doi:10.1016/j.jmb.2004.09.051 J. Mol. Biol. (2004) 344, 781–795

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Protein–Protein Interactions: Hot Spots and Structurally Conserved Residues often Locate in Complemented Pockets that Pre-organized in the Unbound States: Implications for Docking

Xiang Li[1], Ozlem Keskin[2,3], Buyong Ma[3], Ruth Nussinov[3,4]* and Jie Liang[1]*

[1]*Department of Bioengineering University of Illinois at Chicago MC-063, Chicago, IL 60607 USA*

[2]*Center for Computational Biology and Bioinformatics and College of Engineering, Koc University, Rumelifeneri Yolu Sariyer, Istanbul 34450, Turkey*

[3]*Basic Research Program SAIC-Frederick, Inc., Laboratory of Experimental and Computational Biology National Cancer Institute Frederick, MD 21702, USA*

[4]*Sackler Inst. of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University Tel Aviv 69978, Israel*

**Corresponding authors*

Energetic hot spots account for a significant portion of the total binding free energy and correlate with structurally conserved interface residues. Here, we map experimentally determined hot spots and structurally conserved residues to investigate their geometrical organization. Unfilled pockets are pockets that remain unfilled after protein–protein complexation, while complemented pockets are pockets that disappear upon binding, representing tightly fit regions. We find that structurally conserved residues and energetic hot spots are strongly favored to be located in complemented pockets, and are disfavored in unfilled pockets. For the three available protein–protein complexes with complemented pockets where both members of the complex were alanine-scanned, 62% of all hot spots ($\Delta\Delta G > 2$ kcal/mol) are within these pockets, and 60% of the residues in the complemented pockets are hot spots. 93% of all red-hot residues ($\Delta\Delta G \geq 4$ kcal/mol) either protrude into or are located in complemented pockets. The occurrence of hot spots and conserved residues in complemented pockets highlights the role of local tight packing in protein associations, and rationalizes their energetic contribution and conservation. Complemented pockets and their corresponding protruding residues emerge among the most important geometric features in protein–protein interactions. By screening the solvent, this organization shields backbone hydrogen bonds and charge–charge interactions. Complemented pockets often pre-exist binding. For 18 protein–protein complexes with complemented pockets whose unbound structures are available, in 16 the pockets are identified to pre-exist in the unbound structures. The root-mean-squared deviations of the atoms lining the pockets between the bound and unbound states is as small as 0.9 Å, suggesting that such pockets constitute features of the populated native state that may be used in docking.

Keywords: protein–protein interactions; residue hot spots; protein docking; residue conservation; binding and folding

Introduction

Protein–protein interactions are critical for practically all biological functions, including signal transduction, metabolism, vesicle transport and mitogenic processes. To comprehend the mechanism of biological processes, the function of the proteins must be considered within the context of other interacting proteins. Considerable efforts have centered on studies of the principles governing protein–protein interactions, including interface residue contacts, morphology, hydrophobic patches, conservation, residue propensities and secondary structures.[1–5] Identification of binding sites is critical for establishing protein–protein interaction networks, cellular pathways and

Abbreviations used: hGH, human growth hormone; hGHbp, human growth hormone receptor; HEL, hen egg lysozyme; BPTI, basic pancreatic trypsin inhibitor.

E-mail address of the corresponding author: jliang@uic.edu

0022-2836/$ - see front matter

regulation. In addition, it can assist in drug design and quaternary structure prediction.

Residues at protein–protein binding sites are more conserved than on the rest of the protein surface.[4,6–12] Aided by a conservation analysis of interfacial residues, predictions of protein binding sites have recently achieved significant success.[6,12,13] On the basis of the work of Tsai *et al*. in 1996,[14] Keskin *et al*. have constructed a substantially enlarged dataset of non-redundant protein–protein interfaces.[15] Unlike other sequence-non-redundant datasets, this dataset is derived through structural comparisons of interfaces, independent of both the sequence order at the interface and the folds of the two parent protein chains.[9] Following a selection of 3799 non-redundant interface clusters,[15] we have obtained 67 clusters with 343 members and over 3600 structurally conserved residues, allowing a statistical analysis. This database is useful, since it may be explored for important features in protein–protein interactions that may be independent of sequence order or backbone folds.

In this study, we examine the geometrical features of the structurally conserved residues in the form of pockets and voids, known to abundantly populate protein interfaces.[16,17] First, we study unfilled pockets (Figure 1). These are pockets present after protein–protein association. They represent geometrical features of packing defects at the protein interface. Second, we study complemented pockets (Figure 1). These pockets are present when the two proteins are separated, but disappear following association. They represent binding regions on interfaces that have non-trivial geometric shape and have tight fitting. Details on pocket identification are described in Materials and Methods.

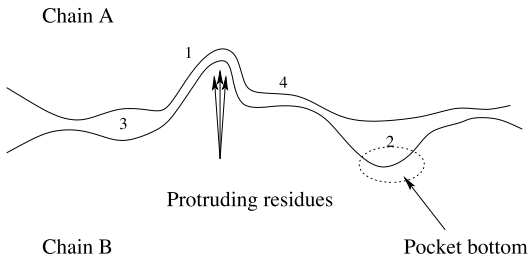

Figure 1. An illustration of a protein–protein interface between chain A and chain B. (1) A complemented pocket is a concave region on the interface part of chain A, which is filled by some residues from chain B upon protein–protein association. A complemented pocket on one protein chain is always filled by protruding residue(s) from its binding partner. (2) An unfilled pocket is a concave region on the interface part of chain A, which is not filled by chain B upon protein–protein association. (3) Unfilled pockets can be formed by shallow depressions on both sides of the interface. Two shallow depressions facing each other form an unfilled pocket. (4) Other regions of the interface, such as matched regions that may be flat or without non-trivial geometry. They make up at least half of the total size of the interface (Figure 2(b) and (d)). The sizes of different interface regions are not drawn in proportion.

Alanine scanning of interfaces has shown that a few key residues can contribute dominantly to the binding free energy of protein–protein complexes.[18–20] A residue is defined as a hot spot if there is a significant binding free energy change ($\Delta\Delta G \geq 2$ kcal/mol) when mutated to alanine (1 cal = 4.184 J). The hot spots collected by Bogan & Thorn[21,22] have been shown to overlap remarkably well with structurally conserved residues.[9,10] Thus, we study the relationship of interfacial pockets and experimental energy hot spots. *In silico*, efficient energy functions have been developed for the prediction of the experimentally measured free energy change by alanine substitution.[23,24] However, there are no general principles to address the question of what makes an interfacial residue a hot spot. With a remarkable foresight, Bogan & Thorn have proposed that some hot spots were largely surrounded by hydrophobic O-rings.[21] Nevertheless, prediction of hot spots remained a difficult task.[18,25] Hydrophobicity, shape, charge and interfacial residue type have been shown to inadequately explain or predict the energy hot spots.[5,23] To address this question, we investigate the geometrical features of the hot spots, particularly focusing on whether they locate in tightly fit interface indentations.

Our analysis indicates that whereas most interfaces have packing defects in the form of unfilled pockets, they also have tight fitting regions characterized by complemented pockets. Importantly, these regions are enriched in structurally conserved residues. For the cases where both proteins in the complex were alanine-scanned,[22] the complemented pockets and protruding residues identify 62% of all known hot spots, and 60% of the residues in complemented pockets are hot spots. We further examine the red-hot residues ($\Delta\Delta G \geq 4$ kcal/mol). We find that 93% (13/14) of the red-hot residues are found as protruding or complemented pocket residue. Our results point toward the crucial role of local tight packing in non-trivial geometrical shapes in protein–protein interactions. Evolution has optimized tight fitting through a complemented pocket hot region organization. These concave indentations may provide sites for drug discovery.

We further analyze a set of 31 protein–protein complexes compiled by Chen *et al*.,[26] which have been crystallized in the unbound state: 18 of these contain complemented pockets. In 16 of these 18, these pockets pre-exist in the unbound state, with a low root-mean-square deviation (RMSD) between the atoms lining the corresponding pockets. This observation supports the suggestion that pre-formed pockets may be a highly populated native state feature. Combined, our results point toward the mechanistic role of the hot spots in protein–protein binding and suggest a possible scheme for identification of a hot spot in an unbound protein structure. As such, they may have applications in protein docking experiments.

Results

Imperfect fit: unfilled pockets on protein–protein interfaces

Shape complementarity between interacting partners is a fundamental aspect of protein–protein interactions. Shape complementarity alone was used by Connolly in 1986[27] to address the docking problem. Achieving a good match of shapes is the focus of many protein docking methods.[28–31] To understand how well protein chains pack together and how shape matching can aid in docking, we analyze the distribution of unfilled pockets. Our result is consistent with the report by Hubbard & Argos,[16] though based on a comprehensive dataset and a different methodology.

Interfaces of native proteins rarely have perfect geometric fit. Unfilled pockets are frequently around 300 Å^2 (Figure 2(a)), accounting for 5% to 20% of the interface area (Figure 2(b)), measured by the change in solvent-accessible surface area. The implied interface areas between 6000 Å^2 and 1500 Å^2 are larger than the average interface size, since a well-defined pocket can be formed only when an interface is relatively large. The total area of unfilled pockets increases with the size of the interface ($R^2=0.81$) (Figure 2(e)). With a few exceptions, the majority of these interfaces with areas greater than 1000 Å^2 have at least one unfilled pocket. Figure 3 shows an example of an unfilled pocket on the rhinovirus coat protein (pdb code 1aym). These imperfect indentations may allow residue dynamic movements, crucial for biological function. Without explicitly including water molecules that might occupy unfilled pockets, it is challenging to obtain correctly docked protein conformations with shape complementarity alone.

