# Article

# Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations

David J. Huggins[1,*]

[1]Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK

ABSTRACT    Protein structural analysis demonstrates that water molecules are commonly found in the internal cavities of proteins. Analysis of experimental data on the entropies of inorganic crystals suggests that the entropic cost of transferring such a water molecule to a protein cavity will not typically be greater than 7.0 cal/mol/K per water molecule, corresponding to a contribution of approximately $+2.0$ kcal/mol to the free energy. In this study, we employ the statistical mechanical method of inhomogeneous fluid solvation theory to quantify the enthalpic and entropic contributions of individual water molecules in 19 protein cavities across five different proteins. We utilize information theory to develop a rigorous estimate of the total two-particle entropy, yielding a complete framework to calculate hydration free energies. We show that predictions from inhomogeneous fluid solvation theory are in excellent agreement with predictions from free energy perturbation (FEP) and that these predictions are consistent with experimental estimates. However, the results suggest that water molecules in protein cavities containing charged residues may be subject to entropy changes that contribute more than $+2.0$ kcal/mol to the free energy. In all cases, these unfavorable entropy changes are predicted to be dominated by highly favorable enthalpy changes. These findings are relevant to the study of bridging water molecules at protein-protein interfaces as well as in complexes with cognate ligands and small-molecule inhibitors.

## INTRODUCTION

Experimental techniques such as x-ray crystallography commonly identify water molecules in the internal cavities of proteins (1,2). A number of analyses on databases of protein structures have shown that this is a common observation and that protein cavities commonly accommodate one to three water molecules (3,4). Such water molecules tend to be stabilized by hydrogen bonding interactions and are thus expected to exhibit strong translational and orientational ordering (5). Numerous computational studies have been performed to quantify the binding free energy of such water molecules using free energy methods (6–9). In this context, the binding free energy is the free energy of transfer of a water molecule from the bulk to the cavity. As expected, the binding free energies for crystallographically observed water molecules in protein cavities are predicted to be favorable. However, the degree of ordering relative to bulk water may mean that this favorable free energy is comprised of a favorable enthalpy component and an unfavorable entropy component (5). Analysis from Dunitz calculated the entropy contribution of individual water molecules to the entropy of inorganic crystals (10). This analysis suggests that an entropic cost of transferring an individual water molecule to a protein cavity will not typically be greater than 7.0 cal/mol/K. This corresponds to a contribution of

approximately $+2.0$ kcal/mol to the free energy. From this data, the conclusion was that water molecules in proteins are unlikely to make an entropic contribution of greater than $+2.0$ kcal/mol to the free energy. In this study, we employ the statistical mechanical method of inhomogeneous fluid solvation theory (IFST) (5,11–14) to quantify the enthalpic and entropic contributions of individual water molecules in protein cavities. IFST has previously been used to rationalize kinase selectivity (15), identify ligand-binding hotspots at protein surfaces (16), and understand the hydrophobic effect (17). It has proven a very useful tool in modeling networks of water molecules in biological systems. IFST has also been used to study water molecules in cavities of IL-1β (18), and in this article we extend such a study to five different proteins and nineteen protein cavities. In addition, we combine the k-nearest neighbors (KNN) (19,20) algorithm with information theory (21–24) to study doubly occupied cavities and explicitly calculate the contributions of changes in water-water correlations and the associated entropy changes. To our knowledge, this approach represents the first estimate of this quantity using mutual information and provides a complete framework to calculate two-particle entropies, as envisioned in the original development of IFST (11,12).

## MATERIALS AND METHODS

Five test systems were chosen for this study: IL-1β (2), T4 Lysozyme (1), FKBP-2, Carbonic Anhydrase (CA-II) (25), and β-Lactamase (26) (see Table 1). From these five structures, 19 internal cavities were identified. Fifteen of these cavities were singly occupied and four were doubly occupied, and thus the analysis considers 23 water molecules in total. Details of

**TABLE 1   The five proteins considered in this study along with crystallographic data, the number of singly and doubly occupied internal cavities, the initial charge of the protein before adding counterions, and the residues altered during the preparation stage**

| Protein | IL-1β | T4 Lysozyme | FKBP-2 | CA-II | β-Lactamase |
|---|---|---|---|---|---|
| PDB ID | 2NVH | 3DKE | 2PBC | 3GZ0 | 2P74 |
| Protein chain | A | X | A | A | A |
| Resolution (Å) | 1.53 | 1.25 | 1.8 | 1.26 | 0.88 |
| Single cavities | 2 | 2 | 1 | 5 | 5 |
| Double cavities | 2 | 1 | 1 | 0 | 0 |
| Initial charge | 0 | +9 | 0 | 0 | +1 |
| Residues altered | Q14 Flip | H31 Protonate | H55 ε-Hydrogen | H4 Protonated | H112 Flip |
|  | H30 Protonate | N68 Flip | Q 72 Flip | H10 ε-Hydrogen | Q31 Flip |
|  | Q48 Flip | Q69 Flip |  | H15 ε-Hydrogen | Q141 Flip |
|  | Q116 Flip | Q123 Flip |  | H17 ε-Hydrogen | Q206 Flip |
|  | N119 Flip | N140 Flip |  | H36 ε-Hydrogen | Q254 Flip |
|  | Q126 Flip | N144 Flip |  | H64 ε-Hydrogen |  |
|  | N129 Flip | N163 Flip |  | H119 ε-Hydrogen |  |
|  | Q149 Flip |  |  | N178 Flip |  |
|  |  |  |  | N253 Flip |  |

the PDB crystal structures and the residue numbers for these 23 water molecules are listed in Table 1.

## System setup

Protein structures were downloaded from the Protein Databank (27). Selenomethionines were changed to methionines and missing sidechains were added using Schrodinger's Preparation Wizard (28), which was also used to check the orientations of the asparagine, glutamine, and histidine residues, as well as the protonation state of all ionizable residues. All heteroatomic species such as buffer solvents and ions were removed, with the exception of the zinc ion in the case of Carbonic Anhydrase. The changes made to each structure to improve the hydrogen bonding patterns are detailed in Table 1. The hydrogen-atom positions were then built using the HBUILD facility of CHARMM (29) with the CHARMM27 energy function (30,31), and the force field parameters and partial charges were assigned from the CHARMM27 force field (30,31). Only the crystallographic water molecules identified in internal cavities were retained, with all others being deleted. Water molecules were modeled with the TIP4P-2005 water model (32). To ensure an overall charge of zero in the system, nine and one additional chloride ions were added in the cases of T4 Lysozyme and Beta-Lactamase respectively. All ions were placed farther than 15 Å from the protein.

