Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding

Tom Young, Robert Abel, Byungchan Kim, Bruce J. Berne*, and Richard A. Friesner

Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027

Contributed by Bruce J. Berne, November 17, 2006 (sent for review October 12, 2006)

The thermodynamic properties and phase behavior of water in confined regions can vary significantly from that observed in the bulk. This is particularly true for systems in which the confinement is on the molecular-length scale. In this study, we use molecular dynamics simulations and a powerful solvent analysis technique based on inhomogenous solvation theory to investigate the properties of water molecules that solvate the confined regions of protein active sites. Our simulations and analysis indicate that the solvation of protein active sites that are characterized by hydrophobic enclosure and correlated hydrogen bonds induce atypical entropic and enthalpic penalties of hydration. These penalties apparently stabilize the protein–ligand complex with respect to the independently solvated ligand and protein, which leads to enhanced binding affinities. Our analysis elucidates several challenging cases, including the super affinity of the streptavidin–biotin system.

binding motifs | hydrophobic effect | streptavidin | dewetting

The hydrophobic interaction is considered to be an important driving force in molecular recognition, yet our understanding of hydrophobicity in enclosed regions, such as those found in protein binding sites, remains incomplete. For example, the binding affinity of biotin to streptavidin is orders of magnitude larger than expected on the basis of most current theoretical models. The inability to predict such "super affinities" and the absence of a molecular understanding of hydrophobic enclosure effects stands as an obstacle to rational design of potent pharmacologically active compounds. A better understanding of the nature of such enclosures is essential to further progress in the area. We show how superaffinity can arise from active sites that have two important molecular recognition motifs: hydrophobic enclosure and correlated hydrogen bonds. Using molecular dynamics, we show that these motifs can induce atypical entropic and enthalpic penalties for hydration of the apostructures of proteins that stabilize the bound state with respect to the hydrated state and, hence, lead to super affinity.

It is widely believed that hydrophobic interactions constitute the principal thermodynamic driving force for the binding of small molecule ligands to their cognate protein receptors. A substantial number of empirical scoring functions aimed at computing protein–ligand binding affinities have been developed; invariably, the largest contribution in such expressions represents a measure of hydrophobic contact between the protein and ligand (1). Underlying these contributions is the idea that replacement of water molecules in the protein cavity by a ligand that is complementary to the protein groups lining the cavity (making hydrogen bonds where appropriate, and hydrophobic contacts otherwise) leads to a gain in binding affinity by releasing water molecules from a suboptimal environment into solution. Standard scoring functions aimed at describing this effect are based on pairwise atom–atom terms or buried surface area terms, which are parameterized by averaging over many different molecular environments. Such an approach is accurate when the molecular environments do not vary significantly from each other, as is often the case. However, two types of protein active site molecular recognition motifs have recently been identified in which displacement of the solvent by the ligand leads to exceptional binding affinities that are significantly underestimated by such standard scoring functions (2), often by several orders of magnitude. The first motif is a strongly hydrophobic cavity that encloses multiple water molecules; that is, the water molecules are surrounded on multiple sides by hydrophobic protein side chains. The second motif involves the formation of one to three hydrogen bonds with the protein by the ligand, where the remainder of the local environment is hydrophobically enclosed. The contribution of this second motif to the binding affinity of the complex was greatest when several hydrogen bonds were made to groups in close proximity on a ring system of the ligand; we refer to such structures in what follows as correlated hydrogen bonds.

The goal of the present paper is to explain the solvation of these molecular recognition motifs at an atomic level of detail. We address this question via all-atom molecular dynamics simulations, focusing on the properties of water molecules in several types of hydrophobically enclosed environments, as found in three different receptors. These receptors are the streptavidin–biotin complex, which contains five hydrophobically enclosed correlated hydrogen bonds; the Cox-2–arachidonic acid complex, which has a tight hydrophobic enclosure; and the antibody DB3–aetiocholanone complex, which has a less-pronounced hydrophobic enclosure. We also chose to simulate HIV protease receptor as a reference case because, according to the scoring function used in ref. 2, it has no hydrophobically enclosed regions when complexed with any cognate ligand.

