interaction, each of them contributing to 22% of the total binding free energy in half of the complexes. The three CDRs H1, L1, and L3 contributed almost equally to the free energy of interaction with a median value of 10%, 9%, and 12%, respectively. Finally, L2 did not interact with the antigen in more than half of the complexes.

Furthermore, the number of CDRs involved in antigen binding was estimated in each complex (Fig. 4b). A CDR was considered as interacting with the antigen when at least one of its residues participated significantly in the interaction ($\Delta\Delta G > 0.8$ kcal/mol). The distribution of the number of used CDRs showed that, in 97% of the cases, at least three CDRs contained one or more such important residues. It is noteworthy that the 206 complexes used in this analysis were all standard antibodies with heavy and light chains, discarding VHH-like antibodies for which the interaction mode, constrained by only three CDRs was very different (Supplementary Fig. 3).

In correlation with the binding free energy contribution of each CDR (Fig. 4a), H2 and H3 contained at least one residue that contributed significantly to the binding energy ($\Delta\Delta G > 0.8$ kcal/mol) in 93% and 90% of the antibody–antigen complexes, respectively, whereas only 39% of the structures showed a significant interaction between L2 and the antigen (Fig. 4c). Finally, H1, L1, and L3 contained a residue with a significant free binding energy in 77%, 67%, and 78% of the cases, respectively.

Side-chain usage and roles

Several studies have shown that particular amino acid side chains play different roles during the formation of a protein complex [25,26], including antibodies [2,12,20,27–29]. Having identified the antibody positions involved in antigen binding and their average binding free energy, we focused on a detailed analysis of the energy contributed by each side-chain type during side chain–side chain interactions. Paratope analysis was restricted to the 15 most contributing CDR positions that represented about 50% of the binding energy (Fig. 3). As a reference, we used 62 surface residues that never interacted with the antigen (mean $\Delta G = 0$ in Fig. 2

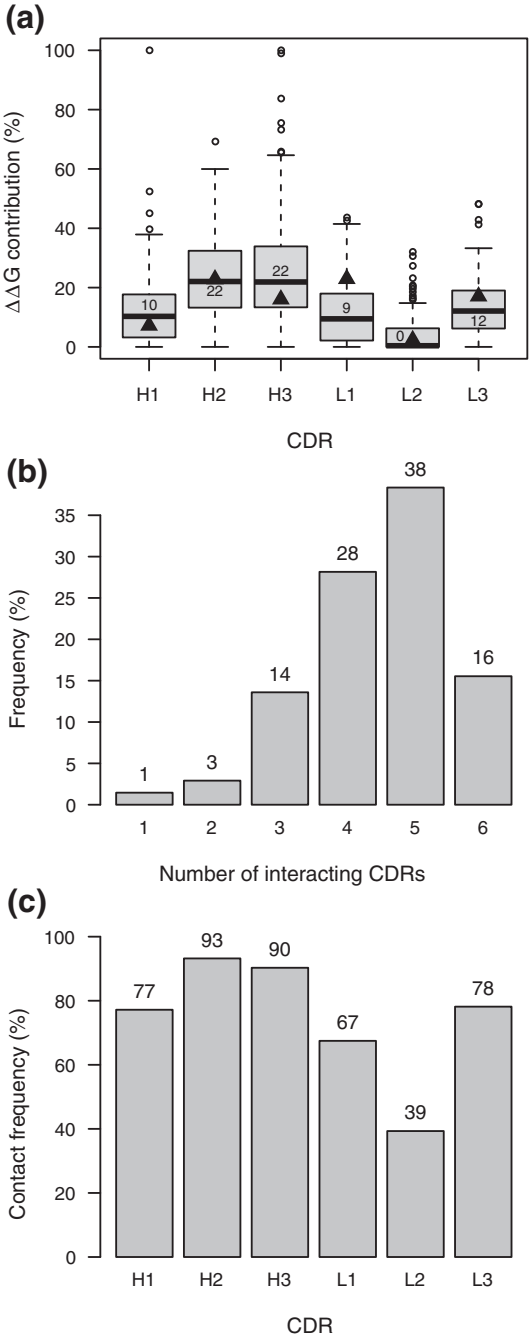

Fig. 4. CDR contributions and usage. (a) Distribution of the CDR binding free energy contributions. The value in the boxes is the median of the distribution. Black-filled triangles are the values obtained for the GkF5 complex. Boxplot whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range. (b) Percentage of structure with a given number of interacting CDR. (c) Percentage of structures in which the indicated CDR contributes to the binding free energy.

and solvent accessibility > 19%). Side- chain usage frequencies were very different for these two classes of surface residues, illustrating a highly specific residue usage at antibody hot spots (Fig. 5). Among the 7 most frequent residues of these two groups, representing 75% of the used residues, only two residues were in common, G and S. Tyrosine, the most frequent residue at hot spot positions with a 24% frequency, was only present with a 2.7% frequency at non-paratope surface positions. This general trend was found at most of the positions of the paratope but H39, L40, and L137 showed a somewhat different distribution, the two former being strongly enriched in Tyr and the latter in Pro and Leu (Supplementary Fig. 4).

We took advantage of the quantitative nature of our database to also calculate the frequency of the positive contributions of each residue of the paratope to the free energy of binding (black bars in Fig. 5). The ranking of the residues was clearly different and two classes could be identified: residues that make frequent interactions with the antigen, essentially aromatics, and neutral residues that were frequently found in the paratope but did not play any role in the binding free energy, as G and S.

Finally, for residues more frequently found at the paratope than at other surface positions, we noticed that their side chains made frequent interactions with the antigen. For instance, Y and W, respectively enriched 9 and 450 times, interacted with the antigen 66% and 74% of the time, respectively, and the other 3-fold enriched residues were also involved in the stability of the complex at a frequency of at least 25% (N: 25%, H: 37%, D: 31%, F: 32%). On the other hand, residues Q, K, P, and T that were 3-fold less frequent in the paratope than on the protein surface were also less likely to make productive interactions with the antigen (Q: 22%, K: 22%, T: 19%), except for proline (51% productive interactions). This abnormal reduced frequency of proline residues at hot spot positions is presumably due to the structural constraint this amino acid would impose on the CDR conformations when its presence is not required for shaping the paratope.

