Available online at www.sciencedirect.com
ScienceDirect
Molecular Immunology xxx (2007) xxx–xxx
Molecular Immunology
www.elsevier.com/locate/molimm

Computational characterization of B-cell epitopes

Nimrod D. Rubinstein [a], Itay Mayrose [a], Dan Halperin [b], Daniel Yekutieli [c], Jonathan M. Gershoni [a], Tal Pupko [a,*]

[a] *Department of Cell Research and Immunology, Tel Aviv University, Tel Aviv 69978, Israel*
[b] *School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel*
[c] *Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel*

Received 12 August 2007; received in revised form 28 September 2007; accepted 4 October 2007

Abstract

Characterizing B-cell epitopes is a fundamental step for understanding the immunological basis of bio-recognition. To date, epitope analyses have either been based on limited structural data, or sequence data alone. In this study, our null hypothesis was that the surface of the antigen is homogeneously antigenic. To test this hypothesis, a large dataset of antibody–antigen complex structures, together with crystal structures of the native antigens, has been compiled. Computational methods were developed and applied to detect and extract physico-chemical, structural, and geometrical properties that may distinguish an epitope from the remaining antigen surface. Rigorous statistical inference was able to clearly reject the null hypothesis showing that epitopes are distinguished from the remaining antigen surface in properties such as amino acid preference, secondary structure composition, geometrical shape, and evolutionary conservation. Specifically, epitopes were found to be significantly enriched with tyrosine and tryptophan, and to show a general preference for charged and polar amino acids. Additionally, epitopes were found to show clear preference for residing on planar parts of the antigen that protrude from the surface, yet with a rugged surface shape at the atom level. The effects of complex formation on the structural properties of the antigen were also computationally characterized and it is shown that epitopes undergo compression upon antibody binding. This correlates with the finding that epitopes are enriched with unorganized secondary structure elements that render them flexible. Thus, this study extends the understanding of the underlying processes required for antibody binding, and reveals new aspects of the antibody–antigen interaction.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Antigen; Antibody; Epitope; Surface

1. Introduction

The interaction between an antibody and its antigen is at the heart of the humoral immune response. Antibodies bind to their corresponding antigens at discrete sites known as antigenic determinants or epitopes, as originally defined by Jerne (1960). The precise localization of an epitope can be essential in the development of biomedical applications such as rationally designed vaccines, diagnostic kits, and immuno-therapeutics (Irving et al., 2001; Westwood and Hay, 2001). Thus, a detailed molecular characterization of epitopes may greatly contribute to such endeavors. In a wider sense, characterizing epitopes is fundamental to the understanding of the basis of immunological discrimination between self and non-self as well as mechanisms of bio-recognition in general. Since proteins are one of the most abundant and diverse classes of antigens, including transplantation antigens, antigens of infectious agents, and allergens, much of the interest in antigen characteristics is focused on antigenic proteins.

Epitopes play a pivotal role in antigen recognition as was illustrated very early on in such studies as those of Arnon and Sela (1969). These investigators demonstrated that a specific and isolated linear peptide, derived from the sequence of a given antigen, was able to elicit antibodies that not only bound the peptide but strongly cross-reacted with the native antigen as

Abbreviations: APBS, adaptive Poisson–Boltzmann solver; ASA, accessible surface area; CDR, complementarity determining region; CE, combinatorial extension; CGAL, computational geometry algorithms library; CSU, contacts of structural units; DSSP, dictionary of secondary structure of proteins; FDR, false discovery rate; PDB, protein data bank; RMSD, root mean square deviation; SCOP, structural classification of proteins.

* Corresponding author. Tel.: +972 3 640 7693; fax: +972 3 642 2046.
E-mail address: talp@post.tau.ac.il (T. Pupko).

0161-5890/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2007.10.016

well. Atassi (1975, 1978) was able to demonstrate this point for conformational epitopes using peptides whose structures mimic discontinuous fragments of the antigen brought to juxtapose one another through protein folding. Thus, the importance of the tertiary conformation of the antigen structure was emphasized. The dependence of the immune response on the antigen structure was further demonstrated by Jemmerson and Margoliash (1979) who showed that even minor modifications of antigens can have a profound effect on the immune response.

What actually constitutes an antigenic determinant has been the center of much debate and at least two opposing views have been suggested. One proposes that the surface of a globular protein is a continuum of potential antigenic sites, all capable of eliciting an immune response. Conversely, the other view argues that protein surfaces contain a very limited number of exclusive sites that are inherently antigenic. The fact that interior parts of the protein, which are naturally not exposed and as such not expected to possess antigenic characteristics, are quite capable of stimulating production of specific antibodies (Benjamin et al., 1984) has been proposed as evidence in support of the first view. Yet, in support of the second view is the fact that during the course of the natural response to a given antigen very few locations on the surface of a protein are required for the generation of the overwhelming majority of antibodies produced against it (Hopp, 1986). A compromising view suggests that all molecular sites could be antigenic, although some have a significantly higher potential to be recognized by the immune system (Berzofsky, 1985). Thus, the question becomes what correlates best for effective antigenicity, or what are the defining traits of strong epitopes?

Hopp and Woods (1981) reported a significant correlation between hydrophilicity and antigenicity. Westhof et al. (1984) then claimed that backbone flexibility is a better criterion for antigenicity than is hydrophilicity. Other works analyzed the few antibody–antigen co-crystal structures available at that time and revealed further structural aspects of epitopes. For example, Novotny et al. (1986) demonstrated that the correlation between accessible surface area (ASA) and antigenicity is superior to the correlation between backbone flexibility and antigenicity. Thornton et al. (1986) showed that antigenic sites protrude considerably from the protein surface and concluded that this property is a strong characteristic of antigenicity. Laver et al. (1990) reported that epitope areas span a narrow range of 650–900 $Å^2$, encompassing 15–22 amino acids. In addition, they also observed striking structural complementarity in the antibody–antigen interface.

The advance of structure determination technologies in the 1990s accelerated the production rate of crystals of protein complexes (Berman et al., 2000), which led to the solution of numerous protein–protein interfaces. Although the amount of antibody–antigen co-crystal data remained limited, these analyses revealed general principles of protein–protein interactions, which also had implications for antibody–antigen interactions. Jones and Thornton (1995, 1996, 1997a,b) compared interfaces and non-interface regions with respect to physico-chemical and structural properties at the amino acid level, such as ASA, amino acid composition, degree of protrusion, and flexibility. Lo Conte et al. (1999) performed a similar analysis, adding further physico-chemical and structural aspects of protein–protein interfaces at the atomic level. Neuvirth et al. (2004) analyzed a much larger dataset of hetero-complexes (57 structures, however antibody–antigen complexes were excluded). Aside from highlighting the importance of the physico-chemical character of interfaces, evolutionary conservation and secondary structure content were also found to be important properties of protein–protein interfaces.

In most of the recent comprehensive analyses of protein–protein co-crystal complexes antibody–antigen complexes were either discarded (e.g., Ma et al., 2003; Neuvirth et al., 2004) or constituted an insignificant portion of the data and then did not receive much attention as a separate module (e.g., Jones and Thornton, 1996, 1997a; Lo Conte et al., 1999). This last point may be important as the nature of the antibody–antigen interaction may be fundamentally different from other types of protein–protein interactions such as subunit–subunit association or enzyme-substrate binding. Furthermore, although some physico-chemical and structural features were suggested to correlate with antigenicity, no thorough statistical examination was performed to assess whether they truly distinguish epitopes from the remaining antigen surface. Thus, the nature of antibody–antigen recognition is still far from being resolved.

With the large increase of solved antibody–antigen co-crystal structures in the protein data bank (PDB) (Shindyalov and Bourne, 1998), it is now possible to perform a large-scale analysis to define epitope characteristics and reveal new aspects of immunological molecular recognition. Furthermore, the abundance of currently available data enables to perform the analysis in a statistically robust manner to reliably determine whether some properties significantly distinguish epitopes from the remaining antigen surface. In addition, the availability of native antigen structures makes it possible to examine the changes that antigens experience due to antibody binding. For all these reasons, a large-scale analysis of all available antibody–antigen complexes was undertaken here. Physico-chemical, structural, and geometrical aspects of epitopes were characterized and rigorous statistical inference was applied to determine which of these properties significantly distinguish epitopes from their surrounding antigen surface.