Tight geometric fit: complemented pockets

A complemented pocket is a concave surface region on one protein chain that is filled by its binding partner, representing a tight fit. Although less frequent than unfilled pockets, there is a considerable number of complemented pockets

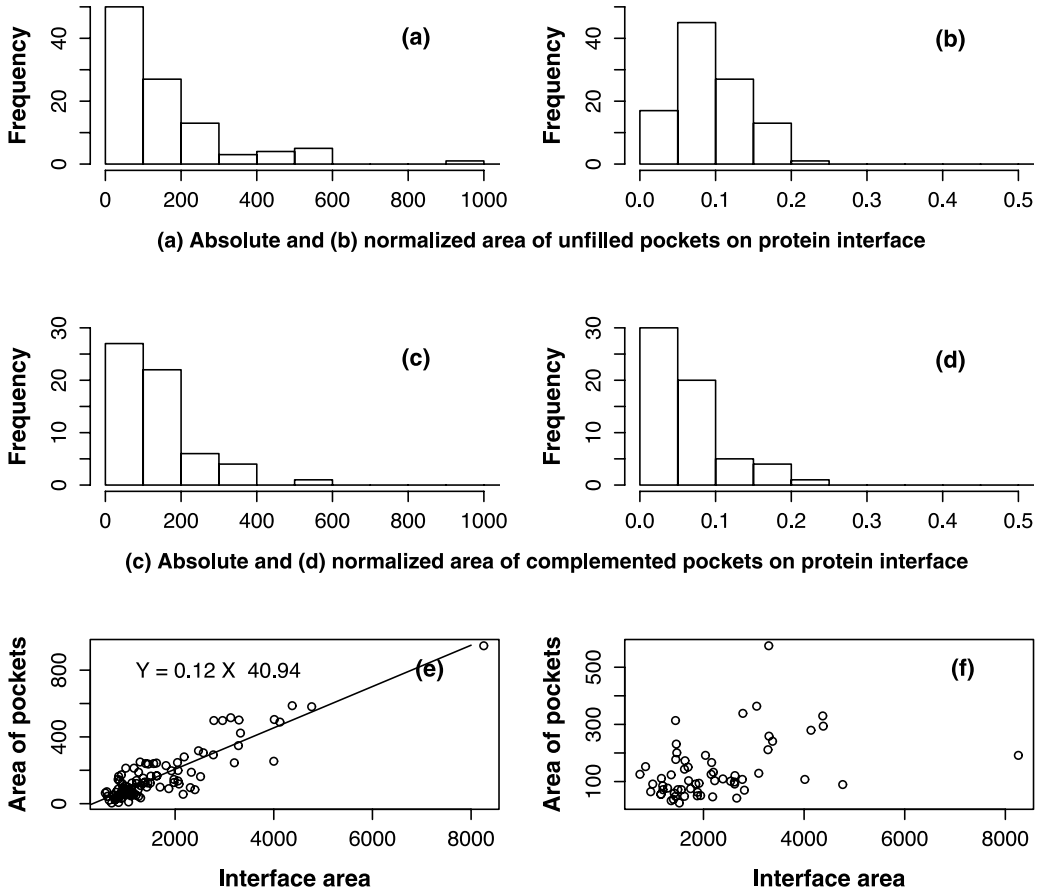

Figure 2. Size distributions of unfilled pockets and complemented pockets in 60 of the 103 structural clusters. (a) The size distributions of unfilled pockets by the absolute values of the surface areas (Å^2). (b) The size distributions of unfilled pockets by the relative values of the surface areas (Å^2). (c) The size distributions of complemented pockets by the absolute values of the surface areas normalized against the whole interfaces. (d) The size distributions of complemented pockets by the relative values of surface areas. (e) Total area of the unfilled pockets plotted against the interface area for each interface cluster. The total area of unfilled pockets increases with the size of the interface ($R^2=0.81$). (f) Total area of complemented pockets plotted against the interface area for each interface cluster. The size of complemented pockets is poorly correlated with the interface size ($R^2=0.16$). As can be seen from (b) and (d), unfilled and complemented pockets together form at most 50% of the interface. The rest of the regions of the interface are flat or shallow.

(Figure 2(c) and (d)). Among the 103 interface clusters where structurally conserved residues can be identified, 60 have complemented pockets. Among the other 43 interface clusters, many are permanent dimers (19 interfaces are formed between two enzyme homodimers, e.g. hydrolases and proteases; seven between the heavy and light chains of immunoglobulins). Twelve of the interfaces form a large pocket and the binding partner is engulfed almost entirely in this pocket. We exclude these pockets from this study.

If two complexed proteins are separated, there would be many surface pockets on each monomeric chain.[32] We characterize the unfilled and complemented pockets regardless of the partner conformation. A noticeable difference is that a complemented pocket is, on average, larger (104.93 Å^2, with a standard deviation (SD) of 72.70 Å^2). Although two small uncomplemented pockets on each chain may combine to form a large pocket (as in Figure 3), the average size of uncomplemented pockets is much smaller (39.24 Å^2, with an SD of 22.48 Å^2).

The size of complemented pockets correlates poorly with the interface size (R^2=0.16, Figure 2(f)). Complemented pockets are locally clustered and are well packed. This is in agreement with hot spots appearing in clusters within hot regions on protein interfaces (O.K. *et al.*, unpublished results). Three examples of complemented pockets on the interface between chain A and chain D of bacterial 3α,20β-hydroxysteroid dehydrogenase (pdb code 1hdc), on the interface between blood coagulation factor VIIa (VIIa) and a mutant of bovine pancreatic trypsin inhibitor 5L15 (pdb code 1fakHI),[33] and on the interface between chain A and chain D in triosephosphate isomerase (pdb code 1b9bAB) are shown in Figures 4–6.

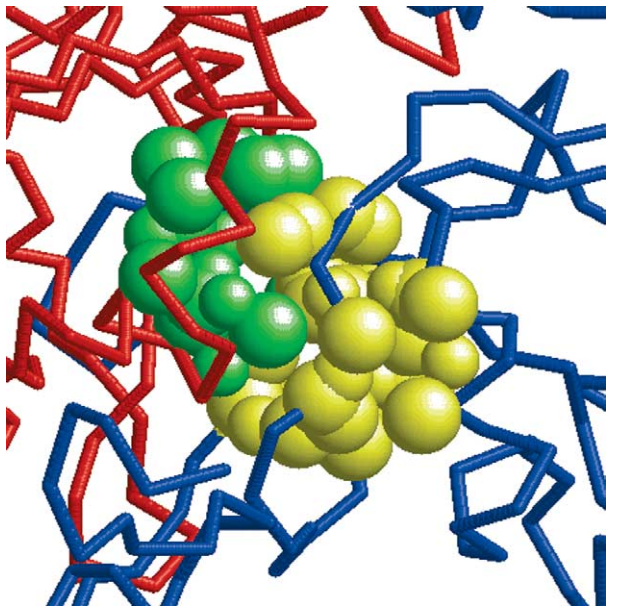

Figure 3. A large unfilled pocket on the interface between chain 1 and chain 2 of rhinovirus coat protein (1aym). This unfilled pocket has a molecular surface area of 478 Å^2, and a volume of 374 Å^3, large enough to contain 12 or 13 water molecules. The backbone of chain 2 is colored red, and the pocket wall of chain 2 is colored in green. The backbone of chain 1 is colored in blue, and the pocket wall of chain 1 is colored in yellow. This unfilled pocket is made up of 56 heavy atoms from 24 residues on both chains. They are: R107:1, R108:1, R249:1, R252:1, T208:1, T261:1, T262:1, Y256:1, Q101:1, E111:1, H260:1, M112:1, M192:1, F193:1, P250:1, P251:1, T177:2, A173:2, D175:2, H130:2, I127:2, L182:2, F185:2, Q131:2.

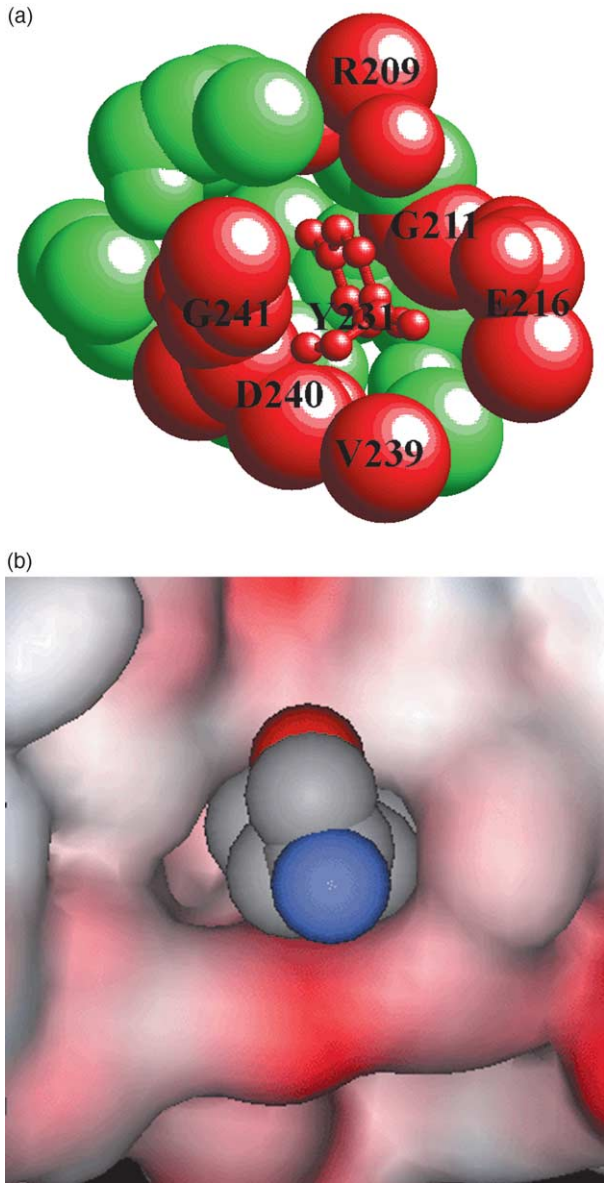

Figure 4. A complemented pocket on the interface between chain A and chain D of bacterial 3α,20β-hydroxysteroid dehydrogenase (pdb code 1hdc). (a) The protruding residue Tyr231 and the complemented pocket. This complemented pocket is composed of 13 residues or 41 heavy atoms from chain D. Eight structurally conserved residues, including seven complemented pocket residues from chain D and one protruding residue Tyr231 from chain A, are labeled and colored in red. Other residues are in green color. Structurally conserved residue Gly183 is at the bottom and not shown. (b) Geometric fit of the protruding residue Tyr231 and the complementary pocket.

