On the Contribution of Water-Mediated Interactions to Protein-Complex Stability[†]

Dana Reichmann, Yael Phillip, Asaf Carmi, and Gideon Schreiber*

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, 76100, Israel

Received September 30, 2007; Revised Manuscript Received November 6, 2007

ABSTRACT: Protein−water interactions have long been recognized as a major determinant of chain folding, conformational stability, binding specificity and catalysis. However, the detailed effects of water on stabilizing protein−protein interactions remain elusive. A way to test experimentally the contribution of water-mediated interactions is by applying double mutant cycle analysis on pairs of residues that do not form direct interactions, but are bridged by water. Seven such interactions within the interface between TEM1 and BLIP proteins were evaluated. No significant interaction free energy was found between either of them. Water can bridge interactions, but also stabilize the structure of the monomer. To distinguish between these, we performed a bioinformatic analysis using AQUAPROT (http://bioinfo.weizmann.ac.il/aquaprot) to determine the degree of water conservation between the bound and unbound states. 29 structures of twelve complexes and 20 related monomers were analyzed. Of the 262 water molecules located within the interfaces, 145 were conserved between the unbound and bound structures. Strikingly, all 50 buried or partially buried waters in the monomer structures were conserved at the same location in the bound structures. Thus, buried waters have an important role in stabilizing the monomer fold rather than contributing to protein−protein binding, and are not replaced by residues from the incoming protein. Taking together the experimental and bioinformatics evidence suggests that exposed waters within the interface may be good sites for protein engineering, while buried or mostly buried waters should be left unchanged.

Water−protein interactions have long been recognized as a major determinant of chain folding, conformational stability, binding specificity and catalysis (*1*, *2*). There are a variety of experimental and theoretical studies acknowledging the active role of solvent in protein stability and dynamics (*3*, *4*). Experimentally, X-ray, NMR (*5*, *6*) and femtosecond fluorescence (*7*) measurements reveal the binding sites, structure, and dynamics of water. Water molecules satisfy hydrogen bonds by direct interactions with protein backbone and side chains, fill cavities (*8*, *9*) and mediate interactions with ligands, nucleic acids and proteins (*10*−*14*). The much longer mean residence time for buried waters compared to other first hydration shell water (10−1000 ns in cavities vs 500 ps in first hydration shell) suggests that buried waters should be considered as an integral part of the tertiary structure (*3*).

Protein crystals, which normally contain substantial amounts of waters (*15*), show a wide range of nonrandom hydrogen-bonding environments (*3*). The location of many of these waters is conserved in the same positions in multiple crystals of the same or homologous proteins (*16*, *17*). These and other studies have shown that spatial conservation of water molecules is common in enzyme active sites, as well as in polar cavities inside the proteins (*18*) or on the protein surface. Furthermore, the analysis and prediction of conservation versus displacement of water upon ligand binding showed a relation between the coordination of water binding and its conservation (*19*, *20*).

Water is also considered an integral part of protein−protein interfaces. On average, about 10 water molecules are found per 1000 Å^2 of interface area, albeit this number varies greatly depending on the type of protein−protein interaction. In general, binding interfaces in homodimers are more solvated than interfaces of heterodimers (*12*). Comparing the architecture and chemical composition of water−protein interactions within specific protein−protein complexes versus crystal packing has shown that the latter is ∼50% more solvated (*12*). Moreover, interfaces of weak and highly transient complexes, such as electron-transfer proteins, contain more water than found in high-affinity complexes. In both crystal packing and weak transient complexes, water permeates the interface, replacing more preferable inter-residue interactions as found in higher affinity protein−protein binding sites.

Water-mediated interactions are believed to be a main reason for the water conservation phenomenon, and their inclusion into theoretical models improved folding predictions (*21*, *22*). Jiang et al. (*23*) developed a method to model water molecules within protein−protein interfaces, which successfully predicted bound water molecules at protein interfaces. However, inclusion of water did not improve prediction of either the free energy of binding or the free energy difference induced by mutations. While intensive theoretical work was done to estimate the importance of water-mediated interactions, little direct experimental evidence exists to support the energetic contribution of specific water molecules toward the affinity of protein complexes. A good approach to study the contributions of specific interactions toward binding and stability is the use of double-mutant cycles (DMC[1]), where interactions between amino

[†] This research was funded by MINERVA Grant 8525 and the Ministry of Science and Technology (MOST) Grant 0263.

* Corresponding author. E-mail: gideon.schreiber@weizmann.ac.il. Phone: +972-8-9343249. Fax: +972-8-9346095.

10.1021/bi7019639 CCC: $40.75
Published on Web 12/28/2007

acids are treated within their native context (*24*−*26*). In such cycles the sum of free energy loss by the two single mutations is compared to that of the double mutation, and when it is greater the residues are defined as interacting. DMC was applied in two studies of protein stability for the analysis of water-mediated interactions, and provided contradictory results (*27*, *28*). Water-mediated interactions within protein−protein interfaces were studied by Ikura et al., who fitted the change in free energy of binding of mutant proteins to the number of lost H-bonds to water, and found on average a contribution of 4 kJ/mol per water−protein interaction (*29*). In the current work we present a direct and extensive assessment of the energetic contribution of water-mediated interactions toward protein−protein binding. Here, we analyzed the energetic contributions of eight water-mediated hydrogen bonds toward protein−protein binding by using DMC. For the seven cases where no direct bonds were found between the residues, no interaction free energy was measured. However, DMC using values of k_{off} instead of K_A showed a weak bond for two of those interactions. A bioinformatic analysis of water conservation within the interfaces of 29 protein−protein complexes and 20 related monomers revealed that buried waters were conserved between the unbound and bound interfaces, suggesting that they are mainly important in stabilizing the folded conformation of the monomer.

RESULTS

Crystallographic Water in Bound versus Unbound Proteins. 10−15% of the interface is occupied by crystal water (*12*) with some of the water detected in the unbound state being found at the same position also in the bound state. As no good correlation was found between the percentage of the wet area and the binding affinity, one may assume that specific waters contribute to binding by forming through water H-bonds (*4*, *30*). In the first part of this study we analyzed the conservation of surface water upon complexation, followed by direct determination of the energetic contribution of through water H-bonds by DMC.

Water molecules are an integral part of X-ray protein structures, albeit their number is a function of the quality of the structure. To extract interface water we developed a new Web based tool, AQUAPROT (http://bioinfo.weizmann.ac.il/aquaprot). AQUAPROT retrieves a matrix containing all the interface interactions between residues from both proteins. Next, all water molecules within a distance of up to 3.5 Å from an interface polar atom were assigned. (Figure 1 and Materials and Methods). The binding site determined on the bound complexes was assigned to the unbound proteins and is called "unbound interface". Twelve different protein−protein complexes were analyzed for their water content using AQUAPROT (Table 1). This set of protein−protein complexes was derived from databases of Neuvirth et al. (*31*) and Mintseris et al. (*32*). For a protein complex to be analyzed it had to fulfill the three following criteria: availability of at least one unbound structure with identical sequence and similar resolution of <2.5 Å, identical structures of the unbound and bound proteins with no major conformational changes accompanying binding, and no cofactors within the binding sites. In the case where more than one copy of the protein exists within the asymmetric unit (as for TEM1−BLIP, barnase−barstar, CheY−CheA and the unbound structures of colicine E9 and barnase), each copy was analyzed independently, and the water molecules located in their binding sites were assigned. Moreover, four different complexes (TEM1−BLIP, trypsin−inhibitor, vitamin D−actin and colicine E9-IM9) and two monomers (trypsin inhibitor and cytochrome *c* peroxidase) have more than one PDB structure that were suitable for the water analysis. The final dataset consisted of 20 unbound structures and 29 complexes, which are related to twelve different protein−protein interactions (Table 1 and Table 1 in Supporting Information). The dataset includes complexes from different types, such as enzyme inhibitors, signal proteins and very weak transient complex (cytochrome *c* peroxidase).