## Molecular dynamics simulations

Equilibration was performed for 1.0 ns in an NVT ensemble at 300 K using Langevin temperature control (33). All systems were brought to equilibrium before continuing, by verifying that the energy fluctuations were stable. MD simulations were performed using an MD time step of 2.0 fs. Electrostatic interactions were modeled with a uniform dielectric and a dielectric constant of 1.0 throughout the equilibration and production runs. Van der Waals interactions were truncated at 11.0 Å with switching from 9.0 Å. Electrostatics were modeled using the particle mesh Ewald method (34), and the systems were treated using rhombic dodecahedral periodic boundary conditions. All non-water atoms were fixed for the entirety of the equilibration and production simulations. MD simulations were performed using NAMD version 2.9 (35). To explore the effect of protein conformation on the IFST results, we generated 10 conformations of T4 Lysozyme by running a 10.0 ns simulation with heavy-atom restraints of 1.0 kcal/mol/Å$^2$ and storing the coordinates every 1.0 ns. Each unique protein conformation was then used for a separate 10.0 ns IFST calculation with a fixed protein structure. For comparison, the 10.0 ns simulation with heavy-atom restraints was also analyzed.

## Free energy perturbation calculations

Free energy perturbation (FEP) calculations were performed for each internal cavity, to calculate the binding energy of the buried water molecules ($\Delta G_{bind}$). The total free energy change for an FEP calculation ($\Delta G_{FEP}$) was calculated as the sum of free energy changes for a series of N small steps between intermediate states a and b (36). The change in free energy was calculated for each small step ($\Delta G_{a \to b}$) using the partition functions ($Q$) for the two states, which are calculated from the Hamiltonians ($H$):

$$\Delta G_{FEP} = \sum_{a=1, \ b=a+1}^{N} \Delta G_{a \to b} \qquad (1)$$

$$\begin{aligned} \Delta G_{a \to b} &= G_b - G_a \\ &= -kT \ln\left(\frac{Q_b}{Q_a}\right) \\ &= -kT \ln\left(\langle \exp(-(H_b - H_a)/kT)\rangle_a\right) \end{aligned} \qquad (2)$$

The results for the forward and backward FEP simulations were combined using the Bennett Acceptance Ratio (BAR) method (37,38). BAR was implemented using the ParseFEP Plugin from Visual Molecular Dynamics and the statistical error was estimated in each case (39). The estimated statistical error in the FEP free energy predictions using BAR was less than 0.1 kcal/mol in all cases. We used 32 equally spaced λ windows for the forward FEP simulations and 32 equally spaced λ windows for the backward FEP simulations. A soft-core potential was employed with a van der Waals radius-shifting coefficient of 5.0 (40,41), electrostatic interactions were scaled down to zero between λ = 0.0 and λ = 0.5, and van der Waals interactions were scaled down to zero between λ = 0.5 and λ = 1.0 (38). Equilibration was performed for 250 ps for each lambda window, and production simulation was performed for 750 ps for each lambda window. An NVT ensemble was used throughout and all nonwater atoms were fixed for the entirety of the FEP simulations. The free energy cycle for calculating $\Delta G_{bind}$ can be seen in Fig. 1. The first step transfers the water molecule from 55 M to a fixed point in solution ($-\Delta G_{liberation}$), and the second step annihilates the water molecule from bulk ($-\Delta G_{insertion}$). $\Delta G_{insertion}$ was calculated to be −6.95 kcal/mol for a fixed TIP4P-2005 water molecule using FEP at 1 atm and 300 K in an NPT ensemble. The third step transfers the water molecule from the fixed location back to 55 M in vacuum ($\Delta G_{liberation}$), and thus the two liberation terms in the cycle cancel one another. The fourth step is to harmonically restrain the oxygen atom of the water molecule to aid convergence of the free energy calculations
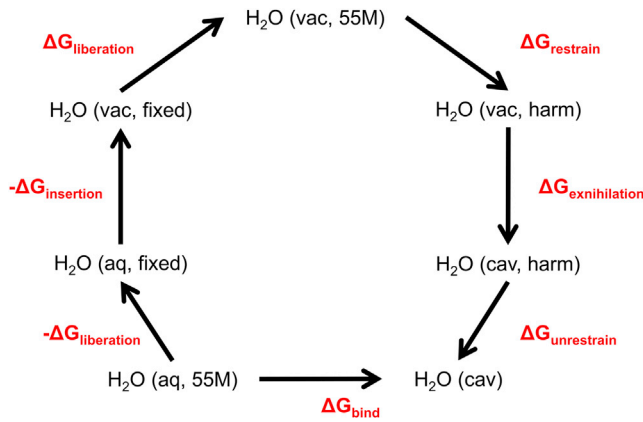
FIGURE 1  The steps in the free energy cycle used to calculate $\Delta G_{bind}$ for a water molecule. The term aq represents a water molecule in aqueous conditions, the term vac represents a water molecule in vacuum, and the term cav represents a water molecule in a protein cavity. 55 M represents a water molecule at the standard bulk concentration of 55 mol/dm$^3$, fixed represents a water molecule that is at a fixed point, and harm represents a water molecule that is harmonically restrained. To see this figure in color, go online.

(7,9,42,43). This harmonic restraint leads to an analytic free energy penalty ($\Delta G_{restrain}$) given by Eq. 3:

$$\Delta G_{restrain} = -RT \ln[C_0 V_{harm}] \quad (3)$$

$$V_{harm} = \left(\frac{2\pi RT}{k_{harm}}\right)^{3/2}, \quad (4)$$

where $C_0$ is the concentration of a water molecule in bulk (55 M), $V_{harm}$ is the volume available to the harmonically restrained water molecule, and $k_{harm}$ is the harmonic force constant. The value of $k_{harm}$ was set to 0.5 kcal/mol/Å$^2$ in all cases, and thus $\Delta G_{restrain}$ was calculated to be +0.23 kcal/mol. The fifth step is exhinilation of the water molecule in the cavity ($\Delta G_{exhinilation}$) using FEP. The final step is to remove the harmonic restraint ($\Delta G_{unrestrain}$) and this contribution is assumed to be zero, as in previous work, and is justified because the dynamics of the water molecule in the cavity are not affected by a force constant that is small in relation to the atomic fluctuations (9,18). For cavities containing two water molecules, the exhinilation is performed on both simultaneously and interactions between the exhinilated water molecules are also scaled to decouple them (18). Considering the steps described above, $\Delta G_{bind}$ can be calculated using Eq. 5:

$$\Delta G_{bind} = \Delta G_{exhinilation} - \Delta G_{insertion} + \Delta G_{restrain}. \quad (5)$$

The symmetry contribution to the binding free energy ($-0.41$ kcal/mol in the case of a water molecule) is only appropriate if there is a difference between the sampling of the symmetry-related states in the bound and unbound states (44). The unbound water molecules are treated as fixed and cannot sample the two symmetry-related states. In this case, the bound water molecules were not observed to sample the two symmetry-related states, presumably because of a large kinetic barrier. Thus, there is no symmetry contribution.

