A "Solvated Rotamer" Approach to Modeling Water-Mediated Hydrogen Bonds at Protein–Protein Interfaces

Lin Jiang,[1,2] **Brian Kuhlman,**[†] **Tanja Kortemme,**[‡] **and David Baker**[1,2*]
[1]*Department of Biochemistry, University of Washington, Seattle, Washington*
[2]*Biomolecular Structure & Design Program, University of Washington, Seattle, Washington*

ABSTRACT   Water-mediated hydrogen bonds play critical roles at protein–protein and protein–nucleic acid interfaces, and the interactions formed by discrete water molecules cannot be captured using continuum solvent models. We describe a simple model for the energetics of water-mediated hydrogen bonds, and show that, together with knowledge of the positions of buried water molecules observed in X-ray crystal structures, the model improves the prediction of free-energy changes upon mutation at protein–protein interfaces, and the recovery of native amino acid sequences in protein interface design calculations. We then describe a "solvated rotamer" approach to efficiently predict the positions of water molecules, at protein–protein interfaces and in monomeric proteins, that is compatible with widely used rotamer-based side-chain packing and protein design algorithms. Finally, we examine the extent to which the predicted water molecules can be used to improve prediction of amino acid identities and protein–protein interface stability, and discuss avenues for overcoming current limitations of the approach. Proteins 2005;58:893–904.   © 2005 Wiley-Liss, Inc.

Key words: hydration; solvent modeling; water-mediated hydrogen bonds; free-energy function; protein design; protein–protein interactions

INTRODUCTION

Many protein–protein interfaces contain specifically bound water molecules that bridge the side chains via multiple hydrogen bonds.[1,2] Side-chain truncation experiments have shown that such water-mediated hydrogen bond networks can contribute significantly to the free-energy of interaction of two proteins: the removal of one of the partners in the network often leads to substantial destabilization.[3,4] Covell and Wallqvist[5] showed that accounting for the loss of water-mediated hydrogen bonds observed in X-ray crystal structures improves the accuracy of prediction of the effects of side chain truncations on the binding free-energy, and Luque and Freire[6] found that treating buried crystallographic waters as part of the ligand improved prediction of the binding enthalpy of small ligands.

Widely used continuum solvation models cannot accurately describe water-mediated hydrogen bond networks. The energetics of these networks are likely to depend sensitively on the precise orientation of the water molecules making multiple hydrogen bonds and more generally reflect the properties of discrete individual molecules rather than continuum solvent properties. The traditional approach to avoiding the problems associated with continuum solvent models is to model all solvent molecules explicitly. However, to obtain free-energy differences for comparison to experiment, it is necessary to adequately sample the large number of possible arrangements of the solvent molecules, which is a formidable challenge. One approach is to compute ensemble averages over the course of a molecular dynamics simulation.[7] Serrano and coworkers[8] carried out molecular dynamics simulations in explicit solvent and computed the frequency with which water molecules were bound at specific sites along the peptide chain. Interaction free energies were estimated from these occupancies and used to improve modeling of the free-energy of alternative peptide conformations. However, because of the computational expense of molecular dynamics simulations in explicit water, this approach is not viable for many prediction and design applications in which the free-energy of very large numbers of alternative structures or sequences must be determined.

Here we present a computationally efficient method for modeling discrete water molecules at protein–protein interfaces in prediction and design calculations. Our approach extends current side-chain packing methods by using a rotamer library including solvated rotamers with one or more water molecules fixed to polar functional groups in probable hydrogen bond orientations, together with a simple energetic description of water-mediated hydrogen bonds. We describe the performance of the model on a series of prediction and design tests: the prediction of 1) the positions of water molecules at protein–protein interfaces, 2) the energetic effects of alanine scanning muta-

---

TABLE I. Placement of Water Molecules Around Functional Groups

Classes	Illustration of water position[a]	Distance d (Å)	Angle θ	Description
Carbonyl oxygen (backbone, Asp, Glu, Gln, and Asn)		Mainchain O: 2.70 O (Gly): 3.20 Sidechain O: 2.80	50° 50° 50°	
Amide nitrogen (backbone, Lys, Arg, Gln, and Asn)		Mainchain N: 1.95 (Gly): 2.30 Sidechain N: 1.95	180° 180° 180°	
Hydroxyl oxygen (in SER, THR)		2.80	180°	Water molecules are placed every 120° along the l axis
Hydroxyl oxygen (in TYR)		2.80	180°	
Aromatic nitrogen (in HIS)		2.95	180°	

[a]The water is in the plane of the functional group, unless otherwise noted.

tions, 3) binding free energies of protein complexes, and 4) amino acid identities at protein–protein interfaces.

METHODS
Definition of Buried Bridging Water Molecules at Protein Interfaces

In this study we only considered buried water-bridging hydrogen bonds in protein–protein interfaces. Buried water molecules were defined by the number of C_β atoms within a sphere of 8 Å radius of the oxygen atom of the water molecule of interest (buried $\geq$ 10). Water molecules were assumed to be at the interface if they were within 4.0 Å of heavy atoms of both partners of protein complexes. The hydrogens on water were not considered in our analysis. Water molecules were defined as bridging if the water formed hydrogen bonds to two polar atoms within the distance and angle constraints described in the Results section.

Solvated Rotamers

We started with a recent Dunbrack rotamer library (http://dunbrack.fccc.edu/bbdep/index.php), supplemented with additional rotamers generated by varying the chi 1 angle by ±1 standard deviation. To these rotamers, we added water molecules based on previous studies of the distribution of water molecules around backbone and polar side-chain atoms as described in the main text. Illustrations of the water positions and the exact geometric parameters used in placement are given in Table I: a single water molecule was associated with the amide nitrogen linearly, and two water molecules with the carbonyl oxygen along the direction of the two lone pairs on the sp^2 hybridized oxygen (backbone and side-chain carbonyl oxygen water placements were slightly different, see Table 1). For serine and threonine, water molecules attached to hydroxyl groups had three different, staggered positions, and for tyrosine hydroxyl groups, water molecules were assumed to be in the plane of the aromatic ring. Waters were associated with both aromatic nitrogens of histidine.

Figure 1 shows the allowed positions of water molecules around different amino acids. Extra backbone–water rotamers were added for polar and nonpolar amino acids. Except for proline, three rotamers were added with a single back-bone water: one amide-nitrogen water rotamer and two carbonyl-oxygen water rotamers, and two rotamers were added with two backbone waters: one rotamer with the amide-nitrogen water and a carbonyl-oxygen water molecule and the other rotamer with the amide-nitrogen water and the other carbonyl-oxygen water molecule. For proline, only two single-carbonyl-oxygen-water rotamers were added.

