

Cation- π , amino- π , π - π , and H-bond interactions stabilize antigen-antibody interfaces

Georgios A. Dalkas, Fabian Teheux, Jean Marc Kwasigroch, and Marianne Rooman*

Department of BioModeling, BioInformatics & BioProcesses, Université Libre de Bruxelles, CP 165/61, 1050 Brussels, Belgium

ABSTRACT

The identification of immunogenic regions on the surface of antigens, which are able to stimulate an immune response, is a major challenge for the design of new vaccines. Computational immunology aims at predicting such regions—in particular B-cell epitopes—but is far from being reliably applicable on a large scale. To gain understanding into the factors that contribute to the antigen-antibody affinity and specificity, we perform a detailed analysis of the amino acid composition and secondary structure of antigen and antibody surfaces, and of the interactions that stabilize the complexes, in comparison with the composition and interactions observed in other heterodimeric protein interfaces. We make a distinction between linear and conformational B-cell epitopes, according to whether they consist of successive residues along the polypeptide chain or not. The antigen-antibody interfaces were shown to differ from other protein-protein interfaces by their smaller size, their secondary structure with less helices and more loops, and the interactions that stabilize them: more H-bond, cation- π , amino- π , and π - π interactions, and less hydrophobic packing; linear and conformational epitopes can clearly be distinguished. Often, chains of successive interactions, called cation/amino- π and π - π chains, are formed. The amino acid composition differs significantly between the interfaces: antigen-antibody interfaces are less aliphatic and more charged, polar and aromatic than other heterodimeric protein interfaces. Moreover, paratopes and epitopes—albeit to a lesser extent—have amino acid compositions that are distinct from general protein surfaces. This specificity holds promise for improving B-cell epitope prediction.

Proteins 2014; 82:1734–1746.

© 2014 The Authors. Proteins: Structure, Function, and Bioinformatics Published by Wiley Periodicals, Inc.

Key words: immunoinformatics; antigen-antibody interactions; protein-protein interactions; B-cell epitope; paratope; salt bridge; hydrogen bond; hydrophobic contacts; cation- π interactions; π - π interactions.

INTRODUCTION

Humans and other animals are constantly exposed to a wide variety of pathogens and infectious agents, and the role of the immune system is to protect them against these infections. A crucial element of the humoral immune responses is played by the B-cells, which secrete antibodies against antigens. Antibodies bind to antigens at sites called B-cell epitopes. The identification and characterization of epitope regions on the antigen surface, which are capable of inducing an efficient immune response, is one of the key goals for the development of new vaccines.¹

B-cell epitopes are classified into two classes: linear or continuous epitopes, and conformational or discontinuous epitopes, depending on whether the amino acids included in the epitope are contiguous in the polypeptide chain or not.² Specifically, linear epitopes are composed of residues that correspond to a single segment of the amino acid sequence of a protein, while conforma-

tional epitopes consist of residues from different regions of the sequence, that come spatially close upon protein folding. Although the distinction between these two categories of epitopes is not totally well-defined, since linear epitopes may sometimes be interrupted by a few residues that are not in contact with the antibody, we may nevertheless state that the majority of B-cell epitopes are conformational (about 90%).³

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License which permits use, distribution and reproduction in any medium, provided that the Contribution is properly cited and is not used for commercial purposes.

Additional Supporting Information may be found in the online version of this article.

Marianne Rooman is Research Director at the Belgian Fund for Scientific Research (FNRS).

*Correspondence to: M. Rooman, Department of BioModeling, BioInformatics & BioProcesses, Université Libre de Bruxelles, 50 Roosevelt Ave, CP 165/61, 1050 Brussels, Belgium. E-mail: mrooman@ulb.ac.be or <http://babylone.ulb.ac.be>

Received 21 November 2013; Revised 23 January 2014; Accepted 28 January 2014
Published online 1 February 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24527

As experimental approaches for B-cell epitope identification⁴ are highly time, money, and resource consuming, the use of alternative, computational, strategies is increasingly used for facilitating epitope recognition. More precisely, these computational techniques are utilized for proposing, on a rational basis, a small number of potential epitope candidates, which remain to be validated experimentally. Several computational methods have been developed in recent years,⁵ for both linear and conformational B-cell epitopes.

The majority of available *in silico* methods for predicting B-cell immunogenic regions focus on linear epitopes, and are based on several amino acid-based propensity scales, including hydrophilicity, solvent accessibility, secondary structure, flexibility, and antigenicity.^{6–11} The predictive performance of approaches using a single propensity scale has been systematically assessed on the basis of the correlation between the score peaks obtained from 484 amino acid propensity scales and the epitope location within a data set of 50 proteins.¹² It was found that even the best set of scales and parameters performed only marginally better than random. In addition, several groups investigated the combination of multiple amino acid propensity scales to predict linear B-cell epitopes, such as PREDITOP,¹³ BEPITOPE,¹⁴ PEOPLE,¹⁵ and BcePred,¹⁶ with no significant improvement in the prediction success rate. This suggests the need for designing more sophisticated approaches for predicting linear B-cell epitopes.

Recently, machine learning methods, such as hidden Markov models, artificial neural networks, and support vector machines, have been used in developing new approaches for linear B-cell epitope predictions.^{17–24} These methods yield better, but still insufficient, performances.

With the increase in publicly available, good resolution, and antigen–antibody crystal structures, it has become possible to perform reliable structural analyses and to devise structure-based B-cell epitope predictions. However, the number of such analyses remains limited.^{25–28} Several groups have investigated various physicochemical, structural, and geometrical features of epitopes in order to determine which of them significantly distinguish epitope from nonepitope antigen residues.^{29–36} Unfortunately, the prediction performance of these approaches is not yet satisfactory.

Most of the existing prediction tools are based on studies indicating that protein–protein interfaces differ in their features (e.g., amino acid composition) from the remaining protein surface.^{37,38} It is also generally assumed that antigen–antibody interfaces are different from other protein–protein interfaces.^{39,40} However, it is often stated that no difference is seen between epitopes and general protein surfaces.^{28,41–43} To further investigate these issues, we compiled nonredundant data sets of experimentally resolved antigen–antibody and other heterodimeric protein complexes, and performed a compre-

hensive analysis of the differences in amino acid composition and secondary structure between antigen, antibody, and other protein surfaces, as well as between epitopes, paratopes, and other protein–protein interfaces. The observed variations were correlated with differences in residue–residue contacts across the interface, in particular salt bridges, hydrogen bonds, hydrophobic contacts, as well as cation– π , amino– π , and π – π interactions. Our goal here is to gain understanding of the specific features of B-cell epitopes that ensure their recognition by antibodies. Our next goal will be to use these features to design more efficient B-cell epitope prediction tools.

MATERIALS AND METHODS

Preparation of the data sets

Data sets of the antigen–antibody complexes

To study the antigenic properties of proteins, a preliminary data set of experimentally determined antigen–antibody structures was obtained from the structural component of the Immune Epitope Database (IEDB-3D).⁴⁴ This data set was then filtered using several criteria: (i) only the antibody–antigen complexes with resolution better than 2.5 Å were considered; (ii) for complexes represented by more than one 3D structure, the structure with the best resolution was selected as the representative structure; (iii) structures in which the antibody binds the antigen but involves no residues from complementarity determining regions (CDRs) have been excluded from the analysis; (iv) all complexes in which the antibody does not contain both the light and heavy chains were discarded; (v) if a structure contained several complexes in one asymmetric unit and there was no structural difference observed between these complexes, only one complex was selected; and (vi) to obtain a nonredundant data set, the sequences were pairwise aligned using the program ClustalW,⁴⁵ and two sequences were considered similar if their sequence identity was higher than 70%. Since antibody sequences are highly similar to each other, the similarity criterion was determined on the basis of the CDRs only. In addition, similar antigens were kept if they were in complex with different antibody CDR sequences. After this procedure, 105 antibody–antigen complexes were selected and the corresponding coordinate files were collected from the Protein Data Bank (PDB).⁴⁶

The complexes were then separated into two groups on the basis of the epitope segmentation. Epitopes (defined in the next subsections) consisting of successive residues along the sequence, with gaps of three nonepitope residues at most, were considered as linear epitopes. The others were defined as conformational epitopes. Based on this definition, antigens containing conformational epitopes were included in the data set denoted as S_{conf} (Supporting Information Table S1) and those with

linear epitopes, in the data set S_{linear} (Supporting Information Table S2). Complexes containing small antigen fragments bound to the antibody were also included in S_{linear} . With these definitions, the data sets S_{conf} and S_{linear} contain 58 and 47 antigen–antibody complexes, respectively.

