SCIENTIFIC REPORTS

Dynamics of Hydration Water Plays a Key Role in Determining the Binding Thermodynamics of Protein Complexes

Song-Ho Chong & Sihyun Ham

Interfacial waters are considered to play a crucial role in protein–protein interactions, but in what sense and why are they important? Here, using molecular dynamics simulations and statistical thermodynamic analyses, we demonstrate distinctive dynamic characteristics of the interfacial water and investigate their implications for the binding thermodynamics. We identify the presence of extraordinarily slow (~1,000 times slower than in bulk water) hydrogen-bond rearrangements in interfacial water. We rationalize the slow rearrangements by introducing the "trapping" free energies, characterizing how strongly individual hydration waters are captured by the biomolecular surface, whose magnitude is then traced back to the number of water–protein hydrogen bonds and the strong electrostatic field produced at the binding interface. We also discuss the impact of the slow interfacial waters on the binding thermodynamics. We find that, as expected from their slow dynamics, the conventional approach to the water-mediated interaction, which assumes rapid equilibration of the waters' degrees of freedom, is inadequate. We show instead that an explicit treatment of the extremely slow interfacial waters is critical. Our results shed new light on the role of water in protein–protein interactions, highlighting the need to consider its dynamics to improve our understanding of biomolecular bindings.

Water is an active and indispensable component of cells. Understanding its versatile roles in determining the structure and dynamics of biomolecules and mediating their interactions is of fundamental importance[1–3]. The versatility of water in biological contexts arises from its ability to alter its characteristics depending on its interaction with biomolecules. For example, the DNA sequence-dependent behavior of hydration water serves as a sequence-specific "hydration fingerprint"[4]; changes in water dynamics during binding of a substrate to an enzyme play a vital role in protein–ligand recognition[5]; and the non-bulk behavior of water inside the translocon strongly affects the partitioning of hydrophobic segments from the translocon to the membrane[6]. However, although our understanding of the behavior of hydration water around biomolecules has advanced significantly in recent years[7–14], it remains a challenge to elucidate the extent to which water molecules located between two biomolecules are modified through concurrent interactions with the two binding surfaces and how such altered water molecules in turn affect the binding affinity.

In this connection, it has been suggested that water-mediated contacts substantially complement direct protein–protein contacts, providing an additional layer of biomolecular recognition[15, 16]. The necessity of an explicit treatment of interfacial water molecules to properly describe such water-mediated interactions has also been noted[17]. Indeed, recent computational studies have reported on the relevance of explicitly handling "key" interfacial waters in protein–protein interaction[18] and protein–ligand binding[19]: for example, including two, rather than all, interfacial water molecules was crucial to correctly obtaining the trends observed in mutation effects on protein–protein binding affinity[18]; in another study, explicitly taking into account interfacial water molecules ranging in number (N_{wat}) from 30 to 70 significantly improved the correlation with the experimental binding affinities for four different systems, where the optimum value of N_{wat} depended on the specific system[19]. What, however, distinguishes those key interfacial water molecules from others? Do any distinctive characteristics of the interfacial water emerge upon protein–protein complex formation?

Department of Chemistry, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul, 04310, Korea. Correspondence and requests for materials should be addressed to S.H. (email: sihyun@sookmyung.ac.kr)
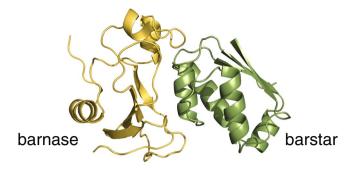
Figure 1. Structure of the barnase–barstar complex.

In this paper, we investigate the dynamic and thermodynamic features of interfacial water in the barnase–barstar complex[15]. This is a well-studied paradigm for protein–protein interactions and is also an ideal system for analyzing the interfacial water because X-ray measurements indicate the presence of waters filling the gap between the binding surfaces[15, 20]. We perform molecular dynamics simulations to explore dynamic characteristics of the interfacial water. We focus on the rearrangements of hydrogen bonds, which are the most important protein–water interaction because the protein–protein binding surfaces comprise mainly polar and charged residues[21]. We then conduct statistical thermodynamic analyses to rationalize the dynamic characteristics of the interfacial water. Finally, we discuss the impact of the interfacial water dynamics on the protein–protein binding affinity. We find that the conventional approach to the water-mediated interaction, which assumes the time-scale separation between the protein and hydration water dynamics, fails owing to the extremely slow dynamics exhibited by the interfacial waters. We show instead that an explicit treatment of those slow waters as an integral part of biomolecules is critical. Thereby, we would like to shed new light on the role of water in protein–protein interactions based on a dynamic view point.

Methods

Molecular dynamics simulations. We conducted explicit-water molecular dynamics simulations for the barnase–barstar complex (Fig. 1) and for the free barnase and barstar proteins. The initial complex structure was modeled based on the X-ray structure (PDB ID: 1 BRS[15]) as detailed in ref. 22. The starting structures of the free barnase and barstar simulations were taken from their NMR structures (1 BNR[23] and 1 BTA[24], respectively). The complex structure was solvated by 23,477 waters and neutralized by 4 counter Na^+ ions in a cubic box of the initial side length 95.4 Å; the free barnase (barstar) was solvated by 14,346 (8,397) waters and neutralized by $2\,Cl^-$ ($6\,Na^+$) ions in a cubic box of the initial size 81.7 Å (69.5 Å). All the simulations were performed using the AMBER14 suite[25] with the FF99SB force field[26] for proteins and the TIP3P model[27] for water. The temperature and pressure were maintained at $T = 300$ K and $P = 1$ bar using the Berendsen's method[28]. We adopted the same simulation procedures as described in ref. 22, and three independent 1 μs production runs were performed for each system starting from different random initial velocities.