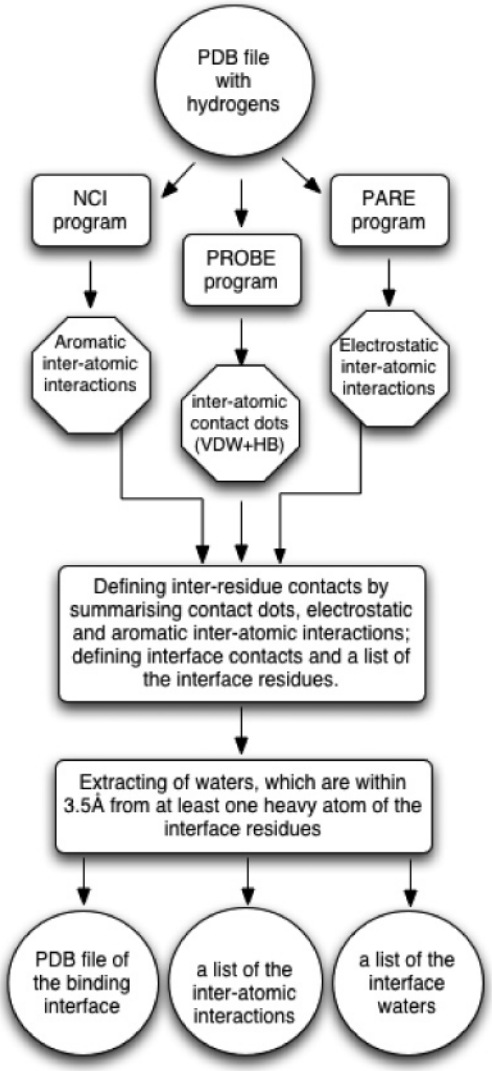

FIGURE 1: A flow diagram of the AQUAPROT server used to produce interface contact maps and to analyze interface waters. HB is for hydrogen bonds, VDW is for van der Waals interactions.

Conservation of Interface Water between the Unbound and Bound State. On average 70% of the binding interface has a high degree of shape complementarity, with crystal water occupying 10−15% of the interface (*12*, *33*). Therefore, one may assume that the incoming protein will replace this percentage of water found in the unbound state. Actually,

[1] Abbreviations: TEM1, TEM1-β-lactamase; BLIP, β-lactamase inhibitor protein; *mc*, backbone; *sc*, side chain; DMC, double-mutant cycle; ASA, accessible surface area.

Table 1: Dataset of the Protein−Protein Complexes

protein name	PDB ID and chains	resolution (Å)	no. of waters in interface[a]	% of interface waters from all protein water[b]	conservation (unbound vs bound)[c] (%)
		A. Monomers			
TEM1	1BTL	1.8	23	12	68
BLIP	BLIP[d]	2.1	8	7	75
trypsin	2PTN	1.5	12	3	83
trypsin inhibitor	4PTI	1.5	10	17	38
	1BPI	1.1	20	12	31
barnase	1B20_B	1.7	15	11	47
	1B20_C		20	9	76
actin α1	1J6Z	1.54	47	23	48
colicine E9	1FSJ_B	1.8	18	17	43
	1FSJ_C		16	15	90
	1FSJ_D		16	15	62
	1FSJ_E		9	9	88
α-amylase	1JAE	1.65	27	11	80
p67 phox	1HH8	1.8	13	9.4	45
cytochrome *c*	1JCI	1.9	16	4	58
peroxidase	1DJ1	1.93	11	3	78
serine proteinase	1SCA	2.0	9	2	22
chemotaxis proteins CheY	1E6K	2.0	3	7.5	100
		B. Complexes			
TEM1−BLIP	1JTG_AB	1.73	44	19	
	1JTG_CD		47	21	
	1SOW_AC	1.9	32	13	
	1SOW_BD		37	15	
	1XXM_AC	1.9	32	15	
	1XXM_BD		32	15	
	2B5R_AB	1.65	44	14	
	2B5R_CD		42	13	
trypsin−inhibitor	2TGP	1.9	27	6	
	1TPA	1.9	22	10	
	2PTC	1.9	21	10	
	3TPI	1.9	21	10	
trypsin−pancreatic secretory inhibitor	1TGS	1.8	22	6	
barstar−barnase	1B2S_AD	1.8	39	19	
	1B2S_BE		37	19	
	1B2S_CF		21	11	
vitamin D binding protein−actin	1KXP	2.1	73	19	
	1LOT	2.5	22	25	
	1MA9	2.4	58	19	
colicine E9−cognate immunity protein IM9	1BXI	2.05	23	22	
	1EMV	1.7	27	18	
	1FR2	1.6	29	17	
α-amylase−RAGI inhibitor	1TMQ	2.5	34	13	
α-amylase−AAI inhibitor	1CLV	2.0	50	20	
Rac GTApase−p67 phox	1E96	2.4	13	19	
cytochrome *c*−cytochrome *c* peroxidase	1S6V	1.88	21	7	
serine proteinase−inhibitor	1CSE	1.2	22	5	
chemotaxis proteins	1FFG_AB	2.1	16	14	
CheY−CheA	1FFG_CD		9	5	

[a] Number of water molecules found in the interface and making polar contact with interface residues, as described in Materials and Methods. [b] Percent of waters located in the interface out of the total number of water molecules interacting with protein residues. Interactions of waters with the protein were defined by the same method as for the interface water. [c] The conservation between waters interacting with the binding site residues found in unbound and bound proteins. The conservation score was calculated by eq 2 (Materials and Methods). [d] The structure of the BLIP protein was obtained from (*47*).

the percentage of replaced water may be even higher, as crystal water indicates preferable locations for H-bonds. The conservation of crystal waters upon protein−protein binding was evaluated by aligning the water molecules and their related residues in the unbound and bound structures (Figure 2A and Table 1). As the resolutions of the bound and unbound structures were similar, water conservation scores are comparable (Table 1). One should remember that the lack of conservation of crystallographic water might also be a result of poor water assignment in the crystal structure of the bound state, and does not have to show a real lack of conservation. Table 2 in the Supporting Information provides the detailed conservation scores as calculated from eq 1. Two water molecules are considered as conserved only when they make the same polar contacts with protein atoms and are located in the same location in the interface (up to 0.5 Å). Waters bound to interface residues with significantly different side chain positions between unbound and bound structures were ignored in the conservation analysis. Table 1 clearly shows that interface waters found in unbound proteins were most often (43−100%) found also in the complex (see conservation score, eq 2 in Materials and Methods). Outliers are complexes of bovine trypsin−inhibitor and serine protease−inhibitor with conservation scores of 31−38% and

