

## REVIEW

# Computer-aided antibody design

Daisuke Kuroda<sup>1,2,3,5</sup>, Hiroki Shirai<sup>4</sup>,  
Matthew P. Jacobson<sup>3</sup> and Haruki Nakamura<sup>1</sup>

<sup>1</sup>Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka, Japan, <sup>2</sup>Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka, Japan, <sup>3</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158, USA and <sup>4</sup>Molecular Medicine Laboratories, Astellas Pharma Co., Ltd, 21 Miyukigaoka, Tsukuba, Ibaraki, Japan

<sup>5</sup>To whom correspondence should be addressed.  
E-mail: dkuroda@protein.osaka-u.ac.jp

Received April 14, 2012; revised April 14, 2012;  
accepted April 19, 2012

Edited by Dan Tawfik

**Recent clinical trials using antibodies with low toxicity and high efficiency have raised expectations for the development of next-generation protein therapeutics. However, the process of obtaining therapeutic antibodies remains time consuming and empirical. This review summarizes recent progresses in the field of computer-aided antibody development mainly focusing on antibody modeling, which is divided essentially into two parts: (i) modeling the antigen-binding site, also called the complementarity determining regions (CDRs), and (ii) predicting the relative orientations of the variable heavy ( $V_H$ ) and light ( $V_L$ ) chains. Among the six CDR loops, the greatest challenge is predicting the conformation of CDR-H3, which is the most important in antigen recognition. Further computational methods could be used in drug development based on crystal structures or homology models, including antibody–antigen dockings and energy calculations with approximate potential functions. These methods should guide experimental studies to improve the affinities and physicochemical properties of antibodies. Finally, several successful examples of *in silico* structure-based antibody designs are reviewed. We also briefly review structure-based antigen or immunogen design, with application to rational vaccine development.**

**Keywords:** antibody design/antibody engineering/protein therapeutics/vaccine design

## Introduction

Computational methods are almost universally accepted as important tools for the invention of small molecule drugs, helping with tasks such as optimizing affinity for a target, minimizing off-target effects and optimizing pharmacokinetic properties. In these contexts, the computational methods are generally not considered substitutes for empirical testing, but rather a way to generate testable hypotheses, helping to interpret and guide experiments. The situation

with antibody therapeutics is strikingly different in this regard. While modeling has in fact contributed to the design of therapeutic antibodies, notably in early attempts at humanizing antibodies, overall the potential impact of computational methods is not as well defined, and the tools not as well developed and less broadly employed, than in small molecule drug discovery. Here we review progress in the development of computational methods that may ultimately be routinely used in antibody drug discovery.

Because we encounter a large variety of foreign molecules in daily life, ‘diversity’ is a key concept in the adaptive immune system in which antibodies take a major role. The sequences, structures and functions of antibodies have been extensively studied due to their growing importance as therapeutics (Carter, 2006; Reichert, 2008; Nelson and Reichert, 2009) and research tools (Nogi *et al.*, 2008; Hattori *et al.*, 2010). At the sequence level, antibody diversity is generated by (i) recombination of V(J)D gene segments (Tonegawa, 1983) and (ii) somatic mutations during antibody maturation, which leads to higher affinity (Besmer *et al.*, 2004; Neuberger, 2008). At a structural level, antibody diversity is manifested primarily in the antigen-binding sites, comprised of six loops, and by the relative orientations between the light chain and heavy chain variable domains ( $V_L$  and  $V_H$ ).

Our growing understanding of sequence–structure relationships in antibodies, and advances in computational protein modeling, has enabled progress toward computational methods that can assist in re-designing antibodies for higher affinity or other desired modifications (Rosenberg and Goldblum, 2006; Lippow and Tidor, 2007; Karanicolas and Kuhlman, 2009). In Fig. 1, we summarize several ways in which computational methods can be deployed in the context of antibody design. One central goal is to accurately predict the structures of antibodies from their sequences, a special case of the comparative protein modeling problem, which is valuable due to the challenges and expenses associated with experimental structure determination. In cases where the binding interaction between antibody and antigen is unknown, protein–protein docking methods can be used to predict the complex structure, although this remains challenging, especially when using homology-modeled structures for either the antibody or the antigen (Gray, 2006). Finally, using experimentally determined or predicted structures of the antibody–antigen complex, computational methods can be used to predict mutations that may improve binding affinity, specificity or other properties such as solubility.

In addition to antibody designs, designing better antigens or immunogens is also expected to elicit neutralizing antibodies (NAbs) for viruses, such as HIV and influenza. Antigens could work as vaccines only if they elicit appropriate antibodies without showing viral activity. In traditional vaccine approaches, antigens are empirically identified and

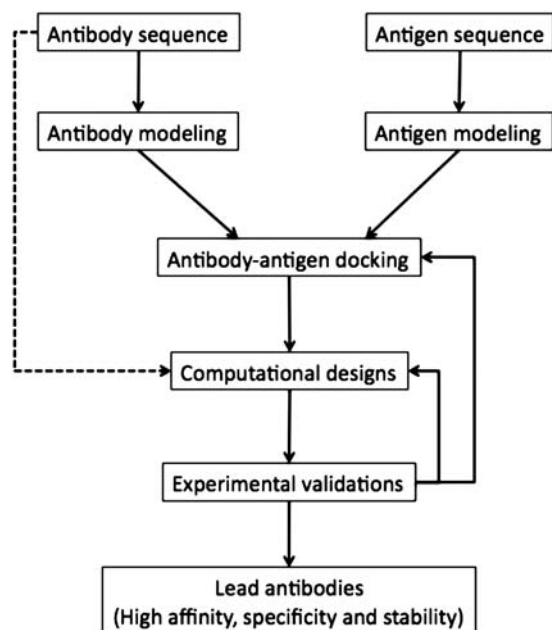


Fig. 1. Flow of computational antibody designs.

are usually isolated, inactivated or attenuated, and then injected so that they cannot cause undesirable infections, but can induce desired antibody response (Rinaudo *et al.*, 2009). Although the principle has not changed dramatically, the accumulation of experimental data and complex structures of NAb–antigens provides us with a better understanding of the structural basis of antigen recognition, and a structure-based rational vaccine design is now becoming possible.

In this review, we discuss successes and challenges of applying computational methods in each of these areas, focusing on areas that we view as particularly critical to achieving routine, successful use of computational methods in antibody design, especially complementarity determining region (CDR)-H3 modeling and predicting the effects of somatic or designed mutations during antibody optimization. Recent successes of structure-based antigen design based on protein scaffolds are also described. Other topics, such as canonical structures (Al-Lazikani *et al.*, 1997; Morea *et al.*, 2000; Vargas-Madrazo and Paz-García, 2002; Chailyan *et al.*, 2011a), especially of CDR-L3 (Kuroda *et al.*, 2009); the antibody numbering schemes (Honegger and Pluckthun, 2001; Abhinandan and Martin, 2008); B-cell epitope predictions (Ponomarenko and Bourne, 2007; Evans, 2008); humanization (Almagro and Fransson, 2008); and details of protein–protein docking (Gray, 2006) are reviewed elsewhere.

### Antibody CDR modeling

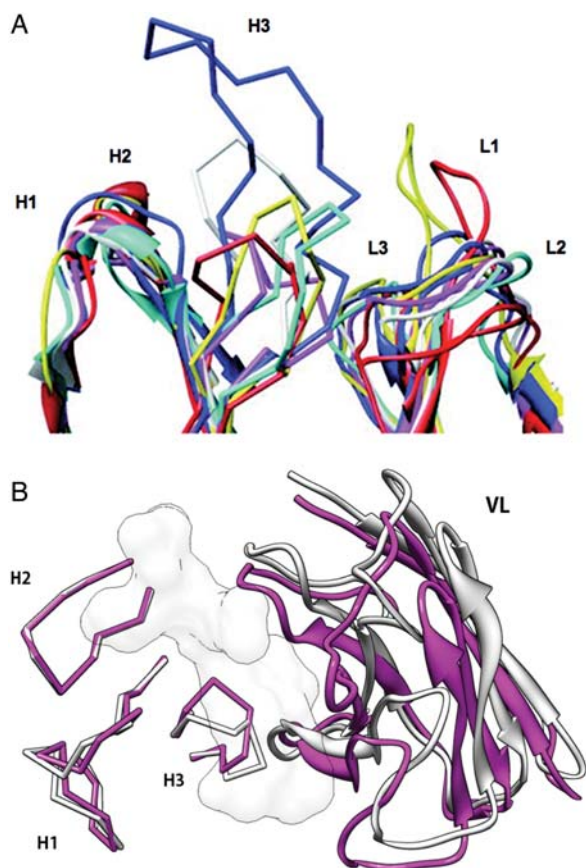
A core challenge for computational antibody engineering is predicting the structure of the antibody from sequence. This is of course a specific case of the more general protein structure prediction problem (Baker and Sali, 2001; Zhang, 2008), but antibodies present both special advantages and special challenges relative to other proteins. Before discussing methods for predicting antibody structures, we first briefly review their structural organization.

An antibody consists of two types of chains, the light and heavy chains, each of which is composed of multiple domains, all of which assume the common immunoglobulin (Ig)-fold (Murzin *et al.*, 1995; Chothia *et al.*, 1998). The antigen-binding site is located in the ‘variable’ domains of each chain,  $V_L$  and  $V_H$ . The antigen-binding sites are called CDRs and are composed primarily of six ‘hypervariable’ loops, three on the light chain (L1, L2 and L3) and three on the heavy chain (H1, H2 and H3). The remainder of the variable domains is structurally well conserved at the backbone level, and aligning antibody sequences is straightforward. For these reasons, a primary focus of antibody modeling is predicting the conformations of the CDR loops from their sequences (Whitelegg and Rees, 2004).

In the 1980s, antibody modeling started with the unexpected discovery that most of the CDR loops (all but H3) adopt a limited number of conformations called canonical structures (Chothia and Lesk, 1987; Chothia *et al.*, 1989; Tramontano *et al.*, 1989, 1990; Tramontano and Lesk, 1992). The most probable canonical structures can be predicted based on the length of the loop and the identities of key residues within or outside the CDR regions. This knowledge has been used in humanization procedures (Riechmann *et al.*, 1988) where key residues should be maintained when non-human CDRs are grafted onto human frameworks; otherwise, the conformation of the antigen-binding sites could change resulting in reduced antibody–antigen-binding affinity (Morea *et al.*, 2000; Almagro and Fransson, 2008). Canonical structures can be defined only when the same sequence motifs discovered from the Protein Data Bank (PDB) appear in the large sequence database (Kuroda *et al.*, 2009). Recent increases in the number of antibody crystal structures have made it possible to define more canonical structures (Kuroda *et al.*, 2009; Teplyakov *et al.*, 2010b; Chailyan *et al.*, 2011a). Definitions of canonical structures are available in the literatures (Al-Lazikani *et al.*, 1997; Morea *et al.*, 2000). A new classification scheme for CDRs has also been proposed based on affinity propagation clustering with a large number of crystal structures (North *et al.*, 2011).

Adopting these empirically determined sequence–structure relationships has made it possible to model the L1, L2, L3, H1 and H2 loops with high accuracy [backbone root mean square deviation (RMSD) between native structures and templates are usually  $<1.0$  Å] and subsequent work has focused mainly on predicting CDR-H3 structures (Shirai *et al.*, 1996; Morea *et al.*, 1997, 1998; Oliva *et al.*, 1998). CDR-H3 loops are located in the center of the antigen-binding site, play an important role in antigen recognitions and gain more mutations during affinity maturation than other portions of antibodies (Clark *et al.*, 2006b). They assume diverse conformations, limiting the ability to derive simple rules for their sequence–structure relationships (Fig. 2A). Structural changes in CDR-H3 also occur upon antigen binding (Fig. 2B), which makes the problem more complex. These challenges have led to many experimental and computational studies to elucidate sequence–structure–function relationships of CDR-H3 (Zemlin *et al.*, 2003; Brooks *et al.*, 2010; Julien *et al.*, 2010; Ofek *et al.*, 2010a,b).

Although their overall conformations are diverse, the conformations of the C-terminal base regions of CDR-H3 are limited and assume ‘kinked’ or ‘extended’ forms. Applying sequence–structure rules for this portion of the loop



**Fig. 2.** Structural diversity of antibodies. (A) Variability of six CDRs. (B) Structural changes of CDR-H3 [root mean square deviation (RMSD) of backbone N, C $\alpha$ , C and O atoms is 2.3 Å] and rearrangement of V<sub>L</sub>/V<sub>H</sub> domain orientation. Anti-HIV-1 peptide antibody, Fab50.1 (magenta (1GGI) and white (1GGC) for antigen bound and free structures, respectively). The white surface model is the antigen, HIV-1 V3 loop peptide. All graphics of protein structures in this article are generated using UCSF Chimera (<http://www.cgl.ucsf.edu/chimera>).

simplifies the problem by reducing the computational search space for the loop modeling. Shirai *et al.* (1996, 1999) first proposed H3-rules, which identified the base form and  $\beta$ -hairpin characteristics from the sequences. Recent advances in structural genomics have increased our knowledge of sequence–structure relationships in CDR-H3, leading to refinements of the H3-rules (Koliashnikov *et al.*, 2006; Kuroda *et al.*, 2008).