TABLE II. Number of Extra Solvated Rotamer per "Normal" Rotamer

| | Number of allowed water positions | | Number of extra rotamers per standard dunbrack rotamer | | | | | | | |
| | | | Number of water molecules attached to backbone | | | Number of water molecules attached to side chain | | | | |
Amino acids	Backbone	Side chain	0	1	2	0	1	2	3	Total
SER, THR	3	3	1	3	2	1	3	0	0	24
TYR	3	2	1	3	2	1	2	0	0	18
TRP	3	1	1	3	2	1	1	0	0	12
LYS	3	3	1	3	2	1	3	0	0	24
ASN, ASP	3	4	1	3	2	1	4	2	0	42
GLN, GLU	3	4	1	3	2	1	4	2	0	42
ARG	3	5	1	3	2	1	5	6	1	78
HIS	3	2	1	3	2	1	2	1	0	24
PRO	2	0	1	2	0	1	0	0	0	3
Others	3	0	1	3	2	1	0	0	0	6

Side-chain–water rotamers were generated from the set of allowed water positions shown in Figure 1 as follows. A single side-chain–water rotamer was added for each of the allowed side-chain water positions for each rotamer for each amino acid. To reduce the combinatoric complexity of multiple side-chain–water rotamers, rotamers with two adjacent water molecules were excluded, and thus asparagines, aspartic acid, glutamine, glutamic acid and arginine had two-side-chain–water rotamers, and arginine one three-side-chain–water rotamer. For each base rotamers in the Dunbrack library, the total number of all extra rotamers per conventional rotamer in the Dunbrack library was the product of the number of backbone–water rotamers and the number of side-chain–water rotamers for each amino acid (Table II). The extra water rotamers were added only if all the attached water molecules were buried (no less than 10 C_β atoms within a sphere of 8 Å radius of the water molecule).

The Free-Energy Function

The total energy is a sum of protein–protein interactions and protein–water interactions. The energy of protein–protein interactions was as previously described.[9–13] The protein–water interactions consist of a Lennard Jones repulsive potential and a water-mediated hydrogen-bonding potential described in the Results section (Fig. 3(e)].

The side-chain–side-chain and side-chain–backbone hydrogen-bonding potentials were scaled by the extent of burial of both participating residues. The extent of burial was converted to a weight using

$$\text{weight} = \begin{cases} 0.2 & nb < 7 \\ (nb - 2.75)/21.25 & 7 \le nb \le 24 \\ 1.0 & nb > 24 \end{cases} \quad (3)$$

where nb is the number of neighbor C_β atoms (within a sphere of 8 Å radius of the C_β atom of the residue of interest or the water oxygen). Side-chain–side-chain and side-chain–backbone hydrogen bonds between two residues were scaled by the average of the burial weights for the two residues involved.

The binding energy ($\Delta\Delta G_{bind}$) is the change in free-energy upon formation of a protein complex from unbound partners:

$$\Delta\Delta G_{bind} = \Delta G_{complex} - \Delta G_{partnerA} - \Delta G_{partnerB} \quad (4)$$

where $\Delta G_{complex}$, $\Delta G_{partnerA}$ and $\Delta G_{partnerB}$ are the free-energies of the complex and the unbound partners.

The binding energy change upon alanine mutation ($\Delta\Delta G_{bind}^{ALA}$) is computed as the difference of the binding energies of wild-type ($\Delta\Delta G_{bind}^{WT}$) and mutant ($\Delta\Delta G_{bind}^{MUT}$) proteins.

$$\Delta\Delta G_{bind}^{ALA} = \Delta\Delta G_{bind}^{MUT} - \Delta\Delta G_{bind}^{WT} \quad (5)$$

where WT and MUT describe wild-type and mutant proteins.

All parameters for the water-mediated hydrogen-bonding potentials can be found in the Supplementary Material.

Parameterizing the Energy Function

The relative contribution of the different terms of the free-energy function was parameterized for the free-energy calculations and the design tests using a conjugate-gradient-based optimization method. For the free-energy calculations, the ProTherm dataset of alanine mutation of monomeric proteins (www.rtc.riken.go.jp/jouhou/protherm/protherm.html) was used and weights were obtained as described previously.[12] For the design tests, energies were computed for all rotamers of all amino acids at all interface positions in a 170 protein complex data set taken from Mintseris and Weng[14] restricting all other amino acids to their native conformation. The weights on all terms were optimized to maximize the probability of the native amino acid type at each position as described in Kuhlman and Baker.[10]

RESULTS
Derivation of Solvated Rotamer Library

We describe the development of a "solvated rotamer" approach for modeling water molecules that make hydrogen bonds to polar protein side-chain and backbone groups. Our approach builds on earlier studies of the distributions of water molecules around polar groups in high-resolution

protein crystal structures.[15,16] These studies showed that water molecules are tightly clustered around functional group in polar side-chains in positions expected based on hydrogen bonding stereochemistry. Roe and Teeter[15] derived centroid positions for the water clusters around five polar side chains, and showed that water positions could be predicted to reasonable accuracy by placing water molecules at these positions.

We use the centroid positions derived in these previous studies to create a "solvated rotamer" library. Water molecules are placed using these centroid positions around the polar functional groups in each of the conformers in Dunbrack's backbone-dependent rotamer library (http://dunbrack.fccc.edu/bbdep/index.php). The allowed positions of water molecules around the backbone and the polar side chains are shown in Figure 1. The geometric parameters used in water placement are illustrated in Table I. For each polar side-chain rotamer in the Dunbrack library, additional rotamers are generated by adding one or more water molecules at the indicated positions. Association of water with the protein backbone is treated by generating additional rotamers with one or more waters associated with the backbone for each side-chain rotamer. To avoid the combinatoric explosion of solvated rotamers, e.g., in the case of arginine, waters are not allowed at adjacent sites on the side chain; this restricts the numbers of solvated rotamers to the values listed in Table II.

Model for Energetics of Water-Mediated Hydrogen Bonding Interactions

To use the solvated rotamer library in the prediction of side-chain conformations and design applications requires a term in the energy function representing water-mediated interactions. Here, we consider only the contribution of the explicitly modeled water molecules to water-mediated hydrogen bonds; we leave discussion of the possible use of these predicted water molecules to improve modeling of solvation more generally to the Discussion section. Based on our previous findings of the efficacy and accuracy of an orientation-dependent hydrogen-bonding potential for hydrogen bonds between protein atoms,[9,12] we develop an orientation-dependent hydrogen-bonding potential for water-mediated hydrogen bonds.

The water-mediated hydrogen-bond energy function was derived from the distribution of water-bridging hydrogen-bond geometries observed in high-resolution protein crystal structures. Statistics on water-mediated hydrogen-bond geometries were complied from a set of 91 crystal structures of two-chain protein complexes with a resolution of 2.0 Å or better containing no less than 10 water molecules at interfaces, which was taken from the Dunbrack culled pdb set.[17] In the statistics, only buried water molecules were considered (see Methods). Since there are no hydrogens on water molecules in crystal structures, the distance between the protein polar atoms (donor/acceptor) and oxygen of water were measured (the distance between donor heavy atom and water oxygen was very similar to that between acceptor atom and water oxygen). Figure 2 illustrates the four geometrical parameters considered: (a)

the distance d_{wp} between the protein polar atoms (donor/acceptor) and oxygen of water, (b) the angle Θ at the hydrogen atom, (c) the angle Ψ at the acceptor atom and (d) the angle Ω at the water oxygen atom.

