JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

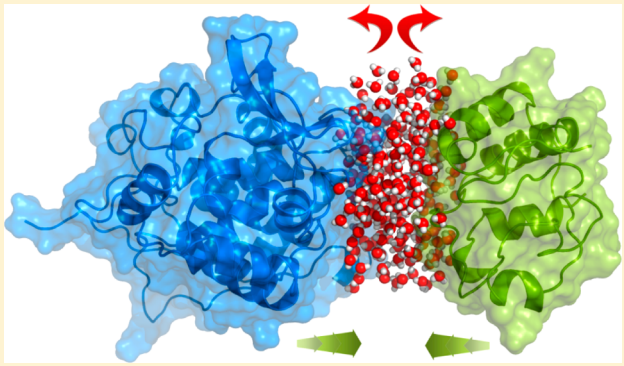
Energetics of Hydrophilic Protein−Protein Association and the Role of Water

Ozlem Ulucan, Tanushree Jaitly, and Volkhard Helms*

Center for Bioinformatics, Saarland University, Saarbruecken, Germany

Supporting Information

ABSTRACT: Hydrophilic protein−protein interfaces constitute a major part of all protein−protein interfaces and are thus of great importance. However, the quantitative characterization of their association is still an ongoing challenge and the driving force behind their association remains poorly characterized. Here, we have addressed the association of hydrophilic proteins and the role of water by means of extensive molecular dynamics simulations in explicit water using three well studied protein complexes; Barnase−Barstar, cytochrome c−cytochrome c peroxidase, and the N-terminal domain of enzyme I-histidine-containing phosphocarrier. The one-dimensional free energy profiles obtained from umbrella sampling simulations are downhill or, in other words, barrierless. Using these one-dimensional free energy profiles, the computed standard free energies of binding are −12.7 ± 1.1 kcal/mol, −9.4 ± 0.7 kcal/mol, and −8.4 ± 1.9 kcal/mol that are in reasonable to very good agreement with the experimental values of −19.6 kcal/mol, −8.8 kcal/mol, and −7.8 kcal/mol. As expected, analysis of the confined water between the hydrophilic complex partners shows that the density and the orientational order parameter deviate noticeably from the bulk values, especially at close separations of the confining proteins.



INTRODUCTION

The capability of proteins to bind each other in a specific manner is essential for a wide variety of biological processes such as the cell cycle, cellular transport, immune response, apoptosis, DNA replication and transcription, and RNA splicing and signal transduction. Even though many proteins perform their functions independently, a large part of all proteins interact with others for proper biological activity. Based on large-scale proteomics studies, it has been estimated that about half of all cellular proteins are permanently or transiently involved in protein complexes where they form on average 6−8 interactions with other proteins.[1] Studies that analyzed the residue composition of protein−protein interfaces showed that protein−protein interfaces are enriched in both charged and polar residues rather than nonpolar residues.[2,3] Yet, hydrophobic residues are found to be scattered over the entire interface where they form small hydrophobic patches within polar and charged residues.[4]

Similar to the protein folding process, the decrease in the total Gibbs free energy (at constant temperature and pressure) of the protein−protein−solvent system upon binding is accompanied by opposing roles of entropy and enthalpy. For the protein folding process, the requirement for lowering the free energy while reducing the conformational space ensures that the energy landscape is funnel-shaped.[5,6] Analogously, spontaneous protein−protein association also lowers the free energy of the full system while reducing the conformational, rotational, and translational entropies of the binding partners.[7] Decades of experimental and theoretical efforts in this field have led to the development of paradigms, theoretical methods, and computational tools. An established concept is the calculation of the standard free energy of binding from a one-dimensional potential of mean force (PMF). The potential of mean force is the work required to bring two particles to a particular relative separation r. Its gradient gives the average force, including direct and indirect contributions. PMF calculations provide a reliable method for determining the absolute free energies of binding of protein−ligand and protein−protein systems.[8] The commonly used method for this purpose is umbrella sampling[9] with the Weighted Histogram Analysis Method (WHAM).[10] This method is able to connect the mechanistic details of the binding process with the underlying free energy surface. Application of these methods is relatively straightforward for small systems. However, it is an ongoing challenge for large systems such as protein−protein complexes.

Another important factor is the water solvent that plays a crucial role in governing the structure, stability, dynamics, and function of biomolecules. In many natural systems, water is confined in an environment where its free movement is restricted and its three-dimensional hydrogen-bonded network

Received: February 28, 2014
Published: June 23, 2014

   dx.doi.org/10.1021/ct5001796 | J. Chem. Theory Comput. 2014, 10, 3512−3524

is partially disrupted. The properties of the confined water are difficult to predict and may be considerably different from those of bulk water. For example, hydrophobic dewetting has been reported as a general mechanism for the association of hydrophobic surfaces.[11] Our understanding how hydrophilic surfaces assemble is less clear. There is some evidence, for example, by McLain et al.,[12] that the association of hydrophilic surfaces results from the direct interactions between the binding partners. Based on neutron diffraction data, these authors showed that the association of some small peptides in aqueous solution is dominated by charge−charge interactions among the solutes. On the contrary, based on thermodynamic models, Ben-Naim argued that solvent-induced interactions play a dominant role for protein−protein association.[13] Using model systems, Berne and co-workers[14] addressed this issue in detail. In their study, plates characterized by large hydrophobic domains interacted via attractive solvent induced interactions. In contrast, a homogeneous distribution of hydrophobic and hydrophilic particles on the plates produced repulsive solvent induced interactions. Ahmad et al.[15,16] extensively studied the process of protein−protein association by unbiased MD simulations in explicit solvent. When studying the spontaneous binding of a proline-rich peptide to an SH3 domain they found a clear dewetting transition upon binding of this hydrophobic interface.[15] With respect to hydrophilic interfaces, they observed that the nature of the water confined in the interfacial gap volume between the hydrophilic protein interfaces of the Barnase−Barstar pair deviated significantly from bulk behavior.[16]

Taking three well-studied protein−protein complexes as model systems, we revisit here the issue of hydrophilic protein−protein association and the role of water using extensive molecular dynamics simulations. The studied systems are Barnase−Barstar (BN−BS), cytochrome c−cytochrome c peroxidase (CC−CCP), and the complex of the N-terminal domain of enzyme I with histidine-containing phosphocarrier (EIN−HPr). The three model complexes were selected due to the availability of a wealth of kinetic and structural data, due to the demonstrated importance of electrostatic interactions on their association, and because none of them exhibits large conformational changes upon complexation.

BN−BS is one of the best studied protein−protein complexes. Barnase is an bacterial RNAase that is lethal to the cell when expressed without its intracellular inhibitor Barstar.[17] The association of the BN−BS pair was addressed extensively in previous studies using, for example, Brownian dynamics (BD) and molecular dynamics (MD) simulations. Using BD simulations, Gabdoulline and Wade[18] computed association rates and their dependence on ionic strength and protein mutations for the BN−BS system. Spaar and Helms[19] characterized the long-range free energy landscape of the BN−BS encounter complexes using BD simulations. Using an analogon of transition-state theory and atomic-detail modeling of proteins, Zhou and colleagues[20] calculated the electrostatic enhancement of the association rate of BN−BS and obtained results that correlated well with experiments. Recently Gumbart et al.[21] calculated the standard binding free energy of the BN−BS system from atomistic MD simulations in explicit solvent. Their newly proposed methodology relies upon PMF calculations where the proteins are restrained in the conformation, relative position, and orientation of the bound state.

Both cytochrome c and cytochrome c peroxidase are located in the intermembrane space of mitochondria. Utilizing two molecules of ferrous CC as specific electron source, CCP catalyzes the two-electron reduction of alkyl hydroperoxides.[22] Using BD simulations, Northrup et al.[23] studied the diffusional association of the CC−CCP pair, and Gabdoulline and Wade[24] investigated the factors that influence association rates of five different protein−protein complexes including the CC−CCP pair as model systems. They found, for example, that the CC−CCP association rate is fast enough to support a two-step electron transfer mechanism.

In bacterial cells, the phosphorylation and the translocation of sugars are coupled by the phosphoenolpyruvate (PEP)−sugar phosphotransferase system (PTS) that consists of two cytosolic proteins, namely enzyme I and HPr, as well as of sugar specific components. The complex between EIN and HPr is a classical example of surface complementarity.[25] Using data from paramagnetic relaxation enhancement and replica exchange simulations, Hummer and co-workers[26] studied the transient encounter complexes in the EIN−HPr association. They reported that, besides the specific complex, distinct nonspecific complexes exist as well that account for ∼10% of relative population.

Here, we present the results from umbrella sampling simulations in explicit water and simulations of water localized between the two proteins for the three systems mentioned above. The one-dimensional free energy profiles of protein−protein association were found to be downhill. Using these one-dimensional free energy profiles, the computed standard free energies of binding are in overall good agreement with the experimental values. Decomposition of the free energy of binding revealed that the direct nonbonded interactions between the complex partners favor the association whereas the solvent-induced interactions turned out to be repulsive. Moreover, the density and the orientational order parameter of confined water deviate noticeably from the bulk values at close separation of the confining proteins.