2. Methods

2.1. Data construction

All antibody–antigen complexes from the SPIN server of protein–protein complexes (http://trantor.bioc.columbia.edu/cgi-bin/SPIN/) were retrieved. To ensure that all available antibody–antigen complexes were indeed obtained, the PDB was also manually searched using appropriate key words. Eventually, a dataset of 246 antibody–antigen co-crystal structures was obtained. This preliminary dataset was then subjected to a filtering process using several criteria. First, all complexes in which the antibody molecule does not contain both the light and heavy chains were discarded as they do not reliably represent a *bona fide* antibody–antigen interaction. Then, complexes in which the

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

Table 1
Complexed, bound, and unbound datasets

Complexed and bound datasets				Unbound dataset	
PDB ID	Antibody chains	Antigen chains	Antigen description	PDB ID[a]	Description
1a14	H,L	N	Influenza A neuraminidase	1iny(A)	Influenza A subtype N9 neuraminidase
1a2y	A,B	C	Hen egg white lysozyme	1hel(A)	Hen egg white lysozyme wild type
1adq	H,L	A	IgG4 REA	1mco(H)	IgG1 (IgG1) (MCG) with a hinge deletion
1afv	H,L	A	HIV-1 capsid		
1ahw	A,B	C	Human tissue factor	1boy(A)	Human tissue factor
1ar1	C,D	A,B	Cytochrome *C* oxidase		
1bgx	H,L	T	Taq dna polymerase	1cmw(A)	Taq dna polymerase I
1bj1	J,K	V	Vascular endothelial growth factor	2vpf(D)	Vascular endothelial growth factor
1bql	H,L	Y	Bobwhite quail lysozyme	1dkj(A)	Bobwhite quail lysozyme
1cic	A,B	C,D	Ig light and heavy chain v regions		
1dqj	A,B	C	Hen egg white lysozyme	1hel(A)	Hen egg white lysozyme wild type
1dvf	C,D	A,B	Fv D1.3	1a7r(L,H)	Monoclonal antibody D1.3
1e6j	H,L	P	HIV-1 capsid protein P24	1a43(A)	HIV-1 capsid protein P24
1egj	H,L	A	Cytokine receptor common beta chain precursor		
1eo8	H,L	A,B	Hemagglutinin	5hmg(E,F)	Hemagglutinin
1ezv	X,Y	E	Ubiquinol-cytochrome *C* reductase iron–sulfur subunit		
1fbi	H,L	X	Guineafowl egg white lysozyme	1hhl(A)	Guineafowl egg white lysozyme
1fe8	H,L	A	Von willebrand factor	1ao3(A)	Von willebrand factor
1fj1	A,B	F	Outer surface protein A		
1fns	H,L	A	Von willebrand factor	1ijb(A)	Von willebrand factor
1fsk	B,C	A	Major pollen allergen bet V 1-A	1bv1(A)	Major pollen allergen bet V 1-A
1g9m	H,L	G	Envelope glycoprotein GP120		
1h0d	A,B	C	Angiogenin	1k59(A)	Angiogenin
1hys	C,D	A,B	HIV-1 reverse transcriptase		
1i9r	H,L	A	CD40 ligand	1aly(A)	CD40 ligand
1iai	I,M	H,L	Idiotypic Fab 730.1.4 (IgG1) of virus neutralizing antibody		
1iqd	A,B	C	Human factor VIII	1d7p(M)	Coagulation factor VIII precursor
1jhl	H,L	A	Pheasant egg lysozyme	1ghl(A)	Pheasant egg lysozyme
1jrh	H,L	I	Interferon-gamma receptor alpha chain		
1k4d	A,B	C	Potassium channel KCSA	1j95(A)	Voltage-gated potassium channel
1kb5	H,L	A,B	KB5-C20 T-cell antigen receptor		
1lk3	H,L	A	Interleukin-10	1ilk(A)	Interleukin-10
1mhp	H,L	A	Integrin alpha 1	1ck4(A)	Integrin alpha 1
1n8z	A,B	C	Receptor protein-tyrosine kinase ERBB-2		
1nca	H,L	N	N9 neuraminidase-NC41	1iny(A)	N9 neuraminidase-NC41
1nfd	E,F	A,B	N15 alpha–beta T-cell receptor	1tcr(A,B)	Alpha–beta T-cell receptor
1nl0	H,L	G	Factor IX		
1oaz	H,L	A	Thioredoxin 1		
1ob1	A,B	C	Merozoite surface protein 1		
1orq	A,B	C	Potassium channel KCSA		
1ots	C,D	A	Voltage-gated CLC-type chloride channel eric	1kpk(A)	Putative channel transporter
1pg7	W,X	H,L	Humanized antibody D3H44		
1pkq	A,B	E	Myelin oligodendrocyte glycoprotein	1pko(A)	Myelin oligodendrocyte glycoprotein
1qfw	H,L	A	Gonadotropin alpha subunit	1hcn(A)	Human chorionic gonadotropin
1qgc	4	5	Gh-loop from virus capsid protein VP1		
1qle	H,L	B	Cytochrome *C* oxidase polypeptide II		
1rvf	H,L	1,2,3,4	Human rhinovirus 14 coat protein	4rhv(1,2,3,4)	Rhinovirus 14
1sy6	H,L	A	T-cell surface glycoprotein CD3 gamma/epsilon chain		
1tpx	C	A	Major prion protein	1uw3(A)	Prion protein
1v7m	H,L	V	Thrombopoietin		
1wej	H,L	F	Cytochrome *C*	1hrc(A)	Cytochrome *C*
2jel	H,L	P	Histidine-containing protein	1poh(A)	Phosphotransferase
2vir	A,B	C	Hemagglutinin	1ha0(A)	Hemagglutinin precursor

[a] Relevant chains are indicated in parentheses.

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

antibody–antigen contact was found to be mediated by antibody residues that are not part of the complementarity determining regions (CDRs) (Allcorn and Martin, 2002) were additionally discarded. Following that, all complexes that contain only small fragments of the antigen, bound to the antibody, were discarded since they do not enable an appropriate comparison between the epitope and the remaining antigen surface. Finally, redundant complexes (i.e., complexes with identical antibody and antigen), which were detected using the structural classification of proteins (SCOP) database (Murzin et al., 1995) were removed. Following these processes 53 complexes remained. Two datasets were derived from the PDB files of these complexes (Table 1): (i) the complexed dataset, containing all antibody–antigen complexes; and (ii) the bound dataset, containing only the antigen structures derived from the complexes.

An antigen structure that is derived from the complex may reflect geometrical changes that occurred following the formation of the complex with the antibody, compared with its native structure. In order to analyze the antigen structure before such geometrical changes took place, the unbound structures of the antigens (i.e., the native structure) were thus searched. To find an identical or homologous representative for a bound antigen, the combinatorial extension (CE) method (Shindyalov and Bourne, 1998), which performs structure alignments of a query protein against the PDB, was used. A filtering criterion of a minimum of 70% sequence identity to the query structure (both for the entire sequence and the epitope sequence alone) was applied, and whenever multiple hits were obtained the structure with the highest *Z*-score (Shindyalov and Bourne, 1998) was chosen. This procedure resulted with a dataset of 32 antigens termed the unbound dataset (Table 1). For every retrieved unbound structure, each of its amino acids was associated with an amino acid from the bound structure according to their sequence alignment.

2.2. Surface analysis

The molecular surfaces of each structure in the three datasets (complexed, bound, and unbound) were computed using the Surface Racer program (Tsodikov et al., 2002), with a probe radius of 1.4 Å. An amino acid was considered to be exposed to the solvent if the sum of the ASAs of its atoms exceeded 5% of its maximal (theoretical) ASA (i.e., relative ASA = ASA/maximal ASA > 0.05). The maximal ASA value of an amino acid was calculated in an extended GXG theoretical tripeptide, where G denotes glycine and X denotes the residue in question (Miller et al., 1987).