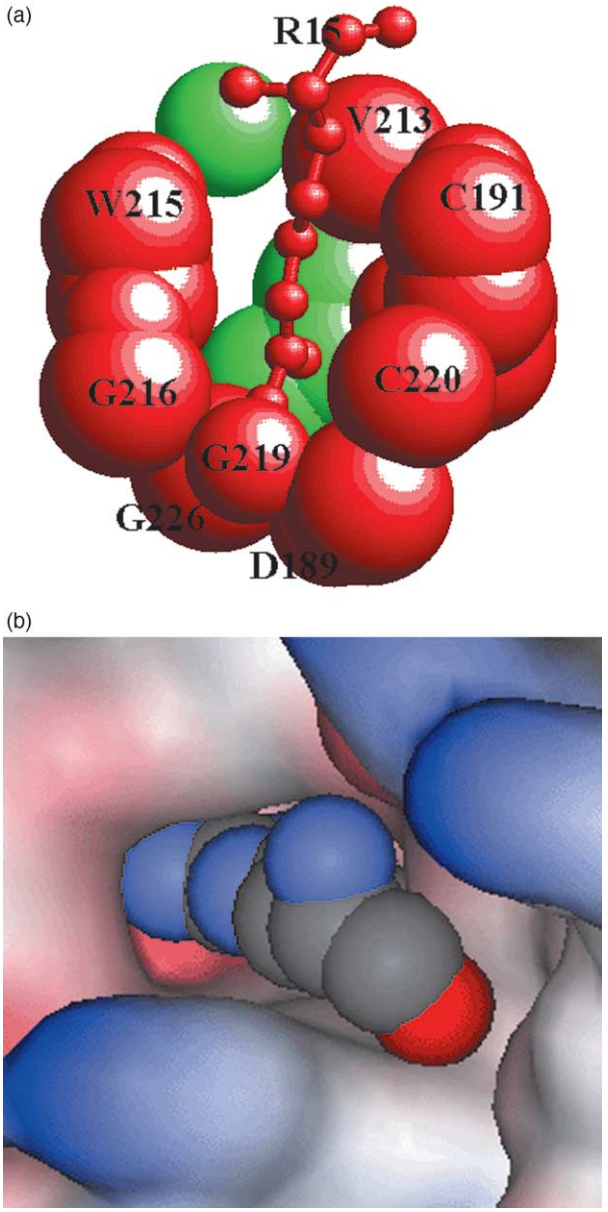

Figure 5. A complemented pocket on the interface between chain H and chain I of 1fak (complex of blood coagulation factor VIIa and mutant BPTI 5L15). (a) The protruding residue Arg15 and the complemented pocket. This complemented pocket consists of 12 residues or 29 heavy atoms from VIIa. 11 structurally conserved residues, including ten complemented pocket residues from VIIa and one protruding residue Arg15 from 5L15, are labeled and colored in red. Structurally conserved residues Ser190 and Val227 are at the bottom and not shown. (b) Geometric fit of the protruding residue Arg15 and the complementary pocket.

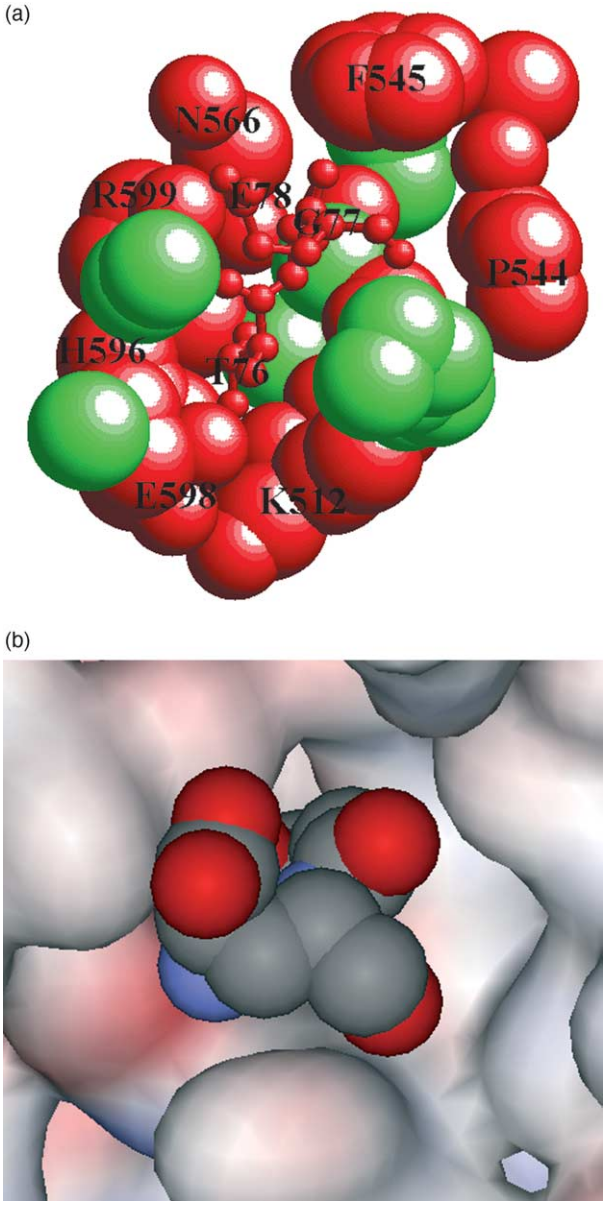

Figure 6. A complemented pocket on the interface between chain A and chain B of 1b9b (triosephosphate isomerase). (a) The protruding residues Thr76-Gly77-Glu78 and the complemented pocket. This complemented pocket consists of 16 residues or 46 heavy atoms from chain A. The 15 structurally conserved residues, including 12 complemented pocket residues from chain A and three protruding residues Thr76, Gly77, Glu78 from chain B, are labeled and colored in red. Structurally conserved residues Asn510, Pro543, Leu548, Gln565 and ILE593 are at the bottom and not shown. Unlike the pockets in Figures 4(a) and 5(a), this complemented pocket protrudes by more than one residue: Thr76-Gly77-Glu78. (b) Geometric fit of the protruding residues Thr76-Gly77-Glu78 and the complementary pocket.

Structurally conserved residues are enriched in complemented pockets but disfavored in unfilled pockets

Structurally conserved residues are often correlated with energy hot spots.[9,10] We examine the locations of structurally conserved residues in terms of unfilled and complemented pockets. Among the 103 interface clusters with a total of 627 interface members, not every cluster or member has sufficient structurally conserved residues for

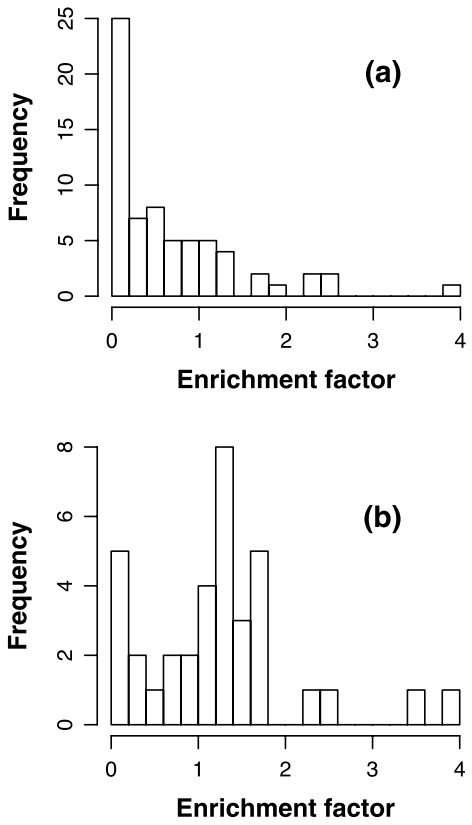

Figure 7. Distribution of structurally conserved residues on protein–protein interfaces by enrichment factor of conserved residues (a) at unfilled pockets. The majority of the conserved residues have an enrichment factor <1.0. It is unfavorable for these residues to be located at the bottom of unfilled pockets. (b) Enrichment factor of conserved residues at complemented pockets. The majority of conserved residues have an enrichment factor >1.0, and these residues are favored to be located in complemented pockets.

analysis. Thus, we take only interfaces with at least one structurally conserved residue for every 30 interfacial residues. There are 67 interface clusters and 343 interface members that satisfy this requirement.