In the computer simulations, the proteins without the ligands were inserted into a water box. The water molecules that sterically overlapped the protein were removed, and counter ions were added to maintain electric neutrality. During the molecular dynamics simulations, the proteins' heavy atoms and the counter ions were harmonically restrained to their initial positions. Although the ligand is not simulated in the molecular dynamics runs, we define the binding-cavity volume as any space that lies within 2 Å of any heavy atom of where the ligand would be. The location of the binding-cavity with respect to the protein remains constant throughout the simulation because the protein is restrained to its initial position. This procedure is analogous to the rigid receptor approximation commonly used in ligand docking studies. The positions, orientations, and energies of interaction of all waters observed in the binding-cavity were recorded, and the binding-cavity solvent density distribution was determined. For simulated systems whose binding cavities were solvated, the water was found to be structured by the protein field. We used a clustering algorithm and applied inhomogenous solvation theory (3) to characterize this ordering. The clustering algorithm partitioned the binding-cavity solvent density distribution into

Author contributions: T.Y. and R.A.F. designed research; T.Y., R.A., and B.K. performed research; T.Y. and R.A. analyzed data; and T.Y., R.A., B.J.B., and R.A.F. wrote the paper.

The authors declare no conflict of interest.

*To whom correspondence should be addressed. E-mail: bb8@chem.columbia.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0610202104

nonoverlapping, 1-Å radius spheres. Spheres circumscribing a region of the solvent significantly denser than the bulk fluid were denoted the principal hydration sites of the binding cavity. The Cox-2 binding cavity was found to be only transiently hydrated in all molecular dynamics runs and was analyzed separately. We calculated approximate thermodynamic properties of water molecules occupying each of these principle hydration sites. The energy of interaction of the water molecules with the entire system (world energies) were calculated directly from simulation. We estimated the entropic penalty of solvent ordering in each hydration site by using an expansion of the entropy in terms of orientational and spatial particle correlation functions (3–6). In this theory, a uniform bulk density distribution of solvent has an entropy of zero; deviations from this uniform distribution represent structure and result in unfavorable local contributions to the excess entropy.

When a ligand binds to a solvated protein, water in the binding cavity is expelled into the bulk. Associated with this expulsion are enthalpic and entropic contributions to the free energy of binding that arise from the differences between the waters entropic and energetic properties in the bulk and its properties in the binding cavity. In many cases of protein–ligand binding, the energy of interaction between the water in the binding cavity and the protein is roughly comparable with the energy of interaction between the docked ligand and the protein. However, the entropy of structuring waters in the protein cavity has no equivalent mapping with the change in entropy from desolvating the ligand. This asymmetry suggests that the net contribution to the free energies of ligand binding when highly ordered waters are expelled from binding cavities is greater than when less-ordered waters are expelled. From a thermodynamic analysis of the principal hydration sites, we were able to characterize how the various molecular recognition motifs affect the excess chemical potentials of the solvating waters and verify the hypothesis of entropy-driven free energy liberation upon displacing the solvating waters structured by the molecular recognition motifs.

Results and Discussion

Using the above solvent analysis approach, we analyzed the thermodynamic properties of the principal hydration sites of the streptavidin, antibody DB3, and HIV-protease receptors. Figs. 1 and 2 plot the interaction energies versus the calculated excess entropies for water molecules occupying the principal hydration sites of each of the binding cavities studied. Although the majority of data points are clustered together, there are four data points that clearly deviate. These points represent data for hydration sites that have unusually high ordering (large excess entropies). Remarkably, each of these outlying data points corresponds to the hydration of molecular recognition motifs identified in ref. 2. No outlying data points were found for the hydration of HIV-protease, for which no motifs were identified. This finding validates the physical chemical significance of these proposed motifs (i.e., they entropically perturb the binding-cavity solvent) and motivates an atomistic description of their solvation.

The most striking feature of the solvation of the streptavidin binding cavity was the formation of a five-membered water ring. This ring persisted throughout the entire 10 ns of simulation (Figs. 3 and 4). All of the streptavidin hydration sites with outlying entropies are members of this ring. Although energetically favorable, five-membered rings are only fleetingly observed in bulk water because of their unfavorable entropies (7). However, in the streptavidin binding cavity, the ring is stabilized by the topographical characteristics of the motifs found in the binding cavity (see Fig. 3). The three correlated hydrogen bonding groups (Asp-116; Ser-33; and the combination of Asn-11, Tyr-31, and Ser-15) are located in the plane of the ring and are ideally positioned in space such that the water molecules