Figure 3(a–d) shows the distributions of d_{wp}, Θ, Ψ, and Ω obtained from the analysis for a total of 2,594 interfacial water molecules. For the angular distributions of hydrogen bonds, only occurrences with a water-acceptor/donor distance between 2.6 Å and 3.6 Å were considered. The distributions were corrected for the differences in volume elements of the bins. The distance distributions show a maximum at around 2.90 Å, which is similar to that observed for side-chain–side-chain hydrogen bonds.[9] The distribution of angle Ψ at the acceptor atom has a peak at around 120°, consistent with a simple picture in which the water molecule interacts with the lone pairs of the acceptor. The distribution of angle Θ at the hydrogen atom is peaked at around 180° showing a clear preference for linear water hydrogen bonds. The angle Ω at the water oxygen atom has a more broad distribution, possibly reflecting the alternative orientations of water molecules. The peak at around 110° in the Ω distribution is consistent with water H-bonds to protein atoms along the direction of the lone pairs and O-H bonds of the sp[3] hybridized water oxygen atom.

Because of errors in the placement of water molecules at the end of discrete polar-side-chain rotamers and the limited amount of data available, we chose to model the angular dependence using simple cutoffs: a water molecule was considered to make a water-mediated hydrogen bond only if $90° \leq \psi \leq 160°, 120° \leq \Theta \leq 180°, 80° \leq \Omega \leq 140°$. If these conditions are satisfied, the hydrogen bond energy was computed using a distance-dependent potential [Fig. 3(e)] obtained from the logarithm of the distance distribution of water-mediated hydrogen bonds in crystal structures.

In addition to the water-mediated hydrogen-bonding term, the water molecules interact with all protein atoms and each other via a Lennard Jones repulsive term to avoid clashes between atoms. The remaining terms in the potential function, the most important of which are a Lennard Jones potential, an implicit solvation model, and an orientation-dependent hydrogen-bonding model, describe the interactions between protein atoms and are the same as in our previous prediction and design work.[10–12]

Test of the Solvated Rotamer Approach

We used a number of different tests to evaluate our treatment of water-mediated hydrogen bonds. The first test is to predict the positions of buried water molecules in X-ray crystal structures, which evaluates the accuracy of water placement. The second and third tests evaluate the energy function: the second test is to calculate binding free-energy changes upon alanine mutations in protein complexes and the third test is to calculate binding free energies of protein complexes. The fourth test, a prelude to protein design calculations using the solvated rotamers, is to assess the recovery of the native amino acid sequence at protein–protein interfaces when redesigning amino acids
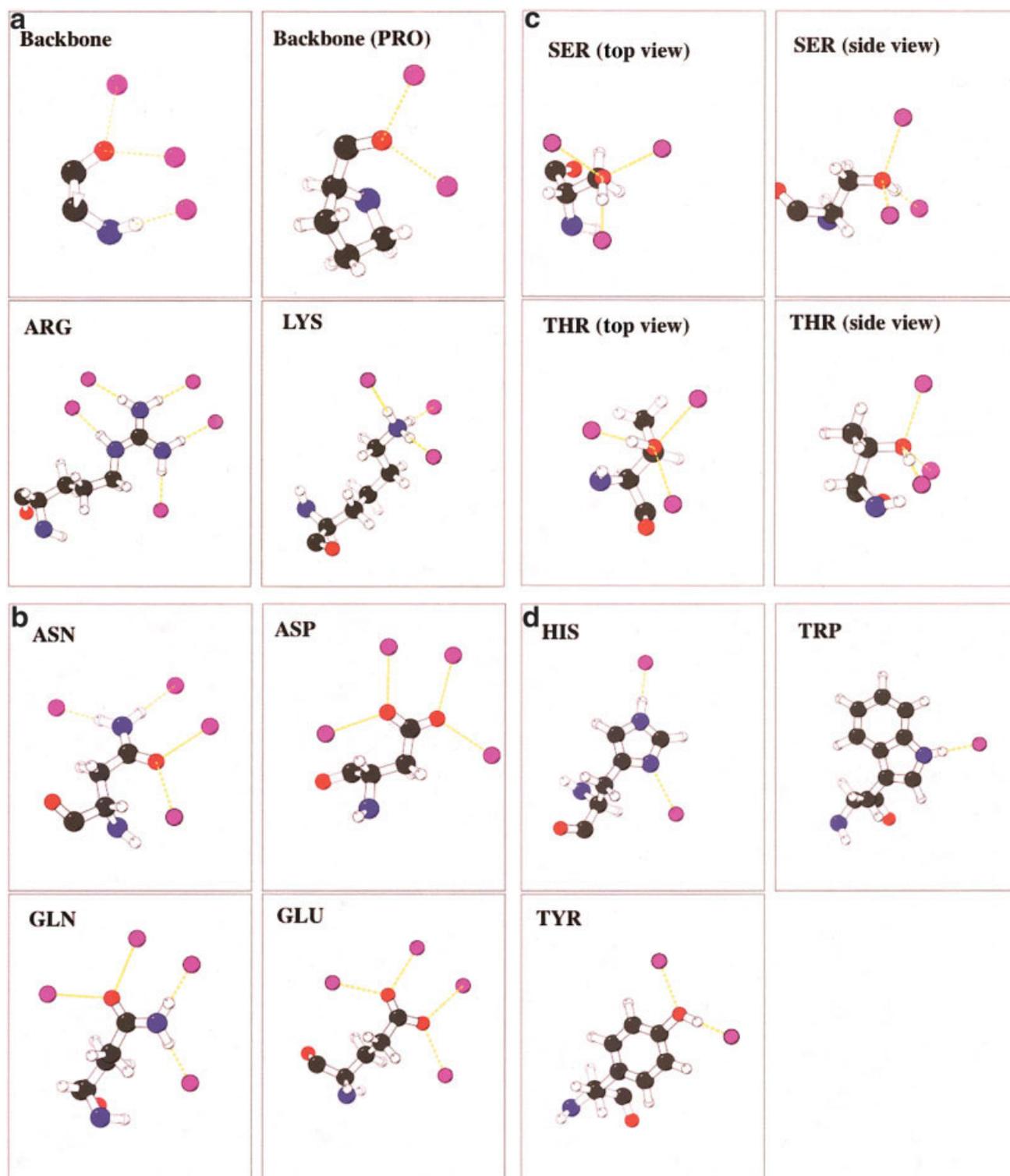
Fig. 1. Water placement in solvated rotamers. Protein atoms are colored using the CPK convention (nitrogen, blue; oxygen, red; carbon, black; hydrogen, white). Oxygen atoms in water molecules are colored in purple. Hydrogen bonds are indicated by yellow dashed lines. For serine and threonine, two views are shown. These figures were prepared with Molscript.[22]

