Hydration of Protein–Protein Interfaces

Francis Rodier,[1†] Ranjit Prasad Bahadur,[2†] Pinak Chakrabarti,[2] and Joël Janin[1*]
[1]*Laboratoire d'Enzymologie et de Biochimie Structurales, Gif-sur-Yvette, France*
[2]*Department of Biochemistry, Bose Institute, Calcutta, India*

ABSTRACT **We present an analysis of the water molecules immobilized at the protein–protein interfaces of 115 homodimeric proteins and 46 protein–protein complexes, and compare them with 173 large crystal packing interfaces representing nonspecific interactions. With an average of 15 waters per 1000 Å^2 of interface area, the crystal packing interfaces are more hydrated than the specific interfaces of homodimers and complexes, which have 10–11 waters per 1000 Å^2, reflecting the more hydrophilic composition of crystal packing interfaces. Very different patterns of hydration are observed: Water molecules may form a ring around interfaces that remain "dry," or they may permeate "wet" interfaces. A majority of the specific interfaces are dry and most of the crystal packing interfaces are wet, but counterexamples exist in both categories. Water molecules at interfaces form hydrogen bonds with protein groups, with a preference for the main-chain carbonyl and the charged side-chains of Glu, Asp, and Arg. These interactions are essentially the same in specific and nonspecific interfaces, and very similar to those observed elsewhere on the protein surface. Water-mediated polar interactions are as abundant at the interfaces as direct protein–protein hydrogen bonds, and they may contribute to the stability of the assembly. Proteins 2005;60:36–45.**

Key words: protein–protein recognition; specific and nonspecific interactions; interface water; crystal packing; hydrogen bonds

INTRODUCTION

When polypeptide chains assemble to form a complex or a protein oligomer, bulk solvent is excluded from the interface, allowing direct protein–protein contacts and creating a favorable hydrophobic effect. However, X-ray structures generally show that solvent remains caught at the interface between the protein components. The possible contribution of bound water molecules to the stability of the assembly and the specificity of recognition has been recognized early.[1,2] The subject is often mentioned by crystallographers reporting the structure of protein–protein complexes, yet general studies have been very few. The present work aims to quantify the number of water molecules associated with protein–protein interfaces, the types of interactions they make, and the amino acid residues to which they are bound. We may compare the interfaces to the protein surface, where bound solvent has been analyzed in more detail.[3–7] For that purpose, we utilize the information stored in the Protein Data Bank (PDB)[8] and compare the interfaces in a set of homodimers and protein–protein complexes to a set of crystal packing interfaces. The former offers examples of stable and specific protein assemblies[9–14]; the latter illustrates nonspecific interactions,[15–18] similar to those all molecules undergo as they collide in solution and in the cell. The specific and nonspecific interfaces have been shown to differ in their size and chemical composition, and in the way protein atoms pack.[19,20] We show here that they also differ in the degree of hydration and the way water molecules are distributed across the protein surface. Residual hydration leaves about 10 waters per 1000 Å^2 of interface area in protein–protein complexes and homodimers, and 50% more on average in crystal packing interfaces. Specific interfaces tend to be "dry," with water molecules forming a ring around interface atoms, whereas a majority of the crystal packing contacts are "wet," as water permeates the interface. However, each category is heterogeneous from these points of view, and no general conclusion can be drawn on the role of the residual hydration in the stability and specificity of protein–protein recognition.

METHODS AND RESULTS

Residual Hydration of Protein–Protein Interfaces

The sample of protein–protein interfaces examined here includes 334 interfaces taken from 295 PDB entries: 46 protein–protein complexes, 115 homodimeric proteins, and 173 large crystal packing interfaces (Table I). The latter are defined as pairwise contacts between monomeric proteins in crystals that have an interface area B larger than 800 Å^2. B is calculated as the sum of the solvent accessible surface areas (ASAs) of the two protein subunits less that of the pair, evaluating the ASAs with program NACCESS by S. Hubbard (University College, London), which implements the Lee and Richards[21] algorithm. The sample of protein–protein interfaces we used here is a subset of the

Grant sponsor: Council of Scientific and Industrial Research of India (senior research fellowship to R. P. Bahadur). Grant sponsor: EIDIPP program of Action Concertée Incitative IMPBio (financial support to J. Janin and F. Rodier).

†F. Rodier and R. P. Bahadur contributed equally to this work.

*Correspondence to: Joël Janin, Laboratoire d'Enzymologie et de Biochimie Structurales, CNRS UPR 9063, 91198-Gif-sur-Yvette, France. E-mail: janin@lebs.cnrs-gif.fr

Received 24 November 2004; Accepted 12 January 2005

Published online 26 April 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20478