Reference data set of protein–protein complexes

A data set of protein–protein heterodimers with pairwise sequence identity $\leq 20\%$ was built from the server 3D complex and denoted as S_{dimer} (Supporting Information Table S3).⁴⁷ Antigen–antibody complexes were excluded from this set, so that both sets have no overlap. This data set was used for comparison and reference.

Definition of surface residues

It is generally agreed that epitopes are located on the surface of the protein.^{25,29,48} In the present study, an amino acid was considered as exposed to the solvent if its solvent accessibility is higher than 5%. The accessibility of a residue X is defined as the sum of the accessible surface area (ASA) of its atoms, multiplied by 100 and divided by its maximal (theoretical) ASA in a GXG tripeptide in extended conformation, where G denotes glycine.⁴⁹ The 5% accessibility cutoff³⁸ can seem quite low, but ensures that all epitope residues are surface residues in the absence of the paratope.

Definition of epitopes

To determine the epitopes, we calculated the solvent accessibility values of the antigen residues without taking the antibody into account (ACC_{unbound}) and compared them with the accessibility of antigen residues in the complex (ACC_{bound}). All antigen residues with a solvent accessibility variation of 5% at least upon antibody binding ($ACC_{\text{unbound}} - ACC_{\text{bound}} \geq 5\%$) were considered as epitope residues.

Computation of solvent accessibility and secondary structure

We have developed a program to compute the solvent accessible surface and the solvent accessibility of protein atoms and residues from their Cartesian coordinates, as well as the secondary structure. It is based on the DSSP algorithm,⁵⁰ with some notable differences regarding the data used from the PDB, some definitions and running options.

In the default mode, information coming from PDB files is carefully processed to take into account only residues that are part of the polypeptide chain, which implies the omission of all ligands (even if these are natural amino acids). Furthermore, modified residues in the polypeptide chain are included, with all their side chain

atoms. Finally, alternate location and occupancy information are managed: the selected atoms are those that have the maximum electronic density. These new features allow a more accurate computation of the solvent accessible surface and secondary structure.

During computation, residues with missing atoms (even in the backbone) are taken into account to keep the results as close as possible to reality. Furthermore, when assigning the secondary structure, a maximum of two different H-bond partners per residue is imposed (one involving the CO group and the other the NH group); the partners that are closest in space are selected.

Finally, additional running options are available to add flexibility to the program:

- All atomic van der Waals radii can be user-defined, and ditto for the solvent radius; for example, using a solvent radius of zero allows the detection of atomic contacts.
- A subset of polypeptide chains can be selected; it is thus possible to compute the solvent accessibility of one monomer without taking into account its interaction with another monomer.
- The hydrogen atoms are in general reconstructed geometrically; however, the hydrogen atoms present in the PDB file can be considered instead.

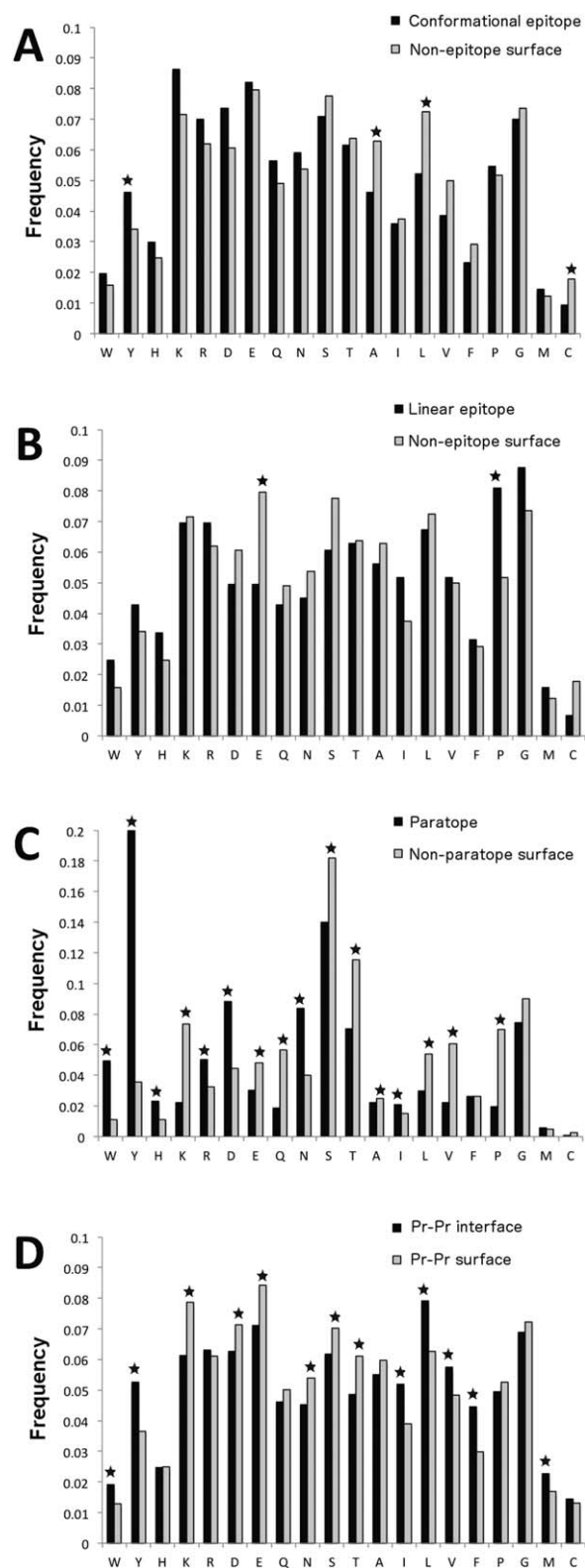
Antigen–antibody interactions

Information regarding the residues involved in hydrogen bonds and salt bridges were obtained using the web-server PISA.⁵¹ Hydrophobic interactions were calculated using the webserver Protein Interactions Calculator.⁵²

Cation– π , amino– π , and π – π interactions were obtained by in-house programs.⁵³ Cation– π and amino– π interactions link an aromatic residue (Trp, Tyr, Phe, and His) to a positively charged residue (Arg, Lys, and His) and to a residue carrying a partial positive charge on its side-chain amino group (Gln and Asn), respectively. They are defined geometrically by a distance criterion and an angle criterion. The distance criterion requires that at least one of the atoms of the aromatic ring is located no further than 4.5 Å from one of the atoms carrying the net or partial positive charge. The angle criterion requires this atom to be situated above the plane defined by the aromatic ring, more precisely, inside a cylinder of height 4.5 Å, whose base includes the ring and has a radius equal to twice the ring's radius.

The π – π interactions between two aromatic residues (Trp, Tyr, Phe, and His) were defined in much the same way, with a distance and angle criterion. The distance criterion requires that at least one of the atoms of the first aromatic ring be separated by 5 Å at most from at least one atom of the second aromatic ring. The angle criterion requires this second atom to be located above the plane defined by the first aromatic ring, more

precisely, inside a cylinder of height 5 \AA , whose base includes the ring and has a radius equal to three times the ring's radius.