METHODS

The coordinates for the bound protein−protein complexes were retrieved from the protein databank:[27] Barnase−Barstar (PDB ID: 1BRS[17]), cytochrome c−cytochrome c peroxidase (PDB ID: 2PCC[22]), and the amino terminal domain of enzyme I and the histidine-containing phosphocarrier protein (PDB ID: 3EZB[25]).

Parameterization of Proteins. The titration states of titratable amino acids were assigned at physiological pH using the program PROPKA (http://nbcr-222.ucsd.edu/pdb2pqr_1.8/).[28] All crystallographically resolved water molecules were kept and the placement of additional water molecules in internal protein cavities was tested using the program DOWSER.[29] and keeping only those with DOWSER energy below −12 kcal/mol. All interactions were modeled by the Amber force field FF99SB-ILDN.[30] Short range nonbonded interactions were computed up to 1.2 nm distance. Long range electrostatic interactions were treated by the particle mesh Ewald (PME)[31] method. Dispersion correction was applied to energy and pressure. Periodic boundary conditions were applied in all directions. Water molecules were modeled by the TIP3P[32] potential that is typically used together with the AMBER force field. All simulations were carried out using the GROMACS package, version 4.5.4.[33]

Heme Center Parameterization. The active sites of CC and CCP contain a heme group each. In CC, the heme is covalently bonded to the polypeptide chain and the iron atom in both oxidized and reduced states is six-coordinate low spin.[34] The central heme iron is ligated to both a histidine and methionine. CCP, in its resting state, involves a noncovalent heme with a five-coordinate high-spin iron.[34] Here, the central iron atom is coordinated by a histidine residue at the fifth position. In our study, we considered CC and CCP in their resting states where the iron atoms are Fe(II) and Fe(III), respectively.

Even though the coordinate set 2PCC (PDB ID) was used to start the MD simulations, we retrieved coordinates for the heme group and its coordinating residues from crystal structures determined at higher resolution as starting positions for the quantum mechanical calculations. The PDB entries 1YYC (1.23 Å) and 1ZBY (1.20 Å) were used for CC and CCP, respectively. The amino acids bonded to the heme centers were truncated at their β-carbons and hydrogens were added. All QM calculations were performed using Gaussian03.[35] For derivation of partial atomic charges, we followed the standardized protocol commonly used in combination with the original AMBER94 force field with some minor changes. The geometry was optimized at B3LYP level using the basis set 6-31G* in two consecutive stages. First, we optimized the geometry of the heme group alone. Then, the coordinating residues were added to the resulting heme configuration in their conformations observed in the crystallographic structure and the geometry of the full system was further optimized without any restraints. Using the optimized geometry, we obtained the molecular electrostatic potential from the HF/6-31G* electron density of the heme centers. Restricted ESP (RESP) charges[36] were obtained using the RESP program under Amber Tools in two steps. In the first stage, charge equivalency on chemically equivalent heavy atoms was imposed and the total charges on the methyl groups that were generated by adding hydrogen atoms to the Cβ atoms of the heme coordinating amino acids were set to zero. In the second stage, the charges of chemically equivalent hydrogen atoms were equated, keeping the constraint on methyl groups and the charges on heavy atoms constant. The remaining excess charge obtained when linking the amino acid Cβ to the rest was equally distributed over all atoms included in the parametrization, ensuring that the overall charge on the heme and the newly defined amino acids is integral.

The missing force constants for bonds, angles and dihedrals of the heme groups were taken from heme parameters in the AMBER parameter database, see http://www.pharmacy.manchester.ac.uk/bryce/amber/ and Shahrokh et al.[37] The derived charges and the used force constants for the heme centers are provided in the Supporting Information.

Molecular Dynamics Simulations. In this study, we conducted two different sets of molecular dynamics simulations for the three protein−protein complexes in explicit solvent. First, we combined umbrella sampling and the weighted histogram analysis method to characterize the one-dimensional binding free energy surface of the three systems. In order to generate equilibrated initial structures for the simulations, each system was placed in a cubic box of TIP3P water. To mimic physiological conditions 100 mM NaCl was added, including neutralizing counterions. This resulted in 50 Na^+ and 46 Cl^- for the BN−BS system, 83 Na^+ and 82 Cl^- for the CC−CCP system and 107 Na^+ and 86 Cl^- for the EIN−HPr system. Following an initial energy minimization of 1000 steps of steepest descent, each system was equilibrated in two steps where the heavy atoms of the proteins were restrained using a force constant of 1000 kJ mol^{-1} nm^{-2}. The first step involved 100 ps of MD in the NVT ensemble, maintaining the temperature at 310 K. Protein and nonprotein atoms were coupled separately to temperature baths using Berendsen's weak coupling algorithm[38] with a coupling time of 0.1 ps. All bonds were constrained using LINCS algorithm.[39] Subsequently, 100 ps of NPT equilibration were performed, keeping the pressure at 1 bar also using Berendsen's weak coupling method[38] with a time constant of 1 ps. During data collection, the Nosé−Hoover thermostat[40,41] was combined with the Parrrinello−Rahman barostat[42] to regulate temperature and pressure, respectively. Equilibration was completed by 20 ns of conventional MD simulation in the absence of any restraints. For integrating Newton's equations of motion, the leapfrog algorithm with a time step of 2 fs was used.

The final coordinates at the end of these trajectories were used as starting configurations for umbrella sampling simulations. The protein coordinates were rotated in order to align the line connecting the centers of mass of the two complex partners with the z-axis. Then, the two proteins were placed in a rectangular box with dimensions sufficient to satisfy the minimum image convention even at the largest separation distance of 3 nm. The solvation box sizes were 6.7 nm × 7.7 nm × 14.7 nm, 9.0 nm × 9.4 nm × 16.0 and 11.0 nm × 7.8 nm × 16.3 nm for BN−BS, CC−CCP, and EIN−HPr, respectively. The initial configurations were generated by translating one of the protein partners along the z-axis up to 3 nm distance between the surfaces of the two proteins, while keeping the other one fixed. The disassociation path between the bound and the 3.0 nm separated states was divided into 0.1 nm intervals up to 1 nm of separation and 0.2 nm intervals between 1 and 3 nm. This resulted in 21 windows. For each window, an umbrella sampling run of 10 ns length was performed with a force constant of 1000 kJ/mol·nm^2 in general. The force constant was increased to 2000 kJ/mol·nm^2 in cases when the protein centers in the windows deviated strongly from the initial configurations and therefore caused sampling problems. For the BN−BS system each window was extended to 40 ns to check whether this led to a better agreement with the experimental binding free energy. The first 0.5 ns of all windows were considered as equilibration and excluded during analysis. Different time intervals were utilized for construction of the PMF curves.

For WHAM analysis, the g_wham[43] utility of GROMACS 4.5.4 was used with default options, except for the convergence tolerance that was set to 10^{-9}. The histograms were carefully analyzed to ensure sufficient overlapping. New windows were added when the overlaps between the histograms were not sufficient and existing windows were deleted in those cases to avoid redundancy. The histograms of the final umbrella windows are displayed in Supporting Information Figure SI 1.

Standard Free Energy of Binding from PMF. In order to calculate the standard free energy of binding, ΔG°, we followed the strategy presented by Henchman and co-workers.[44] They assumed that a PMF is sampled along a 1D reaction coordinate, whereas concurrently the orthogonal translational movement is restricted by a harmonic confinement potential. No restraints are applied to the orientations. The standard free energy of binding is computed as the sum of three terms.

$$\Delta G^{\circ} = \Delta G_{PMF} + \Delta G_{V} + \Delta G_{R}$$

where ΔG_{PMF} stands for the binding free energy change obtained as the difference between the bound and unbound states retrieved from the PMF, ΔG_{V} stands for the free energy change from the unbound volume to the standard state volume and ΔG_{R} accounts for the change in free energy associated with the introduction of translational confinement restraints.

ΔG_{PMF} and ΔG_{V} are computed as

$$\Delta G_{PMF} = -RT \ln\left(\frac{Q_b}{Q_u}\right)$$

and

$$\Delta G_{v} = -RT \ln\left(\frac{V_u}{V_o}\right)$$

where, R is the ideal gas constant, T is the absolute temperature, V_u is the unbound volume, and V_o is the standard state volume. Q_b and Q_u are the partition functions for the bound and unbound regions, respectively. Their ratio is computed by the following equation:

$$\frac{Q_b}{Q_u} = \frac{l_b}{l_u} e^{(-\Delta W/RT)}$$

Here, the PMF depth, ΔW, is defined as the lowest point minus the exponential average over the entire unbound region of the PMF. $W(z)$ is the PMF as a function of z and defined to be zero at its lowest point when the proteins are bound.

$$\Delta W = RT \ln\left[\frac{\int_{unbound} e^{-W(z)/RT} dz}{\int_{unbound} dz}\right]$$

l_b and l_u are the configurational integral of the PMF and are given by the following equations:

$$l_b = \int_{bound} e^{-W(z)/RT}\, dz$$

and

$$l_u = \int_{unbound} dz$$

The unbound volume V_u is the area explored in the xy plane times the distance sampled along the z axis, l_u, by the protein and computed as follows:

$$V_u = l_u \frac{2\pi RT}{k_{xy}}$$

where k_{xy} is the force constant of the applied harmonic restraint potential along x and y directions.