TABLE I. PDB Entries for Specific and Nonspecific Interfaces

Specific interfaces									
Homodimers (115)									
12as	1a3c	1a4i	1a4u	1aa7	1ad3	1ade	1afw	1ajs	1alo
1amk	1aor	1aq6	1auo	1b3a	1b5e	1b67	1b68	1b69	1b70
1bbh	1bd0	1bif	1biq	1bis	1bjw	1bkp	1bmd	1brw	1bsl
1bsr	1buo	1bxg	1bxk	1cdc	1cg2	1chm	1cmb	1cnz	1coz
1csh	1ctt	1cvu	1czj	1daa	1dor	1dpg	1dqs	1dxg	1e98
1ebh	1f13	1fip	1fro	1gvp	1hjr	1hss	1hxp	1icw	1imb
1isa	1ivy	1jhg	1jsg	1kba	1kpf	1m6p	1mkb	1mor	1nox
1nse	1nsy	1oac	1opy	1pgt	1qfh	1qhi	1qr2	1r2f	1reg
1rpo	1ses	1slt	1smn	1sox	1tcl	1tox	1trk	1uby	1utg
1vfr	1vok	1wtl	1xso	2arc	2ccy	2hdh	2ilk	2lig	2mcg
2nac	2ohx	2spc	2sqc	2tct	2tgi	3dap	3grs	3sdh	3ssi
4cha	5csm	5rub	8prk	9wga					
Complexes (46)									
1a2k	1acb	1ak4	1avw	1brs	1bth	1cbw	1cho	1cse	1dan
1dfj	1dhk	1dvf	1efn	1efu	1fin	1fle	1gg2	1got	1gua
1hia	1hwg	1igc	1kb5	1mct	1mel	1mlc	1nca	1nmb	1osp
1ppf	1stf	1tgs	1tx4	1vfb	1ycs	1ydr	2kai	2pcc	2ptc
2sic	2sni	2trc	3sgb	3tpi	4htc				
Large crystal packing interfaces									
Two-fold symmetry (93)									
13pk	1a7v_1	1a7v_2	1afk	1ag9	1ah7_1	1ako_1	1ako_2	1amu_1	1atl
1aw7	1ay1	1b1j_1	1b1j_2	1bc2	1bea	1bin	1bkz	1byo	1c02
1caq	1cki_1	1clu	1cqx	1dsu	1dys	1e0s_1	1e0s_2	1ehy	1epa
1ewf_1	1feh	1fgk	1fjm	1fkd	1fmt_1	1fmt_2	1g2a	1gar	1gjm
1hf8	1ilr_1	1kpt	1kwa_1	1mpg_1	1mpg_2	1mss	1naw_1	1naw_2	1np4
1pbg_1	1pbg_2	1pda_1	1ppo_1	1ppo_2	1ppo_3	1qaz	1qci	1qha	1qjp
1qme_1	1qpa	1qtq_1	1qtq_2	1rb3	1rhs_1	1rne	1shk	1the_1	1the_2
1tht	1toa	1ton	1urp_1	1urp_2	1vbt	1xgs	256b	256l	2acy
2atj_1	2bc2	2bls_1	2g3p	2ihl	2mbr	2scp	2shp	2tps	2ugi
3pmg	830c_1	8pti							
No 2-fold symmetry (80)									
1al2_1	1a12_2	1a39	1a3y_1	1a3y_2	1a6q	1a7t	1ae9	1af7	1ah7_2
1ajk	1akz	1amu_2	1aq0	1aqz	1ayi	1b24	1b80	1be0	1bf6
1bg0	1bgc	1cfy	1cki_2	1dxm	1dz4_1	1dz4_2	1elp	1ewf_2	1ewf_3
1fmt_3	1fvk_1	1fvk_2	1hrn_1	1hrn_2	1hsl	1i4g_1	1i4g_2	1ihb	1ilr_2
1inp_1	1inp_2	1jfr	1kfs_1	1kfs_2	1kwa_2	1mbl	1ml1	1mwc	1nmt
1nuc	1ome	1pda_2	1pmi	1pva	1qme_2	1qnt_1	1qnt_2	1qs8	1qsn
1rge	1rhs_2	1sw6_1	1sw6_2	1trn	1vjw	1vlz_1	1vlz_2	1xca	1ygh
1zin	2atj_2	2bls_2	2end	2nap	2rn2	3cms	3ngl_1	3ngl_2	830c_2

Interfaces in complexes and homodimers are taken from Chakrabarti and Janin[13] and Bahadur et al.[14] Large crystal packing interfaces are taken from Bahadur et al.[20] and bury > 800 Å^2; in the present study, the 2-fold/no 2-fold categories are merged. When a PDB entry contains several large packing interfaces, they are labeled _1, _2, and so on.

sets described elsewhere,[13,14,20] removing entries with resolution worse than 2.6 Å or lacking solvent information.

We define interface atoms as protein atoms that lose ASA in the assembly and contribute to B, and interface waters as crystallographic solvent molecules that are within 4.5 Å from interface atoms of both sides of the assembly after checking all crystallographic transformations. The 334 interfaces contain a total of 9903 water molecules, representing 9.5% of all the solvent molecules reported in the 295 PDB entries. As the interfaces of this sample cover about 10% of the solvent-accessible protein surface, this suggests that immobilized water molecules are almost as abundant at interfaces as elsewhere on the protein surface.

The number N_w of waters per interface is 30 on average (Table II). N_w is about twice larger in homodimers than in complexes and in large crystal packing interfaces, but its standard deviation is large in each category, and the range is extremely wide. Whereas only 1 water molecule is reported at the kallikrein–pancreatic trypsin inhibitor (PTI) interface (2kai), the copper amine oxidase dimer interface (1oac) has 190. A reason for this wide range is the lack of consistency in the way solvent positions are described in crystallographic studies. Notably, fewer interface waters are reported in lower resolution X-ray studies. Figure 1 shows that relative to the set in Table I, the average of N_w increases by 15% if we limit the sample to PDB entries at resolution 2 Å or better, and by 44% for the 40 entries at resolution 1.6 Å or better. Nevertheless, the range of N_w remains wide, 8–120, in the 1.6 Å set, which strongly suggests that there are real differences between interfaces in terms of hydration in spite of the inconsistent reporting.

TABLE II. Statistics on Interface Water Molecules

Interfaces	Homodimers	Complexes	Crystal packing	All
Number of interfaces	115	46	173	334
Interface area B (Å^2)	3900	1970	1480	2380
s.d.	2250	820	490	1790
Water molecules[a]				
Mean number per interface	44	20	23	30
s.d.	33	13	14	25
Per 1000 Å^2 interface area	11.3	10.0	15.3	12.4
Distance ratio d_r	1.19	1.18	1.01	1.09
s.d.	0.16	0.15	0.18	0.19
Buried fraction				
Interface atoms[b]	0.36	0.34	0.21	0.28
Interface waters	0.31	0.29	0.25	0.29
s.d.	0.16	0.16	0.18	0.17
Temperature factor (B factor)				
Mean B of protein atoms (Å^2)	24.6	28.1	25.9	25.7
Interface atoms/all atoms	0.87	0.93	1.07	0.98
Interface water/all atoms	1.29	1.22	1.36	1.32

[a]Water molecules less than 4.5 Å from atoms of both subunits; s.d., standard deviation.
[b]Data from Table 2 of Bahadur et al.[20]

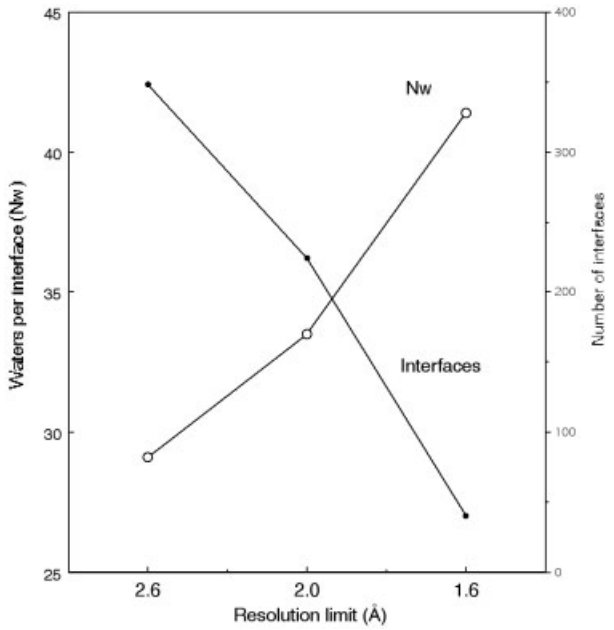

Fig. 1. Interface waters and the resolution of the X-ray structure. Plot of the number of interfaces and the average number N_w of interface water molecules in the PDB entries of Table I (2.6 Å-resolution) and two high-resolution subsets.