Note that all the π - π interactions between the aromatic residues Tyr, Trp, and Phe are also identified as hydrophobic contacts. The converse is not true, as the angle criterion makes the π - π interactions more specific than the hydrophobic contacts. To avoid any overlap between these interactions, we exclude all π - π interactions from the set of hydrophobic interactions.

Our in-house programs for detecting cation- π , amino- π , and π - π interactions have several advantages compared to other programs. First, they are not only based on a distance criterion but also an angle criterion, which makes them more accurate. Second, amino- π interactions are in general not considered. Finally, our programs also identify chains of successive cation/amino- π and π - π interactions, which have been shown to be quite interesting in other contexts.⁵⁴

The statistical significance of the differences in calculated interaction frequencies in the different protein sets was evaluated with a *z*-test, using the common threshold of statistical significance, that is a *P* value ≤ 0.05 .

Amino acid composition

Investigation of enrichment or depletion of amino acids between epitope and non-epitope surface residues, as well as the evaluation of their statistical significance on the basis of a *t*-test and a threshold *P* value ≤ 0.05 , was carried out using the Composition Profiler server.⁵⁵ The same tool was used to analyze the antibodies and the protein-protein complexes.

RESULTS AND DISCUSSION

Amino acid composition in epitopes and antigen surfaces

A frequently investigated characteristic of the B-cell epitopes is the amino acid composition. We deepen this analysis here by comparing systematically the amino acid composition of linear epitopes, conformational epitopes, non-epitope antigen surfaces, paratopes, nonparatope antibody surfaces, other protein-protein interfaces, and noninterface surfaces. We used for that purpose a definition of surface residues that consistently includes all epitopes and paratopes, as described in Materials and Methods section. The results are given in Figure 1.

Figure 1

Amino acid frequencies of linear and conformational epitopes, paratopes, and protein-protein interfaces. (A) Conformational epitopes of set S_{conf} and (B) linear epitopes of set S_{linear} , compared with non-epitope antigen surfaces of set S_{conf} ; (C) paratopes compared with nonparatope antibody surfaces of sets S_{conf} and S_{linear} ; (D) protein-protein interfaces compared with the remaining protein surfaces of set S_{dimer} . Statistically significant differences of frequencies using a *z*-test are marked with an asterisk. The amino acids are sorted by their physical chemical features.

The first trend we observe is that the overall amino acid composition on the surfaces of antigens and heterodimeric proteins differ, but much less than the composition on the antibody surfaces, where some residues are extremely frequent or almost absent. In particular, Tyr in paratopes and Ser and Thr on the remaining antibody surface reach frequencies of 10–20%. This reflects the fact that antibodies form a class of quite particular proteins, with some conserved features, whereas antigens are more variable.

Conformational epitopes are enriched, compared to the nonepitope antigen surface, in positively and negatively charged residues (Lys, Arg, His, Glu, and Asp), in residues carrying a partially charged side chain amino group (Gln and Asn) and in the aromatic residues Trp and especially Tyr, but not Phe [Fig. 1(A)]. In contrast, they are depleted of Cys residues, and of the aliphatic amino acids Ile, Val, and especially Leu and Ala [Fig. 1(A)].

Linear B-cell epitopes differ in amino acid composition from conformational epitopes, as seen from the comparison between Figure 1(A,B): less charged residues, and especially negatively charged ones, less Asn and Gln with a partially charged amino group, much more Pro and Gly residues, and somewhat more aliphatic hydrophobic residues. The amount of aromatic residues Trp and Tyr is roughly the same.

These characteristics were compared with the amino acid composition in other types of protein heterodimers [Fig. 1(D)]. Clearly, protein–protein interfaces are depleted of polar residues, in particular Ser, Thr, Gln, and Asn and of charged residues Lys, Glu, and Asp, compared with the rest of the protein surface. In contrast, they are enriched in hydrophobic residues, in particular in the aliphatic residues Ile, Val, and Leu, and the aromatic residues Trp, Tyr, and Phe. Thus, compared with the conformational epitope regions, the protein–protein interfaces contain more hydrophobic residues and less polar and charged residues; the amount of aromatic residues remains comparable, with the notable difference that Phe is abundantly present in protein–protein but not in antigen–antibody interfaces. The case of linear epitopes is somewhat intermediate: more aliphatic and less positively charged residues than in conformational epitopes, and less hydrophobic and more positively charged residues than in protein–protein interfaces. Moreover, the linear epitopes exhibit fewer negatively charged residues and more Pro residues than all other interfaces.

Finally, the paratope amino acid composition was found to be very different from the other interfaces [Fig. 1(C)]: very few Pro and aliphatic residues, and a huge amount of Trp and Tyr aromatic residues. Moreover, paratopes contain very few Lys and Glu and many Asp residues, and a normal amount of Arg. They also contain an unusually large amount of Ser, albeit less than on the

rest of the surface. This result indicates that, though the paratope region is highly variable among antibodies, this variation is nevertheless limited. This result is not surprising at all. If it were not so, antibodies would be able to bind basically all protein surfaces, which is obviously (and fortunately for us) not the case.

These results are in overall agreement with previous results, although the exact comparison is difficult. Indeed, no other analysis has compared the composition among linear and conformational epitopes, protein–protein interfaces, and the noninterface surfaces of antigens, antibodies and heterodimeric proteins. To summarize, other groups found epitopes to be depleted in aliphatic amino acids and enriched in polar amino acids and often in the aromatic amino acids Tyr or Trp.^{25–29,56,57} Several groups have also investigated the amino acid composition in paratopes. The amino acids Asn, Trp, and especially Tyr were systematically found to be overrepresented, and so were sometimes Ser, Asp, and His according to whether the comparison was with proteins in general, protein–protein interfaces, or antibody surfaces.^{26,28,43,57–59}

A related issue is whether significant differences in amino acid composition are observed between the different types of interfaces and general protein surfaces. This question is quite important for prediction purposes. Indeed, if B-cell epitopes were indistinguishable from protein surfaces, it would be impossible to predict them on the basis of their composition. Such indistinctness has been reported earlier.^{41–43} However, the use of our nonredundant data sets, our relaxed definition of surface residues, and the differentiation between linear and conformational epitopes allowed us to identify statistically significant differences, as can be seen in Figure 2.

The differences between paratopes and general protein surfaces are clearly the most obvious ones: less Pro, less positively and negatively charged residues, less aliphatic residues, more Ser and Thr, and above all, more Trp and especially Tyr. However, conformational B-cell epitopes also present statistically significant differences compared to protein surfaces: more positively charged residues, more aromatic residues Trp and Tyr, and less of the most hydrophobic residues (aliphatic, Phe, and Met). Linear epitopes have more Pro and less negatively charged residues. Protein–protein interfaces show less negatively charged residues, less Ser, Thr, Asn, and Gln, and more aromatic and other hydrophobic residues.

