

# Spatial Decomposition of Translational Water–Water Correlation Entropy in Binding Pockets

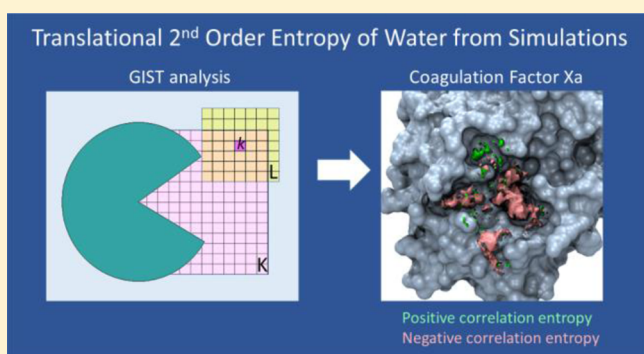
Crystal N. Nguyen,<sup>†</sup> Tom Kurtzman,<sup>‡,§</sup> and Michael K. Gilson<sup>\*,†</sup>

<sup>†</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0736, United States

<sup>‡</sup>Department of Chemistry, Lehman College, The City University of New York, 250 Bedford Park Blvd. West, Bronx, New York, New York 10468, United States

<sup>§</sup>Ph.D. Program in Chemistry, The Graduate Center of The City University of New York, New York 10016, United States

**ABSTRACT:** A number of computational tools available today compute the thermodynamic properties of water at surfaces and in binding pockets by using inhomogeneous solvation theory (IST) to analyze explicit-solvent simulations. Such methods enable qualitative spatial mappings of both energy and entropy around a solute of interest and can also be applied quantitatively. However, the entropy estimates of existing methods have, to date, been almost entirely limited to the first-order terms in the IST's entropy expansion. These first-order terms account for localization and orientation of water molecules in the field of the solute but not for the modification of water–water correlations by the solute. Here, we present an extension of the Grid Inhomogeneous Solvation Theory (GIST) approach which accounts for water–water translational correlations. The method involves rewriting the two-point density of water in terms of a conditional density and utilizes the efficient nearest-neighbor entropy estimation approach. Spatial maps of this second order term, for water in and around the synthetic host cucurbit[7]uril and in the binding pocket of the enzyme Factor Xa, reveal mainly negative contributions, indicating solute-induced water–water correlations relative to bulk water; particularly strong signals are obtained for sites at the entrances of cavities or pockets. This second-order term thus enters with the same, negative, sign as the first order translational and orientational terms. Numerical and convergence properties of the methodology are examined.



## 1. INTRODUCTION

Molecular recognition in water cannot be understood purely in terms of the direct interactions of the two molecules that bind each other. Water itself plays an intimate role, such as by effectively weakening charge–charge interactions, by competing with the solutes for hydrogen-bonding opportunities, and by providing hydrophobic interactions that drive the association of nonpolar parts of the solutes. Much is known about the physical principles by which water influences binding,<sup>1–12</sup> and the effects of water can, to a large extent, be captured in both implicit<sup>13–21</sup> and explicit<sup>22–29</sup> representations of water. However, there are also still puzzles and gaps in our understanding of water's role in molecular recognition. For example, although the hydrophobic effect may be reasonably well understood for simple nonpolar surfaces, many solutes, particularly biomolecules like proteins and DNA, present surfaces with complicated shapes and complex patterns of polarity, where water might have properties quite different from water at simpler, better understood surfaces. Thus, small surface clefts can lead to the formation of water clusters with a reduced set of hydrogen-bond arrangements, making such waters particularly easy to displace into bulk, where they regain many more hydrogen bonding opportunities.<sup>30–34</sup>

Classical notions of the hydrophobic effect<sup>35–37</sup> do not necessarily envision such distinctive cases.

In recent years, new computational tools have come online to compute and visualize the structure and thermodynamics of water at complex surfaces, by applying inhomogeneous solvation theory (IST)<sup>38–44</sup> to simulations of solutes immersed in explicit water. Some of these tools, such as WaterMap,<sup>30,45</sup> STOW,<sup>46</sup> and others<sup>47–49</sup> gain simplicity by modeling the density distribution of water in terms of discrete hydration sites. Alternatively, Grid Inhomogeneous Solvation Theory (GIST)<sup>34,50–52</sup> discretizes water density and thermodynamic contributions onto a three-dimensional grid, in order to gain a more comprehensive and quantitative description of solute hydration. Such tools offer insights into the properties and role of surface water and can also be useful in computer-aided drug design, by pointing out regions in a protein binding pocket where a ligand may gain affinity by displacing thermodynamically unfavorable water.<sup>30,34,45,53</sup>

Implementations of IST have so far concentrated almost entirely on calculating first-order contributions to the entropy of water.

Received: October 1, 2015

Published: November 11, 2015

The first-order translational entropy reflects, in effect, the bumpiness of the density distribution of water induced by the solute, while the first-order orientational entropy reflects the degree to which the solute reduces the orientational freedom of nearby water. However, it is likely that higher-order terms also contribute significantly to the hydration entropy. Higher-order terms relate to water–water correlations, which certainly exist in bulk water<sup>54,55</sup> and are expected to be modified by the presence of a solute. Note that greater correlation implies lower entropy and that the higher-order entropy terms in the IST series expansion are akin to the mutual information terms in the mutual information expansion<sup>56,57</sup> which has been used to estimate changes in the configurational entropy of small molecules and proteins,<sup>58–61</sup> and have been tied to the multibody expansion of IST.<sup>62</sup> Indeed, Velez-Vega et al. have recently obtained encouraging results for pure water with a novel method that treats water on an atomistic, rather than molecular, basis and is based on the mutual information expansion.<sup>63</sup>

Several prior studies have addressed the contributions of water–water correlations to hydration entropy. An early work used IST to study the contributions of water–water correlations to the hydration entropy of methane, modeled as a single Lennard-Jones atom, for which numerical convergence is facilitated by exploitation of spherical symmetry.<sup>64</sup> Interestingly, the signs of the water–water translational and orientational terms were found to be negative (increased correlation relative to bulk water) at room temperature but found to become positive at high temperature, whereas the first order terms remained negative at all temperatures considered. This study also tested an approximation in which the water–water correlation function near the solute was approximated by the bulk water–water correlation function; however, the resulting entropies were found to deviate substantially from those computed without the approximation. Nonetheless, small molecule hydration free energies computed with this approximation were shown to agree reasonably well with corresponding values obtained by a reference free energy perturbation method,<sup>65</sup> and the same approximation was used to estimate water–water correlations within clusters of water molecules buried at a protein–ligand binding interface.<sup>66</sup> Two other studies have taken an alternative approach to obtaining these numerically challenging terms for water near small molecules<sup>67</sup> and protein binding sites:<sup>52</sup> instead of computing the water–water terms from actual correlation data, they were instead heuristically set as proportional to the first order entropy, where the constant of proportionality is on the order of  $-0.3$  to  $-0.5$ . Finally, a recent study used the water–water correlations obtained from simulations of the system of interest, rather than of bulk water, to compute the contributions of these correlations to the hydration entropy contributions of pairs of waters buried in protein cavities.<sup>62</sup> The water–water correlation contribution was found to be negligible, however, apparently because the waters were tightly bound to the protein and hence had minimal fluctuations available to correlate with each other. Thus, considerable work has already been done to study or estimate the contributions of water–water correlations to the entropy of hydration. However, correlations computed from simulations have not yet been used to quantify or visualize second order terms for water at well-hydrated, complex surfaces, such as the binding sites of host–guest systems and proteins.

The present paper addresses this challenge with a three-dimensional grid formulation of the second-order translational entropy of water, which is related to the two-point spatial correlation function of water density. Rewriting the second order

translational entropy in terms of a conditional density, instead of a two-point density, leads to an intuitively interpretable form of this term and also facilitates spatial mapping and visualization. The present method also makes use of the efficient nearest-neighbor entropy-estimation approach.<sup>50,67,68</sup> We present an evaluation of the numerical properties of the method and analyses of the second-order translational entropy of water in and around the synthetic host molecule cucurbit[7]uril and the blood clotting cascade enzyme Factor Xa (FXa). These second order results are placed in the context of the first-order translational and orientational entropy around these solutes, and visualization is used to understand the spatial distribution of the second order term, including molecular features that generate particularly high water–water correlations and hence particularly low second-order translational entropies.

## 2. METHODS

**2.1. Theory.** Here, we express the second order (or pairwise) translational entropy,  $S_{\text{ww}}$  in terms of conditional distribution functions, then show how the resulting integrals can be estimated from simulation data by using a three-dimensional grid to discretize space and then estimating water densities with either a histogram or a nearest neighbor method. We also review how the first order (or single-body) translational entropy,  $S_{\text{sw}}$  can be estimated with the nearest neighbor method,<sup>67</sup> as this offers numerical advantages over the histogram approach used previously.<sup>50</sup>

**2.1.1. Pairwise Translational Entropy in the Inhomogeneous Solvation Theory: Background.** The solvation entropy of a flexible solute may be written as

$$\begin{aligned} \Delta S_{\text{solvation}} = & \int \rho_{\text{sol}}(q) \Delta S_{\text{sol}}(q) dq \\ & - k_{\text{B}} \left\{ \int \rho_{\text{sol}}(q) \ln[\rho_{\text{sol}}(q)] dq \right. \\ & \left. - \int \rho_{\text{vac}}(q) \ln[\rho_{\text{vac}}(q)] dq \right\} \end{aligned} \quad (1)$$

where  $\rho_{\text{sol}}(q)$  is the equilibrium probability distribution function of the solute over its internal coordinates  $q$  when it is in solution and  $\rho_{\text{vac}}(q)$  is the corresponding function for the solute *in vacuo*,  $\Delta S_{\text{sol}}(q)$  is the solvation entropy the solute would have if it were constrained in conformation  $q$ ,<sup>69</sup> and  $k_{\text{B}}$  is Boltzmann's constant. The first term in eq 1 is the solvation entropy averaged over the Boltzmann ensemble of solute configurations, and the second term is the change in solute configurational entropy on being transferred from a vacuum to solvent. Here, as in prior applications of IST, we consider the solvation of a solute in a given conformation,  $q$ , or in a narrow range of conformations, although one could, at some computational cost, explore solvation over a range of conformations by applying IST to each one separately and appropriately weighting each conformation. For simplicity, we will refer to  $\Delta S_{\text{sol}}(q)$  as the solvation entropy and write  $\Delta S_{\text{sol}}$  instead of  $\Delta S_{\text{sol}}(q)$ .