Our results show that whereas all 67 interface clusters have unfilled pockets, for 50 of these the structurally conserved residues are disfavored to be located at the bottom of the unfilled pockets (Figure 7(a)). The overall average enrichment of structurally conserved residues is 0.65 (with an SD of 0.64), indicating that they are generally disfavored in unfilled pockets. For the majority of the 67 clusters where conserved residues are found in unfilled pockets, there are no substantial contacts between the pocket residues and their binding partners. The conservation of these residues may relate to functions other than stability.[18,24]

Our results show that structurally conserved residues are favored to be located at complemented pockets (Figure 7(b)). Among the 67 interface clusters, there are 36 clusters that contain complemented pockets. Structurally conserved residues are found to be favored in complemented pockets in 24 of the 36 clusters (Figure 7(b)). The average enrichment factor for these 24 interface clusters is 1.56 (SD of 0.74).

We calculated the fractions of conserved residues in the complemented pockets, unfilled pockets and other regions for the 36 clusters (Figure 8). We find that, on average, 30% of the residues are conserved in complemented pockets, but only 11% and 20% are conserved in unfilled pockets, and in other regions of the interface, respectively. The residue conservation in complemented pockets is statistically significantly higher than that in the unfilled pockets ($p=2.01\times10^{-5}$ for Student's one-tailed t-test), or in other regions on interfaces ($p=9.29\times10^{-3}$).

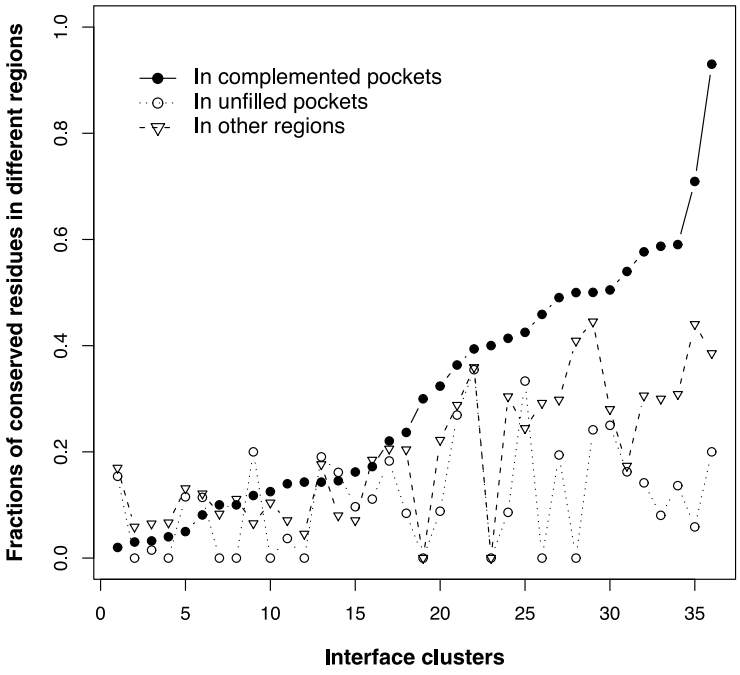

Figure 8. Fractions of structurally conserved residues in complemented pockets, unfilled pockets and other regions. Complemented pockets have a greater fraction of conserved residues than unfilled pockets ($p=2.01\times10^{-5}$ for Student's one-tailed t-test), or other regions on interfaces ($p=9.29\times10^{-3}$). The clusters are sorted in ascending order by fraction of conserved residues in complemented pockets.

We have further compared the conserved residues between the clusters with and without complemented pockets. For the set of the 60 clusters containing complemented pockets, on average, there are 2.5 conserved residues for every ten interface residues. In contrast, there are only 1.9 conserved residues for every ten interface residues for the set of 43 clusters that do not contain complemented pockets. Although this difference appears small, it is statistically significant, with a p-value of 0.001 by t-test.

Complemented pockets are less hydrophilic than other surface pockets

Pockets are distributed widely on protein surfaces.[34] Compositional difference between complemented pockets and other surface pockets may be useful for predicting complemented pockets from the unbound states. Figure 9 shows the amino acid composition of complemented pockets and of other surface pockets for the complexes in the interface dataset. Complemented pockets contain smaller fractions of hydrophilic residues than the other surface pockets. For complemented pockets, on average, 18% of the pocket residues are ionizable residues (Arg, Lys, Glu, Asp). For the other surface pockets, on average, 25.6% of the pocket residues are ionizable. This observation is similar to that obtained from a larger-scale analysis.[34]

We further investigated the hydrophilicity of complemented pockets and other surface pockets at the atomic level. We consider all oxygen and nitrogen atoms as polar. For complemented pockets, on average, 29% of the pocket atoms are polar. For other surface pockets, on average, 36% of the pocket atoms are polar. Although this difference is small, it is statistically significant with a p-value of 3.0×10^4 by t-test. This result suggests that the desolvation barrier for protruding residues to anchor into the complemented pocket is not high, since there are few polar residues, especially ionizable residues, in complemented pockets. Comparing with the other surface pockets, complemented pockets may expel water molecules more readily when binding the protruding residues from the complementary chain. Polar and ionizable residues located at the bottom of the complemented pockets are important for binding stability, enhancing polar–polar interactions (such as H-bonds and salt-bridges) in a hydrophobic environment. As shown below, these residues are preferred to be conserved, and are often associated with hot spots in alanine-scanning mutagenesis experiments.

Preference of conserved residue types in complemented pockets

Given a complemented pocket, it is useful to predict residues that are structurally conserved and thus likely to be important for stability and function. The propensity of a residue to be conserved in complemented pockets can be useful for this purpose and is plotted against the propensity to be conserved on the whole interfaces (not shown). The latter propensity values are obtained

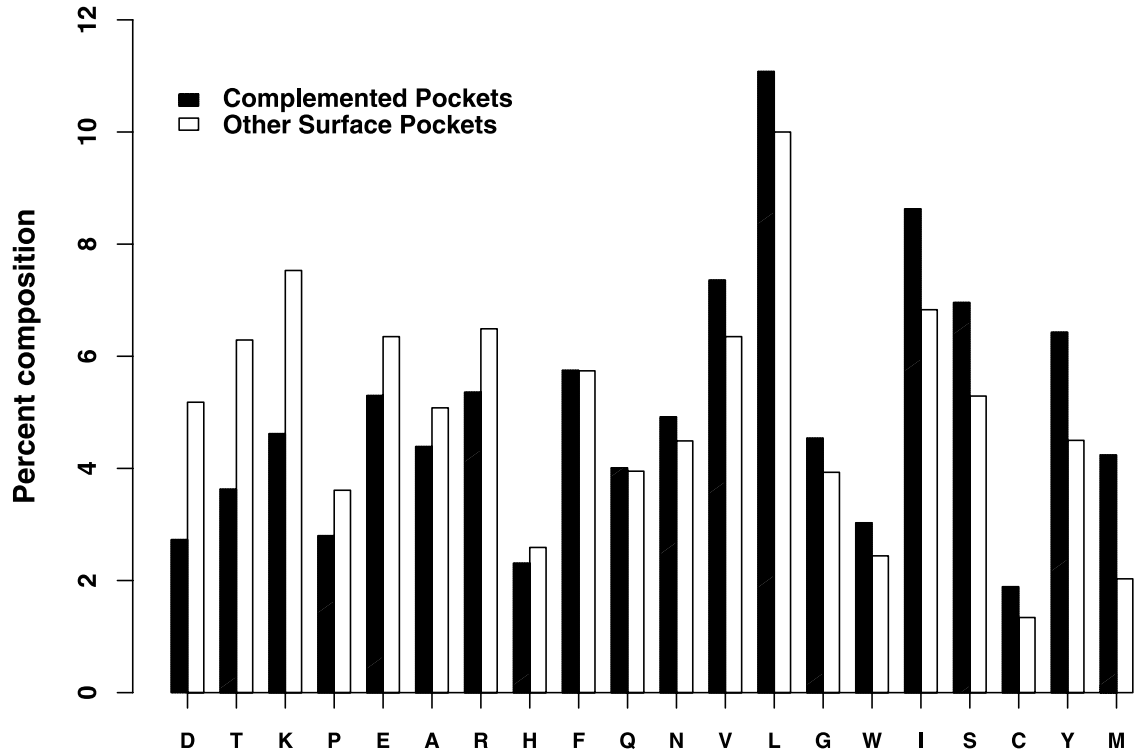

Figure 9. Fractions of different types of amino acids in complemented pockets, and in other surface pockets. A surface pocket is defined to have over 80% pocket atoms exposed to the solvent.

from Keskin *et al.*[15] For residues located in complemented pockets, Trp, Gly, Pro, Cys, Tyr and Glu are likely to be more conserved. Gly, Trp, Pro and Glu are significantly more conserved in complemented pockets *versus* the rest of the interface. Among these residues, Trp is often on the wall of the complemented pocket and may function to protect interactions inside the pocket from exposure to the solvent. Trp is a large residue with many neighbors, contributing to tight packing. Cys is more likely to be conserved as it is often involved in disulfide bonds, which provides stability to the necessary structural support of complemented pockets. Gly is far more conserved if located at a complemented pocket than if located in the rest of the interface. Gly lacks a side-chain and a mutation may reduce the size of the complemented pocket. Its mutation may prevent protruding partner residues from accessing the pocket. Gly's flexibility allows tight packing. Its conservation was observed to be coupled with aromatic, polar and small hydrophobic residues in the interacting chain, with likely backbone H-bonding across the interface.[35]

Energetic hot spots: importance of complemented pockets

Interfaces often contain a few hot spots with binding free energy change $\Delta\Delta G \geq 2$ kcal/mol when mutating to alanine.[21] Here, we examine the energetic hot spots in complemented pockets.