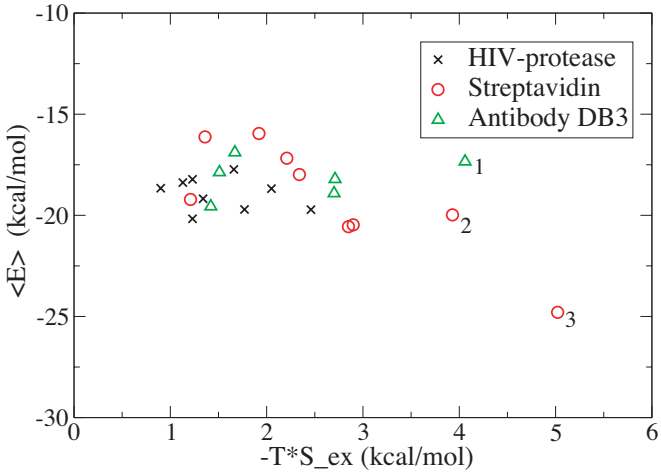

Fig. 1. World energies and excess entropies of water molecules in the principal hydration sites of the binding cavities. The world energy is the energy of interaction of the water molecules with the entire system. Shown are data for principle hydration sites that are proximal to hydrophylic protein groups. The points labeled 1, 2, and 3 represent data for hydration sites with unusually high ordering.

hydrating each group can form unstrained hydrogen bonds with each other when in the five-membered ring configuration. Because the ring is enclosed above and below by hydrophobic groups, the only orientations for which water molecules in this region can maintain the maximal number of hydrogen bonds are those consistent with the ring formation (a water molecule's hydrogen atoms pointed toward hydrophobic groups cannot form hydrogen bonds). The ring configuration was so energetically dominant that no other stable configurations were observed throughout the entire 10 ns of simulation. The reduction in the accessible phase space for these five water molecules solvating the streptavidin binding cavity is what leads to the entropic penalties captured in Figs. 1 and 2.

A simple estimate taken by summing the difference between the calculated excess entropy of the five waters found in the ice-like ring and the average excess entropy of all other active-

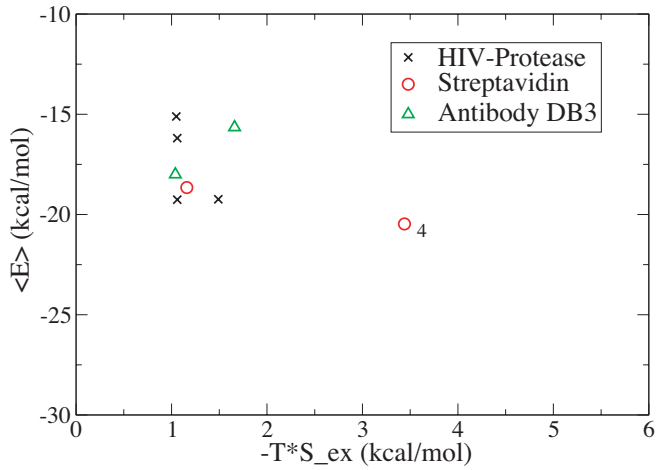

Fig. 2. World energies and excess entropies of water molecules in the principal hydration sites of the binding cavities. The world energy is the energy of interaction of the water molecules with the entire system. Shown are data for principle hydration sites that are proximal to hydrophobic protein groups. The point labeled 4 represents data for a hydration site with unusually high ordering.

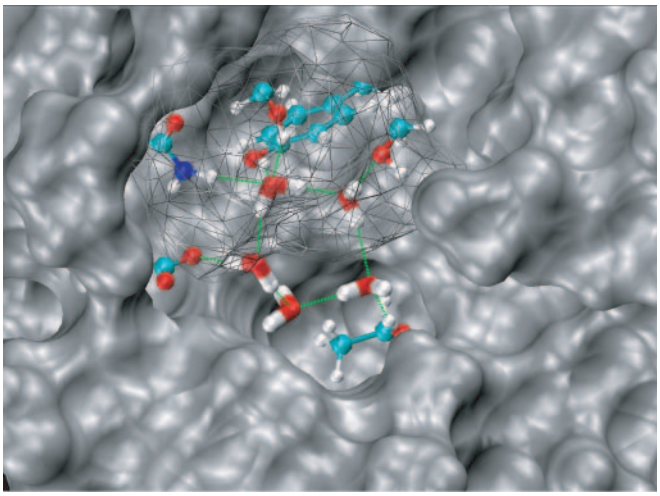

Fig. 3. The binding cavity of streptavidin and a typical solvating water configuration. Also shown is the protein structure that stabilizes the ring. The green lines represent hydrogen bonds. The hydrogen bonds between the ring water molecules and the protein are the correlated hydrogen bonds referred to in the text. The gray scaffolding is the protein that encloses the ring from above.