The number of interface waters also increases with the size of the interfaces (Fig. 2). The correlation between N_w and the interface area is good in homodimers ($R^2 = 0.82$), weaker in complexes and crystal contacts ($R^2 = 0.45–0.5$). On average, homodimers have interfaces that are about twice larger than those in complexes or the crystal packing interfaces we selected (Table II). These interfaces are in turn much larger than the average pair wise interface in protein crystals[15], which has $B = 570$ Å^2. Although the average homodimer interface contains more water than crystal contacts, the surface density of water is less due to its larger size. The interfaces of complexes and homodimers contain about 10 solvent molecules per 1000 Å^2 of interface area, and crystal packing interfaces, about 15. We find that, irrespective of the interface type, 94% of the waters are within 3.5 Å from at least 1 polar protein atom, counting O, N, and S as polar, and all carbon-containing groups as nonpolar. Homodimer interfaces are 65% nonpolar and 35% polar on average,[14] whereas crystal packing interfaces and interfaces in complexes are 57–58% nonpolar and 42–43% polar on average, the same proportions as the protein accessible surface.[11,20] Thus, the lower density of immobilized waters at homodimer interfaces may reflect their hydrophobic character. To account for the difference in chemical composition, we may relate N_w to the surface area contributed to the interfaces by the polar groups only (polar area). Then, the density of water molecules becomes similar in homodimers and crystal contacts: On average, 32 immobilized waters are reported per 1000 Å^2 polar interface area. In the 2 Å-resolution sample, the density is 37 waters per 1000 Å^2 polar area. As each C, N, or O interface atom contributes an average 9.9 Å^2 to the interface area irrespective of the interface type,[20] the average protein–protein interface contains about 1 immobilized water molecule for every 3 polar atoms.

We also estimated the area covered by interface water molecules on the protein surface by calculating the ASA of a complex or subunit pair twice, with and without the interface waters, and taking the difference of the 2 values. The difference correlates linearly with N_w. The slope of the plot is 18 Å^2 per water molecule in homodimers ($R^2 = 0.97$) and 20 Å^2 in complexes and crystal packing interfaces ($R^2 = 0.94–0.89$). This suggests that, irrespective of the type of interface, each interface water covers 9–10 Å^2 of the surface of each subunit, which is equivalent to the ASA of the average protein C, N, or O atom.

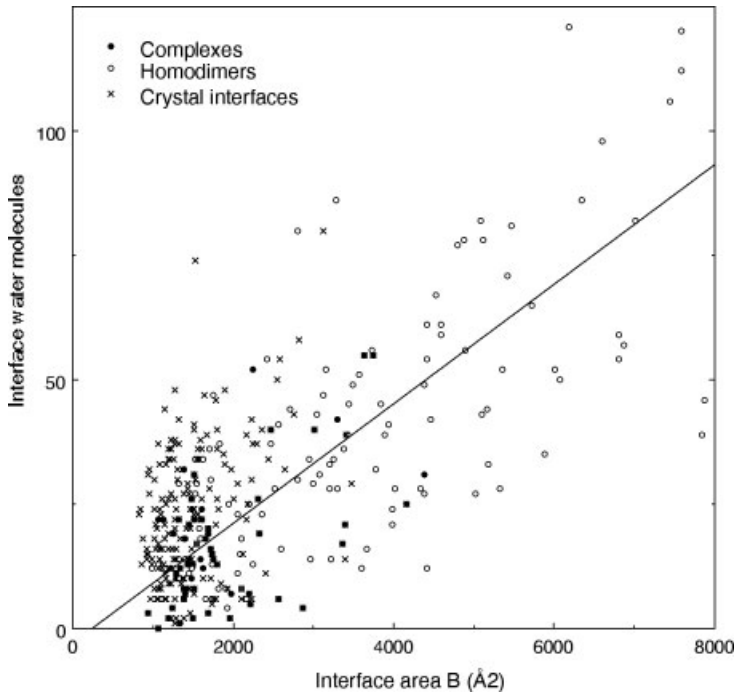

Fig. 2. Interface size and water molecules. The number N_w of water molecules is plotted against the interface area *B* in protein–protein complexes (●), homodimers (○), and large crystal packing interfaces (x).

Patterns of Interface Hydration

Figure 3 shows 6 protein–protein interfaces with the solvent molecules drawn as spheres: 2 in homodimers, 2 in enzyme–inhibitor complexes, and 2 in crystal packing. Both homodimer interfaces are dry inside. The one between the subunits of the protein kinase C (PKC) interacting protein (1kpf) has solvent all around and none inside; that of malate dehydrogenase (1bmd) is in two patches, and solvent lines both patches. The patches were defined by applying to interface atoms a geometric clustering procedure.[13] This procedure ignores water molecules and therefore introduces no bias in their distribution. In the same way, the elastase–ovomucoid inhibitor interface (1ppf) contains water around a dry patch. Patterns of hydration that leave most of a protein–protein interface dry have been reported in many other structures.[29] However, the dry pattern is not general. "Wet" interfaces where water permeates the contact are the rule in protein–DNA interaction, and they also occur in protein–protein complexes.[30] Two well-defined examples are the α-amylase–bean inhibitor interface (1dhk) and the 1i4g_2 crystal packing interface illustrated in Figure 3. These 2 interfaces are typically wet, with many water molecules spread evenly across them. Intermediate cases are also common: Immobilized solvent surrounds a patch of protein surface that is incompletely dehydrated. The myoglobin crystal packing interface (1mwc) of Figure 3 illustrates that situation.

To quantify the spatial distribution of solvent molecules within an interface, we calculate the distance of each interface water to the center of mass of that interface and compare its average value r_{wat} to one, r_{atom}, calculated for the protein atoms that form the interface. In multipatch interfaces, patches being defined as in Chakarabarti and Janin,[13] the distances are to the center of the patch, and each water molecule is attributed to the patch with which it had the shortest contact. The ratio $d_r = r_{\mathrm{wat}}/r_{\mathrm{atom}}$ describes how water is distributed relative to other interface atoms. In a wet interface where water is evenly distributed throughout the interface, d_r should be 1 or less; if the interface is dry and water localized at its edge, d_r should be larger than 1. The interfaces shown in Figure 3 follow this simple rule. In our sample of 334 interfaces, the range of d_r is 0.5–1.7. In crystal packing interfaces, the mean value is very close to 1; in homodimers and complexes, it is 1.18, calculated on all single-patch interfaces and the larger patch of each multipatch interface (Table II). As in Figure 3, the mean values of d_r suggest that crystal contacts tend to be wet, and specific interfaces, dry, but a wide distribution of d_r is observed in both types (Fig. 4), and there are obvious exceptions to this rule. In practice, half of the crystal packing interfaces and 13% of the specific interfaces have d_r less than 1.0; these interfaces are typically wet. Conversely, if we take $d_r > 1.2$ to indicate a dry interface, 44% of the specific interfaces, but only 12% of the crystal packing interfaces, are dry. Values of d_r between 1 and 1.2 are ambiguous and more refined criteria are needed to distinguish between the dry, wet, and intermediate situations.