The free energy term ΔG_R to remove the orthogonal restraints in the bound state is computed from additional 10 or 40 ns long umbrella windows without orthogonal restraints, which cover the bound region along the reaction coordinate, using the following equation

$$\Delta G_R = RT \ln\langle e^{-k_{xy}(\Delta x^2+\Delta y^2)/2RT}\rangle_{k_{xy}=0}$$

Here, Δx and Δy are the observed displacements relative to the minimum in the restrained simulations.

In this study, we chose the cutoff between the bound and unbound regions to be the value of z where the PMF becomes constant within statistical noise.

Rotational Entropy Calculation. We calculated the rotational entropy based on the distribution of the orientations of BS with respect to BN, of CC with respect to CCP, and of HPr with respect to EIN.

For calculating the rotational entropy, we used the umbrella sampling trajectories after removing the rotational motion of one of the complex partners. First of all, the reaction coordinate was divided in equal bins of 0.1 nm length (distance bins). Then, the snapshots were assigned to the distance bins according to the protein−protein COM distance. Afterward, for each distance bin we calculated the entropy value as explained below.

The rotation matrices needed for calculating protein orientations were obtained by the g_rotmat routine of GROMACS 4.5.4 that is based on least-squares fitting. Utilizing the resulting matrices the three Euler angles (ϕ, θ, ψ) were computed.[45] The sampled distribution of the Euler angles was used to compute the entropy as follows:[46]

$$S^{rot} = -R\int_{\phi=0}^{360} p(\phi)\ln p(\phi)d\phi - R\int_{\theta=0}^{180} p(\theta)\ln p(\theta)\sin\theta d\theta - R\int_{\psi=o}^{360} p(\psi)\ln p(\psi)d\psi$$

In order to define the states for the entropy calculation, the range for each Euler angle was equally divided in angular bins of 3.6° bin size. For every angular bin we counted the observed frequency, p. Summing up the contribution of each state according to the above formula gave the rotational entropy. Instead of the absolute rotational entropy we reported the change in rotational entropy, ΔS^{rot}, relative to the uniform distribution of the three Euler angles what corresponds to the ideal freely rotating case.

Entropy of Binding. The entropy loss of one of the protein partners upon association was also estimated using the SF (system-frame) method introduced by Irudayam and Henchman[47] for protein ligand systems. Here we give a brief description of the method. For the detailed derivations please see reference.[47] Even though in our systems both binding partners are proteins, we refer to the smaller binding partner (Barstar, cytochrome c and histidine-containing phosphocarrier) as ligand. We ignored the change in internal entropy and, thus, assume that the ligand only loses translational and rotational entropy upon complexation. In the unbound state, the translational entropy of the ligand, $S^{o,tr}_{L(aq)}$ is the sum of vibrational and cratic entropies and is given by

$$S^{o,tr}_{L(aq)} = R \ln\left(\frac{1}{x^{o}_{L(aq)}}\prod_{i=1}^{3}\frac{2k_{\beta}Te}{F^{i}_{L(aq)}\Lambda_L}\right)$$

where $x^{o}_{L(aq)}$ is the mole fraction of L, e is the natural logarithm base, $F^{i}_{L(aq)}$ is half of the average force magnitudes along the principal axes of the ligand and Λ_L is the translational thermal de Broglie wavelength. The rotational entropy of the ligand in the unbound state consists of orientational and librational terms and is given by

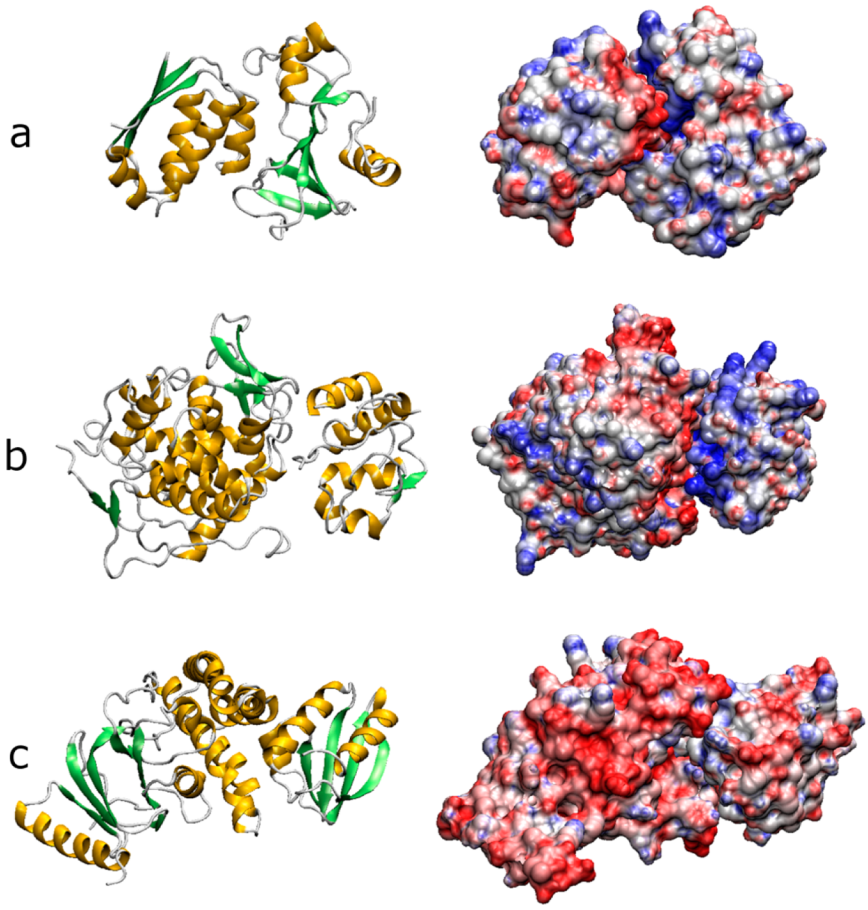


Figure 1. Cartoon (left panel) and electrostatic potential surface representation (right panel) of the studied complexes. (a) Barnase−Barstar, (b) cytochrome c−cytochrome c peroxidase, (c) N-terminal domain of enzyme I−histidine-containing phosphocarrier. The electrostatic potential was computed utilizing the Adaptive Poisson−Boltzmann Solver (APBS)[66] setting the dielectric constant to 78.5 and 2.0 for the solvent and the proteins, respectively. The electrostatic potential was mapped to the protein surface using the VMD software.[67] The color scale data ranges from $-7k_BT$ (red) to $+7k_BT$ (blue).

$$S^{\mathrm{rot}}_{\mathrm{L(aq)}} = R\ \ln\left(\frac{8\pi^2}{\sigma_{\mathrm{L}} V_{\mathrm{w(l)}}} \prod_{i=1}^{3} \frac{2k_\beta Te}{r^i_{\mathrm{L(aq)}} \tau^i_{\mathrm{L(aq)}} \Lambda^i_{\mathrm{L}}}\right)$$

In the above equation, σ_{L} is the symmetry number, $V_{\mathrm{w(l)}}$ is the volume of a single water molecule, r^i_{L} is the radius of the ligand along the ith principal axis, τ^i_{L} is half of the torque magnitude about the ith principal axis of the ligand and Λ^i_{L} represents the rotational thermal de Broglie wavelengths.

In the bound state the ligand was assumed to have no cratic and orientational entropy. Hence, the translational entropy of the bound ligand, $S^{\mathrm{tr}}_{\mathrm{(comp,aq)}}$, is only vibrational

$$S^{tr}_{\mathrm{(comp,aq)}} = R\ \ln\left(\prod_{i=1}^{3} \frac{2k_\beta Te}{F^i_{\mathrm{L(comp,aq)}} \Lambda_{\mathrm{L}}}\right)$$

and the rotational entropy, $S^{\mathrm{rot}}_{\mathrm{(comp,aq)}}$, is solely librational.

$$S^{\mathrm{rot}}_{\mathrm{(comp,aq)}} = R\ \ln\left(\prod_{i=1}^{3} \frac{2k_\beta Te}{\tau^i_{\mathrm{(comp,aq)}} \Lambda^i_{\mathrm{L}}}\right)$$

Taking the difference of the bound and free ligand entropies and assuming that the thermal rotational de Broglie wavelengths do not change between solution and the complex, the final equations for translational and rotational entropy changes upon complexation are as follows:

$$\Delta S^{\mathrm{tr}}_{\mathrm{L}} = R\ \ln\left(\frac{V_{\mathrm{w(l)}}}{V^o} \prod_{i=1}^{3} \frac{F^i_{\mathrm{L(aq)}}}{F^i_{\mathrm{L(comp,aq)}}}\right)$$

$$\Delta S^{\mathrm{rot}}_{\mathrm{L}} = R\ \ln\left(\frac{\sigma_{\mathrm{L}} V_{\mathrm{w(l)}}}{8\pi^2} \prod_{i=1}^{3} \frac{\tau^i_{\mathrm{(aq)}}}{r^i_{\mathrm{L(aq)}} \tau^i_{\mathrm{(comp,aq)}}}\right)$$

The average force and torque magnitudes were extracted from the first and last windows of umbrella sampling simulations corresponding to the bound and unbound states, respectively.