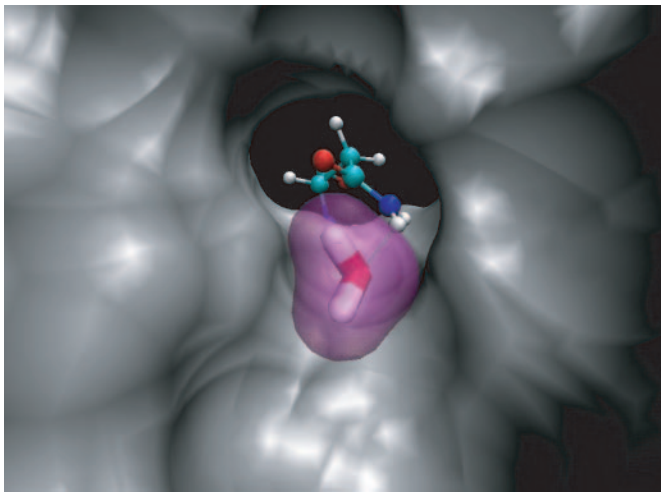

Fig. 5. A typical configuration for a water molecule in principle hydration site 1 from Fig. 1 in the 1DBJ binding cavity. The molecule is orientationally constrained such that its oxygen atom maintains a hydrogen bond with Asn-35. It is also flanked on three sides (to the left, right, and below) by hydrophobic groups. The two hydrogen bond vectors point toward additional solvent with which the water molecule can hydrogen bond. The purple shading is to the scale of a Van der Waals radius for a water molecule.

site waters suggests that the entropic contribution to the free energy of solvent expulsion due to these enclosure effects may be as large as −7 kcal/mol (five orders of magnitude of binding affinity). Several mutagenesis studies support this analysis. When mutated to Ala, residues Asp-128 and Ser-45 exhibit 1.4 kcal of binding cooperativity, driven by a 3.0 kcal/mol entropic term. This finding is consistent with a destabilization of the water ring solvating the active site due to the loss of the hydrogen binding groups maintaining it (8). Circular deletion of the mobile loop formed by residues 47–51 that hydrophobically enclose the ligand from above, most notably by Val-47, loses 8 kcal/mol lower than the unmutated protein; this is much more than continuum methods predict but in agreement with our estimate (9). Mutating Trp-79 to Phe was found to enthalpically stabilize biotin binding by 1.5 kcal/mol but entropically destabilize it by 2.4 kcal/mol (10). This mutation effectively enlarges the cavity and partially removes the hydrophobic enclosure, resulting in more entropically favorable binding-cavity solvation. This result explains why Poisson–Boltzmann-based methods, which cannot capture molecular-length scale solvation physics, underestimate the binding affinity, as measured by the disassociation constant, of the streptavidin–biotin complex by three to six orders of magnitude, whereas explicit solvent simulations predict the binding affinity within chemical accuracy (11–13).

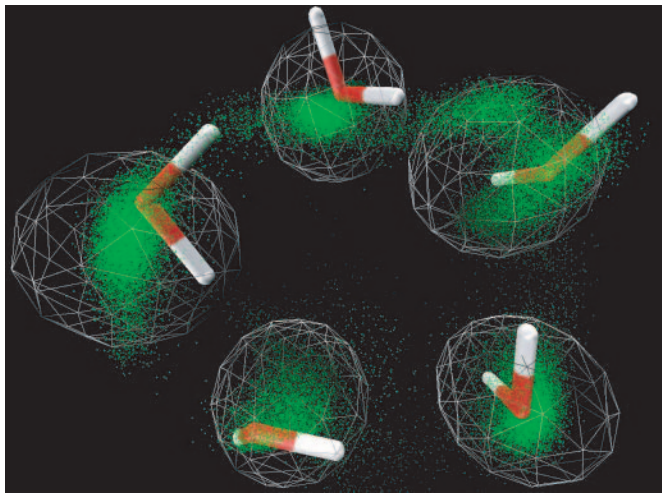

Fig. 4. Shown is the same conformation of the water molecules as is shown in Fig. 3 (from a different perspective) with the solvent density averaged over all simulations in green. The hydration sites determined by the clustering of this density are shown in wireframe. Note the near absence of water molecules in the inner part of the five-membered ring and in between the ring sites.