Universality of the characteristic rules, a structure case study

Most of the antibodies present in the current Protein Data Bank (PDB) and analyzed above have been obtained from immunized animals and antibody phage libraries made from human donors. In both cases, the diversity of the paratope is natural and originates from V gene diversity, V(D)J recombination and somatic hypermutation events. We demonstrated that these antibodies interact with their antigen using a very limited number of residues, both in position and in nature. Such a restricted diversity could originate from constraints imposed either by the immune system or by the structure of the antibody paratope itself. To address this question, we isolated a single chain Fv (scFv) from a

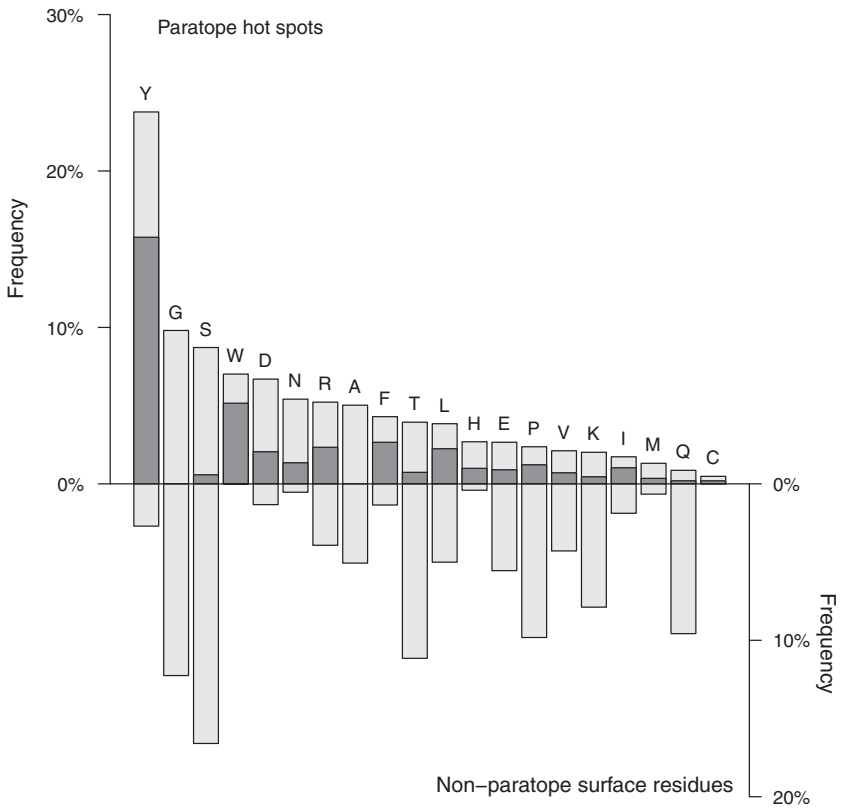

Fig. 5. Residue usage at paratope and non-paratope positions. The height of the bars is the frequency of a given residue at the 15 most contribution positions of the paratope (left axis) or at 62 solvent-accessible and non-interacting residues (right axis). The dark-gray bars are the frequency of the considered residue contributing to at least 0.8 kcal/mol to the free energy of binding.

synthetic strongly biased antibody library by panning on Gankyrin antigen. This scFv binds its antigen with a moderate affinity that has been estimated to 100–200 nM using Biacore [30] and 450 nM by competition ELISA [31]. The complex is however stable enough to be separated from the free antigen and the scFv by gel-filtration chromatography [31]. In the used library [32], diversity was introduced only in the two CDR3s, the four other CDRs being fixed to those of an anti-β-galactosidase antibody [33]. In addition, whereas the introduced H3 sequence was close to a human one, the introduced L3 was unnatural since its N-terminal part did not correspond to any natural germline. We then solved and analyzed the Gankyrin–scFv F5 (GkF5) complex to determine whether the library design may bias the paratope location and chemical composition.

The crystal structure of the GkF5 was solved at 2.5 Å resolution (Table 1). The Gankyrin consists of seven ankyrin repeats, ANK1–ANK7. Its conformation is similar to the previously solved Gankyrin structures [34,35]. The C^{α} atoms of the bound and the uncomplexed Gankyrin structures can be superimposed with a RMSD of 0.95 Å, showing no large conformational rearrangements due to scFv binding. The scFv F5 adopts the β-sheet-rich canonical antibody fold but the glycine-rich linking region between VH and VL chains was not seen in the electron density map because of its flexibility. The antibody F5 interacts with the α-helices of the ANK4–ANK6 repeats, which are part of the convex face of Gankyrin (Fig. 6a).

The antibody paratope is not limited to the two diversified CDR3s but instead includes other CDRs in contact with the antigen (Fig. 6a). Computational alanine scanning analysis further specified that 12 antibody residues located in 5 CDRs and 1 FR were implicated in the stability of the complex (Fig. 6b). Eleven of these residues were located within the 30 best hot spot positions determined above from the antibody–antigen database (Table 2 and Fig. 3) and their binding free energy profiles fitted well with the average values obtained in the database (filled triangles in Fig. 1). The missed residue is N167 (L39) with a $\Delta\Delta G$ of 0.9 kcal/mol. This position has a low average energy in the database due to its low usage in κ light chains. However, if we only consider Fv domains with a light chain of the same λ class

Table 1. Data collection and refinement statistics.

	Gankyrin–F5 complex
Data processing	
Resolution (Å)	30–2.5 (2.59–2.50)
Crystal space group	$P3_221$
Cell parameters (Å)	
a	135.6
b	135.6
c	68.7
Unique reflections	25,162 (2438)
Mean redundancy	2.8 (2.7)
R_{sym} (%)[a]	9.7 (45.4)
Completeness (%)	98.9 (97.9)
Mean I/σ	10.8 (2.2)
Wilson *B*-factor (Å^2)	41.2
Refinement	
Resolution (Å)	30–2.5
Number of non-hydrogen atoms	
Gankyrin	1688
F5 antibody	1684
Water molecules	428
RMSD bond length (Å)	0.003
RMSD bond angles (°)	0.69
R_{cryst} (%)[b]	16.4
R_{free} (%)[c]	20.8
Averaged *B*-factor for non-hydrogen atoms (Å^2)	
Gankyrin	26.6
Antibody F5	55.1
Water molecules	45.4
Ramachandran plot (%)	
Core	87.5
Allowed	11.7
Generous	0.5
Disallowed	0.3

[a] $R_{sym} = 100 \times \sum_{hj}|I_{hj} - \langle I_h\rangle|/\sum_{hj}I_{hj}$, where I_{hj} is the *j*th measurement of the intensity of reflection *h* and $\langle I_h\rangle$ is its mean value.
[b] $R_{cryst} = 100 \times \sum||F_o| - |F_c||/\sum|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.
[c] Calculated using a random set containing 5% of observations that were not included throughout refinement [62].

than the scFv F5, L39 ranks 18th with an average $\Delta\Delta G$ of 0.6 kcal/mol (Supplementary Fig. 2). Furthermore, the CDR contributions to the GkF5 complex were very similar to the averaged values obtained from the database (filled triangle in Fig. 4a). The binding free energy of the two diversified CDRs (H3 and L3) added up to 32% of the total $\Delta\Delta G$, whereas the three interacting but fixed CDRs (H1, H2, and L1) and the FRs contributed 68% to the total $\Delta\Delta G$. This higher contribution of the fixed part of the paratope in the binding free energy is also found in the number of contacts since the two CDR3s and the non-diversified parts of the scFv (other CDRs and FRs) formed 4 and 8 contacts with the antigen, respectively. Finally, 83% of the total binding free energy is due to only 8 residues as in natural antibodies (Fig. 1c), 6 of them being Tyr.