one at a time.[10] Wherever appropriate, we compare results in these tests with 1) no explicit solvent model, 2) the water molecules observed in the X-ray crystal structure, 3) water molecules predicted using the geometric parameters in Table I and fixing the side chains in their native conformations, and 4) water molecules predicted by full
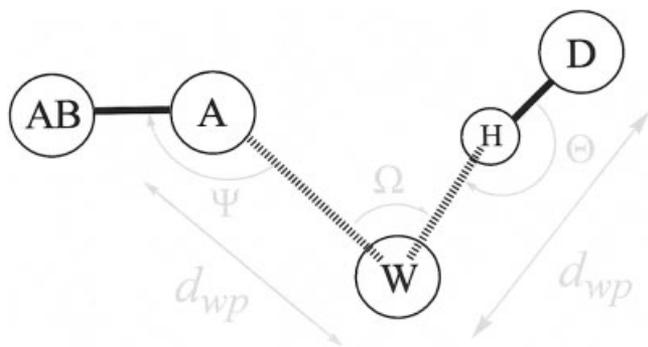
Fig. 2. Schematic representation of the parameters used to describe water-mediated hydrogen bond geometry. d_{wp}, distance between the water oxygen and polar (acceptor/donor) atoms; Θ, angle at the hydrogen atom; Ψ, angle at the acceptor atom; Ω, angle at the water oxygen atom. A, acceptor; D, donor; H, hydrogen; AB, acceptor base; W, water oxygen.

side chain repacking (simultaneous alterations of all side chain conformations modeled as rotamers) using the solvated rotamer library.

Prediction of water positions at protein interfaces

The dataset for testing water prediction was taken from the nonredundant set of protein–protein complexes compiled by Conte et al.[18] Only structures with a resolution of 2.0 Å or better and containing more than eight water molecules buried at the interface (and hence likely to contain water-mediated hydrogen bonding networks) were selected, which left 15 protein complexes containing 264 experimental water molecules. The positions of buried waters were predicted using two approaches. In one set of calculations the side chains were fixed in their native conformations and only the water positions were optimized, and in a second set of calculations the side chain positions were optimized in addition to the waters. The total number of predicted water molecules (241 water molecules when fixing and 302 molecules when repacking the side-chain conformations) was comparable to the number of bound water molecules observed in the experimental crystal structures (264 water molecules). The numbers of predicted water molecules for each individual interface are shown in the Supplementary Material. Figure 4 shows the cumulative percentage of closest distances between native water and predicted water when fixing and repacking the side-chain conformations. Many predicted water molecules are within 1.5 Å of native water molecules.

Predicted interfacial water positions in the barnase-barstar complex are shown in Figure 5(a) and in greater detail in Figure 5(b, c). Figure 5(b, c) illustrates a water-mediated H-bond network involving a central aspartate (Asp35 from barstar) and three other residues (Lys62, Asn58, and Arg59 from barnase) (left) and one water molecule bridging three side-chains (Glu73 from barnase, Lys27 from barnase and Asp39 from barstar) (right). In both cases, the water predictions with fixed side-chain conformations [Fig. 5(b)] and with repacked side chains [Fig. 5(c)] reproduced the native water positions accurately.

If structures of the unbound partners are available, knowledge of the position of bound water molecules in these structures could likely be used to improve prediction of water positions in the complex using the method for predicting conservation of water positions developed by Rayme et al.[19]

Prediction of free-energy changes associated with alanine mutations at protein–protein interfaces

We have previously described the prediction of the effects of alanine mutations on protein–protein binding free energies using a simple computational model dominated by Lennard Jones interactions, an implicit solvation model, and an orientation-dependent hydrogen bonding potential.[10–12] To incorporate the new treatment of water-mediated hydrogen bonding into this model, the relative contributions of the different terms of the free-energy function were parameterized as described previously.[12] Repulsive interactions between the water and the protein were given the same weight as repulsive interactions between protein atoms, and water-mediated hydrogen-bonding interactions were weighted the same as side-chain hydrogen bonds. As there are no additional parameters in the predicted water model, the performance of models excluding and including explicit water molecules can be compared directly.

The free-energy function was used to compute binding free-energy changes upon alanine mutations for a dataset of 19 complexes with known crystal structures and experimentally alanine scanning data,[12] using models with either no water, native water molecules, or predicted water molecules (leaving all side chains in their native conformations). Figure 6 shows the results for mutations involving residues which make [Fig. 6(a)] or are predicted to make [Fig. 6(b)] water-mediated hydrogen bonds. The correlation between observed and calculated free-energy changes was higher for the model using the native water molecules than the no water model (correlation coefficient of 0.64 versus 0.55 over the whole data set of 19 complexes and 378 mutations), but inclusion of the predicted waters showed little improvement over the no water model. As indicated in Figure 6(b), the benefit from correctly predicted water molecules appears to be offset by overprediction of free-energy changes in cases involving spurious water-mediated hydrogen bonds.

The accuracy of the predictions for different interfaces was quite variable. Figure 7 shows the result for the barnase–barstar complex. The correlation coefficient for predicted versus observed DDG changes is 0.68 using no water, 0.86 using native water, and 0.75 using solvated rotamers.