Properties of Interfacial Water. Subsequently, we performed a second set of MD simulations for the three protein−protein systems to characterize the density and the tetrahedral order parameter of the interfacial water localized between the two protein interfaces. These simulations were performed under the same conditions used in the umbrella sampling simulations. Additionally, harmonic position restraints were applied to the backbone atoms of the proteins to maintain their interfacial distance and a fixed relative orientation. The starting configurations were obtained in the same way as previously mentioned for the umbrella sampling simulations. The interfacial distance was varied between 0.35 and 5.0 nm. Each simulation was 20 ns long. The snapshots for analysis were collected every 0.25 ps. Analyses were carried out using

dx.doi.org/10.1021/ct5001796 | J. Chem. Theory Comput. 2014, 10, 3512−3524

Table 1. Some Global and Interface Properties of the Three Protein−Protein Complexes[a]

	BN−BS	CC−CCP	EIN−HPr
no. of amino acids in protein I	110	108	249
no. of amino acids in protein II	89	296	85
area of binding interface (Å^2)	778	570	1002
no. of interface residues in protein I	16	13	33
no. of interface residues in protein II	14	10	24
no. of H-bonds across interface	14	4	6
no. of salt-bridges across interface	12	2	5
total charge of protein I [e]	+2	−7	−19
total charge of protein II [e]	−6	+6	−2
total charge of interface I [e]	+3	+5	−5
total charge of interface II [e]	−4	−2	+4
binding constant	1.3×10^{-14} M^{-1}[b]	6×10^{-7} M^{-1}[c]	3.1×10^{-6} M^{-1}[d]

[a]The interface area and number of interface residues were retrieved from the ABC^2 database.[48] The number of H-bonds and salt-bridges across the interfaces were taken from the PDBe Pisa database (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). [b]The disassociation constant was retrieved from reference.[68] It was measured at 25 °C, in pH 8 and 100 mM NaCl. [c]The disassociation constant was taken from reference.[69] It was measured at 25 °C, in pH 7 and 104 mM ionic strength. [d]The disassociation constant was converted from the association constant reported in reference.[70] It was measured at 30 °C, in pH 7.5 and 100 mM KCl.

the snapshots from the last 15 ns. A cubic box of pure water was also simulated in order to compare the properties of interfacial water to the bulk values.

Interfacial Gap Definition. The interface residues on the protein surfaces were retrieved from the ABC^2 database.[48] There, protein interfaces contain those residues exhibiting a certain change in their solvent accessible surface area (SASA) when comparing the values of the unbound state to those of the complexed forms. The interfacial gap was defined by a rectangular box, which spans the center of geometry of both interfaces along the z-axis (the length). The width and height of the box were calculated by considering minimum and maximum coordinate values of the interfaces along x- and y-axes. The values, which gave the smaller dimensions, were taken for defining the interfacial gap. Those water molecules, having oxygen positions inside the interfacial gap, were considered as interfacial water molecules. The volume of the interfacial water gap was calculated by subtracting the overlapping protein volume from the volume of the aforementioned rectangular box. The overlapping protein volume was calculated from the protein mass inside the rectangular box and assuming an average value of the protein density. For the protein density, we used the generally accepted value of 1.35 g/cm^3 that was deduced from hydrodynamic[49,50] and adiabatic compressibility[51] experiments. This procedure was applied to all frames along a trajectory and the quantities were averaged subsequently.

Tetrahedral Order Parameter. Water molecules have a general propensity for tetrahedral coordination, owing to their hydrogen-bonding network. The tetrahedral order parameter is a three-body order parameter that measures the degree to which the nearest-neighbor molecules are tetrahedrally coordinated with respect to a given molecule.[52,53] The tetrahedral order parameter is defined as

$$q = 1 - \frac{3}{8}\sum_{i=1}^{3}\sum_{j=i+1}^{4}\left(\cos\psi_{ikj} + \frac{1}{3}\right)^2$$

where ψ_{ikj} is the angle formed by the lines connecting the oxygen atoms of a given water molecule k and those of its nearest neighbors i and j. We computed the orientational order parameter in two different ways: first, we only considered water molecules as potential neighbors; second, in addition to the water molecules, we also took into account the nearby protein oxygen and nitrogen atoms within a cutoff 3.5 Å as potential neighbors. In both cases, we considered the four nearest neighbors of interfacial water molecule k, based on their Euclidian distances and calculated the orientational order parameter using the formula above. These calculations were performed for each interfacial water molecule k and then averaged.

RESULTS AND DISCUSSION

Global and Interface Properties of the Systems of Interest. In this work we studied the association/dissociation of three well studied protein−protein complexes; Barnase−Barstar, cytochrome c−cytochrome c peroxidase and the complex of the N-terminal domain of enzyme I with histidine-containing phosphocarrier. Figure 1 shows cartoon and electrostatic surface representations of the systems studied. Visual inspection easily reveals the favorable electrostatic complementarity between the protein partners, especially of the BN−BS and CC−CCP complexes.

Table 1 summarizes general and interface properties of the three model systems. As seen in Table 1, the proteins and, more importantly, the interfaces carry nonzero electrostatic net charges at pH = 7. The charge values given in Table 1 were computed after assignment of the titration states of amino acids by PROPKA[28] and may thus differ from the standard charge values reported in the literature. In all systems, the interfaces are oppositely charged. In the BN−BS and CC−CCP complexes, the proteins also have an opposite overall charge whereas in the EIN−HPr complex both proteins carry a negative overall charge. Another noteworthy property is the area of the binding interfaces. The EIN−HPr system has the largest binding area which is commonly a characteristic of permanently assembled proteins.[54] Here, however, it is the system with the weakest binding constant. This may result from the overall negative charges of the complex partners. Among the known protein complexes, BN−BS is one of the tightest complexes with 14 hydrogen bonds and 12 salt bridges formed across the interfaces, bearing a K_d value of 1.3×10^{-14} M^{-1}. The CC−CCP complex shows perfect electrostatic complementarity[24] but has the smallest binding interface of all three

dx.doi.org/10.1021/ct5001796 | J. Chem. Theory Comput. 2014, 10, 3512−3524

systems. The relatively weak stability of the bound complex may be connected to the transient nature of the electron transfer step taking place between this interaction pair.

One Dimensional Free Energy Surface of Protein−Protein Association. In this study, we combined umbrella sampling with the WHAM method to obtain the PMF curve for protein−protein disassociation. Here, the reaction coordinate corresponds to the *z*-axis, coincident with the distance between the centers of mass (COM) of the complex partners. We partitioned the trajectory into pieces and computed the PMF and subsequently the standard free energy of binding. Figure 2

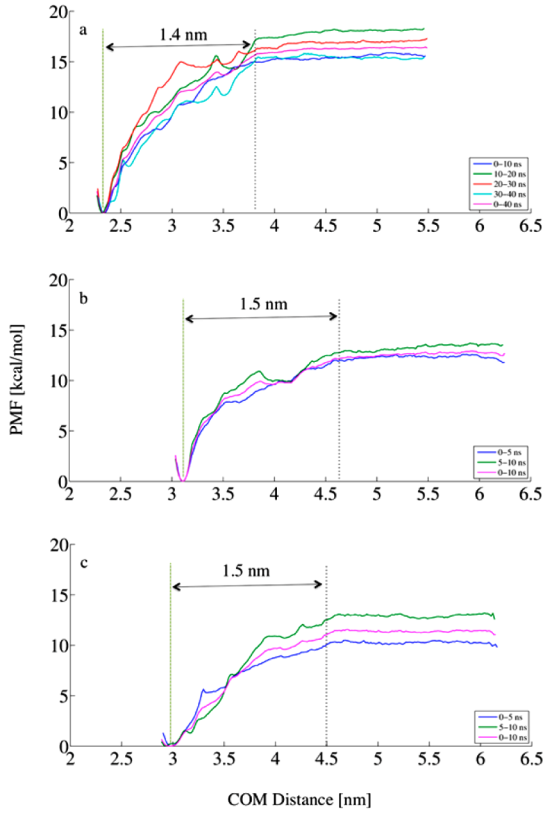


Figure 2. Potential of mean force calculated from different time intervals of the umbrella sampling simulations. (a) Barnase−Barstar, (b) cytochrome c−cytochrome c peroxidase, (c) N-terminal domain of enzyme I−histidine-containing phosphocarrier. The black dashed line that is parallel to the *y*-axis represents the cutoff that separates the bound region from the unbound region. The green dashed line left of it marks the position of the bound state.

shows PMF curves for the three systems obtained using different time intervals of the full-length simulation windows. As seen in the figure all one-dimensional free energy surfaces of association are downhill or in other words barrierless. For BN−BS, this was reported before.[21,55,56] Even though the maximum value of the PMF varies among different time intervals, the PMF curves behave similar for the three systems. Among the systems, the sharpest PMF curve belongs to the BN−BS complex. Here the PMF becomes flat at about 1.4 nm of separation what corresponds to 3.6 nm along the reaction coordinate. For the CC−CCP and EIN−HPr systems the PMF curves flatten beyond 1.5 nm of separation and coincide with the data points at 4.6 and 4.5 nm on the reaction coordinate, respectively. The dashed line parallel to the *y*-axis in Figure 2 indicates this critical separation for each system.