In the absence of simple sequence–structure relationships for the remainder of CDR-H3, the conformations can be predicted by exploiting the great many methods that have been developed for the more general problem of loop prediction. These methods generally work by first enumerating large numbers of plausible loop conformations, and then predicting which of these is likely to be correct by using conformational energy or a scoring function derived from the database of crystal structures (Soto *et al.*, 2008).

One approach to enumerating plausible loop conformations for CDR-H3, and other loops for which no simple sequence–structure relationships exist, is to mine the known loop structures in the PDB (Jones and Thirup, 1986). For example, Levy and coworkers surveyed the coverage of short (7-, 9- and 15-residue) fragments in the PDB against fragments derived from a novel protein, and showed that most of

them are covered with reasonable accuracy (RMSD  $\leq$  2.0 Å) (Du *et al.*, 2003). Some other studies have also suggested that existing loop conformations in the PDB are sufficient to model most loops (Fernandez-Fuentes and Fiser, 2006; Choi and Deane, 2010).

With respect to CDR-H3 specifically, Kuroda *et al.* developed a knowledge-based sampling method using fragments derived from both CDR-H3 loops in antibodies and loops in non-homologous proteins (Kuroda *et al.*, submitted for publication). They have also demonstrated that the usefulness of the database approach is enhanced by energy-based refinement. Similarly, Gray and coworkers proposed antibody-modeling software, RosettaAntibody, based on their Rosetta design suite (Rohl *et al.*, 2004; Sircar *et al.*, 2009). They modeled CDR-H3 loops using fragment assembly and the cyclic coordinate descent approach and minimized them using the Rosetta protocol (Sivasubramanian *et al.*, 2009). They also showed that antibody models constructed by their protocols could be used in subsequent computational docking procedures using RosettaDock (Gray *et al.*, 2003; Wang *et al.*, 2007), although they noted a number of challenges mainly associated with CDR-H3 modeling.

A distinct approach to enumerating possible loop conformations is the so-called ‘*ab initio*’ approach, which uses generic ‘Ramachandran’ preferences to generate possible backbone conformations, and not specific loop conformations from the PDB. In one of the earliest antibody modeling methods, Martin *et al.* combined database searching and *ab initio* loop prediction using the CONGEN program (Bruccoleri *et al.*, 1988; Martin *et al.*, 1989). A neural network machine learning method was also used to tackle the H3 modeling problem (Reczko *et al.*, 1995).

More recently, Jacobson and coworkers have developed an *ab initio* loop modeling protocol, which searches conformational space by backbone torsion-angle sampling with subsequent energy-based refinement and scoring based on the all-atom optimized potentials for liquid simulations force field and an implicit solvent model (Jacobson *et al.*, 2004; Sellers *et al.*, 2008). This method has been shown to perform well for loop modeling in generic proteins (Rossi *et al.*, 2007, 2009) and the protocol can be applied to CDR-H3 modeling with accuracies high enough that the models could be used as surrogates for experimental structures in many applications (Sellers *et al.*, 2010).

In general, loop prediction becomes more difficult as the loops become longer, especially for those longer than  $\sim$ 12 amino acids (Zhu *et al.*, 2006), which includes a significant fraction of CDR-H3 loops, which have a median length of 12 residues (Zemlin *et al.*, 2003). This is largely because the sampling problem rapidly becomes more difficult with increasing loop length. However, it should also be kept in mind that long surface loops in proteins are often intrinsically flexible, such that the observed conformation can vary depending on its interactions with antigen or due to ‘artifacts’ associated with crystal packing (Rapp and Pollack, 2005). These considerations should be born in mind when evaluating or applying loop prediction methods to CDR-H3.

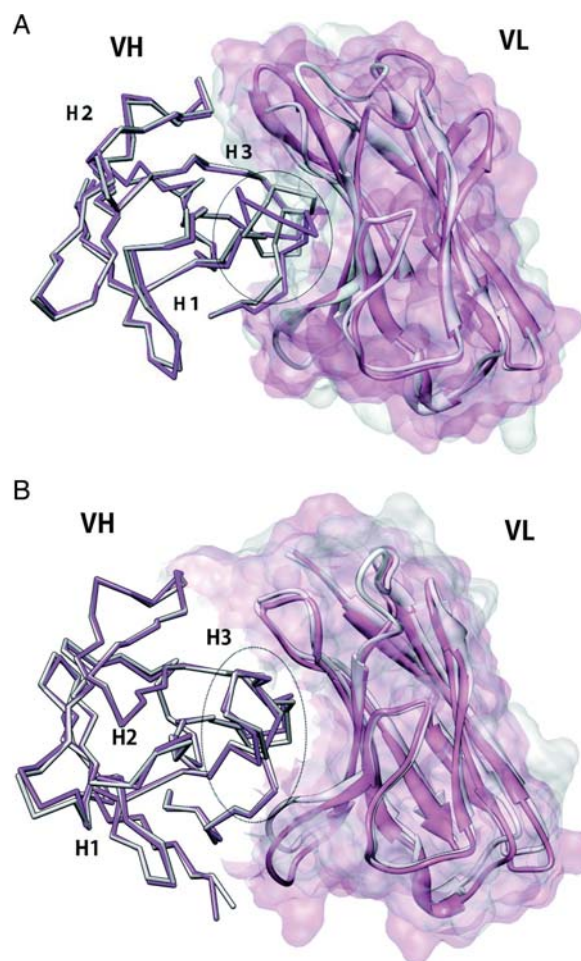
Theoretically, the most reliable approach should be to identify the loop conformations with the lowest free energy values at room temperature. For that purpose, Shirai *et al.* (1998) and Kim *et al.* (1999) performed the multicanonical molecular dynamics simulations (Nakajima *et al.*, 1997;

Higo *et al.*, 2012) to probe the free energy landscapes, and they obtained the stable structural ensembles of the CDR-H3 loops at 300 K, which were found to include the corresponding crystal structures. In addition, characterizing the free energy landscapes should aid in predicting structural changes of CDR-H3 upon antigen binding based on the ‘population shift’ principle that suggests that bound conformations should be relatively close in free energy to the unbound state (Watanabe *et al.*, 2006; Okazaki and Takada, 2008). Higo *et al.* (2011) recently observed both the population shift and the induced fit mechanisms during the coupled folding and binding in an intrinsically disordered protein. However, although conformational sampling methods have matured, the underlying energy models—generally all-atom force fields—remain imperfect, limiting the accuracy of the results. We expect that, in the future, with the further refinement of force field parameters, methods based on generating structural ensembles will be routinely used to accurately predict CDR-H3 loop conformations with and without antigens.

### Predicting $V_L/V_H$ domain orientations

Another challenge in antibody modeling is predicting the  $V_L/V_H$  domain orientations. The orientations of these two domains vary among different antibodies, and can also change upon antigen binding as seen in early work by Stanfield *et al.* (1993; Fig. 2B). When modeling antibody structures from sequence, the relative orientations of the domains will affect the structure of the antigen-binding surface. Thus, understanding sequence–structure relationships as well as the energetics of  $V_L/V_H$  domains interface is crucial not only to elucidate the structural diversity of antibodies, but also to accurately model antibody structures.

A three-layered structure was proposed for the  $V_L/V_H$  interface based on analyzing antibody crystal structures and sequences (Chothia *et al.*, 1985; Vargas-Madrado and Paz-Garcia, 2003). The key residues in these layers are also important for determining the conformations of CDR-H3 (Kuroda *et al.*, 2008). Therefore it is apparent that the accuracy of domain orientation prediction can affect the accuracy of CDR-H3 modeling and vice versa, which makes sense because CDR-H3 is located close to the interface with the  $V_L$  domain. In one striking example, a crystal structure of a ‘chimeric’ antibody, which had a different  $V_L$  domain from the original antibody, was solved, and a structural change of CDR-H3 was observed upon the  $V_L$  domain substitution (RMSD 2.88 Å) whereas the canonical structures of the parental antibodies are maintained (Fig. 3A; Pei *et al.*, 1997). In another example, however, the  $V_H$  domain of an antibody accommodated a different  $V_L$  domain from another antibody without any significant structural changes of CDR-H3 (RMSD 0.54 Å) or the other portions of the  $V_H$  domain (Fig. 3B; Teplyakov *et al.*, 2010a). These studies, in which artificial  $V_H/V_H$  domain pairs form stable complexes, shed light on the possibility of multi-specific antibodies (Bostrom *et al.*, 2009), which recognize two distinct antigens with non-overlapping functional paratopes. That is, variable domains of light and heavy chains, which recognize an antigen, respectively, independently of the other chain, can be combined, leading to a multi-specific antibody. Considering the large number of  $V_L$  and  $V_H$  sequences in nature, computer-guided approaches to explore the potential pairing of these



**Fig. 3.** Effect of  $V_L$  domain substitution on CDR-H3 conformation. (A) B1-8 (1A6V; magenta ( $V_H$ ) and pink ( $V_L$ )) and B1-8/NQ11 (1NQB; gray ( $V_H$ ) and white ( $V_L$ )). The  $V_L$  domain of B1-8 (pink) was replaced by that of another antibody, NQ11, which results in a ‘chimeric’ antibody termed B1-8/NQ11. Structural change of CDR-H3 is observed (RMSD 2.9 Å). (B) C836 (3L7E; magenta ( $V_H$ ) and pink ( $V_L$ )) and X836 (3MBX; gray ( $V_H$ ) and white ( $V_L$ )). The  $V_L$  domain of C836 (pink) was replaced by that of another antibody, which results in a ‘chimeric’ antibody termed X836. Structural change of CDR-H3 is not observed (RMSD 0.5 Å).

two domains and the domain-wise contributions to antigen binding, which would be difficult to examine by experiments, are expected.

Exploring the possibility of the prediction of correct domain orientations based on energetics, Jacobson and coworkers examined the variations of  $V_L/V_H$  domain orientations in crystal structures and how they relate to sequence (Narayanan *et al.*, 2009). In this study, the difference in domain orientation was defined as the non-CDR  $C\alpha$  RMSD of the  $V_L$  domain when structural superposition is performed using the  $V_H$  domain. They also calculated molecular mechanics energies between the two domains, showing that the native  $V_L/V_H$  orientations correspond to those of their energy minima in conformational space.

In another study, Abhinandan and Marin defined the  $V_L/V_H$  packing angles as a torsion angle at the interface using four pseudo-coordinates derived from a set of structurally conserved residues, and determined the distribution of the packing angle, which varies from  $-31.0$  to  $-60.8^\circ$  following approximately a normal distribution with a mean of

$-45.6^\circ$  (Abhinandan and Martin, 2010). The authors proposed a method to predict the  $V_L/V_H$  packing angles using neural networks trained against several interface residues in antibody crystal structures. This work also identified some potential key residues (L38, L40, L41, L44, L46, L87, H33, H42, H45, H60, H62, H91, H105), which are located in the framework regions and are extracted by the genetic algorithm, for the prediction of the packing angle and for antibody engineering (Abhinandan and Martin, 2010). Here it is noteworthy that the evaluation of  $V_L/V_H$  domain orientations only by the packing angles could neglect another aspect of the variations: translation of each domain.