2.3. Epitope definition

In order to determine the epitope from each complex structure, the contacts of structural units (CSU) program (Sobolev et al., 1999), which lists all atoms that are in contact between two proteins in a complex, was used. Only solvent exposed amino acids for which at least one atom was found to be in contact with the antibody were regarded as epitope amino acids.

2.4. Generation of overlapping patches from the antigen surface

In order to examine similarities and differences between epitopes and other areas on the antigen surface, overlapping patches derived from the antigen surface were generated. A patch was defined as the group of $n - 1$ residues with the shortest distance to a central residue, where n equals the number of residues in the corresponding epitope (Jones and Thornton, 1997a). The distance between two residues was defined as the minimal Euclidean distance between the centers of any of their exposed atoms. To extract all non-epitope overlapping patches, each exposed residue of the antigen was selected as a central residue around which a patch was constructed. To avoid sampling of the epitope, each patch which overlapped the epitope with one or more residues was discarded.

2.5. Statistical inference

The *G*-test for goodness of fit (Sokal and Rohlf, 1995) was used in order to test whether a certain property in epitopes and the remaining antigen surfaces is sampled from the same distribution. Thus, first all instances of epitopes and remaining antigen surfaces are combined into two separate groups, respectively. Then, the *G*-test is applied to compare the property between the two combined data.

A statistical caveat of the *G*-test may be encountered when the data instances are not homogeneous. For example, epitopes can only be considered as homogeneous with respect to alanine frequency if the alanine frequency in all epitopes is sampled from the same distribution. Non-homogeneous data may lead to the Simpson's paradox, which results in erroneous conclusions (Simpson, 1951). The Mantel–Haenszel test (Lilienfeld and Stolley, 1994) overcomes this limitation by accounting for the possible heterogeneity among data instances. Thus, the Mantel–Haenszel test was always applied in addition to the *G*-test.

Whenever a multiple testing procedure was applied, the false discovery rate (FDR) correction for multiple testing (Benjamini and Hochberg, 1995) was used.

3. Results

The fundamental question underlying this study is whether B-cell epitopes have physico-chemical and structural-geometrical characteristics that can render them more immunogenic compared to the remaining antigen surface. Whereas each co-crystal examined clearly defines an epitope, one cannot assume that the remaining surface of the given antigen is necessarily non-immunogenic in its entirety. It may well be that the remaining surface of the antigen not occupied by the specific antibody is mixed with epitope surfaces of alternative antibodies. The null hypothesis in this investigation assumes that the entire surface of an antigen is equally immunogenic and could be effective B-cell epitopes. Therefore, in the comparisons of epitope versus non-epitope surfaces of defined co-crystals it is not expected to find differences for each property evaluated. However, rejection of the null hypothesis clearly implies that there are unique

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

Table 2
Ranking of the contribution of the CDR loops to the contact area with the epitope

CDR loop	Average contact area with epitope (Å^2)	Average length (number of amino acids)
L1	126.157	11.62
L2	68.3	7
L3	144.375	9.15
H1	104.171	5.13
H2	221.757	16.84
H3	268.445	10.2

traits for "epitopeness". Before conducting comparative analyses we first characterize basic traits of the epitope surfaces per se.

3.1. Size and area distributions of epitopes

The size of an epitope was defined as the number of amino acids comprising it. The area of an epitope was computed by subtracting the ASA of antigen-derived amino acids of the complex from the ASA of the bound antigen. This analysis revealed that 75% of the epitopes constitute 15–25 amino acids spanning an area range of 600–1000 Å^2 (with a median size of 20 amino acids and a median area of 790 Å^2). This agrees well with previously published results from analyses of smaller datasets of antibody–antigen complexes (Chakrabarti and Janin, 2002; Laver et al., 1990; Lo Conte et al., 1999).

3.2. CDR- and non-CDR-bound epitope regions

The narrow ranges of epitope size and area may reflect global structural constraints on CDRs of antibodies. We thus characterized in detail the contribution of the CDR loops to the interaction with the antigen. Measuring the contact area between each of the six Fab CDR loops and the antigen revealed that the third heavy chain CDR loops make the largest average contact area with epitopes (Table 2). Respectively the distribution of the CDR contact areas, where the difference in the average lengths of the CDRs (number of amino acids) is accounted for, showed a significant deviation from a uniform distribution ($P < 10^{-16}$; χ^2-test). The CDR analysis further showed that, on average, 90% of an epitope area is in contact with CDR residues. In other words, 10% of the epitope is bound by antibody residues outside the CDR loops. This analysis further showed that there is no significant difference in the average distance to the antibody molecule between these two epitope regions. If one defines those amino acids that are in close proximity to the antibody as "the core epitope" (reviewed in Shoemaker and Panchenko, 2007) it follows that the core is not enriched with CDR-bound amino acids (data not shown).

3.3. The segmented structure of epitopes

Epitopes are traditionally classified as either linear (i.e., continuous) or conformational (i.e., discontinuous) (Berzofsky, 1985). According to this classification, all epitopes that are composed of a single continuous segment of amino acids are regarded as linear and the remaining epitopes are regarded as conformational. Using this definition, all epitopes analyzed in this study were found to be conformational, where the majority of them (over 70%) consisted of six or more short (1–3 amino acids) linear segments (Fig. 1). However, many of these segments were found to be separated by only a few amino acids, which do not directly contribute to the epitope (e.g., two segments interrupted by a single buried amino acid). It may be more informative to allow a less stringent definition for continuity, where a segment may include three or less "non-epitope" amino acids. Using this refinement, it was found that still all epitopes are conformational, however now the majority of them (over 70%) where found to be composed of 1–5 segments of longer lengths (1–6 amino acids) (Fig. 1).

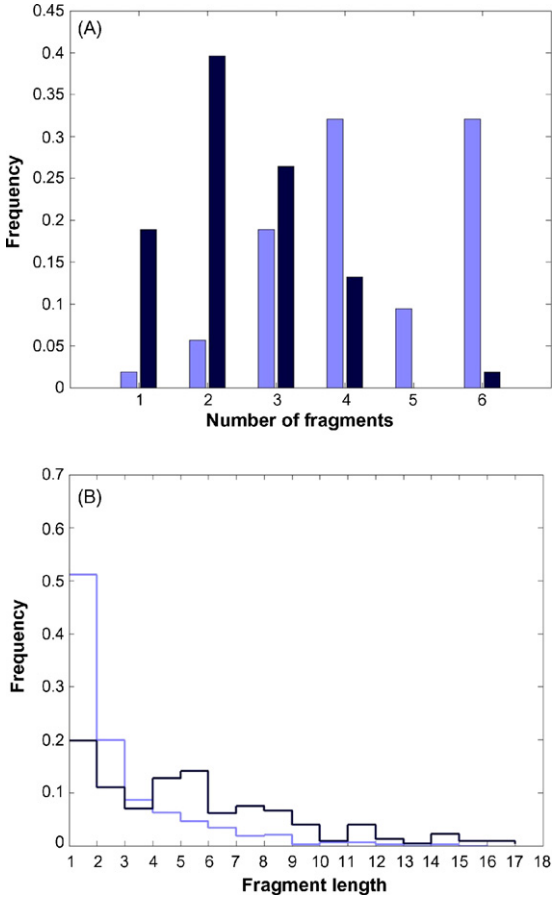

Fig. 1. The segmented structure of epitopes. (A) Distribution of the number of linear segments per epitope. (B) Frequency of segments lengths. In light blue, linear segments, which are not interrupted by non-epitope amino acids. In dark blue, segments which are interrupted by 0–3 non-epitope amino acids. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

The following analyses compare the composition and structure of the epitopes with the non-epitope surfaces.