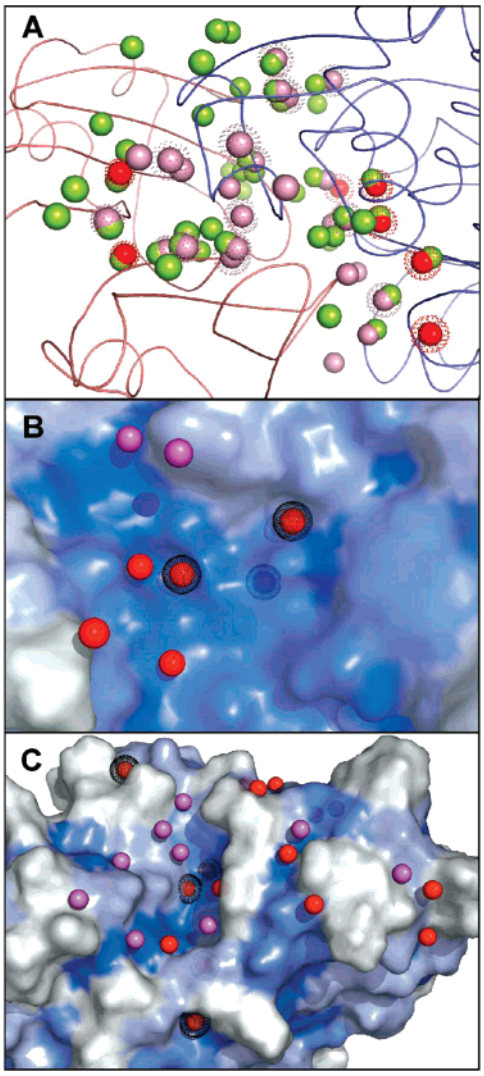

FIGURE 2: Crystal waters within the TEM1−BLIP interface. (A) An overlay of the complex waters (green) and water found in the unbound structures of TEM1 and BLIP (pink). The conserved water molecules are marked by dotted presentation, while buried waters are in red. Ribbons of BLIP and TEM1 are in pink and blue, respectively. The surface of the BLIP (B) and TEM1 (C) proteins are colored by an accessibility gradient. The conserved waters are in red, and nonconserved waters are in pink. Water molecules with relative ASA less than 16% are shown in dotted presentation.

22%, respectively. This may be due to the tight fit needed for this complex to avoid chain degradation, emphasizing the energetic advantage of the intra-protein interactions relatively to water-mediated interactions. The actual conservation percent may be even higher, as water molecules may be missing from the bound structures, which were solved at lower resolutions compared to the unbound structures (Table 1). A detailed analysis of the TEM1−BLIP interface waters showed that the conservation defined only by specific protein−water contacts, and not by location, is even higher. Of the 31 water molecules found in the interfaces of the unbound TEM1 and BLIP structures, 22 were conserved in the complex, 3 water molecules from one protein (TEM1) were replaced upon complexation by residues of the other protein (BLIP) and for 3 waters no density was found in the complex. The remaining 3 water molecules were replaced by water at different positions, but binding the same residues (Table 3 Supporting Information). Two examples for water relocation upon complexation, in conjunction with their adjacent side chains, are shown in Figure 3A and 3B. Figure 3C shows a case where the water molecule in the bound complex moved to a new location, although the side chain did not move. In both cases the same protein atoms are hydrated in the unbound and bound states. Still, they were considered as nonconserved in the above analysis, further increasing the actual conservation degree above the values given in Table 1.

In addition to the twelve complexes presented in Table 1, the trimeric complex of lysozyme (HEL) with light and heavy chains of the D1.3 antibody was analyzed for its water conservation. The PDB ID of the complex is 1VFB; the unbound structures of the lysozyme proteins are 132L and 194L. 29 waters are located at the binding site of the unbound proteins; of those only one water molecule is buried and conserved (ASA of 2%). The two other conserved water molecules have ASA values of 8 and 43%. All other waters are nonconserved and exposed. Comparison between different complexes of D1.3−HEL showed no water conservation within the bound interfaces. Sunberg and Mariuzza (*34*) showed that binding of HEL to the D1.3 antibody involves flexibility and structural rearrangements, which may affect the water architecture within the interface. This structural flexibility is one of the main reasons that we did not further consider anitigen−antibody complexes for water conservation analysis.

Buried Waters Are Conserved between the Unbound and Bound States. Next, we wanted to relate conservation to ASA, taking into account crystal contact corrections as described in Materials and Methods. 262 water molecules were found in the interface of the unbound proteins. These were divided into two groups: conserved (145 molecules) and nonconserved (117 molecules). Water was defined as conserved if it is found in at least one bound structure at the same position and interacting with the same atoms as in the unbound protein. Figures 4A and B show that all buried and mostly buried waters (up to ASA of 5%) are conserved between at least one unbound and bound structure. Thus, the incoming protein does not replace buried waters located within the binding site. For the water molecules exposed to the surrounding solvent, no preference for either conservation in the bound state or replacement with an incoming amino acid was found. This is seen by the similar bar height in Figure 4A for water exposed over 10% (ASA), as well as in Figure 4B by the parallel trends of the conserved versus nonconserved waters for over 10% exposure. In the case of the TEM1−BLIP complex, which has the largest number of analyzed structures, the accessible area of the nonconserved water molecules is in all cases >16%. The given analysis suggests that many of the buried water molecules found in the interface contribute to the stability of the monomer fold, rather than to binding.

Analyzing the Physical Parameters of Water−Protein Interactions. The distances between the water oxygen and the protein polar atoms are shown in Figure 5. A separate analysis of hydrogen bonds to backbone (*mc*) and side chain (*sc*) atoms shows a similar distribution of *mc*-HOH interactions for the unbound and bound states, with a maximum at 2.9 Å (Figure 5A). The distribution of *sc*-HOH (Figure 5B) shows an optimal distance of 2.8 Å for the bound versus 2.9 Å for the unbound interfaces. These values are similar

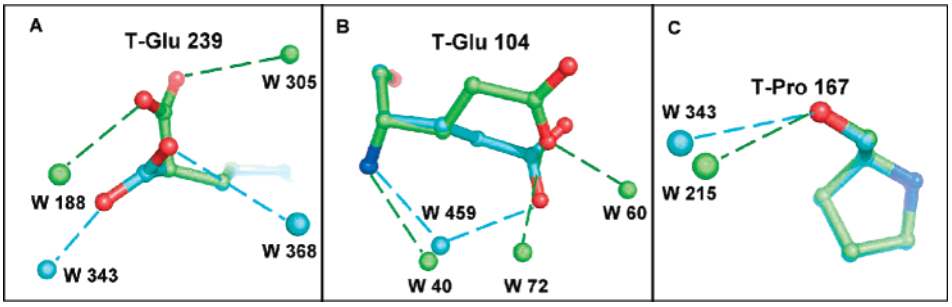

FIGURE 3: Examples for water relocation upon complexation. Panels A and B show the water molecules relocated upon changes in rotamers of the interacting residues Glu 239 (A) and Glu 104 (B) in the TEM1 protein. In panel A the distances between nonconserved waters are 4.63 Å (between W188 and W343) and 5.44 Å (between W305 and W368). In panel B the distances between unbound water W459 and three related bound waters, W40, W72 and W60, are 0.55 Å, 2.3 Å and 5.3 Å, respectively. (C) The water molecule in the bound complex moved to a new location, although the side chain did not move. The distance between the water molecules 343 and 215 is 0.8 Å. Residues and water molecules drawn from the structure of the complex are in green, and from the unbound structures are in cyan.