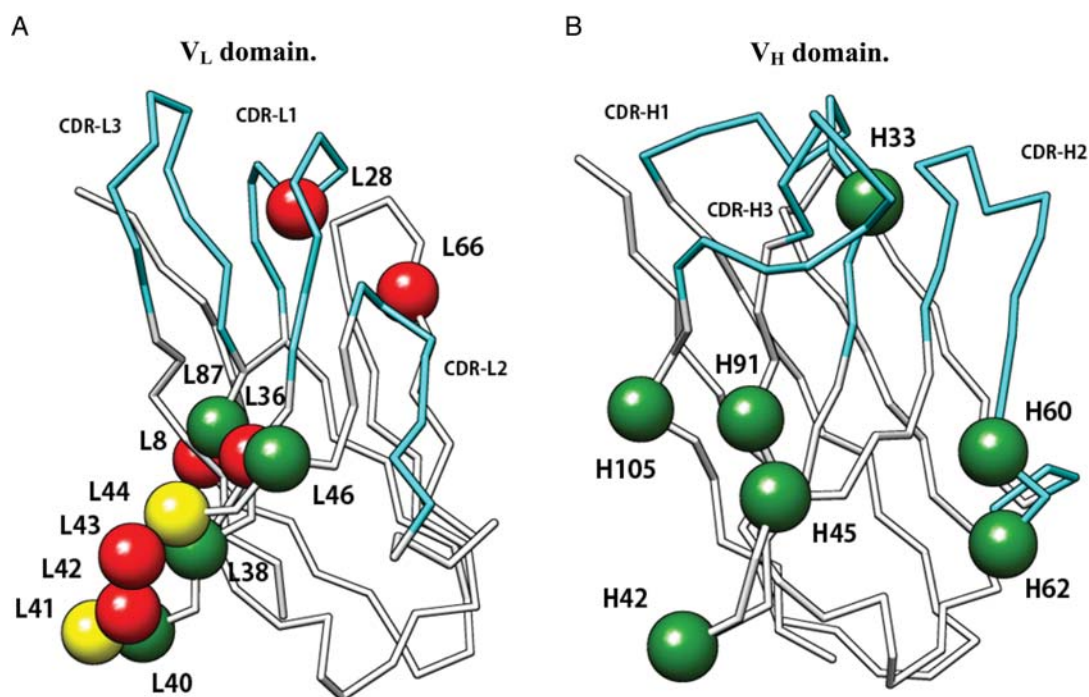
Tramontano and coworkers also analyzed the variability of  $V_L/V_H$  domain angles using cluster analysis based on a distance by LGA software (Chailyan *et al.*, 2011b). The number of antibody crystal structures with  $\kappa$  light chain in the PDB is much larger than that of antibodies with  $\lambda$  light chain, and the authors used a ‘balanced light chain set’, which included similar numbers of  $V_\kappa$  and  $V_\lambda$  antibody sequences, respectively. They found that  $V_L/V_H$  pairings can be divided into two major clusters where one cluster is only observed in mouse antibodies and the other is observed both in mouse and in human, indicating the importance of template selection in humanization procedure. They pointed out that these clusters could be assigned from the sequences by looking at some specific positions (L8, L28, L36, L41, L42, L43, L44, L66) in the sequences. Interestingly, only two of these residues, L41 and L44, are shared with the residues proposed by Abhinandan and Marin as the packing determinants. The AbNum scheme (Abhinandan and Martin, 2008) was used in both studies for antibody numbering. Structural positions of these residues are described in Fig. 4. In another finding of Tramontano and coworkers, the antibodies in one of the clusters tend to recognize smaller antigens, such as haptens,

whereas ones in the other cluster tend to show more promiscuous binding, indicating the correlation between the relative orientation and the antigen-binding capability.

The studies discussed above imply that  $V_L/V_H$  domain orientations are encoded by the sequence of each domain, but the dependence seems to be subtle. The domain orientations can also change somewhat upon antigen binding and may fluctuate in solution. Another approach might be to use the predicted domain orientation to select the best templates for antibody modeling. Also currently unexplored is the correlation between the packing angle and other properties such as solubility and stability.

Wang *et al.* (2009) performed covariation analyses using a large multiple sequence alignment of Ig-fold derived from NR-database at National Center for Biotechnology Information as well as from the PDB. Their calculations suggested the existence of conserved amino acid networks in  $V_H$  (H37, H38, H44, H45, H47, H103) and  $V_L$  (L36, L37, L43, L44, L46, L98) domains, respectively, and the networks clustered in  $V_L/V_H$  domain interface. They also identified the conserved network in  $V_H/C_H$  domain interface, but not in  $V_L/C_L$  domain interface. Protein designs based on these conserved amino acid networks showed some successes in the context of generating thermostable antibodies.

Some software for antibody modeling is available on the web. Tramontano and coworkers developed the PIGS web server, which provides various options for choosing templates for the light and heavy chains (Marcatili *et al.*, 2008). It constructs CDRs simply by grafting canonical structures and CDR-H3 loops of other antibodies onto the modeled framework. RosettaAntibody uses a  $V_L/V_H$  docking procedure based on the Rosetta energy function and rigid-body minimization to refine the orientations following CDR-H3 modeling (Sircar *et al.*, 2009; Sivasubramanian *et al.*, 2009).



**Fig. 4.** Structural positions of residues that were identified as important for  $V_L/V_H$  packing. AbNum numbering scheme was used both in Abhinandan and Martin (2010) and in Chailyan *et al.* (2011b). Consensus residues proposed in both studies are shown as yellow spheres while residues proposed only by Abhinandan and Martin (2010) and Chailyan *et al.* (2011b) are shown as green and red spheres, respectively. (A)  $V_L$  domain. (B)  $V_H$  domain.

Web Antibody Modeling (WAM), which was a pioneering software package for antibody modeling developed by Rees and coworkers, is also available on the web (Whitelegg and Rees, 2000). Recently, Almagro and coworkers performed blind tests of antibody modeling methods including PIGS, RosettaAntibody and the other two modeling strategies independently developed in Chemical Computer Group and Accelrys Inc., using nine unpublished antibody crystal structures at that time (Almagro *et al.*, 2011). They demonstrated that, overall, the results of each method showed similar trends. In each case, the conformations of the framework regions were predicted quite well (average RMSDs are  $\sim 1.0$  Å) whereas they tended to fail to predict the conformations of CDR-H3 (average RMSDs  $> 3.0$  Å). These results reconfirmed the difficulty of modeling CDR-H3, and there seems to be much room to improve the methods. More importantly, this work may lead to additional blind-prediction exercises, analogous to CASP (Critical Assessment of protein Structure Prediction) and CAPRI (Critical Assessment of Prediction of Interactions).

There are a few public databases that contain antibody data from the PDB (Allcorn and Martin, 2002; Kaas *et al.*, 2004) or from the large sequence database (Retter *et al.*, 2005; Giudicelli *et al.*, 2006). Considering the rapid increase in antibody structural and sequence data, these public databases will play important roles in organizing the information.

### Antibody–antigen recognition

Needless to say, the best way to gain atomic-level insights into antibody–antigen recognition is to determine crystal structures of the complexes by X-ray crystallography. However solving the structures of antibody–antigen complexes is frequently challenging. Computational protein–protein docking methods provide an alternative way to study antibody–antigen interactions when structures are not available. Although computational protein–protein docking methods are rapidly improving, there are still limitations, especially when modeled structures are used as starting points (Gray, 2006). On the other hand, there are a few advantages in applying protein–protein docking to antibody–antigen complexes. First, we know that antigens bind to the CDRs of antibody structures. In addition, the residues in antigens that the antibody binds to, called epitopes, can be predicted by computational methods to some extent (Evans, 2008). This information can help to more accurately predict antigen–antibody complex structures, along with any other experimental data such as from mutagenesis that can be used to restrain the protein–protein docking.

In general, antigenic epitopes in proteins are discontinuous in the protein sequences (Barlow *et al.*, 1986; Van Regenmortel, 1996). One problem in epitope prediction is that many existing methods focus on protein sequences to predict continuous epitopes. The increase in the number of antibody–antigen complex structures has enabled us to better understand the amino acid compositions of antibody–antigen interfaces and how they differ from those of other heterodimer proteins. Structure-based epitope prediction algorithms have begun to be developed based on such information (Liang *et al.*, 2009). Another problem is that current methods, including structure-based algorithms (Haste Andersen *et al.*, 2006), are not designed to predict epitopes

for specific antibodies. Recently Soga *et al.* (2010) have developed a method to predict epitope residues for individual antibodies from the sequence composition of the antibody–antigen interfaces.

In general, protein–protein interactions can be classified into permanent and transient interactions (Ozbabacan *et al.*, 2011). Antibody–antigen interactions are often considered to be transient (Cho *et al.*, 2006). Although we might expect high complementarity of antibody–antigen interfaces since antibodies often show high specificities and affinities, a previous analysis based on a small dataset of available structures showed that the shape complementarity of antibody–antigen interfaces was lower on average than that of protein–protein interfaces in general (Lawrence and Colman, 1993). From an evolutionary perspective, antibodies could evolve independently of antigens in our body, while other proteins involved in protein–protein interactions can evolve with their counterparts, which is often called coevolution, so that the shape complementarity of general proteins could be optimized against their partners. However, a later analysis with high-resolution crystal structures suggested the possibility that the previous observation that antibody–antigen complexes had lower shape complementarity was due to the low quality of the crystal structures (Cohen *et al.*, 2005). It has been also found that overall topography of antigen-binding sites is roughly classified into a limited number of groups associated with different types of antigens (MacCallum *et al.*, 1996; Lee *et al.*, 2006). For example, anti-protein antibodies tend to have flat antigen-binding sites while those of anti-peptide antibodies have a groove-like shape, in which peptides are buried. Electrostatic complementarity of antibody–antigen interfaces seemed to show a similar trend to that of protein–protein interface in general (McCoy *et al.*, 1997). Since the discussions above were based on a limited number of antibody–antigen structures, a more comprehensive analysis of the complementarities of the interfaces is required to understand the essential nature of antibody–antigen recognitions, helping to guide the rational design of antibody therapeutics.

Lysozyme has frequently been used as a model antigen to gain insights into antibody–antigen interactions (Davies and Cohen, 1996). In computational studies, Smith-Gill and coworkers investigated antibody–antigen interactions using homology-modeled anti-lysozyme antibody structures, HyHEL8 and HyHEL26, and a crystal structure, HyHEL10 (Sinha *et al.*, 2002; Mohan *et al.*, 2003). They calculated the binding energies and performed kinetics experiments, showing that a greater electrostatic contribution to antigen binding leads to higher specificity while hydrophobic interactions favor conformational flexibility and cross-reactivity. Their subsequent work based on molecular dynamics (MD) simulations with a crystal structure, HyHEL63, also supported the importance of electrostatic forces in the antibody–antigen association (Sinha *et al.*, 2007).

In a recent application of protein–protein docking to antibody–antigen complexes, Gray and coworkers used antibody models generated by WAM and antigen structures to propose the potential binding modes of antibody–antigen complexes using RosettaDock (Sivasubramanian *et al.*, 2006, 2008). The protein–protein docking used a rigid-body Monte Carlo (MC) search and a low-resolution interaction potential including an alignment score specific for antibodies in an

initial stage, and, in the subsequent refinement stage, an all-atom scoring function was used to select a solution (Gray *et al.*, 2003; Gray, 2006). In one of the examples, an EGFR-binding antibody, they also incorporated computational and experimental mutagenesis information into their docking procedures (Sivasubramanian *et al.*, 2006).

The flexibility of protein–protein interactions should be incorporated when considering the possibility of structural change upon binding (Zacharias, 2010). Structural errors in comparative models could also lead to serious errors in the subsequent steps in protein–protein docking. Methods are available for treating side chain flexibility, for instance by taking advantage of rotamer libraries, while dealing with backbone flexibility is much more difficult. Although implicit, one approach to incorporate backbone flexibility in protein–protein docking is the so-called ensemble docking, which uses several structures generated by sampling methods as starting points. MD simulations can also be used to generate such conformational ensembles. An advantage of ensemble docking is that it uses rapid rigid protein–protein docking methods, which have shown good performance in CAPRI contests so far, to identify the correct near-native models (Bonvin, 2006).

Gray and coworkers examined how ensemble docking is effective by varying the treatment of backbone flexibilities in each ensemble in protein–protein docking simulations in general (Chaudhury and Gray, 2008). They took advantage of two ensembles for docking: computational ensembles generated by RosettaRelax (Misura and Baker, 2005) and nuclear magnetic resonance (NMR) ensembles provided by NMR solution state studies. Although the backbone flexibilities were restricted only to the ligand proteins, they showed better performance using ensemble dockings than using a rigid-body protein–protein docking. Recently, for the application of modeling antibody–antigen complexes, they also developed a novel docking procedure called SnugDock, which simultaneously optimizes paratopes, i.e. CDR loops, and  $V_L/V_H$  domain orientations during antibody–antigen docking (Sircar and Gray, 2010). In a subsequent study, this method was applied to protein–protein docking in general, showing the versatility of this protocol (Sircar *et al.*, 2010). In another innovative study, Simonelli *et al.* (2010) combined experimental data with a computational protein–protein docking method. They used several antibody models created by the PIGS and RosettaAntibody programs as an ensemble, and docked them to the antigen, a surface protein of dengue virus, using RosettaDock. They took advantage of NMR chemical shift data to validate the docking results, showing that NMR epitope mapping improved the accuracy of computational docking.

Looking forward, incorporating backbone flexibility in the docking procedure is a promising and an essential approach to take into account the structural change upon binding and to overcome the small structural errors expected when using homology-modeled structures.