3.4. Amino acid preference of epitopes

Previous works have reported that the amino acid preference of epitopes differs from that of the remaining antigen surface (Jones and Thornton, 1995; Lo Conte et al., 1999). Here, the

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

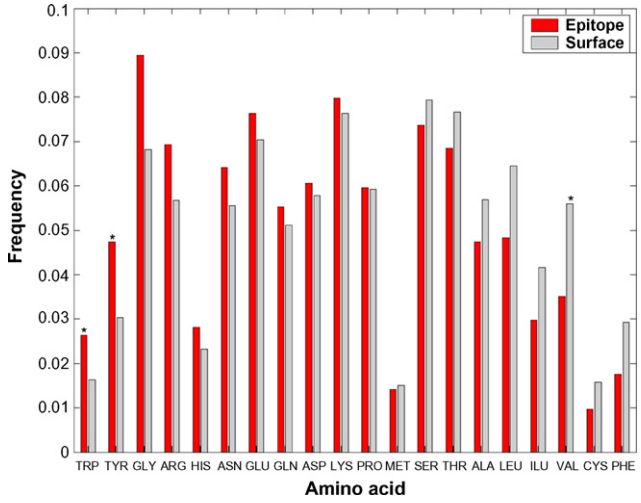

Fig. 2. Amino acid preference of epitope and non-epitope surfaces. Asterisks mark statistically significant differences between epitope and surface.

amino acid preference was evaluated using the amino acid frequencies (Fig. 2).

The overall amino acid composition was found to differ significantly between epitope and non-epitope surfaces ($P < 10^{-6}$; *G*-test). In order to explain this difference, the test was repeated for each one of the amino acids, separately. The results show that epitopes are significantly overrepresented by tyrosine and tryptophan, and generally enriched with charged and polar amino acids (in all cases $P < 0.002$; *G*-test subject to the FDR correction). They are underrepresented by the aliphatic hydrophobic amino acids, with a significant depletion of valine ($P < 0.005$; *G*-test subject to the FDR correction). These findings are supported by previous reports (Bogan and Thorn, 1998; Jackson, 1999) claiming that tyrosine, tryptophan and charged residues, are generally preferred in protein–protein interfaces due to their capability to form a multitude of interactions.

Evaluating amino acid preference using amino acid frequencies may nevertheless be biased if the ASA contribution of the amino acids is not accounted for. For example, an amino acid with a low relative ASA of 0.06 is counted similarly as an amino acid with an extremely high relative ASA of 0.99. To account for this possible bias, amino acid frequencies in epitope and non-epitope surfaces were repeatedly measured using increasing relative ASA cutoffs, from 0.05 to 0.25 with a 0.05 increment. In this manner, the area contribution is reflected since amino acids with minor relative ASAs are filtered out. This analysis produced essentially the same results as shown in Fig. 2.

The physico-chemical character of an amino acid is defined by the composition of its side chain. Thus, if certain epitope-favored amino acids make contact with the antibody only through their backbones it is possible that there is no real physico-chemical preference for them over other amino acids. For this reason, a more refined analysis on the amino acid preference of epitopes was performed. In this analysis, only amino acids for which side-chain atoms are exposed were considered. As before, the analysis was performed with increasing ASA cutoffs. The results obtained reveal that much of the same trend observed in the earlier analyses is retained (in all cases $P < 0.01$; *G*-test and Mantel–Haenszel test; data not shown).

Establishing that epitopes are enriched with specific amino acids, it was next tested whether epitopes have a higher concentration of residues with exposed side chains versus remaining antigen surfaces. The percentage of the amino acids with an exposed side chain (relative to the total number of amino acids) was compared between epitopes and remaining antigen surfaces. The unbound dataset was used for this purpose in order to prevent a possible bias concerned with the possibility that the epitopes of the bound dataset reflect structural changes experienced due to the antibody binding. The comparison indicated that the epitope surface is significantly enriched with amino acids that have exposed side chains, relative to remaining antigen surfaces ($P = 0.008$; paired *t*-test). This suggests that residues with exposed side chains play an important role in forming the antibody–antigen complex and hence surface regions in which a large fraction of the amino acids expose their side chains are favorable for antibody binding.

3.5. Amino acid cooperativity in epitopes

It has been suggested that amino acids that are proximal on the epitope act cooperatively, thus enhancing certain traits important for the binding interaction (Bublil et al., 2007; Enshell-Seijffers et al., 2003; Neuvirth et al., 2004). To test this hypothesis, the composition of amino acids was analyzed again, to check for overrepresentation of amino acid pairs (see Supplementary material for a detailed explanation of the statistical test).

As observed in Fig. 3, a signal pointing at cooperativity between spatially adjacent amino acids in epitopes is indeed apparent. Interestingly, tyrosine, which is significantly abundant in epitopes, is also dominant in pairs of cooperativity. Conversely, tryptophan, which is also significantly abundant in epitopes, does not seem to play an important role in cooperativity. In addition, whereas proline is not significantly overrepresented in epitopes, it seems to be important when paired with either cysteine or aspartate.

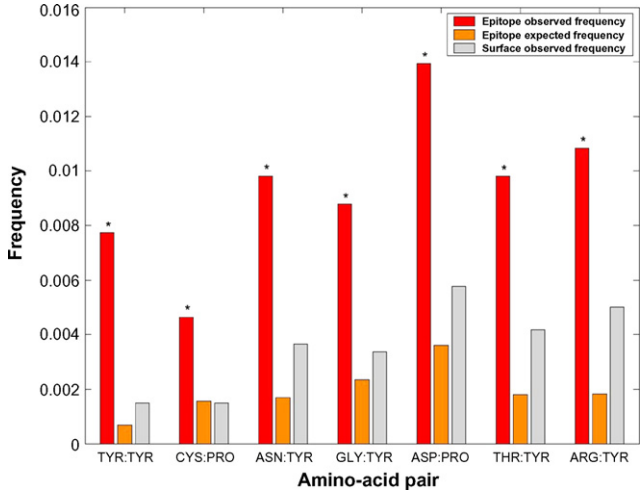

Fig. 3. Frequencies of amino acid pairs. Asterisks mark pairs which are observed in epitopes significantly more than the random expectation.

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

3.6. Accessible surface area of epitope residues

It is expected that epitope residues should be highly accessible to facilitate their contact with the antibody. To test this, the surface accessibility (measured by residue mean relative ASAs) was compared between epitope and non-epitope surfaces. This analysis was performed on the unbound dataset as it is possible that epitopes undergo structural modifications upon complex formation that increase their surface accessibility, which may thus bias the analysis. The epitope portion of the surface was found to be readily more exposed compared to the remaining surface ($P < 10^{-9}$; paired *t*-test). The same results ($P < 10^{-7}$; paired *t*-test) were obtained when a probe with a radius of 9 Å (approximating a CDR rather than the 1.4 Å radius approximating a water molecule) was used to measure surface accessibility (Novotny et al., 1986). These results indicate that a typical epitope is characterized with a significant accessibility compared to the remaining antigen surface.

3.7. Epitope geometry

3.7.1. At the atom level

The surface shape of atoms of residues with higher solvent accessibility is expected to be more bulgy (convex). The average surface curvature of exposed atoms was compared between epitope and non-epitope surfaces of the unbound dataset using the Surface Racer program (Tsodikov et al., 2002). The results obtained reveal that the shape of epitope atoms is significantly more convex than that of non-epitope atoms ($P < 10^{-6}$; paired *t*-test). Hence, it seems that the surface shape of epitopes at the atom level can best be viewed as a rugged terrain.

3.7.2. At the patch level

The two previous sections both characterize epitopes at the amino acid microenvironment resolution and thus do not portray the geometrical shape of the epitope as in its entirety. The shape of an epitope, considered as a single entity, can assume two possible conformations, either flat or curved. To examine whether epitopes are flatter relative to other patches on the antigen surface, two measures were computed: (i) the width of the patch measured by computing the minimal distance between two parallel planes encompassing the centers of all exposed patch atoms using the computational geometry algorithms library (CGAL) (http://www.cgal.org); (ii) root mean square deviation (RMSD) of the centers of exposed patch atoms from the least squares plane fitted to the centers of these atoms (Jones and Thornton, 1997a). These measures (illustrated in Fig. 4) were compared between epitopes and other patches on the antigen surface for the unbound dataset (see Supplementary material for a detailed explanation of the statistical test).