For solvation in water, inhomogeneous solvation theory (IST) provides an expansion of  $\Delta S_{\text{sol}}$  in terms of one-water, two-water, and higher order terms.<sup>38,39,41–43</sup> The one-water term,  $S_{\text{sw}}$ , accounts for the patterning of water density around the solute, while the two-water term,  $\Delta S_{\text{ww}}$ , accounts for the difference between the two-point, water–water correlations around the solute and those in bulk; subsequent terms capture still higher-order correlations. Thus, neglecting terms above pairwise leads to the following approximation:

$$\Delta S_{\text{sol}} \approx S_{\text{sw}} + \Delta S_{\text{ww}} \quad (2)$$

The one-water term has been considered in some detail in prior studies (see [Introduction](#)) and is considered here only in relation to its calculation via a nearest-neighbor (NN) method ([section 2.5](#)). The pairwise term, which is of primary interest here, may be broken down further into orientational and translational parts:<sup>46,66</sup>

$$\Delta S_{\text{ww}} = \Delta S_{\text{ww}}^{\text{trans}} + \Delta S_{\text{ww}}^{\text{orient}} \quad (3)$$

The present study focuses exclusively on the translational part,  $\Delta S_{\text{ww}}^{\text{trans}}$ , which accounts for the solute-induced change in the two-point correlation of water density, without reference to the spatial orientation of the waters. Using a form provided by Morita and Hiroike<sup>42</sup> and clearly rendered in eq 12 of a subsequent paper by Lazaridis,<sup>43</sup> one may write this quantity as

$$\begin{aligned} \Delta S_{\text{ww}}^{\text{trans}} &= S_{\text{ww}}^{\text{trans}} - S_{\text{ww,bulk}}^{\text{trans}} \\ S_{\text{ww}}^{\text{trans}} &= \frac{-k_{\text{B}} \rho^2}{2} \int \int g(\mathbf{r}) g(\mathbf{r}') [g(\mathbf{r}, \mathbf{r}') \ln g(\mathbf{r}, \mathbf{r}') \\ &\quad - g(\mathbf{r}, \mathbf{r}') + 1] \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ S_{\text{ww,bulk}}^{\text{trans}} &= \frac{-k_{\text{B}} \rho^2}{2} \int \int g^{\circ}(\mathbf{r}) g^{\circ}(\mathbf{r}') [g^{\circ}(\mathbf{r}, \mathbf{r}') \ln g^{\circ}(\mathbf{r}, \mathbf{r}') \\ &\quad - g^{\circ}(\mathbf{r}, \mathbf{r}') + 1] \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \end{aligned} \quad (4)$$

The first equation of [eq 4](#) writes  $\Delta S_{\text{ww}}^{\text{trans}}$  as the difference between the pairwise translational entropy of the solute-perturbed water from the pairwise translational entropy of unperturbed (bulk) water. The second and third equations provide expressions for the solute-perturbed system and bulk water. Here,  $\rho$  is the number density of bulk water,  $g(\mathbf{r})$  is the unitless ratio of the number density of water at  $\mathbf{r}$  in the presence of the solute to the bulk density,  $g(\mathbf{r}) \equiv \frac{\rho(\mathbf{r})}{\rho}$ , and  $g(\mathbf{r}, \mathbf{r}')$  is the two-point correlation

function in the presence of the solute,  $g(\mathbf{r}, \mathbf{r}') \equiv \frac{\rho(\mathbf{r}, \mathbf{r}')}{\rho(\mathbf{r}) \rho(\mathbf{r}')}$ , where  $\rho(\mathbf{r}, \mathbf{r}')$  is the two-point number density of water at locations  $\mathbf{r}$  and  $\mathbf{r}'$ . The solute is considered to be immobilized in the lab frame of reference, so that coordinates  $\mathbf{r}$  and  $\mathbf{r}'$  represent both lab-frame coordinates and solute-frame coordinates. The integrals range over the whole system or, at any rate, to points  $\mathbf{r}$  and  $\mathbf{r}'$  far enough from the solute to effectively capture the entire effect of the solute on the solvent distribution.

The expression in line three for the pairwise translational entropy of pure water uses the corresponding distribution functions,  $g^{\circ}(\mathbf{r})$ ,  $g^{\circ}(\mathbf{r}')$ , and  $g^{\circ}(\mathbf{r}, \mathbf{r}')$  for unperturbed bulk water. Note, however, that in the absence of the perturbing solute (or, in the terms of Morita and Hiroike, in the absence of an external field),  $g^{\circ}(\mathbf{r}) = g^{\circ}(\mathbf{r}') = 1$ . We include these terms here to make clear that derivations and algorithms which apply in the presence of a solute fixed in the lab frame are equally applicable to pure water, where no solute is present.

**2.1.2. Conditional Formulation of the Pairwise Translational Entropy.** We now develop an expression for  $S_{\text{ww}}^{\text{trans}}$  in terms of conditional densities. As noted in the prior subsection, the same derivation carries through in the case of pure water, so the pure water case is not included explicitly. The resulting formulation enables an informative spatial decomposition and provides a convenient basis for evaluating this correlation term with nearest-neighbor<sup>51,68</sup> or histogram methods. We first use the definitions in the prior paragraph to write the starting expression in terms of number densities:

$$\begin{aligned} S_{\text{ww}}^{\text{trans}} &= -\frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}, \mathbf{r}') \ln \rho(\mathbf{r}, \mathbf{r}') \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ &\quad + \frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}, \mathbf{r}') \ln [\rho(\mathbf{r}) \rho(\mathbf{r}')] \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ &\quad - \frac{k_{\text{B}}}{2} \int \int [-\rho(\mathbf{r}, \mathbf{r}') + \rho(\mathbf{r}) \rho(\mathbf{r}')] \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \end{aligned} \quad (5)$$

Because, in the canonical ensemble,  $\int \rho(\mathbf{r}) \, \mathbf{d}\mathbf{r} = N$  and  $\int \int \rho(\mathbf{r}, \mathbf{r}') \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' = N(N-1)$ , where  $N$  is the number of water molecules, the second line in [eq 5](#), which we call the nonlogarithmic term, reduces to  $-\frac{Nk_{\text{B}}}{2}$ . For the sake of brevity, this simple term will be omitted from subsequent expressions, except as otherwise noted. We now use the fact that the two-point density may be written in terms of either of the conditional two-point densities,  $\rho(\mathbf{r}, \mathbf{r}') = \rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}') = \rho(\mathbf{r}'|\mathbf{r}) \rho(\mathbf{r})$ , to rewrite  $S_{\text{ww}}^{\text{trans}}$ , without the nonlogarithmic term, as

$$\begin{aligned} S_{\text{ww}}^{\text{trans}} &= -\frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}') \ln [\rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}')] \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ &\quad + \frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}') \ln [\rho(\mathbf{r}) \rho(\mathbf{r}')] \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ &= -\frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}') \ln \rho(\mathbf{r}|\mathbf{r}') \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ &\quad + \frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}') \ln \rho(\mathbf{r}) \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ &= -\frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}') \ln \rho(\mathbf{r}|\mathbf{r}') \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \\ &\quad + \frac{k_{\text{B}}}{2} \int \int \rho(\mathbf{r}'|\mathbf{r}) \rho(\mathbf{r}) \ln \rho(\mathbf{r}) \, \mathbf{d}\mathbf{r} \, \mathbf{d}\mathbf{r}' \end{aligned} \quad (6)$$

Intuitively, the first term in the last line of [eq 6](#) reports how the presence of a water at  $\mathbf{r}'$  influences the entropy of the water in the system as a whole, while the second term essentially subtracts out the entropy of the water for the distribution not conditioned on the presence of water at  $\mathbf{r}'$ . In the absence of any correlation between the density of water at  $\mathbf{r}$  and at  $\mathbf{r}'$ , then  $\rho(\mathbf{r}|\mathbf{r}') = \rho(\mathbf{r})$  and  $\rho(\mathbf{r}'|\mathbf{r}) = \rho(\mathbf{r}')$ , so the integrands of both terms in the last line of [eq 6](#) become equal, and the volume elements at these locations will make zero contribution to the pairwise entropy.

The nearest neighbor entropy estimation method provides averages of the form  $\langle \ln \rho \rangle$ ,<sup>68</sup> so it is of interest to express [eq 6](#) in a form which uses such averages. We use the facts that  $\int \rho(\mathbf{r}) \, \mathbf{d}\mathbf{r} = N$  and  $\int \rho(\mathbf{r}|\mathbf{r}') \, \mathbf{d}\mathbf{r} = N-1$ , to recognize that  $\frac{1}{N} \int \rho(\mathbf{r}) \, \mathbf{d}\mathbf{r}$  and  $\frac{1}{N-1} \int \rho(\mathbf{r}|\mathbf{r}') \, \mathbf{d}\mathbf{r}$  are normalized probability density functions and hence can be cast as probability weighting factors of the corresponding logarithm terms in [eq 6](#). Multiplying and dividing the first and second terms in the last two lines of [eq 6](#) by the respective normalization constants  $(N-1)$  and  $N$  and integrating the second term over  $\mathbf{r}'$  to obtain a factor of  $N-1$  yields

$$\begin{aligned} S_{\text{ww}}^{\text{trans}} &= -\frac{k_{\text{B}}}{2} (N-1) \int \rho(\mathbf{r}') \int \left[ \frac{\rho(\mathbf{r}|\mathbf{r}')}{(N-1)} \ln \rho(\mathbf{r}|\mathbf{r}') \, \mathbf{d}\mathbf{r} \right] \, \mathbf{d}\mathbf{r}' \\ &\quad + \frac{k_{\text{B}}}{2} N(N-1) \int \frac{\rho(\mathbf{r})}{N} \ln \rho(\mathbf{r}) \, \mathbf{d}\mathbf{r} \\ &= -\frac{k_{\text{B}}}{2} (N-1) \int \rho(\mathbf{r}') \langle \ln \rho(\mathbf{r}|\mathbf{r}') \rangle \, \mathbf{d}\mathbf{r}' \\ &\quad + \frac{k_{\text{B}}}{2} N(N-1) \langle \ln \rho(\mathbf{r}) \rangle \end{aligned} \quad (7)$$



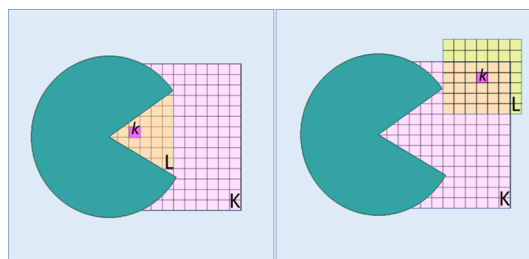
where angle brackets indicate averages over the Boltzmann distribution.