### Affinity maturation by somatic mutations and computational design

Antibodies can evolve in a short time in response to antigens, so that they are more specific to their antigens and have higher affinity, mainly by improving the complementarity of the

antibody–antigen interfaces (Li *et al.*, 2003; Cauerhff *et al.*, 2004; Acierno *et al.*, 2007). Somatic mutations can occur not only in the antigen-binding site, but also in the framework region, which antigens usually do not contact physically (Lavoie *et al.*, 1992). Computational methods have played an important role in understanding how these mutations improve binding affinity, and more recently, have also been used to predict novel mutations to improve affinity or specificity.

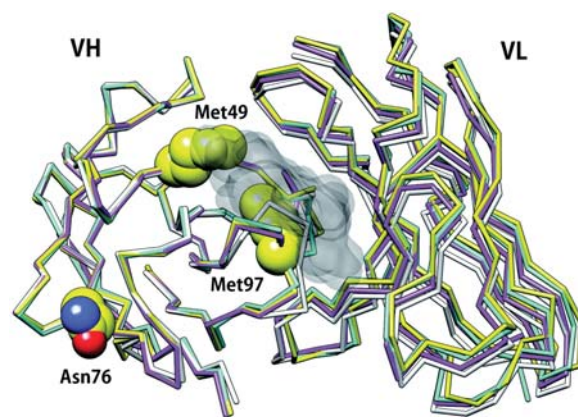
Based on the analysis of a large number of sequences, Clark *et al.* (2006) examined trends in amino acid substitutions during the somatic maturation process. Specifically, using a gene-fitting procedure with codon probability tables, they examined mutation probabilities in 23 116 heavy chains and 11 095 light chains. By comparing the probabilities with the conventional BLOSUM matrix, they concluded that amino acid substitutions in somatic mutation process have similar trends to those observed in protein evolution in general. They also showed that mutations tend to occur in antibody–antigen interfaces and in exposed surface regions, rather than the framework region. Finally, they examined changes of amino acid composition in the antibody–antigen interface, showing that the numbers of tyrosine, serine and tryptophan residues decrease while those of histidine, proline and phenylalanine increase during the maturation process.

Several computational studies have been conducted to investigate the relationship between somatic mutations and conformational flexibilities of antibodies. A plausible hypothesis is that antibodies acquire rigidity during maturation in order to increase binding affinity (by minimizing entropic losses upon binding) and to improve specificity. In fact, crystal structures of antibody drugs available in the PDB are predicted to assume rigid conformations based on their sequence features (Kuroda *et al.*, 2008). In these studies, the rigidification was hypothesized to occur within the entire variable domains rather than within only CDRs, by forming regular secondary structures, such as  $\beta$ -turns, and hydrogen bonds (Kuroda *et al.*, 2008), or by making residue contact networks, during the maturation process (Zimmermann *et al.*, 2010). Thermodynamic experiments and MD simulations demonstrated that all residues in variable domains may participate in the rigidification rather than the limited number of CDR residues (Jimenez *et al.*, 2003; Thielges *et al.*, 2008; Zimmermann *et al.*, 2010).

Kollman and coworkers (Chong *et al.*, 1999) used MD simulations to study the germline and matured anti-hapten antibodies, 48G7 (Patten *et al.*, 1996; Wedemayer *et al.*, 1997). They showed that maturation of 48G7 was driven by electrostatic optimization of the antigen-binding site, which resulted in geometric reorganization and rigidification, as manifested by smaller root mean square fluctuations of matured 48G7 compared with those of the germline antibody. Demirel and Lesk (2005) investigated the same antibodies using an elastic network model and concluded that the fluctuations of the germline antibody were similar to those of matured one. Rather, the structural difference observed in crystal structures was thought to result from the different binding mechanisms of germline and matured antibodies; induced fit could occur upon hapten binding in the germline antibody while hapten binding to the mature one was explained by a lock-and-key model. A second well-studied pair of germline and mature antibodies are those of the anti-hapten antibody, 4-4-20. Experimental results and

MD simulations again suggested that the somatic mutations, several of which are not in direct contact with the antigen, restrict the conformational fluctuations in the mature antibody (Zimmermann *et al.*, 2006; Thorpe and Brooks, 2007). More recently, Babor and Kortemme analyzed several germline and matured antibody structures in the PDB using their multi-constraint computational design protocol (Humphris and Kortemme, 2007) with the Rosetta scoring function, and confirmed that native sequences of germline CDR-H3 are designed for conformational flexibility while those of matured ones are highly optimized for single conformations (Babor and Kortemme, 2009). On the other hand, rigidification upon maturation is most likely not universal, because a matured anti-hapten antibody, SPE-7, is known to assume several different conformations in response to binding different antigens (James *et al.*, 2003; James and Tawfik, 2003). In addition, a series of studies of anti-hapten antibodies has demonstrated that rapid maturation based on CDR rigidification tend to arrive at a moderate affinity ceiling whereas flexibilities in CDR-H3 may play an important role in higher affinity maturation (Furukawa *et al.*, 1999, 2001).

To more directly determine how somatic mutations modulate CDRs flexibility in nature, Jacobson and coworker also analyzed a set of antibodies, whose crystal structures were solved both in germline and matured forms, without their antigens in the crystal structures, using MD simulations with an implicit solvent model (Wong *et al.*, 2011). Significant rigidification was observed for CDR residues that contact the antigen, mainly in CDR-H3, while the other five CDRs frequently displayed some increase in flexibility. The observed rigidification was explained by new hydrogen bonds, salt bridges and side chain packings that were improved by somatic mutations. Interestingly, conformations similar to the germline antigen free- and bound forms of 7G12 (1NGZ and 1N7M, respectively) were reproduced when using the germline free form (1NGZ) as the initial structure in the simulation, whereas the germline bound form was not sampled when the matured free form (1NGY) was used as the initial structure. This observation suggested a population shift mechanism of antigen binding in the 7G12 antibody. Consistent with his hypothesis, a structural change of CDR-H3 upon antigen binding was observed in the germline crystal structures (RMSD of CDR-H3 between the antigen bound- and free forms is 2.22 Å) while that of matured form is not observed (RMSD 0.39 Å), and CDR-H3 of the germline bound form is almost identical to those of matured antigen free- and bound forms (RMSD of CDR-H3 in the germline form against those in matured antigen free- and bound forms is 0.39 and 0.42 Å, respectively) (Fig. 5). In this antibody, the S97M mutation in the  $V_H$  domain contributed to improve hydrophobic contacts and modulated the flexibility of CDR-H3 by anchoring the loop to its antigen free form, thus shifting the equilibrium toward the conformation most suitable for binding, so that the matured form could bind the antigen without entropic penalty. Another mutation that modulates flexibility in this antibody was the mutation of Ser to Asn at the position 76 in the  $V_H$  domain. This S76N mutation anchored CDR-H1 loop via hydrogen bonds during the simulations. In the simulations of the germline form, the Ser residue at the position 76 rarely formed hydrogen bonds with the CDR-H1 whereas, in the simulations of the matured form, the Asn at the corresponding position



**Fig. 5.** Positions of somatic mutations in a metal chelataze catalytic antibody, 7G12. The mutations, R49M, S76N and S97M in  $V_H$  domain, are represented as spheres. Germline and matured antigen bound forms are represented by magenta (1N7M) and yellow (3FCT) model, respectively. Germline and matured antigen-free forms are represented by white (1NGZ) and cyan (1NGY) model, respectively. The antigen is shown as a black surface model. The residues at position 49 and 97 contact the antigen whereas the residue at position 76 is far from the antigen-binding site. Structural change of CDR-H3 is observed only in the germline form (magenta and white; RMSD 2.2 Å. See text.).

often formed multiple hydrogen bonds via the side chain with the CDR-H1, which results in the rigidification of the CDR. It is noteworthy that position 76 in the  $V_H$  domain is far from the antigen-binding site (Fig. 5). Therefore, this example shows how mutations distant from the antigen-binding site can modulate binding.

Although antibodies can evolve *in vivo* through affinity maturation, arguments have emerged that binding affinities and specificity could be further improved by using artificial systems, such as synthetic antibody libraries (Soderlind *et al.*, 2000; Borrebaeck and Ohlin, 2002; Bond *et al.*, 2005; Sidhu and Fellouse, 2006; Michnick and Sidhu, 2008; Babor *et al.*, 2011) and computational protein designs (Babor and Kortemme, 2009). Here we review some recent successes of the latter approach, in which the affinities of natural antibodies were improved to varying extents (Table I). Details of protein design methodologies are reviewed elsewhere (Rosenberg and Goldblum, 2006; Lippow and Tidor, 2007; Karanicolas and Kuhlman, 2009).

Most of the current examples of computational antibody design start from antigen-antibody complex structures in the PDB, that is, the studies can be described as 're-design' of antigen-antibody interfaces. Protein interface re-design relies on estimating free energy changes due to amino acid substitutions, which can be accomplished using physics-based force fields (Huang *et al.*, 2006) or knowledge-based potentials derived from the structural database (Ota *et al.*, 2001; Russ and Ranganathan, 2002; Clark and van Vlijmen, 2008). Similar to comparative modeling methods, re-design also requires efficient conformational search algorithms, such as dead-end elimination (DEE) (Desmet *et al.*, 1992; Dahiyat and Mayo, 1996, 1997; De Maeyer *et al.*, 2000) and MC searches (Koehl and Levitt, 1999; Kuhlman and Baker, 2000), to sample conformations of modified structures. In most studies, protein backbones have been fixed during the procedures due to computational limitations. However, it has become possible to incorporate explicit backbone flexibility



**Table I.** Results of computational designs

Initial structures <sup>a</sup>	Resolution (Å)	Antigen	$K_D^{\text{WT,b}}$	$K_D^{\text{Mut,b}}$	Affinity improvement <sup>c</sup>	Designed mutations <sup>d</sup>	Ref.
AQC2 (1MHP)	2.80	Integrin	7 nM	850 pM	10	L: S28Q, N52E, H: T50V, K64E	Clark <i>et al.</i> (2006b)
D44.1 (1MLC)	2.50	Lysozyme	4.4 nM	30 pM	140	L: N92A, H: T28D, E35S, S57V, T58D, G99D	Lippow <i>et al.</i> (2007)
TA4	N/A <sup>c</sup>	Gastrin	6 mM	13.2 nM	454	L: H91F, Q92F, R94P, V96A, H: I100L, S102V	Barderas <i>et al.</i> (2008)
E2 (3BN9)	2.17	Protease	4.8 nM	340 pM	14	H: T98R	Farady <i>et al.</i> (2009)

<sup>a</sup>Antibody structures to be designed. PDB codes are shown in parentheses.

<sup>b</sup>Reported values in the papers. For E2,  $K_I$  value is shown.

<sup>c</sup>Reported values in the papers.

<sup>d</sup>There are a number of designed mutations in each works. Only representatives are shown. L and H represent light and heavy chains, respectively.

<sup>e</sup>Homology modeled structure was used.

in computational protein design (Hu *et al.*, 2007; Murphy *et al.*, 2009).

Clark *et al.* optimized an antibody AQC2 for binding to the I domain of human integrin VLA1 using the CHARMM force field (Brooks *et al.*, 1983), the ICE software (Kangas and Tidor, 1998), which calculates electrostatic interactions and desolvation energies, and a DEE-based side chain repacking software, DEZYMER, developed by Hellinga and coworkers (Looger and Hellinga, 2001). This approach succeeded in improving the binding affinity even though the original antibody has the relatively high affinity ( $K_D \sim 7$  nM) (Clark *et al.*, 2006a). They tried >80 designed variants including multiple mutations, in which the combination of four mutations (L: S28Q, N52E, H: T50V, K64E; Fig. 6A) displayed the highest affinity of 850 pM, which was a 10-fold improvement over the wild type. The designed structure was solved and the rearrangement of a hydrogen bond network in the interface was elucidated.

In another successful antibody–antigen re-design, Lippow *et al.* (2007) focused on electrostatic interactions. One of the obvious advantages to optimize electrostatic forces is that long-range interactions could be incorporated into design process. They designed single mutations in all CDR positions using the DEE and A\* search algorithms with the CHARMM force field, ranked them by Poisson–Boltzmann electrostatics and demonstrated that four mutations (L: N92A, H: T28D, S57V, T58D; Fig. 6B) together generate 100-fold improvement of the affinity over wild type (43 pM vs. 4.4 nM for wild type).