We also conducted pure-water simulations to obtain dynamical quantities for bulk water. Three independent 100 ns simulations were performed at $T = 300$ K and $P = 1$ bar with 2,539 waters.

Hydrogen-bond rearrangement dynamics. We analyzed the hydrogen-bond time-correlation function, which quantifies the extent to which hydrogen bonds found at time $t = 0$ survive to subsequent times t[29], to investigate hydrogen-bond rearrangements between protein and hydration water. A hydrogen bond is considered formed when the water oxygen is located within 3.5 Å from heavy atoms in a protein. The hydration water is classified as follows (see Fig. 2a for an illustration). A water molecule forming a single hydrogen bond to a protein is referred to as single HB water. The locations of single HB waters in a simulation snapshot are indicated by cyan spheres in Fig. 2b, and their average number ($\pm$standard deviation) computed from the whole simulation trajectories for the complex is 325.7 ± 13.7 (Table 1 summarizes the number of water molecules and the number of water–protein hydrogen bonds). A water molecule making two or more hydrogen bonds to a protein is termed double HB water: the positions of double HB waters in a snapshot are shown by orange spheres in Fig. 2b. There are 133.0 ± 8.0 double HB waters in the system, and the average number of hydrogen bonds to a protein is 2.4 ± 0.1. Finally, a water molecule forming concurrent hydrogen bonds with two proteins is called bridging water. By definition, bridging waters are present only at the interface between two proteins (red spheres in Fig. 2b). We find 19.6 ± 3.0 bridging waters located at the interface, with the average number of water–protein hydrogen bonds being 2.9 ± 0.2.

We investigate the hydrogen-bond time-correlation function defined by

$$C_{HB}^{\alpha}(t) = \frac{\langle C_{HB,indiv}^{\alpha}(t) \rangle}{\langle C_{HB,indiv}^{\alpha}(0) \rangle} \tag{1}$$

with the individual contributions $C_{HB,indiv}^{\alpha}(t)$. Here, α refers to the type of water, and the brackets denote the average over water molecules of this type. For bulk and single HB water, $C_{HB,indiv}^{\alpha}(t)$ is 1 if a hydrogen bond found at time 0 survives at time t. For double and bridging water, $C_{HB,indiv}^{\alpha}(t)$ is 1 if two or more hydrogen bonds found at time $t = 0$ survive at time t.
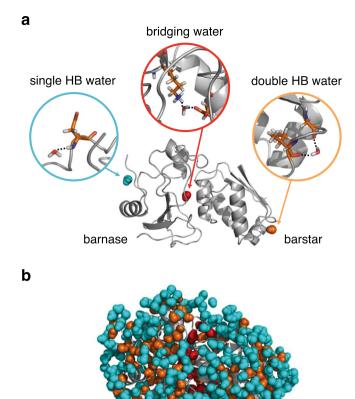
a

b

Figure 2. (**a**) Illustration of single HB, double HB, and bridging waters; dotted lines denote hydrogen bonds. (**b**) Snapshot of the barnase–barstar complex structure with hydration water, showing the distribution of single HB water (cyan), double HB water (orange), and bridging water (red).

	Single HB water	Double HB water	Bridging water
# of waters	325.7 ± 13.7	133.0 ± 8.0	19.6 ± 3.0
# of H-bonds	1	2.4 ± 0.1	2.9 ± 0.2

Table 1. Average number of waters and water–protein hydrogen bonds[a]. [a]Average $\pm$ standard deviation.

Trapping free energy. We introduce the trapping free energies of individual hydration waters to quantify how strongly they are bound to the protein surface. The trapping free energy refers to the reversible work (i.e., the potential of mean force) for transferring a water molecule from an infinite separation to a specific position and orientation relative to the protein–protein complex. We consider here the transfer process to a fixed position and orientation relative to the solute for two reasons. First, this allows us to compute the trapping free energies of individual hydration waters solely based on the simulation snapshots such as the one presented in Fig. 2b. Second, we are interested in whether the trapping free energies so computed at time $t = 0$ serve as a descriptor of the degree of retardation of the subsequent ($t > 0$) dynamics of individual water molecules. The thermodynamic cycle shown in Fig. 3 is used to obtain this quantity; in this cycle, we consider the transfer of the i-th water molecule to a specific position and orientation around the solute u, which includes the hydration water molecules of interest (e.g., all the waters displayed in Fig. 2b). The solute from which the i-th water molecule is excluded is denoted as u'.