Accessibility and Mobility of Interface Water

The solvent molecules that permeate wet interfaces occupy sites in grooves or cavities between the 2 protein surfaces.[31] One might expect them to be buried within the

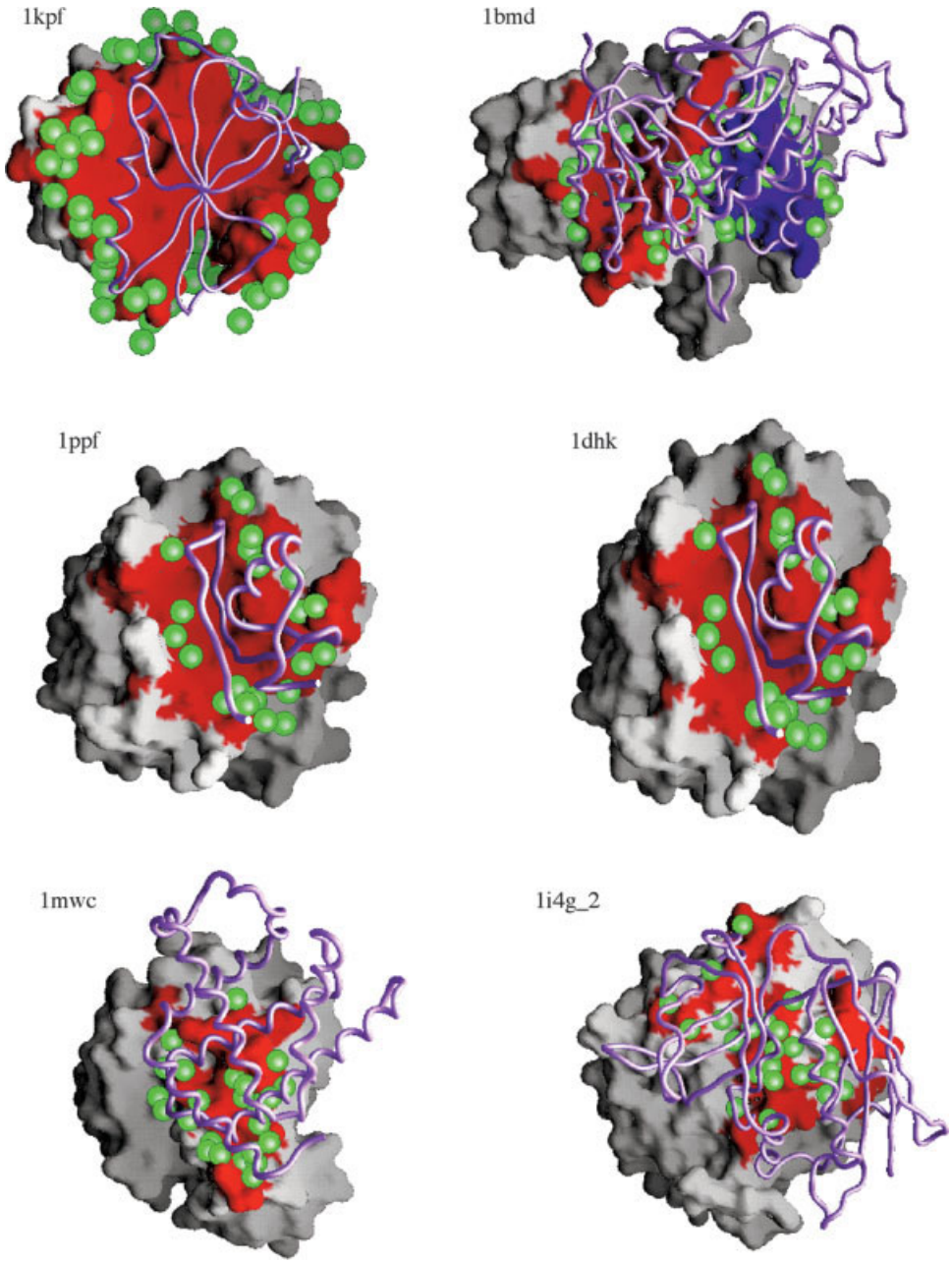

Fig. 3. Patterns of interface hydration. In each panel, one protein subunit is shown as a backbone worm, the other as a molecular surface with the interface in color. Green spheres are water molecules. Drawn with GRASP.[22] 1kpf: The subunit interface of the dimeric PKC interacting protein[23] is typically dry; its interface area is $B = 3700$ Å^2; $N_w = 56$; $d_r = 1.34$. 1bmd: The subunit interface of malate dehydrogenase[24] is in 2 patches colored red and blue; interface waters line both patches; $B = 3100$ Å^2; $N_w = 47$; $d_r = 1.35$ (red patch) and 1.15 (blue patch). 1ppf: Elastase–ovomucoid inhibitor complex[25]; $B = 1320$ Å^2; $N_w = 22$; $d_r = 1.5$. 1dhk: α-amylase–bean inhibitor complex[26]; $B = 3010$ Å^2; $N_w = 40$; $d_r = 0.85$. 1mwc: A wet packing interface in myoglobin crystals[27]; $B = 840$ Å^2; $N_w = 16$; $d_r = 1.0$; 1i4g_2: A wet packing interface in crystals of enterotoxin A[28]; $B = 1460$ Å^2; $N_w = 20$; $d_r = 0.7$.

protein assembly, whereas those that form rings around the interface patches, as in 1kpf or 1ppf, remain in contact with bulk solvent. We calculated the ASA of each interface water including all protein atoms and other interface waters in the calculation. A water was considered as buried at the interface if it was within 3.5 Å from both the subunits and had an ASA < 5 Å^2, which corresponds to 5% of the ASA of a probe of radius 1.4 Å. On that basis, about 30% of the interface water molecules in homodimers and complexes are buried, but only 25% in crystal packing interfaces (Table II). The difference between the two mean values is much less than their standard deviations, yet it recalls a similar difference that is observed for protein atoms in the same interfaces: 34–36% of the interface

