Bioinformatics, 32(16), 2016, 2451–2456
doi: 10.1093/bioinformatics/btw197
Advance Access Publication Date: 19 April 2016
Original Paper

Structural bioinformatics

Shape complementarity and hydrogen bond preferences in protein–protein interfaces: implications for antibody modeling and protein–protein docking

Daisuke Kuroda[1,2] and Jeffrey J. Gray[1,3,*]

[1]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA, [2]Department of Analytical and Physical Chemistry, Showa University School of Pharmacy, Tokyo, 142-8555, Japan and [3]Program in Molecular Biophysics, Johns Hopkins University, Baltimore, MD 21218, USA

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on July 16, 2015; revised on February 28, 2016; accepted on April 3, 2016

Abstract

Motivations: Characterizing protein–protein interfaces and the hydrogen bonds is a first step to better understand proteins' structures and functions toward high-resolution protein design. However, there are few large-scale surveys of hydrogen bonds of interfaces. In addition, previous work of shape complementarity of protein complexes suggested that lower shape complementarity in antibody–antigen interfaces is related to their evolutionary origin.

Results: Using 6637 non-redundant protein–protein interfaces, we revealed peculiar features of various protein complex types. In contrast to previous findings, the shape complementarity of antibody–antigen interfaces resembles that of the other interface types. These results highlight the importance of hydrogen bonds during evolution of protein interfaces and rectify the prevailing belief that antibodies have lower shape complementarity.

Contact: jgray@jhu.edu

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Interactions are one of the most fundamental activities of biomolecules. Disturbance of these interactions underlie biological disorders including cancers and neurodegenerative diseases. Characterizing interactions is important to understand the detailed mechanisms of life (Russell and Aloy, 2008).

Complementarity is a key concept in molecular recognition. Previous studies have extracted common features of protein–protein interfaces and revealed peculiar features of particular classes of interfaces. Traditional structural surveys of protein–protein interactions often divided complexes into hetero-dimers and homo-dimers and into different functional classes, such as antibody–antigen, enzyme–inhibitor/substrate and 'other' complexes (Jones, 2012; Lo Conte *et al.*, 1999; Ofran and Rost, 2003; Vreven *et al.*, 2015). A key driving force of molecular recognition and specificity is forming hydrogen bonds and salt bridges. Unlike salt bridges, hydrogen bonds are orientation-dependent. Hence, modulating hydrogen bond networks is critical for specificity in protein engineering and design (Stranges and Kuhlman, 2013). Hydrogen bonds are also crucial for protein folding where the patterns formed define secondary structure elements. Based on 319 protein complexes, Nussinov and co-workers reported that backbone–backbone hydrogen bonds are dominant within protein structures whereas side-chain–side-chain hydrogen bonds are most common across protein–protein interfaces (Xu *et al.*, 1997).

Another important factor in the complementarity of molecular recognition is the shape of the interfaces. There are several analyses on shape complementarity of protein interfaces (Lawrence and Colman, 1993; Tsuchiya *et al.*, 2006). In such an analysis based on

15 of crystal structures including six antibody–antigen complexes, it was proposed that the shape complementarity of antibody–antigen interfaces was lower on average than that of protein–protein interfaces in general (Lawrence and Colman, 1993). Antibodies evolve independently of foreign antigens, while other proteins involved in protein–protein interactions typically evolve with their counterparts, so that the shape complementarity of general interfaces is optimized by both partners. However, a later analysis with two high-resolution crystal structures suggested that the observed lower shape complementarity of antibody–antigen complexes was due to the low quality of the crystal structures (Cohen *et al.*, 2005). Since the discussions were based on a limited number of protein–protein complexes, a more comprehensive analysis of the complementarities of the interfaces is required to understand the essential nature of molecular recognition and to guide the rational design of protein therapeutics. The question is relevant today as computational design efforts strive to create *de novo* interfaces by optimizing only one protein in the pair, as antibodies do. Antibodies provide a natural system that can demonstrate what level of shape complementarity is possible in single-sided interface design.