**2.1.3. Localized Perturbation Approximation.** The integrals in the equations above range over the entire system, but in practice we will often be most interested in the region of solvent near the solute. For one thing, such a region may have particular functional importance, such as binding of a ligand by the solute, if it is a receptor. Furthermore, the influence of the solute on the solvent distribution functions dies off with distance from the solute,<sup>41,43</sup> so distant regions contribute little, and ignoring distant regions can make calculations more tractable without introducing much error. Accordingly, it is useful to restrict the integrals over  $\mathbf{r}'$  to a local region of interest,  $K$ , such as a receptor binding site. Furthermore, the contribution of the water at  $\mathbf{r}'$  to  $S_{\text{ww}}^{\text{trans}}$  is directly connected with the conditional density function  $\rho(\mathbf{r}|\mathbf{r}')$ , which reports on the perturbation of water density at  $\mathbf{r}$  by the presence of a water at  $\mathbf{r}'$ , and this perturbation also dies off with the distance of  $\mathbf{r}$  from  $\mathbf{r}'$ . As a consequence, it is often reasonable to restrict the integrals over  $\mathbf{r}$  to finite regions centered on  $\mathbf{r}'$ , which will be termed  $L_{r'}$ . With these localizing approximations, eq 6 may be written in either of the following equivalent forms:

$$\begin{aligned} S_{\text{ww}}^{\text{trans}} &\approx S_{\text{ww,KL}}^{\text{trans}} \\ &= -\frac{k_B}{2} \int_K d\mathbf{r}' \rho(\mathbf{r}') \int_{L_{r'}} \rho(\mathbf{r}|\mathbf{r}') \ln \rho(\mathbf{r}|\mathbf{r}') d\mathbf{r} \\ &\quad + \frac{k_B}{2} \int_K d\mathbf{r}' \int_{L_{r'}} \rho(\mathbf{r}|\mathbf{r}') \rho(\mathbf{r}') \ln \rho(\mathbf{r}) d\mathbf{r} \\ &= -\frac{k_B}{2} \int_K d\mathbf{r}' \rho(\mathbf{r}') \int_{L_{r'}} \rho(\mathbf{r}|\mathbf{r}') \ln \rho(\mathbf{r}|\mathbf{r}') d\mathbf{r} \\ &\quad + \frac{k_B}{2} \int_K d\mathbf{r}' \int_{L_{r'}} \rho(\mathbf{r}'|\mathbf{r}) \rho(\mathbf{r}) \ln \rho(\mathbf{r}) d\mathbf{r} \end{aligned} \quad (8)$$

In physical terms, eq 8 provides the contribution to the total pairwise entropy from the degree to which the water density in regions  $L_{r'}$  is modified by conditioning on the presence of a water at each location  $\mathbf{r}'$  in  $K$ . Note that eq 6 is recovered when the  $K$  and  $L$  regions span the entire system.

**2.2. Using a 3D Grid to Estimate Distribution Functions from Simulations with Explicit Water.** Following the existing first-order GIST approach,<sup>51</sup> we use a three-dimensional grid to compute the spatial distribution of the second-order translational entropy from explicit-water simulations for a region of interest, such as a protein binding site. Thus, we discretize eq 8 by breaking space into cubic voxels of side-length  $R_{\text{vox}}$  and volume  $V_{\text{vox}} = R_{\text{vox}}^3$  which are assumed to be small enough that a voxel can contain at most one water molecule. (Here, the translational coordinates of a water molecule are identified with those of its oxygen atom.) As a matter of notation, location  $\mathbf{r}'$  maps to voxel index  $k$ , and the region  $K$  corresponds to a grid in the region of interest, such as a binding site (Figure 1), while  $\mathbf{r}$  maps to voxel index  $l$  and the regions  $L_{r'}$ , which are centered around  $\mathbf{r}'$ , map to regions  $L_k$ , which are centered around voxels  $k \in K$ ; see Figure 1. The following subsections describe a histogram method and a nearest-neighbor method of using this discretization to estimate  $S_{\text{ww,KL}}^{\text{trans}}$  through analysis of  $n_f$  frames, or snapshots, from a molecular dynamics (MD) simulation of explicit water molecules around the solute; an additional subsection explains how the nonlogarithmic term in eq 5 is computed.



**Figure 1.** Diagram of a receptor (green) in a water-filled simulation box (blue), with grid corresponding to region  $K$  in pale pink and two examples of regions  $L_k$  in yellow (orange where overlapping with  $K$ ). In each case, the  $L_k$  grid is centered in one  $K$ -grid voxel,  $k$ , which is highlighted in pink.

**Histogram Method.** The histogram method uses MD data to estimate the various densities in the first line of eq 8 by counting the instances of water molecules in voxels:

$$\begin{aligned} \rho(\mathbf{r}) &= \frac{n_l}{n_f V_{\text{vox}}} & \rho(\mathbf{r}') &= \frac{n_k}{n_f V_{\text{vox}}} \\ \rho(\mathbf{r}'|\mathbf{r}) &= \frac{n_{kl}}{n_l V_{\text{vox}}} & \rho(\mathbf{r}|\mathbf{r}') &= \frac{n_{lk}}{n_k V_{\text{vox}}} \end{aligned} \quad (9)$$

Here,  $n_l$  and  $n_k$  are the numbers of MD frames for which a water is found in voxels  $l$  and  $k$ , respectively, and  $n_{kl} = n_{lk}$  are the numbers of MD frames for which a water is found in both voxels  $k$  and  $l$ . Note that this approximation treats the water densities as constant over the volume of each voxel, and, as noted above, we have assumed that the voxels are small enough that at most one water molecule can occupy a voxel. In going from a double integral to a double sum over voxels, two factors of the volume element  $V_{\text{vox}}$  appear. Thus, the histogram method estimates  $S_{\text{ww,KL}}^{\text{trans}}$  as

$$\begin{aligned} S_{\text{ww,KL}}^{\text{trans}} &\approx S_{\text{ww,KL}}^{\text{trans,hist}} \equiv V_{\text{vox}} \sum_{k \in K} s_k^{\text{cond,hist}} - V_{\text{vox}} \sum_{k \in K} s_k^{\text{hist}} \\ s_k^{\text{cond,hist}} &\equiv \frac{n_k}{n_f V_{\text{vox}}} s_k^{\text{cond,norm,hist}} \\ s_k^{\text{hist}} &\equiv \frac{n_k}{n_f V_{\text{vox}}} s_k^{\text{norm,hist}} \\ s_k^{\text{cond,norm,hist}} &\equiv -\frac{k_B}{2} \sum_{l \in L} \left[ \frac{n_{lk}}{n_k} \ln \left( \frac{n_{lk}}{n_k V_{\text{vox}}} \right) \right] \\ s_k^{\text{norm,hist}} &\equiv -\frac{k_B}{2} \sum_{l \in L} \left[ \frac{n_{lk}}{n_k} \ln \left( \frac{n_l}{n_f V_{\text{vox}}} \right) \right] \end{aligned} \quad (10)$$

Here, the contribution of regions  $K$  and  $L$  to the pairwise translational part of the solvation entropy is estimated in terms of the quantities summed over voxels  $k \in K$ . The contribution of voxel  $k$ ,  $V_{\text{vox}}(s_k^{\text{cond,hist}} - s_k^{\text{hist}})$ , which is the product of the voxel volume,  $V_{\text{vox}}$ , and the difference between the conditional and unconditional entropy densities, is computed as the probability of finding a water in voxel  $k$ ,  $\frac{n_k}{n_f}$ , multiplied by the per water normalized quantity  $s_k^{\text{cond,norm,hist}} - s_k^{\text{norm,hist}}$ . Writing the pairwise translational entropy in terms of a sum over voxels, in this manner, allows a spatial decomposition of this term which is useful for generating graphical representations. In practice, the

expressions in eq 10 are evaluated by analyzing the  $n_f$  available MD frames to determine  $n_k$ , the number of frames with a water in voxel  $k$ ;  $n_l$ , the number of frames with a water in voxel  $l$ ; and  $n_{lk}$ , the number of frames with a water in both voxels  $l$  and  $k$ , where voxels  $k$  and  $l$  are required to reside in regions  $K$  and  $L$ , respectively.

**Nearest-Neighbor Method.** Starting with the final expression in eq 8, we multiply and divide the first term by  $\int_{L_{r'}} \rho(\mathbf{r}\mathbf{r}') \, \mathbf{dr} \approx \sum_{l \in L_{r'}} \frac{n_{lk}}{n_k} \equiv N_{L_{r'},lk}$ , which is the mean number of waters on the  $L_{r'}$  grid given a water in the  $k$  voxel, and we multiply and divide the second term by  $N_{L_{r'}} = \int_{L_{r'}} \rho(\mathbf{r}) \, \mathbf{dr}$ , which is the mean number of waters on the  $L_{r'}$  grid without any condition, in order to convert the corresponding number density functions into probability density functions, much as previously done in deriving eq 7. The present expression for  $N_{L_{r'},lk}$  is an approximation, due to the discretization of  $\mathbf{r}'$  into voxels, so the following expressions are, accordingly, also written as approximations. We obtain

$$\begin{aligned} S_{\text{ww},KL}^{\text{trans}} &= -\frac{k_B}{2} \int_K \mathbf{dr}' \rho(\mathbf{r}') \int_{L_{r'}} \rho(\mathbf{r}\mathbf{r}') \, \mathbf{dr} \\ &\quad \times \int_{L_{r'}} \frac{\rho(\mathbf{r}\mathbf{r}')}{\int_{L_{r'}} \rho(\mathbf{r}\mathbf{r}') \, \mathbf{dr}} \ln \rho(\mathbf{r}\mathbf{r}') \, \mathbf{dr} \\ &\quad + \frac{N_{L_{r'}} k_B}{2} \int_K \mathbf{dr}' \int_{L_{r'}} \frac{\rho(\mathbf{r})}{\int_{L_{r'}} \rho(\mathbf{r}) \, \mathbf{dr}} \rho(\mathbf{r}'\mathbf{r}) \ln \rho(\mathbf{r}) \, \mathbf{dr} \\ &= -\frac{k_B}{2} \int_K \rho(\mathbf{r}') \int_{L_{r'}} \rho(\mathbf{r}\mathbf{r}') \, \mathbf{dr} \langle \ln \rho(\mathbf{r}\mathbf{r}') \rangle_{L_{r'}} \, \mathbf{dr}' \\ &\quad + \frac{k_B}{2} \int_K \mathbf{dr}' N_{L_{r'}} \langle \rho(\mathbf{r}'\mathbf{r}) \ln \rho(\mathbf{r}) \rangle_{L_{r'}} \end{aligned} \quad (11)$$

Here,  $\langle \rangle_{L_{r'}}$  indicates an average over the volume of the  $L_{r'}$  grid. These integrals are now discretized via the approximations in eq 9:

$$\begin{aligned} S_{\text{ww},KL}^{\text{trans}} &\approx S_{\text{ww},KL}^{\text{trans,NN}} \\ &= -\frac{k_B}{2} \sum_{k \in K} \left[ \frac{n_k}{n_f} N_{L_k|k} \langle \ln \rho(\mathbf{r}k) \rangle_{L_k} \right] \\ &\quad + \frac{k_B}{2} \sum_{k \in K} \left[ N_{L_k} \left\langle \frac{n_{kll}}{n_l} \ln \rho(\mathbf{r}) \right\rangle_{L_k} \right] \end{aligned} \quad (12)$$