Process (1) in Fig. 3 is the independent solvation processes of the i-th water molecule and the solute u'; hence, the associated Gibbs free energy change is given by $\Delta G_{(1)} = G_{\text{water}}^{\text{solv}} + G_{u'}^{\text{solv}}$, which consists of the solvation free energies of a single water molecule ($G_{\text{water}}^{\text{solv}}$) and of the solute u' ($G_{u'}^{\text{solv}}$). In process (2), the i-th water molecule is transferred to a specific position and orientation around the solute u' from an infinite separation in vacuum. The reversible work required for this process is given by the interaction energy $E_{u'-i}$ between the solute u' and the i-th water molecule, $\Delta G_{(2)} = E_{u'-i}$. Process (3) is the solvation of the solute $u(=u'+i)$, and we obtain $\Delta G_{(3)} = G_{u(=u'+i)}^{\text{solv}}$. From the thermodynamic cycle, one obtains the trapping free energy from $-\Delta G_{(1)} + \Delta G_{(2)} + \Delta G_{(3)}$; that is,

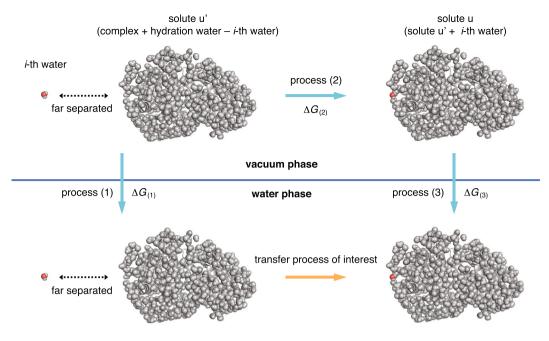$$G_i^{\text{trap}} = E_{u'-i} + G_{u(=u'+i)}^{\text{solv}} - G_{u'}^{\text{solv}} - G_{\text{water}}^{\text{solv}} \qquad (2)$$

Figure 3. Thermodynamic cycle for obtaining the trapping free energy. Process (1) is the separate solvation processes of the i-th water molecule and the solute u' (barnase–barstar complex + hydration water − i-th water molecule). Process (2) is the transfer process of the i-th water molecule to a specific position and orientation around the solute u' in vacuum. Process (3) is the solvation process of the solute u (=solute u' + i-th water molecule). From the Gibbs free energy changes associated with these processes, the transfer free energy of interest can obtained as $-\Delta G_{(1)} + \Delta G_{(2)} + \Delta G_{(3)}$.

We computed the interaction energy $E_{u'-i}$ from the force field, whereas the solvation free energy G_u^{solv} is obtained using the 3D-RISM theory[30], whose details are presented in Supplementary Methods. An efficient method for computing the contribution $G_{u(=u'+i)}^{\text{solv}} - G_{u'}^{\text{solv}}$ which is based on the atomic decomposition of the solvation free energy[31, 32] is also provided there. The trapping free energies for the hydration waters surrounding free barnase and barstar proteins can be obtained in a similar manner.

Recently, several computational methods have been developed for evaluating thermodynamic functions of individual hydration waters[33–41]. However, these methods typically demand performing additional distinct simulations. For example, the application of the inhomogeneous solvation theory[33–35] requires conducting simulations in which restrains are added on protein atoms to sample waters' positions and orientations for a given protein conformation. Further complications in analysis will arise when hydration waters exchange with bulk waters during those additional simulations. On the other hand, our computational method for the trapping free energy that employs the integral-equation theory (3D-RISM) is applicable solely based on snapshots taken from unrestrained equilibrium simulations, and it is in this sense more computationally efficient.

Standard binding free energy. *Conventional approach.* The statistical thermodynamic expression for the standard binding free energy is given by refs 22 and 42

$$\Delta G_{\text{bind}}^0 = \overline{\Delta f_u} - T(\Delta S_{\text{config}} + \Delta S_{\text{ext}}) \tag{3}$$

Here, ΔX denotes the change in X upon complex formation from two free proteins (labeled 1 and 2), $\Delta X = X_{\text{complex}} - (X_1 + X_2)$; $f_u = E_u + G_u^{\text{solv}}$ comprises the gas-phase energy (E_u) and the solvation free energy (G_u^{solv}) of the solute u (here, u refers to the complex or one of the two free proteins and excludes hydration waters); the bar denotes the average over the simulated configurations; S_{config} is the configurational entropy associated with the solute's internal degrees of freedom; and ΔS_{ext} is the entropy change originating from the reduction in the external (positional and orientational) degrees of freedom upon complex formation. ΔS_{ext} carries the standard-state dependence, which is chosen here to be the one of the standard concentration (1 M).

We computed the gas-phase energy E_u from the force field adopted in the simulations. (E_u for free proteins represents only the intra-protein energy, but E_u for the complex includes the inter-protein interaction energy as well). For the solvation free energy G_u^{solv}, we employed the 3D-RISM theory[30] (see Supplementary Methods). For the configurational entropy S_{config}, we used an energetic approach[43, 44] that expresses S_{config} in terms of the statistical properties of f_u. In particular, when the probability distribution $W(f_u)$ of f_u is Gaussian, the following holds:

$$T\,S_{\text{config}} = \frac{1}{2k_{\text{B}}T}\overline{\delta f_u^2} \tag{4}$$
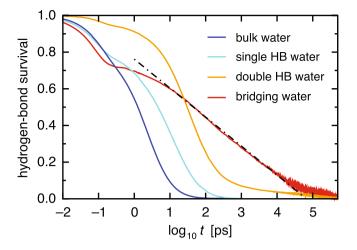
Figure 4. Hydrogen-bond time-correlation functions for bulk water (blue), single HB water (cyan), double HB water (orange), and bridging water (red) versus the logarithmic time axis. Dashed-dotted line denotes the fit by a logarithmic function.