In this study, we focus on hydrogen bond networks using 6637 non-redundant protein complexes. Our main dataset includes 547 antibody V_L–V_H (light and heavy variable domain) pairs, 191 antibody–protein antigen complexes, 104 antibody–peptide antigen complexes, 88 enzyme–inhibitor/substrate complexes, 102 'other' complexes and 92 obligate complexes. For a reference, we also analyzed 2251 and 3262 protein–protein interfaces of hetero- and homodimers. The analyses reveal peculiar features of each protein complex type. We discuss the impact on antibody modeling, docking simulations and interface designs.

2 Methods

2.1 Protein interfaces analyzed

The non-redundant antibody structures and the complexes, enzyme–inhibitor/substrate, 'other' complexes, obligate complexes, hetero- and homo-dimer complexes were taken from the PDB through SAbDab (Dunbar *et al.*, 2014), Docking Benchmark 5.0 (Vreven *et al.*, 2015), a work of obligate complexes (Mintseris and Weng, 2005) and the 3D-complex database (Levy *et al.*, 2006), respectively. Antibody structures were renumbered by Chothia's standard. Details of the dataset preparation are described in the Supplementary Materials.

2.2 Hydrogen bond detection and shape complementarity

Hydrogen bonds and shape complementarity by the S_c measure were calculated using the Rosetta libraries (Leaver-Fay *et al.*, 2011). To detect hydrogen bonds in protein–protein interfaces, we also used HBPLUS (McDonald and Thornton, 1994). Since the results of Rosetta and HBPLUS were comparable, we report only the Rosetta results. The Rosetta command lines are available in the supplementary text.

3 Results

3.1 Asymmetric usages of the backbone and side chains in antibody hydrogen bond networks

To study the pattern of hydrogen bonds in interfaces, we first focused on the usage of backbone and side-chain polar atoms. As previously observed (Xu *et al.*, 1997), all interfaces other than enzyme–inhibitor/substrate complexes showed dominant usage of side-chain–side-chain hydrogen bonds (34–60%) (Fig. 1). Interestingly, asymmetric usage of backbone and side chain was observed in the **V_L–V_H interfaces of antibodies**. That is, the V_H domain uses the backbone while the V_L domain mainly uses the side chain to form hydrogen bonds in the interface (Fig. 1A). Although the number of hydrogen bonds is small in V_L–V_H interfaces (average 3.7 ± 1.7), this trend is observed broadly across different antibodies in the dataset with varying germline genes. We analyzed the V_L–V_H interfaces as a separate category since V_L–V_H antibodies are a family of closely related proteins. Proteins sequences evolve over time. Likewise, but in much shorter time scale, the V_L and V_H sequences diverge from germline genes of light and heavy chains, respectively. The sequence identities in each domain family typically vary between 40 and >90%. To eliminate a potential bias caused by germline gene family redundancy, we repeated the same analysis using the 153 antibodies that have unique germline gene family pairings in our dataset. Even after removing the redundancy, the observation still holds (Supplementary Figure S1).

Visual inspection revealed that 94% (595/634) of these V_H backbone–V_L side chain hydrogen bonds involved the backbone of complementarity determining region (CDR) H3 (H95–H102). This observation can be viewed in light of the antibody evolution. Relative to V_L, V_H, especially the CDR-H3, is more diverse in terms of both sequence and conformation and more often involved with antigen recognition. Thus, to maintain the V_L–V_H association, a V_H domain must use the backbone to form proper hydrogen bonds. On the other hand, the light chain is relatively conserved in terms of sequence including CDRs, and it can use the side chain to form hydrogen bonds without losing these interactions during the evolution. This implies that the sequences, especially in V_L domain, encode the