The outlying hydration site of antibody DB3 is depicted in Fig. 5. This water molecule is hydrophobically enclosed on three sides (below and to the left and right in Fig. 5). On a fourth side, it is bordered by Asn-35, to which it can form a hydrogen bond. To form a hydrogen bond with the protein, the oxygen of the water must face Asn-35, leaving two hydrogen atoms pointing away from the Asp residue and very few orientations that the molecule can take such that both hydrogen atoms can point toward other water molecules and thereby hydrogen bond with them. The orientation shown in Fig. 5 has one hydrogen facing toward the reader and the other hydrogen facing up. This orientation is representative of the most energetically favorable because the molecule can make hydrogen bonds with two other water molecules. If the molecule were rotated significantly about any axis, it would no longer be able to simultaneously form hydrogen bonds with both the protein and its two hydrogen atoms. Therefore, the molecule has very few energetically accessible configurations that leads to entropic penalties of hydration.

The Cox-2 active site was found to contain no persistent hydration sites and is in fact entirely devoid of solvent in 80% of the simulation, despite the cavity sterically accommodating approximately seven water molecules. The high excess chemical potential of the binding-cavity solvent is due to an inability of the water molecules to make hydrogen bonds with the surrounding hydrophobic protein residues and other water molecules. This results in an extreme enthalpic perturbation, >8 kcal/mol, which drives the dewetting of the cavity (Fig. 4). The active site water molecules of an artificially hydrated Cox-2 structure were evacuated within 100 ps of explicitly solvated NPT dynamics. The active site of Cox-2 is predominantly a narrow paraffin-like tube and is therefore in line with other studies of hydrophobically induced dewetting (14–18). What is perhaps most remarkable

about the Cox-2 system is that the active-site cavity does not collapse in the absence of a ligand to form hydrophobic contacts between the enclosing hydrophobic groups of the protein (19). The creation of such a cavity, against the hydrophobic forces promoting collapse, clearly requires substantial evolutionary engineering.

Conclusions

Our simulations suggest that the hydrophobic enclosures found in these systems aid molecular recognition by perturbing the solvation of the binding cavity, which in turn results in a relative stabilization of the bound complex. Such hydrophobic enclosures sterically allow for very few energetically competitive water configurations, yielding entropic penalties of solvation that are not observed in larger ligand cavities. The streptavidin/biotin system demonstrates that severe entropic constraints, without corresponding energy gain, can be manifested even in the presence of polar groups if specific enclosed geometrical requirements are met; such structures constitute a well defined and thermodynamically substantial molecular recognition motif. An extreme case of hydrophobic enclosure is observed in the Cox-2 binding cavity, where no energetically stable solvent configurations appear to exist; insertion of ligand hydrophobic groups into such a region of persistent vacuum will result in substantially larger free-energy liberation than would be expected if the binding cavity were treated as solvated.

Such hydrophobically enclosed regions are compelling targets for drug design because it is possible with suitable ligands to obtain exceptionally large enhancements of potency with a minimal increase in molecular weight. We have also demonstrated here how explicitly solvated trajectories of a receptor can be analyzed by a solvent clustering procedure and inhomogenous solvation theory to identify solvent regions of anomalously unfavorable solvent entropy. This information can actively guide drug design by suggesting regions of the solvent proximal to known active compounds that will be maximally (free-energetically) beneficial to displace by adding new chemical groups to the ligand.

Systems and Simulation

The starting structures of streptavidin [Protein Data Bank (PDB) ID code 1STP], Cox-2 (PDB ID code 1CVU), antibody DB3 (PDB ID code 1DBJ), and HIV protease (PDB ID code 1HPX) were taken from the PDB (20–23). All nonprotein molecules were then removed. For antibody DB3 and Cox-2 systems, we eliminated residues located far away from active sites for computational efficiency. For 1DBJ, residues 109–211 of chain A and 114–228 of chain B were eliminated. For Cox-2, residues 33–84 were removed. Protonation states were assigned assuming the systems are at pH 7.0. Two Asp residues (Asp-25 of both chain A and chain B) within 3 Å of each other in the HIV-protease active site were handled specially. The Asp residue in chain B was protonated and made to hydrogen bond with the Asp residue in chain A, which substantially reduced the strain of the system.

The proteins without the ligands were inserted into water boxes, and water molecules that sterically overlapped with the proteins were removed. The size of each system was chosen to accommodate a minimum of 10 Å of water between the protein surface and the box walls. Counter ions were added to maintain electric neutrality. The systems were then equilibrated for a minimum of 1 ns by using the Nose–Hoover chains thermostat and Andersen–Hoover barostat (24, 25). The OPLS-AA force field (26) was used for the protein, and the TIP4P (27) water model was used for the solvent, with a cut-off of 10 Å for Lennard–Jones interactions and a Particle-Mesh Ewald (28) for electrostatic interactions.