	Bulk water	Single HB water	Double HB water	Bridging water
Relaxation time $\langle \tau_i \rangle^b$	5.1 ± 1.6	28.1 ± 10.1	163.3 ± 83.7	4954 ± 4418
$(\langle \log_{10} \tau_i \rangle)^c$	(0.69 ± 0.13)	(1.42 ± 0.16)	(2.17 ± 0.20)	(3.52 ± 0.39)
Trapping free energy^d	0	-4.5 ± 3.4	-8.0 ± 4.1	-12.5 ± 4.7

Table 2. Average relaxation time and trapping free energy^a. ^{a}Average $\pm$ standard deviation; ^{b}Relaxation time in ps computed with individual relaxation times τ_i; ^{c}Statistics computed with $\log_{10} \tau_i$; ^{d}Trapping free energy in kcal/mol.

where k_B is Boltzmann's constant, and $\delta f_u = f_u - \overline{f_u}$. For the external entropy ΔS_{ext}, we use the estimate $T\Delta S_{ext} = -6.8 \pm 0.1$ kcal/mol for the barnase–barstar complex, which was computed in ref. [22]. by extending the energetic approach to the binding process and is close to the value reported in ref. [45].

Explicit inclusion of the water molecules of interest. A statistical thermodynamic formulation of the binding free energy which allows one to explicitly include certain solvent molecules was also derived in ref. [42]. In essence and using our notation, what is required is to replace $f_u = E_u + G_u^{solv}$, which appears in equation (3) for ΔG_{bind}^0, with

$$f_u = E_u + G_u^{solv} - nG_{water}^{solv} \tag{5}$$

In this expression, the solute u now explicitly includes the water molecules of interest (e.g., E_u now contains interactions with and among those water molecules), n is the number of water molecules included, and G_{water}^{solv} is the single water molecule's solvation free energy. S_{config} then needs to be evaluated by combining equations (4) and (5).

Results and Discussion

Hydrogen-bond rearrangement dynamics.
We study the dynamic and thermodynamic features of the hydration water surrounding the barnase–barstar complex by conducting molecular dynamics simulations and statistical thermodynamic analyses. In particular, we aim to uncover the distinctive characteristics of the interfacial water between two proteins that emerge upon complex formation. This is done by contrasting the dynamics of the interfacial water with that of the hydration water surrounding free proteins; for this purpose, we also perform simulations and analyses for the free barnase and barstar proteins. We focus on the rearrangements of hydrogen bonds, which are the most important protein–water interactions because of the largely hydrophilic nature of the protein–protein binding surfaces[21].

Figure 4 shows the hydrogen-bond time-correlation functions, which quantify the extent to which hydrogen bonds found at time $t = 0$ remain at subsequent times t. For water molecules making a single hydrogen bond to a protein (referred to as single HB water; see Fig. 2 and Table 1), we observe profound slowing down of the relaxation dynamics compared to those of bulk water (see Table 2 for a comparison of the average relaxation times). For water molecules making two or more hydrogen bonds to a protein (double HB water), the hydrogen-bond rearrangement is even slower. For bridging water molecules, that is, interfacial water molecules making concurrent hydrogen bonds with two proteins, the relaxation is extraordinarily slow (~1,000 times slower than the relaxation in bulk water). Furthermore, the relaxation curve is anomalous, exhibiting a logarithmic decay over three orders of magnitude in time.
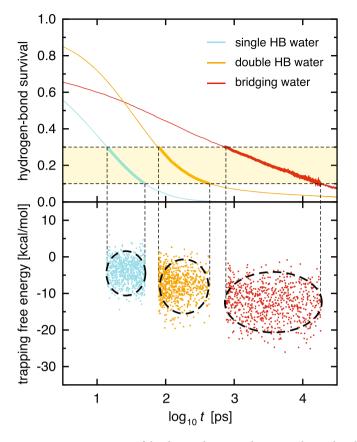
Figure 5. Construction of the thermodynamic–dynamic relationship diagram. Upper panel: Hydrogen-bond time-correlation functions for single HB water (cyan), double HB water (orange), and bridging water (red) taken from Fig. 4, focusing on the time regime where the functions decay from 0.3 to 0.1 (light yellow). Lower panel: Scatter plots of the hydrogen-bond survival times (τ_i) and trapping free energies (G_i^{trap}) of individual water molecules. Centers of ellipsoids are determined by the averages, and the width and hight are determined by 3.6 σ (where σ is the standard deviation) along each axis.

Thermodynamic–dynamic relationship diagram. To rationalize the slow relaxations of hydration waters, we conducted statistical thermodynamic analyses. We focused on the long-time region where the time-correlation functions decay from 0.3 to 0.1 (light yellow region in the upper panel of Fig. 5) and extracted the water molecules contributing to the relaxation there by examining the hydrogen-bond survival times (τ_i) of individual molecules. For each of those water molecules, we computed the trapping free energy (G_i^{trap}) using the simulation snapshot at $t = 0$. The trapping free energy can be considered as the effective potential characterizing how strongly each water molecule is captured by the biomolecular surface: a more negative trapping free energy means that a water molecule is more stable near the protein complex than in the bulk, and hence, is more favorably "trapped" by the protein complex. The lower panel of Fig. 5 shows scatter plots of the relaxation times and trapping free energies of individual water molecules. (The average relaxation times and trapping free energies listed in Table 2 are obtained from these plots. Since the distribution of the individual waters' relaxation times τ_i is well represented on the logarithmic axis as shown in the lower panel of Fig. 5, Table 2 also provides the statistics computed with $\log_{10} \tau_i$). The resulting "thermodynamic–dynamic relationship diagram" clearly illustrates that the trapping free energy (G_i^{trap}) at $t = 0$ serves as a good descriptor of the degree of retardation (τ_i) of the subsequent dynamics of hydration water.