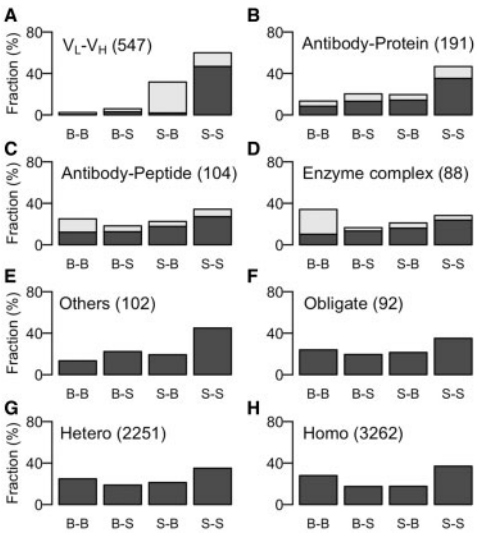

Fig. 1. Hydrogen bond type across protein–protein interfaces. The number of interfaces is given in the parenthesis. **(A)** V_L–V_H antibodies. **(B)** Antibody–protein antigens. **(C)** Antibody–peptide antigens. **(D)** Enzyme–inhibitor/substrates. **(E)** 'Other' complexes. **(F)** Obligate complexes. **(G)** Hetero dimers. **(H)** Homo dimers. CDR-H3-mediated and all other hydrogen bonds in antibodies are colored in gray and black, respectively. Protease-mediated and all other hydrogen bonds in enzyme complexes are colored in gray and black respectively. B-B, B-S, S-B and S-S correspond to backbone–backbone, backbone–side-chain, side-chain–backbone and side-chain–side-chain interactions, respectively. The first letter indicates V_L domain, antibody, enzyme, or other proteins and the second letter indicates V_H domain, antigen, inhibitor/substrate, or other proteins

correct V_L–V_H orientation during the antibody evolution, also emphasizing the importance of the CDR-H3 backbone conformation on the domain orientations.

A dominant hydrogen bond in V_L–V_H domains was between the side chains of two glutamines at positions L38 and H39. This pair is conserved in 86% (471/547) of the V_L–V_H dataset, and 88% (413/471) of those side-chain pairs form hydrogen bonds. Despite the high conservation in the sequences, an experiment has demonstrated that the L38:Gln–H39:Gln hydrogen bond does not correlate with antibody stability (Tan *et al.*, 1998). Analysis on the relative orientation of the V_L–V_H domains of antibodies has become an active area of research (Abhinandan and Martin, 2010; Chailyan *et al.*, 2011; Dunbar *et al.*, 2013; Narayanan *et al.*, 2009), along with the demand for high-resolution antibody modeling (Almagro *et al.*, 2011). In light of this trend in the field, our analysis characterized V_L–V_H interfaces by analyzing hydrogen bond networks, and our observation may be useful for the correct sampling of V_L–V_H domain orientations. All hydrogen bonds in V_L–V_H interfaces are enumerated in Supplementary Table S2.

Structure prediction of CDR-H3 is also a hot topic in antibody modeling field (Almagro *et al.*, 2011). Conformations of CDR-H3 can be classified into kink or extend forms (Kuroda *et al.*, 2008), and a few hydrogen bonds, such as L36:Tyr, appear to contribute these conformations. However, these residues are not necessarily conserved when CDR-H3 stems assume the kink conformation (Kuroda *et al.*, 2008), and these hydrogen bonds in the V_L–V_H interfaces alone do not define either kinked or extended conformations (Weitzner *et al.*, 2015). The kinked or extended region is defined as four consecutive Cα atoms (H100X–H103) in antibodies having CDR-H3 loops longer than four residues. In our V_L–V_H set, there are 529 such antibodies, where kinked or extended conformations are definable. When hydrogen bonds are broken down in the V_L–V_H interfaces, 18.1% (354/1951) of them involve the kinked or extended regions, 30.5% of them involve the rest of the H3 loop and 39.4% of them are between L38:Gln–H39:Gln. The remaining hydrogen bonds are in the other parts of the V_L–V_H domains.

The difficulty of CDR-H3 modeling arises mainly from the fact that CDR-H3 is governed by V(D)J recombination. Although D and J germline gene segments assignments are difficult, we can predict the V genes using IgBLAST (Ye *et al.*, 2013). We found that, in 10 897 somatic mutations of V genes in our V_L–V_H set, only 0.95% (104/10 897) of the mutations contribute to hydrogen bonds in the V_L–V_H interface. In other words, most of the hydrogen bonds in V genes are encoded already in germline sequences, in agreement with our observations that V_L domains, which are less variable, tend to use their side chains to form hydrogen bonds.