The average width of an epitope was found to be 12.45 Å with a standard deviation of 4.17 Å (compared to 14.19 Å with a standard deviation of 6.16 Å for non-epitope patches). The statistical analysis revealed that epitopes are significantly flat according to the width measure ($P = 0.02$). However, statistical significance was not achieved when the RMSD measure was used to test flatness ($P = 0.24$). This apparent inconsistency can be used to provide a refined insight into the geometry of an epitope. The width measure is only affected by the location of the centers of the most outlying atoms. If an epitope is visualized as a terrain the width corresponds to the difference in height between the highest and lowest points. The RMSD measure however, is affected by the location of the centers all atoms. According to the terrain visualization, the RMSD measure is affected not only by the highest mountain and lowest basin, but also by local hills and valleys. Taken together the emerging geometrical model of an epitope is of a flat yet rugged surface.

3.7.3. At the molecule level

An exposed atom on the surface of the antigen may reside in a depressed area of the protein such as a pocket, in a bulgy area, or, in a relatively flat area. For an epitope to be able to interact with the CDR of an antibody, one would expect it to reside in a bulgy area that is easily accessible to another macromolecule. To test this hypothesis, the convex hull (see illustration in Fig. 5) was constructed for the centers of the atoms comprising the antigen. Informally, the convex hull can be described as a sheet

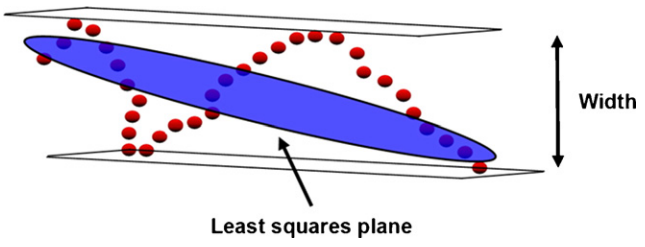

Fig. 4. Illustration of the width and RMSD measures. The epitope atom centers are presented as red dots. The width is the distance between the two parallel planes encompassing all epitope atoms (the minimum over all possible two parallel planes). The least squares plane fitted to the epitope atoms is colored blue. The RMSD is calculated as the root mean square deviation of all atom centers from the least squares plane. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

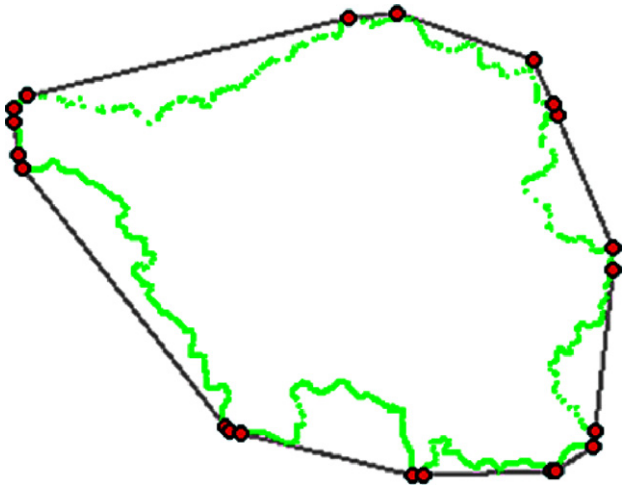

Fig. 5. A two-dimensional convex hull. The structure is shown in green. In black, is the convex hull wrapping the structure. In red, are points of the structure that reside on the convex hull. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

that wraps the most exterior points of a structure. Formally, the convex hull of a set of points S in the three-dimensional space is the smallest convex set containing S, thus creating a convex polytope which contains all the extreme points of S. The convex hull was constructed using CGAL (http://www.cgal.org).

If an epitope resides on a convex area, its atoms are expected to be within a short distance from the boundary of the convex hull. Thus, the fraction of atom centers that lie within a certain cutoff distance from the boundary of the convex hull was compared between epitopes and the remaining antigen surfaces in the unbound dataset. This analysis revealed that for any distance cutoff (0, 2, 4, 6, 8, and 10 Å) the fraction of atom centers that reside near the convex hull is significantly larger in epitopes than in the remaining antigen surfaces ($P < 0.001$ for all cutoffs; paired *t*-test). This demonstrates that epitopes reside on areas that are easily accessible to other macromolecules such as antibodies. In summary, the results of all the geometrical analyses provide a strong indication that epitopes are regions with distinguishable structural properties.

3.8. Epitopes secondary structure

An important structural aspect of an epitope that may distinguish it from the remaining antigen surface is its secondary structure composition. To test whether epitopes are enriched with respect to specific secondary structure elements versus non-epitope surfaces, each amino acid was assigned to either of the following three secondary structure groups according to its description from the dictionary of secondary structure of proteins (DSSP) (Kabsch and Sander, 1983): (i) alpha-helices, 3/10 helices, and pi-helices were grouped as helices; (ii) isolated beta-bridges and extended beta strands were grouped as strands; and (iii) turns, bends, and irregular structures were grouped as loops. This analysis revealed that epitopes are significantly enriched with loops and significantly depleted of helices and strands, compared to non-epitope surfaces (in all cases $P < 0.001$; *G*-test and Mantel–Haenszel test subject to the FDR correction) (Fig. 6). Since loops tend to be more flexible than other organized secondary structure elements (Jemmerson and Paterson, 1985; Pellequer et al., 1991), these results suggest that epitopes are relatively flexible.

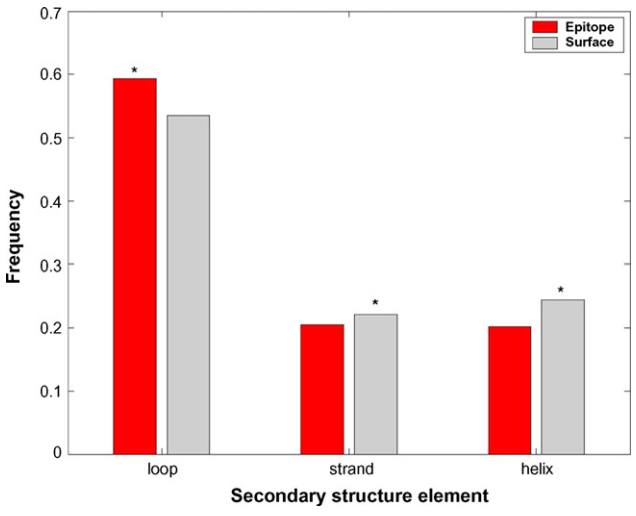

Fig. 6. Distributions of secondary structure elements in epitope vs. non-epitope surfaces. Asterisks mark significant frequency differences between epitope and surface according to the *G*-test subject to the FDR correction.

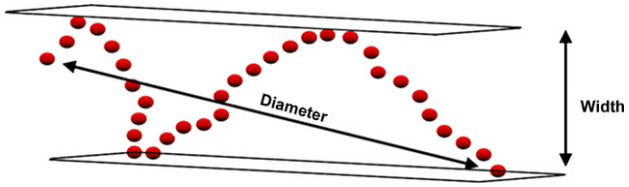

Fig. 7. Illustration of the concepts of the width and three-dimensional diameter measures. The epitope atom centers are presented as red dots. The width is the distance between the two parallel planes encompassing all epitope atoms (the minimum over all possible two parallel planes). The three-dimensional diameter is the maximal distance between any two-epitope atoms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.9. The structural effect on epitopes upon complex formation

The availability of both unbound and bound structures of the same antigen allows the characterization of the structural changes that epitopes undergo upon antibody binding. For 32 pairs of such unbound and bound antigen structures compiled in this work (Table 1), two geometrical measures were computed. The width measure (described above) and the three-dimensional diameter measure, which is the maximal distance between the centers of any two-epitope atoms (as illustrated in Fig. 7).

This analysis revealed that bound epitopes are significantly wider than unbound epitopes (average widths of bound and unbound epitopes = 14.46, 11.73 Å, respectively; $P = 0.002$; paired *t*-test). In addition, it was also found that the three-dimensional diameter of bound epitopes is smaller than that of unbound epitopes (average diameters of bound and unbound epitopes = 30.52, 35.23 Å, respectively; $P = 0.001$; paired *t*-test). This may indicate that epitopes undergo compression to a certain degree upon antibody binding, as if the CDR acts like a vice-grip. As a case in point, this structural compression is demonstrated for the complex of CD40 ligand and the Fab fragment of its neutralizing antibody, humanized 5C8 (PDB identifiers 1i9r and 1aly for the bound and unbound structures, respectively) in Fig. 8. The abundance of flexible secondary structure elements in epitopes may facilitate the capacity demonstrated by epitopes to undergo conformational adjustments upon antibody binding.