## IFST calculations

IFST calculates the hydration free energy of a solute ($\Delta G_{IFST}$) by computing the difference in free energy between a solution and the same number of solvent molecules ($n$) modeled as the pure bulk solvent (11,12). $\Delta G_{IFST}$ can be calculated for small subvolumes of the system, allowing the contribution of specific regions to be estimated (13,45,46). In

the context of protein systems, water molecules tend to cluster at distinct locations termed hydration sites, and it is natural to compute the contribution of individual water molecules to the hydration free energy (14,18,47,48). For the IFST calculations, 100.0 ns of production simulation in an NVT ensemble were performed at 300 K for each system. System snapshots were saved every 5.0 ps, yielding 20,000 snapshots in total for each system. We calculated a mean and standard deviation for each of the IFST quantities by considering 10 blocks from the 100 ns simulation, each derived from 2000 randomly selected snapshots. $\Delta G_{IFST}$ is calculated from the contributions of the hydration energy ($\Delta E_{IFST}$) and hydration entropy ($\Delta S_{IFST}$) to the hydration free energy:

$$\Delta G_{IFST} = \Delta E_{IFST} - T\Delta S_{IFST}. \quad (6)$$

$\Delta E_{IFST}$ is calculated from the mean solute-water interaction energy ($E_{sw}$), the mean water-water interaction energy ($E_{ww}$), and the mean interaction energy of a bulk water molecule ($E_{bulk}$):

$$\begin{aligned} \Delta E_{IFST} &= E_{sw} + E_{ww} - nE_{bulk} \\ &= E_{sw} + \Delta E_{ww} \end{aligned}. \quad (7)$$

$E_{bulk}$ and $E_{ww}$ are defined as half the interaction energy of a water molecule with all other water molecules in the system. For the TIP4P-2005 water model, $E_{bulk}$ is calculated to be $-11.5748$ kcal/mol. The total hydration entropy is calculated as the sum of local ($S_{local}$) and nonlocal contributions ($S_{nonlocal}$). $S_{local}$ is the summed contribution of each local subvolume and $S_{nonlocal}$ is the sum of the volume entropy ($S_{ve}$) and the change in liberation entropy ($\Delta S_{lib}$) (12). $S_{local}$ is calculated from the two-particle correlation function ($g_{sw}$) that is a function of the position ($r$) and orientation ($\omega$) of the water molecule. The number density of bulk TIP4P-2005 water ($\rho$) is calculated to be 0.03324 molecules/Å$^3$:

$$\begin{aligned} \Delta S_{local} &= -R\rho \int [g_{sw} \ln g_{sw} - g_{sw} + 1]dV + \Delta S_{ww} \\ &= -R\rho \int g_{sw} \ln g_{sw} dV - R\rho \int [1 - g_{sw}]dV + \Delta S_{ww} \\ &= \Delta S_{IFST} - R\rho \int [1 - g_{sw}]dV \end{aligned} \quad (8)$$

$$\begin{aligned} \Delta S_{nonlocal} &= S_{ve} + \Delta S_{lib} \\ &= -R(1 - \rho V_s) + R(\alpha T - \rho\kappa kT) \\ &= -R + R\rho V_s + R\alpha T - R\rho\kappa kT \\ &= -R + R\rho\left(\kappa kT + \int[1 - g_{sw}]dV\right) + R\alpha T - R\rho\kappa kT \\ &= R(\alpha T - 1) + R\rho \int[1 - g_{sw}]dV \end{aligned} \quad (9)$$

$$\begin{aligned} \Delta S_{hydration} &= S_{local} + S_{nonlocal} \\ &= \Delta S_{IFST} + R(\alpha T - 1) \\ &\approx \Delta S_{IFST} - R \end{aligned}, \quad (10)$$

where $V_s$ is the partial molar volume of the solute, $\alpha$ is the thermal expansion coefficient of the solvent, and $\kappa$ is the isothermal compressibility of the solvent. Equation 9 uses the Kirkwood-Buff relationship for $V_s$ (49). The two remaining integrals in Eqs. 8 and 9 can be understood as a local term corresponding to a reduction in the volume accessible to the solvent ($V_s$) and a non-local term corresponding to an increase in the volume accessible to the solvent ($V_s$) as the system expands under constant pressure. These two terms are equal and cancel one another. $\Delta S_{IFST}$ is calculated from the two-particle correlations using the solute-water entropy ($S_{sw}$)

and the difference in water-water entropy ($\Delta S_{ww}$); higher-order correlations are not considered:

$$\begin{aligned}\Delta S_{IFST} &= S_{sw} + S_{ww} - nS_{bulk} \\ &= S_{sw} + \Delta S_{ww}\end{aligned}. \qquad (11)$$

For the TIP4P-2005 water model, $S_{bulk}$ is calculated from the values for $E_{bulk}$ and $\Delta G_{insertion}$ to be -15.5097 cal/K/mol. In recent work, we developed a novel KNN approach to calculate $S_{sw}$ using the translational and orientational distance metrics (50). The translational distance ($d_{trans}$) between two water molecules is simply the Euclidean norm between the Cartesian coordinates of the two water molecules. The orientational distance ($d_{orient}$) between two water molecules is the distance between the rotations required to bring the two orientations to the same reference orientation. The correct distance metric for the rotation group is twice the geodesic distance on the unit sphere (51). The KNN algorithm provides an unbiased estimate of the absolute entropy ($H_{abs}$) from the general expression in Eq. 12:

$$H_{abs} = \frac{1}{n}\sum_{i=1}^{n} \ln\left[\frac{nd_{i,k}^{p}\pi^{p/2}}{\Gamma(p/2+1)}\right] - L_{k-1} + \gamma \qquad (12)$$

$$L_j = \sum_{i=1}^{j}\frac{1}{i}, \qquad (13)$$

where $n$ is the number of samples, $d_{i,k}$ is the distance between sample point $i$ and its kth nearest neighbor, $p$ is the number of degrees of freedom, $\Gamma$ is the gamma function, $L_0$ is 0, and $\gamma$ is Euler's constant. In the context of an individual water molecule, the relative solute-water thermodynamic entropy ($S_{sw}$) is calculated using the total distance ($d_{total}$) between two water molecules in six dimensions ($p = 6$) (50). It is calculated from the difference between the absolute entropy of the distribution ($H_{abs}$) and the absolute entropy of a uniform distribution ($H_{uni}$). We disregard the symmetry number (two in this case) as it is present in both $H_{abs}$ and $H_{uni}$. We use the first nearest neighbor in this work ($k = 1$) and the number of samples is equal to the number of frames where a water molecule is present in the hydration site:

$$\begin{aligned}S_{sw} &= R\{H_{abs} - H_{uni}\} \\ &= R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{nd_{total}^{6}\pi^{3}}{\Gamma(4)}\right] - L_0 + \gamma - \ln\left[\frac{8\pi^2}{\rho}\right]\right\} \\ &= R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{nd_{total}^{6}\pi^{3}}{6}\right] + \gamma - \frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{8\pi^2}{\rho}\right]\right\} \\ &= R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{nd_{total}^{6}\pi\rho}{48}\right] + \gamma\right\}\end{aligned}$$

$$(14)$$

$$d_{total} = \sqrt{d_{trans}^2 + d_{orient}^2}, \qquad (15)$$

$$d_{trans} = \sqrt{\left(x_{ij}-x_{kl}\right)^2 + \left(y_{ij}-y_{kl}\right)^2 + \left(z_{ij}-z_{kl}\right)^2}, \qquad (16)$$

$$d_{orient} = 2 \times \mathrm{acos}\left(\left|q_{ij}\cdot q_{kl}\right|\right), \qquad (17)$$

where $R$ is the gas constant and converts the entropy to thermodynamic units and $\Gamma(4)$ is equal to 6. The Cartesian coordinates of water molecule $j$ in frame $i$ and its nearest neighbor water molecule $k$ in frame $l$ are denoted by $x_{ij}, y_{ij}, z_{ij}$ and $x_{kl}, y_{kl}, z_{kl}$, respectively. The quaternion representations of the rotations for water molecule $j$ in frame $i$ and its nearest neighbor water

molecule $k$ in frame $l$ are denoted by $q_{ij}$ and $q_{kl}$, respectively. In this work, we extend the nearest neighbor approach to calculate $S_{ww}$ from the two-particle correlation functions ($g_{sw}$, $g_{sw'}$, and $g_{ww'}$) and the triplet correlation function ($g_{sww'}$), which is a function of the 12 variables representing the positions and orientations of a pair of water molecules ($p = 12$):

$$\begin{aligned}S_{ww} &= -\frac{1}{2}R\rho^2\iint g_{sw}g_{sw'}[g_{ww'}\ln g_{ww'} - g_{ww'} + 1]dwdw' \\ &= -\frac{1}{2}R\rho^2\iint g_{sw}g_{sw'}g_{ww'}\ln g_{ww'}dwdw' \\ &\quad -\frac{1}{2}R\rho^2\iint g_{sw}g_{sw'}[1 - g_{ww'}]dwdw' \\ &= -\frac{1}{2}R\rho^2\iint g_{sww'}\ln\left[\frac{g_{sww'}}{g_{sw}g_{sw'}}\right]dwdw' \\ &\quad -\frac{1}{2}R\rho^2\iint g_{sw}g_{sw'}[1 - g_{ww'}]dwdw' \\ &= -\sum_{pairs}[I_{ww} + S_{vol}]\end{aligned}.$$

$$(18)$$

$I_{ww}$ is best understood as a mutual information (MI) term as it represents the additional correlation between two water molecules that is not captured by the solute-water entropy term. The $S_{vol}$ term accounts for the exclusion of solvent from the volume occupied by other solvent molecules and is related to the Kirkwood-Buff integral (11,49). For a single water molecule, $S_{ww}$ can be calculated by considering the pair correlation with all other water molecules in the system. In this work, we restrict the sum to pairs of water molecules within 4.0 Å. One can easily consider $I_{ww}$ as a combination of two- and three-particle entropy terms:

$$\begin{aligned}I_{ww} &= \frac{1}{2}R\rho^2\iint g_{sww'}\ln\left[\frac{g_{sww'}}{g_{sw}g_{sw'}}\right]dwdw' \\ &= \frac{1}{2}R\rho^2\iint [g_{sww'}\ln g_{sww'} \\ &\quad - g_{sww'}\ln g_{sw} - g_{sww'}\ln g_{sw'}]dwdw' \\ &= S_{sww'} - S_{sw}^{*} - S_{sw'}^{*}\end{aligned}, \qquad (19)$$

where $S^{*}_{sw}$ and $S^{*}_{sw'}$ represent the two-particle entropies computed from the pair data and can be calculated using Eq. 14. The three-particle entropy ($S_{sww'}$) can be calculated from the total distance ($d_{pair}$) between two pairs of water molecules. In this case, both hydration sites are occupied in every snapshot for each doubly occupied cavity and thus all $n$ frames contain pair data:

$$\begin{aligned}S_{sww'} &= R\{H_{abs} - H_{uni}\} \\ &= R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{nd_{pair}^{12}\pi^{6}}{\Gamma(7)}\right] - L_0 + \gamma - \ln\left[\frac{64\pi^4}{\rho^2}\right]\right\} \\ &= R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{nd_{pair}^{12}\pi^{6}}{720}\right] + \gamma - \frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{64\pi^4}{\rho^2}\right]\right\} \\ &= R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{nd_{pair}^{12}\pi^{2}\rho^{2}}{46080}\right] + \gamma\right\}\end{aligned}$$

$$(20)$$

$$d_{pair} = \sqrt{d_{total}^2 + d_{total'}^2}. \tag{21}$$

$\Gamma(7)$ is equal to 720. In practice, problems arise from combining KNN terms of different dimensionality in Eq. 19 (52). Thus, we use the method of permuted fill modes as described by Hensen et al. (53) The permuted set of distances ($d_{perm}$) captures the correlation of the individual water molecules with the solute but decouples the correlation between the water molecules by computing the entropy of the artificially decorrelated data:

$$
\begin{aligned}
I_{ww} &= S_{sww'} - S_{sww'}^{perm} \\
&= R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{\rho^2\pi^2 n d_{pair}^{12}}{46080}\right] + \gamma\right\} \\
&\quad - R\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left[\frac{\rho^2\pi^2 n d_{perm}^{12}}{46080}\right] + \gamma\right\}. \\
&= \frac{R}{n}\sum_{i=1}^{n}\ln\left[\frac{d_{pair}^{12}}{d_{perm}^{12}}\right]
\end{aligned}
\tag{22}
$$

Whereas the individual entropy terms, such as the one defined in Eq. 20, obey a power law convergence (as previously observed), the MI estimate in Eq. 22 does not appear to obey a power law convergence. It is also interesting to note that this MI estimate is not biased, as the bias terms from the two entropy estimates cancel one another. Within a singly occupied cavity, $I_{ww}$ and $S_{vol}$ are negligible because there are no significant water-water pair correlations. In a cavity with two or more water molecules, only pairs of subvolumes in which $g_{sw}$ and $g_{sw'}$ are nonzero will make a contribution to $S_{vol}$. These regions are very small and $S_{vol}$ is thus expected to be negli-

gible. For this reason, we assume that $S_{vol}$ is zero in all cases. As a comparison, the magnitude of $S_{vol}$ in bulk water can be calculated from the radial distribution function and has a value of $-0.99$ cal/mol/K for the TIP4P-2005 water model, making a contribution of $+0.29$ kcal/mol to the excess free energy.