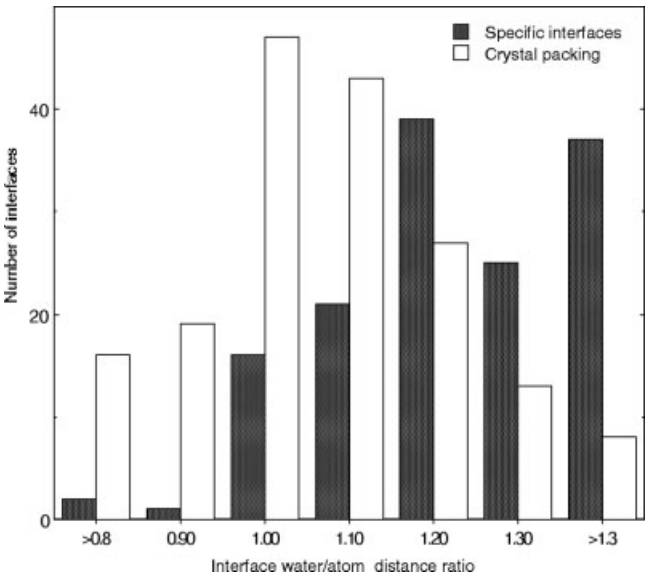

Fig. 4. Histogram of the interface water/atom distance ratio d_r. Crystal packing interfaces (white bars) and interfaces in complexes and homodimeric proteins (shaded). In cases where the latter are multipatch, only the largest patch is considered.

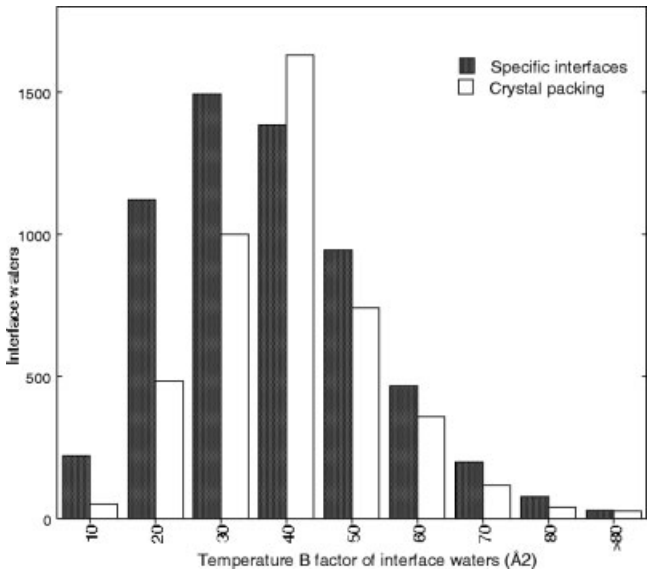

Fig. 5. Histogram of the temperature factors (*B* factors) of interface waters. Crystal packing interfaces (white bars) and interfaces in complexes and homodimeric proteins (shaded).

atoms are fully buried (zero ASA) in homodimers and complexes, 21% in crystal packing interfaces.[13,14,20] As the latter proportions refer to ASA calculations that ignore crystallographic solvent, an interface atom that interacts with an interface water molecule may count as accessible, while the water molecule itself may be buried. Thus, in a wet interface like that 1dhk in Figure 3, over half of the water molecules are buried, but only 30% of the protein atoms. In general, the buried fraction is larger for water (25%) than for protein atoms (21%) in crystal packing interfaces. In contrast, protein atoms are more often buried than water in the interfaces of homodimers and complexes, although 1dhk is a counterexample.

In proteins, the accessibility to solvent generally correlates with the mobility of atoms expressed in crystallographic temperature factors (B factors).[32] However, the mean value of atomic *B* factors in a crystal, which determines its diffracting power, is a property of the crystal as a whole, not of the protein. To account for systematic differences between crystals and PDB entries, we divide the *B* factor reported for individual atoms by its mean value in each entry, before averaging in atom categories or interface classes. Table II indicates that the interface atoms of homodimers and complexes are somewhat less mobile than the average protein atom, whereas in crystal packing interfaces, they are more mobile. From that point of view, as from many others, interface atoms resemble atoms of the protein core in homodimers and complexes, whereas atoms at crystal packing interfaces resemble the protein surface.[15–18] Figure 5 is a histogram of the *B* factors of the interface water molecules in our set. The mean value is 1.3 times the mean *B* factor of the protein atoms, indicating that interface water molecules are only slightly more mobile than these atoms. This can safely be attributed to the fact that solvent is not constrained by covalent bonds like protein atoms. Moreover, the mean value of the ratios in Table II shows no clear difference between the specific and nonspecific interfaces.

In addition to measuring mobility, *B* factors are a test of the correct identification of water by crystallographers in electron density maps. Very low values might indicate the presence of a metal ion or a molecule other than water, high values, noise in the electron density. Figure 5 shows that nearly all interface water molecules have *B* factors in the range 10–70 Å^2: Only 2.8% are below, and 1.8% are above that range. Thus, the great majority is correctly interpreted.

Polar Interactions

Solvent molecules that can be located in electron density maps are immobilized and therefore likely to form hydrogen bonds with polar groups on the protein surface. These interactions occur in addition to direct protein–protein hydrogen bonds, which have been often reviewed.[7,11,33–35] We identified water interactions with polar (N-, O-, or S-containing) protein groups based simply on distance, with a cutoff at 3.5 Å more restrictive than the 4.5 Å cutoff that defines water as part of the interface. With this criterion, the average number of water–protein polar interactions is 28 per interface (Table III), close to the number of interface waters. As on average each interface water makes 1 polar interaction to protein groups, waters that bridge the interface by interacting with polar groups on both sides are a minority, 8 out of 30. The fraction of bridging waters is the same, about 30%, in the interfaces of specific and nonspecific interfaces. There are 10 direct hydrogen bonds on average in the same set of interfaces[20]; therefore, water-mediated polar interactions are similar in abundance to protein–protein hydrogen bonds.