3.10. Evolutionary conservation of epitopes

Functional regions on protein surfaces tend to be evolutionarily conserved relative to other regions (Nimrod et al., 2005; Zhou and Shan, 2001). Epitopes may overlap such functional regions due to shared constraints imposed by the nature of protein–protein interactions. If so, epitopes should be more evolutionary conserved than remaining antigen surfaces. To test this

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

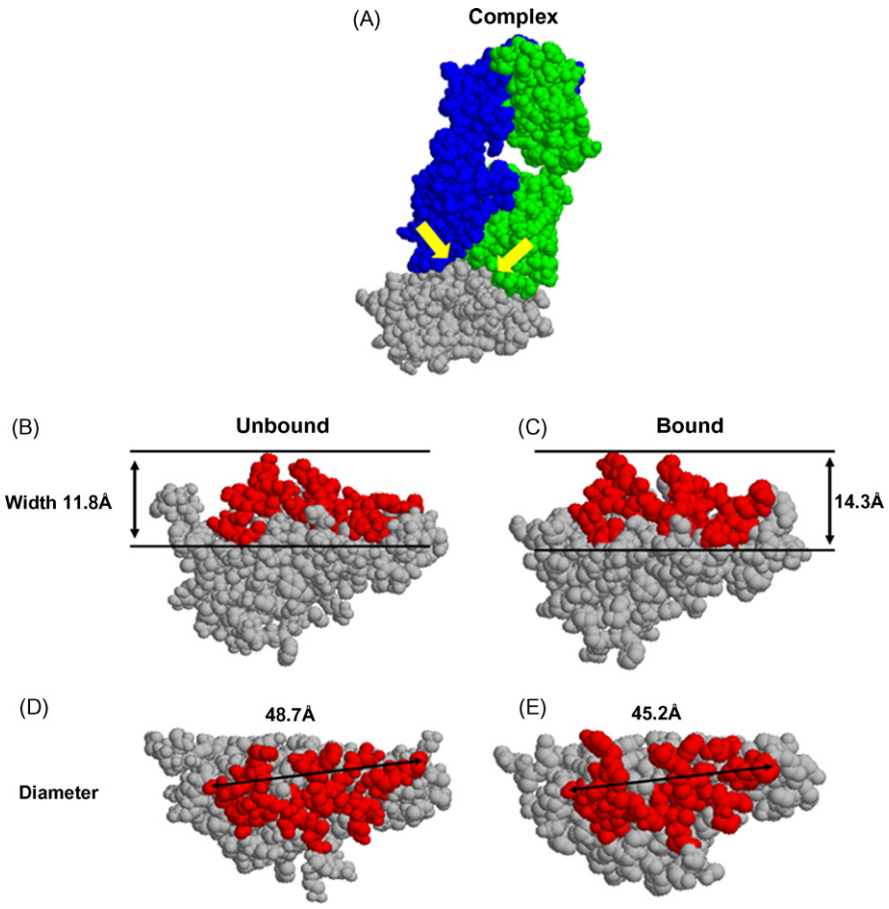

Fig. 8. Illustration of the structural effect on an epitope upon complex formation. (A) Complex between CD40 ligand and the Fab fragment of its neutralizing antibody, humanized 5C8 (PDB identifiers 1i9r and 1aly, for the bound and unbound structures, respectively). Antibody Fab light and heavy chains are colored blue and green, respectively, and the antigen is colored grey. Yellow arrows indicate the axes of the compression force which the CDRs supposedly exert on the epitope. Width of the (B) unbound epitope and (C) bound epitope shown from a side view. Three-dimensional diameter of the (D) unbound epitope and (E) bound epitope shown from a top view. Epitopes are colored in red and their widths and three-dimensional diameters are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

hypothesis, a Bayesian conservation measure (Mayrose et al., 2004) for each amino acid site was computed, and the average conservation was compared between epitope and non-epitope surfaces. In contrast to the above expectation, the results revealed that epitopes are significantly less evolutionarily conserved than non-epitope surfaces ($P = 0.002$; paired t-test).

As described above, epitopes are enriched with unorganized secondary structures (loops). It is claimed that amino acid replacements in surface loops usually do not perturb the three-dimensional structure of the protein since surface loops are relatively flexible (Saunders and Baker, 2002). Thus, the conservation variability of epitopes might be biased by the abundance of loops in epitopes. To examine this option, the conservation analysis was performed again, this time for loops, and non-loops (helices and strands), separately. This analysis showed again the significant variability in epitopes compared to the remaining antigen surfaces. These results imply that epitopes do not tend to overlap other types of functional patches, but rather encompass separate regions (see Section 4).

Several additional analyses which were performed did not find epitopes to be significantly different from non-epitope surfaces. To examine whether electrostatic interactions are a major driving force of the antibody–antigen interaction, the electrostatic potential was compared between unbound epitope and non-epitope surfaces, using the adaptive Poisson–Boltzmann solver (APBS) (Baker et al., 2001). In addition, to test whether epitopes are characterized with high backbone flexibility, the average temperature factor value along the polypeptide chain was compared between unbound epitope and non-epitope surfaces. Moreover, to explore the possibility that water molecules contribute to the chemical complementarity of an interacting antibody–antigen pair, the disposition of water molecules surrounding unbound epitope and non-epitope surfaces was compared. The detailed methodology of the above three analyses are not given.

4. Discussion

The key question addressed in this study is what are the features of epitopes that distinguish them from the remaining antigen surface? Although earlier works were able to define key features of antigenicity, they were based on limited data. Thus, they were limited in their ability to statistically determine whether a specific feature truly distinguishes an epitope from the

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

Table 3
Epitope properties analyzed in this study

Epitope property	Results based on 53 complexes	Previously published results	Remarks
Size and area	75% of the data span 15–25 residues and an area of 600–1000 Å^2; medians: 20 residues and 790 Å^2	Based on 18 complexes, (Chakrabarti and Janin, 2002) measured the sum of the epitope and paratope interfaces to span 38–66 residues and an area of 1250–2320 Å^2	Based on 3 antigens, (Laver et al., 1990) reported epitopes to span 15–22 residues and an area of 650–900 Å^2
CDR-bound region	90% of the epitope area is bound by CDR residues with no preference for a specific CDR loop		
Segment composition	The majority of epitopes are composed of up to five approximately linear segments	Epitopes are strictly classified as linear or conformational (Berzofsky, 1985)	A novel classification of linear vs. conformational epitopes is suggested in this study
Amino acid preference, cooperativity, and side-chain contribution	Epitopes are enriched with Y, W, charged, and polar amino acids, and with specific amino acid pairs: Y:Y, Y:N, Y:G, Y:T, Y:R, P:C, P:D[a]. The epitope surface is enriched with amino acids that have exposed side chains, relative to the non-epitope surface	Based on 15 complexes, (Jackson, 1999) reported high frequency of polar, charged, and aromatic residues in epitopes; (Jackson, 1999) also reported that epitopes interact with antibodies mainly through their backbones	This study is the first to apply robust statistical analysis for amino acid preference, and the first to report cooperativity between epitope amino acids
Surface accessibility	Epitope surfaces are significantly more accessible than non-epitope surfaces	Based on 3 antigen structures, (Novotny et al., 1986) reported that peaks of accessibility correlate with epitope locations	
Geometry at the atom, patch, and molecule level	Epitope atoms are more convex than non-epitope atoms. Epitopes are flatter yet rugged and reside on more convex regions of the antigen surface compared to non-epitope surfaces	Based on 6 complexes, (Jones and Thornton, 1997a) reported that epitopes are the most planar and protruding patches on the antigen surface	Novel methods for measuring flatness and convexity were developed in this study
Secondary structure preference	Epitopes are enriched in loops and depleted from helices and beta strands		First reported in this study
Structural effect upon complex formation	Complex formation induces epitope compression		First reported in this study
Evolutionary conservation	Epitopes are more evolutionary variable compared to non-epitope surfaces		First reported in this study
Electrostatic potential	No significant difference between epitope and non-epitope surfaces		First analyzed in epitopes in this study
Backbone flexibility	No significant difference between epitope and non-epitope surfaces	Based on 3 antigen structures, (Westhof et al., 1984) reported that peaks of temperature-factor values correlate with epitope locations	
Water molecules disposition	No significant difference between epitope and non-epitope surfaces		First analyzed in epitopes in this study

[a] Single letter abbreviation of amino acids.

remaining antigen surface. In this work, a comprehensive analysis of epitope characteristics was conducted combined with the development of novel computational techniques for this purpose. The main findings of this study are summarized in Table 3.