The distributions of the number of hydrogen bonds in all types of the interfaces are available in Supplementary Figure S2. Overall, the number of hydrogen bonds in the V_L–V_H interfaces (average 3.7) was smaller compared with other interfaces (average 6.2–15.6), reflecting the hydrophobic nature of the V_L–V_H interfaces. This observation is consistent with the higher fraction of buried non-polar atoms in V_L–V_H interfaces compared with other interfaces (Supplementary Figure S3).

The average hydrogen bond densities (number of hydrogen bonds per 100 $Å^2$ buried surface area) in each interface are given in Supplementary Table S3. V_L and V_H domains and other obligate complexes do not exist in isolation, i.e. they are always associated to form a complex. However, a critical difference between V_L–V_H and obligate interfaces is that the V_L–V_H domain pairing is generally considered to be random, and V_H domains can associate with several different V_L domains, which is known as receptor editing (de Wildt *et al.*, 1999). The smaller number of hydrogen bonds may enable the quasi-promiscuous binding.

In **antibody–protein antigen interfaces**, as expected, there were more hydrogen bonds to the V_H domain than to the V_L domain (5.2 ± 3.0 and 2.4 ± 2.0, respectively) (Supplementary Figure S2), emphasizing the importance of the V_H domain in antigen recognition. Forty two percent (418/987) of the V_H-mediated hydrogen bonds involved CDR-H3, followed by CDR-H2 (H52–H56) and FR_H-mediated hydrogen bonds (24 and 23%, respectively). In the FR_H cases, most of the hydrogen bonds were formed by the residues adjacent to CDRs, such as H58 (in 42 of 231 FR_H-mediated hydrogen bonds, or 18%), H50 (12%), H33 (12%) and H61 (6%). Similarly to the V_L–V_H interfaces, we also found that, in 4,460 somatic mutations of V genes in our antibody-protein set, only a small fraction ($163/4460 = 3.7\%$) of the mutations contribute to hydrogen bonds in the V genes. Note that, in the antibody-protein interfaces, about half (49%) of the hydrogen bonds are mediated by CDR-H3 and L3, both of which are generated by V(D)J recombination, and we investigated only V genes here. Unlike V_L–V_H interfaces, most of the interactions in antibody-antigen interfaces are via CDRs. Hence, it is still possible that hydrogen bonds to antigens would be formed via somatic mutations in CDR-H3 and L3.

In **antibody–peptide interfaces**, the number of backbone–backbone hydrogen bonds was comparable to the number of backbone–side-chain or side-chain–backbone hydrogen bonds (Fig. 1C). A possible explanation is that peptide antigens are smaller than protein antigens, and the number of residues that can contribute to interactions with the antibody is smaller. As a result, the backbone of peptide antigens needs to contact the antibody for favorable binding. On the antibody side, as previously reported in Bates *et al.* (1998), peptide recognition of antibodies can be divided into two structural classes depending on the length of CDR-H3; when H3 is short (<7 residues), the peptide conformations are extended on the pocket formed by the six CDR loops whereas when H3 is long (≥7 residues), the peptides are buried into the pocket (Supplementary Figure S4). Due to the spatial constraint, the backbones of antibodies more easily contact the peptides, especially in the latter case, where, on average, there are more hydrogen bonds between backbones of antibodies and the peptide antigens than the first class of antibody–peptide complexes (1.6 ± 1.3 and 2.9 ± 2.1 for short and long H3, respectively).

It is sometimes difficult for epitope scanning (Jemmerson, 1987) to recognize a conformational epitope since tertiary interactions within the protein stabilize the peptide conformation; these interactions are absent in the lone peptide. So segmented proteins have unsatisfied backbone hydrogen bond donors/acceptors, and antibodies that recognize peptides may have to make backbone contacts to stabilize the peptide conformation. Some antibodies may bind a lone epitope peptide in a different conformation than the binding to the epitope in the parent-protein context.