## RESULTS AND DISCUSSION

We begin by considering the IFST estimates of enthalpy, entropy, and free energy contributions of each water molecule to the protein hydration free energy. The calculations for the 23 sites are presented in Table 2.

The standard deviations are all below 0.1 kcal/mol and generally much smaller. The predicted free energies are in the range from $-2.6$ to $-16.0$ kcal/mol, and this is in excellent agreement with previous IFST studies that showed a range from $-1.9$ to $-17.2$ kcal/mol (5,14). The predicted enthalpies are larger in magnitude than the entropies and make the dominant contributions to the free energies for all 23 water molecules. This is also in agreement with previous studies. The mean contribution of $+1.82$ kcal/mol is in good agreement with the contribution of $+2.0$ kcal/mol estimated by Dunitz, and the maximum contribution of $+2.67$ kcal/mol is not far above this value. As expected, the most favorable $\Delta E_{IFST}$ and unfavorable $-T\Delta S_{IFST}$ are found for protein cavities containing charged amino acid sidechains. Table 2 also shows that the $I_{ww}$ term is small in magnitude for each pair of water molecules in the four

**TABLE 2** The results of the IFST calculations for the 23 hydration sites, with the mean and standard deviation from 10 blocks reported

| System | PDB water ID | $E_{sw}$ (kcal/mol) | $E_{ww}$ (kcal/mol) | $\Delta E_{IFST}$ (kcal/mol) | $-TS_{sw}$ (kcal/mol) | $-TS_{ww}$ (kcal/mol) | $-T\Delta S_{IFST}$ (kcal/mol) | $\Delta G_{IFST}$ (kcal/mol) |
|---|---|---|---|---|---|---|---|---|
| IL-1β | 202 | $-15.93 \pm 0.03$ | $-2.69 \pm 0.01$ | $-7.05 \pm 0.03$ | $6.35 \pm 0.04$ | $0.07 \pm 0.02$ | $1.80 \pm 0.05$ | $-5.25 \pm 0.04$ |
| IL-1β | 204 | $-16.73 \pm 0.02$ | $-2.72 \pm 0.01$ | $-7.88 \pm 0.02$ | $6.41 \pm 0.03$ | $0.07 \pm 0.02$ | $1.86 \pm 0.04$ | $-6.02 \pm 0.03$ |
| IL-1β | 203 | $-14.56 \pm 0.05$ | $-1.97 \pm 0.02$ | $-4.96 \pm 0.04$ | $6.63 \pm 0.05$ | $0.10 \pm 0.03$ | $2.11 \pm 0.06$ | $-2.85 \pm 0.04$ |
| IL-1β | 207 | $-13.75 \pm 0.02$ | $-1.99 \pm 0.02$ | $-4.16 \pm 0.03$ | $5.77 \pm 0.03$ | $0.10 \pm 0.03$ | $1.25 \pm 0.03$ | $-2.91 \pm 0.03$ |
| IL-1β | 200 | $-21.21 \pm 0.02$ | | $-9.48 \pm 0.02$ | $6.53 \pm 0.04$ | | $1.91 \pm 0.04$ | $-7.57 \pm 0.03$ |
| IL-1β | 209 | $-19.04 \pm 0.03$ | | $-7.33 \pm 0.03$ | $5.08 \pm 0.04$ | | $0.46 \pm 0.04$ | $-6.87 \pm 0.02$ |
| T4 Lysozyme | 902 | $-22.13 \pm 0.02$ | $-2.95 \pm 0.01$ | $-13.50 \pm 0.02$ | $6.72 \pm 0.03$ | $0.05 \pm 0.02$ | $2.16 \pm 0.03$ | $-11.34 \pm 0.04$ |
| T4 Lysozyme | 905 | $-18.67 \pm 0.02$ | $-2.97 \pm 0.01$ | $-10.06 \pm 0.02$ | $6.45 \pm 0.03$ | $0.05 \pm 0.02$ | $1.88 \pm 0.03$ | $-8.18 \pm 0.03$ |
| T4 Lysozyme | 904 | $-16.23 \pm 0.03$ | | $-4.65 \pm 0.03$ | $6.27 \pm 0.03$ | | $1.65 \pm 0.03$ | $-3.00 \pm 0.01$ |
| T4 Lysozyme | 920 | $-21.10 \pm 0.02$ | | $-9.54 \pm 0.02$ | $6.40 \pm 0.03$ | | $1.78 \pm 0.03$ | $-7.76 \pm 0.01$ |
| FKBP-2 | 207 | $-17.96 \pm 0.01$ | $-2.37 \pm 0.01$ | $-8.76 \pm 0.01$ | $6.01 \pm 0.03$ | $0.08 \pm 0.01$ | $1.48 \pm 0.03$ | $-7.29 \pm 0.02$ |
| FKBP-2 | 208 | $-19.54 \pm 0.03$ | $-2.44 \pm 0.01$ | $-10.40 \pm 0.02$ | $6.09 \pm 0.05$ | $0.08 \pm 0.01$ | $1.56 \pm 0.06$ | $-8.85 \pm 0.05$ |
| FKBP-2 | 203 | $-26.49 \pm 0.02$ | | $-14.91 \pm 0.02$ | $7.03 \pm 0.04$ | | $2.41 \pm 0.04$ | $-12.50 \pm 0.02$ |
| CA-II | 2004 | $-20.77 \pm 0.03$ | | $-9.47 \pm 0.03$ | $6.47 \pm 0.03$ | | $1.85 \pm 0.03$ | $-7.62 \pm 0.02$ |
| CA-II | 2015 | $-27.52 \pm 0.03$ | | $-15.92 \pm 0.03$ | $6.85 \pm 0.04$ | | $2.23 \pm 0.04$ | $-13.70 \pm 0.01$ |
| CA-II | 2031 | $-22.40 \pm 0.02$ | | $-11.09 \pm 0.02$ | $6.28 \pm 0.02$ | | $1.66 \pm 0.02$ | $-9.44 \pm 0.02$ |
| CA-II | 2042 | $-24.11 \pm 0.02$ | | $-12.56 \pm 0.02$ | $6.59 \pm 0.03$ | | $1.97 \pm 0.03$ | $-10.58 \pm 0.02$ |
| CA-II | 2055 | $-17.80 \pm 0.02$ | | $-6.20 \pm 0.02$ | $6.66 \pm 0.03$ | | $2.04 \pm 0.03$ | $-4.17 \pm 0.03$ |
| β-Lactamase | 2023 | $-16.08 \pm 0.03$ | | $-3.95 \pm 0.03$ | $6.02 \pm 0.04$ | | $1.40 \pm 0.04$ | $-2.55 \pm 0.02$ |
| β-Lactamase | 2048 | $-27.94 \pm 0.01$ | | $-16.33 \pm 0.01$ | $7.09 \pm 0.03$ | | $2.47 \pm 0.03$ | $-13.87 \pm 0.03$ |
| β-Lactamase | 2073 | $-28.60 \pm 0.03$ | | $-16.47 \pm 0.03$ | $6.52 \pm 0.03$ | | $1.90 \pm 0.03$ | $-14.57 \pm 0.02$ |
| β-Lactamase | 2105 | $-24.70 \pm 0.02$ | | $-13.08 \pm 0.02$ | $5.93 \pm 0.05$ | | $1.31 \pm 0.05$ | $-11.77 \pm 0.04$ |
| β-Lactamase | 2327 | $-30.27 \pm 0.02$ | | $-18.70 \pm 0.02$ | $7.29 \pm 0.02$ | | $2.67 \pm 0.02$ | $-16.03 \pm 0.01$ |
| Minimum | | | | $-18.70$ | | | $0.46$ | $-16.03$ |
| Maximum | | | | $-3.95$ | | | $2.67$ | $-2.55$ |
| Mean | | | | $-10.28$ | | | $1.82$ | $-8.46$ |