Secondary structure of epitopes

To test whether the secondary structure of the epitope regions is distinct from that of the antigen surface, we assigned each amino acid to three secondary structure classes: (i) helices that group α -, 3_{10} -, and π -helices, (ii) strands, that is, isolated β -bridges and extended β -strands, and (iii) loops, consisting of turns, bends, and

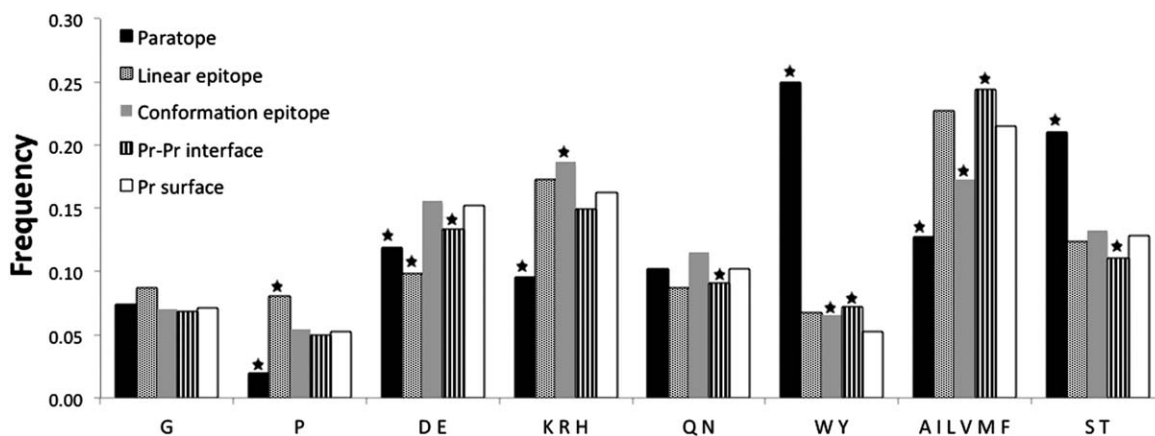


Figure 2

Amino acid frequencies in linear epitopes, conformational epitopes, paratopes, and protein–protein interfaces compared to arbitrary protein surfaces (including protein interfaces). Statistically significant differences, identified by a *z*-test, between the frequencies of a given type of interface and those of general protein surfaces are marked with an asterisk above the interface bar.

irregular structures. We compared the secondary structure composition of linear and conformational epitopes with that of the non-epitope antigen surfaces, paratopes, nonparatope antibody surfaces, protein–protein interfaces, and noninterface surfaces. The results are given in Figure 3.

Analysis of the results reveals that linear epitopes and paratopes essentially interact through loop regions. So do conformational epitopes albeit to a lesser extent. In contrast, all protein–protein interfaces, and all nonparatope, non-epitope, and noninterface surfaces have almost the same secondary structure content. The major presence of loops in the epitope–paratope interface suggests an important flexibility of these regions, in agreement with previous studies.^{25,26} This is in general not the case for other types of protein–protein interfaces.

The helix content of paratope and nonparatope antibody surfaces is very low, and their strand content very high especially on the nonparatope surface. This reflects the fact that antibodies are all- β proteins. The linear epi-

topes have quite a low content in both helices and strands, which is not surprising as they are in general too short to be very structured. The conformational epitope and non-epitope regions have a higher helical content and lower strand content than antibody surfaces and paratopes, but a lower helical content and slightly higher strand content than usual protein surfaces and protein–protein interfaces. So, antigens may be α -, $\alpha\beta$ -, and β -proteins, but are more often β -proteins than average.

Analysis of the epitope–paratope interactions

All the interactions across the epitope–paratope interface, that is, salt bridges, hydrogen bonds, hydrophobic interactions, disulfide bonds, cation– π , amino– π , and π – π contacts, were recorded and analyzed. We considered three types of interfaces: (i) conformational epitope–paratope interfaces, (ii) linear epitope–paratope interfaces, and (iii) protein–protein interfaces. The detailed results

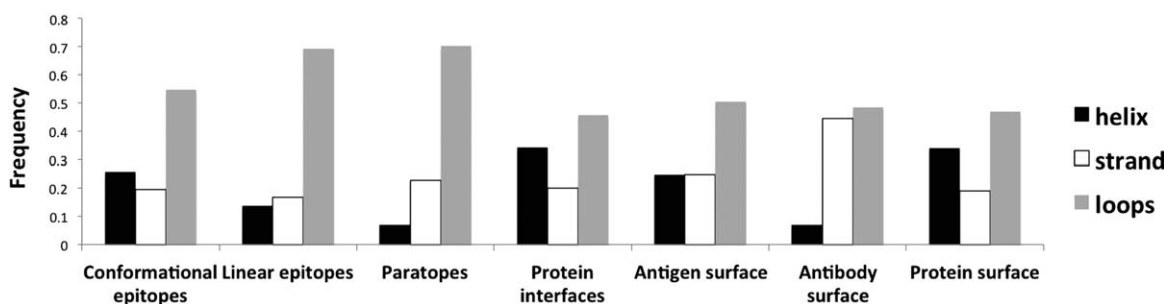


Figure 3

Secondary structure content of conformational and linear epitopes, paratopes, protein–protein interfaces, antigen (non-epitope) surfaces, antibody (nonparatope) surfaces, and protein (noninterface) surfaces.

Table I
Frequency of Different Types of Interaction Across the Interfaces

| Interactions | Interface | | | | | | Statistical significance ^a |
|-----------------------------|---------------------------------|---------------|-------------------------|---------------|-----------------|---------------|---------------------------------------|
| | Conformational epitope–paratope | | Linear epitope–paratope | | Protein–protein | | |
| | Occurrence | Frequency (%) | Occurrence | Frequency (%) | Occurrence | Frequency (%) | |
| 1. Salt bridges | 99 | 7.3 | 74 | 9.0 | 410 | 7.4 | ○■△ |
| 2. Hydrogen bonds | 661 | 48.7 | 356 | 43.2 | 1888 | 34.2 | ○■△ |
| 3. Hydrophobic ^b | 468 | 34.5 | 334 | 40.5 | 2970 | 53.9 | ○■△ |
| 4. Cation/amino- π | 75 | 5.5 | 32 | 3.9 | 102 | 1.9 | ■△ |
| 5. π - π contacts | 54 | 4.0 | 29 | 3.5 | 134 | 2.4 | ■ |
| 6. Disulfide bonds | 0 | 0 | 0 | 0 | 9 | 0.2 | |
| Total | 1357 | | 825 | | 5513 | | |

^aThe symbols ○, ■, and △ stand for the statistical significance, computed with the z-test, of the difference in average interaction frequency between conformational and linear epitopes, conformational epitopes and protein–protein interfaces, and linear epitopes and protein–protein interfaces, respectively.

^b π - π contacts are not included.

for all the complexes are listed in Supporting Information Tables S4–S6, and a summary of the interactions across the interfaces is shown in Tables (I–V).

The most frequent interactions across the interfaces are obviously hydrogen bonds and hydrophobic contacts, which represent together about 85% of all interactions (Table I). These interactions thus bring the largest contribution to the stability of the complexes. However, the relative weight of the hydrophobic and H-bond contacts vary among the complexes, as well as the contribution of the other types of interactions which, albeit less frequent, add to the specific recognition of the complexes. We will focus on these differences in what follows.

The differences in contact preferences between the three types of interfaces are seen in Table I. A significant preference for hydrophobic contacts is observed in linear epitope–paratope interfaces and even more in protein–protein interfaces, when compared with conformational epitope–paratope interfaces. Protein–protein interfaces are thus much more stabilized by hydrophobic contacts than conformational epitope–paratope interfaces; linear epitope–paratope interfaces are intermediary.

In contrast, hydrogen bonds, cation/amino- π , and π - π interactions are found to be significantly more frequent

in conformational epitope–paratope interfaces, and less frequent in protein–protein interfaces; again linear epitope–paratope interfaces are intermediary.

Disulfide bonds are never observed across both conformational and linear epitope–paratope interfaces, and only very rarely across protein–protein interfaces.

Salt bridges are observed in all types of interfaces, but are somewhat more frequent in linear epitope–paratope interfaces.

Another important difference, shown in Table II, is the size of the interfaces. Obviously, the smallest interfaces are the linear epitopes with 9 residues on the average, followed by the conformational epitopes with 18 residues on the average. The protein–protein interfaces are much larger and involve 29 residues on the average. Note that the standard deviation on the latter value is equal to 20 and is thus very high, which indicates the large variability in size of the heterodimeric protein interfaces.