On the protein side, the chemical group most often involved in a polar interaction with interface water mol-

TABLE III. Interaction of Interface Water Molecules With Protein Polar Groups

Interfaces	Homodimers	Complexes	Crystal packing	All
Mean number per interface				
Water molecules	44	20	23	30
Bridging waters[a]	13	6	6	8
Water polar interactions	43	21	21	28
s.d.	16	32	14	25
Direct H-bonds[b]	18	10	5	10
Fraction of water polar interactions with				
main-chain NH	0.12	0.11	0.11	0.12
main-chain CO	0.33	0.32	0.32	0.33
charged side-chains	0.29	0.28	0.32	0.30
neutral side-chains	0.25	0.30	0.25	0.26

[a]Water molecules less than 3.5 Å from a polar (N, O, or S) protein atom are assumed to make a polar interaction with that group; those that interact with groups of both subunits are "bridging"; s.d., standard deviation.
[b]Hydrogen bonds between protein groups taken from Table 2 of Bahadur et al.[20]

ecules is the main-chain carbonyl. This group alone makes one-third of the interactions (Table III); the main-chain NH makes 12%. The water molecules are hydrogen bond donors in the first case, acceptors in the second, and the proportion of the two is similar for water at interfaces and elsewhere on the protein surface.[3–6] Other interactions involve protein side-chains: neutral side-chains, 26%, and charged side-chains, 30%. The carboxylates of Asp and Glu, and the guanidinium group of Arg contribute 8–9% each, Lys, 5–6%. The distribution between the chemical types of protein groups is not significantly different in specific and in nonspecific interfaces. The amino acid composition of the protein surface that forms hydrogen bonds with the interface water molecules is cited in Table IV. Asp, Glu, Arg, Ser, and Tyr are abundant both in specific and in nonspecific interfaces,[9–11,18–20] and these residues also frequently interact with water elsewhere on the protein surface.[3,4] A depletion of Lys and an excess of Arg and Tyr at interfaces relative to the solvent accessible protein surface have often been noted in the overall composition of protein–protein interfaces.[9–14] Together with a slight excess of Ser, these are the only significant differences in composition that can be observed in Table IV between the interfaces, specific or not, and the protein surface.

TABLE IV. Amino Acid Residues Making Hydrogen Bonds With Interface Water Molecules

Interfaces Residue type	Homodimers	Complexes	Crystal packing	Protein surface[a]
Ala	4.1	2.5	3.5	5.8
Arg	9.2	7.6	7.6	6.4
Asn	6.3	7.4	6.8	6.1
Asp	9.4	10.6	11.1	11.4
Cys	1.4	1.5	0.6	1.5
Gln	5.4	3.5	6.7	5.6
Glu	10.3	7.8	12.1	8.2
Gly	4.4	7.5	4.3	7.4
His	4.8	4.7	4.0	4.4
Ile	1.6	1.7	1.7	2.8
Leu	2.9	2.1	2.8	3.0
Lys	5.0	6.2	6.0	8.2
Met	1.1	0.9	1.4	1.1
Phe	1.9	0.7	1.0	1.9
Pro	1.9	1.7	2.0	2.9
Ser	10.4	10.5	10.9	6.9
Thr	7.9	8.7	8.4	6.1
Trp	1.3	1.6	0.7	1.6
Tyr	8.2	10.8	6.1	5.6
Val	2.6	2.1	2.2	3.0
Total	100.1	100.1	99.9	99.9

Percentage composition of protein residues forming hydrogen bonds (identified using the program HBPLUS[33]) with water. Values in each column add up to 100.
[a]Values derived from the data in Table 4 of Thanki et al.[3] on the distribution of water around amino acid residues. The percentage composition was obtained by dividing the IntRes value for each residue type by the sum over all types.

DISCUSSION

The data presented here confirm that immobilized solvent is a general feature of protein–protein interfaces, and one that is widely observed in X-ray structures. For protein–protein complexes, the average number of 20 waters per interface that are reported by crystallographers amounts to 10 molecules per 1000 Å^2 of interface area. There are more waters at homodimer interfaces, but their surface density is about the same. Crystal packing interfaces, which are more polar in their chemical composition, immobilize 50% more water per unit area. Because water molecules tend to be under-reported in crystallographic studies, these numbers certainly under-rate the residual hydration of protein–protein interfaces. They refer to a data set that contain medium-resolution (2.6 Å) data, and the reality may be closer to what we observe in high (1.6 Å) resolution structures, where the interfaces contain 44% more immobilized water. This bias, which affects all our data, is likely to be the same in protein–protein complexes, homodimers, and crystal packing interfaces.

Albeit underestimated, the solvent that remains bound at the interface when the subunits of a protein–protein complex or oligomeric protein assemble represents only

about 10–15% of the solvent that is removed upon assembly. This ratio is calculated by assuming that a water molecule covers about 10 Å^2 of the protein ASA, like the waters that remain at the interface. In other terms, the average protein–protein interface is 85–90% dehydrated, justifying the assumption that the hydrophobic effect is a major contributor to the stability of protein–protein assemblies in general.[36–39] The polar interactions formed by immobilized water molecules may also stabilize the assembly. We find many water-mediated interactions at protein–protein interfaces. They are as abundant as direct hydrogen bonds between protein groups, and possibly more abundant if we correct for under-reporting. Moreover, these interactions are similar to the ones made by water bound at other sites of the protein surface: The same chemical groups and the same types of amino acid residues are involved.

Our analysis of the PDB data indicates that there are large differences between individual interfaces in their degree of hydration and the way immobilized solvent is distributed around or across an interface. The heterogeneity exists within each category—complexes, homodimers, and crystal packing—and it cannot be entirely attributed to inconsistent reporting by crystallographers. The number of waters at interfaces of similar size and polar character is highly variable. There is a "standard size" for the interfaces of protein–protein complexes. For instance, most antigen–antibody or enzyme–inhibitor complexes, and many other complexes, have B in the range 1200–2000 Å^2.[11] An example is the elastase–ovomucoid inhibitor (1ppf) of Figure 3. Thirty of the 46 protein–protein complexes in Table I have standard-size interfaces. These interfaces contain between 1 and 34 water molecules; those in structures determined at 2 Å resolution or better (16 entries) contain between 8 and 34. Thus, there is no standard level of hydration. Moreover, the patterns of hydration vary. One-half of the standard-size interfaces have a distance ratio $d_r > 1.2$, suggesting that they are mostly dry, but the range of d_r in that sample is 0.9–1.5, and there are some typical wet interfaces among them. The concept of a standard-size interface does not apply to homodimers. Only 20% of those in Table I have B in the range 1200–2000 Å^2; all others are larger, and often much larger.[10,14] Even though a majority is dry, their level of hydration is highly variable, and the distribution of d_r is just as wide in homodimers as in complexes.

Most pairwise interfaces in protein crystals are small and contain too few protein atoms or water molecules for significant statistics.[15] This is why we selected packing interfaces with $B > 800$ Å^2, expecting the smaller interfaces to obey the same rules.[20] Among those of Table I, a majority (55%) has B in the range 1200–2000 Å^2. They are even more heterogeneous than protein–protein interfaces in terms of hydration. Those in PDB entries at 2 Å-resolution or better contain between 1 and 74 interface water molecules. Although many are wet and very few have a $d_r > 1.2$, the whole range of d_r values is represented among them. It should be noted that in addition to being smaller than specific interfaces, crystal packing interfaces bury fewer protein atoms and are often split into many small patches,[20] in which case the d_r parameter is a poor representation of their mode of hydration.