Here,  $\rho(\mathbf{r}k)$  represents the number density of water at  $\mathbf{r}$  conditioned on the presence of a water in voxel  $k$ , and  $L_k$  is the  $L$  region associated with, and typically centered on, voxel  $k$ . This may be put into a form analogous to the histogram formulation in eq 10, as follows:

$$\begin{aligned} S_{\text{ww},KL}^{\text{trans}} &\approx S_{\text{ww},KL}^{\text{trans,NN}} = V_{\text{vox}} \sum_{k \in K} s_k^{\text{cond,NN}} - V_{\text{vox}} \sum_{k \in K} s_k^{\text{NN}} \\ s_k^{\text{cond,NN}} &\equiv \frac{n_k}{n_f V_{\text{vox}}} s_k^{\text{cond,norm,NN}} \\ s_k^{\text{NN}} &\equiv \frac{n_k}{n_f V_{\text{vox}}} s_k^{\text{norm,NN}} \\ s_k^{\text{cond,norm,NN}} &\equiv -\frac{k_B}{2} \langle \ln \rho(\mathbf{r}k) \rangle_{L_k} \sum_{l \in L_k} \frac{n_{lk}}{n_k} \\ s_k^{\text{norm,NN}} &\equiv \frac{N_{L_k} k_B}{2} \left( \frac{n_k}{n_f} \right)^{-1} \left\langle \frac{n_{kll}}{n_l} \ln \rho(\mathbf{r}) \right\rangle_{L_k} \end{aligned} \quad (13)$$

Both average log terms inside eq 13 can be found using the NN method, with the gamma correction for the bias, as follows:

$$\begin{aligned} \langle \ln \rho(\mathbf{r}k) \rangle_{L_k} &= \frac{1}{N_{L_k|k}} \sum_{(ik) \in L_k} \left( \ln \frac{3}{n_k 4\pi (r_i^{\text{NN}})^3} - 0.5772 \right) \\ \left\langle \frac{n_{kll}}{n_l} \ln \rho(\mathbf{r}) \right\rangle_{L_k} &= \frac{1}{N_{L_k}} \sum_{i \in L_k} \left( \frac{n_{kll}}{n_l} \left[ \ln \frac{3}{n_f 4\pi (r_i^{\text{NN}})^3} - 0.5772 \right] \right) \end{aligned} \quad (14)$$

Here, the index  $i$  runs over water instances and  $r_i^{\text{NN}}$  is the distance between water instance  $i$  and its nearest-neighbor water instance.

The water instances are defined as follows. For the second expression in eq 14, the water coordinates found in all  $n_f$  MD frames are merged into one “superframe,” containing  $N n_f$  sets of water coordinates, so that each of the  $N$  water molecules in the system appears  $n_f$  times, each time with a different set of coordinates; each resulting set of water coordinates in the superframe is considered one instance of a water. The symbol  $\sum_{i \in L_k}$  indicates a sum over the  $N_{L_k} n_f$  water instances falling within the  $L_k$  grid, but the nearest neighbors of these water instances may be water instances which fall just outside the  $L_k$  grid. For the conditional term in the first line of eq 14, the water coordinates in all  $n_k$  frames for which a water molecule is found in  $k$  are merged into one superframe, which is now conditioned on the presence of a water at  $k$ , and each set of water coordinates in this conditional superframe is considered one instance of a water. The symbol  $\sum_{(ik) \in L_k}$  indicates a sum over the water instances in this superframe which fall within the  $L_k$  grid; again, the nearest neighbors of these conditional water instances may be conditional water instances which fall just outside the  $L_k$  grid.

**Nonlogarithmic Term.** The nonlogarithmic term in eq 5 may readily be computed for regions  $K$  and  $L$ , within the grid approach. Thus,

$$\begin{aligned} S_{\text{ww},KL}^{\text{trans,nonlog}} &= \frac{-k_B}{2} \int_K \int_{L_{r'}} [-\rho(\mathbf{r}, \mathbf{r}') + \rho(\mathbf{r}) \rho(\mathbf{r}')] \, \mathbf{dr} \, \mathbf{dr}' \\ &= \frac{k_B}{2} \int_K \mathbf{dr}' \rho(\mathbf{r}') \int_{L_{r'}} \rho(\mathbf{r}\mathbf{r}') \, \mathbf{dr} \\ &\quad - \frac{k_B}{2} \int_K \mathbf{dr}' \rho(\mathbf{r}') \int_{L_{r'}} \mathbf{dr} \rho(\mathbf{r}) \\ &\approx V_{\text{vox}} \sum_{k \in K} s_k^{\text{nonlog}} \end{aligned} \quad (15)$$

where

$$\begin{aligned} s_k^{\text{nonlog}} &\equiv \frac{n_k}{n_f V_{\text{vox}}} s_k^{\text{nonlog,norm}} \\ s_k^{\text{nonlog,norm}} &\equiv \frac{k_B}{2} [N_{L_k|k} - N_{L_k}] \end{aligned}$$

This term reflects the difference between the two-point density of water around the solute and the product of the corresponding one-point densities. As detailed in the Results section, once referenced to bulk, its contribution becomes negligible, at least for the cases analyzed here.

**2.3. Referencing the Results to Bulk Water.** As explained in section 2.1.1, the derivation and algorithms described above for water in the presence of a solute fixed in the lab frame are equally applicable to pure water. In order to reference the entropy of water in the presence of solute to pure water values matched numerically to those computed in the presence of the solute, we analyze a simulation of neat water with a small  $K$  grid

near the center of the simulation box and use the same grid spacing and  $L$ -grid parameters as used to analyze water in the presence of the solute to compute the value of  $TS_{\text{ww,bulk}}^{\text{trans}}$  for the water on the  $K$  grid. This quantity is then normalized by the mean number of water molecules on the  $K$  grid, for the pure-water calculation, to obtain  $S_{\text{ww,bulk}}^{\text{trans, norm}}$ , the mean pairwise translational entropy per water in bulk. Note that the dimensions of the pure water  $K$  grid need not be matched to the one used in the presence of solute: every voxel  $k$  is equivalent to every other, for the pure water case, so there is no need to cover any particular region. Now, if the  $K$  region in the presence of solute contains an average of  $N_K$  waters, again in the presence of solute, then one obtains the net pairwise entropy difference of the waters in the presence of solute versus in bulk from the following expression:

$$T\Delta S_{\text{ww}}^{\text{trans}} = TS_{\text{ww,KL}}^{\text{trans}} - N_K TS_{\text{ww,bulk}}^{\text{trans, norm}} \quad (16)$$

In the formulations detailed above, each voxel  $k$  on the  $K$  grid is associated with an entropy density. For example, for the nearest neighbor approach, this quantity is obtained from eq 13 as  $s_k^{\text{cond, NN}} - s_k^{\text{NN}}$ . This density may be referenced to bulk by subtracting the corresponding entropy density for bulk water; when the nonlogarithmic term is also included, the resulting entropy difference density is  $\Delta s_k^{\text{NN}} \equiv s_k^{\text{cond, NN}} - s_k^{\text{NN}} + s_k^{\text{nonlog}} - \rho S_{\text{ww,bulk}}^{\text{trans, norm}}$ . The corresponding per-water normalized entropy difference is  $\Delta s_k^{\text{norm, NN}} \equiv \frac{V_{\text{vox},f}}{n_k} \Delta s_k^{\text{NN}}$ . These voxel quantities may be visualized in terms of three-dimensional contours, as illustrated in the Results section.

**2.4. Validation of Grid Methods by Comparison with the Radial Distribution Method for Pure Water.** The applicability of the present methods to pure water provides an opportunity to validate them by comparing their results with those obtained by an independent computational approach. We use the fact that the second order translational entropy of a homogeneous liquid may be obtained from the one-dimensional radial distribution function (RDF) centered on a given water in the pure liquid:

$$S_{\text{ww}}^{\text{trans, RDF}}(R) = -k_B 2\pi\rho \int_0^R g(r) \ln g(r) r^2 dr \quad (17)$$

Formally, the upper limit of the integral,  $R$ , corresponds to the size of the entire system, but far smaller values of  $R$  yield good approximations, because water–water correlations decay quickly with values of  $r$  on the scale of the mean water–water distances, as detailed in the Results. Here,  $g(r)$ , the radial correlation function between water molecules, is estimated from  $n_f$  MD frames as

$$g(r) = \frac{3n}{4\pi n_f \rho \left[ \left( r + \frac{\Delta r}{2} \right)^3 - \left( r - \frac{\Delta r}{2} \right)^3 \right]} \quad (18)$$

where  $n$  is the number of water instances in the spherical shell of thickness  $\Delta r$  between  $r - \Delta r/2$  and  $r + \Delta r/2$ . In practice, we maximize statistical convergence by averaging the values of  $g(r)$  over all  $N$  waters in a simulation of pure water. The result from eq 17 was compared with that obtained by the grid methods as described in the prior subsection.

**2.5. NN Estimation of the First Order Translational Entropy.** The first order translational entropy of the water around a solute is given by

$$\begin{aligned} S_{\text{sw}}^{\text{trans}} &= -k_B \rho \int g(\mathbf{r}) \ln g(\mathbf{r}) d\mathbf{r} \\ &= -k_B \int \rho(\mathbf{r}) \ln \rho(\mathbf{r}) d\mathbf{r} + k_B N \ln \rho \\ &= -N k_B \langle \ln \rho(\mathbf{r}) \rangle + k_B N \ln \rho \end{aligned} \quad (19)$$

The contribution of region  $K$  to  $S_{\text{sw}}^{\text{trans}}$  may, for the sake of visualization and analysis, be decomposed into contributions from voxels  $k \in K$ , as follows:

$$\begin{aligned} S_{\text{sw,K}}^{\text{trans}} &= -k_B \int_K \rho(\mathbf{r}) \ln \rho(\mathbf{r}) d\mathbf{r} + k_B N_K \ln \rho \\ &= -k_B \sum_{k \in K} N_k [\langle \ln \rho(\mathbf{r}) \rangle_k - \ln \rho] \end{aligned} \quad (20)$$

where  $N_k = \frac{n_k}{n_f}$  is the mean number of waters in voxel  $k$ , so that  $N_K = \sum_k N_k$ , and  $\langle \ln \rho(\mathbf{r}) \rangle_k$ , the mean log density of water in voxel  $k$ , may be computed by the nearest neighbor method as

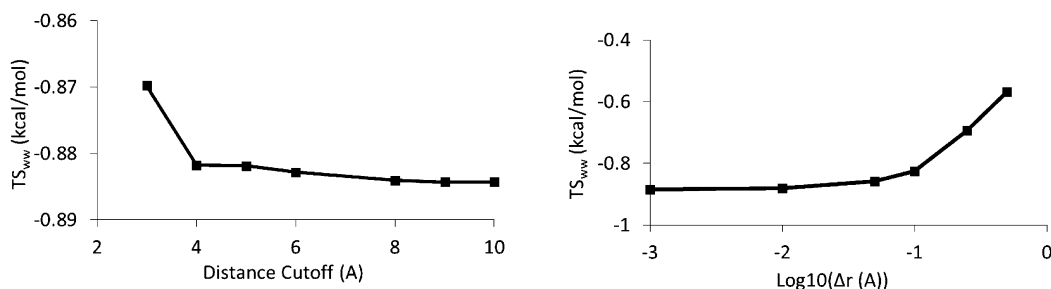
$$\langle \ln \rho(\mathbf{r}) \rangle_k = \frac{1}{n_k} \sum_{\text{water } i}^{n_k} \left( \ln \frac{3}{n_f 4\pi (r_i^{\text{NN}})^3} - 0.5722 \right) \quad (21)$$

**2.6. Implementation of Entropy Calculations.** The  $K$  and  $L$  regions used for both the histogram and NN entropy calculations were sized to fit within the simulation boxes (see details below), and the grid spacings were 0.5 Å, except as otherwise noted. The efficient ANN algorithm<sup>70</sup> was used to find nearest neighbors of water instances. Because the difference between the unconditional density and the density conditioned on a water in voxel  $k$  diminishes rapidly with distance from  $k$ , each voxel  $k$  was assigned its own cubic  $L$  grid, centered around  $k$ , in order to efficiently capture the local contributions of these differences to the overall second-order entropy. The side-length of the  $L$  grid was set to twice the desired cutoff distance within which correlations would be captured. For example, a cutoff distance of 5 Å leads to a 10 Å  $L$  grid. The first-order orientational entropy was computed with the previously described GIST NN method,<sup>51</sup> and the first order translational entropy was computed by the NN method described in the subsection above.