Calculation of binding free-energy of protein complexes

The total free-energy of binding with and without predicted water molecules was computed for a database of 52 complexes with known crystal structures and experimentally measured binding free-energy collected by Yaqui Zhou and coworkers (pers. commun.). Figure 8 shows that
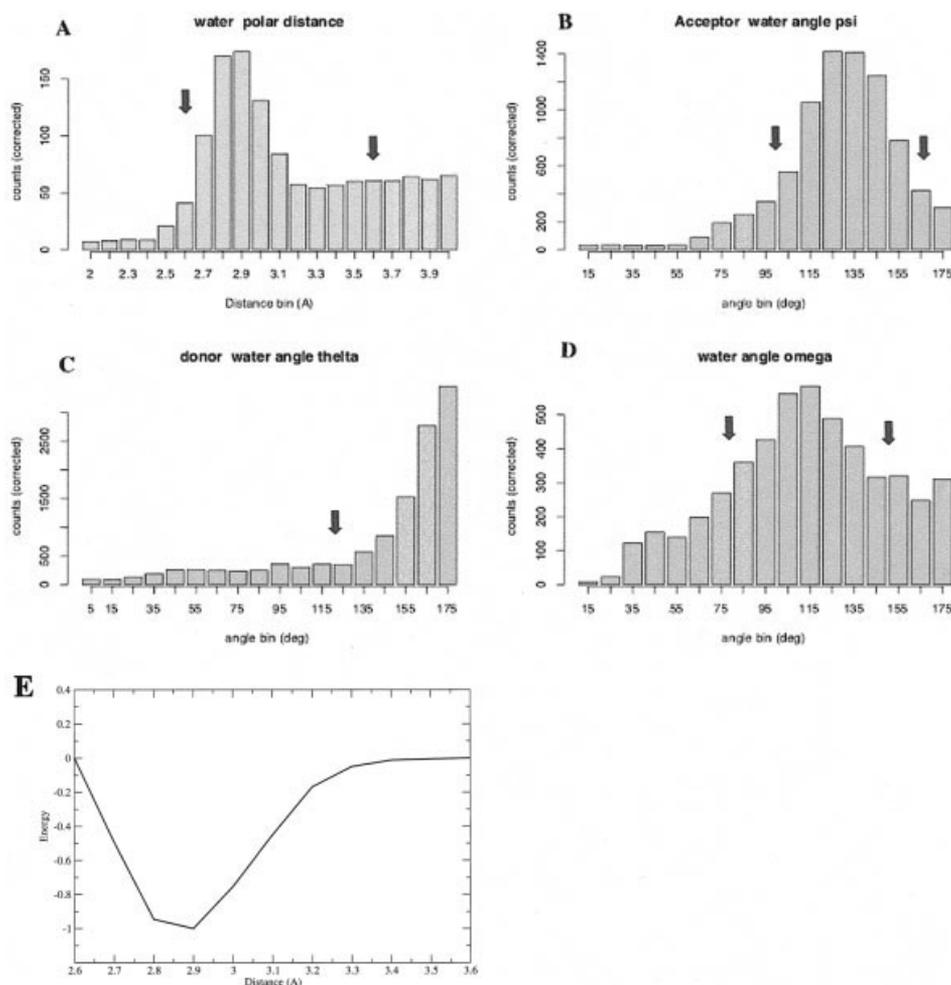
Fig. 3. Distribution of hydrogen bonding parameters obtained from 91 protein crystal structures. Raw counts were corrected for the different volume elements encompassed by the bins [angular correction: sin(angle); distance correction: (distance)^2]. **a:** water oxygen–acceptor/donor distance d_{wp}; **b:** angle Θ at the hydrogen; **c:** angle Ψ at the acceptor; **d:** angle Ω at the water oxygen; **e:** the distance-dependent potential obtained from (a). Water-mediated hydrogen bonds were considered to exist if all geometric parameters were within the bounds indicated by the arrows.

there is a modest improvement using the predicted waters; the correlation between experimentally measured and predicted binding free energies is 0.79 for using no water and 0.84 using the water molecules predicted by fixing the side chains. The slope of the best fit line is considerably less than 1, indicating that parameterizing on single point mutations greatly overestimates overall binding free energies; this may be because the free-energy cost of mutation includes both the loss of favorable interactions and a free-energy cost associated with cavity formation; for the overall binding free-energy only the first term contributes.

Prediction of side-chain identities at protein interfaces

The prediction of side chain identities was carried out by evaluating for one sequence position at a time the energy of each possible rotamer of each of the 20 amino acids (identity recovery test) and selecting the lowest energy

amino acid. The dataset for this test was taken from a nonredundant set of protein complexes compiled by Mintseris and Weng;[14] structures with missing side chain or backbone atoms were excluded and the structures having more than 620 total residues were ignored (because of limitations in computer memory) leaving 170 protein complexes with 5092 sequence position at interfaces (a residue is defined to be at the interface if its C_β atom is within 8 Å of any C_β atom from the other partner of protein complex). The data set was divided into 10 equal parts each containing 17 protein complexes. For each part, the other nine parts were used to obtain weights on the components of the energy function that maximize the probability of the native amino acid,[10] and the weighted energy function was used to predict the amino acid identities and side chain conformations of the interfacial residues. The prediction results are reported as averages of these 10 calculations.

A

Closest Distance from native to predicted water

B

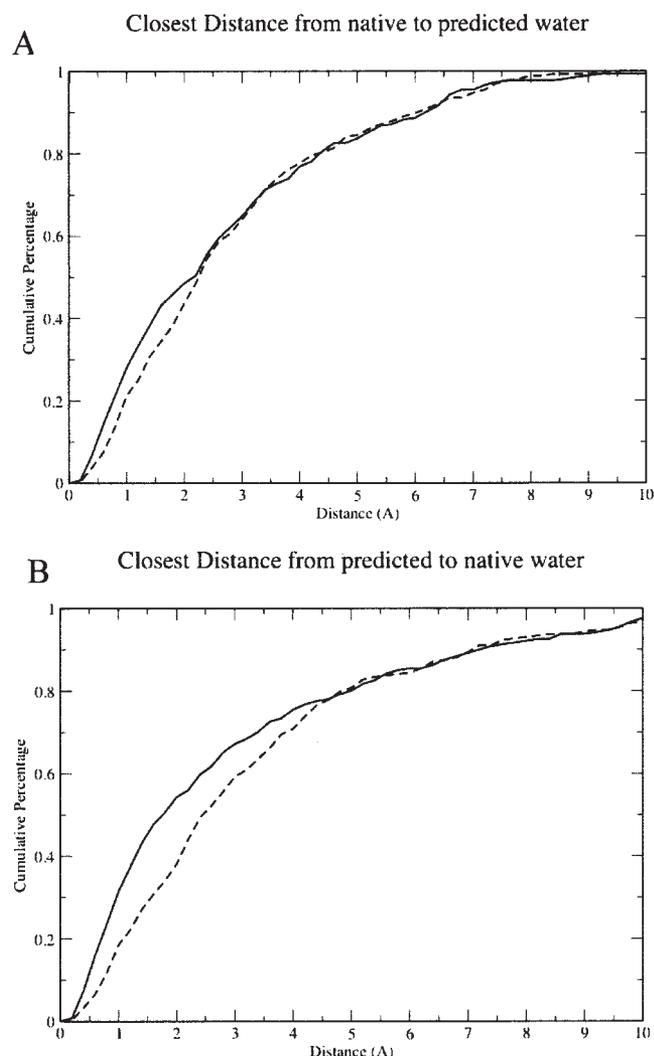Closest Distance from predicted to native water

Fig. 4. Prediction of buried water positions at protein interfaces. Cumulative percentage of the closest distances between native water and predicted water at 15 high resolution (2.0 Å or better) protein interfaces are shown. The solid lines show predicted water position using native side-chain conformations and the dashed lines show predicted water positions using the solvated rotamers. **a:** Closest distances from native to predicted water molecules. **b:** Closest distances from predicted to native water molecules.

Three sets of experiments were used to determine if our model for water-mediated hydrogen bonds increases the recovery of native amino acids at protein interfaces: 1) simulations without an explicit water model, 2) simulations using the native water positions, and 3) simulations using predicted water positions. Success was measured as the fraction of cases, for each amino acid, that the observed (native) amino acid was predicted to have the lowest energy. Because of the similarity of the solvated rotamers for aspartic acid and asparagine, and glutamine and glutamate, these two pairs are treated as the same type in computing percentage correct.