There are eight complexes with experimental data on mutations in both interface sides.[35] Three of them have complemented pockets and protruding residues. Two of these have been well studied in the literature. These are the complex of the human growth hormone (hGH) and human growth hormone receptor (hGHbp) (pdb code 3hhr),[20] and the barnase/barstar complex (pdb code 1brs).[19] hGH has a complemented pocket composed of Thr175 ($\Delta\Delta G = 2.0$ kcal/mol), Lys172 ($\Delta\Delta G = 2.0$ kcal/mol) and Phe176 ($\Delta\Delta G = 1.9$ kcal/mol). It is tightly packed by the protruding Trp104 on hGHbp, which is also energetically crucial for binding ($\Delta\Delta G = 4.5$ kcal/mol).[20] For barnase/barstar, there is a complemented pocket on barstar, occupied by the side-chain of a protruding His102 ($\Delta\Delta G = 6.0$ kcal/mol) from the active site of barnase.[36]

The third complex is antibody D1.3, which offers another interesting example of a complemented pocket. There are two complex structures in the database where D1.3 binds to hen egg lysozyme (HEL, pdb code 1vfb), and an anti-idiotopic antibody E5.2 (pdb code 1dvf), respectively. The interfaces of both complexes have overall similar conformations. A pocket on the surface of D1.3 is complemented by a protruding Tyr98 from E5.2 (Figure 10(a)). The $\Delta\Delta G$ of Tyr98 upon mutation to Ala is a dramatic 4.7 kcal/mol for D1.3 with E5.2 (pdb code 1dvf). Mutation of either Trp52 or Glu98 to alanine also leads to a large $\Delta\Delta G$ (4.2 kcal/mol). The same pocket is unfilled by any residues from HEL (Figure 10(b)). It was suggested that this pocket is filled with water and that solvation makes important contributions to binding.[37] For the complex of D1.3 and HEL, the $\Delta\Delta G$ values of Trp52 and Glu98 are only 1.2 kcal/mol and 1.1 kcal/mol, respectively. Trp52 and Glu98 are much less important in the unfilled pocket of D1.3-HEL.

We examined complexes where mutational data for residues are available for only one side of the interface. There are seven structures with solved

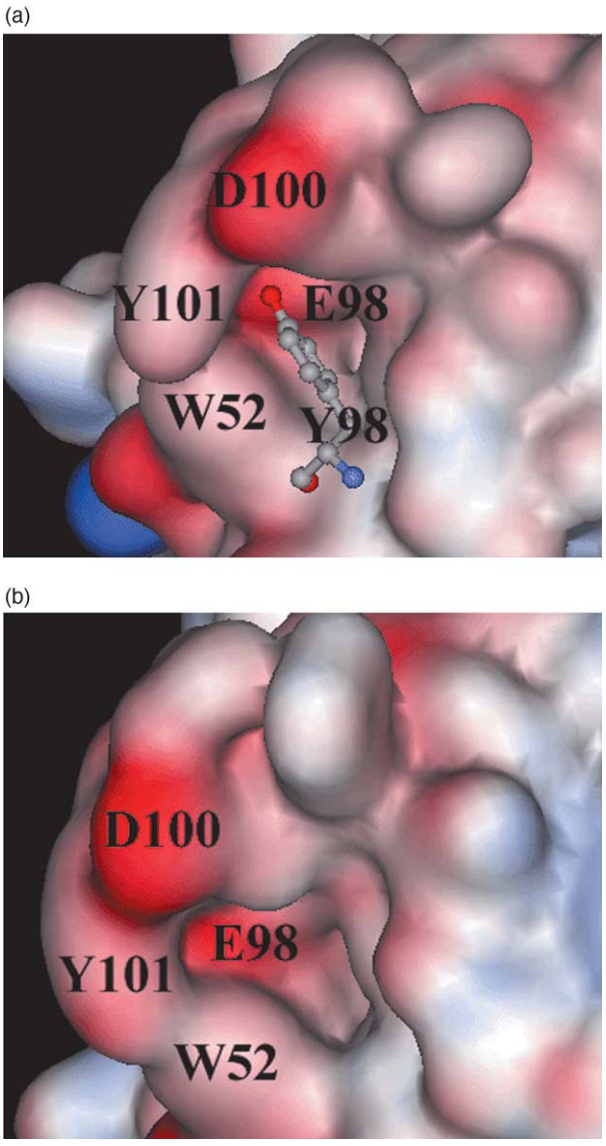

Figure 10. A pocket on the interface of antibody D1.3. (a) This pocket is occupied by Tyr98 from E5.2 (pdb code 1dvf) and forms a complemented pocket. The residues involved are energetic hot spots: $\Delta\Delta G$(Tyr98) = 4.7 kcal/mol, $\Delta\Delta G$(Trp52) = 4.2 kcal/mol, $\Delta\Delta G$(Glu98) = 4.2 kcal/mol, $\Delta\Delta G$(Tyr101) = 4.0 kcal/mol. (b) This pocket is not occupied by any residues from HEL (1vfb). Trp52 and Glu98 are no longer hot spots: $\Delta\Delta G$(Trp52) = 1.2 kcal/mol, $\Delta\Delta G$(Glu98) = 1.1 kcal/mol, $\Delta\Delta G$(Tyr101) = 4.0 kcal/mol.

structures and with hot spots reported in ASEdb.[22] In two of these, we detected complemented pockets on the interfaces. These are the basic pancreatic trypsin inhibitor (BPTI) with chymotrypsin (pdb code 1cbw), and BPTI with trypsin (pdb code 2ptc) (a volume of 256.68 $Å^3$ and 225.65 $Å^3$ for complemented pocket on chymotrypsin–BPTI and on trypsin–BPTI, respectively, Figure 11). This complemented pocket is the S1 active site of the chymotrypsin/trypsin enzymes. Lys15 at the P1 position in BPTI is the protruding residue, filling the complemented pocket with an up position in the chymotrypsin–BPTI complex (Figure 11(a)), and in the down position in trypsin–BPTI (Figure 11(b)).[38]

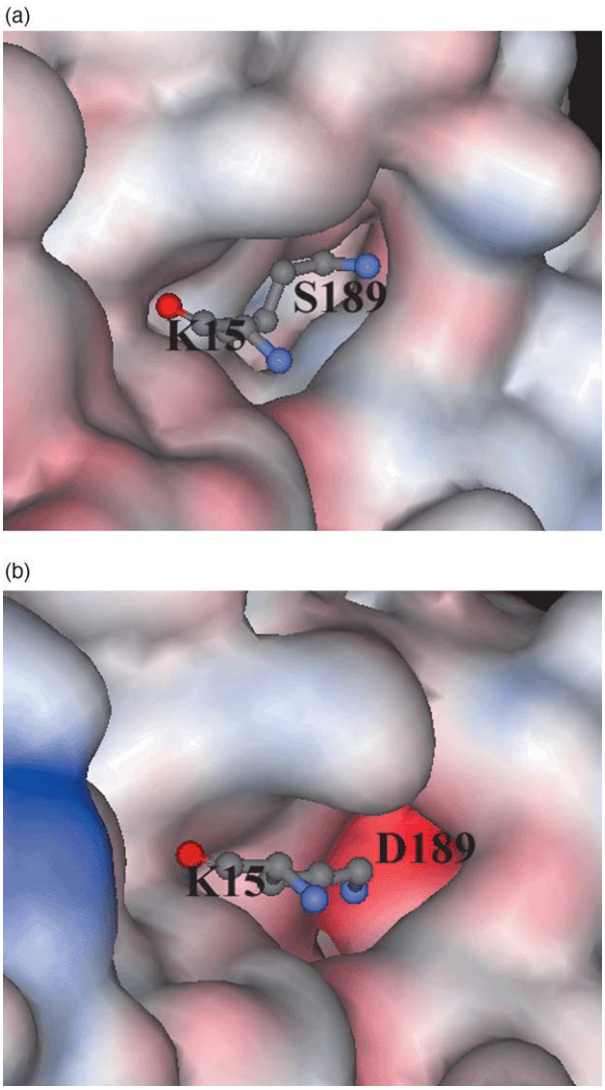

Figure 11. Two different enzymes bound to the same inhibitor BPTI. (a) The bottom of the complemented pocket is Ser189 from chymotrypsin (pdb code 1cbw), and the environment is neutral. The side-chain of Lys15 covers the pocket, instead of protruding into the interior. In this case, we have $\Delta\Delta G$(Lys15) = 2.0 kcal/mol. (b) The bottom of the complemented pocket of trypsin (pdb code 2ptc) contains Asp189, and the environment is negative. The side-chain of Lys15 reaches into the pocket deeply and forms a salt-bridge with Asp189 at the bottom. In this case, $\Delta\Delta G$(Lys15) = 10.0 kcal/mol.

There are no alanine-scanning data on chymotrypsin/trypsin. For the protruding Lys15, the $\Delta\Delta G$ values are available and are drastically different for the two complexes: 10.0 kcal/mol for the trypsin–BPTI complex; the largest free energy change reported in the ASEdb for a single mutation. The $\Delta\Delta G$ of Lys15 for the chymotrypsin–BPTI complex is only 2.0 kcal/mol. The topographic and hydrophobic differences of the S1-P1 binding site between these two complexes are discussed in great detail in the literature.[38,39] In the chymotrypsin–BPTI complex, Lys15 in the up position is bending away from the bottom residue Ser189 (Figure 11(a)). In the trypsin–BPTI complex, it is in the down position and reaches deep into the complemented S1 binding pocket, forming two salt-bridges with Asp189 at the bottom of the pocket (Figure 11(b)). Our calculation shows that Lys15 makes only 18 atomic contacts in the chymotrypsin–BPTI complex with residues at the bottom of the S1 site. On the other hand, the protruding Lys15 in the trypsin–BPTI complex forms 36 atomic contacts with the bottom residues on the S1 site. The number of contacts and the physico-chemistry of the interactions lead to very different binding free energies for the two complemented pockets.