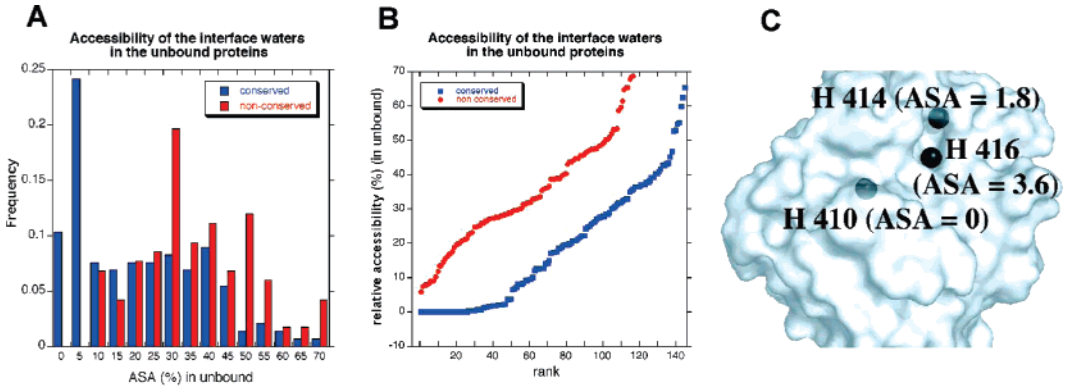

FIGURE 4: Correlation between water conservation and accessible surface area (ASA) of the waters in the unbound state, taking into account crystal contact corrections as described in the methods section. (A) Histogram of the relative accessibility values for conserved and nonconserved interface water molecules. (B) Plot of the relative accessibility of the conserved and nonconserved interface waters ordered by their relative accessibility (normalized by the number of conserved and nonconserved waters respectively). (C) An example of buried and partially buried water molecules interacting with the interface residues of serine proteinase (PDB ID 2ptn).

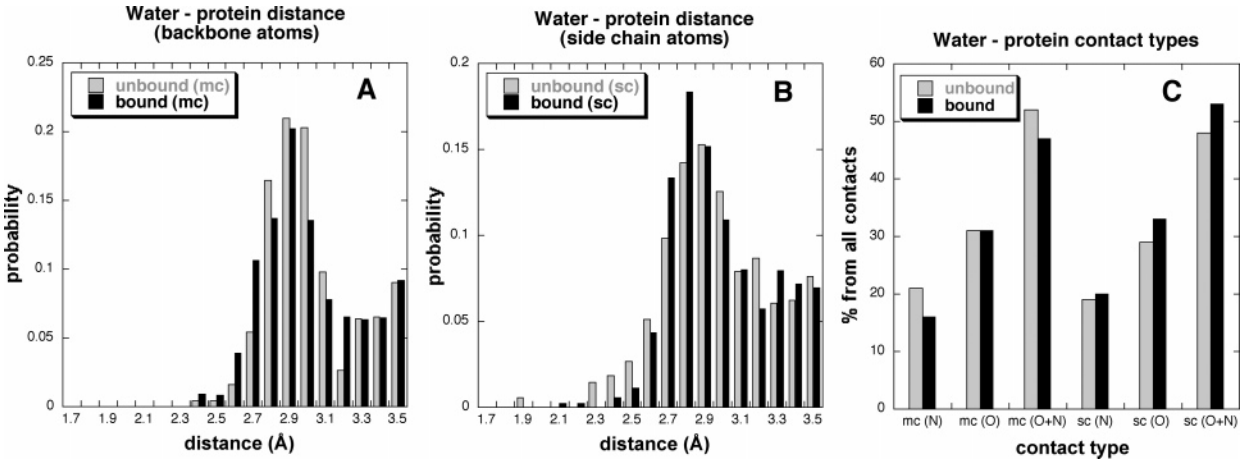

FIGURE 5: Histogram of the distances between the water oxygen and the protein heavy atoms of the interface residues in the bound and unbound structures. A separate analysis of hydrogen bonds to backbone (*mc*) and side chain (*sc*) atoms is shown in (A) and (B), respectively. (C) The distribution of the H-bonds of *sc*-HOH and *mc*-HOH in the bound and unbound interfaces. The hydrogen bonds between the water oxygen and the protein nitrogen atom are marked by N (sc(N) and mc(N)), and with the oxygen atom are signed by O (sc(O) and mc(O)). The distributions were corrected for the differences in volume elements of the bins by $(\text{distance})^2$.

to those observed around polar groups in high-resolution protein structures (*23*), and to inter-protein hydrogen bonds (*35*). Interestingly, there is a minor preference of water to interact with *mc* polar atoms in the unbound state while in bound interfaces the preference is opposite (Figure 5C). Similarly, Park showed that buried waters tend to bind *mc* polar atoms of amino acids in globular proteins (*8*). An interesting outcome of the current analysis is that on average a water molecule makes ~1.4 hydrogen bonds with polar atoms, independent of whether the unbound or bound structure was analyzed. Thus, no new interactions are formed between water molecules and the protein upon complexation.

Double-Mutant Cycle Analysis of Water-Mediated Interactions. The high degree of water conservation between the unbound and bound states of the proteins, as well as the similar number of H-bonds per water molecule in both states,

Table 2: Coupling Binding Energies of Water-Mediated Interactions[a]

	TEM residue	BLIP residue	water no. (1JTG)	distance between residues (Å)	$\Delta\Delta G_{k_{off}(int)}$ (kJ/mol)	$\Delta\Delta G_{K_A(int)}$ (kJ/mol)	water no. in unbound	ASA (%) bound	ASA (%) unbound	conservation (%) unbound vs bound	conservation (%) within bound
1	Glu110	Ser71	144	2.6		1.2	408	7.3	24	50	40
2[b]	Arg243	Asp49	154	2.8	−5.6	−8.8	405	6.5	27	50	40
3	Glu168	Arg160	84	5.8	−2.67	−1.31	379	5.5	35	83	80
4	Glu168	Trp162	84	4.3	−1.9	−1.25	379	5.5	35	83	80
5	Glu104	Ser146	72	5.6	1.06	1.45		2.6			20[c]
6	Ser124	Ser35	91	4.8	−0.03	0.9		3.1			100
7	Thr128	Ser35	91	5.2	0.19	−0.65		3.1			100
8	Gln99	Ser128	111	5.0	0.14	0.79		4.3			100

[a] Binding energies (in kJ/mol) were measured with TEM1 immobilized to the sensor chip. $\Delta\Delta G_{k_{off}(int)}$ is for interaction energies, calculated using double-mutant cycles based on k_{off} ($2 \times SE = 1.2$ kJ/mol). $\Delta\Delta G_{K_A(int)}$ is for interaction energies, calculated using double-mutant cycles based on K_A ($2 \times SE = 2.4$ kJ/mol). Conservation between unbound and bound was calculated by eqs 1 and 2, respectively, as described in Materials and Methods. Conservation of 100% indicates that this specific water is found in other 6 structures of TEM1−BLIP complex in the same position and creates the same contacts as in 1JTG_AB. Conservation within the bound complexes was calculated by overlay between 1JTG_AB and 5 other TEM1−BLIP structures. Structure 2B5R was not used for this analysis due to conformational changes in the analyzed interface area (*36*). [b] Coupling energies obtained from (*24*). [c] Residue Glu104 is mutated to Ala in 1XXM structure and has non wild-type position in 1SOW structure, therefore water bound to Glu104 is not conserved. However, this water molecule is conserved in two copies of the wild-type structure.