Furthermore, the residues that are part of linear epitopes make many more contacts across the interface than those in conformational epitopes and in heterodimeric protein interfaces (Table II). Indeed, the number of interactions per residue is equal to 1.9 for the linear epitopes, and to 1.3 for the conformational epitopes and protein

Table II
Average Number of Interactions Per Interface Residue, Multiplied by 100

| | N^a | Salt bridges | H-bonds | Hydrophobic ^b | Cation/ amino- π | π - π | Disulfide bonds | Cation- π chains | π - π chains | R^c |
|---------------------------|-------------|--------------|---------|--------------------------|-------------------------|---------------|--------------------|-------------------------|----------------------|-------|
| Conformational epitopes | 17.6 (9.0) | 8.7 | 57.8 | 40.9 | 6.6 | 4.7 | 0 | 1.7 | 7.4 | 1.28 |
| Linear epitopes | 9.4 (2.2) | 16.7 | 80.6 | 75.6 | 7.2 | 6.6 | 0.0 | 2.1 | 4.3 | 1.93 |
| Proteins | 28.6 (20.4) | 9.8 | 44.9 | 70.6 | 2.4 | 3.2 | 0.2 | 0.3 | 2.1 | 1.34 |
| Significance ^d | | ○△ | ○■△ | ○■△ | ■△ | ■△ | | ■△ | ○■△ | |

^aAverage number of residues per interface, with the standard deviation in parentheses.

^b π - π contacts are not included.

^cAverage number of contacts per residue

^dThe symbols ○, ■, and △ stand for the statistical significance, computed with the z-test, of the difference in mean number of interactions per interface residue between conformational and linear epitopes, conformational epitopes and protein–protein interfaces, and linear epitopes and protein–protein interfaces, respectively.

Table III
Frequency of Salt Bridge Interactions Across the Interfaces

| Conformational epitope–paratope | | | | Linear epitope–paratope | | | | Protein–protein | | | |
|---------------------------------|----|------------|----|-------------------------|----|------------|----|-----------------|----|------------|----|
| Ag | Ab | Occurrence | % | Ag | Ab | Occurrence | % | Pr | Pr | Occurrence | % |
| R | D | 29 | 29 | K | D | 23 | 31 | R | E | 115 | 28 |
| K | D | 17 | 17 | R | D | 15 | 20 | R | D | 101 | 25 |
| E | R | 12 | 12 | E | R | 8 | 11 | K | E | 86 | 21 |
| R | E | 10 | 10 | D | R | 7 | 9 | K | D | 73 | 18 |
| D | R | 10 | 10 | R | E | 4 | 5 | H | E | 19 | 5 |
| D | K | 4 | 4 | D | K | 3 | 4 | H | D | 16 | 4 |
| E | H | 4 | 4 | E | H | 3 | 4 | | | | |
| H | D | 4 | 4 | H | D | 3 | 4 | | | | |
| K | E | 4 | 4 | K | E | 3 | 4 | | | | |
| E | K | 3 | 3 | D | H | 2 | 3 | | | | |
| D | H | 2 | 2 | E | K | 2 | 3 | | | | |
| H | E | 0 | 0 | H | E | 1 | 1 | | | | |

interfaces. This can be understood as follows: the linear epitopes are small peptides, which do not possess a core. All their residues try therefore to make contacts with the paratope, whereas in conformational epitopes and other protein interfaces, part of the interface residues make stabilizing contacts with the core. This implies that the number of interactions per residue—that is, of salt bridges, H-bonds, hydrophobic contacts, cation/amino- π , and π - π interactions—is higher in linear epitopes than in other interfaces.

In Table II, the number of cation/amino- π and of π - π chains, consisting of several successive cation/amino- π or π - π interactions, respectively, is also indicated. Clearly, the frequency of this type of interactions is higher in conformational and linear epitopes than in heterodimeric complexes. An illustration of these interactions in antigen–antibody interfaces is shown in Figure 4.

We detail hereunder the different types of interactions across the interface.

Salt bridges

All types of salt bridges across the different interfaces are listed in Table III. Strikingly, the positively charged amino acid is located in the antigen and the negatively charged in the antibody in the majority of the cases, that is, 65% for both conformational and linear epitopes. This result is consistent with the finding that the epitope is enriched in positively charged residues and the paratope in negatively charged residues (Fig. 1).

The Arg-Asp pair constitutes the most frequent salt-bridge pair in conformational epitopes (29% of the pairs) and Lys-Asp the most frequent pair in linear epitopes (31%). Considering the Arg-Asp and Asp-Arg pairs together yields a frequency of 39% in conformational epitopes and 29% in linear epitopes, whereas this pair constitutes only 25% in protein–protein interfaces. The combined Lys-Asp and Asp-Lys pairs exhibit a frequency as high as 35% in linear epitopes, 21% in conformational epitopes, and only 18% in protein–protein interfaces.

Table IV
Frequency of π - π Interactions Across the Interfaces

| Conformational epitope–paratope | | | | Linear epitope–paratope | | | | Protein–protein | | | |
|---------------------------------|----|------------|----|-------------------------|----|------------|----|-----------------|----|------------|----|
| Ag | Ab | Occurrence | % | Ag | Ab | Occurrence | % | Pr | Pr | Occurrence | % |
| H | Y | 10 | 19 | Y | W | 9 | 31 | F | Y | 33 | 25 |
| F | Y | 8 | 15 | F | W | 3 | 10 | F | F | 20 | 15 |
| H | W | 8 | 15 | F | Y | 3 | 10 | F | H | 19 | 14 |
| Y | Y | 7 | 13 | W | F | 2 | 7 | F | W | 17 | 13 |
| W | Y | 5 | 9 | H | H | 2 | 7 | W | Y | 13 | 10 |
| F | F | 3 | 6 | H | W | 2 | 7 | H | Y | 11 | 8 |
| H | F | 3 | 6 | W | W | 2 | 7 | Y | Y | 8 | 6 |
| W | W | 2 | 4 | Y | Y | 2 | 7 | H | W | 5 | 4 |
| Y | H | 2 | 4 | H | F | 1 | 3 | W | W | 4 | 3 |
| Y | W | 2 | 4 | Y | H | 1 | 3 | H | H | 3 | 2 |
| F | H | 1 | 2 | H | Y | 1 | 3 | | | | |
| F | W | 1 | 2 | W | Y | 1 | 3 | | | | |
| W | F | 1 | 2 | | | | | | | | |
| W | H | 1 | 2 | | | | | | | | |

Table V
Frequency of Cation- π and Amino- π Interactions Across the Interfaces

| Conformational epitope-paratope | | | | Linear epitope-paratope | | | | Protein-protein | | | |
|---|----|------------|----|-------------------------|----|------------|----|-----------------|----|------------|----|
| Ag | Ab | Occurrence | % | Ag | Ab | Occurrence | % | Pr | Pr | Occurrence | % |
| Cation-π interactions | | | | | | | | | | | |
| K | Y | 11 | 23 | K | Y | 8 | 36 | R | Y | 22 | 27 |
| R | Y | 11 | 23 | R | Y | 4 | 18 | R | F | 11 | 13 |
| H | Y | 5 | 11 | H | Y | 2 | 9 | R | W | 9 | 11 |
| H | W | 3 | 6 | W | K | 2 | 9 | K | W | 8 | 10 |
| K | W | 3 | 6 | H | F | 1 | 5 | K | Y | 7 | 8 |
| Y | R | 3 | 6 | H | K | 1 | 5 | H | F | 6 | 7 |
| H | F | 2 | 4 | H | W | 1 | 5 | H | Y | 6 | 7 |
| R | W | 2 | 4 | K | W | 1 | 5 | K | F | 6 | 7 |
| F | R | 1 | 2 | R | H | 1 | 5 | R | H | 4 | 5 |
| H | K | 1 | 2 | W | R | 1 | 5 | K | H | 2 | 2 |
| H | R | 1 | 2 | | | | | H | W | 2 | 2 |
| K | F | 1 | 2 | | | | | | | | |
| R | F | 1 | 2 | | | | | | | | |
| W | H | 1 | 2 | | | | | | | | |
| Y | H | 1 | 2 | | | | | | | | |
| Amino-π interactions | | | | | | | | | | | |
| N | Y | 10 | 36 | N | Y | 3 | 30 | N | Y | 7 | 37 |
| Q | Y | 7 | 25 | N | W | 2 | 20 | N | F | 4 | 21 |
| N | W | 4 | 14 | W | N | 2 | 20 | Q | W | 4 | 21 |
| N | F | 2 | 7 | N | F | 1 | 10 | N | W | 2 | 11 |
| Q | W | 2 | 7 | Y | N | 1 | 10 | Q | F | 1 | 5 |
| Y | N | 1 | 4 | H | N | 1 | 10 | Q | Y | 1 | 5 |
| W | N | 1 | 4 | | | | | | | | |
| Y | Q | 1 | 4 | | | | | | | | |

This is in agreement with the finding that the paratopes are enriched in Asp (Fig. 1).