Complemented pockets are energetically important

Among the three complexes with mutations on both interface sides, we find that 62% (16 out of 26) of all hot spots with $\Delta\Delta G \geq 2$ kcal/mol are either at complemented pockets or are protruding residues. In all, 60% (19/31) of the residues in complemented pockets are hot spots ($\Delta\Delta G \geq 2$ kcal/mol). We further examine the red-hot residues (defined as those with $\Delta\Delta G \geq 4$ kcal/mol), in the three complexes. The results show that 93% (13/14) can be identified as protruding residues or residues in complemented pockets. All five protruding residues that we have identified have $\Delta\Delta G \geq 4$ kcal/mol: Tyr98 in 1dvf, Trp104 in 3hhr, Trp169 in 3hhr, His102 in 1brs, and Asp35 in 1brs.

Do pockets pre-exist in the unbound state?

We analyze protein–protein interfaces in their complexed states. We are interested in finding out whether the complemented pockets' geometries are pre-formed features that are populated in the unbound state. To address this question, we examine the Benchmark1.0 database compiled by Chen *et al.*[26] This benchmark dataset contains a listing of proteins crystallized in their complexed and in their unbound states. Among the 31 complex structures where the unbound structures of both binding partners are known, there are 16 enzyme–inhibitor, five antibody–antigen and ten belong to other categories. 18 (58%) have complemented pockets: 13 enzyme–inhibitor, two antibody–antigen and three belong to other categories. For all of

Table 1. Pockets in the unbound state that become complemented in the bound state

Complex	Receptor/ligand	No. complemented pockets	No. pocket atoms	cRMSD[a]	Protruding residues	Anchor residues[b]
1ACB(E:I)	Chymotrypsin/Eglin C	1	54	0.44	Leu45	Leu45
1AVW(A:B)[c]	trypsin/soybean trypsin inhibitor	2	57	1.03	Tyr62; Arg63	Arg63
1BRC(E:I)	Trypsin/APPI	1	54	0.40	Arg15	Arg15
1BRS(A:D)	Barnase/barstar	2	33	1.30	His102:A; Asp35:D	N/A
1CGI(E:I)	Chymotrypsinogen/PSTI	1	46	2.41	Tyr18	Tyr18
1CHO(E:I)	α-Chymotrypsin/ovomucoid 3rd domain	1	53	0.30	Leu18	Leu18
1CSE(E:I)	Subtilisin Carlsberg/Eglin C	2	57	0.51	Pro42; Leu45	Leu45
1TGS(Z:I)	trypsinogen/PSTI	1	42	0.77	Lys18	Lys18
1UGH(E:I)	UDG/UGI	1	25	1.37	Leu272	Leu272
2KAI(AB:I)	Kallikrein A/trypsin inhibitor	1	40	0.79	Lys15	N/A
2PTC(E:I)	β-Trypsin/pancreatic trypsin inhibitor	1	35	0.80	Lys15	Lys15
2SIC(E:I)	Subtilisin BPN/subtilisin inhibitor	2	67	0.58	Met70; Met73	Met70
2SNI(E:I)	Subtilisin Novo/chymotrypsin inhibitor 2	2	66	0.65	Ile56; Met59	Met59
1FIN(A:B)	CDK2 cyclin-dependent kinase 2/cyclin	1	47	1.05	Arg122, Phe152, Gly153, Val154	N/A
1BTH(LH:P)	Thrombin mutant/PTI	1	43	0.80	Lys15	N/A
1WQ1(G:R)	RAS activating domain/RAS	2	113	0.93	Arg789; Tyr32	N/A
1BVK(DE:F)[d]	Antibody Hulys11 Fv/lysozyme	1	19	0.99	Gln121	Gln121
1MLC(AB:E)[d]	Fab fragment/lysozyme	1	12	1.01	Arg68	Arg68
Average	–	–	48	0.90	–	–

APPI, Alzheimer's amyloid beta-protein precursor; PSTI, pancreatic secretory trypsin inhibitor; PTI, pancreatic trypsin inhibitor.
[a] rmsd of atoms at complemented pockets between unbound and bound states.
[b] Interface residues with the side-chains losing the largest solvent-accessible surface area after forming the complex. Taken from Rajamani *et al.*[40]
[c] The underlined complexes are included in the dataset compiled by Keskin *et al.*[15]
[d] The complemented pocket on the antibody side is not identified by the automated alpha shape program, but is found to pre-exist at unbound structure after manual inspection. Same as 1MLC(AB:E).

the 18 complexes containing complemented pockets, these pockets already exist in the unbound states (Table 1). We further examine the atomic details of the complemented pockets prior to and following binding. The root-mean-squared deviations of atoms appearing in both the unbound and the bound states of the complemented pockets in an optimal alignment (cRMSD) are listed in column 5 of Table 1. For the majority of these complemented pockets, which contain, on average, 48 atoms, the average cRMSD values are as small as 0.9 Å. This suggests that hot spots and conserved residues enriched in complemented pockets that exist prior to binding do not experience much conformational change upon binding. That is, the environment for favorable packing interactions exists before the binding events take place. This interpretation is consistent with the results of a recent study in which Rajamani *et al.*[40] found that a subset of residues anchoring on the protein interface are preconditioned to take biased, bound state side-chain conformations. Such anchoring residues may coincide with the protruding residues identified in this study. Here, we further observe that residues in complemented pockets in the unbound state adopt conformations similar to those populated in the bound state. Both complemented pockets and protruding residues are enriched in energetic hot spots and conserved residues.

These observations suggest that binding hot spots and conserved residues are special. They have preferred conformations in the unbound states and are often located in pre-existing geometric pocket environments that will be occupied by their binding partners. The mechanism of protein–protein interactions often does not involve spatial conformational changes that abolish locally pre-existing pocket organizations.

Discussion and Conclusions

Protein–protein interfaces are "porous" and are rarely packed perfectly.[16] Unfilled and complemented pockets are often scattered on the interfaces. Most large protein–protein complexes have a significant amount of unfilled pockets. The size of unfilled pockets in the interfaces is correlated with the interface size. Experimentally, it is unclear whether unfilled pockets contain water molecules or how the dynamics of water molecules entering and escaping these pockets may affect binding stability. If all water molecules are expelled, formation of large "water-empty" pockets on the protein–protein interface is likely to incur a significant desolvation penalty. If water molecules are contained in an unfilled pocket, polar interactions, including electrostatic and H-bonds between the proteins in the interface interior, are likely to be weakened significantly, and the desolvation penalty for residues surrounding the unfilled pocket is likely to be insignificant. In either case, unfilled

pockets are likely to reduce the binding stability gained from an increase in the interface size. On the other hand, these lesser packed areas might be crucial for binding site flexibility. In contrast, interfacial residues at complemented pockets have extensive contacts and are well-packed with the binding partner.

Structurally conserved residues in interfaces are correlated strongly with the experimental hot spots.[10] Here, we show that there is an enrichment in conserved residues and hot spots in complemented pockets; and, conversely, most residues in such pockets are conserved structurally. These include both the receiving residues in pockets and the protruding residues on the binding partner. For complemented pockets abundant with enriched conserved residues, the corresponding protruding residues are also frequently conserved. Residues across the protein–protein interface often co-evolve.[41–44]

Not every complemented pocket contains conserved residues. Some pockets without conserved residues are located at the periphery of protein–protein interfaces, and are easily accessible to solvent molecules (e.g. L-lactate dehydrogenase, pdb code 1ez4AC; data not shown). These are unlikely to be energetically critical.[21] There are examples where protruding residues do not appear to be structurally conserved. In some cases, this is due to poor alignment, as in the case of the kallikren/hirustasin complex (pdb code 1hia, Figure 12). In some other cases, protruding residues are located at flexible loop regions and cannot be aligned structurally at all. Some protruding residues become conserved if we use a reduced alphabet set where residues of similar properties are grouped, for example, A=(Trp, Tyr), B=(Arg, Lys); and C=(Asp, Glu). Finally, since MultiProt, the multiple structure alignment software used here simultaneously aligns multiple structures, there may be occasional inaccuracies in the multiple superpositions.

Hot spots and conserved residues form clusters, rather than being scattered over the entire interface. These densely packed local hot regions form a network of conserved interactions (O.K. *et al.*, unpublished results). It has been proposed that hot spots are surrounded by O-rings.[21] Our observations here suggest that these may, at least partly, consist of complemented pockets, with snugly fitting bottoms, walls and rims. In such geometries, water may be easily expelled, strengthening the backbone H-bonds and charge interactions.