shows that the desolvation of unbound interface waters is only partial, and skips completely over the more buried waters. This may be interpreted as a thermodynamic advantage for having interface water. Despite the importance of this question, very little experimental work was done to directly address it. One way of measuring the interaction free energy of water-mediated interactions is by performing a double-mutant cycle between the involved residues. First we identified all crystal waters that bridge side chains from TEM1 and BLIP using AQUAPROT, allocating twelve inter-protein water-mediated *sc*−*sc* interactions (Table 4 in the Supporting Information). This number is similar to that found for a dataset of 161 interfaces (*12*). Five of the water molecules mediating TEM1 and BLIP were also found in the unbound structures with conservation scores of 50−100% between the unbound and bound states (Table 2). The surface exposure of these waters in the unbound state was among the highest for conserved water molecules (24−39%).

Eight of the twelve water-mediated inter-protein interactions in the interface of TEM1 and BLIP were experimentally analyzed using DMC (Table 2). The remaining four interactions involve residues that are binding hotspots, resulting in too weak affinities to be accurately determined. Figure 6 and Movie 1 in the Supporting Information show the structural details of the analyzed interactions (superscript B is for BLIP and superscript T for TEM1). The distances between all these residues, except ^{T}E110−^{B}S71 and ^{T}R243−^{B}D49, are above that of a direct H-bond distance, making them good representatives of water-mediated interactions (*23*). The DMC free energies ($\Delta\Delta G_{K_A(int)}$) for all the measured water-mediated interactions, except for ^{T}R243A−^{B}D49A, were between −1.31 and 1.45 kJ/mol. As the experimental error of the double-mutant cycles as determined from values of K_A ($\Delta\Delta G_{K_A(int)}$) was 2.4 kJ/mol (for two SE), all measured binding free energy values are not significantly different from zero. The average $\Delta\Delta G_{K_A(int)}$ value for the seven water-mediated H-bonds was 0.16 kJ/mol (positive value is for repulsion), which is within the experimental error for this experiment. Performing the same cycles using only the dissociation rate constants (k_{off}) showed that four cycles (^{T}E104−^{B}S146, ^{T}S124−^{B}S35, ^{T}T128−^{B}S35 and ^{T}Q99−^{B}S128) were additive, while two interactions bridged by the same water molecule w84 (^{T}E168−^{B}R160 and ^{T}E168−^{B}W162) were nonadditive ($\Delta\Delta G_{k_{off}(int)}$ values were −2.67 kJ/mol and −1.9 kJ/mol, respectively, Table 2). The experimental error for double-mutant cycles evaluated from measurements of k_{off} ($\Delta\Delta G_{k_{off}(int)}$) is 1.2 kJ/mol (for two SE), and therefore these values are significant (albeit small). The only strong DMC free energy was measured between ^{B}D49 and ^{T}R243. However, these two residues are only 2.6 Å apart, sponsoring a direct salt bridge between these residues (Figure 6) (*24*). The waters analyzed by DMC are all well conserved between the six different TEM1−BLIP structures (40−100%). The high degree of conservation points toward the nonrandom positioning of the evaluated waters, indicating that their location is not a crystallographic artifact.

DISCUSSION

Water is abundantly found within the interfaces between proteins filling cavities and bridging residue−residue interactions. In this manuscript we aimed to determine the structural and energetical role of interface water by combining bioinformatic analysis and experimental measurements. Double-mutant cycles were first developed to determine the coupling free energy between pairs of residues. Thus, results of measurements of interaction free energies between residues separated by water can be related to the energetic contribution of these water-mediated bonds. DMC analysis of seven water-mediated hydrogen bonds between residues that are too far to form direct contacts within the TEM1−BLIP interface showed their interaction energy being either weak or within the experimental error of the measurements (Table 2 and Figure 6). The lack of strong contribution of water-mediated H-bonds toward protein−protein binding is different from the data presented by Ikura and Jang, but in line with the data of Langhorst (*27*−*29*), who did not observe a significant through water interaction for the one analyzed DMC. In the work of Jang et al. (*27*) an interaction energy of 5 kJ/mol was measured for a water-mediated bond. However, with one of the mutations being D99L and the cycle calculated between residues located in a highly buried and hydrophobic environment causing structural rearrangement, the interpretation of these results toward the direct

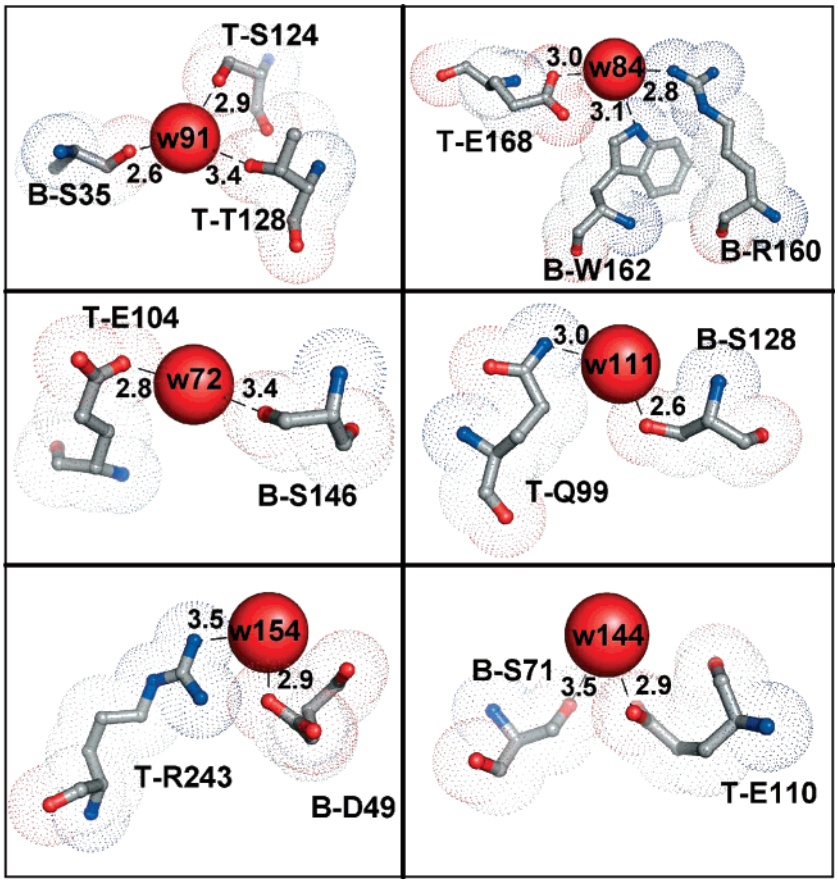

FIGURE 6: Eight *sc*−*sc* water-mediated inter-protein interactions in the interface of TEM1 and BLIP. First letter represents the protein (B-BLIP, T-TEM1). Water molecules bridging the protein residues are assigned as *W*, with the atom number from the TEM1−BLIP wt structure (PDB ID is 1JTG, chains A and B). Distances in the figure are between waters and active side chain atom.

contribution of the water-mediated interactions on protein stability may be problematic.