To confirm this result using a different approach, we determined antibody contacts using the Crystallography and NMR System software suite that uses geometric rules to identify interactions (Table 2) [36,37]. Ten of the twelve residues identified using computational alanine scanning were also present in this analysis, the two missed residues being Y34 and Y185. In addition, among the 19 identified residues, 14 are in the list of the 30 most contributing positions presented in Fig. 3. Altogether, this demonstrated using two approaches that, despite its synthetic and biased origin, scFv F5 paratope conforms well to the general rules defined from our database analysis.

Universality of the characteristic rules, prediction of important paratope residues

We next tested if our database can be used to predict important paratope residues in the absence of structural and experimental data. As a reference, we used the experimental data generated by Muller *et al.* [38] on a humanized anti-VEGF Fab. In this publication, authors mutated one by one the 68 CDR residues to alanine and measured the relative affinity of the resulting mutant Fab. We grouped the CDR residues in three categories: 8 major contributors that affected the affinity by more than a 150-fold factor, 11 residues decreasing the affinity by a more moderate factor (between 5- and 66-fold), and 49 neutral or weak contributors. These residues are displayed with dark green, light green, and white backgrounds, respectively, in the line Muller1998 in Fig. 7. If we compare the location of these residues with the positions identified as the main hot spots in our study (Fig. 2), 7 of the 8 Muller's major contributors have an average binding free energy higher than 0.8 kcal/mol. Since only 12 hot spot positions have such a high average binding free energy, this means that using our data and targeting mutations to these 12 residues would have identified 7 (88%) out of 8 of the Fab major paratope residues. If we decrease the binding free energy cutoff value to 0.4 kcal/mol, our database identified 27 potential interaction sites in the Fab paratope (red and pink backgrounds in Fig. 7). Among these 27 sites, 15 positions were also identified (moderate or major contributors) in Muller's data. In that case, experimental alanine scanning of only 27 positions would have identified 15 (79%) out of 19 of the paratope residues. This good correlation between the identified positions and the experimental data was further confirmed using a second antibody for which the whole CDRs were also analyzed using shotgun scanning mutagenesis [23]. Again, among the 12 positions with an average binding free energy higher than 0.8 kcal/mol, all but two had a strong deleterious effect ($F_{wt/ala} > 20$) when mutated to alanine (Supplementary Fig. 5). Both of the two positions with a high energy that did not affect binding were in the light chain (L40 and L137) that marginally participates to the interaction in this particular antibody. In addition, only 2 of the 12 positions affecting the binding by more than a 50-fold factor have an

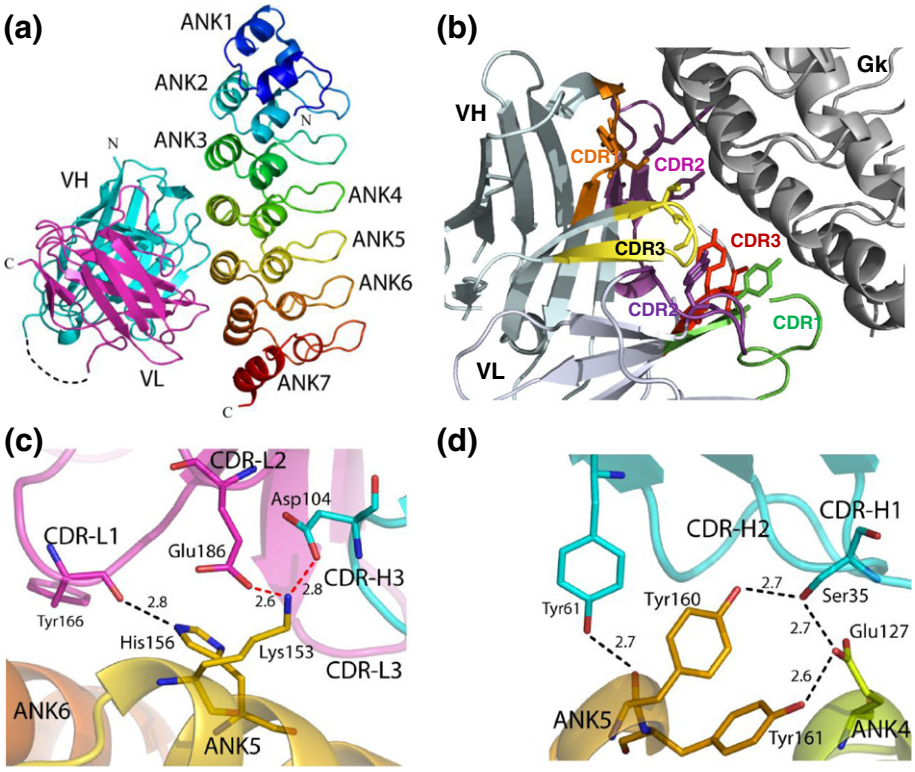

Fig. 6. GkF5 structure and analysis. (a) Overall view of the GkF5 complex. Each ankyrin repeat of Gankyrin is indicated by ANK1–ANK7 and different colors (from blue to red). The VH and VL domains of the antibody F5 are colored cyan and magenta, respectively. Dotted line represents unstructured linking region of the antibody. (b) GkF5 paratope. The 12 interacting residue side chains identified by computational alanine scanning (Table 2) are represented. The six CDRs are colored orange (H1), pink (H2), yellow (H3), green (L1), purple (L2), and red (L3). Ionic and hydrogen bonds between scFv F5 and Gankyrin are represented in the vicinity of the VL (c) and the VH (d) domains (Table 2). VL, VH, ANK4, ANK5, and ANK6 repeats are colored magenta, cyan, yellow, orange, and brown, respectively. Oxygen and nitrogen atoms are colored red and blue, respectively. Red and black dotted lines indicate ionic and hydrogen bonds, respectively.

average binding energy lower than 0.4 kcal/mol in our database (H33 and H135).