3.2 Backbone–backbone hydrogen bonds are common in protease–inhibitor complexes, but not in the other enzyme–inhibitor/substrate complexes

Unlike the other interfaces, the most frequent hydrogen bonding pattern in enzyme complexes was the backbone–backbone pair (Fig. 1D). When dividing the enzymes into proteases and other types, the dominant hydrogen bonds in the protease-inhibitor complexes were backbone–backbone hydrogen bonds whereas those in the other types of enzymes came from the side-chain–side-chain pairs, as seen in other interfaces. This observation is probably a

reflection of the difference in the biological roles: proteases cleave peptide backbones of substrates, whereas other enzymes typically recognize the other types of bonds, such as glycosidic or phosphodiester ones.

In a previous work (Jackson, 1999), Jackson computed interaction energies comprising van der Waals and electrostatic terms in a molecular mechanics formulation, showing that, in the 12 protease–inhibitor complexes studied, the energies from backbone–backbone interactions account for the highest fraction (41%) of the total energies. However, although hydrogen bonds could be considered as electrostatic interactions in general, in his calculations, the electrostatic component of the protease–inhibitor interactions only account for 24% of the total energies. Perhaps the molecular mechanics formulation could not fully describe hydrogen bonds because of the lack of directionality in the functional form. Indeed, an orientation-dependent hydrogen bonding potential in Rosetta discriminated native interfaces better than molecular mechanics calculations (Kortemme *et al.*, 2003), particularly due to hydrogen bond specificity.

The dominant contribution of backbone–backbone hydrogen bonds is reminiscent of formation of secondary structure within protein structures (Pauling and Corey, 1951). Supplementary Figure S5 shows the examples of these hydrogen bonds. In these cases, the backbones of enzymes and inhibitors become closer upon binding. Thus, we next analyzed the shape complementarity of the interfaces.

3.3 Shape complementarity of antibody-protein antigen complexes is close to those of generic proteins

Figure 2 shows the shape complementarity by the S_c measure of eight types of protein–protein interfaces. Contrary to previous findings (Lawrence and Colman, 1993), these data indicate that the shape complementarity of antibody–protein interfaces is comparable to those of the enzyme complex, obligate, hetero- and homodimer interfaces while the 'other' interfaces show less complementarity (Wilcoxon rank-sum test, $P < 0.01$). It has been proposed that the antibody–antigen complex interface shape depends on antigen types: groove-like binding is common in antibody–peptide interactions whereas antibody–protein interfaces tend to be more flat (Lee *et al.*, 2006; MacCallum *et al.*, 1996). Although antibody–peptide interfaces do exhibit the highest shape complementarity (0.75 ± 0.06, $P < 0.01$), this observation may arise from the fact that S_c can be affected by edge effects. When S_c is computed, grid points of molecular surface are defined and a local geometric quantity is extracted from them. S_c is the median value of the local geometric quantity after removing grid points on the edge of the interface. Many more edge points could be removed in small interfaces or the elongated shape of peptide interfaces than in larger protein interfaces (Janin *et al.*, 2008), which may lead to the higher S_c values of peptide-mediated interfaces. Shape complementarity of V_L–V_H domain interface are higher than the other interfaces (other than antibody–peptide ones) ($P < 0.01$), possibly due to the obligate nature of the interfaces.

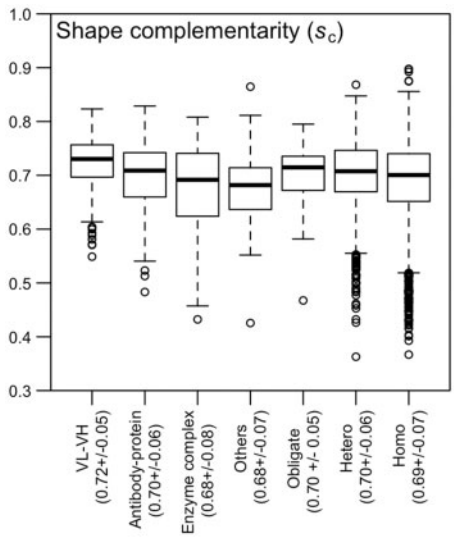

Fig. 2. Shape complementarity (S_c) of protein–protein interfaces. Higher values mean better complementarity. The result of antibody-peptide interfaces was not shown because S_c measure can suffer from the edge effect of small interfaces. For box-and-whiskers plots, medians are shown by the thick line, boxes show the range from the first (Q1) to third (Q3) quartile, whiskers extend to the most extreme data point within 1.5 times the interquartile range of the Q1 and Q3, and points show additional outliers

Protease–inhibitor complexes have higher shape complementarity than the other enzyme–inhibitor/substrate complexes ($P < 0.01$), probably due to the larger number of backbone–backbone hydrogen bonds in the interfaces.