It should be noted that although all available antibody–antigen co-crystal structures were assembled in this study, the data bear several inherent limitations. First, the data may be biased towards specific proteins of interest and those for which the crystal structure could be obtained, such as pathogenic globular proteins. Second, based on these data the results of this study cannot separate immunogenic important traits from antigenic ones. In other words, it cannot be concluded that the epitope characteristics highlighted here are precisely those which allow an immune response to be realized. Third, the role of post translation modifications in epitopes cannot be assessed since it is absent from the crystal data. Forth, solved crystal structures are embedded in a lattice where crystal contacts can potentially modify the surface of the molecule. If epitopes are either enriched or depleted of such contacts relative to the remaining antigen surface, this may bias the structural analyses preformed on the unbound dataset. Regarding this limitation however it was previously reported that unlike biological interfaces, crystal contacts do not show unique characteristics (Valdar and Thornton, 2001). Furthermore, no significant difference was found when comparing the occurrences of crystal contacts in epitope versus non-epitope surfaces (see Supplementary material for details).

All through the analyses performed in this investigation the epitope was compared to the remaining antigen surface. It is

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

likely that the remaining antigen surface encompasses additional epitopes, not accounted for in the data. Ideally, all epitopes should be compared to all non-epitope regions. The possibility that additional epitopes exist in the remaining antigen surface is expected to bias the results towards the null hypothesis, i.e., that epitope and non-epitope surfaces are equally antigenic. The fact that epitopes were found to be significantly different from the remaining surfaces, in spite of the potential bias, strengthens the notion that epitopes have distinguishable characteristics which are in fact underestimated.

To exemplify the essence of the distinction between epitope and non-epitope surfaces we chose to focus on the lysozyme antigen. Our data include five structures of lysozymes from different species (all from the Aves class), co-crystallized with different antibodies (PDB identifiers: 1a2y, 1bql, 1dqj, 1fbi, 1jhl). Superposition of the five epitopes onto an unbound lysozyme structure (PDB identifier 1hel) reveals that they mildly overlap and together cover a contiguous region that is approximately two thirds of the entire lysozyme surface. Thus, the epitopes-excluded area also forms a contiguous patch on the lysozyme surface. Comparison of the epitopes-excluded area to the epitopes area reveals that the two regions are essentially different. Compared to the epitopes area, the epitopes-excluded area is less accessible to the solvent, less convex at the atom level, and lies on relatively depressed areas of the molecule. In addition, the proportions of charged, polar, and aromatic residues (such as arginine, aspartate, aspargine, tryptophan, and tyrosine) and of unorganized secondary structure elements, are much lower in the epitopes-excluded area. Moreover, the epitopes-excluded area tends to include residues that are more evolutionary conserved, either near the active site (Boeckmann et al., 2003) or at other locations. Assuming that these characteristics are important for antibody binding, this comparison thus stresses the point that the protein surface is not homogeneously antigenic, but is rather composed of regions that vary considerably according to their antigenic potential.

Integrating the characteristics examined above reveals a relatively concise image of a typical epitope. Such a protein surface region covers approximately 20 amino acids and is composed of 2–6 linear fragments. It is enriched with tyrosine, tryptophan, charged, and polar residues and is depleted of hydrophobic ones. These epitope-enriched residues expose their side chains to a higher extent compared to the remaining antigen surface, thus rendering them chemically active and available for interaction. Epitopes reside on convex and flat parts of the antigen surface and thus are highly accessible to the CDR of the antibody. The flexibility of an epitope, inferred by comparing the epitope geometry before and after antibody binding, is manifested by its enrichment with unorganized secondary structures. A flexible protein region is probably preferred by the antibody as it maintains a movement capacity needed to form the strong interaction bond. We speculate that the ruggedness of the epitope surface further supports this strong chemical interaction.

The inability to distinguish epitopes from non-epitope surfaces according to temperature factor values may be viewed as inconsistent with results presented above pointing at epitope flexibility. Assuming the temperature factor is a reliable measure of flexibility, perhaps epitopes are not as flexible as suggested. Yet, it has been claimed that the temperature factor may not provide a reliable indication to flexibility (Saunders and Baker, 2002). Alternatively, crystallized protein structures might be biased towards non-flexible proteins that are easy to crystallize. However, temperature factor analysis can only be performed on crystal structures hence this possible bias is unavoidable. Altogether, the conclusion that epitopes are highly flexible is not necessarily undermined by the lack of support of the temperature factor analysis. Another property according to which epitopes cannot be distinguished from remaining antigen surfaces is the disposition of water molecules. Although water molecules probably play a role in the antibody–antigen interaction, it is uncertain whether their distribution around the unbound antigen promotes antibody binding.

Comparison between epitopes and other types of protein–protein interfaces in transient hetero-complexes with respect to physico-chemical and structural characteristics (Bogan and Thorn, 1998; Jones and Thornton, 1995, 1996, 1997a; Neuvirth et al., 2004), revealed general similarities between the two types of protein regions. Namely, preference for tyrosine and tryptophan residue and unorganized secondary structures was found to characterize both types of interactions. However, whereas hydrophobic amino acids are abundant in protein–protein interfaces and thus probably play an important role in the interaction (Lo Conte et al., 1999; Neuvirth et al., 2004), this is not the case in epitopes as they are specifically depleted of hydrophobic residues. Other than that, the main differences between protein–protein interfaces and epitopes are the geometrical properties and the evolutionary signal.

It is well documented that evolutionary conservation sharply distinguishes surface regions that have a functional role, such as protein–protein interfaces (Aloy et al., 2001; Landgraf et al., 2001; Neuvirth et al., 2004). The opposite pattern was detected here for epitopes. The lack of conservation can be partially explained by the enrichment of loops in epitopes, as they are relatively tolerant to amino acid replacements. In protein–protein interfaces, the lack of selection is probably balanced by purifying selection acting to maintain functional interactions, whereas in epitopes this purifying selection force is irrelevant. The lack of evolutionary conservation was found both for epitopes originating from pathogens and species with an adaptive immune system (see Supplementary material). An additional explanation for the lack of epitope conservation, in the latter case, involves self-tolerance. Self-tolerance is the result of the clonal selection process by which the potential of the host's immune system to react against self-determinants is eliminated. As conserved antigen regions may be present in the host itself they are expected to have a lower potential of eliciting an immune response.

Antibody–antigen interaction is a type of protein–protein interaction that potentially spans an infinite range of participating molecules. Nevertheless, the analysis performed here was clearly able to underline characteristics that are inherent to this type of molecular recognition. On the theoretical side, the conclusions derived in this work clearly extend the understanding of what are the ingredients for antibody recognition.

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016

Acknowledgments

This work was supported by the Wolfson Foundation, an ISF grant no. 1208/04, and a grant from the Israeli Ministry of Science (to T.P.); an ISF grant (to J.M.G.); by the IST Programme of the EU as Shared-cost RTD (FET Open) Project under contact no. IST-006413 (ACS, Algorithms for Complex Shapes), by an ISF grant no. 236/06, and by the Hermann Minkowski-Minerva Center for Geometry at Tel Aviv University (for D.H.); by the Edmond J. Safra Program in Bioinformatics at Tel Aviv University (for N.D.R.). We thank Efi Fogel and Angela Enosh for their assistance with the geometrical analyses, and Joel Hirsch for his suggestions and insightful comments. We also thank Adi Doron-Faigenboim, Adi Stern, David Burstein, Eyal Privman, Julian Dutheil, Ofir Cohen, and Osnat Penn for critically reading the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molimm.2007.10.016.