Interaction energies calculated from double-mutant cycles are not an exact measure of the bond strength, but rather a quantitative evaluation of the difference in binding energy between the two single mutations and the double mutant. Secondary effects of mutations as well as structural rearrangements are not accurately taken into account by this analysis. For example, the F142A mutation on BLIP makes an interaction with E104A on TEM1 of $\Delta\Delta G_{int}$ of −3.8 kJ/mol (*26*). This is despite the lack of a "good" physical contact between these residues. This apparent interaction is a result of the structural rearrangement accompanying the F142A mutation on E104 (*26*). Thus, one may claim that the lack of significant interaction energies measured for water-mediated H-bonds are due to second-site effects. However, in this work we analyzed the simplest case, where the two residues in question do not make direct contact with one another and no new interactions are expected to be formed between these residues upon mutation to alanine. Rather, the mutations simply increase the size of the cavity within the interface, replacing an interaction between a polar residue and water with a cavity that is expected to be filled with additional water (see Figure 6 for details). Moreover, the methyl group of the mutant alanine is hydrophobic and not expected to form new specific H-bonds with water.

How can we explain the differences in DMC values between $\Delta\Delta G_{k_{off}(int)}$ and $\Delta\Delta G_{K_A(int)}$? k_{off} is related to the strength of the short-range interactions between two proteins, while k_{on} is a measure of the rate of complex formation (K_A is the ratio k_{on}/k_{off}). We have previously shown that k_{off} is affected by short-range interactions, while k_{on} is mostly affected by long-range interactions (*24*). Therefore, the significant (albeit weak) DMC values for $\Delta\Delta G_{k_{off}(int)}$ determined for two of the water-mediated H-bonds may signify the existence of short range through water interactions, which are canceled out by other, less specific effects such as electrostatics, entropy etc. The overall net result of the $\Delta\Delta G_{K_A(int)}$ values is close to zero, with no significant contribution of water-mediated interactions toward complex stability.

It is interesting to note that weak water-mediated H-bonds (for $\Delta\Delta G_{k_{off}(int)}$) were observed only through water 84, bridging ^{T}E168 with ^{B}R160, ^{B}W162 and the backbone oxygen of ^{T}E166 (Table 2). Thus, w84 makes four H-bonds with protein, while all other analyzed waters make one or two H-bonds also with other adjacent water molecules (observed in the X-ray structure). It may be possible that the more confined environment of w84 contributes to this interaction. Still, most interface waters do not play a crucial role in the formalization of the final complex stability, but rather fill the holes within the interface, which are not occupied by amino acids. The question is asked whether this would mean that increasing the percentage of dry areas within an interface would increase the binding energy. The conservation of waters from the unbound to bound state points toward a nonrandom distribution of water in these structures, which allows one to derive general rules for water architec-

ture within binding interfaces, and sheds further light on the role of water in binding. Comparing the conservation of water upon binding shows that buried and mostly buried waters in the interface of the unbound proteins are always conserved upon complexation (Figure 4). This suggests that these water molecules are important in stabilizing the monomer fold, rather than contributing to the protein−protein interaction. These results are in line with a recent MD simulation using explicit water performed by us, where we saw that buried interface water has a lower entropy in respect to other surface water molecules (Lapid, H., Gottschalk, K., and G.S., unpublished). They also fit the findings of Amadasi and Barillari that water with higher coordinate number of H-bonds and tighter binding energy is less likely to be replaced by an incoming ligand (*19*, *20*). Examples of buried and partially buried waters are shown in Figures 2 and 4C. Although these waters are on the surface of the proteins and are partially exposed, they should be treated as the "twenty-first" amino acid (*3*). Thus, care should be taken when attempting to replace them with amino acids, which may decrease the stability of the monomer and thus increase the overall free energy of the system upon complexation. By contrast, exposed waters within the interface would be good candidates for replacement with specific amino acids to increase complex stability. This result may also explain why all interfaces of transient protein−protein complexes have also "wet" regions (in the order of 25−35%), as opposed to the protein core, which is "dry". The protein interface has a dual life, solvated in the unbound state and bound to a protein in the bound state. Protein surfaces are not flat, with specific waters occupying the polar cavities. These waters are important to the protein stability, and hence are conserved upon complexation. Within the core of proteins cavities filled with water are rare, and the unfolded state is basically unstructured and hence has no water containing cavities that are conserved upon folding.

Another take-home message from this study is that interface waters do not contribute much to the overall stability of the complex. This should make life simpler when creating a "good" force field, as one can basically ignore their contribution to binding (albeit not to protein stability). Indeed, current force fields work quite well without taking explicit water into account, and the attempt to add them did not result in much improvement in calculating the effects of mutations on binding.

MATERIALS AND METHODS

Deriving a Dataset of Unbound and Bound Structures. A set of protein−protein complexes and related monomers was derived from databases of Neuvirth et al. (*31*) and Mintseris et al. (*32*). The final set of complexes was selected according to the following criteria: availability of at least one unbound structure with identical sequence, being longer than 30 amino acids and similar resolution of <2.5 Å, identical structures of the unbound and bound proteins with no major conformational changes accompanying binding and no cofactors within the binding sites. In order to compare water occupancy within unbound and bound structures, resolution of the unbound and bound structures has to be similar. The major difference in the resolution (up to 1 Å) was found in two complexes, vitamin D−actin and α-amylase−RAGI inhibitor. Antigen−antibody complexes were not included in the dataset, since their binding involves structural flexibility (*34*) and their evolutionary process is significantly different from that of other proteins. In the case where more than one copy of the protein exists within the asymmetric unit (as for TEM1−BLIP, barnase−barstar, Chey−Chea and the unbound structures of colicine E9 and barnase), each copy was analyzed independently, and the water molecules located in their binding sites were assigned. Moreover, four different complexes (TEM1−BLIP, trypsin−inhibitor, vitamin D−actin and colicine E9−IM9) and two monomers (trypsin inhibitor and cytochrome *c* peroxidase) have more than one PDB structure suitable for water analysis. The final dataset consisted of 20 unbound structures and 29 complexes, which are related to twelve different protein−protein interactions (Table 1 and Table 1 in Supporting Information).