During the molecular dynamics simulations, the proteins' heavy atoms and the counter ions were harmonically restrained to their initial positions. Multiple configurations of each system were sampled from the 1 -s initial run at constant pressure and temperature and used as initial configurations for constant-energy, constant-volume molecular dynamics simulations. Data were taken from >10 ns of simulation time for each system. All simulations were run with the SIM molecular dynamics program, which was developed in the Berne group (29).

In the constant pressure equilibration runs, water quickly filled the vacuum left by the removal of the ligands in all systems except for Cox-2. For Cox-2, it was found that few water molecules entered the protein active site during the equilibration step and left quickly after entering. For this protein, we artificially solvated the active site by transforming seven heavy atom sites of the ligand to water molecules. Starting with the carboxylic acid head of the arachidonic acid (ligand), every third carbon atom was changed to a water molecule (seven total). The remaining ligand atoms were removed. The average spacing between the resulting water molecules was 3.3 Å after this procedure. The structure was locally minimized before simulation. In these artificially hydrated simulations, the inserted water molecules vacated the cavity within 100 ps in the equilibration runs.

Analysis

Much of our analytical effort focused on applying inhomogenous solvation theory to properties of water molecules solvating the active sites of the proteins. Application of inhomogenous solvation theory to systems with an open enclosure where water molecules can exchange with the bulk solvent posed a number of challenges. In particular, the fluctuating number of molecules solvating the areas of interest required a clear and sensible definition of which water molecules would be studied; the rough topography of the active sites made the definition of an orientational frame of reference particularly difficult because, even over smaller subvolumes, the orientational distributions were highly position-dependent; the large size of the protein active sites necessitated the partitioning of the solvent density into well defined subvolumes for both numerical integration of the orientational contribution to the entropies and for physical interpretation.

Binding Cavity. The starting point for each simulation was the protein–ligand complex. The ligand was then removed, and water was allowed to fill the vacated volume. We refer to this vacated volume as the binding cavity. Although the ligand was not simulated in the molecular dynamics runs, the binding cavity volume is defined as any space that lies within 2 Å of any heavy atom of where the ligand would be. The location of the binding cavity remains constant throughout the simulation, and, because the protein is harmonically restrained to its initial positions, it maintains its spatial relation to the protein throughout the simulation.

Density Profile. Throughout the course of the molecular dynamics simulations, any water molecule whose oxygen atom was in the binding cavity at a given time was tagged, and the positions and orientations of these water molecules were recorded. All of these water molecules together provided the water density profile inside each protein's binding cavity. The spatial distribution inside part of the streptavidin cavity is shown in Fig. 4. This distribution was considered to be the equilibrium distribution of water molecules inside the binding cavity and is the distribution functions used for the clustering and inhomogenous solvation theory described below.

Clustering Algorithm. Because of the inhomogeneity of the protein surface, the orientational distribution was highly dependent on the position inside the binding cavities. This necessitated the partitioning of the binding cavities into small subvolumes for which the distributions could be treated as independent of position. We identified subvolumes of the binding cavities with high densities by using a clustering algorithm. This algorithm cycles through the positions of the oxygen atom of every water molecule composing the water density profile in the binding cavity and finds the position that has the greatest number of water neighbors within a 1-Å radius. We denote this position as a principal hydration site and remove it and all of the oxygen positions within 1 Å of it from the solvent density distribution. The process is then repeated, cycling through the remaining positions. This process terminated when a hydration site is found with a water density in the 1-Å sphere that is less than twice that of the expected value in the bulk system. We should note that the 1-Å sphere is small enough such that at any given time only one water molecule occupies a given principle hydration site.

Application of this clustering algorithm resulted in nonoverlapping 1-Å radius spheres corresponding to the regions of high water density in each the binding cavities of the proteins. The five wireframe spheres shown in Fig. 4 encompass five of the principal hydration sites identified by this clustering algorithm for the streptavidin binding cavity. The principal hydration sites are well defined subvolumes of the binding cavities that have ideal convergence properties, i.e., sparse water density near the edges of the cluster, for the inhomogenous solvation theory machinery.

World Energy. The energy of interaction of each water molecule with the entire system was calculated for each water molecule in each principle hydration site. These energies are simply the difference in energy between the system with the water molecule and the system without the water molecule. The average of these quantities for all molecules in each principle hydration site are the energies shown in Figs. 1 and 2.

Inhomogeneous Solvation Theory. We estimated the entropic cost of solvent ordering due to the protein field separately for each hydration site by following the inhomogenous solvation theory of Lazaridis (3). This theory uses an expansion of the entropy in terms of orientational and spatial particle correlation functions (3–6). In this expansion, a uniform bulk density distribution of solvent has an entropy of zero; deviations from this uniform distribution represent structure and result in unfavorable local contributions to the excess entropy.