This demonstrated that our compiled database of free binding energy can be used to pinpoint important residues in an antibody paratope, decrease the number of residues to be experimentally tested by site-directed mutagenesis, and increase the chances to identify quickly the critical binding residues.

Discussion

The aim of this study was to describe quantitatively at the molecular level the mode of binding of antibody molecules to their cognate antigens, identify the critical residues of the paratope both in position and in nature, and derive general rules to help in antibody engineering and library design.

Large-scale alanine scanning analysis on an aligned non-redundant structural database of antibody–antigen complexes enabled the first quantitative description of the role of all the antibody residues in the free energy of binding. These data provide a quantitative description of paratopes from which characteristic rules can be derived. (1) Eighty percent of the binding energy is due to only a few significant residues, typically 8 (between 4 and 13). (2) These residues localize among 25–30 fixed hot spot positions that contribute to 75–80% of the energy in more than half of the complexes. (3) Between 3 and 6 CDRs are involved in the complex formation. (4) Hot spot positions are occupied by a restricted number of side chains (Y, G, S, W, D, and N in 61% of the cases), (5) among which essentially aromatic residues contribute to the binding free energy. These rules are valid not only for natural antibodies that constitute the majority of the structures used in this analysis but also to a synthetic antibody selected from a highly biased library with a focused and restricted diversity, and thus, these represent constraints present in most antibody–antigen complexes.

It is known since the first solved antibody–antigen structure that only a small subset of contact residues dominates the energetics of the association reaction. Several antibody–antigen complexes have been studied using experimental alanine scanning

Table 2. F5 hot spot positions and GkF5 contacts.

Rank[a]	AHo	CDR	scFv F5[b]	$\Delta\Delta G$[c]	Interaction[d]	Gankyrin[e]
1	L40	L1	Y168	3.2	VDW	ANK5:L152/K153/H156
2	H40	H1	S35	1.0	H-bond	ANK4:E127 ANK5:Y160
3	H59	H2	S54		VDW	ANK4:E127
6	H39	H1	Y34	0.9		
7	H112	H3	D104		Ionic bond	ANK5:K153
9	L57	FR	Y185	3.1		
10	H111	H3	I103	2.4	VDW	ANK5:K153/H156/I157/Y160
12	H69	H2	Y61	3.1	H-bond	ANK5:Y160
					VDW	ANK5:K162
13	L109	L3	Y227	2.3	VDW	ANK5:H156
16	H110	H3	V102	1.9	VDW	ANK5:Y160
17	L58	L2	E186		Ionic bond	ANK5:K153
18	L33	L1	G165		VDW	ANK6:E186
19	H60	H2	G55		H-bond	ANK4:E127
20	L110	L3	I228	1.5	VDW	ANK5:H156
24	H67	H2	Y59	2.7	VDW	ANK5:Y161/K162
27*[f]	L38	L1	Y166	2.0	H-bond	ANK5:H156
					VDW	ANK5:L152 ANK6:E186/K189/L190/S193
32*[f]	L112	L3	S230		VDW	ANK5:Y160
34	L111	L3	S229		VDW	ANK6:Q194
41	H33	H1	N33		VDW	ANK4:E127
42	L39	L1	N167	0.9	VDW	ANK5:H156
54	H30	FR	T30		VDW	ANK3:K90

[a] Rank of the residue as defined in Fig. 3.
[b] Nature of the residue in scFv F5. Numbering scheme is the sequential numbering used in the PDB ID: 4NIK structure. Equivalence between the different numbering schemes is given in Supplementary Table 3.
[c] $\Delta\Delta G$ of the residue in the scFv F5–Gankyrin complex, determined by computational alanine scanning, in kilocalories per mole. Only $\Delta\Delta G > 0.8$ kcal/mol are shown.
[d] Type of contact.
[e] Interacting residue in Gankyrin determined using contact.inp command in CNS. ANKn is the ankyrin repeat to which the following residue belongs.
[f] Position occupied in less than 40% of the antibody sequences and excluded from the ranking in Fig. 3. The rank given is that of the residue with the closest $\Delta\Delta G$ and present in more than 40% of the sequences.

mutagenesis and the effects precisely quantified. With the use of different models, it has been shown that the interaction involves between 5 and 10 main residue contributors [3,21]. Even in the case of an antibody with a very high affinity of 50 pM for its antigen, only 4–5 residues of the light chain contributed significantly to the binding energy [39], in good agreement with the values obtained from our database (Fig. 1b). Some known antibodies deviate however from this general rule as the anti-idiotypic antibody E5.2 in which most of the contact residues play a significant role in the binding free energy [3,40]. This discontinuous nature of the paratope may play a functional role since it makes the antibody more tolerant to mutations of the antigen, particularly because a loss of contact can be compensated by new water molecules, as demonstrated in the case of HEL D18A mutants in interaction with D1.3 antibody [40]. However, the novelty of our study is to show that there is a common mode of binding for almost all the antibody molecules present in the PDB. In other words, not only few residues contribute to the binding free energy but also these residues are mainly located at about 25 conserved positions. This restricted number of paratope residues may explain why D1.3 bound with the lysozyme and the anti-idiotypic antibody D5.2 showed a similar mode of binding with 13 common hot spot residues among 17 contacts [41]. Again, in some special cases, antibodies may adapt their binding in an unconventional manner as demonstrated in some highly mutated cross-neutralizing anti-human immunodeficiency virus antibodies [42].