Assigning Interface Water Molecules. A flow diagram of the process of assigning interface water molecules as implemented in AQUAPROT (http://bioinfo.weizmann.ac.il/aquaprot) is presented in Figure 1. All water molecules within 3.5 Å from at least one interface polar atom (oxygen or nitrogen) of both chains were extracted, named "interface waters". The interface residues were defined by an approach based on the "all-atom contact" method, where the detailed atomic contacts between protein residues were calculated using modified REDUCE, PROBE, NCI and PARE programs (*36*). REDUCE (*37*) adds hydrogens to all heavy atoms, and allows side chain rearrangement for H-bond optimization. PROBE (*38*) defines interactions by rolling a ball with a radius of 0.25 Å along the van der Waals surface and defining the points of interatomic contacts. While this method gives an excellent quantitative description of H-bonds and van der Waals interactions, it misses some of the aromatic and electrostatic interactions because their distance and angle restrictions on physically meaningful bonds are not within the simple definitions of van der Waals distances probed by PROBE. Therefore, we added a modified version of the NCI (*39*) server to identify pi−pi and cation−pi interactions, and the PARE (*40*) program to identify charge−charge interactions. Once all interatomic contacts were listed, they were integrated into their respective amino acids and summarized in a matrix containing all the interface interactions between residues from both proteins. After definition of the interface residues, water molecules interacting with the interface residues were extracted. Only water molecules that form hydrogen bonds with the interface atoms were analyzed in this paper. Each water molecule can make up to four interactions, with the number and type of hydrogen bonds per water molecule being recorded. In cases where more than two donors and two acceptors were found, the closest four were chosen. We ignored water−aromatic interactions (*41*) and nonpolar interactions such as water−CH interactions. Still, the number of waters making only hydrophobic or aromatic contacts was very low (0−2 per interface in the unbound and bound sets). Hydrogens were not added to the water molecules, but were added to all other polar atoms in the proteins. The accessible surface area (ASA) of water molecules was calculated by a modified NACCESS 2.1.1 program (*42*), with the ASA of each analyzed water oxygen being calculated separately (probe size of 1.4 Å), taking into consideration only protein atoms. The relative accessible area was derived as a percentage from the maximum ASA value of an oxygen atom (98.5 Å). The

ASA values were corrected by accessible area occupied by crystal contacts and not by atom−atom interactions using the CryCo server (*43*) (http://ligin.weizmann.ac.il/~lpgerzon/cryco5.0/cryco/cryst1.cgi). In addition to water analysis, AQUAPROT provides interface contact maps and their analysis within the Mavisto program. The output files include a PDB file of the interface with related water molecules, a list of interatomic contacts within the proteins and between the interacting proteins and a PyMol script for visualization the binding interface with interface waters.

Conservation Analysis. The conservation between unbound and bound waters was analyzed by the overlay between the unbound structure and related complexes. Two water molecules were considered to be conserved if they are located at up to 0.5 Å from one another and form hydrogen bonds with the same atoms on the protein. To avoid inaccuracy in the global structural alignment due to small structural changes, the residues related to specific interface water were aligned locally, and conservation of this water molecule was analyzed. The conservation score for each interface water molecule in the bound proteins was calculated according to

$$S = (\mathrm{Nw}/\mathrm{Np}) \times 100 \quad (1)$$

where S is the conservation score, Nw is its occurrence in the bound state and Np stands for the number of aligned protein complexes. Waters bound to interface residues where the side chain position changed completely between unbound and bound structures were ignored. To evaluate a general conservation score of all interface residues found in the unbound state versus bound state (Table 1), the score was calculated by

$$S_t = (N_{cons}/N_{total}) \times 100 \quad (2)$$

where S_t is the total conservation score of all unbound waters versus waters in bound state, N_{cons} is number of waters conserved in at least one bound structure, and N_{total} is the total number of waters in unbound state. The conservation between waters found in different bound states was calculated by the same way, using one of the complexes as the reference (in case of TEM1−BLIP, chains A and B). The interface waters of other complexes were compared to the reference interface.

Site Directed Mutagenesis. Mutagenesis was performed using Kunkel mutagenesis (*44*) as described in (*36*). The insertion of the mutations was verified by DNA sequencing.

Protein Expression and Purification. Expression and purification of TEM1 and BLIP were carried out as previously described (*45*). Protein quality (degree of purity and activity) as well as its active concentration was determined by analytical gel filtration chromatography and its absorption in respect to its extinction coefficient.

Kinetic Measurements. Binding constants were determined by surface plasmon resonance (SPR) detection using the ProteOn XPR36 Protein Interaction Array System (BioRad) in PBS pH = 7.4 with 0.005% surfactant P20 at 25 °C as previously described (*36*, *46*). The change in free energy ($\Delta\Delta G_{K_A}$) upon mutation was calculated from

$$\Delta\Delta G_{K_A} = -RT \ln \frac{K_A^{mut}}{K_A^{wt}} \quad (3)$$

with K_A values being determined from $K_A = k_{on}/k_{off}$ and by following the refractive index (RU) change at equilibrium, fitting the data to the mass action expression [termed $K_{A(ma)}$],

$$\mathrm{RU} = [C \cdot K_{A(ma)} \cdot R_{max}]/[C \cdot K_{A(ma)} + 1] \quad (4)$$

where C represents the protein concentration. Values of $\Delta\Delta G_{K_A}$ and $\Delta\Delta G_{K_{A(ma)}}$ (determined from eq 3) are highly correlated (with a correlation coefficient of 0.99 (*36*)). Error analysis shows that significant values of $\Delta\Delta G$ (2SE) calculated from k_{off} were >0.7 kJ/mol and from K_A >1.4 kJ/mol.

Double Mutant Cycle Analysis. The interaction binding energy, $\Delta\Delta G_{int}$, for a pair of residues (1 and 2) was calculated from

$$\Delta\Delta G_{int} = \Delta\Delta G^{mut1mut2} - \Delta\Delta G^{mut1} - \Delta\Delta G^{mut2} \quad (5)$$

with all the single and double mutations being to alanine. A lack of interaction results in values close to zero, attractive interactions result in negative $\Delta\Delta G_{int}$ values, and repulsive interactions result in positive values. Only values of $\Delta\Delta G_{int}$ as calculated from K_A >2.4 kJ/mol were treated as significant.

ACKNOWLEDGMENT

We are grateful to Dr. Ora Schueler-Furman, Dr. Yaakov Levy and Eyal Kalie for their critical comments, stimulating discussions and editing this manuscript.

SUPPORTING INFORMATION AVAILABLE

Table 1 summarizes the analyzed dataset of 20 unbound structures and 29 complexes, which are related to twelve different protein−protein interactions. Conservation of crystallographic water molecules between unbound and bound structures of the TEM1−BLIP complex is shown in Table 2. Table 3 presents cases where the interface waters found in the unbound TEM1 and BLIP proteins are replaced either by other water located in a different position but interacting with the same protein atom, or by a residue of the interacting protein. Water-mediated interactions found in the wild type complex of the TEM1 and BLIP proteins are summarized in Table 4. Movie 1 shows the water-mediated interactions within the TEM1−BLIP interface. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Colombo, M. F., Rau, D. C., and Parsegian, V. A. (1992) Protein solvation in allosteric regulation: a water effect on hemoglobin, *Science 256*, 655−659.
2. Pocker, Y. (2000) Water in enzyme reactions: biophysical aspects of hydration-dehydration processes, *Cell. Mol. Life Sci. 57*, 1008−1017.
3. Levy, Y., and Onuchic, J. N. (2006) Water mediation in protein folding and molecular recognition, *Annu. Rev. Biophys. Biomol. Struct. 35*, 389−415.
4. Raschke, T. M. (2006) Water structure and interactions with protein surfaces, *Curr. Opin. Struct. Biol. 16*, 152−159.
5. Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M., and Clore, G. M. (1995) Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR, *Science 267*, 1813−1817.
6. Otting, G., Liepinsh, E., and Wuthrich, K. (1991) Protein hydration in aqueous solution, *Science 254*, 974−980.
7. Pal, S. K., Peon, J., and Zewail, A. H. (2002) Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution, *Proc. Natl. Acad. Sci. U.S.A. 99*, 1763−1768.