S_c values can suffer poor quality of crystal structures (Cohen *et al.*, 2005). Hence, we repeated the same calculation for 3487 high-resolution crystal structures (Resolution ≤ 2.0 Å), confirming that the shape complementarity of antibody–protein antigen complexes is close to the general interfaces (Supplementary Figure S6).

Shape complementarity of protein interfaces could be defined in various ways (Bahadur *et al.*, 2004). Therefore, we did a second computation based on the gap volume index (Laskowski, 1995) as another metric of shape complementarity (Supplementary Figure S7). Gap volume index also suggested that shape complementarity of antibody–protein interfaces is close to enzyme and 'other' interfaces while these interfaces are less complementary than obligate complexes ($P < 0.01$). The lower complementarities of the antibody, enzyme and 'other' complexes would be explained by the observations that these interfaces are more polar than the obligate ones (Supplementary Figure S3), and polar interfaces often accommodate water molecules to fill the gap regions (Ahmad *et al.*, 2011; Cohen *et al.*, 2005; Lo Conte *et al.*, 1999; Rodier *et al.*, 2005). These gaps or cavities could be filled in by water molecules. By definition, S_c values exploit the median of local geometric quantities based on normal products in interfaces to quantify the relative shape of the surfaces with each other. On the other hand, the gap volume index measures the volumes of the cavities in interfaces and normalizes them by the buried surface areas. Hence, S_c values and gap volume index provide different information on the complementarity of protein–protein interfaces, and both methods may help characterize interfaces in protein docking and design.

4 Discussion and conclusions

What has been already known about protein–protein interfaces in general was that obligate interfaces are more hydrophobic, have better shape complementarity, and evolve slower than transient ones (Keskin *et al.*, 2008; Mintseris and Weng, 2005). Amino acid compositions of interfaces are also well studied; generally, compared with the other surfaces on proteins, protein–protein interfaces are enriched in non-polar and aromatic residues whereas charged residues tend to be depleted (De *et al.*, 2005; Janin *et al.*, 2008). Arg residues are a notable exception in that it is one of the common hot spot residues in interfaces, as well as Trp and Tyr residues (Moreira *et al.*, 2007). Based on 1150 two domain proteins, 583 homodimers and 94 heterodimers, residue-level local network patterns were also

proposed in protein–protein interfaces (Luo *et al.*, 2013). Furthermore, by exploiting all the 57 944 PDB entries available at that time (June 2009), 20 034 atomic-resolution motifs (i.e. 3D organization of atoms) were identified in protein–protein interfaces (Kinjo and Nakamura, 2010). In the past, few studies investigated the hydrogen bonds in interfaces beyond the average numbers and the individual geometries, such as angles and distances (Lo Conte *et al.*, 1999; McDonald and Thornton 1994; Xu *et al.*, 1997). In the case of antibodies specifically, there are some studies about hydrogen bonds in antibody–antigen (protein) interfaces (Dalkas *et al.*, 2014; Peng *et al.*, 2014; Ramaraj *et al.*, 2012), but few discussed the insights into evolution (i.e. somatic maturations) and peptide immunizations, nor analyzed V_L–V_H domain interface, as we have done in this study. The results in this report have implications for antibody modeling and protein–protein docking problems. Since hydrogen bonds via backbones can be preserved after mutations, asymmetry of hydrogen bonding patterns in V_L–V_H interfaces suggests that mutations in V_H domain might be more tolerated than those in V_L domain in engineering better antibodies. Hydrogen bond density of V_L–V_H interfaces is almost half of those of the other interfaces, highlighting the hydrophobic, promiscuous nature of the domain associations. Although there seems to be no obvious correlations between the number of hydrogen bonds and the geometric descriptors of V_L–V_H orientations (Supplementary Figure S8) (Dunbar *et al.*, 2013), constraining the conserved L38:Gln–H39:Gln hydrogen bonds during the domain orientation sampling would be a possible approach to better capture correct V_L–V_H orientations. Packing V_L side chains could be critical in CDR-H3 modeling and V_L–V_H positioning. Finally, epitope scanning by fragmentation of antigens may lead to unusual recognition of peptides and/or binding that differs from the binding of the full antigen.