Table III shows a summary of the results for each amino acid type for each type of simulation (cysteine residues were excluded because potential disulfide bonds were not modeled). In general, including water-mediated interactions improves the recovery of native amino acids. Without a water model, 44% of the native amino acids were recovered, while 48% and 46% were recovered with native and predicted waters respectively. For each amino acid type, a substitution profile, depicting how often each of the other 16 amino acid types was chosen to be the most favorable replacement, is shown in the supplemental material. For all amino acids the native amino acid was predicted with the highest frequency. In most cases, the inclusion of the water-mediated hydrogen bonding term was useful in discriminating the native amino acid type from others. The largest improvement was observed in predictions of the identities of arginine and tyrosine residues.

DISCUSSION

Our orientation and distance-dependent model of water-mediated hydrogen bond energetics clearly improves prediction of the effects of alanine mutations on protein–protein interface stability [Fig. 6(a)] and recovery of native amino acid sequences in design tests when the positions of the water molecules are known from X-ray crystallography (Table III). While the positions of bound water molecules can be predicted reasonably using our solvated rotamer approach (Fig. 4), the improvement in the prediction and design tests are much more modest than those using the waters from the X-ray crystal structure.

Why are the results with the predicted waters in for example the prediction of the interface alanine scanning results significantly worse than using the X-ray crystallographic waters? An obvious explanation is that the errors in the energy model are compounded by uncertainties in the positions of the water molecules themselves. Improvement in the prediction of water positions and subsequently improvement in modeling using the predicted waters should be achievable through improvements in both the sampling procedure and the energy function. Improved sampling could be obtained by considering more water molecule placements around the side-chain rotamers and backbone, using larger rotamer libraries,[20] or by optimizing the torsion angles of each rotamer in the context of a fixed binding partner.[21] Increasing the number of rotamers puts large demands on computer memory, but with continuing developments in computer technology this will become less of an obstacle.

The energy function can be improved by utilizing the explicit water molecules to improve modeling of more general aspects of solvation beyond the water-mediated hydrogen bonds that are the focus of this paper. A shell of predicted water around a protein or protein–protein complex can be rapidly generated using our method. As an illustration, a predicted water shell for the barnase–barstar complex is shown in Figure 9. Continuum solvation models, such as the Lazaridis-Karplus model[11] used in our studies, estimate the extent of desolvation of polar atoms through the density of surrounding protein atoms, and hence an atom may be considered to be largely desolvated even when it can in fact contact the solvent.
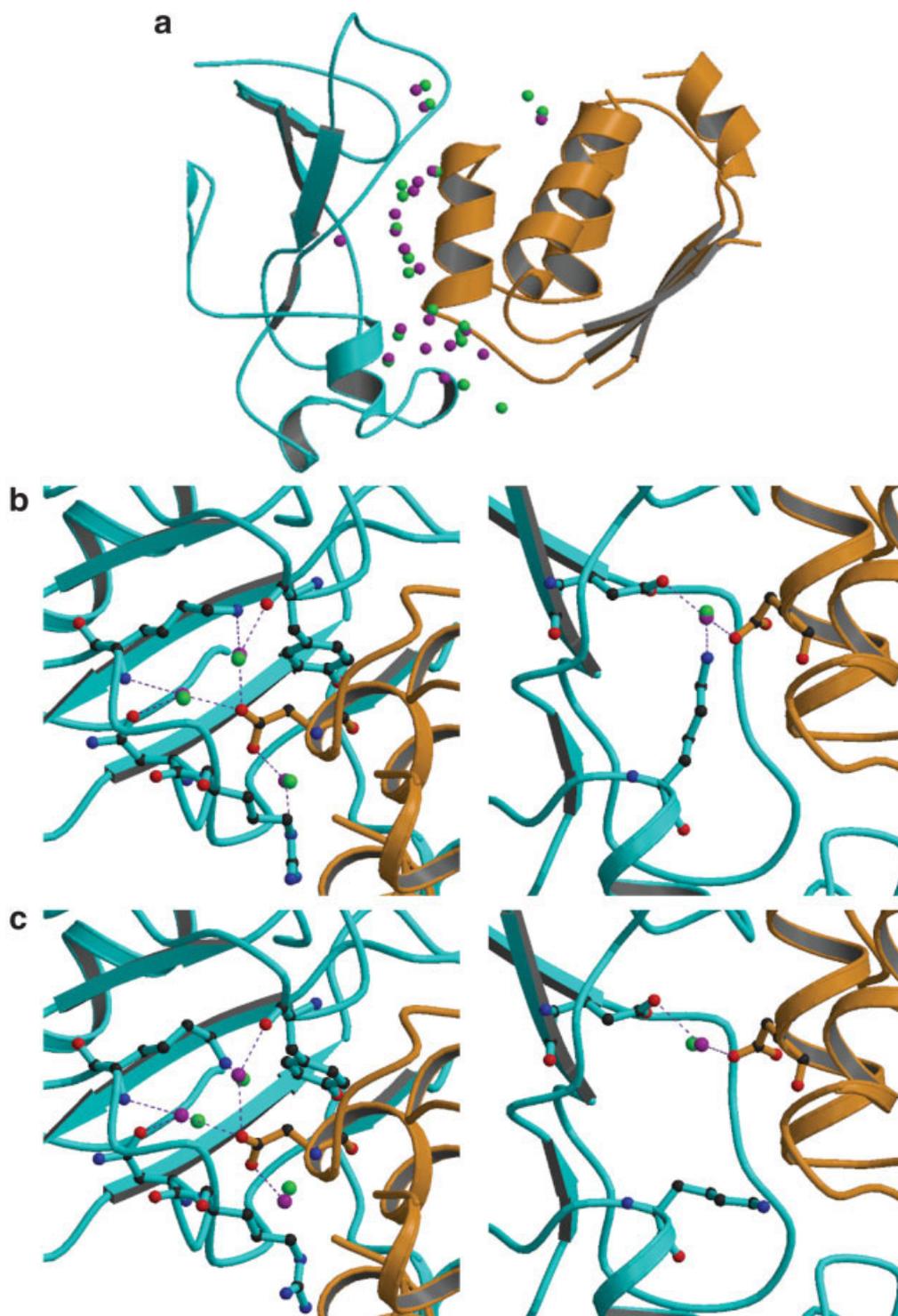
Fig. 5. Water prediction at the barnase–barstar interface. **a:** Comparison of native and predicted water positions at the barnase–barstar interface. **b:** Detail of predicted water molecules at the barnase–barstar interface when fixing the side-chains in their native conformations. **c:** Detail of predicted water molecules at the barnase–barstar interface after repacking using the solvated rotamer library. Two different sites in the interface are shown (left and right panels). Orange, Barstar; Cyan, Banase; Purple, Native water molecules; Green, predicted water molecules. Water-mediated hydrogen bonds between native water and the protein structure are displayed as purple dashed lines. The corresponding protein side-chain atoms are displayed in ball-and-stick mode. Figures are prepared with Molscript[22] and Raster3D.[23]
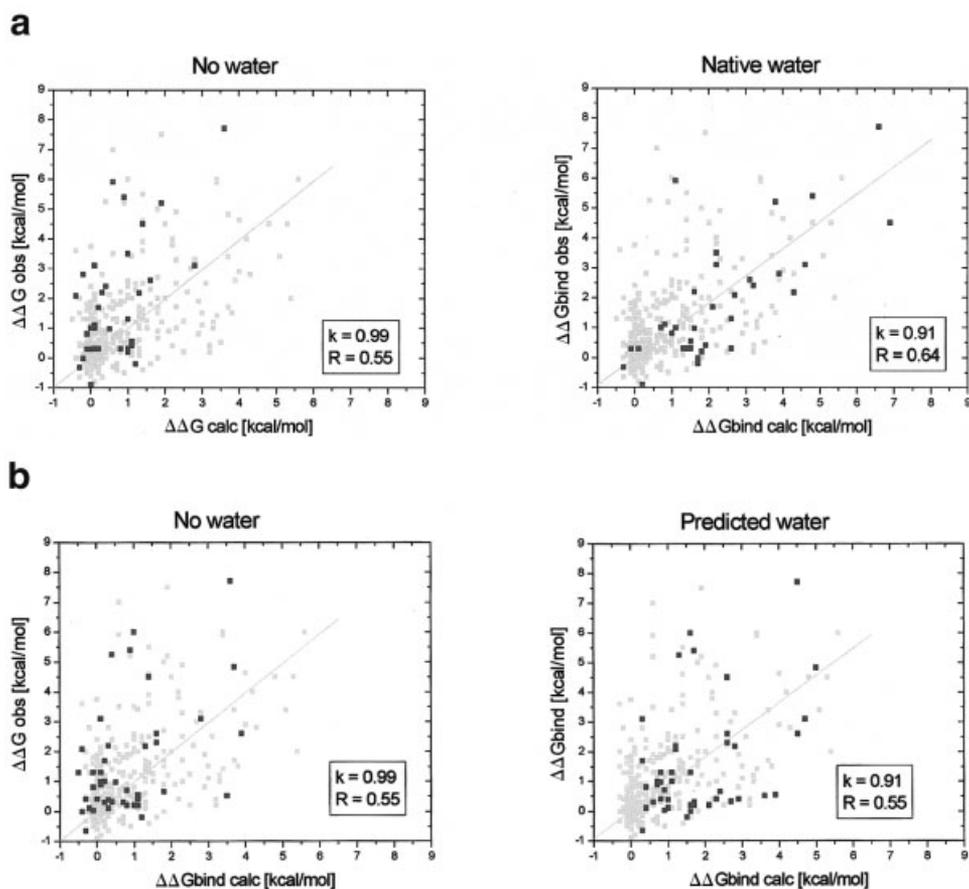
a