Thermodynamic–dynamic relationship diagram is presented schematically in Fig. 6a. Single HB water exhibits slower dynamics than bulk water because it is more stable near the protein surface, which in turn reflects the fact that the hydrogen bond between water and protein is stronger than the one between waters. The even slower dynamics of double HB water can be understood similarly; further stabilization originates from an additional water–protein hydrogen bond. Why, then, is bridging water, which has the comparable number of hydrogen bonds with proteins as double HB water (Table 1), more strongly trapped than double HB water? We also notice here that the dynamic, as well as thermodynamic, characteristics of single and double HB water molecules are nearly the same, irrespective of whether they are placed near the isolated monomer or protein–protein complex (Supplementary Figs S1 and S2). The emergence of the "red region" for bridging water in the diagram (Fig. 6a) thus arises solely from the formation of the protein–protein interface. Is there any special factor that is effective only at the interface?

We notice in this regard the electrostatic complementarity of the barnase–barstar binding surfaces (Fig. 6b), which creates a strong electrostatic field that is exerted on the interfacial water. Indeed, we find that the magnitude
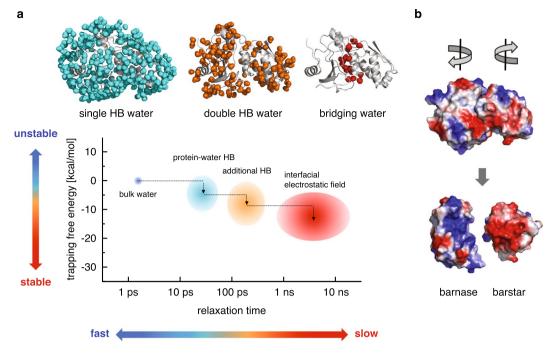
Figure 6. (**a**) Schematic representation of the thermodynamic–dynamic relationship diagram of hydration water. Here, the lower panel from Fig. 5 is schematically redrawn for single HB water (cyan), double HB water (orange), and bridging water (red), along with the position for the bulk water (blue). (**b**) Protein surfaces color-coded with the charge distribution (blue and red for positively and negatively charged regions, respectively).

of the electrostatic field is stronger and the water's dipole vector is more oriented along the electrostatic field for bridging water than for single and double HB water (Supplementary Fig. S3). Thus, whereas essentially no change in the hydration water dynamics is observed in the non-interfacial region before and after the binding (Supplementary Figs S1 and S2), the strong electrostatic field created at the binding interface produces an extra stabilizing factor for bridging water, causing it to exhibit extremely slow (nanosecond timescale) hydrogen-bond relaxations (Table 2). It would be interesting to investigate how the transition in the hydration water dynamics occurs during the binding process, but for this purpose, one needs to perform spontaneous binding simulations. The trapping free energy introduced in the present work serves as a valuable quantity not only to characterize the formation of the binding interface from the water's perspective, but also to discuss how and whether the hydration water is rearranging to go from the unbound protein to bound complex. While the barnase–barstar complex studied here is known to be a system in which the interfacial waters are particularly immobile[46], we anticipate the emergence of the extremely slow water relaxations to be a generic feature of hydrophilic protein–protein interfaces because electrostatic complementarity of the binding surfaces has been observed in numerous protein complexes[47, 48].

"Conventional" binding thermodynamics. Among the terms that contribute to ΔG_{bind}^0 (see equation (3)), the quantity $\Delta f_u = \Delta E_u + \Delta G_u^{\text{solv}}$ plays a special role for two reasons. First, it is generally difficult to compute the configurational (ΔS_{config}) and external (ΔS_{ext}) entropies for complex macromolecules such as proteins. Second, ΔS_{config} and ΔS_{ext} are usually negative upon complex formation and thus make unfavorable positive contributions to ΔG_{bind}^0; hence, the driving force for binding must originate from Δf_u. Indeed, Δf_u is the central quantity, termed the effective binding free energy[49], in computational approaches to biomolecular bindings such as the molecular-mechanics Poisson-Boltzmann surface area (MM-PBSA) method[50–52].

We computed $\overline{\Delta f_u}$ by averaging Δf_u over the simulated protein conformations. (Our approach is referred to as the three-trajectory approach because we conducted separate computations for the complex and for two free proteins. Numerical values for the binding thermodynamics shall be reported with standard errors computed based on the respective independent trajectories of the complex and free proteins and on the rule of error propagation). The energetic contributions (ΔE_u) were calculated directly from the force field, and the solvation contributions (ΔG_u^{solv}) were computed using the 3D-RISM theory (see Supplementary Methods). We obtained $\overline{\Delta f_u} = +25.7 \pm 2.6\,\text{kcal/mol}$; this result leads to an unphysical positive value of ΔG_{bind}^0, which is in obvious disagreement with the experimental observation ($\Delta G_{\text{bind}}^0 = -18.9\,\text{kcal/mol}$)[53]. Interestingly, positive effective binding free energy has also been reported based on the MM-PBSA calculations for the barnase–barstar complex: $\overline{\Delta f_u} = +14\,\text{kcal/mol}$ in ref. 54 and $\overline{\Delta f_u} = +3.6\,\text{kcal/mol}$ in ref. 55. (The difference in these values may originate from the use of the one-trajectory approach in the MM-PBSA calculations, in which both the complex and monomer configurations were taken from simulations of the complex; the use of different force fields; and the use of different approximations for the solvation free energy).