To obtain insight into the energetic contributions that lead to these PMF profiles, we analyzed the components of the nonbonded interaction terms (see Figure 3). The separation distances where the PMFs start to flatten coincide with the distance where the direct Lennard-Jones (LJ) interactions between the proteins almost vanish along the reaction coordinate. In contrast, the direct electrostatic interactions are still very strong at these separations (see Figure 3). Interestingly, these separations are also where the rotational entropy of the proteins starts to converge to a constant value (see Figure 4).

We assessed the quality of the PMF values obtained from different time intervals of the trajectory by inspecting the quality of the sampling histograms and made sure that there is sufficient overlapping between two consecutive windows. Not surprisingly, longer sampling times led to smoother histograms and better overlap. Supporting Information Figure SI 1 shows the histograms of the full trajectories for all systems. Assured by this analysis, we decided to use the full trajectories for further analysis.

Standard Free Energy of Binding. The relation between the equilibrium constant of a chemical reaction and the associated free energy is well established. When the reaction involves a change in the number of components, one has to relate the obtained results to a standard state, see ref 57. Based on the PMF values given in Figure 2, we calculated the standard free energies of binding, as explained in the Methods section. The results are provided in Table 2 together with the experimental values for comparison.

The cutoff between bound and unbound regions was set to the value of the reaction coordinate where the PMF becomes constant within some error interval (see Methods). In Figure 2, the cutoff distances are marked as black dashed lines parallel to the *y*-axis. Although the definition of the cutoff is arbitrary, the lowest values of the PMF in the binding site contribute the most to the integrals and make the calculation of ΔW insensitive to the cutoff.[44] Moreover, the rotational entropy values beyond these cutoffs do not change considerably meaning that the proteins are effectively in bulk (see Figure 4). This is more pronounced for the BN−BS system due to the better convergence during windows of 40 ns in length.

The standard free energy of binding for the different time intervals varies between −11.8 and 14.2 kcal/mol, −8.9 and 9.9 kcal/mol, and −7.0 and 9.8 kcal/mol for BN−BS, CC−CCP, and EIN−HPr, respectively. These give average values of −12.7 ± 1.1, −9.4 ± 0.7 and −8.4 ± 1.9 kcal/mol, which are very close to the free energy of binding computed from the whole trajectory (−12.6 kcal/mol, −9.3 kcal/mol, and −8.3 kcal/mol). Our estimates of the standard error of the mean standard free energy are based on the four 10 ns long trajectory parts for BN−BS and 5 ns long trajectory parts for CC−CCP and EIN−HPr. We carefully searched the literature for the best matching experimental conditions to our simulations. The calculated standard free energies of binding are in good agreement with these experimental values, except for BN−BS. For the CC−CCP and EIN−HPr systems, the computed standard free energies of binding from 10 ns long trajectories, −9.3 kcal/mol and −8.3 kcal/mol, match very closely the experimental values of −8.8 kcal/mol and −7.8 kcal/mol, respectively. However, for the BN−BS system, the value computed from 10 ns long trajectories, −11.8 kcal/mol, is 7.8 kcal/mol less favorable than the experimental value of −19.6 kcal/mol. Extension of the simulations to 40 ns did not bring the computed value much

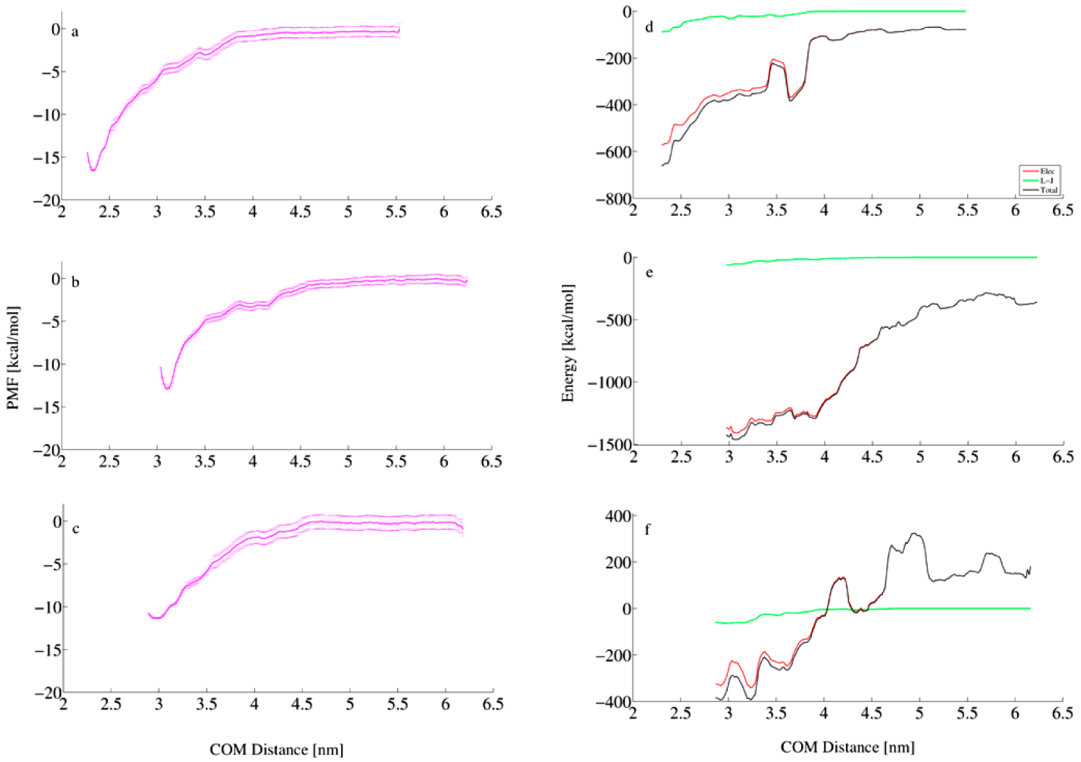


Figure 3. Potential of mean force (PMF) (left panel) and the direct interactions between the two proteins (right panel) calculated from full-length umbrella sampling simulation windows along the COM distance. (a, d) Barnase−Barstar, (b, e) cytochrome c−cytochrome c peroxidase, (c, f) N-terminal domain of enzyme I−histidine-containing phosphocarrier. The error analyses for PMF values were performed using a bootstrap method introduced previously.[43] In panels (d)−(f), green lines refer to Lennard-Jones interactions between the two proteins, red lines to electrostatic interactions, and black lines to the sum of LJ and electrostatic interactions.

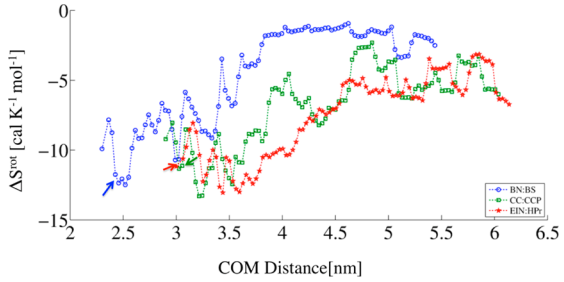


Figure 4. Rotational entropy change along the COM distance: Barnase−Barstar (blue), cytochrome c−cytochrome c peroxidase (green), and N-terminal domain of enzyme I−histidine-containing phosphocarrier (red). The rotational entropy is calculated based on the distribution of the orientations of BS with respect to BN, of CC with respect to CCP and of HPr with respect to EIN. The change in rotational entropy, ΔS^{rot}, is defined relative to the uniform distribution of the three Euler angles. The positions of the bound complexes are marked by arrows.

closer to the experimental correspondent. The largest computed value is −14.2 kcal/mol, which is still 5.4 kcal/mol higher than the experimental value (see Table 2). It is well-known that the computed standard free energy of binding based on the method we followed here is dependent on the applied orthogonal force constant,[44] the sampling time, the number of windows used and the window-width.[58] Therefore, attaining a binding free energy, which compares well to the experimentally determined binding free energy may require further optimization of these parameters for a specific system. Since this is not the scope of this work, we did not perform any further analysis utilizing different parameter sets for BN−BS. Further possible explanations for the deviation from experiment could be inaccuracies of the force field, neglect of explicit polarization, conformational changes of the protein that are not captured during ns-scale time simulations etc. We note that previous, shorter simulations by Neumann and Gottschalk[56] as well as by Wang and Helms[55] also resulted in too low PMF profiles.