Fig. 6. Prediction of free-energy change of mutation to alanine at protein interfaces. The free-energy change upon mutation is calculated using no water, native water, and predicted water models. The x axis is the calculated free-energy change and the y axis the experimental free-energy change. **a:** Comparison of $\Delta\Delta G$ calculations with/without native water for 378 mutations for 19 protein complexes. Mutations involving residues making hydrogen bonds with native water molecules ($\Delta E_{h2ohb} \geq 0.1$kcal/mol) are colored in black, and the others are in light gray. **b:** Comparison of $\Delta\Delta G$ calculations with/without predicted water for the same set as in (a). Mutations involving residues which are predicted to make hydrogen bonds with predicted water molecules ($\Delta E_{h2ohb} \geq 0.1$kcal/mol) are colored in black, and the others are in light gray. The gray line represent linear fits with a fixed zero intercept for all the 378 mutations, k is the slope of the line, and R is the correlation coefficient.
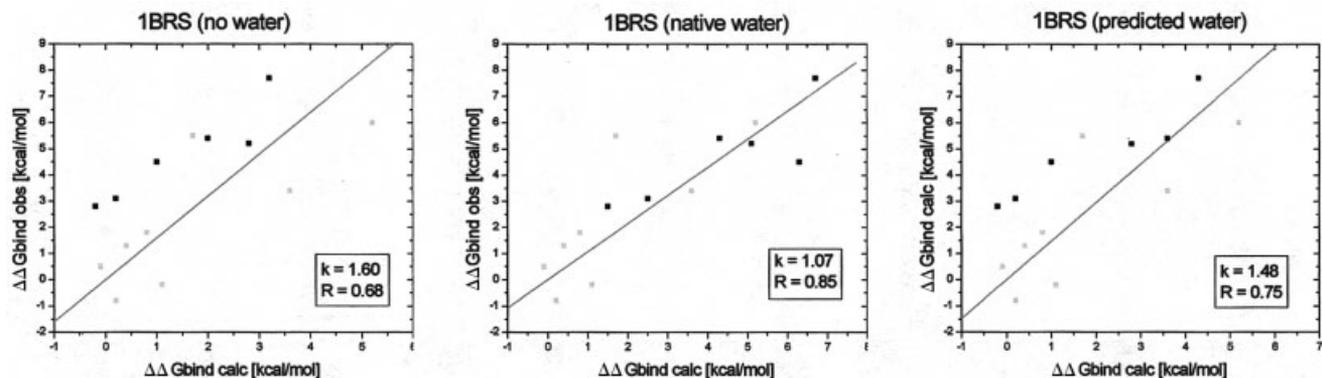