Our reported values of the excess entropy were calculated by a full numerical evaluation of the first and partial evaluation of the second term in the expansion of the entropy in terms of powers of the density:

$$S_e = -\frac{k_b\rho_\omega}{\Omega}\int g_{sw}(\mathbf{r},\ \omega)\ \ln g_{sw}(\mathbf{r},\ \omega)d\mathbf{r}d\omega$$

$$-\frac{k_b\rho_w^2}{2\Omega^2}\int g_{sww}(\mathbf{r}^2\omega^2)\ \ln\delta g_{sww}(\mathbf{r}^2,\ \omega^2)d\mathbf{r}^2d\omega^2 — \ \ldots\ ,$$

where $\mathbf{r}$ and ω describe the Cartesian position and Euler angle orientation of a water molecule, $g_{sw}(\mathbf{r})$ is the single-body distribution of water (w) at $\mathbf{r}$ and ω in the fixed reference frame of the solute protein (s) and ρ_w is the density of the neat TIP4P system. We will refer to the first term on the RHS of Eq. **1** as the one-body term and the second term as the two-body term.

One-Body Terms. The translational one-body terms were straightforward to evaluate and were numerically integrated by using a length of 0.03 Å for $\mathbf{r}$, 15° along θ, and 30° along ϕ in spherical coordinates. The one-body terms were evaluated independently for each hydration site. To account for possible position-dependence of the orientations within each hydration site, we divided each hydration site into two subvolumes by using a quaternion-based angular clustering algorithm.

The orientational distributions were then obtained by using a numerically exact quaternion formalism. The evaluation of the orientational terms were then numerically integrated with 10° bins for each of the Euler angles.

The algorithm we used to partition each hydration site into two subvolumes clustered water molecules in orientational space with a quaternion distance metric. This procedure required the computation of a master quaternion (q) for each water molecule, where q was defined as the quaternion that rotated the water molecule onto a specified reference water orientation. The distance between each water in angular space was defined as $c = 1 - |q_1 \cdot q_2|$. This metric, derived by Kuffner (30), obeys the triangle inequality, is efficient to compute, and is strictly bounded between 0 and 1, with 0 implying identical orientations. The two largest angular clusters were identified by using a clustering algorithm identical to the radial one described above, except the distance $c \leq 0.1$ was used instead of a radius of 1 Å.

The plane orthogonal to the vector connecting the two cluster centroids and equidistant from the cluster centroids was then used to then divide the hydration site into two subvolumes. This hydration site subdivision was found to add precision to the calculation of rotational entropies when several protein hydrogen bonds were found in close proximity, which in turn caused the electrostatic environment to vary greatly with small changes in intracluster position. For most hydration sites, however, this partitioning had little effect. This partition resulted in the expansion

$$g(\mathbf{r},\ \omega) \simeq g_{sw}^{V_1}(\mathbf{r})g_{sw}^{V_1}(\omega) + g_{sw}^{V_2}(\mathbf{r})g_{sw}^{V_2}(\omega),$$

where V_i refer to the two new subvolumes of each hydration site.

$$g_{sw}^{V_1}(\mathbf{r}) = g_{sw}(\mathbf{r}) \text{ for } \mathbf{r} \in V_1 \text{ and 0 otherwise}$$

$$g_{sw}^{V_2}(\mathbf{r}) = g_{sw}(\mathbf{r}) \text{ for } \mathbf{r} \in V_2 \text{ and 0 otherwise}$$

$$g_{sw}(\mathbf{r}) = g_{sw}^{V_1}(\mathbf{r}) + g_{sw}^{V_2}(\mathbf{r}).$$

With the corresponding normalizations due to the division of space into subvolumes, the first term in the entropy expansion was evaluated as

$$-\frac{k_b\rho_\omega}{\Omega}\int g_{sw}(\mathbf{r},\ \omega)\ \ln g_{sw}(\mathbf{r},\ \omega)d\mathbf{r}d\omega$$

$$\simeq -k_b\rho_\omega\int g_{sw}(\mathbf{r})\ \ln g_{sw}(\mathbf{r})d\mathbf{r}$$

$$-\frac{k_bN_\omega^{V_1}}{\Omega}\int_{V_1} g_{sw}^{V_1}(\omega)\ \ln g_{sw}^{V_1}(\omega)d\omega$$

$$-\frac{k_bN_\omega^{V_2}}{\Omega}\int_{V_2} g_{sw}^{V_2}(\omega)\ \ln g_{sw}^{V_2}(\omega)d\omega.$$

The orientational distribution functions were assumed to be invariant within each subvolume of the hydration sites. The orientational distribution functions were then obtained by using a mixed quaternion/Euler angle method for which the quaternions needed to rotate all of the water molecules onto the reference water molecule were analytically computed, the Euler

angle was analytically extracted, and the resulting Euler angle distribution was integrated numerically.