For reversible protein associations, the binding free energy increases with the interface size up to some threshold.[45–47] Further, the energy is not distributed evenly at the protein–protein binding interface.[21] These observations agree well with the lack of correlation between the area of the complemented pockets and the interface size, and with the clustering of hot spots in complemented pockets. The monotonic increase in unfilled pockets with the size of the interface explains why for large, three-state folding proteins there may be an upper threshold in the binding free energy: when the interface is large, packing is more difficult to optimize.[10] The hot spot clusters further offer an explanation as to why the number of hot spots may increase with the interface size, yet the binding free energy may reach a limit.[10] Within the tightly packed cluster, hot spots are in contact, forming a network of interactions. Hence, within a cluster,

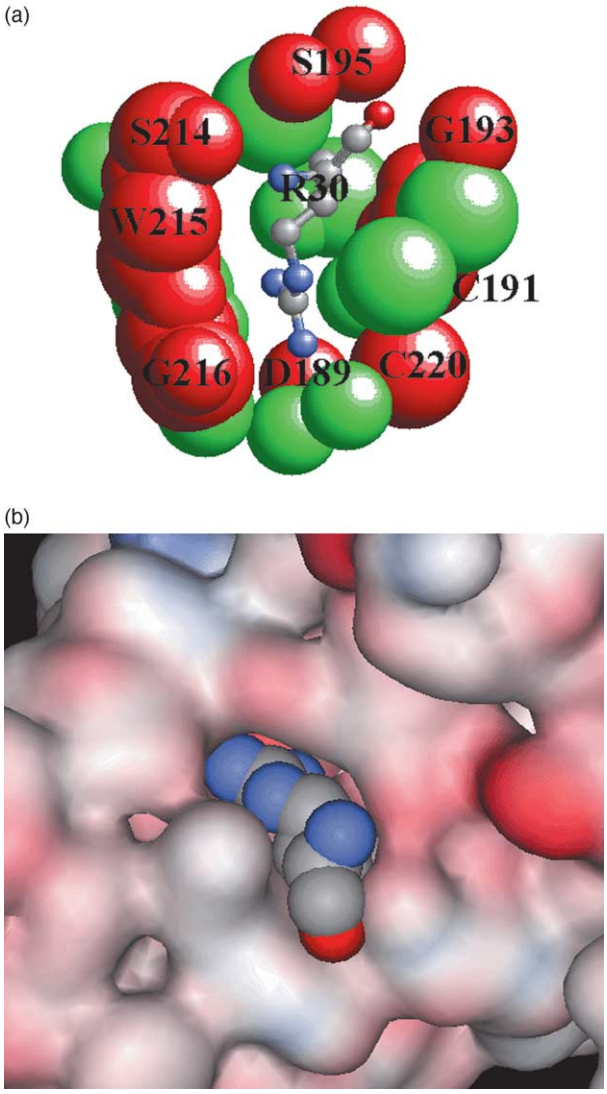

Figure 12. A complemented pocket on the interface of the complex of bacterial 3α,20β-hydroxysteroid dehydrogenase and hirustasin (pdb code 1hia). (a) The protruding residues Thr76-Gly77-Glu78 and the complemented pocket. This complemented pocket consists of 16 residues or 33 heavy atoms from the dehydrogenase. Nine structurally conserved residues, including eight complemented pocket residues from the dehydrogenase and one protruding residue, Arg30, from hirustasin, are labeled and colored in red. Unlike the protruding residues in Figures 4(a), 5(a), and 6(a), the protruding residue Arg30 (colored in CPK) here is not registered as structurally conserved. (b) Geometric fitting of the protruding residue Arg30 and the complementary pocket.

their contribution is cooperative, rather than additive.

The geometrical analysis reveals that many key regions on the interface can be characterized by complemented pockets. Structurally conserved residues are enriched in complemented pockets where local packing is tight, and are disfavored in unfilled pockets. The unfilled pockets and other imperfectly fitted regions may provide the crucial interface flexibility.[10] Based on limited experimental data, we find that complemented pockets are energetically critical for protein–protein interactions, as they account for about 60% of all hot spot residues, and about 90% of the red-hot residues with large free energy change upon Ala mutation. Protein–protein interactions are thus well engineered to follow physico-chemical principles similar to those in protein folding, adapting for their biological binding functions.

Protein–protein binding depends crucially on the energetic hot spots, as they provide most of the binding energy. Hence, the questions arise as to what is the mechanistic role of the hot spot residues in protein–protein binding, and how can we identify hot spots from the single unbound state. Our results suggest that pre-existing pockets are important, and they are often complemented in binding. They further indicate that hot spots and conserved residues are enriched in complemented pockets. The mechanism of binding in most of the cases that we have examined does not involve changes that abolish pockets pre-existing in unbound structures. Furthermore, the complemented pockets existing in the 18 complexes of the Benchmark 1.0 dataset are significantly less hydrophilic than other surface pockets, which exist in all 31 complexes at the unbound states. This result is consistent with that obtained on the entire interface dataset. The lower hydrophilicity of the complemented pockets might be used to predict the complemented pockets out of all surface pockets that exist in the unbound states.

To conclude, rigid docking encounters difficulties, as many parts of the protein–protein interface experience conformational changes. Dynamics is an inherent property of macromolecules, particularly of their surfaces. Nevertheless, encouragingly, our results and those reported by Rajamani *et al.*[40] show that hot spots and conserved residues important for protein binding are often pre-organized to take shapes similar to those in the bound state. If these important residues can be identified from their unbound state, these pocket and protrusion residues can provide valuable assistance to even relatively rigid docking strategies.

In this study, we have focused on geometry, which inherently implies rigid conformations. Nevertheless, protein binding site dynamics is extremely important, allowing multiple different ligands to bind at a single protein site. Backbone and side-chains move, and cooperative optimization of the molecular associations is crucial. The results presented here suggest that in the populated native state, the key residues contributing in a significant way to the free energy of the association tend to adopt preferred states within certain hot regions. In the bound state, these regions will be well packed, often involving complemented interface pocket geometries.

During the revision stage of this paper, the paper by Camacho *et al.*[40] was published. Although different methods have been employed, and the details of the observations differ, the conclusions regarding preferred conformational states of some key residues are in agreement with those reached in this work.

Materials and Methods

Dataset and definition of protein–protein interfaces

The dataset of structurally conserved residues are taken from Keskin *et al.*[15] These entries consist of 21,686 two-chain interfaces. An iterative pairwise structural comparison[14,48–50] and a heuristic clustering procedure were employed to cluster these interfaces. Following six cycles, 3799 clusters are obtained. To further eliminate redundancy, if one of two sequences in a cluster shares a sequence similarity greater than 50%, it is deleted from the cluster. This yields a dataset with structurally similar but sequentially dissimilar member-interfaces. The cluster is required to have at least five members (ten chains). These additional procedures reduce the number of clusters from 3799 to 103. Details have been provided by Keskin *et al.*[15] Finally, MultiProt,[51,52] a multiple structure alignment method, structurally aligns the members of each cluster simultaneously, to identify structurally conserved positions in the interface cluster. A residue is defined as conserved if it occurs in 50% of the cluster members at that geometrical superimposed site.

For the geometrical characterization, we employ the same definition of a protein–protein interface as that used by Tsai *et al.*,[9] which was also used to construct the dataset.[15] Interface residues include both contacting and nearby residues, which correspond to water-mediated residues.[53] Two residues across the interface are considered to be contacting if there is at least a pair of heavy atoms, one from each residue, at a distance less than the sum of their van der Waals radii plus a threshold of 0.5 Å. Residues whose C^{α} atoms are within a distance of 6.0 Å from a C^{α} atom of a contacting residue are considered to be nearby residues. Exposed nearby residues are included, since residues not directly in contact with the binding partner may still be accessible to the solvent, such as the residues at the bottom of unfilled pockets. Those residues may contribute to the binding free energy.[18,24] To exclude completely buried inaccessible residues from our analyses, a residue must experience a solvent-accessible surface area change (>2 Å^2) upon complexation, regardless of whether it is a contacting residue or a nearby residue.

Search for unfilled pockets and complemented pockets

In this study, a pocket is defined as an empty concavity with or without access to solvent molecules. Protein pockets are computed with the weighted Delaunay triangulation, alpha shape, and the pocket algorithm.[54–57]

As depicted in Figure 1, there can be two types of pockets on protein–protein interfaces: unfilled pockets and complemented pockets. Unfilled pockets are those that remain as pockets after protein–protein association. They can appear on protein–protein interfaces, since the interchain packing is not perfect. For example, a pocket on the interface part of one chain can constitute an unfilled pocket if it is not complemented by residues from the other chain. Two shallow depressions on the interfacial parts of two chains can form an unfilled pocket when they face each other upon complexation. Unfilled pockets can contain solvent molecules, which are excluded in this analysis. Complemented pockets are pockets that would be on the interface part of one chain if the protein complexes disassociate, but are complemented or filled by some protruding residues from the binding partner following protein–protein complexation. Protruding residues have heavy-atom contacts with at least three residues from the binding partner, which are located at the bottom of a pocket. We consider only unfilled pockets or complemented pockets that satisfy the following three criteria: (1) at least 80% of the pocket residues belong to interface residues. This is to ensure that we are examining residues relevant to protein–protein interactions. (2) A pocket can occupy no more than 70% of the total interface surface. A very large unfilled pocket is frequently a functional feature of a special protein complex, such as those involved in DNA binding. A very large complemented pocket is frequently the binding interface of an enzyme–inhibitor complex, where an inhibitor is engulfed almost entirely by this large pocket. (3) The surface area of a pocket must be greater than 20 Å^2. This assists in excluding spurious protein–protein interactions due to small conformational fluctuations.

Enrichment of structurally conserved residues

The enrichment factors E_κ for a cluster κ of structurally conserved residues at unfilled pockets or complemented pockets are calculated as:

$$E_\kappa = \frac{\sum_{i \in \kappa} \frac{p_i}{P_i} / \frac{s_i}{S_i}}{|\kappa|}$$

where p_i is the number of conserved pocket residues on interface i, which is a member of cluster κ, P_i is the number of all pocket residues on interface i, s_i is the number of conserved residues on interface i, S_i is the number of all residues on interface i, $|\kappa|$ is the total number of interface members within the cluster κ. $(p_i/P_i)/(s_i/S_i)$ is the enrichment factor of conserved residues for one interface with index i. E_κ is the average of all member interfaces of a cluster.

Analysis of hot spot residue

Here, we follow the definition of a hot spot suggested by Bogan & Thorn. A hot spot is a residue the mutation of which to alanine leads to a $\Delta\Delta G$ of at least 2.0 kcal/mol. This definition is commonly accepted and has been used by us and others in the past.[10] Moderate changes in this definition are unlikely to affect our main conclusion.

Conservation propensities of residues at complemented pockets

The conservation propensities of an amino acid residue a at complemented pockets is calculated as:

$$E(a) = \frac{\sum_{\kappa \in I} E_\kappa(a)}{|I|}$$

where $E_\kappa(a)$ is the enrichment factor of a structurally conserved residue a in complemented pockets for cluster κ of protein–protein interfaces, I is the set of all clusters, and $|I|$ is the total number of interface clusters.