A somewhat different use of computational modeling is to identify favorable positions for experimental random mutagenesis, rather than relying on computational prediction to identify specific favorable mutations. In such a study, Barderas *et al.* (2008) used computational modeling, including constrained antigen docking and MD simulations, to select favorable residues for random mutagenesis in CDR-H3 and -L3 by phage display. This approach resulted in mature anti-gastrin antibodies with binding affinities as potent as 13.2 nM, a 454-fold improvement over wild type based on six mutations.

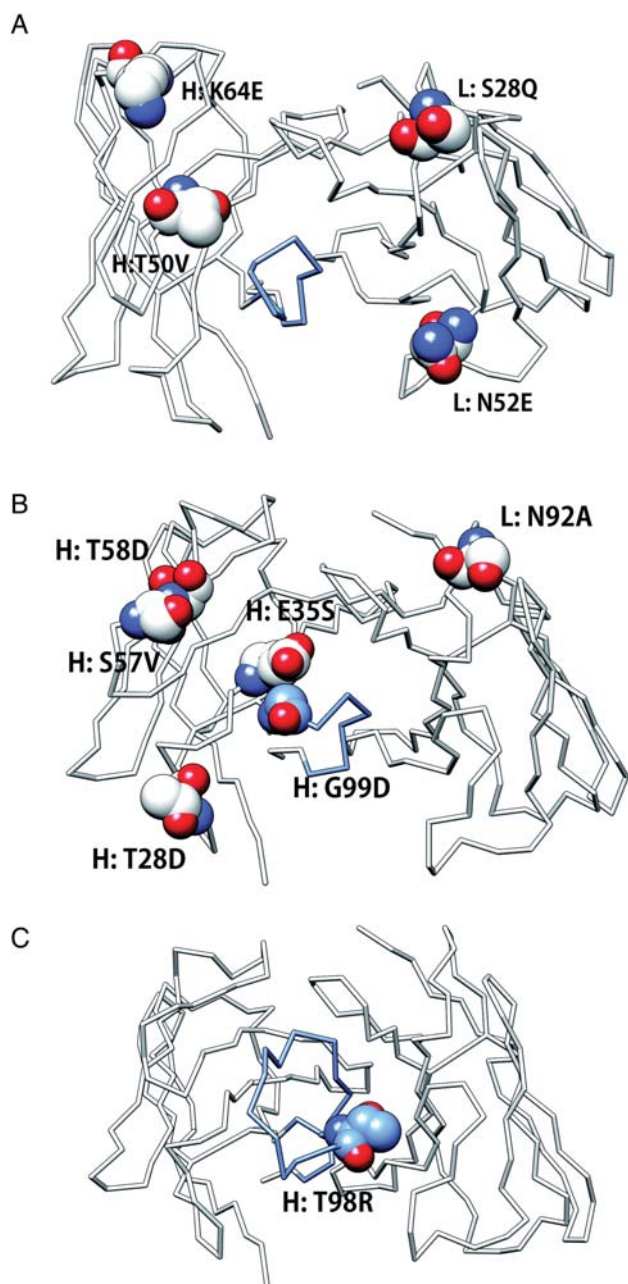
In some cases, specificity, and not affinity, is the major challenge addressed by computational design, as in the work by Farady *et al.* (2009), which aimed to improve the species cross-reactivity of an anti-protease antibody, E2. The problem is that antibodies derived from mice immunized by human antigens often display low efficacy in pre-clinical animal models because the structures and sequences of the

antigen targets are not completely conserved between humans and mice. In the case of the protease antigen, called MT-SP1 in human and epithin in mouse, they share 87% sequence identity, and three surface residues that contact the antibody differ. Based on the complex structure of MT-SP1/E2, an epithin/E2 structure was modeled. Residues in the antibody that contacted the antigen were selected for *in silico* mutations and several mutations were predicted to improve the affinity for epithin on the basis of a molecular mechanics energy function. The T98R mutation (Fig. 6C) in CDR-H3 showed the largest predicted free energy change, and subsequent experimental tests confirmed that the mutation results in a 14-fold affinity improvement over wild type ( $K_I$  reduced from 4.8 nM to 340 pM). The mutation is believed to improve epithin binding by introducing new hydrogen bonds in the antigen–antibody interface.

What can we learn from these successes? First, high-resolution crystal structures are not necessarily required for computational design. As shown in Table I, initial structures, which have moderate resolution from 2.17 to 2.80 Å, could be exploited for computational design, as well as homology models of antibodies and antigens. These successes imply that current antibody modeling methods might be sufficiently accurate to be used in computational antibody design. Second, computational protein–protein docking could be used as a first step in design procedure, particularly in cases where experimental information can be used to help to guide the docking. Third, significant improvement in binding affinity frequently requires multiple amino acid substitutions. Finally, as seen in Figs 5 and 6, it is clear that not only CDR-H3 but also the other five CDR loops can be candidates for affinity-enhancing mutations. Conformational changes that occur upon antigen binding, especially in CDR-H3, present a challenge for antibody design. In addition, current protocols mainly focus on optimizing antigen-contacting residues in CDRs. Mutations during somatic maturations, however, often occur in the framework regions, which antigens cannot directly contact (Lavoie *et al.*, 1992). Exploiting mutations of framework regions might be possible when incorporating backbone flexibility during computational designs.

### Stability design and aggregation in antibodies

Another challenging application of computational designs in antibody engineering is to predict aggregate-prone regions



**Fig. 6.** Positions of designed mutations in antibodies. CDR-H3 is colored blue. Positions of designed mutations are represented as sphere models. Original residues are shown. (A) Anti-integrin antibody, AQC2 (1MHP). (B) Anti-lysozyme antibody, D44.1 (1MLC). (C) Anti-protease antibody, E2 (3BN9).

(APRs) from amino acid sequences and to design aggregate-resistant antibodies by introducing mutations in those regions. One mechanism of aggregation is the formation of amyloid fibrils, which are rich in  $\beta$ -sheet structure, and have been associated with diseases such as Alzheimer's or Parkinson's diseases (Dobson, 1999). Experimental observations suggest that short segments of a protein could facilitate aggregation (Ventura *et al.*, 2004), and these are termed as APRs. Currently, several computational methods are available to predict APRs and rates of aggregations, mainly based on the sequence composition and on propensities such as hydrophobicity, charge and secondary structure propensity (Caffisch, 2006; Conchillo-Sole *et al.*, 2007; Trovato

*et al.*, 2007; Garbuzynskiy *et al.*, 2010). Empirical or phenomenological models have been proposed for the effect of mutations on the aggregation rate on the basis of the experimental data (Chiti *et al.*, 2003; DuBay *et al.*, 2004). MD simulations have also been used to interpret and complement experimental studies of protein aggregation and amyloidosis (Ma and Nussinov, 2006; Sharma *et al.*, 2008). A review of aggregation in protein therapeutics is available elsewhere (Agrawal *et al.*, 2011), so here we limit ourselves to recent studies on design of aggregation-resistant antibodies.

Aggregation in therapeutic proteins can lead to problems with immunogenicity (Rosenberg, 2006; Kumar *et al.*, 2011) and in addition, antibodies have been known to aggregate in high concentration of formulations for therapy and in storage (Shire *et al.*, 2004). Many experimental studies have been performed to investigate antibody stability and resistance to aggregation, primarily using single-chain Fv fragments (scFv) (Bird *et al.*, 1988; Glockshuber *et al.*, 1990; Worm and Pluckthun, 1998, 1999, 2001).

Designed mutations for aggregation-resistant antibodies may induce the loss of desired affinities. Wang *et al.* examined how APRs in antibody structures contribute to antigen recognition using TANGO (Fernandez-Escamilla *et al.*, 2004) and PAGE (Tartaglia *et al.*, 2005), which are computational tools to predict APRs in protein sequences, based on 29 antigen-antibody complexes in the PDB (Wang *et al.*, 2010). They found that the APRs most frequently appeared in CDR-H2, and less frequently in CDR-H3, although residues in the N-terminal framework region of CDR-H3 are predicted to be APRs. Based on changes of buried surface area upon antigen binding, they also showed that aromatic residues, Tyr and Trp, are favored both in antigen-binding sites and APRs, indicating that aggregation may be associated with antigen binding via these residues.

In a different, structure-based approach to predicting APRs in antibodies, Trout and coworkers proposed a novel measure called spatial aggregation propensity (SAP), which quantifies the exposure of hydrophobic residues, averaged over snapshots from a MD simulations with explicit water (Chennamsetty *et al.*, 2009a,b; Voynov *et al.*, 2009). Using this measure, they identified 14 aggregation-prone motifs in constant regions of human IgG molecules and, by comparing the amino acid sequences, they noticed that those APRs are not conserved among the other antibody classes (IgA1, IgD, IgE and IgM) (Chennamsetty *et al.*, 2009a). In these works, antibody structures were used to derive SAP values but antibody models could potentially be used when crystal structures are unavailable. Subsequent work from the same lab demonstrated that the SAP measure could be applied to antibody fragments, such as Fab or Fc, and homology-modeled structures, and that MD simulations could be performed with implicit solvent models to obtain SAP values with reasonable computational time and tolerable accuracies (Chennamsetty *et al.*, 2010).

In an experimental study, Perchiacca *et al.* (2011) performed mutational analysis of a human  $V_H$  domain antibody to investigate the molecular origin of the aggregation by comparing the biophysical properties of an aggregation-prone antibody and the aggregation-resistant counterpart. The difference between the two antibodies is only in the sequence composition of each CDR. Although the introduction of

charged mutations distant from the CDRs failed to confer aggregation-resistance property for the aggregation-prone antibody, the mutations within the first CDR loop (H31–H33) or near the CDR (H29) successfully produced an aggregation-resistant antibody. It is noteworthy that, while the results of some computational methods that were used to predict APRs in this antibody contain several false positives, the consensus APR of those prediction tools (H28–H32) is almost identical to the aggregation hot spots in this experiment (H29, H31–H33). This indicates that computational meta-approaches that integrate prediction results from several methods may help to improve prediction of APRs in antibodies.

Wang and Duan (2011) performed MD simulations with the aim of proposing a strategy for improving the thermal stability of an anti-VEGF scFv antibody. The simulations demonstrated that, at least in the scFv antibody, when the temperature is elevated, the dissociation of  $V_L/V_H$  domain interface occurred at first followed by unfolding of  $V_L$  domain and then partial unfolding of  $V_H$  domain, suggesting that stability of the  $V_L/V_H$  domain interface may govern the stability of entire antibody structure. If this is correct, then engineering of the  $V_L/V_H$  domain interface is the first choice for improving the thermal stability of the antibody. The extensive analyses of  $V_L/V_H$  domain interfaces reviewed in the previous section may provide guidance for producing more stable antibodies based on  $V_L/V_H$  interface designs *in silico*.

### Antigen design to elicit NABs

Advances in computational protein design methods, and the increasing number of protein crystal structures, may make it possible to develop vaccine candidates more rationally and efficiently. Vaccine injections elicit desired NABs for antigens or viruses targeted in therapy (Burton, 2002). HIV-1 has attracted much attention because of the lack of appropriate vaccines and because of the increasing number of patients. NABs work either by binding to the virus surface and then blocking subsequent cellular response, or by binding after virion attachment to a target cell so that they can inhibit viral entry. HIV-1 shows significant variability of the sequences and glycosylation patterns, which results in its escaping from human immune systems (Burton *et al.*, 2004). A few regions in the HIV-1 envelope spike regions are structurally conserved among diverse isolates, including the CD4-binding site on gp120, the coreceptor binding site on gp120 and the membrane proximal external region (MPER) of gp41, all of which can be targets for binding by antibodies. Thus, eliciting broadly NABs by those conserved epitopes or by their mimics is challenging, and advanced computational approaches could conceivably help tackle this issue.

Ofek *et al.* (2010a,b) used computational-guided approach to design epitope scaffolds to elicit anti-HIV-1 antibodies. They introduced a known linear epitope of MPER, which is usually recognized by 2F5 antibody, into scaffolds selected from the PDB. The existing scaffolds were re-designed by RosettaDesign suit to introduce mutations so that the targeted epitope is accommodated in the scaffolds, leading to stable proteins that consist of the HIV-1 epitope and the scaffolds derived from unrelated proteins. These designs succeeded in eliciting some antibodies with similar binding modes of 2F5 with the peptide epitope but with shorter CDR-H3 loops

than 2F5. Despite this apparent success, the resulting antibodies did not neutralize the virus in an experimental test probably due to the lack of a long CDR-H3 loop with the appropriate hydrophobic nature. However, crystal structures demonstrated that the elicited antibodies induced the desired conformational change of the HIV epitope, which is encouraging for computational immunogen design based on epitope scaffolds already existing in the PDB. They also suggested a correlation between epitope flexibility and immunogenicity based on the structures obtained.