Backbone–backbone hydrogen bonds are common in the protease–inhibitor complexes. An important problem in protein–protein docking is in global sampling i.e., identification of potential-binding sites of protein surfaces. An application of our observations may be to use backbone complementarity based on hydrogen bonds and, possibly, to use information of dehydration of backbone atoms (Fernandez and Scheraga, 2003) to identify binding sites in protease–inhibitor systems.

In contrast to the previous discussion, the shape complementarity of antibody–protein interfaces seems to be similar to those of other interfaces. Through somatic maturation, antibodies can evolve over short time scales, leading to antibodies having higher affinity and stability, even though the evolution toward binding is only on one side of the interface, it is sufficient to reach high shape complementarity typical of other protein–protein interface types (Li *et al.*, 2003). Still, higher gap volume index being considered, water molecules could also play a role to complete the better fit of interfaces. Although modulating hydrogen bonds in computational design is still challenging, mimicking the maturation process, designing proteins with higher shape complementarity by modulating backbone–backbone hydrogen bonds, as observed in the protease–inhibitor complexes, is a promising approach.

Funding

This work was supported by the National Institutes of Health [grant number R01 GM078221] and by the Japan Society for the Promotion of Science [grant number 15H06606 to D.K.].

Conflict of Interest: none declared.

References

Abhinandan,K.R. and Martin,A.C. (2010) Analysis and prediction of VH/VL packing in antibodies. *Protein Eng. Des. Sel.*, **23**, 689–697.

Ahmad,M. *et al.* (2011) Adhesive water networks facilitate binding of protein interfaces. *Nat. Commun.*, **2**, 261.

Almagro,J.C. *et al.* (2011) Antibody modeling assessment. *Proteins*, **79**, 3050–3066.

Bahadur,R.P. *et al.* (2004) A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.*, **336**, 943–955.

Bates,P.A. *et al.* (1998) Conformational analysis of the first observed non-proline cis-peptide bond occurring within the complementarity determining region (CDR) of an antibody. *J. Mol. Biol.*, **284**, 549–555.

Chailyan,A. *et al.* (2011) The association of heavy and light chain variable domains in antibodies: implications for antigen specificity. *FEBS J.*, **278**, 2858–2866.

Cohen,G.H. *et al.* (2005) Water molecules in the antibody-antigen interface of the structure of the Fab HyHEL-5-lysozyme complex at 1.7 A resolution: comparison with results from isothermal titration calorimetry. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 628–633.

Dalkas,G.A. *et al.* (2014) Cation-pi, amino-pi, pi-pi, and H-bond interactions stabilize antigen-antibody interfaces. *Proteins*, **82**, 1734–1746.

De,S. *et al.* (2005) Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. *BMC Struct. Biol.*, **5**, 15.

de Wildt,R.M. *et al.* (1999) Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire. *J. Mol. Biol.*, **285**, 895–901.

Dunbar,J. *et al.* (2013) ABangle: characterising the VH-VL orientation in antibodies. *Protein Eng. Des. Sel.*, **26**, 611–620.

Dunbar,J. *et al.* (2014) SAbDab: the structural antibody database. *Nucleic Acids Res.*, **42**(**Database issue**), D1140–D1146.

Fernandez,A. and Scheraga,H.A. (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc. Natl. Acad. Sci. USA*, **100**, 113–118.

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

Janin,J. *et al.* (2008) Protein-protein interaction and quaternary structure. *Quart. Rev. Biophys.*, **41**, 133–180.

Jemmerson,R. (1987) Antigenicity and native structure of globular proteins: low frequency of peptide reactive antibodies. *Proc. Natl. Acad Sci. USA*, **84**, 9180–9184.