Fig. 7. Prediction of free-energy changes for alanine mutations at the barnase–barstar interface. $\Delta\Delta G$ predictions for 14 alanine mutations at the barnase–barstar interface are shown. The x axis is the calculated free-energy change and the y axis the experimental free-energy change. Black dots are the mutations involving water-mediated hydrogen bonds as identified in Schreiber and Fersht,[3] and the other dots are colored in light gray. The predicted free-energy using no water (left), native water (middle), and water predicted by fixing the side chains in their native conformations (right panel) are compared with experimental data. Lines represent linear fits with a fixed zero intercept.
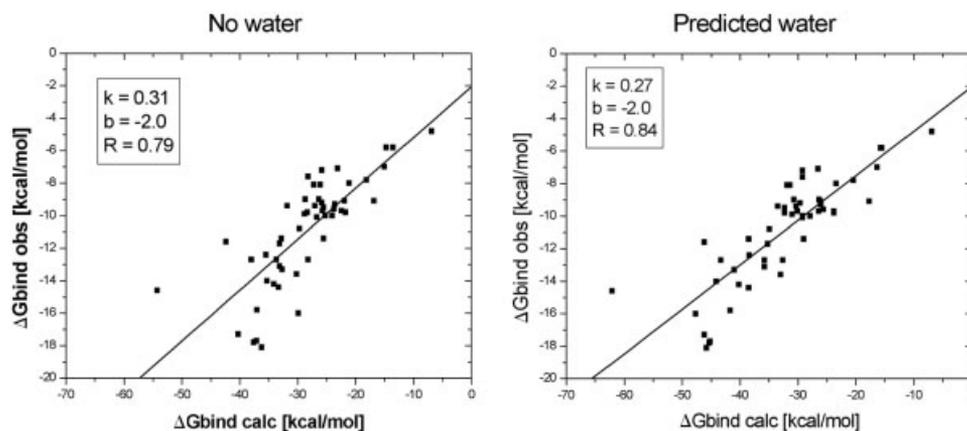
Fig. 8. Binding free-energy predictions for 52 protein complexes. The free-energy change upon protein binding was calculated excluding water-mediated interactions (left), and including water-mediated interactions using water molecules predicted using native side chains (right). The x axis is the calculated binding energy values and the y axis the experimental binding free-energy. Lines represent linear fits, k is the slope of the line, b is the intercept of the line and R is the correlation coefficient.

TABLE III. Recovery of Native Amino Acid Identity [%]

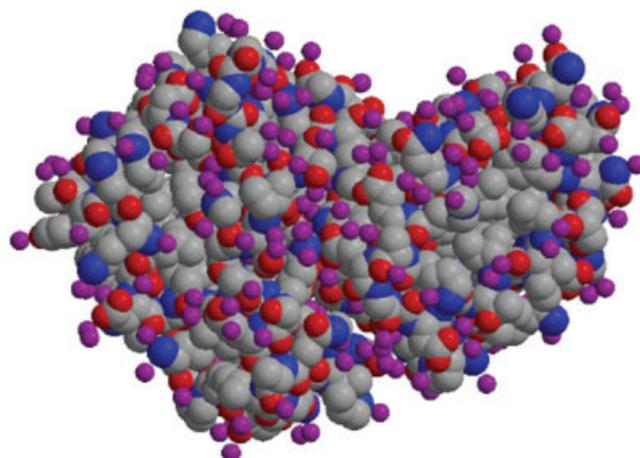|	No water	Native water	Water rotamer
ALA	0.28	0.32	0.25
ASP/ASN	0.48	0.52	0.52
GLU/GLN	0.37	0.42	0.39
PHE	0.26	0.31	0.28
GLY	0.78	0.78	0.81
HIS	0.21	0.21	0.26
ILE	0.43	0.48	0.44
LYS	0.32	0.35	0.30
LEU	0.52	0.57	0.53
MET	0.16	0.18	0.14
PRO	0.74	0.78	0.76
ARG	0.22	0.30	0.28
SER	0.40	0.46	0.42
THR	0.37	0.43	0.35
VAL	0.50	0.51	0.49
TRP	0.17	0.28	0.23
TYR	0.36	0.44	0.39
ALL	0.44	0.48	0.46

Fig. 9. Prediction of first solvation shell around barnase–barstar complex. The heavy atoms of the protein structure are shown in spacefill (nitrogen: blue, carbon: grey; oxygen: red). Predicted water molecules are represented as purple balls (the water radus is 1.4 Å).

The large electrostatic desolvation cost associated with burying atoms away from solvent can be moderated for polar atoms that are contacted by the waters attached to the solvated rotamers. In the simplest implementation of such a strategy, polar atoms in contact with explicit water molecules would be considered to be completely solvated independent of the density of surrounding protein atoms. Likewise, treatment of water–protein van der Waals interactions could be improved using the predicted waters: cavities that can be filled by predicted water molecules that interact via Lennard Jones interactions with the protein would be more favorable than cavities that cannot accommodate predicted water molecules.

In conclusion, the solvated rotamer approach described here provides a rapid means to predict water positions at protein–protein interfaces. The approach should also be useful for modeling protein–nucleic acid interactions, which often involve highly solvated interfaces. The simple water-mediated hydrogen-bonding model provides significant improvements in prediction and design tests when used in conjunction with crystallographically defined water molecules, and small improvements when used in conjunction with the predicted waters. With further improvements in the sampling methodology and the energy model, the solvated rotamer approach should become an efficient way to model the often critical contributions of specifically bound water molecules to the energetics of macromolecular interactions.

REFERENCES

1. Williams MA, Goodfellow JM, Thornton JM. Buried waters and internal cavities in monomeric proteins. Protein Sci 1994;3:1224–1235.
2. Janin J. Wet and dry interfaces: the role of solvent in protein–protein and protein-DNA recognition. Structure Fold Des 1999;7:R277–279.
3. 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.
4. Burnett JC, Kellogg GE, Abraham DJ. Computational methodology for estimating changes in free energies of biomolecular association upon mutation. The importance of bound water in dimer-tetramer assembly for beta 37 mutant hemoglobins. Biochemistry 2000;39:1622–1633.
5. 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.
6. Luque I, Freire E. Structural parameterization of the binding enthalpy of small ligands. Proteins 2002;49:181–190.
7. Wang W, Donini O, Reyes CM, Kollman PA. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein–protein, and protein-nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct 2001;30:211–243.
8. Petukhov M, Cregut D, Soares CM, Serrano L. Local water bridges and protein conformational stability. Protein Sci 1999;8:1982–1989.
9. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. J Mol Biol 2003;326:1239–1259.
10. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci USA 2000;97:10383–10388.
11. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
12. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. Proc Natl Acad Sci USA 2002;99:14116–14121.
13. Dantas G, Kuhlman B, Callender D, Wong M, Baker D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. J Mol Biol 2003;332:449–460.
14. Mintseris J, Weng Z. Atomic contact vectors in protein–protein recognition. Proteins 2003;53:629–639.
15. Roe SM, Teeter MM. Patterns for prediction of hydration around polar residues in proteins. J Mol Biol 1993;229:419–427.
16. Thanki N, Thornton JM, Goodfellow JM. Distributions of water around amino acid residues in proteins. J Mol Biol 1988;202:637–657.
17. Wang G, Dunbrack RL, Jr. PISCES: a protein sequence culling server. Bioinformatics 2003;19:1589–1591.
18. Lo Conte L, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. J Mol Biol 1999;285:2177–2198.
19. Raymer ML, Sanschagrin PC, Punch WF, Venkataraman S, Goodman ED, Kuhn LA. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. J Mol Biol 1997;265:445–464.
20. Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. J Mol Biol 2001;311:421–430.
21. Desjarlais JR, Handel TM. De novo design of the hydrophobic cores of proteins. Protein Sci 1995;4:2006–2018.
22. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Cryst 1991;24:946–950.
23. Merritt EA, Bacon DJ. Raster3D Photorealistic Molecular Graphics. Meth Enzymol 1997;277:505–524.