In a related study, Lapelosa *et al.* (2009, 2010) analyzed the MPER epitope with human rhinovirus (HRV) scaffolds using replica exchange molecular dynamics (REMD) simulation techniques with an implicit solvent model. Their intention was to identify the conformational preferences of the epitope, which are considered to be recognized by 2F5 antibody, and to identify the virus scaffold that maximizes the epitope exposure to solvent. They proposed a structural model of the induced antibody and HRV complex. This model was then used for further design of virus constructs. Using a focused library, the anchor of each end of the 2F5 epitope in the modeled HRV sequence was optimized, and the binding affinity of the optimized models was evaluated experimentally. Subsequently, the authors used REMD simulations of the optimized models to show a correlation between the binding affinity and the conformational preferences of the epitope, showing that conformations that closely mimicked ones in crystal structures displayed high affinity, confirming the validity of their proposed computational model and the strategy used.

Another NAB that recognizes the MPER of gp41 is the 4E10 antibody. Correia *et al.* (2010b) designed an epitope scaffold based on the binding mode of this antibody, again by using Rosetta. They searched for appropriate backbone scaffolds from the PDB, and designed the amino acid sequences with the goal of designing scaffolds that are soluble and stable with the epitope peptide in solution. Their subsequent experimental validation showed that the designed proteins consisting of the MPER epitope and the scaffolds bind to 4E10 with higher affinities (from pM to nM levels) compared with only the MPER peptide epitope itself (nM level). Then, they determined the crystal structures, demonstrating the similarity between the inserted epitope in the scaffold and the epitope recognized by 4E10 in the antibody–peptide crystal structures. Based on these designed chimeric proteins, they also proposed two distinct computational technologies: flexible backbone remodeling and resurfacing (Correia *et al.*, 2010a). These methods were used to modify the structure of the designed proteins with the goal of the minimizing immunogenicity while retaining the thermal stability, solubility and high affinity toward cognate antibody obtained in the earlier study (Correia *et al.*, 2010b). By using the backbone remodeling method, a globular domain, which assumes  $\alpha$ – $\beta$ – $\alpha$  topology, of the designed protein scaffold was trimmed, then the sequence of the remaining scaffold was designed by the resurfacing method or side chain replacement, resulting in a chimeric protein having a better melting temperature and a similar affinity to the parent protein. They also determined the crystal structure of the re-designed proteins, showing close agreement between the structure and the model. However they also pointed out that the loop region that was replaced from a

domain in the parent protein in the remodeling stage had a slightly different backbone conformation from the crystal structure, which results in packing errors in the region, demonstrating that even a small displacement of a backbone conformation could lead to an inaccurate prediction.

In a related study, Oomen *et al.* (2003) designed an immunogen by synthesizing cyclic peptides, which reduce the peptide flexibility and mimic the  $\beta$ -turn conformation in the peptide–antibody complex structure, as an alternative to using an epitope scaffold to elicit an antibody. They assessed conformational preferences and the stability of four designed peptides using MD simulations. The most stable and the least flexible peptide evaluated by the simulations actually induced a cross-reactive potent antibody in an experiment.

In the studies discussed here, the common strategy for antigen designs is to produce epitope mimics of the NABs identified from crystal structures, and a challenge is maximizing the exposure of the mimics, which is supposed to be recognized by an antibody, to solvent in a certain protein environment. Advanced computational techniques have been utilized in this context. Currently, a huge number of protein structures are available in public, and the number is still growing due to the structural genomics efforts. Hence searching for a protein scaffold from the structural database seems to be a promising approach for rational vaccine designs. *De novo* scaffolds for protein design were also proposed (MacDonald *et al.*, 2010). In addition, MD simulations are required to analyze conformational preferences of epitopes in solution, ideally in conjunction with experiments such as NMR.

Looking forward, we envision that computational methods may be particularly useful to study antibody–peptide complexes, that is, the recognition of sequential epitopes. The peptides often assume  $\beta$ -turn conformations in an antigen-binding site, and the recognition mechanism seems to depend on the length of CDR-H3 (Bates *et al.*, 1998). Considering their smaller size compared with protein antigens, the evaluation of free energy of binding for designed peptides by computer simulations is a promising approach for rational vaccine design. Furthermore, information concerning the behavior of mutant viral antigens may be quite useful for rational vaccine design since pandemic could occur because of antigenic drifts. Structures and simulations of antibody–antigen complexes provide essential information to analyze such behavior.

### Perspectives and concluding remarks

Recent advances in computational technologies combined with dramatic advances in high-throughput technology (Reddy and Georgiou, 2011) have the potential to make important contributions to the development of biological therapeutics, including those based on antibodies and antigens. A central challenge is accurately predicting antibody structures from their sequences, and significant progress has been made toward this goal. The prediction of antibody–antigen-binding modes, by computational protein–protein docking, has also made great progress, especially by exploiting knowledge of the antigen-binding site and experimental data. Accurate prediction of epitopes and paratopes should further improve accuracy. Finally, we have reviewed several examples of computational affinity maturation (Clark *et al.*, 2006a;

Lippow *et al.*, 2007; Farady *et al.*, 2009), all of which have taken advantage of crystal structures of antibody–antigen complexes in the PDB except for Barderas *et al.* (2008).

However, there are still no striking examples where all of these computational capabilities have been combined, i.e. both an antibody and an antigen are modeled from their sequences, their binding mode is predicted by protein–protein docking, and then the resultant model is used to predict mutations that improve binding affinity and/or specificity. In such an ambitious undertaking, relatively small errors in each step could be amplified in subsequent steps. However, we anticipate continuing improvements in all of these computational approaches; for example, the recent emergence of methods for incorporating backbone flexibility in protein docking and protein engineering will be potentially useful in computational antibody design. Treating loop flexibility is particularly important in antibody design because large conformational changes can occur upon antigen binding, and affinity maturation can modulate both the conformations and dynamics of the CDR loops.

Beyond improvements in conformational sampling methods, continuing improvements in ‘scoring functions’ used to estimate free energies are needed. Approximations in the scoring function can lead to deleterious effects, especially when using implicit solvent models, and quantitative estimations of free energy changes upon antigen binding are still very difficult, even in the case of small ligands. Most current force fields for protein design include many heuristic terms. Although qualitative estimations might be sufficient for some applications, rigorous free energy evaluations explicitly including the contributions of conformational entropy and solvent polarization are challenging issues in computational chemistry and biology. Knowledge-based potentials will likely also continue to improve as the number of protein structures and sequences continues to increase exponentially.

In terms of improvements of the properties in antibodies, such as immunogenicity and solubility, sequence-based methods have shown promise in antibody engineering (Abhinandan and Martin, 2007; David *et al.*, 2010; Thullier *et al.*, 2010), in addition to structure-based methods. The sequence-based methods can take advantage of the rapid advances in sequencing technology.

Structure-based vaccine design approaches will also be enhanced with increasing knowledge of antibody–antigen structures as well as improved computational methods. Scaffold-based vaccine design is a promising approach. Although high specificity is desired for an antibody drug, the importance of poly-reactivity of antibodies as a vaccine candidate has been pointed out (Dimitrov *et al.*, 2011) because such antibodies might accommodate high variability of single antigens. Thus, understanding the origins of high specificity and cross-reactivity of antibodies is important in terms of developing protein therapeutics, and further structural analyses of antibody–antigen complex structures should provide hints to elucidate such mechanisms.

Overall, although there has been much progress in understanding sequence–structure relationships in antibodies, there is still much to learn, especially regarding the mechanisms of antibody affinity maturation, such as the relationship between somatic mutations and conformational heterogeneity. Other types of protein scaffolds have been explored as an alternative to antibodies, and computational methods

are likely to play an important role in these undertakings. Another area that requires greater attention is the role of water molecules in antibody–antigen interfaces (Yokota *et al.*, 2003). Water molecules can improve the complementarity of the interface as seen in the case of affinity maturation by somatic mutations. Explicit treatment of water molecules in the design process will require novel theories or technologies, and some such efforts are underway (Jiang *et al.*, 2005; Schymkowitz *et al.*, 2005). Finally, it will be challenging to simultaneously optimize binding affinity, specificity, stability, solubility and other properties needed for a successful therapeutic protein.

## Acknowledgements

D.K. was supported by a Research Fellowship for Young Scientists and an Excellent Young Researcher Overseas Visit Program of the Japan Society for the Promotion of Science. M.P.J. thanks the Sandler Program in Basic Sciences for supporting work in his lab on antibody modeling, and NIH GM81710 for supporting work on homology modeling. M.P.J. is a consultant to Schrodinger LLC.

## Conflict of interest

We have read the journal's policy and have the following conflicts: M.P.J. is a member of the scientific advisory board of Schrodinger LLC.

## Funding

Funding to pay the Open Access publication charges for this article was provided by Japan Biological Informatics Consortium.