Basic assumption behind the conventional approach. At this point, we critically examine the basic assumption behind the expression (3) for the standard binding free energy. To simplify the discussion, we work in the canonical ensemble and ignore the external entropy, which would not alter the essential point here. We start from the configuration integral, the potential part of the partition function, for a solute-solvent system:

$$Z_{\text{tot}} = \int d\mathbf{r}_u \int d\mathbf{r}_v e^{-\beta[E_u(\mathbf{r}_u) + E_{uv}(\mathbf{r}_u, \mathbf{r}_v) + E_v(\mathbf{r}_v)]} \tag{6}$$

Here, $\mathbf{r}_u$ and $\mathbf{r}_v$ collectively denote the solute and solvent degrees of freedom, respectively; $\beta = 1/(k_{\text{B}}T)$ is the inverse temperature; and E_u, E_{uv}, and E_v are the solute energy, solute-solvent interaction energy, and solvent-solvent interaction energy, respectively. Z_{tot} is the principal object in free energy simulations: the change in the free energy $F_{\text{tot}} = -k_{\text{B}}T \log Z_{\text{tot}}$, e.g., upon mutation, is computed from simulations in which both the solute and solvent degrees of freedom are explicitly handled. However, equation (6) does not serve as a basis of equation (3): for example, by introducing the probability distribution $p_{\text{tot}}(\mathbf{r}_u, \mathbf{r}_v) = e^{-\beta E_{\text{tot}}(\mathbf{r}_u, \mathbf{r}_v)}/Z_{\text{tot}}$ with $E_{\text{tot}} = E_u + E_{uv} + E_v$, the entropy that naturally arises from equation (6) is $S_{\text{tot}} = -k_{\text{B}} \int d\mathbf{r}_u \int d\mathbf{r}_v p_{\text{tot}}(\mathbf{r}_u, \mathbf{r}_v) \log p_{\text{tot}}(\mathbf{r}_u, \mathbf{r}_v)$, and it is non-trivial to partition the solute and solvent terms from this total entropy. This is in contrast to equation (3) where the solute (S_{config}) and solvent (contained in G_u^{solv}) entropies are separated.

To arrive at equation (3) from equation (6), one has to resort to a pre-averaging of solvent degrees of freedom. For a given solute configuration $\mathbf{r}_u$, this pre-averaging can be performed as

$$e^{-\beta G_u^{\text{solv}}(\mathbf{r}_u)} = \frac{1}{Z_v} \int d\mathbf{r}_v e^{-\beta[E_{uv}(\mathbf{r}_u, \mathbf{r}_v) + E_v(\mathbf{r}_v)]} \tag{7}$$

in terms of the solute-configuration dependent solvation free energy $G_u^{\text{solv}}(\mathbf{r}_u)$. Here, $Z_v = \int d\mathbf{r}_v e^{-\beta E_v(\mathbf{r}_v)}$ is the configuration integral for the pure solvent. Now, the configuration integral after the pre-averaging of the solvent degrees of freedom is given by

$$Z \equiv Z_{\text{tot}}/Z_v = \int d\mathbf{r}_u e^{-\beta f_u(\mathbf{r}_u)} \tag{8}$$

with $f_u = E_u + G_u^{\text{solv}}$. By introducing $p(\mathbf{r}_u) = e^{-\beta f_u(\mathbf{r}_u)}/Z$, the associated entropy is given by $-k_{\text{B}} \int d\mathbf{r}_u p(\mathbf{r}_u) \log p(\mathbf{r}_u)$, which is the defining equation for the solute configurational entropy S_{config}. It is therefore clear that equation (3) is based on the pre-averaging of solvent degrees of freedom (see refs 22 and 42 for a complete derivation of equation (3) from equation (8)). By the "conventional" approach, we mean the one that is based on this pre-averaging, and do not refer to specific methods such as PBSA and 3D-RISM.

In practical applications of the conventional approach, one takes only the protein conformations from simulation trajectories, replacing all the explicit water molecules by the equilibrium continuum model (PBSA) or molecular distribution function (3D-RISM). Such a treatment is usually justified because of the timescale separation between the typical water dynamics (picoseconds) and the protein conformational motions (nanoseconds)[16], i.e., because of the rapid equilibration of surrounding waters. However, the extreme slowness of the bridging-water relaxation may invalidate such a naive treatment of the water at biomolecular interfaces, and we conjectured that this might be the origin of the unphysical positive value of $\overline{\Delta f_u}$.

Explicit inclusion of bridging water. We therefore investigated the impact of explicit inclusion of the slow bridging waters. For this purpose, we not only take the protein configurations from the simulation trajectories for the complex, but also bridging waters located at the interface: the number (n) of bridging waters depends on the simulation snapshot, and its average value is 19.6 ± 3.0 (Table 1). Now, those bridging waters are considered as a structural part of the complex, and we apply equation (5) to compute f_u for the complex. We obtain a negative value $\overline{\Delta f_u} = -34.2 \pm 2.1 \, \text{kcal/mol}$, which now serves as the driving force for binding. This result indicates the necessity of considering the dynamics of the interfacial water in the binding thermodynamics.