Jones,S. (2012) Computational and structural characterisation of protein associations. *Advances in Exp. Med. Biol.*, **747**, 42–54.

Keskin,O. *et al.* (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.*, **108**, 1225–1244.

Kinjo,A.R. and Nakamura,H. (2010) Geometric similarities of protein-protein interfaces at atomic resolution are only observed within homologous families: an exhaustive structural classification study. *J. Mol. Biol.*, **399**, 526–540.

Kortemme,T. *et al.* (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.*, **326**, 1239–1259.

Kuroda,D. *et al.* (2008) Structural classification of CDR-H3 revisited: a lesson in antibody modeling. *Proteins*, **73**, 608–620.

Laskowski,R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.*, **13**, 323–330. 307-328.

Lawrence,M.C. and Colman,P.M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol.*, **234**, 946–950.

Leaver-Fay,A. *et al.* (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.*, **487**, 545–574.

Lee,M. *et al.* (2006) Shapes of antibody binding sites: qualitative and quantitative analyses based on a geomorphic classification scheme. *J. Org. Chem.*, **71**, 5082–5092.

Levy,E.D. *et al.* (2006) 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.*, **2**, e155.

Li,Y. *et al*. (2003) X-ray snapshots of the maturation of an antibody response to a protein antigen. *Nat. Struct. Biol*., **10**, 482–488.

Lo Conte,L. *et al*. (1999) The atomic structure of protein-protein recognition sites. *J. Mol. Biol*., **285**, 2177–2198.

Luo,Q. *et al*. (2013) Local network patterns in protein-protein interfaces. *PloS One*, **8**, e57031.

MacCallum,R.M. *et al*. (1996) Antibody-antigen interactions: contact analysis and binding site topography. *J. Mol. Biol*., **262**, 732–745.

McDonald,I.K. and Thornton,J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol*., **238**, 777–793.

Mintseris,J. and Weng,Z. (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc. Natl. Acad. Sci. USA*, **102**, 10930–10935.

Moreira,I.S. *et al*. (2007) Hot spots–a review of the protein-protein interface determinant amino-acid residues. *Proteins*, **68**, 803–812.

Narayanan,A. *et al*. (2009) Energy-based analysis and prediction of the orientation between light- and heavy-chain antibody variable domains. *J. Mol. Biol*., **388**, 941–953.

Ofran,Y. and Rost,B. (2003) Analysing six types of protein-protein interfaces. *J. Mol. Biol*., **325**, 377–387.

Pauling,L. and Corey,R.B. (1951) Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl. Acad. Sci. USA*, **37**, 235–240.

Peng,H.P. *et al*. (2014) Origins of specificity and affinity in antibody-protein interactions. *Proc. Natl. Acad. Sci. USA*, **111**, E2656–E2665.

Ramaraj,T. *et al*. (2012) Antigen-antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures. *Biochim. Biophys. Acta*, **1824**, 520–532.

Rodier,F. *et al*. (2005) Hydration of protein-protein interfaces. *Proteins*, **60**, 36–45.

Russell,R.B. and Aloy,P. (2008) Targeting and tinkering with interaction networks. *Nat. Chem. Biol*., **4**, 666–673.

Stranges,P.B. and Kuhlman,B. (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Prot. Sci*., **22**, 74–82.

Tan,P.H. *et al*. (1998) Contributions of a highly conserved VH/VL hydrogen bonding interaction to scFv folding stability and refolding efficiency. *Biophys. J*., **75**, 1473–1482.

Tsuchiya,Y. *et al*. (2006) Analyses of homo-oligomer interfaces of proteins from the complementarity of molecular surface, electrostatic potential and hydrophobicity. *Protein Eng., Des. Sel*., **19**, 421–429.

Vreven,T. *et al*. (2015) Updates to the integrated protein-protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *J. Mol. Biol*., **427**, 3031–3041.

Weitzner,B.D. *et al*. (2015) The origin of CDR H3 structural diversity. *Structure*, **23**, 302–311.

Xu,D. *et al*. (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng*., **10**, 999–1012.

Ye,J. *et al*. (2013) IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res*., **41(Web Server issue)**, W34–W40.