## References

- Abhinandan,K.R. and Martin,A.C. (2007) *J. Mol. Biol.*, **369**, 852–862.
- Abhinandan,K.R. and Martin,A.C. (2008) *Mol. Immunol.*, **45**, 3832–3839.
- Abhinandan,K.R. and Martin,A.C. (2010) *Protein Eng. Des. Sel.*, **23**, 689–697.
- Acierno,J.P., Braden,B.C., Klinke,S., Goldbaum,F.A. and Cauerhff,A. (2007) *J. Mol. Biol.*, **374**, 130–146.
- Agrawal,N.J., Kumar,S., Wang,X., Helk,B., Singh,S.K. and Trout,B.L. (2011) *J. Pharm. Sci.*, **100**, 5081–5095.
- Al-Lazikani,B., Lesk,A.M. and Chothia,C. (1997) *J. Mol. Biol.*, **273**, 927–948.
- Allcorn,L.C. and Martin,A.C. (2002) *Bioinformatics*, **18**, 175–181.
- Almagro,J.C., Beavers,M.P., Hernandez-Guzman,F., *et al.* (2011) *Proteins*, **79**, 3050–3066.
- Almagro,J.C. and Fransson,J. (2008) *Front. Biosci.*, **13**, 1619–1633.
- Babor,M. and Kortemme,T. (2009) *Proteins*, **75**, 846–858.
- Babor,M., Mandell,D.J. and Kortemme,T. (2011) *Protein Sci.*, **20**, 1082–1089.
- Baker,D. and Sali,A. (2001) *Science*, **294**, 93–96.
- Barderas,R., Desmet,J., Timmerman,P., Meloen,R. and Casal,J.I. (2008) *Proc. Natl Acad. Sci. USA*, **105**, 9029–9034.
- Barlow,D.J., Edwards,M.S. and Thornton,J.M. (1986) *Nature*, **322**, 747–748.
- Bates,P.A., Dokurno,P., Freemont,P.S. and Sternberg,M.J. (1998) *J. Mol. Biol.*, **284**, 549–555.
- Besmer,E., Gourzi,P. and Papavasiliou,F.N. (2004) *Curr. Opin. Immunol.*, **16**, 241–245.
- Bird,R.E., Hardman,K.D., Jacobson,J.W., *et al.* (1988) *Science*, **242**, 423–426.
- Bond,C.J., Wiesmann,C., Marsters,J.C., Jr and Sidhu,S.S. (2005) *J. Mol. Biol.*, **348**, 699–709.
- Bonvin,A.M. (2006) *Curr. Opin. Struct. Biol.*, **16**, 194–200.
- Borrebaeck,C.A. and Ohlin,M. (2002) *Nat. Biotechnol.*, **20**, 1189–1190.
- Bostrom,J., Yu,S.F., Kan,D., *et al.* (2009) *Science*, **323**, 1610–1614.
- Brooks,B.R., Brucoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, **4**, 187–217.
- Brooks,C.L., Blackler,R.J., Sixta,G., *et al.* (2010) *Glycobiology*, **20**, 138–147.
- Bruccoleri,R.E., Haber,E. and Novotny,J. (1988) *Nature*, **335**, 564–568.
- Burton,D.R. (2002) *Nat. Rev. Immunol.*, **2**, 706–713.
- Burton,D.R., Desrosiers,R.C., Doms,R.W., *et al.* (2004) *Nat. Immunol.*, **5**, 233–236.
- Cafisch,A. (2006) *Curr. Opin. Chem. Biol.*, **10**, 437–444.
- Carter,P.J. (2006) *Nat. Rev. Immunol.*, **6**, 343–357.
- Cauerhff,A., Goldbaum,F.A. and Braden,B.C. (2004) *Proc. Natl Acad. Sci. USA*, **101**, 3539–3544.
- Chailyan,A., Marcatili,P., Cirillo,D. and Tramontano,A. (2011a) *Proteins*, **79**, 1513–1524.
- Chailyan,A., Marcatili,P. and Tramontano,A. (2011b) *FEBS J.*, **278**, 2858–2866.
- Chaudhury,S. and Gray,J.J. (2008) *J. Mol. Biol.*, **381**, 1068–1087.
- Chennamsetty,N., Helk,B., Voynov,V., Kayser,V. and Trout,B.L. (2009a) *J. Mol. Biol.*, **391**, 404–413.
- Chennamsetty,N., Voynov,V., Kayser,V., Helk,B. and Trout,B.L. (2009b) *Proc. Natl Acad. Sci. USA*, **106**, 11937–11942.
- Chennamsetty,N., Voynov,V., Kayser,V., Helk,B. and Trout,B.L. (2010) *J. Phys. Chem. B*, **114**, 6614–6624.
- Chiti,F., Stefani,M., Taddei,N., Ramponi,G. and Dobson,C.M. (2003) *Nature*, **424**, 805–808.
- Cho,K.I., Lee,K., Lee,K.H., Kim,D. and Lee,D. (2006) *Proteins*, **65**, 593–606.
- Choi,Y. and Deane,C.M. (2010) *Proteins*, **78**, 1431–1440.
- Chong,L.T., Duan,Y., Wang,L., Massova,I. and Kollman,P.A. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 14330–14335.
- Chothia,C., Gelfand,I. and Kister,A. (1998) *J. Mol. Biol.*, **278**, 457–479.
- Chothia,C. and Lesk,A.M. (1987) *J. Mol. Biol.*, **196**, 901–917.
- Chothia,C., Lesk,A.M., Tramontano,A., *et al.* (1989) *Nature*, **342**, 877–883.
- Chothia,C., Novotny,J., Brucoleri,R. and Karplus,M. (1985) *J. Mol. Biol.*, **186**, 651–663.
- Clark,L.A., Boriack-Sjodin,P.A., Eldredge,J., *et al.* (2006a) *Protein Sci.*, **15**, 949–960.
- Clark,L.A., Ganesan,S., Papp,S. and van Vlijmen,H.W. (2006b) *J. Immunol.*, **177**, 333–340.
- Clark,L.A. and van Vlijmen,H.W. (2008) *Proteins*, **70**, 1540–1550.
- Cohen,G.H., Silverton,E.W., Padlan,E.A., Dyda,F., Wibbenmeyer,J.A., Willson,R.C. and Davies,D.R. (2005) *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 628–633.
- Conchillo-Sole,O., de Groot,N.S., Aviles,F.X., Vendrell,J., Daura,X. and Ventura,S. (2007) *BMC Bioinformatics*, **8**, 65.
- Correia,B.E., Ban,Y.E., Friend,D.J., *et al.* (2010a) *J. Mol. Biol.*, **405**, 284–297.
- Correia,B.E., Ban,Y.E., Holmes,M.A., *et al.* (2010b) *Structure*, **18**, 1116–1126.
- Dahiyat,B.I. and Mayo,S.L. (1996) *Protein Sci.*, **5**, 895–903.
- Dahiyat,B.I. and Mayo,S.L. (1997) *Science*, **278**, 82–87.
- David,M.P., Concepcion,G.P. and Padlan,E.A. (2010) *BMC Bioinformatics*, **11**, 79.
- Davies,D.R. and Cohen,G.H. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 7–12.
- De Maeyer,M., Desmet,J. and Lasters,I. (2000) *Methods Mol. Biol.*, **143**, 265–304.
- Demirel,M.C. and Lesk,A.M. (2005) *Phys. Rev. Lett.*, **95**, 208106.
- Desmet,J., Demaeyer,M., Hazes,B. and Lasters,I. (1992) *Nature*, **356**, 539–542.
- Dimitrov,J.D., Kazatchkine,M.D., Kaveri,S.V. and Lacroix-Desmazes,S. (2011) *PLoS Pathog.*, **7**, e1002095.
- Dobson,C.M. (1999) *Trends Biochem. Sci.*, **24**, 329–332.
- Du,P., Andrec,M. and Levy,R.M. (2003) *Protein Eng.*, **16**, 407–414.
- DuBay,K.F., Pawar,A.P., Chiti,F., Zurdo,J., Dobson,C.M. and Vendruscolo,M. (2004) *J. Mol. Biol.*, **341**, 1317–1326.
- Evans,M.C. (2008) *Curr. Opin. Drug Discov. Devel.*, **11**, 233–241.
- Farady,C.J., Sellers,B.D., Jacobson,M.P. and Craik,C.S. (2009) *Bioorg. Med. Chem. Lett.*, **19**, 3744–3747.
- Fernandez-Escamilla,A.M., Rousseau,F., Schymkowitz,J. and Serrano,L. (2004) *Nat. Biotechnol.*, **22**, 1302–1306.
- Fernandez-Fuentes,N. and Fiser,A. (2006) *BMC Struct. Biol.*, **6**, 15.
- Furukawa,K., Akasako-Furukawa,A., Shirai,H., Nakamura,H. and Azuma,T. (1999) *Immunity*, **11**, 329–338.
- Furukawa,K., Shirai,H., Azuma,T. and Nakamura,H. (2001) *J. Biol. Chem.*, **276**, 27622–27628.
- Garbuzynskiy,S.O., Lobanov,M.Y. and Galzitskaya,O.V. (2010) *Bioinformatics*, **26**, 326–332.
- Giudicelli,V., Duroux,P., Ginestoux,C., Folch,G., Jabado-Michaloud,J., Chaume,D. and Lefranc,M.P. (2006) *Nucleic Acids Res.*, **34**, D781–784.

- Glockshuber,R., Malia,M., Pfiztinger,I. and Pluckthun,A. (1990) *Biochemistry*, **29**, 1362–1367.
- Gray,J.J. (2006) *Curr. Opin. Struct. Biol.*, **16**, 183–193.
- Gray,J.J., Moughon,S.E., Kortemme,T., Schueler-Furman,O., Misura,K.M., Morozov,A.V. and Baker,D. (2003) *Proteins*, **52**, 118–122.
- Haste Andersen,P., Nielsen,M. and Lund,O. (2006) *Protein Sci.*, **15**, 2558–2567.
- Hattori,T., Umetsu,M., Nakanishi,T., Togashi,T., Yokoo,N., Abe,H., Ohara,S., Adschiri,T. and Kumagai,I. (2010) *J. Biol. Chem.*, **285**, 7784–7793.
- Higo,J., Ikebe,J., Kamiya,N. and Nakamura,H. (2012) *Biophys. Rev.*, in press.
- Higo,J., Nishimura,Y. and Nakamura,H. (2011) *J. Am. Chem. Soc.*, **133**, 10448–10458.
- Honegger,A. and Pluckthun,A. (2001) *J. Mol. Biol.*, **309**, 657–670.
- Hu,X., Wang,H., Ke,H. and Kuhlman,B. (2007) *Proc. Natl Acad. Sci. USA*, **104**, 17668–17673.
- Huang,N., Kalyanaraman,C., Bernacki,K. and Jacobson,M.P. (2006) *Phys. Chem. Chem. Phys.*, **8**, 5166–5177.
- Humphris,E.L. and Kortemme,T. (2007) *PLoS Comput. Biol.*, **3**, e164.
- Jacobson,M.P., Pincus,D.L., Rapp,C.S., Day,T.J., Honig,B., Shaw,D.E. and Friesner,R.A. (2004) *Proteins*, **55**, 351–367.
- James,L.C., Roversi,P. and Tawfik,D.S. (2003) *Science*, **299**, 1362–1367.
- James,L.C. and Tawfik,D.S. (2003) *Protein Sci.*, **12**, 2183–2193.
- Jiang,L., Kuhlman,B., Kortemme,T. and Baker,D. (2005) *Proteins*, **58**, 893–904.
- Jimenez,R., Salazar,G., Baldridge,K.K. and Romesberg,F.E. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 92–97.
- Jones,T.A. and Thirup,S. (1986) *EMBO J.*, **5**, 819–822.
- Julien,J.P., Huarte,N., Maeso,R., Taneva,S.G., Cunningham,A., Nieva,J.L. and Pai,E.F. (2010) *J. Virol.*, **84**, 4136–4147.
- Kaas,Q., Ruiz,M. and Lefranc,M.P. (2004) *Nucleic Acids Res.*, **32**, D208–210.
- Kangas,E. and Tidor,B. (1998) *J. Chem. Phys.*, **109**, 7522–7545.
- Karanicolas,J. and Kuhlman,B. (2009) *Curr. Opin. Struct. Biol.*, **19**, 458–463.
- Kim,S.T., Shirai,H., Nakajima,N., Higo,J. and Nakamura,H. (1999) *Proteins*, **37**, 683–696.
- Koehl,P. and Levitt,M. (1999) *J. Mol. Biol.*, **293**, 1161–1181.
- Koliasnikov,O.V., Kiral,M.O., Grigorenko,V.G. and Egorov,A.M. (2006) *J. Bioinform. Comput. Biol.*, **4**, 415–424.
- Kuhlman,B. and Baker,D. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 10383–10388.
- Kumar,S., Singh,S.K., Wang,X., Rup,B. and Gill,D. (2011) *Pharm. Res.*, **28**, 949–961.
- Kuroda,D., Shirai,H., Kobori,M. and Nakamura,H. (2008) *Proteins*, **73**, 608–620.
- Kuroda,D., Shirai,H., Kobori,M. and Nakamura,H. (2009) *Proteins*, **75**, 139–146.
- Lapelosa,M., Arnold,G.F., Gallicchio,E., Arnold,E. and Levy,R.M. (2010) *J. Mol. Biol.*, **397**, 752–766.
- Lapelosa,M., Gallicchio,E., Arnold,G.F., Arnold,E. and Levy,R.M. (2009) *J. Mol. Biol.*, **385**, 675–691.
- Lavoie,T.B., Drohan,W.N. and Smith-Gill,S.J. (1992) *J. Immunol.*, **148**, 503–513.
- Lawrence,M.C. and Colman,P.M. (1993) *J. Mol. Biol.*, **234**, 946–950.
- Lee,M., Lloyd,P., Zhang,X., Schallhorn,J.M., Sugimoto,K., Leach,A.G., Sapiro,G. and Houk,K.N. (2006) *J. Org. Chem.*, **71**, 5082–5092.
- Li,Y., Li,H., Yang,F., Smith-Gill,S.J. and Mariuzza,R.A. (2003) *Nat. Struct. Biol.*, **10**, 482–488.
- Liang,S., Zheng,D., Zhang,C. and Zacharias,M. (2009) *BMC Bioinformatics*, **10**, 302.
- Lippow,S.M. and Tidor,B. (2007) *Curr. Opin. Biotechnol.*, **18**, 305–311.
- Lippow,S.M., Witttrup,K.D. and Tidor,B. (2007) *Nat. Biotechnol.*, **25**, 1171–1176.
- Looger,L.L. and Hellinga,H.W. (2001) *J. Mol. Biol.*, **307**, 429–445.
- Ma,B.Y. and Nussinov,R. (2006) *Curr. Opin. Chem. Biol.*, **10**, 445–452.
- MacCallum,R.M., Martin,A.C. and Thornton,J.M. (1996) *J. Mol. Biol.*, **262**, 732–745.
- MacDonald,J.T., Maksimiak,K., Sadowski,M.I. and Taylor,W.R. (2010) *Proteins*, **78**, 1311–1325.
- Marcatili,P., Rosi,A. and Tramontano,A. (2008) *Bioinformatics*, **24**, 1953–1954.
- Martin,A.C., Cheetham,J.C. and Rees,A.R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9268–9272.
- McCoy,A.J., Chandana Epa,V. and Colman,P.M. (1997) *J. Mol. Biol.*, **268**, 570–584.
- Michnick,S.W. and Sidhu,S.S. (2008) *Nat. Chem. Biol.*, **4**, 326–329.
- Misura,K.M. and Baker,D. (2005) *Proteins*, **59**, 15–29.
- Mohan,S., Sinha,N. and Smith-Gill,S.J. (2003) *Biophys. J.*, **85**, 3221–3236.
- Morea,V., Lesk,A.M. and Tramontano,A. (2000) *Methods*, **20**, 267–279.
- Morea,V., Tramontano,A., Rustici,M., Chothia,C. and Lesk,A.M. (1997) *Biophys. Chem.*, **68**, 9–16.
- Morea,V., Tramontano,A., Rustici,M., Chothia,C. and Lesk,A.M. (1998) *J. Mol. Biol.*, **275**, 269–294.
- Murphy,P.M., Bolduc,J.M., Gallaher,J.L., Stoddard,B.L. and Baker,D. (2009) *Proc. Natl Acad. Sci. USA*, **106**, 9215–9220.
- Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) *J. Mol. Biol.*, **247**, 536–540.
- Nakajima,N., Nakamura,H. and Kidera,A. (1997) *J. Phys. Chem.*, **101**, 817–824.
- Narayanan,A., Sellers,B.D. and Jacobson,M.P. (2009) *J. Mol. Biol.*, **388**, 941–953.
- Nelson,A.L. and Reichert,J.M. (2009) *Nat. Biotechnol.*, **27**, 331–337.
- Neuberger,M.S. (2008) *Immunol. Cell Biol.*, **86**, 124–132.
- Nogi,T., Sangawa,T., Tabata,S., Nagae,M., Tamura-Kawakami,K., Beppu,A., Hattori,M., Yasui,N. and Takagi,J. (2008) *Protein Sci.*, **17**, 2120–2126.
- North,B., Lehmann,A. and Dunbrack,R.L., Jr (2011) *J. Mol. Biol.*, **406**, 228–256.
- Ofek,G., Guenaga,F.J., Schief,W.R., Skinner,J., Baker,D., Wyatt,R. and Kwong,P.D. (2010a) *Proc. Natl Acad. Sci. USA*, **107**, 17880–17887.
- Ofek,G., McKee,K., Yang,Y., et al. (2010b) *J. Virol.*, **84**, 2955–2962.
- Okazaki,K. and Takada,S. (2008) *Proc. Natl Acad. Sci. USA*, **105**, 11182–11187.
- Oliva,B., Bates,P.A., Querol,E., Aviles,F.X. and Sternberg,M.J. (1998) *J. Mol. Biol.*, **279**, 1193–1210.
- Oomen,C.J., Hoogerhout,P., Bonvin,A.M., Kuipers,B., Brugghe,H., Timmermans,H., Haseley,S.R., van Alphen,L. and Gros,P. (2003) *J. Mol. Biol.*, **328**, 1083–1089.
- Ota,M., Isogai,Y. and Nishikawa,K. (2001) *Protein Eng.*, **14**, 557–564.
- Ozbabacan,S.E., Engin,H.B., Gursoy,A. and Keskin,O. (2011) *Protein Eng. Des. Sel.*, **24**, 635–648.
- Patten,P.A., Gray,N.S., Yang,P.L., Marks,C.B., Wedemayer,G.J., Boniface,J.J., Stevens,R.C. and Schultz,P.G. (1996) *Science*, **271**, 1086–1091.
- Pei,X.Y., Holliger,P., Murzin,A.G. and Williams,R.L. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 9637–9642.
- Perchiacca,J.M., Bhattacharya,M. and Tessier,P.M. (2011) *Proteins*, **79**, 2637–2647.
- Ponomarenko,J.V. and Bourne,P.E. (2007) *BMC Struct. Biol.*, **7**, 64.
- Rapp,C.S. and Pollack,R.M. (2005) *Proteins*, **60**, 103–109.
- Reczko,M., Martin,A.C., Bohr,H. and Suhai,S. (1995) *Protein Eng.*, **8**, 389–395.
- Reddy,S.T. and Georgiou,G. (2011) *Curr. Opin. Biotechnol.*, **22**, 584–589.
- Reichert,J.M. (2008) *Curr. Pharm. Biotechnol.*, **9**, 423–430.
- Retter,I., Althaus,H.H., Munch,R. and Muller,W. (2005) *Nucleic Acids Res.*, **33**, D671–674.
- Riechmann,L., Clark,M., Waldmann,H. and Winter,G. (1988) *Nature*, **332**, 323–327.
- Rinaudo,C.D., Telford,J.L., Rappuoli,R. and Seib,K.L. (2009) *J. Clin. Invest.*, **119**, 2515–2525.
- Rohl,C.A., Strauss,C.E., Misura,K.M. and Baker,D. (2004) *Methods Enzymol.*, **383**, 66–93.
- Rosenberg,A.S. (2006) *AAPS J.*, **8**, E501–507.
- Rosenberg,M. and Goldblum,A. (2006) *Curr. Pharm. Des.*, **12**, 3973–3997.
- Rossi,K.A., Nayeem,A., Weigelt,C.A. and Krystek,S.R., Jr (2009) *J. Comput. Aided Mol. Des.*, **23**, 411–418.
- Rossi,K.A., Weigelt,C.A., Nayeem,A. and Krystek,S.R., Jr (2007) *Protein Sci.*, **16**, 1999–2012.
- Russ,W.P. and Ranganathan,R. (2002) *Curr. Opin. Struct. Biol.*, **12**, 447–452.
- Schymkowitz,J.W., Rousseau,F., Martins,I.C., Ferkinghoff-Borg,J., Stricher,F. and Serrano,L. (2005) *Proc. Natl Acad. Sci. USA*, **102**, 10147–10152.
- Sellers,B.D., Nilmeier,J.P. and Jacobson,M.P. (2010) *Proteins*, **78**, 2490–2505.
- Sellers,B.D., Zhu,K., Zhao,S., Friesner,R.A. and Jacobson,M.P. (2008) *Proteins*, **72**, 959–971.
- Sharma,S., Ding,F. and Dokholyan,N.V. (2008) *Front. Biosci.*, **13**, 4795–4807.
- Shirai,H., Kidera,A. and Nakamura,H. (1999) *FEBS Lett.*, **455**, 188–197.
- Shirai,H., Kidera,A. and Nakamura,H. (1996) *FEBS Lett.*, **399**, 1–8.
- Shirai,H., Nakajima,N., Higo,J., Kidera,A. and Nakamura,H. (1998) *J. Mol. Biol.*, **278**, 481–496.