Several studies have analyzed the distribution of side chains in the paratope of antibody molecules and concluded that there is notably a strong bias toward some aromatic residues [2,12,20,27–29]. However, in these analyses, authors used a definition of the paratope that spanned most of the CDR residues. For instance, Collis *et al.* analyzed the amino acid distribution at positions [12] they previously defined as contact residues [13]. Though being shorter than classical CDRs, these regions are still large and contain more than 50 residues. In our case, we restricted the analysis to the positions that formed an energetically favorable contact with the antigen in most of the analyzed structures since we only considered the 15 positions with the highest average free binding energy ($\Delta\Delta G > 0.87$ kcal/mol). In addition, we used surface residues of antibody molecules as a reference whereas Collis used loop

L1	AHo	24	25	26	…	29	30	31	32	…	39	40	41	42
VL mean ΔΔG		0.00	0.00	0.00		0.05	0.05	0.09	0.18		0.18	1.73	0.00	0.27
aa hot propensity		0.07	0.00	0.07		0.22	0.31	0.59	0.07		0.25	0.66	0.58	0.25
VL anti-VEGF		S	A	S		Q	D	I	S		N	Y	L	N
Muller1998		1.3	1.1	1.5		1.2	1.2	1.4	1.5		1.7	1.9	2.2	3.7

L2	AHo	58	…	67	68	69	70	71	72
VL mean ΔΔG		0.66		0.06	0.04	0.22	0.07	0.14	0.07
aa hot propensity		0.62		0.19	0.07	0.07	0.58	0.37	0.07
VL anti-VEGF		F		T	S	S	L	H	S
Muller1998		1.40		0.78	0.75	0.76	0.86	0.98	0.85

L3	AHo	107	108	109	110	111	…	135	136	137	138
VL mean ΔΔG		0.19	0.10	0.96	0.66	0.33		0.59	0.19	0.81	0.00
aa hot propensity		0.22	0.22	0.66	0.07	0.19		0.33	0.51	0.74	0.19
VL anti-VEGF		Q	Q	Y	S	T		V	P	W	T
Muller1998		3.7	2.7	14	0.90	0.87		1.5	3.5	150	1.4

H1	AHo	27	28	29	30	31	32	33	…	39	40	41	42
VH mean ΔΔG		0.00		0.45	0.10	0.10	0.13	0.24		0.99	1.54	0.01	0.54
aa hot propensity		0.00		0.66	0.19	0.62	0.19	0.25		0.66	0.00	0.27	0.25
VH anti-VEGF		G		Y	T	F	T	N		Y	G	M	N
Muller1998		2.3		34	1.3	16	1.3	150		150	6.1	6.3	66

H2	AHo	57	58	59	60	61	…	65	66	67	68	69	70	71	72	73	74	75	76
VH mean ΔΔG		1.34	0.19	1.67	0.66	0.50		0.59	0.03	0.65	0.14	0.91	0.24	0.02	0.03	0.01	0.00	0.03	0.00
aa hot propensity		0.74	0.59	0.25	0.19	0.66		0.19	0.00	0.34	0.51	0.19	0.66	0.00	0.00	0.31	0.62	0.22	0.45
VH anti-VEGF		W	I	N	T	Y		T	G	E	P	T	Y	A	A	D	F	K	R
Muller1998		150	3.8	150	8.6	8.7		4.4	1.1	1.7	1.5	2.6	2.0	1.2	1.4	1.4	0.97	1.2	1.2

H3	AHo	109	110	111	112	113	114	115	…	132	133	134	135	136	137	138
VH mean ΔΔG		0.96	0.71	0.97	1.02	0.79	0.61	0.61		0.84	0.51	1.03	0.31	0.40	0.26	0.29
aa hot propensity		0.66	0.51	0.37	0.66	0.66	0.00	0.07		0.07	0.37	0.74	0.66	0.62	0.31	0.33
VH anti-VEGF		Y	P	H	Y	Y	G	S		S	H	W	Y	F	D	V
Muller1998		150	38	4.1	3.8	4.8	1.8	0.7		150	2.4	150	19	25	1.9	1.3

Fig. 7. Analysis of an anti-VEGF paratope. Data were generated using the spreadsheet "complexes.xlsm" supplied as Supplementary Material. The six regions corresponding to the six CDRs and mutagenized in Muller's publication [38] are shown. First line: AHo numbering. A black background indicates Kabat's CDRs. VH mean $\Delta\Delta G$: average binding free energy (Fig. 2). Positions with a $\Delta\Delta G > 0.8$ and 0.4 kcal/mol are in red and pink, respectively. aa hot propensity: for the considered amino acid, this is the frequency at which this side chain makes an energetically favorable bond at the 15 strongest hot spot positions (Fig. 5). VH/VL anti-VEGF: sequence of the anti-VEGF Fab-12. Muller1998: fold change of the binding affinity when the position is mutated to alanine in Muller's publication. Main paratope residues (150-fold decrease in affinity) are in dark green and moderate contributors (between 5- and 66-fold) are in light green.

regions from the CATH database. Despite these differences, we share the main observation that Tyr and Trp are strongly enriched in the paratope of antibodies. However, this enrichment is larger in our case presumably because of the prominent role of these residues in the free energy of binding. There are also subtle discordances that may be due to the different sets of residues analyzed. We noted an enrichment of the two other aromatic residues (F and H), Asp, and Asn whereas Collis did not see any significant enrichment of these amino acids in the paratope. As for Tyr and Trp, this discrepancy could be attributed to the important contribution to the free binding energy of these side chains in our analysis.

The discontinuous nature of the paratope presented here is different from the classical CDR definition. Indeed, most of the main contributing residues are part of Kabat's hypervariable regions but the reverse is not true, since only 25% of these regions contribute to more than 70% of the binding free energy. However, many of the non-interacting residues play an important structural role in avoiding steric clashes and in shaping the paratope [9,13]. The importance of these residues in the antigen binding affinity has been demonstrated in several studies. However, they did not show up in our analysis for two main reasons. First, the used FoldX script restrained the analysis to the residues present in the interface between the antibody and the antigen. Second, the importance of these residues is highly variable between antibody molecules and we averaged the energy on the whole database of 227 structures. Such regions are clearly visible in Fig. 2 as, for example, framework positions H82–H84 but their contributions are restricted to few antibodies. Nevertheless, our study shows that the interacting residues remain broadly conserved, provided that a proper numbering of the loops is used. The important positions identified here may help in the design of several experiments, as antibody–antigen modeling and docking, affinity maturation, paratope identification, CDR and SDR grafting [43], and library design. Restraining diversity to the identified positions and using a genetic code focused toward the most important side chains should allow the construction of smaller yet efficient antibody libraries, as already demonstrated in a series of papers by Dr. Sidhu [29,44,45].

As a direct application, we show that our data may be used to quickly pinpoint important residues in an antibody paratope. This can be then experimentally confirmed but with only few mutations. For instance, as shown in Fig. 7 in the case of an anti-VEGF Fab, 7 of the 8 most important residues of the paratope could have been identified using only 12 mutations and an almost complete description of the paratope would have been obtained by testing only 27 positions

instead of the 68 mutations necessary to cover the six CDRs (Supplementary Fig. 6). This information could be subsequently used to evolve an antibody for better affinity using different techniques or assumptions. For instance, the identified positions could be targeted first for optimization using mutagenesis and selection or, on the opposite, these positions can be kept unchanged and only surrounding residues mutagenized [46–49]. A spreadsheet is available as a supplementary material (complexes.xlsm) to help the reader in aligning his own antibody sequence with our data.