ΔG_{PMF} contributes the most to the standard free energy of binding because it accounts for all direct interactions, solvent contributions, rotational and translational entropies of the proteins. The term ΔG_v mainly accounts for translational entropy since among the enthalpy and entropy terms translational entropy is the only one that has a concentration dependence.[59,60] The area explored by the protein in the *xy* plane, 1.61 Å^2, was computed analytically (see Methods) and, therefore, is the same for all three systems. Since the unbound length, l_u, is also almost the same for all systems, the free energy change ΔG_v is the same up to second decimal. Because the unbound volume, V_u, is smaller than the standard state volume, the rescaling free energy term ΔG_v yields a positive contribution (2.5 kcal/mol) to the standard free energy of binding. The free energy to remove the orthogonal constraints, ΔG_R, makes a contribution of about −0.9 kcal/mol for the BN−BS and EIN−HPr systems and of about −1.2 kcal/mol for the CC−CCP system. The ΔG_R values computed using different time frames do not differ much (see Table 2) presumably due to the exponential averaging.

Determinants of Binding Affinity. The riddle in protein−protein association is that there exist no strong chemical bonds formed between the proteins, yet they do form stable

dx.doi.org/10.1021/ct5001796 | J. Chem. Theory Comput. 2014, 10, 3512−3524

Table 2. Standard Free Energies of Binding (ΔG°) for the BN−BS, CC−CCP, and EIN−HPr Complexes[a]

time interval	ΔW (kcal/mol)	ΔG_{PMF} (kcal/mol)	ΔG_v (kcal/mol)	ΔG_R (kcal/mol)	ΔG° (kcal/mol)	ΔG°_{exp} (kcal/mol)
			BN−BS			
0−10 ns	−12.9	−13.5	2.5	−0.8	−11.8	−19.6[b]
10−20 ns	−15.3	−15.8	2.5	−0.9	−14.2	
20−30 ns	−14.2	−14.7	2.5	−0.8	−13.0	
30−40 ns	−12.9	−13.4	2.5	−0.9	−11.8	
					−12.7 ± 1.1[c]	
0−40 ns	−13.7	−14.2	2.5	−0.9	−12.6	
			CC−CCP			
0−5 ns	−9.7	−10.3	2.5	−1.1	−8.9	−8.8[b]
5−10 ns	−10.6	−11.2	2.5	−1.2	−9.9	
					−9.4 ± 0.7[c]	
0−10 ns	−10.2	−10.6	2.5	−1.2	−9.3	
			EIN−HPr			
0−5 ns	−7.7	−8.6	2.5	−0.9	−7.0	−7.8[b]
5−10 ns	−10.3	−11.3	2.5	−0.9	−9.8	
					−8.4 ± 1.9[c]	
0−10 ns	−8.8	−9.9	2.5	−0.9	−8.3	

[a]The experimental values ΔG°_{exp} are based on the binding constants given in Table 1. The calculated values; ΔW, ΔG_{PMF}, ΔG_v, and ΔG_R stand for the PMF depth, the free energy change of binding between the bound and unbound section of the PMF, the free energy change from the unbound volume to the standard state volume and the free energy change to remove the orthogonal restraints, respectively. [b]Converted from the disassociation constants reported in Table 1, using the simulation temperature (310 K). [c]The mean value and the standard deviation of the standard free energies of binding (ΔG°) from 10 ns (BN−BS) and 5 ns (CC−CCP and EIN−HPr) time intervals.

Table 3. Solvent-Induced Contribution (δG) to the Standard Free Energy of Binding Was Derived As the Difference of the Computed ΔG° from Full-Length Windows (40 ns for BN−BS and 10 ns for CC−CCP and EIN−HPr) and the Direct Interactions ΔU Value Computed from Full-Length Windows[a]

systems	ΔG° (kcal/mol)	ΔU (kcal/mol)	$-T\Delta S_L^{tr}$ (kcal/mol)	$-T\Delta S_L^{rot}$ (kcal/mol)	ΔG_{T-R} (kcal/mol)	δG (kcal/mol)
BN−BS	−12.7	−594.6	2.4	4.2	6.6	575.3
CC−CCP	−9.4	−1178.5	2.5	5.0	7.5	1161.6
EIN−HPr	−8.4	−718.4	2.5	4.8	7.3	702.7

[a]ΔS_L^{tr} and ΔS_L^{rot} are translational and rotational entropy changes upon complexation based on the SF approach. ΔG_{T-R} is the sum of the two entropic terms $-T\Delta S_L^{tr}$ and $-T\Delta S_L^{rot}$, where T stands for the temperature.

assemblies in aqueous solution. As explanation for these strong driving forces Ben-Naim[13] suggested that solvent-induced interactions make a large, favorable contribution to the standard free energy of binding. Ben-Naim decomposed the binding free energy into the following three terms:[13]

$$\Delta G^\circ = \Delta G_{T-R} + \Delta U + \delta G$$

where ΔG_{T-R} is the contribution to the driving force (ΔG°) due to the changes in translational and rotational degrees of freedom of all species (the monomers and the complex), ΔU is the energy change for bringing the two proteins from infinite separation to the final configuration of the complex in vacuum, and δG is the solvent induced contribution to the binding free energy.

Figure 3d−f show the direct interaction energy between the complex partners along the reaction coordinate obtained from full-length windows. As for the entropies, the direct interaction energies were computed from the full-length windows by dividing the reaction coordinate into bins and averaging the energies inside each bin. As expected for protein pairs carrying nonzero net electric charges, the direct interactions are overwhelmingly large compared to the corresponding PMF and always favorable except for EIN−HPr at larger distances. The observed humps in the electrostatic energy along the dissociation path (Figure 3d−f) are due to protein rotations and reveal the pronounced dipolar character of the proteins. In the EIN−HPr system the proteins have a negative overall electrostatic charge even though the interfaces are oppositely charged. Therefore, the proteins may adopt configurations with positive overall electrostatic interaction energies even in the early stages of dissociation. Since these effects are compensated by respective protein−solvent interactions (see below), the strong changes in the electrostatic interactions along the reaction coordinate do not lead to spikes in the corresponding PMF values (Figure 3a−c). Even though the PMF and Lennard-Jones interactions almost vanish beyond the cutoff, the electrostatic interactions are still very strong at this distance. Thus, there must be some effect, which counteracts these interactions and lowers the PMF to its actual value and converges to zero beyond the cutoff.

By decomposing the standard free energy of binding in the manner described above, we obtained the results listed in Table 3. The free energy change ΔG_{T-R}(see Table 3) was computed as the sum of translational and rotational entropy terms $-T\Delta S_L^{tr}$ and $-T\Delta S_L^{rot}$, which were predicted from the SF approach. The free energy contributions due to the loss of translational entropy upon binding are almost equal; 2.4 kcal/mol for BN−BS and 2.5 kcal/mol for the other two systems. The $-T\Delta S_L^{rot}$ term that accounts for the free energy due to rotational entropy loss upon complexation is almost twice the contribution from the translational entropy loss, namely 4.2 kcal/mol, 5.0 and 4.8 kcal/mol for BN−BS, CC−CCP, and EIN−HPr, respectively.

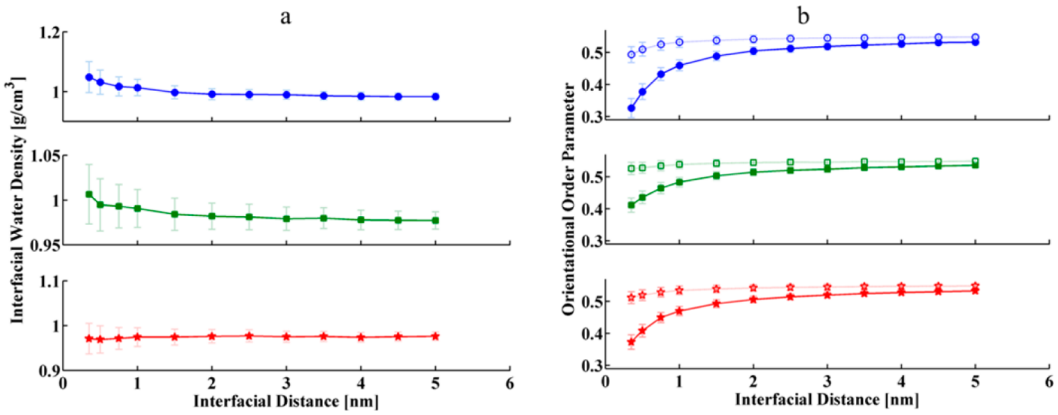


Figure 5. Density (panel a) and orientational order parameter (panel b) of the interfacial water for the three systems: Barnase−Barstar (blue circles), cytochrome c−cytochrome c peroxidase (green squares), and N-terminal domain of enzyme I−histidine-containing phosphocarrier (red stars). In panel b, the solid symbols are the orientational order parameter when only including neighboring water molecules. The open symbols stand for a modified orientational order parameter of the interfacial water molecules when considering their four closest water oxygens or interfacial protein atoms (O and N) within a 3.5 Å distance cutoff.