- Shire,S.J., Shahrokh,Z. and Liu,J. (2004) *J. Pharm. Sci.*, **93**, 1390–1402.
- Sidhu,S.S. and Fellouse,F.A. (2006) *Nat. Chem. Biol.*, **2**, 682–688.
- Simonelli,L., Beltramello,M., Yudina,Z., Macagno,A., Calzolari,L. and Varani,L. (2010) *J. Mol. Biol.*, **396**, 1491–1507.
- Sinha,N., Li,Y., Lipschultz,C.A. and Smith-Gill,S.J. (2007) *Cell Biochem. Biophys.*, **47**, 361–375.
- Sinha,N., Mohan,S., Lipschultz,C.A. and Smith-Gill,S.J. (2002) *Biophys. J.*, **83**, 2946–2968.
- Sircar,A., Chaudhury,S., Kilambi,K.P., Berrondo,M. and Gray,J.J. (2010) *Proteins*, **78**, 3115–3123.
- Sircar,A. and Gray,J.J. (2010) *PLoS Comput. Biol.*, **6**, e1000644.
- Sircar,A., Kim,E.T. and Gray,J.J. (2009) *Nucleic Acids Res.*, **37**, W474–479.
- Sivasubramanian,A., Chao,G., Pressler,H.M., Wittrup,K.D. and Gray,J.J. (2006) *Structure*, **14**, 401–414.
- Sivasubramanian,A., Maynard,J.A. and Gray,J.J. (2008) *Proteins*, **70**, 218–230.
- Sivasubramanian,A., Sircar,A., Chaudhury,S. and Gray,J.J. (2009) *Proteins*, **74**, 497–514.
- Soderlind,E., Strandberg,L., Jirholt,P., *et al.* (2000) *Nat. Biotechnol.*, **18**, 852–856.
- Soga,S., Kuroda,D., Shirai,H., Kobori,M. and Hirayama,N. (2010) *Protein Eng. Des. Sel.*, **23**, 441–448.
- Soto,C.S., Fasnacht,M., Zhu,J., Forrester,L. and Honig,B. (2008) *Proteins*, **70**, 834–843.
- Stanfield,R.L., Takimoto-Kamimura,M., Rini,J.M., Profy,A.T. and Wilson,I.A. (1993) *Structure*, **1**, 83–93.
- Tartaglia,G.G., Cavalli,A., Pellarin,R. and Caffisch,A. (2005) *Protein Sci.*, **14**, 2723–2734.
- Tepljakov,A., Obmolova,G., Carton,J.M., Gao,W., Zhao,Y. and Gilliland,G.L. (2010a) *Mol. Immunol.*, **47**, 2422–2426.
- Tepljakov,A., Obmolova,G., Rogers,A. and Gilliland,G.L. (2010b) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **66**, 229–232.
- Thielges,M.C., Zimmermann,J., Yu,W., Oda,M. and Romesberg,F.E. (2008) *Biochemistry*, **47**, 7237–7247.
- Thorpe,I.F. and Brooks,C.L., III (2007) *Proc. Natl Acad. Sci. USA*, **104**, 8821–8826.
- Thullier,P., Huish,O., Pelat,T. and Martin,A.C. (2010) *J. Mol. Biol.*, **396**, 1439–1450.
- Tonegawa,S. (1983) *Nature*, **302**, 575–581.
- Tramontano,A., Chothia,C. and Lesk,A.M. (1989) *Proteins*, **6**, 382–394.
- Tramontano,A., Chothia,C. and Lesk,A.M. (1990) *J. Mol. Biol.*, **215**, 175–182.
- Tramontano,A. and Lesk,A.M. (1992) *Proteins*, **13**, 231–245.
- Trovato,A., Seno,F. and Tosatto,S.C. (2007) *Protein Eng. Des. Sel.*, **20**, 521–523.
- Van Regenmortel,M.H.V. (1996) *Methods*, **9**, 465–472.
- Vargas-Madrado,E. and Paz-Garcia,E. (2002) *Proteins*, **47**, 250–254.
- Vargas-Madrado,E. and Paz-Garcia,E. (2003) *J. Mol. Recognit.*, **16**, 113–120.
- Ventura,S., Zurdo,J., Narayanan,S., *et al.* (2004) *Proc. Natl Acad. Sci. USA*, **101**, 7258–7263.
- Voynov,V., Chennamsetty,N., Kayser,V., Helk,B. and Trout,B.L. (2009) *MAbs*, **1**, 580–582.
- Wang,C., Schueler-Furman,O., Andre,I., London,N., Fleishman,S.J., Bradley,P., Qian,B. and Baker,D. (2007) *Proteins*, **69**, 758–763.
- Wang,N., Smith,W.F., Miller,B.R., Aivazian,D., Lugovskoy,A.A., Reff,M.E., Glaser,S.M., Croner,L.J. and Demarest,S.J. (2009) *Proteins*, **76**, 99–114.
- Wang,T. and Duan,Y. (2011) *Protein Eng. Des. Sel.*, **24**, 649–657.
- Wang,X., Singh,S.K. and Kumar,S. (2010) *Pharm. Res.*, **27**, 1512–1529.
- Watanabe,Y.S., Fukunishi,H. and Nakamura,H. (2006) *BIOPHYSICS*, **2**, 1–12.
- Wedemayer,G.J., Wang,L.H., Patten,P.A., Schultz,P.G. and Stevens,R.C. (1997) *J. Mol. Biol.*, **268**, 390–400.
- Whitelegg,N. and Rees,A.R. (2004) *Methods Mol. Biol.*, **248**, 51–91.
- Whitelegg,N.R. and Rees,A.R. (2000) *Protein Eng.*, **13**, 819–824.
- Wong,S.E., Sellers,B.D. and Jacobson,M.P. (2011) *Proteins*, **79**, 821–829.
- Worn,A. and Pluckthun,A. (1998) *Biochemistry*, **37**, 13120–13127.
- Worn,A. and Pluckthun,A. (1999) *Biochemistry*, **38**, 8739–8750.
- Worn,A. and Pluckthun,A. (2001) *J. Mol. Biol.*, **305**, 989–1010.
- Yokota,A., Tsumoto,K., Shiroishi,M., Kondo,H. and Kumagai,I. (2003) *J. Biol. Chem.*, **278**, 5410–5418.
- Zacharias,M. (2010) *Curr. Opin. Struct. Biol.*, **20**, 180–186.
- Zemlin,M., Klinger,M., Link,J., Zemlin,C., Bauer,K., Engler,J.A., Schroeder,H.W., Jr and Kirkham,P.M. (2003) *J. Mol. Biol.*, **334**, 733–749.
- Zhang,Y. (2008) *Curr. Opin. Struct. Biol.*, **18**, 342–348.
- Zhu,K., Pincus,D.L., Zhao,S. and Friesner,R.A. (2006) *Proteins*, **65**, 438–452.
- Zimmermann,J., Oakman,E.L., Thorpe,I.F., Shi,X., Abbyad,P., Brooks,C.L., III, Boxer,S.G. and Romesberg,F.E. (2006) *Proc. Natl Acad. Sci. USA*, **103**, 13722–13727.
- Zimmermann,J., Romesberg,F.E., Brooks,C.L., III and Thorpe,I.F. (2010) *J. Phys. Chem. B*, **114**, 7359–7370.