Many of the results presented here are dependent on the numbering scheme used. The relative benefits of the different schemes have been discussed elsewhere [50] and we decided to use the AHo scheme [17] for two main reasons. First, length variations in CDR1 are accommodated with two gaps, separated by the conserved hydrophobic residue 31. Second, insertions and deletions are placed symmetrically around a key position (marked in yellow in Supplementary File complexes.xlsm), whereas in the other numbering schemes, insertions are growing unidirectionally. These modifications result in a better alignment of the loops and ensure that structurally equivalent residues are identically numbered not only in the FRs but also in the CDRs. However, long loops provide more potential interaction sites. This is indicated in Fig. 2 by the white bars that represent hot spots only present in long CDRs. These positions are located at the top of the loops and are more frequent in the L1 and H3 CDRs. Nevertheless, these contacts present in the longest loops usually add to but not substitute for the other contacts present in the shorter loops, resulting in a stronger interaction of the considered CDR with the antigen. The case of the VH CDR3 is presumably somewhat different because of the very high diversity of H3 loop structures [24], but however, some positions seem to be preferentially used in certain loop lengths (Supplementary File complexes.xlsm).

A trend in the selection of binders and library design is to expand the natural diversity to alleviate innate binding restrictions. This has been one of the main arguments in favor of synthetic and semisynthetic antibody libraries over natural ones to derive antibodies against self-proteins. However, there is no experimental evidence showing that such synthetic libraries are indeed more capable of generating such binders than natural repertoires, and several anti-self antibodies have been successfully selected from libraries made from human donors. We directly tested the possibility of escaping the natural antibody mode of binding by selecting an scFv from a synthetic and strongly biased library and solving the structure of the complex. Contrary to usual expectations, most of the binding free energy (68%) is due to non-diversified positions. The antibody almost perfectly fulfilled the rules defined above since most of the interacting residues were located at the hot spot positions identified in this study, and the CDR contributions were similar to the averaged values obtained from the database. In a recent study, Persson *et al.* isolated a series of 10 clones sharing the same H3 sequence despite interacting with different antigens [51]. As in our case, they showed that, in 80% of these clones, several conserved invariant H3 side chains interacted with the antigen. They also isolated two antibodies in which the H3 loop was not involved in the interaction with the antigen, in good agreement with the frequency of 10% of such antibodies in our database (Fig. 4c). The main conclusion of their study is that antibody libraries can be designed to give a dominant role to the L3 loop. Although this remains true but taken individually, such antibodies exist in nature. Indeed, as shown in Fig. 4a, in some antibody–antigen complexes, the L3 loop contributes to more than 40% of the binding energy. This shows again that the described rules are due to antibody and antigen structural constraints and cannot be easily escaped even with highly biased and unnatural libraries.

In summary, our results demonstrate the parsimonious approach followed by nature to introduce diversity in antibody paratopes. Restraining the diversity to key positions and using a restricted set of side chains are presumably required since the immune system cannot sample the large number of potential diversity represented by random sequences in the six CDRs (about 20^{60}–10^{78}). However, we showed that this is also true in the case of a synthetic antibody repertoire. Our analysis adds to the understanding of antibody–antigen interactions and may help the future design of improved antibodies and libraries with optimized functional diversities.

Materials and Methods

Generating the set of antibody–antigen complexes

The structures of the antibody/antigen complexes from the PDB were retrieved from the IMGT/3Dstructure-DB database (IMGT®, the international ImMunoGeneTics information system®[†]) [52]. A total of 506 PDB files of immunoglobulin antibody or antibody fragments in complex either with protein or with peptide ligands were retrieved from the database (data from May 2011). We extracted 484 VH, 404 VL-κ, and 59 VL-λ sequences from these files. The difference between the number of PDB files and antibody sequences extracted is due to files containing Fc domains only (lacking variable chains), which we removed from the set. VH sequences were clustered using program cd-hit [53] with default parameters and a 92% identity cutoff value, and the sequence

with the best resolution was kept in each cluster, resulting in 238 VH, 183 VL-κ, and 30 VL-λ. All files were renumbered using AHo numbering scheme [17] that minimizes the average deviation of residues in aligned three-dimensional structures of immunoglobulin domains, using a set of already aligned variable domains present in the original publication [17], clustalw software, and in-house python scripts. Unconventional sequences that were incompatible with AHo numbering were discarded resulting in 234 VH, 183 VL-κ, and 30 VL-λ sequences.

When multiple molecules were present in the crystal asymmetric unit, only a single copy of the variable domain was kept and renamed "H", "L", or "K" for VH, VL-λ, and VL-κ, respectively (in-house python script). Ligand interacting chains were defined as possessing at least one contact with antibody CDR residues using NCONT program from CCP4 suite [54]. Structures devoid of any ligand chain interacting with the variable domains were discarded. Inspection of these particular files showed that either the ligands were haptens mis-annotated as peptide in the IMGT (thus ignored by NCONT) or the interaction was through the Fc domain. The final antibody–antigen database consisted of 227 VH, 175 VL-κ, and 31 VL-λ, including the structure solved in this study.

Binding free energy calculations using FoldX

Prior to change in binding free energy calculations, all complexes were optimized using the repair function of FoldX 3.0beta6 [22]. Following optimization of the complexes, we computed the free energies of each of the antibody side chains contributing to the antibody–antigen interactions using the Complex_alascan command of the FoldX program, whereby each position (except glycine and alanine) of a given antibody–antigen interface was truncated to alanine and the positions of the neighboring side chains were optimized. The result provided was the difference in binding free energy $\Delta\Delta G$, in kilocalories per mole, between the "mutant" and the "wild-type" structures, corresponding to the binding free energy of each of the antibody side chain present in the interface and contributing to the antibody–antigen interaction (intermolecular bonds).