The rotational entropy contribution to the standard free energy of binding for BN−BS computed here (4.2 kcal/mol) is in quite good agreement with the value reported in a recent study[21] (5.8 kcal/mol), which was calculated by applying a series of orientational restraints. ΔU is simply defined as the difference between minimum and maximum value of the total direct interactions.

Table 3 illustrates that according to the energy decomposition suggested by Ben-Naim[13] the overall solvent-induced contribution to the standard free energy of binding is positive. This contradicts the reasoning of Ben-Naim who expected ΔU to be in the order of only 0.5 kcal/mol, which is not the case. We note, however, that the systems studied here are charged so that this behavior is quite expected. For the EIN−HPr system, there exist some regions along the reaction coordinate where the total direct interactions are repulsive (see Figure 3f) what would lead to a favorable negative δG. However, when considering the end points along the association path, the solvent-induced interactions are repulsive for this system as well. Therefore, we conclude that the solvent-induced interactions counteract the large and favorable total direct interactions and are repulsive overall. This issue, whether the solvent-induced interactions are attractive or repulsive was addressed before. Based on some model systems, Berne and coworkers[14] found that plates with homogeneous hydrophilic and hydrophobic sites give rise to repulsive solvent-induced interactions. However, for plates with large hydrophilic domains, they reported attractive solvent-induced interactions coupled with a dewetting transition. The plates with homogeneous hydrophobic and hydrophilic sites resemble hydrophilic protein−protein interfaces quite well, since the hydrophilic protein−protein interfaces do not only bear hydrophilic residues, but also contain hydrophobic residues.

We also computed the rotational entropy along the reaction coordinate using the distribution of the three Euler angles (see Methods) and provide the results in Figure 4. Our aim here was mainly to assess the convergence of rotational motions. As seen in the figure, the term ΔS^{rot} converges to a constant value at large distances that we defined as unbound region. We also computed the rotational entropy contribution using different angle bin sizes. Although the numbers changed slightly at some regions along the reaction coordinate, the overall change of ΔS^{rot}, is negligible (see Supporting Information Figure SI 2). The slight deviation in ΔS^{rot} at these regions is largely due to the inadequate sampling, since every distance bin contains different numbers of snapshots. If we consider the rotational entropy of binding as the difference between minimum (in the bound region) and maximum (in the unbound region) values we obtain −11.0 cal K^{-1} mol^{-1} for BN−BS, −10.7 cal K^{-1} mol^{-1} for CC−CCP, and −9.7 cal K^{-1} mol^{-1} for EIN−HPr. These entropy changes give $T\Delta S^{rot}$ values of 3.4 kcal/mol, 3.3 and 3.0 kcal/mol free energy changes at 310 K, respectively, which are smaller than the contributions obtained using the SF approach (see Table 3).

Properties of Interfacial Water. At the atomic level, water appears to be one of the simplest molecules, yet understanding the structure and dynamics of liquid water is an ongoing challenge.[61] Our second set of simulations consisted of MD simulations that address the behavior of water localized between two hydrophilic proteins. It is well-known that confining surfaces can exert a profound influence on the structure and dynamics of water.[61,62] Here, we investigated how the properties of water localized between two hydrophilic proteins differ from those of bulk water by means of density and orientational order parameter. Figure 5 shows that the interfacial water density increases by a few percent at close interfacial distances for the BN−BS and CC−CCP systems. However, for the EIN−HPr complex we observed slight dewetting at the same interfacial distances. In all cases, the density reaches the bulk value (0.98 g/cm^3 for TIP3P water model) beyond separation distances of a few nanometers (see Figure 5). A similar behavior was reported previously for protein−protein[16] and model systems.[14]

In order to investigate the local structure of the interfacial water, we computed the orientational order parameter q. The possible values for this orientational order parameter for a single water molecule range from −3 to 1. However, the average value for a collection of molecules varies between 0 in a random network (no order) and 1 in a tetrahedral network.[53] We conducted this calculation in two different ways, without and with considering the nearby polar protein atoms (N and O) within a cutoff of 3.5 Å. We found that the average value for the orientational order parameter in bulk TIP3P is 0.55. As shown in Figure 5b, the orientational order parameter decreases to values between 0.33 and 0.4 at small separation between the protein surfaces. This was the case for all three systems when

we did not consider protein atoms for the analysis. However, when we took into account the nearby polar protein atoms that are potential H-bond donors and acceptors the decrease in the orientational order parameter stops at values of about 0.5. For the CC−CCP system, the orientational order parameter is even almost constant. We observed the relatively strongest decrease in this modified orientational order parameter for the BN−BS pair (Figure 5b, open circles). The small overall decrease in the orientational order parameter when considering the protein atoms compared to the analysis that only considered neighboring water molecules manifests that most water molecules that are in the vicinity of polar interfaces form hydrogen bonds with the protein atoms as expected.

The small decrease in the orientational order parameter for interfacial water appears to be somehow connected with the increase in density, especially for the BN−BS pair (see Supporting Information Figure SI 3). Structural order decrease with increasing density at constant temperature was previously reported. Errington et al.[53] and Yan et al.[63] increased the density of water upon compression and observed a decrease in structural order. Another way of looking at this is to consider the electric field created by the confining proteins. The effects of electric fields on water are well documented and are manifold; bending or breaking of hydrogen bonds due to reorientation, phase transition,[64] lowering of the dielectric constant,[16,65] etc. have been reported. At small separations the effect of an exerted electric field on the interfacial water molecules is stronger compared to large separations. This is the case because there are only few water molecules in the interfacial gap that can generate a counteracting electric field. The water molecules therefore align with the exerted electric field and do not have enough orientational freedom to rotate into a proper orientation for forming tetrahedral structure.

CONCLUSIONS

In this study, we addressed the association of hydrophilic proteins and the role of water using extensive molecular dynamics simulations. This was done on the example of three well studied complexes; Barnase−Barstar, cytochrome c−cytochrome c peroxidase and the complex of the N-terminal domain of enzyme I with histidine-containing phosphocarrier. We found that the one-dimensional free energy surfaces of association are downhill or in other words barrierless. However, we note that free energy profiles depend on the reaction coordinate, and therefore, our results do not show conclusively that the real binding paths are barrierless. Using the obtained potential of mean force curves along the association path, the standard free energies of binding were computed to be in reasonable to very good agreement with their experimental correspondents. Second, we focused on the role of water in the protein−protein association. Decomposing the standard free energy of binding revealed that the favorable electrostatic and Lennard-Jones interactions between the protein pairs render the solvent-induced interactions repulsive. Analysis of the water localized between the two proteins showed that the orientational order parameter of confined water deviates to a small degree but noticeably from bulk values, especially at close separations of the confining proteins. The water at the interfacial gap is found to be more dense compared to bulk water at close separations of the complex partners. This study showed that different hydrophilic protein−protein interfaces seem to bind according to similar physicochemical principles. Atomistic MD simulations in explicit solvent proved to be a reliable method to investigate overall principles as well as fine details of such binding processes.

ASSOCIATED CONTENT

Supporting Information

Figure SI 1: Histograms of the configurations within the umbrella sampling windows for each system. Figure SI 2: Rotational entropy calculated using different angle bin size for the BN−BS system. Figure SI 3: Interfacial density vs orientational order parameter. Text SI: The newly derived atomic partial charges and the parameters used in this study for heme groups and heme coordinating residues to be integrated to the AMBER force field under the GROMACS package. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Email: volkhard.helms@bioinformatik.uni-saarland.de.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (He 3875/11-1).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de).

REFERENCES

(1) Tong, A. H. Y.; Lesage, G.; Bader, G. D.; Ding, H. M.; Xu, H.; Xin, X. F.; Young, J.; Berriz, G. F.; Brost, R. L.; Chang, M.; Chen, Y. Q.; Cheng, X.; Chua, G.; Friesen, H.; Goldberg, D. S.; Haynes, J.; Humphries, C.; He, G.; Hussein, S.; Ke, L. Z.; Krogan, N.; Li, Z. J.; Levinson, J. N.; Lu, H.; Menard, P.; Munyana, C.; Parsons, A. B.; Ryan, O.; Tonikian, R.; Roberts, T.; Sdicu, A. M.; Shapiro, J.; Sheikh, B.; Suter, B.; Wong, S. L.; Zhang, L. V.; Zhu, H. W.; Burd, C. G.; Munro, S.; Sander, C.; Rine, J.; Greenblatt, J.; Peter, M.; Bretscher, A.; Bell, G.; Roth, F. P.; Brown, G. W.; Andrews, B.; Bussey, H.; Boone, C. Global mapping of the yeast genetic interaction network. *Science* **2004**, *303*, 808−813.

(2) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic aspects of protein−protein interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153−159.

(3) Ansari, S.; Helms, V. Statistical analysis of predominantly transient protein−protein interfaces. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 344−355.

(4) Larsen, T. A.; Olson, A. J.; Goodsell, D. S. Morphology of protein−protein interfaces. *Structure* **1998**, *6*, 421−427.

(5) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding—A synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.

(6) Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, *4*, 10−19.

(7) Tsai, C. J.; Kumar, S.; Ma, B. Y.; Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **1999**, *8*, 1181−1190.