Two-Body Terms. Because of the limitation in the data, we were not able to fully evaluate the two-body terms in the entropy expansion. Instead, we limited ourselves to calculating correlations in the interparticle distances between water molecules in adjacent hydration sites and correlations in the hydrogen–oxygen–oxygen angles formed between water molecules in adjacent hydration sites. The contribution of these terms are the part that is not described by the one-body terms. The overall contribution of these terms was relatively small, and the two-body terms were mainly determined by the product of the one-body terms.

1. Wang R, Lu Y, Fang X, Wang S (2004) *J Chem Inf Comp Sci* 44:2114–2125.
2. Murphy RB, Halgren TA, Friesner RA (2006) *J Med Chem* 49:6177–6196.
3. Lazaridis T (1998) *J Phys Chem B* 102:3531–3541.
4. Baranyai A, Evans D (1989) *Phys Rev A* 40:3817–3822.
5. Morita T, Hiroike K (1961) *Prog Theor Phys* 25:537–578.
6. Lazaridis T, Paulaitis ME (1992) *J Phys Chem* 96:3847–3855.
7. Rahman A, Stillinger F (1973) *J Am Chem Soc* 95:7943–7948.
8. Hyre D, Trong I, Merritt E, Eccleston J, Green N, Stenkamp R, Stayton P (2006) *Protein Sci* 15:459–467.
9. Chu V, Freitag S, Trong I, Stenkamp R, Stayton P (1998) *Protein Sci* 7:848–859.
10. Chilkoti A, Stayton P (1995) *J Am Chem Soc* 117:10622–10628.
11. Lazaridis T, Masunov A, Gandolfo F (2002) *Proteins Struct Funct Genet* 47:194–208.
12. Luo H, Sharp K (2002) *Proc Natl Acad Sci USA* 99:10399–10404.
13. Dixit S, Chipot C (2001) *Proc Natl Acad Sci USA* 105:9795–9799.
14. Wallqvist A, Gallicchio E, Levy RM (2001) *J Phys Chem B* 105:6745–6753.
15. Huang X, Margulis C, Berne B (2003) *Proc Natl Acad Sci USA* 100:11953–11958.
16. Sriraman S, Kevrekidis I, Hummer G (2005) *Phys Rev Lett* 95:13603.
17. Wolde P, Chandler D (2002) *Proc Natl Acad Sci USA* 99:6539–6543.
18. Collins M, Hummer G, Quillin M, Matthews B, Gruner S (2005) *Proc Natl Acad Sci USA* 102:16668.
19. Kurumbail RG, Stevens AM, Gierse JK, Joseph J McDonald RAS, Pak JY, Gildehaus D, iyashiro JM, Penning TD (1996) *Nature* 384:644–648.
20. Weber P, Ohlendorf D, Wendoloski J, Salemme F (1989) *Science* 243:859–863.
21. Kiefer J, Pawlitz J, Moreland K, Stegeman R, Hood W, Gierse J, Stevens A, Goodwin D, Rowlinson S, Marnett L, *et al*. (2000) *Nature* 405:97–101.
22. Arevalo J, Taussig M, Wilson I (1993) *Nature* 365:859–863.
23. Baldwin E, Bhat T, Gulnik S, Liu B, Topol I, Kiso Y, Mimoto T, Mitsuya H, Erickson J (1995) *Structure (London)* 3:581–590.
24. Tuckerman ME, Martyna GJ, Klein ML (1992) *J Chem Phys* 97:2635.
25. Andersen HC (1980) *J Chem Phys* 72:2384.
26. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) *J Phys Chem B* 105:6474.
27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein M (1983) *J Chem Phys* 79:926.
28. Darden T, York D, Pedersen L (1993) *J Chem Phys* 98:10089.
29. Stern H, Berne BJ, Rittner F, Pavese M, Harder E, Xu H, Kim B (2001) SIM: Molecular Dynamics Simulation Program (Columbia University, New York).
30. Kuffner J (2004) *Proc IEEE Int Conf Robot Automat* (Inst Electric Electron Eng, Piscataway, NJ).