Data collection, structure determination, and refinement

The proteins were expressed and purified as previously described [31]. Crystals of the complexes of Gankyrin and scFv F5 were obtained at 17 °C by vapor diffusion in sitting drops by mixing 0.2 μl of the protein solution and 0.2 μl of reservoir solution containing 0.1 M Hepes (pH 7.0) and 8% polyethylene glycol 8000. The crystal was mounted in a fiber loop and flash-cooled in liquid nitrogen after cryoprotection with the reservoir solution plus 15% ethylene glycol. Data collection from a frozen single crystal was performed at 100 K on the beamline ID29 of the European Synchrotron Radiation Facility (Grenoble, France). The crystal belongs to the trigonal space group $P3_221$, with one Gankyrin–scFv F5 complex per asymmetric unit. The data were integrated and scaled using HKL2000 [55] (see statistics in Table 1). The structures were solved by molecular replacement using the program AMoRe [56]. The structure of Gankyrin and a neutralizing antibody F10 [PDB ID: 1UOH and PDB ID: 3FKU (chain X), respectively] were used as starting models [34,57]. Refinement involved iterative cycles of manual building and refinement calculations. The programs phenix.refine [58] and Coot [59] were used throughout structure determination and refinement. Several terminal residues and 16 residues (119–134) of the linker of the scFv F5 are not modeled as electron density map was poor in the corresponding regions. Anisotropic scaling, a bulk solvent correction, and TLS restrains were used. Five TLS groups for Gankyrin and four groups for scFv F5 were generated by using the program TLSMD [60]. Individual *B* atomic factors were refined isotropically. Solvent molecules were then placed according to unassigned peaks in the electron density map. In the Gankyrin–scFv F5 complex, refined at 2.5 Å with no σ cutoff, the final model contains 224 residues (3–226) for Gankyrin, 225 residues (3–119 and 134–244) for antibody scFv F5, and 428 water molecules. Free R value was calculated as described previously [61]. Structural figures were generated by using the program PyMOL [62].

Accession numbers

The Worldwide PDB accession number of the scFv F5–Gankyrin structure is PDB ID: 4NIK.

Acknowledgments

This work and the salary of G.R. were supported by the SELAS Docteur l'Hermite Collaboration Agreement No. 10561A10. Authors thank Labex MabImprove for its continuous support (French National Research Agency under the program "Investissements d'Avenir" Grant Agreement LabEx MAbImprove: ANR-10-LABX-53), François Stricher for his help with FoldX, and Pierre Poussin-Courmontagne (Structural Biology and Genomics Platform, Institut Génétique Biologie Moléculaire Cellulaire) for help in crystallization. N.R. thanks support by institutional funds from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Université de Strasbourg, Instruct, part of the European Strategy Forum on Research Infrastructures, and by national member subscriptions, the French Infrastructure for Integrated Structural Biology. Authors thank Piona Dariavach for critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.08.013.

Received 28 May 2014;
Received in revised form 31 July 2014;
Accepted 9 August 2014
Available online 29 August 2014

Keywords:
protein–protein interaction;
X-ray crystal structure analysis;
binding free energy;
antibody library

†http://www.imgt.org.

Abbreviations used:
scFv, single chain Fv; FR, framework region; CDR, complementarity-determining region; PDB, Protein Data Bank; GkF5, Gankyrin–scFv F5.

References

[1] Newman MA, Mainhart CR, Mallett CP, Lavoie TB, Smith-Gill SJ. Patterns of antibody specificity during the BALB/c immune response to hen eggwhite lysozyme. J Immunol 1992;149:3260–72.
[2] Chen S-WW, Van Regenmortel MHV, Pellequer J-L. Structure-activity relationships in peptide-antibody complexes: implications for epitope prediction and development of synthetic peptide vaccines. Curr Med Chem 2009;16:953–64.
[3] Dall'Acqua W, Goldman ER, Eisenstein E, Mariuzza RA. A mutational analysis of the binding of two different proteins to the same antibody. Biochemistry 1996;35:9667–76.
[4] James LC, Roversi P, Tawfik DS. Antibody multispecificity mediated by conformational diversity. Science 2003;299:1362–7.
[5] Johnson G, Wu TT. Kabat Database and its applications: 30 years after the first variability plot. Nucleic Acids Res 2000;28:214–8.
[6] Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, et al. IMGT/LIGM-DB, the IMGT(R) comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. Nucleic Acids Res 2006;34:D781–4.
[7] Jones PT, Dear PH, Foote J, Neuberger MS, Winter G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature 1986;321:522–5.
[8] Ewert S, Honegger A, Plückthun A. Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering. Methods 2004;34:184–99.
[9] Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 1997;273:927–48.
[10] Kunik V, Peters B, Ofran Y. Structural consensus among antibodies defines the antigen binding site. PLoS Comput Biol 2012;8:e1002388.
[11] Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Acqua W, et al. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. Proc Natl Acad Sci USA 1994;91:1089–93.
[12] Collis AVJ, Brouwer AP, Martin ACR. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J Mol Biol 2003;325:337–54.
[13] MacCallum RM, Martin AC, Thornton JM. Antibody-antigen interactions: contact analysis and binding site topography. J Mol Biol 1996;262:732–45.
[14] Lee M, Lloyd P, Zhang X, Schallhorn JM, Sugimoto K, Leach AG, et al. Shapes of antibody binding sites: qualitative and quantitative analyses based on a geomorphic classification scheme. J Org Chem 2006;71:5082–92.
[15] Dunbar J, Fuchs A, Shi J, Deane CM. ABangle: characterising the VH-VL orientation in antibodies. Protein Eng Des Sel 2013;26:611–20.
[16] Shiroishi M, Tsumoto K, Tanaka Y, Yokota A, Nakanishi T, Kondo H, et al. Structural consequences of mutations in interfacial Tyr residues of a protein antigen-antibody complex. The case of HyHEL-10-HEL. J Biol Chem 2007;282:6783–91.
[17] Honegger A, Plückthun A. Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. J Mol Biol 2001;309:657–70.
[18] Li Y, Urrutia M, Smith-Gill SJ, Mariuzza RA. Dissection of binding interactions in the complex between the anti-lysozyme antibody HyHEL-63 and its antigen. Biochemistry 2003;42:11–22.
[19] Ofran Y, Schlessinger A, Rost B. Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes. J Immunol 2008;181:6230–5.
[20] Sela-Culang I, Kunik V, Ofran Y. The structural basis of antibody-antigen recognition. Front Immunol 2013;4:302.
[21] Dall'Acqua W, Goldman ER, Lin W, Teng C, Tsuchiya D, Li H, et al. A mutational analysis of binding interactions in an antigen–antibody protein–protein complex. Biochemistry 1998;37:7981–91.
[22] Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX Web server: an online force field. Nucleic Acids Res 2005;33:W382–8.
[23] Vajdos FF, Adams CW, Breece TN, Presta LG, de Vos AM, Sidhu SS. Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. J Mol Biol 2002;320:415–28.
[24] Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. J Mol Biol 1998;275:269–94.
[25] Tsai CJ, Nussinov R. Hydrophobic folding units at protein–protein interfaces: implications to protein folding and to protein–protein association. Protein Sci 1997;6:1426–37.
[26] Zhou H-X, Qin S. Interaction-site prediction for protein complexes: a critical assessment. Bioinformatics 2007;23:2203–9.
[27] González-Muñoz A, Bokma E, O'Shea D, Minton K, Strain M, Vousden K, et al. Tailored amino acid diversity for the evolution of antibody affinity. MAbs 2012;4:664–72.
[28] Ramaraj T, Angel T, Dratz EA, Jesaitis AJ, Mumey B. Antigen–antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures. Biochim Biophys Acta Proteins Proteomics 2012;1824:520–32.
[29] Birtalan S, Zhang Y, Fellouse FA, Shao L, Schaefer G, Sidhu SS. The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. J Mol Biol 2008;377:1518–28.
[30] Rinaldi A-S, Freund G, Desplancq D, Sibler A-P, Baltzinger M, Rochel N, et al. The use of fluorescent intrabodies to detect endogenous gankyrin in living cancer cells. Exp Cell Res 2013;319:838–49.
[31] Desplancq D, Rinaldi A-S, Stoessel A, Sibler A-P, Busso D, Oulad-Abdelghani M, et al. Single-chain Fv fragment antibodies selected from an intrabody library as effective