**2.7. Simulation Details.** The pure water calculations used in section 3.1 were carried out in cubic simulation boxes with side lengths of about 32 Å and periodic boundary conditions. The temperature was maintained at 300 K with the Langevin thermostat and a collision frequency of 2.0 ps<sup>-1</sup>; pressure was maintained at 1 atm by isotropic position scaling with a pressure relaxation time of 0.5 ps. A 9 Å cutoff distance was employed for all nonbonded interactions in the pairlist, and the particle-mesh Ewald method was used to account for long ranged electrostatic interactions. The simulations were run for 1.5 μs on a single GPU, with the pmemd.cuda component of AMBER 12<sup>71</sup> or AMBER 14, with a 2 fs time step. The SHAKE algorithm<sup>72</sup> was used to constrain the lengths of all bonds involving hydrogen atoms. Frames were saved for analysis every 0.5 ps.

Parameters were assigned to cucurbit[7]uril (CB7) with the GAFF<sup>73</sup> force field and RESP<sup>74</sup> with the HF 6-31G\* basis set. The starting structure was solvated with 1096 water molecules, resulting in a 10 Å buffer of water surrounding the solute, using the program Tleap in AmberTools.<sup>75</sup> The system was then briefly energy-minimized with Cartesian positional restraints on CB7, with force constants of 10 kcal/mol/Å<sup>2</sup>. The minimization comprised 1500 steps of the steepest descents algorithm,





**Figure 2.** Second-order water–water entropy, provided as a free energy contribution  $TS$  (kcal/mol), computed for pure TIP3P water with the RDF method. Left: entropy as a function of the distance cutoff,  $R$  in eq 17, for a spherical shell thickness,  $\Delta r$  in eq 16, of 0.001 Å. Right: entropy as a function of spherical shell thickness, for a distance cutoff of 10 Å; the two right-most points are for shell thicknesses of 0.25 and 0.5 Å. These calculations used 50 000 MD frames; extending to 200 000 frames changes the results by only a few thousandths of a percent.

followed by a maximum of 2000 steps of the conjugate gradient method. The crystal structure of FXa complexed with the inhibitor ZK-807834 (alternatively called CI-1031), from Protein Data Bank entry 1FJS,<sup>76</sup> was prepared and simulated as described previously;<sup>34</sup> the ligand was deleted to leave an empty binding site, and two ions observed in the crystal structure ( $\text{Ca}^{2+}$  and  $\text{Cl}^-$ ) were included as part of the solute and hence were restrained along with the protein atoms, as detailed below. TIP3P<sup>23</sup> water molecules were added to solvate the protein using the program Tleap,<sup>77</sup> and the AMBER99sb force field<sup>78</sup> was used to assign the atomic parameters. The final system comprised 29 338 atoms, including the 8557 water molecules. The MD simulation started with an energy minimization of 1500 steps by the steepest descents algorithm, followed by conjugate gradient energy minimization for a maximum of 2000 steps. During the initial minimization, all protein atoms were harmonically restrained to their initial positions with a force constant of 100 kcal/mol/Å<sup>2</sup>. The second minimization further relaxed the system by keeping only non-hydrogen protein atoms restrained with the same force constant. The energy-minimized systems were heated in increments of 50 K lasting 20 ps each, at constant volume and temperature. After the system reached 300 K, it was then equilibrated for 10 ns at a constant temperature and constant pressure of 1 atm. The equilibrated system was further run for an additional 5 ns at constant volume, and data were then collected at NVT. During the heating process and thereafter, all solute atoms were harmonically restrained to their energy-minimized positions with a force constant of 100 kcal/mol/Å<sup>2</sup> for FXa and 10 kcal/mol/Å<sup>2</sup> for CB7. Other simulation parameters were the same as for the pure water simulations described above. A 400 ns production simulation was run for FXa, with coordinates saved every 1 ps, for a total of 400 000 snapshots. For CB7, the production run was 600-ns-long, with coordinates saved every 0.5 ps, for a total of 1 200 000 snapshots.

### 3. RESULTS

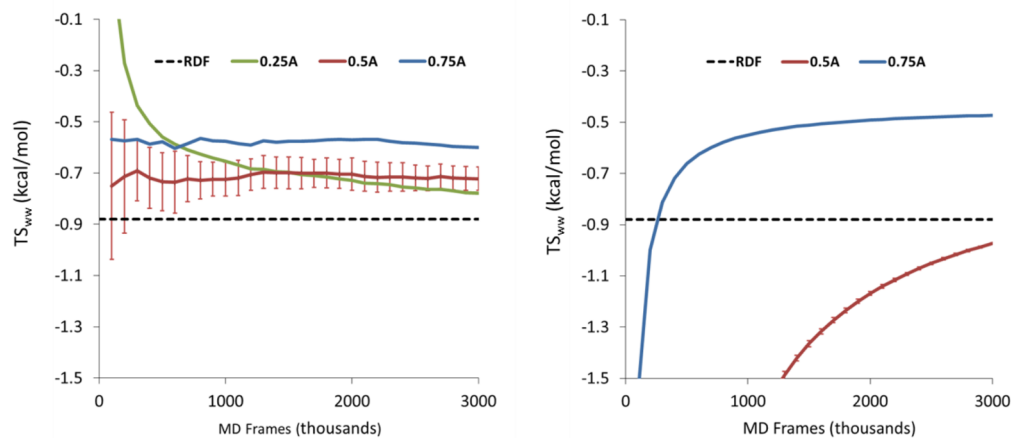
#### 3.1. Second Order Translational Entropy of Pure Water.

**3.1.1. Radial Distribution Function Method.** Here, we study the second order translational entropy of pure water, for which the RDF method is available for reference. (It is worth noting that the RDF method is not applicable to the solute-water systems, for which the grid methods were developed.) First, as a verification of the present methodology, we used the RDF method, with spherical shells of thickness  $\Delta r = 0.01$  Å, and 200 000 MD frames, to estimate  $S_{\text{ww}}^{\text{trans,RDF}}$  for the TIP3P, TIP4P,<sup>23</sup> TIP4PEW,<sup>27</sup> and SPC/E<sup>25</sup> water models; the results all agree with prior calculations<sup>79</sup> to within 0.5%, with the exception of TIP3P, for which the difference was 1.9%. (A replicate TIP3P

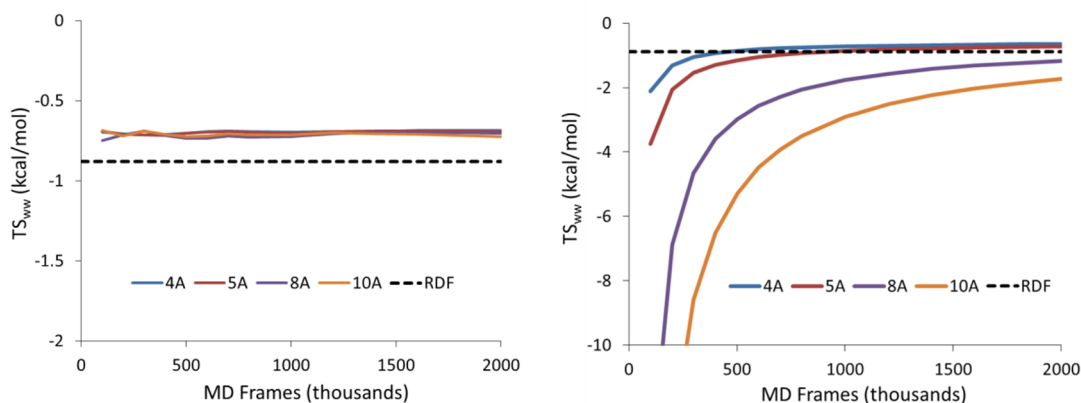
simulation yielded the same result.) No significant changes were observed on extending any of the simulations to 400 000 frames. We then examined the RDF results as a function of the two numerical parameters  $R$ , the upper limit of the integral in eq 17, and  $\Delta r$ , the width of the spherical shells in eq 16. As detailed in Figure 2 (left panel), for the TIP3P water model, the entropy from the RDF method converges quickly as the upper limit of the integral increases, changing by less than 0.3% when  $R$  increases from 4 to 10 Å. This observation supports the use of local  $L(k)$  grids (section 2.2) of modest size. The RDF result also depends on the width of the spherical shell, closely approaching its presumed asymptote for values less than or equal to 0.1 Å (Figure 2, right panel). Intuitively, the entropy falls as the shells become thinner and thereby reveal finer structure in the water distribution. This result suggests that voxels larger than about 0.1 Å in linear dimension are likely to overestimate the entropy somewhat, due to smoothing of the density distributions.

**3.1.2. Grid Methods.** The grid methods, both histogram and NN, were used to compute the same quantity, by defining region  $K$  as a cube of side length 1.5 Å, centered in a pure water simulation box and divided into cubic voxels  $k$ , each of side length  $R_{\text{vox}}$ . Each voxel  $k$  was associated with its own cubic region  $L_k$  (see Figure 1) centered on  $k$  and extending a fixed distance  $R_L$  from  $k$  along each grid axis, so that the side of each region  $L_k$  was of length  $2R_L$ . For the histogram method, the cubic  $L$  regions were divided into voxels  $l$  of side-length  $R_{\text{vox}}$  and the  $K$  and  $L$  grids were registered so that each  $k$  voxel was also an  $l$  voxel. However, not every  $l$  voxel was a  $k$  voxel, because  $L$  extended outside  $K$ . The agreement of the grid method with the RDF method is examined below, as a function of the voxel size,  $R_{\text{vox}}$  and the size of the  $L$  grids,  $R_L$ .