To support our explicit inclusion of bridging waters through a comparison with experiment, we computed the binding free energy ΔG_{bind}^0. To this end, we need to estimate the configurational (ΔS_{config}) and external (ΔS_{ext}) entropy contributions. For the configurational entropy, we used an energetic approach[43, 44] that expresses ΔS_{config} in terms of the fluctuations in f_u. In particular, when the probability distribution $W(f_u)$ of f_u is Gaussian, S_{config} is simply given by the mean-squared fluctuations of f_u (see equation (4)). Indeed, $W(f_u)$ of the barnase–barstar complex with bridging water is well approximated by Gaussian, as well as that of the free barnase and barstar proteins (Supplementary Fig. S4), from which $T\Delta S_{\text{config}} = -4.5 \pm 18.5 \, \text{kcal/mol}$ is obtained. For the external entropy ΔS_{ext}, we use the estimate $T\Delta S_{\text{ext}} = -6.8 \pm 0.1 \, \text{kcal/mol}$, which was obtained using the method developed in ref. 22 and is close to the value reported in ref. 45. Combining all these results, we obtain $\Delta G_{\text{bind}}^0 = -22.9 \pm 18.6 \, \text{kcal/mol}$, which is in reasonable accord with experiment ($-18.9 \, \text{kcal/mol}$)[53]. (As can be inferred from the numerical values presented above, the large standard error for ΔG_{bind}^0 originates from that for $T\Delta S_{\text{config}}$; indeed, the protein configurational entropy is known as the most difficult thermodynamic parameter to estimate).

Conclusions

Water molecules are ubiquitously found at the interfaces between biomolecules, and it is often stated that the interfacial water must be considered as an integral part of biomolecules. The work presented here sheds new light on this statement based on the dynamic viewpoint. We demonstrate the emergence of "special" waters in the interfacial region that bridge two biomolecules through concurrent hydrogen bonds and exhibit extremely slow hydrogen-bond rearrangements. By analyzing the thermodynamic–dynamic relationship diagram, we find that the extremely slow nature of bridging water is due to not only the number of hydrogen bonds involved,

but also the additional stabilization resulting from the strong electrostatic field between the binding surfaces of electrostatic complementarity. The role of such slow interfacial waters in determining the binding affinity cannot be described using the conventional approach to the water-mediated interaction, which assumes rapid equilibration of the waters' degrees of freedom. Indeed, we observe that a meaningful estimate of the binding affinity is achieved only with a unified treatment of both the biomolecules and the interfacial bridging water. Our work thus demonstrates the impact of the hydration dynamics on the protein–protein binding thermodynamics.

References

1. Chaplin, M. Do we underestimate the importance of water in cell biology? *Nat. Rev. Mol. Cell Biol.* **7**, 861–866 (2006).
2. Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108 (2008).
3. Whitesides, G. M. Reinventing chemistry. *Angew. Chem. Int. Ed.* **54**, 3196–3209 (2015).
4. Fuxreiter, M., Mezei, M., Simon, I. & Osman, R. Interfacial water as a "hydration fingerprint" in the noncognate complex of bamhi. *Biophys. J.* **89**, 903–911 (2005).
5. Grossman, M. *et al.* Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site. *Nat. Struct. Mol. Biol.* **18**, 1102–1108 (2011).
6. Capponi, S., Heyden, M., Bondar, A.-N., Tobias, D. J. & White, S. H. Anomalous behavior of water inside the secy translocon. *Proc. Natl. Acad. Sci. USA* **112**, 9016–9021 (2015).
7. Nucci, N. V., Pometun, M. S. & Wand, A. J. Site-resolved measurement of water–protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* **18**, 245–249 (2011).
8. Sterpone, F., Stirnemann, G. & Laage, D. Magnitude and molecular origin of water slowdown next to a protein. *J. Am. Chem. Soc.* **134**, 4116–4119 (2012).
9. King, J. T. & Kubarych, K. J. Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2d-ir spectroscopy. *J. Am. Chem. Soc.* **134**, 18705–18712 (2012).
10. Heyden, M. & Tobias, D. J. Spatial dependence of protein–water collective hydrogen-bond dynamics. *Phys. Rev. Lett.* **111**, 218101 (2013).
11. Hussain, S., Franck, J. M. & Han, S. Transmembrane protein activation refined by site-specific hydration dynamics. *Angew. Chem. Int. Ed.* **52**, 1953–1958 (2013).
12. Nibali, V. C. & Havenith, M. New insights into the role of water in biological function: Studying solvated biomolecules using terahertz absorption spectroscopy in conjunction with molecular dynamics simulations. *J. Am. Chem. Soc.* **136**, 12800–12807 (2014).
13. Lewandowski, J. R., Halse, M. E., Blackledge, M. & Emsley, L. Direct observation of hierarchical protein dynamics. *Science* **348**, 578–581 (2015).
14. Bellissent-Funel, M.-C. *et al.* Water determines the structure and dynamics of proteins. *Chem. Rev.* **116**, 7673–7697 (2016).
15. 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* **33**, 8878–8889 (1994).
16. Papoian, G. A., Ulander, J. & Wolynes, P. G. Role of water mediated interactions in protein–protein recognition landscapes. *J. Am. Chem. Soc.* **125**, 9170–9178 (2003).
17. Li, Z. & Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **9**, 573–581 (2007).
18. Wong, S., Amaro, R. E. & McCammon, J. A. Mm-pbsa captures key role of intercalating water molecules at a protein–protein interface. *J. Chem. Theory Comput.* **5**, 422–429 (2009).
19. Maffucci, I. & Contini, A. Explicit ligand hydration shells improve the correlation between mm-pb/gbsa binding energies and experimental activities. *J. Chem. Theory Comput.* **9**, 2706–2717 (2013).
20. Ikura, T., Urakubo, Y. & Ito, N. Water-mediated interaction at a protein–protein interface. *Chem. Phys.* **307**, 111–119 (2004).
21. Ansari, S. & Helms, V. Statistical analysis of predominantly transient protein–protein interfaces. *Proteins* **61**, 344–355 (2005).
22. Chong, S.-H. & Ham, S. New computational approach for external entropy in protein–protein binding. *J. Chem. Theory Comput.* **12**, 2509–2516 (2016).
23. Bycroft, M., Ludvigsen, S., Fersht, A. R. & Poulsen, F. M. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry* **30**, 8697–8701 (1991).
24. Lubienski, M. J., Bycroft, M., Freund, S. M. V. & Fersht, A. R. Three-dimensional solution structure and 13c assignments of barstar using nuclear magnetic resonance spectroscopy. *Biochemistry* **33**, 8866–8877 (1994).
25. Case, D. A. *et al.* AMBER 14 (University of California, San Francisco, 2014).
26. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).
27. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
28. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
29. Bagchi, B. Water dynamics in the hydration layer around proteins and micelles. *Chem. Rev.* **105**, 3197–3219 (2005).
30. Imai, T., Harano, Y., Kinoshita, M., Kovalenko, A. & Hirata, F. A theoretical analysis on hydration thermodynamics of proteins. *J. Chem. Phys.* **125**, 024911 (2006).
31. Chong, S.-H. & Ham, S. Atomic decomposition of the protein solvation free energy and its application to amyloid-beta protein in water. *J. Chem. Phys.* **135**, 034506 (2011).
32. Chong, S.-H. & Ham, S. Distinct role of hydration water in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis. *Acc. Chem. Res.* **48**, 956–965 (2015).
33. Li, Z. & Lazaridis, T. Thermodynamic contributions of the ordered water molecule in hiv-1 protease. *J. Am. Chem. Soc.* **125**, 6636–6637 (2003).
34. Abel, R., Young, T., Farid, R., Berne, B. J. & Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor xa ligand binding. *J. Am. Chem. Soc.* **130**, 2817–2831 (2008).
35. Nguyen, C. N., Young, T. K. & Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys* **137**, 044101 (2012).
36. Gerogiokas, G. *et al.* Prediction of small molecule hydration thermodynamics with grid cell theory. *J. Chem. Theory Comput.* **10**, 35–48 (2014).
37. Michel, J. *et al.* Evaluation of host–guest binding thermodynamics of model cavities with grid cell theory. *J. Chem. Theory Comput.* **10**, 4055–4068 (2014).
38. Gerogiokas, G. *et al.* Assessment of hydration thermodynamics at protein interfaces with grid cell theory. *J. Phys. Chem. B* **120**, 10442–10452 (2016).
39. Hamelberg, D. & McCammon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *J. Am. Chem. Soc.* **126**, 7683–7689 (2004).
40. Bodnarchuk, M. S., Viner, R., Michel, J. & Essex, J. W. Strategies to calculate water binding free energies in protein–ligand complexes. *J. Chem. Inf. Model.* **54**, 1623–1633 (2014).