doubly occupied cavities. The IFST and FEP results are reported in Table 3.

The agreement between IFST and FEP is extremely good, with an $R^2$ coefficient of determination of 0.995 and a mean unsigned difference (MUD) of 0.45 kcal/mol. The accuracy of the estimates for $I_{ww}$ are supported by the close agreement of $\Delta G_{IFST}$ and $\Delta G_{bind}$. In addition to analyzing the protein conformation from the crystal structure, we considered the effect of the protein conformation on the predictions of IFST in the case of T4 Lysozyme. This was achieved by analyzing 10 different protein conformations generated from a simulation with harmonically restrained heavy atoms. The results are presented in Table 4.

Comparing the results from the fixed crystal structure with the results from the harmonically restrained structure, the $-T\Delta S_{IFST}$ term is reduced from an average of 1.87 kcal/mol to an average of 0.90 kcal/mol, respectively. This is in agreement with previous work showing that the $-T\Delta S_{IFST}$ term is smaller in magnitude when using data from a harmonically restrained protein simulation and is the expected result, because of a blurring of the probability densities when the protein is mobile (18). The $\Delta E_{IFST}$ and $\Delta G_{IFST}$ terms also differ to a small extent, with the most notable difference being that the binding free energy is more favorable by 2.15 kcal/mol in the case of water 904. Comparing the IFST results from the fixed crystal structure with the results from the ensemble of 10 structures, the differences are lower than 1.5 kcal/mol in all cases, and the qualitative picture remains the same. However, it is worth noting that the standard deviations from the 10 simulations are relatively large, with a maximum of 1.40 kcal/mol for $\Delta E_{IFST}$ in the case of water 904.

## CONCLUSIONS

In this work, we have predicted the free energy of transferring water molecules from the bulk into a buried protein cavity using the methods of IFST and FEP. This can be viewed either as the binding free energy of the water molecule ($\Delta G_{bind}$) or the contribution of the water molecule to the hydration free energy of the protein ($\Delta G_{IFST}$). The free energy contributions are all strongly favorable and dominated by a favorable enthalpy component. The entropy contributions to the free energy are all unfavorable, but relatively small in magnitude. The water molecules with the strongest binding affinities tend to be in hydrophilic cavities making one (β-Lactamase W2073) or two (β-Lactamase W2327) hydrogen bonds with charged amino acids. Conversely, the water molecules with the weakest binding affinities tend to be in hydrophobic cavities (T4 Lysozyme W904) with very little potential for hydrogen bonding, as expected. The agreement between IFST and FEP for the 19 protein cavities in the five test systems is extremely good, with an $R^2$ coefficient of determination of 0.995 and a MUD of 0.45 kcal/mol. To our knowledge, this study also represents the first calculation of the total two-particle entropy term ($\Delta S_{IFST}$) using information theory. The excellent agreement between IFST and FEP indicates that the mutual information terms between the pairs of water molecules ($I_{ww}$) have been calculated correctly and that this term is negligible for water molecules in protein cavities. This is an important result because it demonstrates that the majority of the correlation is captured by the solute-water entropy ($S_{sw}$) term. This means that the total change in water-water entropy ($\Delta S_{ww}$) is approximately equal to $-S_{bulk}$ (rather than zero) and makes a significantly favorable contribution of $-4.62$ kcal/mol to the binding free energy for the

**TABLE 3** The results of the FEP and IFST calculations for the 19 cavities

| System | PDB water IDs | $\Delta G_{bind}$ (kcal/mol) | $\Delta G_{IFST}$ (kcal/mol) | Signed difference (kcal/mol) | Unsigned difference (kcal/mol) |
|---|---|---|---|---|---|
| IL-1β | 202 and 204 | −11.77 | −11.27 | −0.49 | 0.49 |
| IL-1β | 203 and 207 | −6.18 | −5.76 | −0.42 | 0.42 |
| IL-1β | 200 | −7.09 | −7.57 | 0.49 | 0.49 |
| IL-1β | 209 | −6.90 | −6.87 | −0.03 | 0.03 |
| T4 Lysozyme | 902 and 905 | −20.41 | −19.52 | −0.89 | 0.89 |
| T4 Lysozyme | 904 | −3.33 | −3.00 | −0.33 | 0.33 |
| T4 Lysozyme | 920 | −8.29 | −7.76 | −0.53 | 0.53 |
| FKBP-2 | 207 and 208 | −16.70 | −16.13 | −0.57 | 0.57 |
| FKBP-2 | 203 | −13.07 | −12.50 | −0.57 | 0.57 |
| CA-II | 2004 | −8.21 | −7.62 | −0.59 | 0.59 |
| CA-II | 2015 | −14.05 | −13.70 | −0.36 | 0.36 |
| CA-II | 2031 | −10.06 | −9.44 | −0.62 | 0.62 |
| CA-II | 2042 | −11.09 | −10.58 | −0.51 | 0.51 |
| CA-II | 2055 | −4.29 | −4.17 | −0.12 | 0.12 |
| β-Lactamase | 2023 | −2.31 | −2.55 | 0.24 | 0.24 |
| β-Lactamase | 2048 | −14.30 | −13.87 | −0.44 | 0.44 |
| β-Lactamase | 2073 | −13.98 | −14.57 | 0.59 | 0.59 |
| β-Lactamase | 2105 | −12.06 | −11.77 | −0.29 | 0.29 |
| β-Lactamase | 2327 | −16.53 | −16.03 | −0.51 | 0.51 |
| Mean | | | | −0.31 | 0.45 |