8. Park, S., and Saven, J. G. (2005) Statistical and molecular dynamics studies of buried waters in globular proteins, *Proteins 60*, 450−463.
9. Williams, M. A., Goodfellow, J. M., and Thornton, J. M. (1994) Buried waters and internal cavities in monomeric proteins, *Protein Sci. 3*, 1224−1235.
10. Miller, M. D., Cai, J., and Krause, K. L. (1999) The active site of Serratia endonuclease contains a conserved magnesium-water cluster, *J. Mol. Biol. 288*, 975−987.
11. Paliwal, A., Asthagiri, D., Abras, D., Lenhoff, A. M., and Paulaitis, M. E. (2005) Light-scattering studies of protein solutions: role of hydration in weak protein-protein interactions, *Biophys. J. 89*, 1564−1573.
12. Rodier, F., Bahadur, R. P., Chakrabarti, P., and Janin, J. (2005) Hydration of protein-protein interfaces, *Proteins 60*, 36−45.
13. Rohrer, M., Prisner, T. F., Brugmann, O., Kass, H., Spoerner, M., Wittinghofer, A., and Kalbitzer, H. R. (2001) Structure of the metal-water complex in Ras x GDP studied by high-field EPR spectroscopy and 31P NMR spectroscopy, *Biochemistry 40*, 1884−1889.
14. Westhof, E. (1988) Water: an integral part of nucleic acid structure, *Annu. Rev. Biophys. Biophys. Chem. 17*, 125−144.
15. Janin, J. (1999) Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition, *Structure 7*, R277−9.
16. Mustata, G., and Briggs, J. M. (2004) Cluster analysis of water molecules in alanine racemase and their putative structural role, *Protein Eng., Des. Sel. 17*, 223−234.
17. Sreenivasan, U., and Axelsen, P. H. (1992) Buried water in homologous serine proteases, *Biochemistry 31*, 12785−12791.
18. Babor, M., Sobolev, V., and Edelman, M. (2002) Conserved positions for ribose recognition: importance of water bridging interactions among ATP, ADP and FAD-protein complexes, *J. Mol. Biol. 323*, 523−532.
19. Amadasi, A., Spyrakis, F., Cozzini, P., Abraham, D. J., Kellogg, G. E., and Mozzarelli, A. (2006) Mapping the energetics of water-protein and water-ligand interactions with the "natural" HINT forcefield: predictive tools for characterizing the roles of water in biomolecules, *J. Mol. Biol. 358*, 289−309.
20. Barillari, C., Taylor, J., Viner, R., and Essex, J. W. (2007) Classification of water molecules in protein binding sites, *J. Am. Chem. Soc. 129*, 2577−2587.
21. Levy, Y., and Onuchic, J. N. (2004) Water and proteins: a love-hate relationship, *Proc. Natl. Acad. Sci. U.S.A. 101*, 3325−3326.
22. Papoian, G. A., Ulander, J., Eastwood, M. P., Luthey-Schulten, Z., and Wolynes, P. G. (2004) Water in protein structure prediction, *Proc. Natl. Acad. Sci. U.S.A. 101*, 3352−3357.
23. Jiang, L., Kuhlman, B., Kortemme, T., and Baker, D. (2005) A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces, *Proteins 58*, 893−904.
24. Albeck, S., Unger, R., and Schreiber, G. (2000) Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges, *J. Mol. Biol. 298*, 503−520.
25. Horovitz, A. (1996) Double-mutant cycles: a powerful tool for analyzing protein structure and function, *Fold. Des. 1*, R121−6.
26. Reichmann, D., Rahat, O., Albeck, S., Meged, R., Dym, O., and Schreiber, G. (2005) The modular architecture of protein-protein binding interfaces, *Proc. Natl. Acad. Sci. U.S.A. 102*, 57−62.
27. Jang, D. S., Cha, H. J., Cha, S. S., Hong, B. H., Ha, N. C., Lee, J. Y., Oh, B. H., Lee, H. S., and Choi, K. Y. (2004) Structural double-mutant cycle analysis of a hydrogen bond network in ketosteroid isomerase from Pseudomonas putida biotype B, *Biochem. J. 382*, 967−973.
28. Langhorst, U., Backmann, J., Loris, R., and Steyaert, J. (2000) Analysis of a water mediated protein-protein interactions within RNase T1, *Biochemistry 39*, 6586−6593.
29. Ikura, T., Urakubo, Y., and Ito, N. (2004) Water-mediated interaction at a protein-protein interface, *Chem. Phys. 307*, 111−119.
30. Buckle, A. M., Schreiber, G., and Fersht, A. R. (1994) Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution, *Biochemistry 33*, 8878−8889.
31. Neuvirth, H., Raz, R., and Schreiber, G. (2004) ProMate: a structure based prediction program to identify the location of protein-protein binding sites, *J. Mol. Biol. 338*, 181−199.
32. Mintseris, J., Wiehe, K., Pierce, B., Anderson, R., Chen, R., Janin, J., and Weng, Z. (2005) Protein-Protein Docking Benchmark 2.0: an update, *Proteins 60*, 214−216.
33. Lawrence, M. C., and Colman, P. M. (1993) Shape complementarity at protein/protein interfaces, *J. Mol. Biol. 234*, 946−950.
34. Sundberg, E. J., and Mariuzza, R. A. (2002) Molecular recognition in antibody-antigen complexes, *Adv. Protein Chem. 61*, 119−160.
35. Kortemme, T., Morozov, A. V., and Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes, *J. Mol. Biol. 326*, 1239−1259.
36. Reichmann, D., Cohen, M., Abramovich, R., Dym, O., Lim, D., Strynadka, N. C., and Schreiber, G. (2007) Binding Hot Spots in the TEM1-BLIP Interface in Light of its Modular Architecture, *J. Mol. Biol. 365*, 663−679.
37. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, *J. Mol. Biol. 285*, 1735−1747.
38. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms, *J. Mol. Biol. 285*, 1711−1733.
39. Babu, M. M. (2003) NCI: A server to identify non-canonical interactions in protein structures, *Nucleic Acids Res. 31*, 3345−3348.
40. Selzer, T., Albeck, S., and Schreiber, G. (2000) Rational design of faster associating and tighter binding protein complexes, *Nat. Struct. Biol. 7*, 537−541.
41. Levitt, M., and Perutz, M. F. (1988) Aromatic rings act as hydrogen bond acceptors, *J. Mol. Biol. 201*, 751−754.
42. Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors, *J. Mol. Biol. 220*, 507−530.
43. Eyal, E., Gerzon, S., Potapov, V., Edelman, M., and Sobolev, V. (2005) The limit of accuracy of protein modeling: influence of crystal packing on protein structure, *J. Mol. Biol. 351*, 431−442.
44. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods Enzymol. 154*, 367−382.
45. Albeck, S., and Schreiber, G. (1999) Biophysical characterization of the interaction of the beta-lactamase TEM-1 with its protein inhibitor BLIP, *Biochemistry 38*, 11−21.
46. Slutzki, M., Jaitin, D. A., Yehezkel, T. B., and Schreiber, G. (2006) Variations in the unstructured C-terminal tail of interferons contribute to differential receptor binding and biological activity, *J. Mol. Biol. 360*, 1019−1030.
47. Strynadka, N. C., Jensen, S. E., Johns, K., Blanchard, H., Page, M., Matagne, A., Frere, J. M., and James, M. N. (1994) Structural and kinetic characterization of a beta-lactamase-inhibitor protein, *Nature 368*, 657−660.

BI7019639