The NN method provides reasonably well-converged results within 1–2 million MD frames, for voxel sizes,  $R_{\text{vox}}$ , of 0.75 and 0.5 Å. Thus, as shown in Figure 3 (left), the mean of  $TS_{\text{ww}}$  across all 27  $k$  voxels (solid lines) changes little with increasing simulation time for both of these grid spacings. In addition, for 0.5 Å voxels (red), the standard deviation of  $TS_{\text{ww}}$  across the 27  $k$  voxels falls to 0.066 kcal/mol at 1 million frames and 0.05 kcal/mol at 2.3 million frames (red error bars). For 0.75 Å voxels (blue), the standard deviation is 0.042 kcal/mol at 1 million frames (data not shown). However, calculations with these grid spacings converge to values that overestimate the reference RDF result somewhat, particularly for a grid spacing of 0.75 Å (Figure 3, blue). Because the NN method does not make use of the discretization of the  $L$  grid into voxels  $l$ , these systematic errors are attributed to the fact that larger  $k$  voxels lead to greater smoothing of the conditional probability density functions over the corresponding regions  $L_k$ . Going to a finer grid



**Figure 3.** Dependence of second order pure water entropy on grid spacing. Left: NN results for grid spacings ( $R_{\text{vox}}$ ) of 0.25 (green), 0.5 (red), and 0.75 Å (blue), with the reference RDF result (dashed line; computed with  $R = 15$  Å,  $\Delta r = 0.01$  Å, and 200 000 MD frames). Error bars for the 0.5 Å results indicate the standard deviations across all 27  $k$  voxels in this  $K$  region. Right: Analogous results from the histogram method; results for a grid spacing of 0.25 Å were computed but are below the scale of the graph. All calculations used  $L$  regions with  $R_L = 8$  Å, and all entropies are reported as free energy contributions, TS, in kcal/mol.



**Figure 4.** Dependence of second order pure water entropy on size of  $L$  regions. Left: Comparison of NN results, for  $L$  regions with  $R_L$  of 4–10 Å, with the reference RDF result (dashed line;  $R = 15$  Å,  $\Delta r = 0.01$  Å, and 200 000 MD frames). Right: Analogous results from the histogram method. All results are for a grid spacing of 0.5 Å, and all entropies are reported as free energy contributions, TS, in kcal/mol.

spacing,  $R_{\text{vox}}$ , of 0.25 Å yields ultimately a closer approach to the reference RDF result (Figure 3, green), but convergence is notably slower. The histogram method converges slowly with the number of MD frames (Figure 3, right). This is evident even for the relatively coarse 0.75 Å grid spacing, and our 0.25 Å results are below the scale of the graph. In addition, it is evident that the 0.75 Å result is converging to a result that is further from the reference RDF result than the NN 0.75 Å result. This difference presumably results from further smearing of the conditional density functions, due to the added discretization of the  $L_k$  regions, which is not required for the NN method.

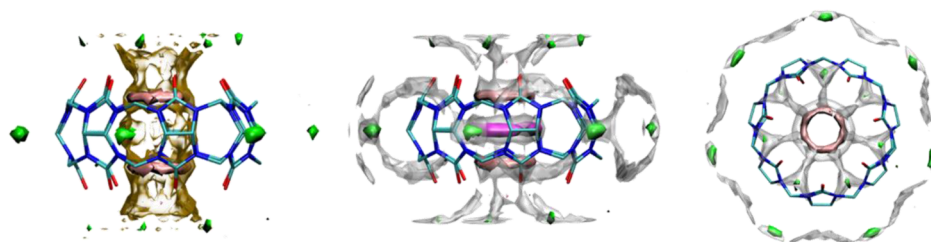
As detailed in the Appendix, further comparison of the NN and histogram methods, with grid spacings down to 0.15 Å, highlights the advantage of NN over the histogram method, in terms of the rate of convergence and the amount of sampling required to reach the reference RDF result. The graphs in the Appendix suggest that the RDF result would be closely approximated by the NN method at a grid spacing of 0.15 Å, given about  $4 \times 10^6$  snapshots, whereas the histogram method would require far more sampling.

The present calculations depend not only on the grid spacing, but also on the size of the  $L_k$  region, in which the perturbation of water density by a water in each voxel  $k$  is evaluated. It is thus

appropriate to evaluate the sensitivity of the second order entropy calculations to the size of the  $L_k$  regions. Based on the RDF study (Figure 2, left), one may anticipate that  $R_L \geq 4$  Å should suffice to capture most of the second-order entropy for the TIP3P water model examined here. This expectation is borne out by the insensitivity of the NN results to  $R_L$ , over a range of 4–10 Å (Figure 4, left); in addition, good numerical convergence is observed over this range. It is worth noting that, because the  $L$  regions are cubic, a value of  $R_L$  equal to 5 Å, for example, includes locations as far as  $\sqrt{3}R_L = 8.9$  Å from the center of voxel  $k$ . On the other hand, the poor convergence of the histogram method (Figure 4, right) makes it difficult to assess the sensitivity of its results to the value of  $R_L$ .

On the basis of this analysis of calculations for pure water, we selected a grid spacing of 0.5 Å as a good working compromise between speed and accuracy for the cucurbituril and protein calculations described below, and we set  $R_L$ , the size of the  $L_k$  regions, to 5 Å. The bulk value used to compute the change in the second order entropy is taken from a pure water calculation under the same conditions and with 2 000 000 MD frames. In addition, given the poor convergence of the histogram method, the following calculations use only the NN method for the second order calculations.





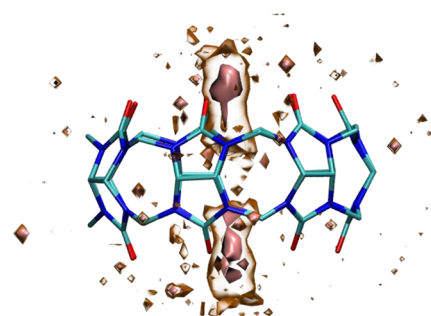
**Figure 5.** Spatial maps of second order translational entropy density of water, relative to bulk, in and around host CB7, based on 1 200 000 MD frames. Left: Entropy contoured at  $-0.02$  kcal/mol/Å<sup>3</sup> (pink),  $-0.01$  kcal/mol/Å<sup>3</sup> (transparent orange), and  $+0.005$  kcal/mol/Å<sup>3</sup> (green). Middle: same, but omitting the contour at  $-0.01$  kcal/mol/Å<sup>3</sup> and additionally showing water number density contoured at  $0.1$  waters/Å<sup>3</sup> (gray) and  $0.175$  waters/Å<sup>3</sup> (magenta). Right: rotated view, showing only entropy density contours at  $-0.02$  and  $0.005$  kcal/mol/Å<sup>3</sup> and number density contour at  $0.01$  waters/Å<sup>3</sup>, with the same colors. Entropies are multiplied by temperature to convert to kcal/mol. These and all other molecular graphics in this paper were generated with the program VMD.<sup>80</sup>

### 3.2. Cucurbit[7]uril (CB7). 3.2.1. Visual Mapping of Second-Order Translational Entropy.

**Figure 5** visualizes the spatial distribution of the second-order translation entropy, relative to bulk, in and around the synthetic host molecule CB7, which we previously studied to first order in entropy.<sup>51</sup> As shown in the left panel, there is a large region of water with reduced second order entropy density, relative to bulk, centered around the symmetry axis of CB7 (yellow), with particular enhancement in two toroids (pink) roughly level with the upper and lower nitrogen-carbon rings of the host. These are regions where more highly correlated water (hence reduced second-order entropy relative to bulk) is present at relatively high number density, leading to notably negative pairwise entropy density relative to bulk. On the other hand, small loci of water with positive pairwise entropy density relative to bulk (i.e., less correlated than bulk) are present at seven symmetry-related sites around the relatively apolar exterior of the host (green), and also at symmetry-related sites between the carbonyl oxygens at the upper and lower portals of the host (green).

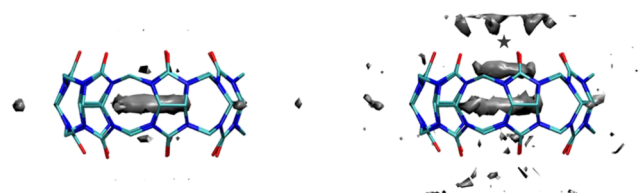
It is of interest to compare this distribution of pairwise entropy density with the simple number density of water, which is contoured in gray and magenta in the middle and right-hand panels of **Figure 5**. Whereas the number density is highest in the central torus (magenta, middle panel), the absence of a central torus in the entropy contours of the left-hand panel indicates that the water molecules in the central torus are not especially correlated with other water molecules. In contrast, water molecules present at intermediate density in the upper and lower tori (middle panel) are highly correlated and therefore appear prominently in the entropy map (pink in the left-hand panel). Similarly, the regions of elevated pairwise entropy (green) appear as a subset of the high number-density regions contoured in gray (middle and right-hand panels).

Further insight may be obtained by examining the pairwise entropy on a per-water-molecule basis, rather than on the basis of entropy density, as shown in **Figure 6**. The contours are more irregular than those in **Figure 5**, because, unlike the density-weighted entropies there, here the contours involve voxels where the number density is low, so the numerical convergence is worse. Nonetheless, it is interesting to observe that the most correlated water molecules are those at the centers of the two portals, as highlighted in the pink and clear orange contours. In contrast, the water in the center of the host is not especially correlated, as indicated by the lack of contouring there. This result is consistent with the absence of a torus of entropy density in the left-hand panel of **Figure 5**.



**Figure 6.** Spatial map of second order translational water entropy per water molecule, relative to bulk, in and around host CB7. Pink: contour at  $-0.5$  kcal/mol/water. Transparent orange:  $-0.3$  kcal/mol/water. No significant positive regions were observed after setting aside voxels with inadequate sampling, e.g., those with number densities less than  $0.001$  water molecules/Å<sup>3</sup>.

The strongly negative pairwise entropy of water molecules at the centers of the portals (pink in **Figure 6**) suggests that the density distribution of water is strongly influenced by the presence of a water at these locations, as discussed in [section 2.1.2](#). Put differently, the density conditioned on the presence of water in the middle of the portal should be quite different from the unconditioned density. This is indeed the case, as shown in **Figure 7**. Here, the left-hand panel shows the baseline density



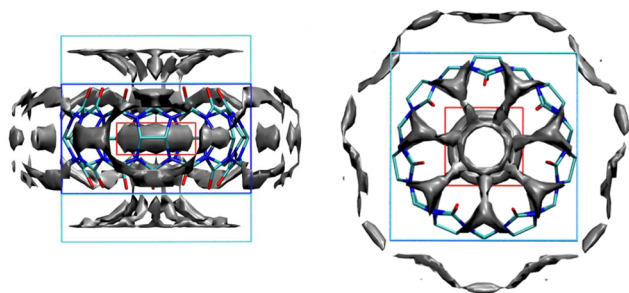
**Figure 7.** Number density of water, contoured at  $0.15$  water molecules/Å<sup>3</sup>, in and around CB7. Left: standard number density derived directly from simulation. Right: number density conditioned on the presence of a water molecule at the location marked by a star.

distribution of water, with a strong central torus and a few other small loci of high density, while the right-hand panel shows water density conditioned on the presence of a water at the center of the portal (star in figure), contoured at the same level as in the left-hand panel. Despite some noise in the conditional contours, which results from a reduction in the number of MD frames available for averaging, due to the condition, one can clearly see a dramatic increase in water structure, with the appearance of an additional torus below the star and a more jagged ring of density

above it. This increase in structure gives a conditional entropy lower than the baseline unconditional entropy and hence a negative contribution to the overall pairwise translational entropy from the voxel corresponding to the star.