Acknowledgements

We thank Dr L. Admin, J. F. Zhang, A. Binkowski, R. Jackups, P. Freeman, and J. Tseng for helpful discussions. We thank Drs C.-J. Tsai, Y. Pan, K. Gunasekaran, D. Zanuy, H.-H (G). Tsai and members of the Nussinov-Wolfson group, in particular Maxim Shatsky for help with MultiProt, and Inbal Halperin for the hot spot collection. We thank Dr Jacob V. Maizel for encouragement. We thank Dr A. Gursoy and S. Aytuna for their helpful discussions. This work is supported by grants from National Science Foundation (CAREER DBI013356, DBI0078270), and National Institute of Health (GM68958). The research of R.N. in Israel has been supported, in part, by the Center of Excellence in Geometric Computing and its Applications funded by the Israel Science Foundation (administered by the Israel Academy of Sciences). This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-12400.

References

1. Tsai, C. J., Lin, S. L., Wolfson, H. J. & Nussinov, R. (1997). Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci.* **6**, 53–64.
2. Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
3. Larsen, T. A., Olson, A. J. & Goodsell, D. S. (1998). Morphology of protein–protein interfaces. *Structure*, **6**, 421–427.
4. Valdar, W. J. & Thornton, J. M. (2001). Protein–protein interfaces: analysis of amino acid conservation in homodimers. *Proteins: Struct. Funct. Genet.* **42**, 108–124.
5. Conte, L. L., Chothia, C. & Janin, J. (1999). The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
6. Lichtarge, O. & Sowa, M. E. (2002). Evolutionary predictions of binding surfaces and interactions. *Curr. Opin. Struct. Biol.* **12**, 21–27.
7. Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18**, 71S–777.
8. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf:

identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
9. Hu, Z., Ma, B., Wolfson, H. & Nussinov, R. (2000). Conservation of polar residues as hot spots at protein interfaces. *Proteins: Struct. Funct. Genet.* **39**, 331–342.
10. Ma, B., Elkayam, T., Wolfson, H. & Nussinov, R. (2003). Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl Acad. Sci. USA*, **100**, 5772–5777.
11. Ma, B., Wolfson, H. J. & Nussinov, R. (2001). Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr. Opin. Struct. Biol.* **11**, 364–369.
12. Zhou, H. & Shan, Y. (2001). Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins: Struct. Funct. Genet.* **44**, 336–343.
13. Lichtarge, O., Bourne, H. R. & Cohen, F. E. (1996). An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* **257**, 342–358.
14. Tsai, C.-J., Lin, S. L., Wolfson, H. J. & Nussinov, R. (1996). A dataset of protein–protein interfaces generated with a sequence-order-independent comparison technique. *J. Mol. Biol.* **260**, 604–620.
15. Keskin, O., Tsai, C.-J., Wolfson, H. & Nussinov, R. (2004). A new, structurally nonredundant, diverse data set of protein–protein interfaces and its implications. *Protein Sci.* **13**, 1043–1055.
16. Hubbard, S. J. & Argos, P. (1994). Cavities and packing at protein interfaces. *Protein Sci.* **3**, 2194–2206.
17. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
18. DeLano, W. L. (2002). Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **12**, 14–20.
19. Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994). Protein–protein recognition: crystal structural analysis of a barnase–barstar complex at 2.0-Å resolution. *Biochemistry*, **33**, 8878–8889.
20. Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone–receptor interface. *Science*, **267**, 383–386.
21. Bogan, A. A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9.
22. Thorn, K. S. & Bogan, A. A. (2001). ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, **17**, 284–285.
23. Guerois, R., Nielsen, J. E. & Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387.
24. Kortemme, T. & Baker, D. (2002). A simple physical model for binding energy hot spots in protein–protein complexes. *Proc. Natl Acad. Sci. USA*, **99**, 14116–14121.
25. Janin, J. (1999). Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. *Struct. Fold. Des.* **7**, R277–R279.
26. Chen, R., Mintseris, J., Janin, J. & Weng, Z. (2003). A protein–protein docking benchmark. *Proteins: Struct. Funct. Genet.* **52**, 88–91.
27. Connolly, M. (1986). Shape complementarity at the hemoglobin alpha 1 beta 1 subunit interface. *Biopolymers*, **25**, 1229–1247.
28. Lamb, M. L. & Jorgensen, W. L. (1997). Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.* **1**, 449–457.
29. Lengauer, T. & Rarey, M. (1996). Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **6**, 402–406.
30. Sternberg, M. J., Gabb, H. A. & Jackson, R. M. (1998). Predictive docking of protein–protein and protein–DNA complexes. *Curr. Opin. Struct. Biol.* **8**, 250–256.
31. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. (2002). Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins: Struct. Funct. Genet.* **47**, 409–443.
32. Liang, J. & Dill, K. A. (2001). Are proteins well-packed? *Biophys. J.* **81**, 751–766.
33. Zhang, E., St. Charles, R. & Tulinsky, A. (1999). Structure of extracellular tissue factor complexed with factor VIIa inhibited with a BPTI mutant1. *J. Mol. Biol.* **285**, 2089–2104.
34. Binkowski, T., Adamian, L. & Liang, J. (2003). Inferring functional relationships of proteins from local sequence and spatial surface patterns. *J. Mol. Biol.* **332**, 505–526.
35. Halperin, I., Wolfson, H. & Nussinov, R. (2004). Protein–protein interactions: coupling of structurally conserved residues and of hot spots across protein–protein interfaces. *Structure*, **12**, 1027–1038.
36. Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994). Protein–protein recognition: crystal structural analysis of a barnase–barstar complex at 2.0-Å resolution. *Biochemistry*, **33**, 8878–8889.
37. Fields, B. A., Goldbaum, F. A., Ysern, X., Poijak, R. J. & Mariuzza, R. A. (1995). Molecular basis of antigen mimicry by an anti-idiotope. *Nature*, **374**, 739–742.
38. Scheidig, A. J., Hynes, T. R., Pelletier, L. A., Wells, J. A. & Kossiakoff, A. A. (1997). Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid {beta}-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities. *Protein Sci.* **6**, 1806–1824.
39. Castro, M. J. M. & Anderson, S. (1996). Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin. *Biochemistry*, **35**, 11435–11446.
40. Rajamani, D., Thiel, S., Vajda, S. & Camacho, C. (2004). Anchor residues in protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **101**, 11287–11292.
41. del Sol Mesa, A., Pazos, F. & Valencia, A. (2003). Automatic methods for predicting functionally important residues. *J. Mol. Biol.* **326**, 1289–1302.
42. Goh, C.-S., Bogan, A. A., Joachimiak, M., Walther, D. & Cohen, F. E. (2000). Co-evolution of proteins with their interaction partners1. *J. Mol. Biol.* **299**, 283–293.
43. Goh, C.-S. & Cohen, F. E. (2002). Co-evolutionary analysis reveals insights into protein–protein interactions. *J. Mol. Biol.* **324**, 177–192.
44. Valencia, A. & Pazos, F. (2002). Computational methods for the prediction of protein interactions. *Curr. Opin. Struct. Biol.* **12**, 368–373.
45. Horton, N. & Lewis, M. (1992). Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**, 169–181.
46. Chothia, C. & Janin, J. (1975). Principles of protein–protein recognition. *Nature*, **256**, 705–708.
47. Brooijmans, N., Sharp, K. A. & Kuntz, I. D. (2002). Stability of macromolecular complexes. *Proteins: Struct. Funct. Genet.* **48**, 645–653.
48. Nussinov, R. & Wolfson, H. (1991). Efficient detection of three-dimensional structural motifs in biological

macromolecules by computer vision techniques. *Proc. Natl Acad. Sci. USA*, **88**, 10495–10499.
49. Fischer, D., Bachar, O., Nussinov, R. & Wolfson, H. J. (1992). An efficient automated computer vision based technique for detection of three dimensional structural motifs in proteins. *J. Biomol. Struct. Dynam.* **9**, 769–789.
50. Bachar, O., Fischer, D., Nussinov, R. & Wolfson, H. (1993). A computer vision based technique for 3-D sequence-independent structural comparison of proteins. *Protein Eng.* **6**, 279–288.
51. Shatsky, M., Nussinov, R. & Wolfson, H. (2002). Multiprot: a multiple protein structural alignment. *Lecture Notes Comput. Sci.* **2452**, 235–250.
52. Shatsky, M., Nussinov, R. & Wolfson, H. (2004). A method for simultaneous alignment of multiple protein structures. *Proteins: Struct. Funct. Genet.* **56**, 143–156.
53. Papoian, G. A., Ulander, J. & Wolynes, P. G. (2003). Role of water mediated interactions in protein–protein recognition landscapes. *J. Am. Chem. Soc.* **125**, 9170–9178.
54. Edelsbrunner, H., Facello, M. & Liang, J. (1996). On the definition and the construction of pockets in macromolecules. *Pac. Symp. Biocomput.*, 272–287.
55. Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V. & Subramaniam, S. (1998). Analytical shape computation of macromolecules. I. Molecular area and volume through alpha shape. *Proteins: Struct. Funct. Genet.* **33**, 1–17.
56. Edelsbrunner, H. & Mucke, E. (1994). Three-dimensional alpha shapes. *ACM Trans. Graph.* **13**, 43–72.
57. Edelsbrunner, H., Facello, M. & Liang, J. (1998). On the definition and the construction of pockets in macromolecules. *Disc. Appl. Math.* **88**, 18–29.

Edited by M. Levitt

(Received 3 May 2004; received in revised form 17 September 2004; accepted 21 September 2004)