Hydrogen bonds

The analysis of the hydrogen bond pairs (Supporting Information Table S7) revealed a preference of the

charged and polar residues of the linear and conformational epitopes to interact with the paratope residues Tyr, Ser, Thr, and Asn. Specifically, there is a predominance of Tyr among the paratope residues: it represents almost one-third (31%) of the total number of paratope residues that form hydrogen bonds with conformational epitopes (Supporting Information Table S9). The frequency

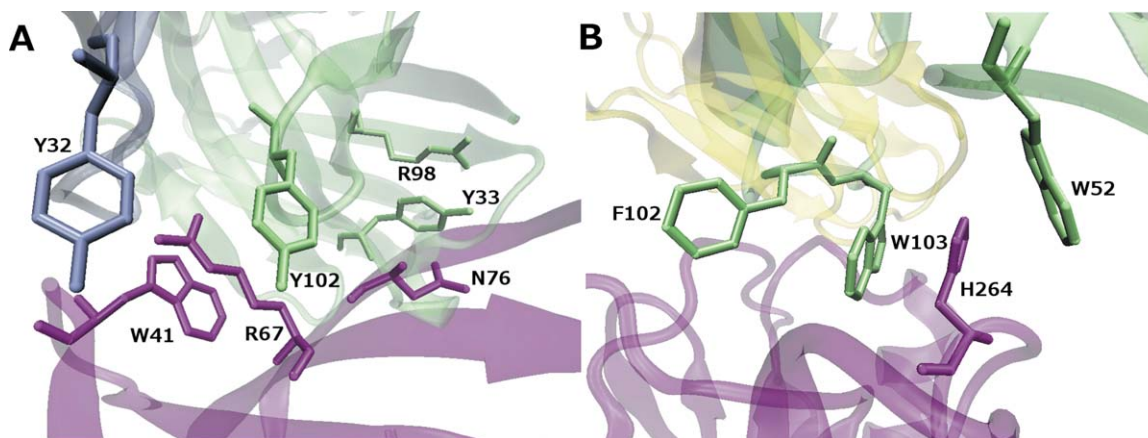


Figure 4

Selected views of antigen-antibody interfaces that show cation/amino- π and π - π chain motifs. Protein chains are displayed as ribbons, and the side chains of interest as sticks. (A) Chains of cation/amino- π interactions in the interface between the major respiratory grass pollen allergen Phl p 2 and specific human immunoglobulin E (PDB code: 2vxq). Epitope residues are shown in purple, the heavy chain residues of the antibody in lime and its light chain residues in blue. The two cation/amino- π chains involve the residues W41/Y32-R67-Y102 and N76-Y33-R98. (B) Chain of π - π interactions in the interface between the intermediate affinity I domain of integrin LFA-1 and its antibody AL-57 (PDB code: 3hi6). Epitope residues are shown in purple and the heavy chain residues of the antibody in lime. The π - π chain involves the residues W52-H264-W103-F102.

of Ser in paratopes that are hydrogen bonded with conformational epitope residues is also high compared with protein–protein interfaces. In conformational epitopes, only 5% of Tyr is involved in H-bonds. The majority of the H-bonds involve positively or negatively charged residues as well as Asn and Gln. In linear epitopes, the preference is less marked. In conformational and especially linear epitopes, some main chain H-bonds are observed.

Hydrophobic contacts

The interactions between hydrophobic residues, from which we exclude the π - π interactions between two aromatic residues, play a critical role in protein–protein interfaces, but are less present in linear epitope–paratope interfaces and even less in conformational epitope–paratope complexes (Table I). The most common hydrophobic residue pairs for the three types of interfaces are presented in Supporting Information Table S8. From the side of both linear and conformational epitopes, the residues involved in hydrophobic packing are first of all Pro, then the aliphatic residues Val, Ile, Leu, and finally aromatic residues. From the side of the paratopes, the main residue involved is undoubtedly again Tyr, with a frequency of 47% and 33% for conformational and linear epitopes, respectively.

In protein–protein interfaces, the majority of the hydrophobic packing is between aliphatic residues. Aromatic residues are less often involved, and Pro even less. So the difference in hydrophobic packing between antigen–antibody and protein–protein interfaces is twofold: there are much less hydrophobic contacts in the former interfaces, and moreover they involve different kinds of residues.

π - π contacts

The occurrence of hydrophobic interactions involving two aromatic rings was investigated in all types of interfaces (Table IV). Such interactions are categorized as parallel or edge-to-face, according to the relative orientation of the aromatic planes. Their importance in molecular recognition has been the subject of intensive study in recent years.⁶⁰ Note that we considered here His both as an aromatic residue capable of forming π - π contacts and as a (sometimes) positively charged residue that can make cation- π interactions. The importance and peculiarities of histidine-aromatic residues have been shown earlier.⁶¹

Table IV indicates a striking preference for Phe in π - π pairs in protein–protein interfaces (at least one partner is a Phe in 67% of the interacting pairs), while Tyr is preferred in epitope–paratope interfaces (64% for conformational epitopes and 57% for linear epitopes). Like for salt bridges, there is an asymmetry between paratopes and epitopes: His is preferred in conformational epitopes (40% of the π - π contacts) and Tyr in the associated paratopes (46%); Tyr is preferred in linear epitopes (41%) and Trp in the associated paratopes (55%).

Cation- π and amino- π interactions

An increased number of experimental and theoretical studies have emphasized the existence of favorable interactions between groups carrying a net or partial positive charge and aromatic systems.^{62,63} These interactions have proven to be important in protein structures and biomolecule associations.^{53,54,64–66} They appear to be an important stabilizing factor of antigen–antibody complexes. They represent 5.6 and 3.9% of the interactions in conformational and linear epitope–paratope interfaces, respectively, against only 1.9% in protein–protein interfaces (Table I). The average number of such interactions per 100 interface residues is 7 for linear and conformational epitope–paratope and only 2 for protein–protein interfaces (Table II). Note that cation- π interactions are roughly twice more frequent than amino- π interactions in epitope–paratope interfaces, whereas they are about four times more frequent in protein–protein interfaces. This indicates that amino- π interactions are perhaps even more important than cation- π interactions in the stabilization of the antigen–antibody interfaces.

The most frequent cation- π and amino- π pairs are listed in Table V. Due to the marked preference for Tyr in paratopes, the cation- π pairs Arg:Tyr, Lys:Tyr, and His:Tyr, with the positive charge in the epitope and the aromatic moiety in the paratope, are the most frequent both in conformational and linear epitope–paratope interfaces (57 and 63% of the cation- π pairs, respectively). Similarly, the amino- π pairs Asn:Tyr and Gln:Tyr are the most frequent in antigen–antibody complexes. As suggested earlier,⁶⁷ the ability of the phenolic OH group of Tyr to form a hydrogen bond could help positioning the phenolic ring properly and participate in favorable cation- π or amino- π contacts.