**3.2.2. Quantifying the Second Order Translational Entropy of Water in and around CB7.** Binding of guest molecules to the CB7 host leads to displacement of water molecules to the bulk from regions in and near the binding cavity. The resulting change in water thermodynamics makes a contribution to the overall thermodynamics of binding, so it is of interest to examine the magnitude and sign of these contributions. In a previous study, we used the histogram method to study the first-order translational entropy, and the NN method to study the orientational entropy, of water in the CB7 binding cavity.<sup>51</sup> Here, we report the second-order translational entropy for regions of interest in and around CB7 and consider it in the context of the first-order terms. We also compare the present NN implementation of the first order translational entropy (section 2.5) with the prior histogram method.

Three regions are considered (Figure 8): the cavity (blue), which comprises the complete interior of the host (including the



**Figure 8.** Regions of interest in and around the CB7 host, shown in side (left) and top (right) views, with the host (stick diagram, omitting hydrogen atoms), and water density contoured at 0.10 water molecules/ $\text{\AA}^3$  (gray). Central blue box: cavity region. Union of upper and lower cyan boxes: portal region. Red box: torus region.

torus, below), with upper and lower cutoffs positioned at a trough in water density near the level of the host's carbonyl oxygens; the torus (red), which is a subset of the cavity region holding only the central torus of high-density water;<sup>51</sup> and the portals (cyan), which extend 2  $\text{\AA}$  above and below the upper and lower borders, respectively, of the cavity region and hold regions of increased water density above the upper and lower rings of carbonyl oxygens. It should be noted that the geometric definitions of the cavity and torus regions used here have been refined, relative to those used previously,<sup>51</sup> so the first-order

thermodynamic results for them differ somewhat. In particular, the cavity is somewhat taller, to reach the troughs in water density at the portals, and the torus is slightly shorter, to capture only its peak density. The results (Table 1) are reported as regional integrals, as indicated by the superscript "R" in each case, and also on a per-water-molecule basis, which provides information on the average properties of water in each region.

For all three regions—cavity, torus, and portal—the net second-order translational entropy has the same sign as the first-order translational and orientational entropy terms. Thus, the second-order term always reinforces, rather than partly cancels, the first order term. This result contrasts with the heuristic assumption that the second-order term tends to partly cancel the first-order one.<sup>52,67</sup> Although the second-order contribution is substantially smaller than the first order terms (Table 1), at 5–8% of their sum, it is large enough to substantially influence the thermodynamics of hydration and of guest-binding to this synthetic host molecule. It is worth noting that the non-logarithmic term contributes minimally to these second order terms, 0.08, 0.04, and  $-0.12$  kcal/mol, for the cavity, torus, and portal regions, respectively.

It is also of interest that the first-order orientational term is about equal to the first-order translational term for the cavity and torus regions, but it is about double the first-order translational term for the portal. Presumably, waters at the portal are particularly well-ordered because they form hydrogen bonds with the carbonyl oxygens of the host.

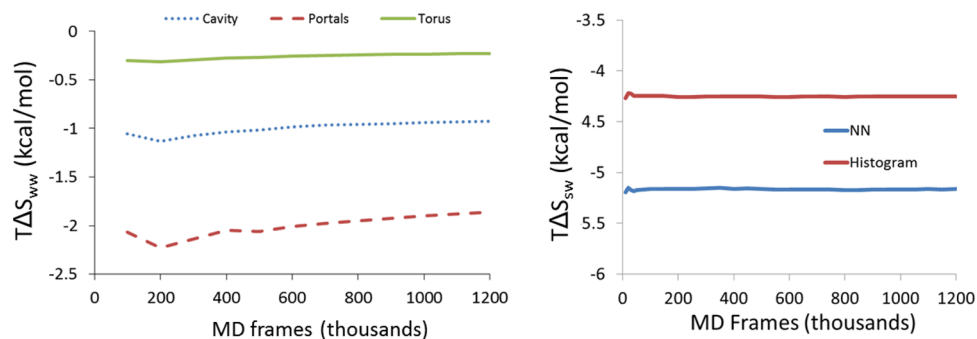
The entropic properties of water molecules in the three regions may be characterized by correcting for the mean number of waters in each region, as done in the last three columns of Table 1. The most correlated waters, based on the value of the second-order translational entropy per water, are found in the cavity region. The degree of correlation within the torus is smaller, and similar to that in the portal region. Since the torus is a subset of the cavity, the high second-order entropy of the cavity water as a whole clearly reflects the particularly high correlation of water above and below the central torus, which is evident in the left panel of Figure 5. The fact that the first-order translational entropy per water of water in the torus is lower ( $-0.86$  kcal/mol/water) than that for the cavity as a whole is simply a reflection of the fact that water is present at higher density in the torus (left panel, Figure 7). The least correlated waters are in the portal region, although the difference relative to the torus is small.

The second-order translational entropy is based on a six-dimensional probability distribution function, the two-point water density  $\rho(\mathbf{r},\mathbf{r}')$ , and thus can require long simulations to achieve acceptable convergence. Reasonably good convergence is obtained here for the various regions in and around CB7

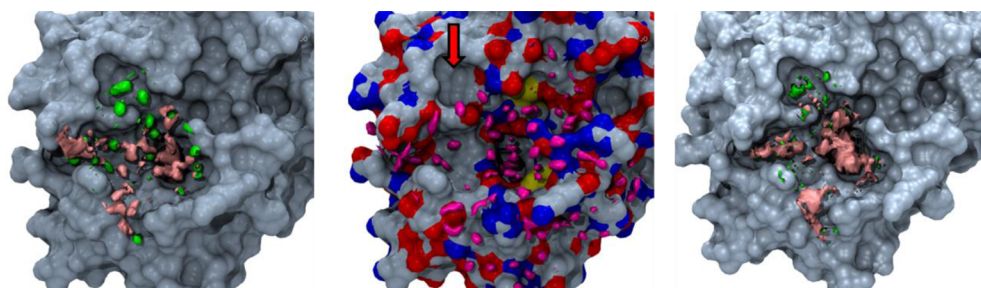
**Table 1. Water Entropy Contributions, As Defined in Text, for Regions in and near the Binding Cavity of the CB7 Host, and for the Region of the FXa Binding Site Occupied by Ligand ZK-807834 (CI-1031)<sup>a</sup>**

	$N_K$	$T\Delta S_{ww}^{R,trans,NN}$	$T\Delta S_{sw}^{R,trans,NN}$	$T\Delta S_{sw}^{R,orient,NN}$	$\frac{T\Delta S_{sw}^{R,trans,NN}}{N_K}$	$\frac{T\Delta S_{sw}^{R,trans,NN}}{N_K}$	$\frac{T\Delta S_{sw}^{R,orient,NN}}{N_K}$
cavity	7.54	-0.96	-5.17	-6.33	-0.13	-0.69	-0.84
torus	2.82	-0.24	-2.42	-2.41	-0.09	-0.86	-0.85
portal	25.85	-1.95	-9.66	-18.32	-0.08	-0.37	-0.71
FXa	16.34	-4.96	-15.31	-26.43	-0.30	-0.94	-1.62

<sup>a</sup>The CB7 regions are defined in the text and in Figure 8.  $N_K$ : mean number of water molecules in each region. The first three entropy columns (kcal/mol) are sums over the respective regions; the second three entropy columns (kcal/mol/water molecule) are normalized by the numbers of waters in the respective regions and thus report on the average property of water in each region. The results are based on  $K$  and  $L$  grids with a 0.5  $\text{\AA}$  spacing and  $L$  grids with  $R_L = 5 \text{\AA}$ , as noted in the text.



**Figure 9.** Convergence of second-order translational entropy of water in and around host CB7 as a function of MD frames (left), and of first order translational entropy for the cavity region, by the NN and histogram methods (right). The regions are defined in the text and Figure 8.



**Figure 10.** Water properties in the active site of enzyme FXa. Left: Second-order translational entropy density of water, scaled by  $T$  to yield kcal/mol/Å<sup>3</sup>, contoured at  $-0.02$  (pink) and  $+0.005$  (green) kcal/mol/Å<sup>3</sup>, with molecular surface uniformly colored. Middle: Water number density contoured at  $0.175$  waters/Å<sup>3</sup>, with protein surface colored by element; blue-gray, carbon; blue, nitrogen; red, oxygen; hydrogens not show. Right: Second-order entropy per water molecule (kcal/mol/water), contoured at  $-0.5$  kcal/mol/water (pink) and  $+0.16$  kcal/mol/water (green). Note that the number density contours were computed for a somewhat larger region than the entropy density contours, especially toward the bottom and right of the active site. All simulations represented here are 400 ns long, with frames analyzed every 1 ps.

(Figure 9, left), though the portal graph, in particular, still has a visible trend after 600 ns of simulation time (1 200 000 frames).

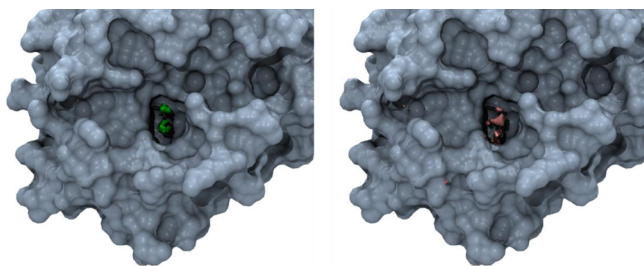
Finally, it is worth commenting here on the NN implementation of the first order translational entropy (section 2.5), a term we previously computed with the histogram method. For the cavity region, both the NN and histogram method with a  $0.5$  Å grid spacing converge well (Figure 9, right), but the histogram result is higher than the NN result. In general, the NN method yields entropies for the three regions 13–22% lower than those provided by the histogram method (results not shown). This is because, unlike the NN method, the histogram method approximates the water density as uniform within each voxel. This approximation smooths the water density and thereby leads to higher entropy estimates.