References

Allcorn, L.C., Martin, A.C., 2002. SACS—self-maintaining database of antibody crystal structure information. Bioinformatics 18, 175–181.

Aloy, P., Querol, E., Aviles, F.X., Sternberg, M.J.E., 2001. Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. J. Mol. Biol. 311, 395–408.

Arnon, R., Sela, M., 1969. Antibodies to a unique region in lysozyme provoked by a synthetic antigen conjugate. Proc. Natl. Acad. Sci. U.S.A. 62, 163–170.

Atassi, M.Z., 1975. Antigenic structure of myoglobin: the complete immunochemical anatomy of a protein and conclusions relating to antigenic structures of proteins. Immunochemistry 12, 423–438.

Atassi, M.Z., 1978. Precise determination of the entire antigenic structure of lysozyme: molecular features of protein antigenic structures and potential of "surface-simulation" synthesis—a powerful new concept for protein binding sites. Immunochemistry 15, 909–936.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. 98, 10037–10041.

Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M., et al., 1984. The antigenic structure of proteins: a reappraisal. Annu. Rev. Immunol. 2, 67–101.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. 57, 289–300.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucl. Acids Res. 28, 235–242.

Berzofsky, J.A., 1985. Intrinsic and extrinsic factors in protein antigenic structure. Science 229, 932–940.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M., 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucl. Acids Res. 31, 365–370.

Bogan, A.A., Thorn, K.S., 1998. Anatomy of hot spots in protein interfaces. J. Mol. Biol. 280, 1–9.

Bublil, E.M., Tarnovitski Freund, N., Mayrose, I., Penn, O., Roitburd-Berman, A., Rubinstein, N.D., Pupko, T., Gershoni, J.M., 2007. Stepwise prediction of conformational discontinuous B-cell epitopes using the mapitope algorithm. Proteins: Struct. Funct. Genet. 68, 294–304.

Chakrabarti, P., Janin, J., 2002. Dissecting protein–protein recognition sites. Proteins: Struct. Funct. Genet. 47, 334–343.

Enshell-Seijffers, D., Denisov, D., Groisman, B., Smelyanski, L., Meyuhas, R., Gross, G., Denisova, G., Gershoni, J.M., 2003. The mapping and reconstitution of a conformational discontinuous B-cell epitope of HIV-1. J. Mol. Biol. 334, 87–101.

Hopp, T.P., 1986. Protein surface analysis. Methods for identifying antigenic determinants and other interaction sites. J. Immunol. Methods 88, 1–18.

Hopp, T.P., Woods, K.R., 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. 78, 3824–3828.

Irving, M.B., Pan, O., Scott, J.K., 2001. Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. Curr. Opin. Chem. Biol. 5, 314–324.

Jackson, R.M., 1999. Comparison of protein–protein interactions is serine protease-inhibitor and antibody–antigen complexes: implications for the protein docking problem. Protein Sci. 8, 603–613.

Jemmerson, R., Margoliash, E., 1979. Topographic antigenic determinants on cytochrome *c*. Immunoadsorbent separation of the rabbit antibody populations directed against horse cytochrome. J. Biol. Chem. 254, 12706–12716.

Jemmerson, R., Paterson, Y., 1985. Mobility and evolutionary variability factors in protein antigenicity. Nature 317, 89–90.

Jerne, N.K., 1960. Immunological speculations. Annu. Rev. Microbiol. 14, 341–358.

Jones, S., Thornton, J.M., 1995. Protein–protein interactions: a review of protein dimer structures. Prog. Biophys. Mol. Biol. 63, 31–65.

Jones, S., Thornton, J.M., 1996. Principles of protein–protein interactions. Proc. Natl. Acad. Sci. U.S.A. 93, 13–20.

Jones, S., Thornton, J.M., 1997a. Analysis of protein–protein interaction sites using surface patches. J. Mol. Biol. 272, 121–132.

Jones, S., Thornton, J.M., 1997b. Prediction of protein–protein interaction sites using patch analysis. J. Mol. Biol. 272, 133–143.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.

Landgraf, R., Xenarios, I., Eisenberg, D., 2001. Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. J. Mol. Biol. 307, 1487–1502.

Laver, W.G., Air, G.M., Webster, R.G., Smith-Gill, S.J., 1990. Epitopes on protein antigens: misconceptions and realities. Cell 61, 553–556.

Lilienfeld, D.E., Stolley, P.D., 1994. Foundations of Epidemiology, third edition, pp. 322–324.

Lo Conte, L., Chothia, C., Janin, J., 1999. The atomic structure of protein–protein recognition sites. J. Mol. Biol. 285, 2177–2198.

Ma, B., Elkayam, T., Wolfson, H., Nussinov, R., 2003. Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc. Natl. Acad. Sci. U.S.A. 100, 5772–5777.

Mayrose, I., Graur, D., Ben-Tal, N., Pupko, T., 2004. Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. Mol. Biol. Evol. 21, 1781–1791.

Miller, S., Janin, J., Lesk, A.M., Chothia, C., 1987. Interior and surface of monomeric proteins. J. Mol. Biol. 196, 641–656.

Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.

Neuvirth, H., Raz, R., Schreiber, G., 2004. ProMate: a structure based prediction program to identify the location of protein–protein binding sites. J. Mol. Biol. 338, 181–199.

Nimrod, G., Glaser, F., Steinberg, D., Ben-Tal, N., Pupko, T., 2005. In silico identification of functional regions in proteins. Bioinformatics 21, i328–i337.

Novotny, J., Handschumacher, M., Haber, E., Bruccoleri, R.E., Carlson, W.B., Fanning, D.W., Smith, J.A., Rose, G.D., 1986. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). Proc. Natl. Acad. Sci. U.S.A. 83, 226–230.

Pellequer, J.L., Westhof, E., Van Regenmortel, M.H., 1991. Predicting location of continuous epitopes in proteins from their primary structures. Methods Enzymol. 203, 176–201.

Saunders, C.T., Baker, D., 2002. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J. Mol. Biol. 322, 891–901.

Shindyalov, I.N., Bourne, P.E., 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. 11, 739–747.

Shoemaker, B.A., Panchenko, A.R., 2007. Deciphering protein–protein interactions. Part I. Experimental techniques and databases. PLoS Comput. Biol., 3.

Simpson, E.H., 1951. The interpretation of interaction in contingency tables. J. Roy. Stat. Soc. 13, 238–241.

Sobolev, V., Sorokine, A., Prilusky, J., Abola, E.E., Edelman, M., 1999. Automated analysis of interatomic contacts in proteins. Bioinformatics 15, 327–332.

Sokal, R.R., Rohlf, J.F., 1995. Biometry, third edition, pp. 686–695.

Thornton, J.M., Edwards, M.S., Taylor, W.R., Barlow, D.J., 1986. Location of 'continuous' antigenic determinants in the protruding regions of proteins. EMBO J. 5, 409–413.

Tsodikov, O.V., Record Jr., M.T., Sergeev, Y.V., 2002. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. J. Comput. Chem. 23, 600–609.

Valdar, W.S., Thornton, J.M., 2001. Conservation helps to identify biologically relevant crystal contacts. J. Mol. Biol. 313, 399– 416.

Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A., Van Regenmortel, M.H., 1984. Correlation between segmental mobility and the location of antigenic determinants in proteins. Nature 311, 123–126.

Westwood, O.M.R., Hay, F.C., 2001. Epitope Mapping: A Practical Approach.

Zhou, H.X., Shan, Y., 2001. Prediction of protein interaction sites from sequence profile and residue neighbor list. Proteins: Struct. Funct. Genet. 44, 336–343.

Please cite this article in press as: Rubinstein, N.D., et al., Computational characterization of B-cell epitopes, Mol. Immunol. (2007), doi:10.1016/j.molimm.2007.10.016