The role of interface solvent in promoting the stability and specificity of recognition has been approached experimentally in a number of cases. Biochemical methods have been used to assess the contribution of polar interactions to the free enthalpy of dissociation ΔG_d, which measures the stability of a protein assembly, in a number of protein–protein complexes. Antigen–antibody,[40–47] enzyme–inhibitor,[48–51] hormone–receptor,[52,53] signal transduction interfaces,[54] and many others have been subjected to the systematic site-directed mutagenesis of interface residues followed by careful measurements of ΔG_d. The contributions of individual protein side-chains, of pairs of interacting side-chains, and in specific cases, of solvent molecules, have been derived from the data. This large body of experimental data is often used in theoretical studies,[55–58] but as only side-chains are accessible to site-directed mutagenesis, we have to assume that the interactions made by the main-chain and by nonprotein components, including solvent, behave like side-chain interactions. Moreover, most of the data are on protein–protein complexes. There are relatively few quantitative studies of subunit association in oligomeric proteins,[24,59–64] and none of a crystal packing interface. Although the PDB contains many examples of surface mutations that change crystal packing, this cannot be analyzed at thermodynamic equilibrium. Nevertheless, conclusions drawn from the study of complexes should apply to other types of protein–protein interfaces. In many cases, polar interactions are neutral in terms of free energy: ΔG_d is about the same when a hydrogen bond or salt bridge is removed by changing both residues involved to alanine. Exceptions are "hot-spots," which make significant contributions (2–4 kcal $\cdot$ mol^{-1}) to ΔG_d. Hot-spots tend to be located in the buried core of the interface.[65] If we assume that this extends to immobilized water molecules, the 71% of the interface waters that are in contact with bulk solvent are unlikely to contribute much to the stability of the assembly, but those that are buried at the interface may do so.

Our data substantiate the "O-ring model" of protein–protein interfaces derived from these studies,[65] but only in part. In its simpler form, this model fits the case of a single-patch interface with a well-defined core from which water is excluded and where hydrophobic and polar interactions coexist. The hydrophobic effect contributes to the stability of the assembly, but not its specificity, whereas polar interactions between the dehydrated protein groups of the core contribute to both. Our sample of specific interfaces offers examples of that situation, especially in complexes with standard-size interfaces. However, it also contains multipatch interfaces, and wet interfaces where water-mediated interactions are abundant and few protein groups are fully dehydrated. Some of these interfaces occur in systems that are just as stable and specific than those that follow the O-ring model. On the other hand, we find that crystal packing generally involves pairwise interfaces that are patchy, poorly packed, and more hydrated

than in specific assemblies. The stability of the crystal results from the implication of the protein molecule in several such interfaces, a situation that is reproduced in biological assemblies that involve a large number of protein molecules instead of just two, as in binary complexes and homodimers. There are many examples of multimolecular assemblies in cells and living organisms. Some are now represented in the PDB, and an analysis of their interface properties should be of great interest.