**TABLE 4   The results of the IFST calculations for 10 conformations of T4 Lysozyme, alongside the results for the first 10 ns of the fixed crystal structure simulation, and the results for the 10 ns harmonically restrained simulation**

| PDB water ID | Fixed crystal structure | | | Ten simulation structures | | | Harmonically restrained | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\Delta E_{IFST}$ (kcal/mol) | $-T\Delta S_{IFST}$ (kcal/mol) | $\Delta G_{IFST}$ (kcal/mol) | $\Delta E_{IFST}$ (kcal/mol) | $-T\Delta S_{IFST}$ (kcal/mol) | $\Delta G_{IFST}$ (kcal/mol) | $\Delta E_{IFST}$ (kcal/mol) | $-T\Delta S_{IFST}$ (kcal/mol) | $\Delta G_{IFST}$ (kcal/mol) |
| 902 | −13.50 | 2.16 | −11.34 | −13.08 ± 0.83 | 2.32 ± 0.32 | −10.76 ± 0.84 | −13.30 | 1.30 | −12.00 |
| 905 | −10.06 | 1.90 | −8.16 | −9.11 ± 0.39 | 2.12 ± 0.17 | −6.99 ± 0.45 | −8.80 | 1.38 | −7.41 |
| 904 | −4.65 | 1.66 | −2.99 | −5.62 ± 1.40 | 1.21 ± 0.25 | −4.41 ± 1.26 | −5.35 | 0.21 | −5.14 |
| 920 | −9.54 | 1.78 | −7.76 | −8.31 ± 0.96 | 1.53 ± 0.29 | −6.79 ± 0.84 | −8.52 | 0.70 | −7.82 |

For the 10 conformations, the means and standard deviations are reported.

TIP4P-2005 water model. This result is expected to hold for highly ordered water molecules in protein active sites. It is important to note that small values of $I_{ww}$ in no way suggest an absence of correlation between the pair of water molecules in a doubly-occupied protein cavity. Solute-water and water-water correlations are explicitly wrapped up in the total two-particle entropy change ($\Delta S_{IFST}$), and the $I_{ww}$ term serves to capture additional correlations not quantified by the $S_{sw}$ term. One could equally well proceed by considering the water-water correlations first, but the choice of a solute reference frame significantly simplifies the calculations. In this study, we have considered cavities containing up to two water molecules. Studies on cavities with three or more water molecules would allow the accuracy of the two-particle approximation to be assessed. This is an important issue that should be addressed in future work. At the same time, it will also be important to estimate the $S_{vol}$ term explicitly, as it may be significant and vary between buried and surface-exposed hydration sites.

The average prediction of $-T\Delta S_{IFST}$ is +1.82 kcal/mol and this is consistent with the work of Dunitz, who noted that water molecules in inorganic salts differ in entropy from bulk solvent by an amount corresponding to a free energy difference of approximately +2.0 kcal/mol. The largest value of $-T\Delta S_{IFST}$ in this study is found for a water molecule near two charged residues and is +2.67 kcal/mol. In comparison, the greatest difference in Dunitz's analysis is +3.5 kcal/mol in the case of zinc sulfate monohydrate (10). It is important to note that the majority of entropy estimates in this work are for a fixed protein structure. The use of a fixed protein allows direct comparison of the IFST and FEP predictions and thus validation of the IFST calculations. It is clear that using a harmonically restrained or fully mobile protein structure within the current framework of IFST will lead to misestimation of the entropies because of a blurring of the probability densities. However, the results from calculations on 10 different fixed conformations of T4 Lysozyme suggest that IFST results should be robust in cases where the solute does not show a significant deviation in size, shape, or electrostatics. This is true for the cavities within these compact protein structures and should extend to many cases of interest in biology. Despite this, it is clear that further work is needed to develop probability-

based statistical mechanical methods such as IFST to make accurate predictions.

The timescales for the IFST and FEP simulations are similar, though the timescales for FEP depend on the number of lambda windows. This makes FEP a technique that requires performing benchmark simulations for each particular case or user input in configuring the calculations. Conversely, the convergence of IFST calculations can be automatically monitored throughout a simulation until the required accuracy is reached. Importantly, a single IFST simulation is informative about every hydration site in a system whereas FEP requires a separate simulation for each water molecule. In addition, IFST can be used to study individual water molecules within a network whereas FEP is best suited to studying single water molecules or groups of water molecules as a whole. This is because annihilating a water molecule from a network requires artificially creating an unphysical vacuum in a hydration site. For this reason, IFST is unique in yielding spatially resolved prediction of water thermodynamics and this is extremely useful in identifying ligand-binding hotspots at protein surfaces (48) and guiding the design of high-affinity small-molecule inhibitors (47). The IFST calculation of the hydration entropy is defined in the context of a fixed solute. The use of a mobile protein has been shown to reduce $\Delta S_{IFST}$, and this can be attributed to a blurring of the probability densities. The effect of the protein conformation on the results of IFST has been investigated by performing simulations for 10 different fixed conformations of T4 Lysozyme. The results are very similar to those for the crystal structure conformation, with no difference greater than 1.5 kcal/mol. This suggests that IFST predictions will be robust if the protein confirmation does not deviate significantly from the conformation observed in the crystal structure. However, the standard deviations are significant and thus the choice of protein confirmation will affect the results if a single-protein confirmation is used. The method of combining results from multiple IFST calculations on an ensemble of protein confirmations allows one to account for molecular flexibility and estimate the coupling of solute and solvent degrees of freedom. Clearly, FEP has the major advantage of considering molecular flexibility and, alongside thermodynamic integration, remains the tool of choice for estimation of absolute and

relative binding free energies. However, when studying water molecules, IFST has a number of unique advantages and these make it a very useful tool for understanding the role of hydration in the structure and function of biological systems.

In summary, to our knowledge, we have developed a new approach combining KNN and MI to calculate the total two-particle entropy in the context of IFST, and we have shown that the resulting predictions for the contribution of water molecules in protein cavities to the hydration free energy agree extremely well with equivalent predictions using FEP. The predicted entropy contributions to the free energy are in the range of +0.46 to +2.67 kcal/mol and this is in excellent agreement with historical estimates. In the future, it will be interesting to apply the entropy estimates developed in this work to extended water networks at protein surfaces and within protein binding sites. In these cases, the coupling of solute and solvent degrees of freedom is expected to be more significant and will need to be treated appropriately.

## ACKNOWLEDGMENTS

## REFERENCES

1. Liu, L., M. L. Quillin, and B. W. Matthews. 2008. Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc. Natl. Acad. Sci. USA.* 105:14406–14411.

2. Quillin, M. L., P. T. Wingfield, and B. W. Matthews. 2006. Determination of solvent content in cavities in IL-1beta using experimentally phased electron density. *Proc. Natl. Acad. Sci. USA.* 103:19749–19753.