The most frequent pairs in protein–protein interfaces are the Asn-Tyr amino- π pair (37%) and the Arg-Tyr cation- π pair (27%), an observation that is in agreement with previous studies stating that Arg:Tyr pair is the most abundant in cation- π interactions.^{66,68} These findings are supported by an earlier study claiming that the larger and less well water-solvated side chain of Arg is likely to benefit from better van der Waals interactions with aromatic rings.⁶⁶ Also, quantum chemistry calculations have shown that Arg contributes to cation- π interactions through electrostatic forces but also through dispersion forces due to the delocalization of the charge on the guanidinium group, whereas Lys has basically only the electrostatic contribution.⁶⁴ There are however slightly more Lys- than Arg-containing cation- π pairs in linear epitopes, but the number is too small to yield reliable statistics.

Cation/amino- π and π - π chain motifs

All the interfaces were searched for chain motifs that involve three or more interacting partners. Cation/

amino- π chains occur when an amino acid is involved simultaneously in two cation/amino- π interactions: either the aromatic ring is sandwiched between two side chains that carry a net or partial charge, or the side chain with a totally or partially charged group is sandwiched between two aromatic rings.⁵⁴ Similarly, π - π chain motifs constitute another type of chain motif extension, where three or more aromatic rings interact. As shown in Table II, the average number of cation/amino- π and π - π chains per interface for both the conformational and linear epitope-paratope interfaces is significantly higher than for protein-protein interfaces. We observe indeed 0.3 cation/amino- π chains per 100 interface residues in protein-protein interfaces, and as much as 1.7 and 2.1 in conformational and linear epitope-paratope interfaces, respectively. For π - π chains, 2.1 per 100 interface residues are found in protein-protein interfaces, against 7.4 and 4.3 in conformational and linear epitope-paratope interfaces, respectively. These observations suggest that cation- π and π - π chain motifs play an important role in the stability of antigen-antibody complexes. Two examples of such interactions are depicted in Figure 4.

CONCLUSIONS

The amino acid composition and secondary structure of the interfaces of antigen-antibody and other heterodimeric protein complexes, as well as the type of residue-residue contacts that stabilize them, were analyzed and compared on the basis of high quality X-ray structures. The goal of this study was to identify the amino acid features and interactions that allow distinguishing the different types of interfaces. These features will be used in a subsequent study for predicting potential B-cell epitope regions.

The first result is that antigen-antibody complexes are significantly different from other protein-protein complexes. Antibodies are all- β proteins that make interactions through specific loop regions (CDRs). Antigens have a higher than average β -strand content and a lower than average helix content. They also have the tendency to interact with the antibody through loop regions, especially in the case of linear epitopes but also in conformational epitopes. The epitope enrichment with loops seems to be important for the antibodies to recognize and bind them.²⁶ Other protein-protein interfaces have on the average a lower loop and a higher helix content.

Moreover, protein-protein interfaces are significantly larger on the average than conformational epitopes (with, however, a high standard deviation), which themselves are larger than linear epitopes. Roughly speaking, linear epitopes are about two times smaller than conformational epitopes and three times smaller than heterodimeric protein interfaces.

H-bonds and hydrophobic contacts are, as usual, the most frequent stabilizing interactions, but their relative

abundance varies according to the type of interfaces. H-bond interactions are found to be much more frequent in conformational epitope-paratope interfaces than in other protein-protein interfaces, whereas the opposite is observed for hydrophobic packing. This is related to the fact that the former interfaces are enriched in residues that are able to make H-bond interactions and depleted of nonaromatic hydrophobic residues. The case of linear epitopes is intermediary. Note that the latter epitopes contain an unusually high number of Pro involved hydrophobic contacts. Also, the number of interactions per residue is much higher in linear than in conformational epitopes: the protein core is inexistent and thus all residues tend to interact tightly with the paratope. Salt bridges are the only interactions that are roughly as frequent in all types of interfaces. Note that, again, the number of salt bridges per residue is higher in linear epitopes, because of their tight packing on the paratope. These interactions are directional in the case of antigen-antibody complexes: the positively charged residue is usually at the antigen side and the negatively charged residue at the antibody side. This is related to the higher frequency of these residues in their respective proteins.

A striking result is the significantly higher number of cation- π , amino- π , and π - π interactions across the antigen-antibody interfaces than across other heterodimeric protein interfaces. In 82% of the conformational epitope-paratope interfaces, there is at least one cation- π , amino- π , or π - π interaction. For linear epitopes, which are much smaller, this percentage drops down to 64%, whereas for heterodimeric protein interfaces, which are much larger, it is also equal to 66%. Per interface residue, the number of cation/amino- π and π - π interactions is almost equal in conformational and linear epitopes and two to three times higher than in other protein interfaces.

Moreover, these interactions are directional in antibody-antigen complexes: Tyr is preferred in the antibody for both cation- π , amino- π , and π - π interactions, whereas the charged amino acids Lys or Arg for cation- π interactions, Gln or Asn with a partially charged amino group for amino- π interactions, and any aromatic group for π - π interactions, are situated in the antigen. Note that His also often appears in this context, with the double role of aromatic and positively charged residue, able to form both cation- π and π - π interactions.⁶¹

Successive cation- π , amino- π , or π - π interactions can combine and form chains of cation/amino- π and π - π interactions, as illustrated in Figure 4. The number of such chains is significantly larger—two to seven times—in linear and conformational epitopes than in other heterodimeric protein interfaces. This result suggests that such motifs make a strong contribution to antigen-antibody complex formation.

Note also that the particularly high occurrence of Tyr in paratopes may be explained by its capability to form

all types of interactions except salt bridges: H-bonds, hydrophobic packing, cation- π , amino- π , and π - π interactions. Tyr can thus accommodate various types of interaction partners and appears to play a privileged role in antigen recognition and binding. However, in spite of the high versatility in possible interactions, the antibodies do not seem capable of recognizing any protein surface. The amino acid composition of B-cell epitopes is indeed significantly different from general protein surfaces, which is in contradiction with some previous results.^{41–43} They contain more positively charged residues, more aromatic residues Trp and Tyr, less aliphatic residues and less Phe.

This leads us to the conclusion that though the CDR's of antibodies are hypervariable and can mutate to accommodate many antigens, they cannot adapt to all possible exogenous proteins. There is indeed a strong bias in epitopes toward certain types of amino acids and certain types of interactions. Of course, this could be due to the limited data sets, which contain 58 antigen-antibody complexes with conformational epitopes and 47 with linear epitopes. But it could also reflect a more fundamental issue: antigens are in a first stage bound with relatively low affinity by antibodies that are not totally specific for them. These antibodies then undergo affinity maturation through extensive mutation, become highly specific for the antigen and able to bind them with high affinity. Antigens thus need to be recognized by existing antibodies in a first step, with however lower affinity. This obviously limits the types of exogenous proteins that can be recognized as antigens by the immune system.

ACKNOWLEDGMENTS

We thank Jean-Louis Ruelle for interesting and motivating discussions. GAD acknowledges financial support from the Walloon Region and Glaxo Smith Kline Biologicals through a First-International project. FT thanks the Walloon Region for support through a FIRST Spin-off project.

REFERENCES

- Irving MB, Pan O, Scott JK. Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Curr Opin Chem Biol* 2001;5: 314–324.
- Barlow DJ, Edwards MS, Thornton JM. Continuous and discontinuous protein antigenic determinants. *Nature* 1986;322:747–748.
- Van Regenmortel MH. Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. *Methods* 1996;9:465–472.
- Morris GE. Epitope mapping: B-cell epitopes. In: eLS. Chichester: John Wiley & Sons, Ltd; 2007.
- El-Manzalawy Y, Honavar V. Recent advances in B-cell epitope prediction methods. *Immunome Res* 2010;6 Suppl 2:S2.
- Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* 1981;78:3824–3828.

- Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 1986;25:5425–5432.
- Pellequer JL, Westhof E, Van Regenmortel MH. [8] Predicting location of continuous epitopes in proteins from their primary structures. In: Langone JJ, editor. *Methods in enzymology*, Vol. 203. San Diego, CA: Academic Press; 1991. pp 176–201.
- Emini EA, Hughes J V, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1985;55:836–839.
- Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften* 1985;72:212–213.
- Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990; 276:172–174.
- Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: under-performance of existing methods. *Protein Sci* 2005;14:246–248.
- Pellequer JL, Westhof E. PREDITOP: a program for antigenicity prediction. *J Mol Graph* 1993;11:204–210.
- Odorico M, Pellequer J-L. BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. *J Mol Recognit* 16: 20–22.
- Alix AJ. Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine* 1999;18:311–314.
- Saha S and Raghava GPS. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: Nicosia G, Cutello V, Bentley PJ, Timis J, editors. *Artif. Immune Syst. Third Int. Conf. (ICARIS 2004)*, LNCS 3239; Springer, 2004. pp 197–204.
- Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2006;2:2.
- Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Bioinformatics* 2006;48: 40–48.
- Chen J, Liu H, Yang J, Chou K-C. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 2007; 33:423–428.
- El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *J Mol Recognit JMR* 2008;21:243–255.
- Wee LJ, Simarmata D, Kam Y-W, Ng LF, Tong JC. SVM-based prediction of linear B-cell epitopes using Bayes feature extraction. *BMC Genomics* 2010;11:S21.
- Wang Y, Wu W, Negre NN, White KP, Li C, Shah PK. Determinants of antigenicity and specificity in immune response for protein sequences. *BMC Bioinformatics* 2011;12:251.
- Gao J, Faraggi E, Zhou Y, Ruan J, Kurgan L. BEST: improved prediction of B-cell epitopes from antigen sequences. *PLoS One* 2012;7: e40104.
- Yao B, Zhang L, Liang S, Zhang C. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tripeptide similarity and propensity. *PLoS One* 2012;7:e45152.
- Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, et al. Computational characterization of B-cell epitopes. *Mol Immunol* 2008;45: 3477–3489.
- 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;181:6230–6235.
- Sun J, Xu T, Wang S, Li G, et al. Does difference exist between epitope and non-epitope residues? Analysis of the physicochemical and structural properties on conformational epitopes from B-cell protein antigens. *Immunome Res* 2011;7:1–11.
- Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol* 2013;53:24–34.

29. Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006;15:2558–2567.
30. Liang S, Zheng D, Standley DM, Yao B, et al. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 2010;11:381.
31. Rubinstein ND, Mayrose I, Martz E, Pupko T. EpiTope: a web-server for predicting B-cell epitopes. *BMC Bioinformatics* 2009;10:287.
32. Rubinstein ND, Mayrose I, Pupko T. A machine-learning approach for predicting B-cell epitopes. *Mol Immunol* 2009;46:840–847.
33. Zhang W, Xiong Y, Zhao M, Zou H, et al. Prediction of conformational B-cell epitopes from 3D structures by random forest with a distance-based feature. *BMC Bioinformatics* 2011;12:341.
34. Sweredoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics* 2008;24:1459–1460.
35. Ponomarenko J, Bui H-H, Li W, Fusseder N, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 2008;9:514.
36. Rapberger R, Lukas A, Mayer B. Identification of discontinuous antigenic determinants on proteins based on shape complementarities. *J Mol Recognit* 2007;20:113–121.
37. Jones S, Thornton JM. Prediction of protein–protein interaction sites using patch analysis. *J Mol Biol* 1997;272:133–143.
38. Jones S, Thornton JM. Analysis of protein–protein interaction sites using surface patches. *J Mol Biol* 1997;272:121–132.
39. Jones S, Thornton J. Principles of protein–protein interactions. *Proc Natl Acad Sci* 1996;93:13–20.
40. Lo Conte L, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. *J Mol Biol* 1999;285:2177–2198.
41. Janin J, Chothia C. The structure of protein–protein recognition sites. *J Biol Chem* 1990;265:16027–16030.
42. Lollier V, Denery-Papini S, Larré C, Tessier D. A generic approach to evaluate how B-cell epitopes are surface-exposed on protein structures. *Mol Immunol* 2011;48:577–585.
43. Kunik V, Ofran Y. The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops. *Protein Eng Des Sel* 2013;26:599–609.
44. Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, et al. IEDB-3D: structural data within the immune epitope database. *Nucleic Acids Res* 2011;39:D1164–D1170.
45. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
46. Berman HM, Westbrook J, Feng Z, Gilliland G, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.
47. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA. 3D complex: a structural classification of protein complexes. *PLoS Comput Biol* 2006;2:e155.
48. Thornton JM, Edwards MS, Taylor WR, Barlow DJ. Location of “continuous” antigenic determinants in the protruding regions of proteins. *EMBO J* 1986;5:409–413.
49. Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. *J Mol Biol* 1987;196:641–656.
50. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–2637.
51. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007;372:774–797.
52. Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. *Nucleic Acids Res* 2007;35:W473–W476.
53. Wintjens R, Liévin J, Rooman M, Buisine E. Contribution of cation– π interactions to the stability of protein–DNA complexes. *J Mol Biol* 2000;302:95–410.
54. Rooman M, Liévin J, Buisine E, Wintjens R. Cation– π /H-bond stair motifs at protein–DNA interfaces. *J Mol Biol* 2002;319:67–76.
55. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* 2007;8:211.
56. Zhao L, Li J. Mining for the antibody–antigen interacting associations that predict the B cell epitopes. *BMC Struct Biol* 2010;10 Suppl 1:S6.
57. Soga S, Kuroda D, Shirai H, Kobori M, Hirayama N. Use of amino acid composition to predict epitope residues of individual antibodies. *Protein Eng Des Sel PEDS* 2010;23:441–448.
58. Fellouse FA, Barthelemy PA, Kelley RF, Sidhu SS. Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J Mol Biol* 2006;357:100–114.
59. Ramaraj T, Angel T, Dratz EA, Jesaitis AJ, Mumeby B. Antigen–antibody interface properties: composition, #residue linteractions, and features of 53 non-redundant structures. *Biochim Biophys Acta* 2012;1824:520–32.
60. Hunter CA, Sanders JKM. The nature of π – π interactions. *J Am Chem Soc* 1990;112:5525–5534.
61. Cauët E, Rooman M, Wintjens R, Liévin J, Biot C. Histidine–aromatic interactions in proteins and protein–ligand complexes: quantum chemical study of X-ray and model structures. *J Chem Theory Comput* 2005;1:472–483.
62. Pletneva EV, Laederach T, Fulton DB, Kostic NM. The role of cation– π interactions in biomolecular association. Design of peptides favoring interactions between cationic and aromatic amino acid side chains. *J Am Chem Soc* 2001;123:6232–6245.
63. Scrutton NS, Raine AR. Cation– π bonding and amino–aromatic interactions in the biomolecular recognition of substituted ammonium ligands. *Biochem J* 1996;319:1–8.
64. Biot C, Buisine E, Kwasigroch J-M, Wintjens R, Rooman M. Probing the energetic and structural role of amino acid/nucleobase cation– π interactions in protein–ligand complexes. *J Biol Chem* 2002;277:40816–40822.
65. Dehouck Y, Biot C, Gilis D, Kwasigroch JM, Rooman M. Sequence-structure signals of 3D domain swapping in proteins. *J Mol Biol* 2003;330:1215–1225.
66. Gallivan JP, Dougherty DA. Cation– π interactions in structural biology. *Proc Natl Acad Sci* 1999;96:9459–9464.
67. Mecozzi S, West AP, Dougherty DA. Cation– π interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. *Proc Natl Acad Sci USA* 1996;93:10566–10571.
68. Crowley PB, Golovin A. Cation– π interactions in protein–protein interfaces. *Proteins* 2005;59:231–239.