REFERENCES

1. Huber R, Kukla D, Bode W, Schwager P, Bartels K, Deisenhofer J, Steigemann W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. J Mol Biol 1974;89:73–101.
2. Janin J, Chothia C. Stability and specificity of protein–protein interactions: the case of the trypsin–trypsin inhibitor complexes. J Mol Biol 1976;100:197–211.
3. Thanki N, Thornton JM, Goodfellow JM. Distributions of water around amino acid residues in proteins. J Mol Biol 1988;202:637–657.
4. Thanki N, Thornton JM, Goodfellow JM. Influence of secondary structure on the hydration of serine, threonine and tyrosine residues in proteins. Protein Eng 1990;3:495–508.
5. Thanki N, Umrania Y, Thornton JM, Goodfellow JM. Analysis of protein main-chain solvation as a function of secondary structure. J Mol Biol 1991;221:669–691.
6. Robert CH, Ho PS. Significance of bound water to local chain conformations in protein crystals. Proc Natl Acad Sci USA 1995;92:7600–7604.
7. Petukhov M, Rychkov G, Firsov L, Serrano L. H-bonding in proteins revisited. Protein Sci 2004;13:2120–2129.
8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
9. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
10. Jones S, Thornton JM. Protein–protein interactions: a survey of protein dimer structures. Prog Biophys Mol Biol 1995;63:31–65.
11. Lo Conte L, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. J Mol Biol 1999;285:2177–2198.
12. Wodak SJ, Janin J. Structural basis for molecular recognition. Adv Prot Chem 2002;61:9–73.
13. Chakrabarti P, Janin J. Dissecting protein–protein recognition sites. Proteins 2002;47:334–343.
14. Bahadur RP, Chakrabarti P, Rodier F, Janin J. Dissecting subunit interfaces in homodimeric proteins. Proteins 2003;53:708–719.
15. Janin J, Rodier F. Protein–protein interaction at crystal contacts. Proteins 1995;23:580–587.
16. Janin J. Specific vs non-specific contacts in protein crystals. Nature Struct Biol 1997;4:973–974.
17. Carugo O, Argos P. Protein–protein crystal-packing contacts. Protein Sci 1997;6:2261–2263.
18. Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. Proteins 1997;28:494–514.
19. Ponstingl H, Henrick K, Thornton JM. Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins 2000;41:47–57.
20. Bahadur RP, Chakrabarti P, Rodier F, Janin J. A dissection of specific and non-specific protein–protein interfaces. J Mol Biol 2004;336:943–955.
21. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
22. Nicholls A, Sharp K, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
23. Lima CD, Klein MG. Hendrickson WA. Structure-based analysis of catalysis and substrate definition in the HIT protein family. Science 1997;278:286–290.
24. Irimia A, Ebel C, Madern D, Richard SB, Cosenza LW, Zaccaï G, Vellieux FMD. The oligomeric states of *Haloarcula marismortui* malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. J Mol Biol 2003;326:859–873.
25. Bode W, Wei AZ, Huber R, Meyer E, Travis J, Neumann S. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. EMBO J 1986;5:2453–2458.
26. Bompard-Gilles C, Rousseau P, Rouge P, Payan F. Substrate mimicry in the active center of a mammalian alpha-amylase: structural analysis of an enzyme–inhibitor complex. Structure 1996;4:1441–1452.
27. Krzywda S, Murshudov GN, Brzozowski AM, Jaskolski M, Scott EE, Klizas SA, Gibson QH, Olson JS, Wilkinson AJ. Stabilizing bound O2 in myoglobin by valine68 (E11) to asparagine substitution. Biochemistry 1998;37:15896–15907.
28. Hakansson M, Antonsson P, Bjork P, Svensson LA. Cooperative zinc binding in a sea mutant mimics the SEA-MHC class II interaction. J Biol Inorg Chem 2001;6:757–762.
29. Larsen TA, Olson AJ, Goodsell DS. Morphology of protein–protein interfaces. Structure 1998;6:421–427.
30. Janin J. Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. Structure 1999;7:R277–R279.
31. Hubbard SJ, Argos P. Cavities and packing at protein interfaces. Protein Sci 1994;3:2194–2206.
32. Parthasarathy S, Murthy MRN. Analysis of temperature factor distribution in high-resolution protein structures. Protein Sci 1997;6:2561–2567.
33. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
34. Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein–protein interfaces. Protein Eng 1997;10:999–1012.
35. Xu D, Lin SL, Nussinov R. Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. J Mol Biol 1997;265:68–84.
36. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–708.
37. Ben-Naim A. Solvent effects on protein association and protein folding. Biopolymers 1990;29:567–596.
38. Young L, Jernigan RL, Covell DG. A role for surface hydrophobicity in protein–protein recognition. Protein Sci 1994;3:717–729.
39. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Study of protein–protein interfaces: a statistical analysis of the hydrophobic effect. Protein Sci 1997;6:53–64.
40. Sundberg EJ, Mariuzza RA. Molecular recognition in antibody–antigen complexes. Adv Prot Chem 2002;61:119–160.
41. Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Acqua W, Souchon H, Schwarz FP, Mariuzza RA. Bound water molecules and conformational stabilization help mediate an antigen–antibody association. Proc Natl Acad Sci USA 1994;91:1089–1093.
42. Braden BC, Fields BA, Poljak RJ. Conservation of water molecules in an antibody–antigen interaction. J Mol Recognit 1995;8:317–325.
43. Ysern X, Fields BA, Bhat TN, Goldbaum FA, Dall'Acqua W, Schwarz FP, Poljak RJ, Mariuzza RA. Solvent rearrangement in an antigen–antibody interface introduced by site-directed mutagenesis of the antibody combining site. J Mol Biol 1994;238:498–500.
44. Fields BA, Goldbaum FA, Dall'Acqua W, Malchiodi EL, Cauerhff A, Schwarz FP, Ysern X, Poljak RJ, Mariuzza RA. Hydrogen bonding and solvent structure in an antigen–antibody interface: crystal structures and thermodynamic characterization of three fv mutants complexed with lysozyme. Biochemistry 1996;35:15494–15503.
45. Goldman ER, Dall'Acqua W, Braden BC, Mariuzza RA. Analysis of binding interactions in an idiotope–antiidiotope protein–protein complex by double mutant cycles. Biochemistry 1997;36:49–56.
46. Dall'Acqua W, Goldman ER, Lin W, Teng C, Tsuchiya D, Li H, Ysern X, Braden BC, Li Y, Smith-Gill SJ, Mariuzza RA. A mutational analysis of binding interactions in an antigen–antibody protein–protein complex. Biochemistry 1998;37:7981–7991.
47. Chacko D, Silverton E, Kam-Morgan L, Smith-Gill S, Cohen G, Davis D. Structure of an antibody–lysozyme complex unexpected effect of a conservative mutation. J Mol Biol 1995;245:261–274.

48. Huang K, Lu W, Anderson S, Laskowski MJ, James MNG. Water molecules participate in proteinase–inhibitor interactions: crystal structure of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. Protein Sci 1995;4:1985–1997.
49. Buckle AM, Schreiber G, Fersht AR. Protein–protein recognition: crystal structure analysis of a barnase–barstar complex at 2.0 Å resolution. Biochemistry 1994;33:8878–8889.
50. Schreiber G, Fersht AR. Energetics of protein–protein interactions: analysis of the barnase–barstar interface by single mutations and double mutant cycles. J Mol Biol 1995;248:478–486.
51. Kuhlmann UC, Pommer AJ, Moore GR, James R, Kleanthous C. Specificity in protein–protein interactions: the structural basis for dual recognition in endonuclease colicin–immunity protein complexes. J Mol Biol 2000;301:1163–1178.
52. Clackson T, Wells JA. A hot spot of binding energy in a hormone–receptor interface. Science 1995;267:383–386.
53. Wells JA. Binding in the growth hormone receptor complex. Proc Natl Acad Sci USA 1996;93:1–6.
54. Kiel C, Serrano L, Herrmann C. A detailed analysis of ras/effector complex interfaces. J Mol Biol 2004;340:1039–1058.
55. Covell DG, Wallqvist A. Analysis of protein–protein interactions and the effects of amino acid mutations on their energetics: the importance of water molecules in the binding epitope. J Mol Biol 1997;269:281–297.
56. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 2002;320:369–387.
57. Kortemme T, Baker D. A simple physical model for the binding energy hot spots in protein–protein complexes. Proc Natl Acad Sci USA 2002;99:14116–14121.
58. Papoian GA, Ulander J, Wolynes PG. Role of water-mediated interactions in protein–protein recognition landscapes. J Am Chem Soc 2003;125:9170–9178.
59. Borchert TV, Abagyan R, Kishan KVR, Zeelen JP, Wierenga RK. The crystal structure of an engineered monomeric triosephosphate isomerase, monoTIM: the correct modelling of a eight-residue loop. Structure 1993;1:205–213.
60. Bailey DL, Fraser ME, Bridger WA, James MN, Wolodko WT. A dimeric form of *Escherichia coli* succinyl-CoA synthetase produced by site-directed mutagenesis. J Mol Biol 1999;285:1655–1666.
61. Kelly CA, Nishiyama M, Ohnishi Y, Beppu T, Birktoft JJ. Determinants of protein thermostability observed in the 1.9-Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus*. Biochemistry 1993;32:3913–3922.
62. Miller MD, Krause KL. Identification of the *Serratia* endonuclease dimer: structural basis and implications for catalysis. Protein Sci 1996;5:24–33.
63. Ottosen MB, Björnberg O, Nørager S, Larsen S, Palfey BA, Jensen KF. The dimeric dihydroorotate dehydrogenase A from *Lactococcus lactis* dissociates reversibly into inactive monomers. Protein Sci 2002;11:2575–2583.
64. Mustat G, Briggs JM. Cluster analysis of water molecules in alanine racemase and their putative structural role. Protein Eng 2004;17:223–234.
65. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.