mono- or bivalent reagents for *in vitro* protein detection. J Immunol Methods 2011;369:42–50.
[32] Philibert P, Stoessel A, Wang W, Sibler A-P, Bec N, Larroque C, et al. A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm. BMC Biotechnol 2007;7:81.
[33] Martineau P, Jones P, Winter G. Expression of an antibody fragment at high levels in the bacterial cytoplasm. J Mol Biol 1998;280:117–27.
[34] Krzywda S, Brzozowski AM, Higashitsuji H, Fujita J, Welchman R, Dawson S, et al. The crystal structure of gankyrin, an oncoprotein found in complexes with cyclin-dependent kinase 4, a 19S proteasomal ATPase regulator, and the tumor suppressors Rb and p53. J Biol Chem 2004; 279:1541–5.
[35] Nakamura Y, Nakano K, Umehara T, Kimura M, Hayashizaki Y, Tanaka A, et al. Structure of the oncoprotein gankyrin in complex with S6 ATPase of the 26S proteasome. Structure 2007;15:179–89.
[36] Brunger AT. Version 1.2 of the Crystallography and NMR system. Nat Protoc 2007;2:2728–33.
[37] Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr Sect D Biol Crystallogr 1998;54:905–21.
[38] Muller YA, Chen Y, Christinger HW, Li B, Cunningham BC, Lowman HB, et al. VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 Å resolution and mutational analysis of the interface. Structure 1998;6:1153–67.
[39] Da Silva GF, Harrison JS, Lai JR. Contribution of light chain residues to high affinity binding in an HIV-1 antibody explored by combinatorial scanning mutagenesis. Biochemistry 2010;49:5464–72.
[40] Goldman ER, Dall'Acqua W, Braden BC, Mariuzza RA. Analysis of binding interactions in an idiotope–antiidiotope protein–protein complex by double mutant cycles†. Biochemistry 1997;36:49–56.
[41] Fields BA, Goldbaum FA, Ysern X, Poijak RJ, Mariuzza RA. Molecular basis of antigen mimicry by an anti-idiotope. Nature 1995;374:739–42.
[42] Zhou T, Georgiev I, Wu X, Yang Z-Y, Dai K, Finzi A, et al. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science 2010;329:811–7.
[43] Gonzales NR, Padlan EA, De Pascalis R, Schuck P, Schlom J, Kashmiri SVS. SDR grafting of a murine antibody using multiple human germline templates to minimize its immunogenicity. Mol Immunol 2004;41:863–72.
[44] Fellouse FA, Li B, Compaan DM, Peden AA, Hymowitz SG, Sidhu SS. Molecular recognition by a binary code. J Mol Biol 2005;348:1153–62.
[45] Fellouse FA, Esaki K, Birtalan S, Raptis D, Cancasci VJ, Koide A, et al. High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. J Mol Biol 2007;373:924–40.
[46] Chen Y, Wiesmann C, Fuh G, Li B, Christinger HW, McKay P, et al. Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen. J Mol Biol 1999;293:865–81.
[47] Chowdhury PS, Pastan I. Improving antibody affinity by mimicking somatic hypermutation *in vitro*. Nat Biotechnol 1999;17:568–72.
[48] Barderas R, Desmet J, Timmerman P, Meloen R, Casal JI. Affinity maturation of antibodies assisted by *in silico* modeling. Proc Natl Acad Sci USA 2008;105:9029–34.
[49] Yang WP, Green K, Pinz-Sweeney S, Briones AT, Burton DR, Barbas CF. CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. J Mol Biol 1995;254:392–403.
[50] Martin ACR. Protein Sequence and Structure Analysis of Antibody Variable Domains. In: Kontermann R, Dübel S, editors. Antibody Engineering. Heidelberg: Springer Berlin; 2010. p. 33–51.
[51] Persson H, Ye W, Wernimont A, Adams JJ, Koide A, Koide S, et al. CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. J Mol Biol 2013;425:803–11.
[52] Ehrenmann F, Lefranc M-P. IMGT/3Dstructure-DB: querying the IMGT database for 3D structures in immunology and immunoinformatics (IG or antibodies, TR, MH, RPI, and FPIA). Cold Spring Harbor Protoc 2011;2011:750–61.
[53] Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a Web server for clustering and comparing biological sequences. Bioinformatics 2010;26:680–2.
[54] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D Biol Crystallogr 2011;67:235–42.
[55] Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter CW, editor. Methods in Enzymology, vol. 276. Fribourg: Academic Press; 1997. p. 307–26.
[56] Navaza J. AMoRe: an automated package for molecular replacement. Acta Crystallogr Sect A Found Crystallogr 1994;50:157–63.
[57] Sui J, Hwang WC, Perez S, Wei G, Aird D, Chen L, et al. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Nat Struct Mol Biol 2009;16:265–73.
[58] Afonine P, Grosse-Kunstleve R, Adams P, et al. The Phenix refinement framework. CCP4 Newsletter on Protein Crystallography 2005;42:8.
[59] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2126–32.
[60] Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr Sect D Biol Crystallogr 2006;62: 439–50.
[61] Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 1992; 355:472–5.
[62] DeLano W. The PyMOL Molecular Graphics System. at http://www.pymol.org; 2002.