3. Hubbard, S. J., K. H. Gross, and P. Argos. 1994. Intramolecular cavities in globular proteins. *Protein Eng.* 7:613–626.

4. Williams, M. A., J. M. Goodfellow, and J. M. Thornton. 1994. Buried waters and internal cavities in monomeric proteins. *Protein Sci.* 3:1224–1235.

5. Li, Z., and T. Lazaridis. 2003. Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.* 125:6636–6637.

6. Olano, L. R., and S. W. Rick. 2004. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.* 126:7991–8000.

7. Roux, B., M. Nina, …, J. C. Smith. 1996. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophys. J.* 71:670–681.

8. Yu, H., and S. W. Rick. 2010. Free energy, entropy, and enthalpy of a water molecule in various protein environments. *J. Phys. Chem. B.* 114:11552–11560.

9. Hamelberg, D., and J. A. McCammon. 2004. 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.

10. Dunitz, J. D. 1994. The entropic cost of bound water in crystals and biomolecules. *Science.* 264:670.

11. Lazaridis, T. 1998. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J. Phys. Chem. B.* 102:3531–3541.

12. Lazaridis, T. 1998. Inhomogeneous fluid approach to solvation thermodynamics. 2. Applications to simple fluids. *J. Phys. Chem. B.* 102:3542–3550.

13. Lazaridis, T. 2000. Solvent reorganization energy and entropy in hydrophobic hydration. *J. Phys. Chem. B.* 104:4964–4979.

14. Li, Z., and T. Lazaridis. 2006. Thermodynamics of buried water clusters at a protein-ligand binding interface. *J. Phys. Chem. B.* 110:1464–1475.

15. Robinson, D. D., W. Sherman, and R. Farid. 2010. Understanding kinase selectivity through energetic analysis of binding site waters. *Chem. Med. Chem.* 5:618–627.

16. Beuming, T., Y. Che, …, W. Sherman. 2012. Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins Struct. Funct. Bioinf.* 80:871–883.

17. Snyder, P. W., J. Mecinović, …, G. M. Whitesides. 2011. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. USA.* 108:17889–17894.

18. Raman, E. P., and A. D. MacKerell, Jr. 2013. Rapid estimation of hydration thermodynamics of macromolecular regions. *J. Chem. Phys.* 139:055105.

19. Singh, H., N. Misra, …, E. Demchuk. 2003. Nearest neighbor estimates of entropy. *Am. J. Math. Manag. Sci.* 23:301–321.

20. Hensen, U., H. Grubmüller, and O. F. Lange. 2009. Adaptive anisotropic kernels for nonparametric estimation of absolute configurational entropies in high-dimensional configuration spaces. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 80:011913.

21. Killian, B. J., J. Yundenfreund Kravitz, and M. K. Gilson. 2007. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* 127:024107.

22. Matsuda, H. 2000. Physical nature of higher-order mutual information: intrinsic correlations and frustration. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 62 (3 Pt A):3096–3102.

23. Hnizdo, V., J. Tan, …, M. K. Gilson. 2008. Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J. Comput. Chem.* 29:1605–1614.

24. Hnizdo, V., E. Darian, …, H. Singh. 2007. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *J. Comput. Chem.* 28:655–668.

25. Avvaru, B. S., S. A. Busby, …, R. McKenna. 2009. Apo-human carbonic anhydrase II revisited: implications of the loss of a metal in protein structure, stability, and solvent network. *Biochemistry.* 48:7365–7372.

26. Chen, Y., R. Bonnet, and B. K. Shoichet. 2007. The acylation mechanism of CTX-M β-lactamase at 0.88 a resolution. *J. Am. Chem. Soc.* 129:5378–5380.

27. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

28. Sastry, G. M., M. Adzhigirey, …, W. Sherman. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 27:221–234.

29. Brooks, B. R., C. L. Brooks, 3rd, …, M. Karplus. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.

30. MacKerell, A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

31. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.

32. Abascal, J. L. F., and C. Vega. 2005. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* 123:234505.

33. Feller, S. E., Y. H. Zhang, …, B. R. Brooks. 1995. Constant-pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* 103:4613–4621.

34. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

35. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.

36. Zwanzig, R. W. 1954. High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.* 22:1420–1426.

37. Bennett, C. H. 1976. Efficient estimation of free energy differences from Monte-Carlo data. *J. Comput. Phys.* 22:245–268.

38. Pohorille, A., C. Jarzynski, and C. Chipot. 2010. Good practices in free-energy calculations. *J. Phys. Chem. B.* 114:10235–10253.

39. Liu, P., F. Dehez, …, C. Chipot. 2012. A toolkit for the analysis of free-energy perturbation calculations. *J. Chem. Theory Comput.* 8:2606–2616.

40. Beutler, T. C., A. E. Mark, …, W. F. van Gunsteren. 1994. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* 222:529–539.

41. Zacharias, M., T. P. Straatsma, and J. A. McCammon. 1994. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* 100:9025–9031.

42. Gilson, M. K., J. A. Given, …, J. A. McCammon. 1997. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 72:1047–1069.

43. Boresch, S., F. Tettinger, …, M. Karplus. 2003. Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem. B.* 107:9535–9551.

44. Gilson, M. K., and K. K. Irikura. 2010. Symmetry numbers for rigid, flexible, and fluxional molecules: theory and applications. *J. Phys. Chem. B.* 114:16304–16317.

45. Huggins, D. J., and M. C. Payne. 2013. Assessing the accuracy of inhomogeneous fluid solvation theory in predicting hydration free energies of simple solutes. *J. Phys. Chem. B.* 117:8232–8244.

46. Nguyen, C. N., T. K. Young, and M. K. Gilson. 2012. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* 137:044101.

47. Young, T., R. Abel, …, R. A. Friesner. 2007. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. USA.* 104:808–813.

48. Huggins, D. J., M. Marsh, and M. C. Payne. 2011. Thermodynamic properties of water molecules at a protein-protein interaction surface. *J. Chem. Theory Comput.* 7:3514–3522.

49. Kirkwood, J. G., and F. P. Buff. 1951. The statistical mechanical theory of solutions. I. *J. Chem. Phys.* 19:774–777.

50. Huggins, D. J. 2014. Estimating translational and orientational entropies using the *k*-nearest neighbors algorithm. *J. Chem. Theory Comput.* 10:3617–3625.

51. Huggins, D. J. 2014. Comparing distance metrics for rotation using the *k*-nearest neighbors algorithm for entropy estimation. *J. Comput. Chem.* 35:377–385.

52. Kraskov, A., H. Stögbauer, and P. Grassberger. 2004. Estimating mutual information. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69:066138.

53. Hensen, U., O. F. Lange, and H. Grubmüller. 2010. Estimating absolute configurational entropies of macromolecules: the minimally coupled subspace approach. *PLoS ONE.* 5:e9179.