41. Ross, G. A., Bodnarchuk, M. S. & Essex, J. W. Water sites, networks, and free energies with grand canonical monte carlo. *J. Am. Chem. Soc.* **137**, 14930–14943 (2015).
42. Gilson, M. K., Given, J. A., Bush, B. L. & McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **72**, 1047–1069 (1997).
43. Chong, S.-H. & Ham, S. Configurational entropy of protein: A combined approach based on molecular simulation and integral-equation theory of liquids. *Chem. Phys. Lett.* **504**, 225–229 (2011).
44. Chong, S.-H. & Ham, S. Protein folding thermodynamics: A new computational approach. *J. Phys. Chem. B* **118**, 5017–5025 (2014).
45. Irudayam, S. J. & Henchman, R. H. Entropic cost of protein–ligand binding and its dependence on the entropy in solution. *J. Phys. Chem. B* **113**, 5871–5884 (2009).
46. Copie, G., Cleri, F., Blossey, R. & Lensink, M. F. On the ability of molecular dynamics simulation and continuum electrostatics to treat interfacial water molecules in protein–protein complexes. *Sci. Rep.* **6**, 38259 (2016).
47. McCoy, A. J., Epa, V. C. & Colman, P. M. Electrostatic complementarity at protein/protein interfaces. *J. Mol. Biol.* **268**, 570–584 (1997).
48. Ulucan, O., Jaitly, T. & Helms, V. Energetics of hydrophilic protein–protein association and the role of water. *J. Chem. Theory Comput.* **10**, 3512–3524 (2014).
49. Gohlke, H. & Case, D. A. Converging free energy estimates: Mm-pb(gb)sa studies on the protein–protein complex ras–raf. *J. Comput. Chem.* **25**, 238–250 (2004).
50. Kollman, P. A. *et al.* Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897 (2000).
51. Sheinerman, F. B., Norel, R. & Honig, B. Electrostatic aspects of protein–protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159 (2000).
52. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed.* **41**, 2644–2676 (2002).
53. Schreiber, G. & Fersht, A. R. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32**, 5145–5150 (1993).
54. Lee, L.-P. & Tidor, B. Optimization of binding electrostatics: Charge complementarity in the barnase–barstar protein complex. *Protein Sci.* **10**, 362–377 (2001).
55. Sheinerman, F. B. & Honig, B. On the role of electrostatic interactions in the design of protein–protein interfaces. *J. Mol. Biol.* **318**, 161–177 (2002).

Acknowledgements

Author Contributions

S.-H.C. and S.H. conducted the research and wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-09466-w

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.