(8) Woo, H. J.; Roux, B. Calculation of absolute protein−ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825−6830.

(9) Kaestner, J. Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 932−942.

(10) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analyses method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(11) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640−647.
(12) McLain, S. E.; Soper, A. K.; Daidone, I.; Smith, J. C.; Watts, A. Charge-based interactions between peptides observed as the dominant force for association in aqueous solution. *Angew. Chem., Int. Ed.* **2008**, *47*, 9059−9062.
(13) Ben-Naim, A. On the driving forces for protein−protein association. *J. Chem. Phys.* **2006**, *125*, 24901.
(14) Hua, L.; Zangi, R.; Berne, B. J. Hydrophobic interactions and dewetting between plates with hydrophobic and hydrophilic domains. *J. Phys. Chem. C* **2009**, *113*, 5244−5253.
(15) Ahmad, M.; Gu, W.; Helms, V. Mechanism of fast peptide recognition by SH3 domains. *Angew. Chem., Int. Ed.* **2008**, *47*, 7626−7630.
(16) Ahmad, M.; Gu, W.; Geyer, T.; Helms, V. Adhesive water networks facilitate binding of protein interfaces. *Nat. Commun.* **2011**, *2*, 261.
(17) Buckle, A. M.; Schreiber, G.; Fersht, A. R. Protein−protein recognition—Crystal structural-analysis of a Barnase−Barstar complex at 2.0-Å resolution. *Biochemistry* **1994**, *33*, 8878−8889.
(18) Gabdoulline, R. R.; Wade, R. C. Simulation of the diffusional association of Barnase and Barstar. *Biophys. J.* **1997**, *72*, 1917−1929.
(19) Spaar, A.; Helms, V. Free energy landscape of protein−protein encounter resulting from Brownian dynamics simulations of Barnase−Barstar. *J. Chem. Theory Comput.* **2005**, *1*, 723−736.
(20) Vijayakumar, M.; Wong, K. Y.; Schreiber, G.; Fersht, A. R.; Szabo, A.; Zhou, H. X. Electrostatic enhancement of diffusion-controlled protein−protein association: Comparison of theory and experiment on Barnase and Barstar. *J. Mol. Biol.* **1998**, *278*, 1015−1024.
(21) Gumbart, J. C.; Roux, B.; Chipot, C. Efficient determination of protein−protein standard binding free energies from first principles. *J. Chem. Theory Comput.* **2013**, *9*, 3789−3798.
(22) Pelletier, H.; Kraut, J. Crystal-structure of a complex between electron-transfer partners, cytochrome-c peroxidase and cytochrome-c. *Science* **1992**, *258*, 1748−1755.
(23) Northrup, S. H.; Boles, J. O.; Reynolds, J. C. L. Brownian dynamics of cytochrome-c and cytochrome-c peroxidase association. *Science* **1988**, *241*, 67−70.
(24) Gabdoulline, R. R.; Wade, R. C. Protein−protein association: Investigation of factors influencing association rates by Brownian dynamics simulations. *J. Mol. Biol.* **2001**, *306*, 1139−1155.
(25) Garrett, D. S.; Seok, Y. J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. Solution structure of the 40 000 M-r phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat. Struct. Biol.* **1999**, *6*, 166−173.
(26) Kim, Y. C.; Tang, C.; Clore, G. M.; Hummer, G. Replica exchange simulations of transient encounter complexes in protein−protein association. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12855−12860.
(27) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. Protein Data Bank—Computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(28) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704−721.
(29) Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433−438.
(30) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(31) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(32) Jorgensen, W. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335−340.
(33) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−18.
(34) Worrall, J. A. R.; Kolczak, U.; Canters, G. W.; Ubbink, M. Interaction of yeast iso-1-cytochrome c with cytochrome c peroxidase investigated by N-15,H-1 heteronuclear NMR spectroscopy. *Biochemistry* **2001**, *40*, 7069−7076.
(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. *Gaussian 03*, Revision C.02. Gaussian, Inc.: Wallingford, CT, 2004.
(36) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational resp methodology to biopolymers—Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(37) Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., III. Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **2012**, *33*, 119−133.
(38) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(39) Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.
(40) Nose, S. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511−519.
(41) Hoover, W. G. Canonical dynamics—Equilibrium phase-space distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695−1697.
(42) Parrinello, M.; Rahman, A. Polymorphic transitions in single-crystals—A new molecular-dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(43) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham-A free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(44) Doudou, S.; Burton, N. A.; Henchman, R. H. Standard free energy of binding from a one-dimensional potential of mean force. *J. Chem. Theory Comput.* **2009**, *5*, 909−918.
(45) Goldstein, H.; Poole, C.; Safko, J. *Classical Mechanics*. 3rd ed.; Addison-Wesley: San Francisco, 2001; pp 150−154.
(46) Lazaridis, T.; Masunov, A.; Gandolfo, F. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 194−208.
(47) Irudayam, S. J.; Henchman, R. H. Entropic cost of protein−ligand binding and its dependence on the entropy in solution. *J. Phys. Chem. B* **2009**, *113*, 5871−5884.
(48) Peter, W.; Sam, A.; Volkhard, H. ABC (Analysing Biomolecular Contacts)-database. *J. Integr. Bioinf.* **2007**, *4*, 50.
(49) Kuntz, I. D., Jr.; Kauzmann, W. Hydration of proteins and polypeptides. *Adv. Protein Chem.* **1974**, *28*, 239−345.
(50) Squire, P. G.; Himmel, M. E. Hydrodynamics and protein hydration. *Arch. Biochem. Biophys.* **1979**, *196*, 165−177.
(51) Gekko, K.; Noguchi, H. Compressibility of globular-proteins in water at 2 °C. *J. Phys. Chem.* **1979**, *83*, 2706−2714.
(52) Chau, P. L.; Hardwick, A. J. A new order parameter for tetrahedral configurations. *Mol. Phys.* **1998**, *93*, 511−518.
(53) Errington, J. R.; Debenedetti, P. G. Relationship between structural order and the anomalies of liquid water. *Nature* **2001**, *409*, 318−321.
(54) Jones, S.; Thornton, J. M. Principles of protein−protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13−20.
(55) Wang, L.; Siu, S. W. I.; Gu, W.; Helms, V. Downhill binding energy surface of the Barnase−Barstar complex. *Biopolymers* **2010**, *93*, 977−985.

(56) Neumann, J.; Gottschalk, K.-E. The effect of different force applications on the protein−protein complex Barnase−Barstar. *Biophys. J.* **2009**, *97*, 1687−1699.
(57) General, I. J. A note on the standard state's binding free energy. *J. Chem. Theory Comput.* **2010**, *6*, 2520−2524.
(58) Buch, I.; Kashif Sadiq, S.; De Fabritiis, G. Optimized potential of mean force calculations for standard binding free energies. *J. Chem. Theory Comput.* **2011**, *7*, 1765−1772.
(59) Yu, Y. B.; Privalov, P. L.; Hodges, R. S. Contribution of translational and rotational motions to molecular association in aqueous solution. *Biophys. J.* **2001**, *81*, 1632−1642.
(60) Zhou, H.-X.; Gilson, M. K. Theory of free energy and entropy in noncovalent binding. *Chem. Rev.* **2009**, *109*, 4092−4107.
(61) Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **2008**, *108*, 74−108.
(62) Rasaiah, J. C.; Garde, S.; Hummer, G. Water in nonpolar confinement: From nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713−740.
(63) Yan, Z.; Buldyrev, S. V.; Kumar, P.; Giovambattista, N.; Debenedetti, P. G.; Stanley, H. E. Structure of the first- and second-neighbor shells of simulated water: Quantitative relation to translational and orientational order. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *76*, 051201.
(64) Danielewicz-Ferchmin, I.; Ferchmin, A. R. Phase diagram of electrostricted H_2O at surfaces of electrodes at 273−373 K: Electric critical point of water. *ChemPhysChem* **2005**, *6*, 1499−1509.
(65) Danielewicz-Ferchmin, I.; Ferchmin, A. R. Water at ions, biomolecules and charged surfaces. *Phys. Chem. Liq.* **2004**, *42*, 1−36.
(66) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(67) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33−38.
(68) Schreiber, G.; Fersht, A. R. Interaction of Barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **1993**, *32*, 5145−5150.
(69) Mei, H. K.; Wang, K. F.; McKee, S.; Wang, X. M.; Waldner, J. L.; Pielak, G. J.; Durham, B.; Millett, F. Control of formation and dissociation of the high-affinity complex between cytochrome c and cytochrome c peroxidase by ionic strength and the low-affinity binding site. *Biochemistry* **1996**, *35*, 15800−15806.
(70) Ginsburg, A.; Szczepanowski, R. H.; Ruvinov, S. B.; Nosworthy, N. J.; Sondej, M.; Umland, T. C.; Peterkofsky, A. Conformational stability changes of the amino terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate−sugar phosphotransferase system produced by substituting alanine or glutamate for the active-site histidine 189: Implications for phosphorylation effects. *Protein Sci.* **2000**, *9*, 1085−1094.