**3.3. Factor Xa.** **3.3.1. Visual Mapping of Second-Order Translational Entropy.** Water in and near the active site of the enzyme FXa adopts a complicated distribution of increased and decreased translational correlation, as manifest in contours of the second-order translational entropy density, relative to bulk, shown in Figure 10 (left panel). (Note that, although the simulations analyzed here placed the protein in a large box of explicit solvent with period boundary conditions, the  $K$  grid covered only part of the surface, so the contours are localized.) Indeed, every hollow in the surface of the enzyme is occupied by contours of negative (pink) or positive (green) entropy density (Figure 10, left), relative to bulk, whereas water at more convex regions of the surface shows fewer entropic features. The tendency of this entropy term to be perturbed in enclosed regions of the FXa surface is consistent with the fact that, for CB7, the greatest perturbations also occurred within the binding cavity.

This patterning may be considered in the context of the polarity of the enzyme's surface and the number density of water (Figure 10, middle). For example, a large, nonpolar pocket without a strong elevation of number density (red arrow, Figure 10, middle) nonetheless has a substantial increase in second-order translational entropy density (Figure 10, left), indicating reduced water–water correlation. Accordingly, the water molecules in this pocket also have a notable increase in the corresponding per water entropy (Figure 10, right). Further examination of the figure may suggest a tendency for elevated water–water entropy (decreased correlation) at other patches of hydrophobic surface too, and a tendency for decreased water–water entropy (increased correlation) in more polar clefts. On the other hand, water in the relatively hydrophobic interior of CB7 showed reduced, rather than increased, water–water translational entropy, so any connection between hydrophobicity of the surface and reduced water–water correlation cannot be viewed as universal.

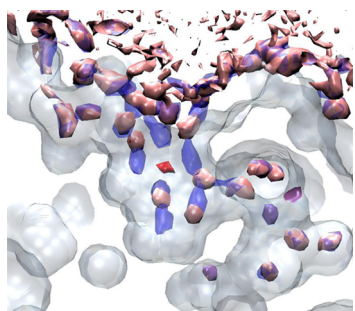
In the case of CB7, water at the center of the portal to the binding-site cavity was found to have particularly low second-order translational entropy per water molecule, implying a high degree of water–water correlation (section 3.2.1). A similar phenomenon, though more pronounced, is also observed in the case of FXa, as the entryway to a deep, water-filled cavity is occupied by water with even lower values of the second-order translational entropy than seen in the case of CB7, on a per water basis (Figure 11, left), and in terms of entropy density (Figure 11, right). Presumably water at such locations is in a particularly good position to influence the density distribution within the nearby cavity or cleft. This was seen previously for CB7





**Figure 11.** Concentration of negative second-order translational entropy at the entry of a deep, water-filled pocket in FXa. Left: second order translational entropy per water molecule contoured at  $-2$  kcal/mol/water. Right: second order translational entropy density, contoured at  $-0.1$  kcal/mol/Å<sup>3</sup>.

(Figure 7) and maybe also be observed in the case of FXa. Thus, the number density distribution of water in the deep cleft of FXa is very different when conditioned on the presence of a water molecule in an entryway site of strongly negative translational entropy per water ( $-3.5$  kcal/mol/water, red) than when it is unconditioned (Figure 12, pink conditioned vs blue un-

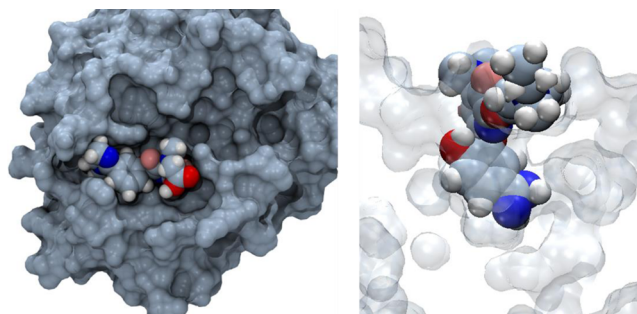


**Figure 12.** Comparison of unconditional (blue) number density distribution of water with number density distribution conditioned on the presence of a water at the red site (pink), in the active site cleft of FXa. This is a side view, relative to the other FXa representations; front and rear clipping and a transparent protein surface allow visualization of density contours deep in the cleft. The water densities are contoured at  $0.1$  water molecules/Å<sup>3</sup>; the red contour of second order translational entropy is contoured at  $-3$  kcal/mol/water molecule.

ditioned density contours). The red site here is at the center of one of the green regions in Figure 11.

**3.3.2. Contribution of the Second Order Translational Entropy to Ligand Binding.** Prior studies have considered the role of the first-order entropy of water in the binding site of FXa to the binding affinity of drugs and drug-like ligands to this enzyme.<sup>34,45</sup> The present methodology now allows examination of the second-order translational entropy of water in this region. We focus on the water displaced by the ligand (PDB HET ID Z34) present in the binding site of FXa in the PDB entry studied here, 1JFS. This ligand occupies part of the complicated active site region (Figure 13, left) and displaces water from the deep cleft that contains the most correlated water (Figures 11, 12, and 13, right).

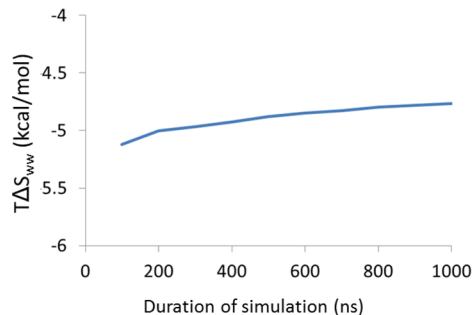
At nearly  $-5$  kcal/mol (Table 1), the integrated water–water translational entropy in the region occupied by this ligand is large enough to have a major impact on the ligand affinity. This value is substantially greater than the corresponding result for the largest regional integral of the pairwise entropy for CB7, which is about  $-2$  kcal/mol, for the portal region (Table 1), even though the number of waters in the portal region is larger. Accordingly, the



**Figure 13.** FXa with cocrystallized ligand (PDB HET ID Z34, InChI Key NPNSVNGQJGRSNR-UHFFFAOYSA-N) seen from outside of active site (left) and in a rotated view with transparent, z-clipped protein surface to show penetration of the ligand into the deep active site cleft also shown in Figure 12.

second-order entropy per water is about 3-fold larger (more negative) in the binding site than in or around CB7. The first order entropy terms are also greater (more negative) for FXa than for CB7, on both a regional and a per water basis (Table 1), but not by as large a factor as the second-order term. As observed in the case of CB7, the nonlogarithmic contribution to the second-order entropy is minimal, at  $0.20$  kcal/mol. Also, again as for CB7, the second-order term contributes with the same, negative, sign as the first-order term, thus reinforcing it rather than balancing it.

Two potentially balancing sources of error in the second order term should be noted. On one hand, this quantity is not fully converged, as shown in Figure 14, and a longer simulation would



**Figure 14.** Convergence of translational water–water entropy for region overlapped by ligand Z34 in the binding site of FXa, reported in kcal/mol.

apparently lead to a somewhat less negative value. On the other hand, as discussed in section 2.3, the use of a  $0.5$  Å grid leads to smoothing of the conditional water densities, and the use of a finer grid, if numerically practical, would lead to a somewhat more negative value. Thus, the combination of a finer grid and a much longer simulation might shift the final result either up or down a bit, and the result in Table 1 is likely a reasonably good estimate of the desired quantity.

#### 4. DISCUSSION

We have described a grid-based methodology, which uses inhomogeneous solvation theory to estimate and regionally map the two-body translational contribution to the hydration entropy of a solute of interest, using the data from a simulation of the solute immersed in explicit water molecules. The present formulation shows that the contribution to this two-body term associated with location  $r'$  tells the degree to which the first-order

translational entropy of water is changed when it is computed from a density distribution conditioned on the presence of a water molecule at  $r'$ . This contribution must also be weighted by the density of water at  $r'$ . Integrating over  $r'$  then yields the total second order translational entropy, which may be referenced to the corresponding quantity for bulk water. Our implementation of this idea combines a gridded discretization of  $r'$  with a nearest neighbor method that efficiently computes the required conditional translational entropies. Evaluating this second order entropy term is substantially more computationally demanding than evaluating the first order translational and orientational terms, but the numerical convergence obtained, both in terms of grid spacing and simulation length, appears good enough to provide at least semiquantitative information and intuition about how water–water correlations contribute to the hydration entropy of solutes. It is worth emphasizing that the NN method used here yields efficiency and accuracy far beyond that afforded by a simple histogram-based method.

The present method has several methodological parameters that affect the trade-off between accuracy and the amount of data needed to achieve a given level of numerical convergence. On the basis of the RDF study, one would ideally use a grid spacing of about 0.1 Å to achieve highly accurate correlation results. However, very long simulations would be required to converge these calculations, so we have used a grid spacing of 0.5 Å, which seems to afford a reasonable balance between accuracy and convergence, based on the RDF study. However, it will be of interest to press for further refinement, via either improved algorithms or brute-force increases in simulation lengths. Another relevant parameter is the size of the  $L_k$  grid positioned around each  $k$  voxel to capture the effects of a water at  $k$  on the local water density distribution. For the TIP3P water model used here, we obtained reasonable results with an  $L$  grid of about 10 Å side length, which hence extends a minimum of 5 Å in all directions from voxel  $k$ . However, the RDF of TIP3P water reaches bulk density at a rather short distance, and other water models, whose RDF deviates from bulk at longer distances, may require a larger  $L$  grid.

We find that this second order term contributes only 5–8% of the first order terms to the hydration entropy associated with regions of interest in and around the synthetic host molecule CB7 and in the active site region of the enzyme FXa. Interestingly, the second-order term is more significant for FXa, with its more polar and geometrically complex binding site, than for CB7. Indeed, for FXa, the integral of the second order term over the region occupied by a high-affinity, small molecule inhibitor is about  $-5$  kcal/mol, which is substantial on the scale of the free energies of ligand-protein binding. Hence, the second-order translational term probably makes a substantial

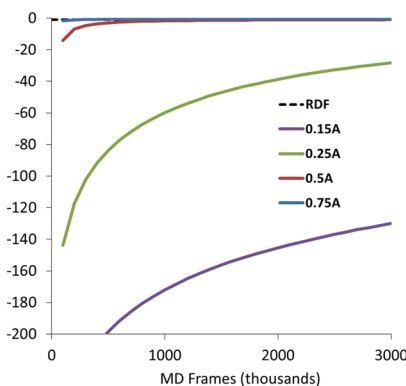
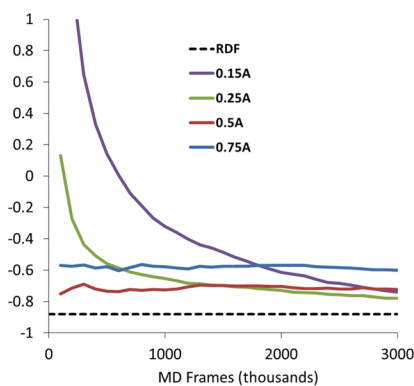
contribution to the change in hydration entropy on protein–ligand binding. It is also worth remarking that the second order term contributes with the same, negative, sign as the first order terms, in the cases studied here; intuitively, the presence of the solute increases water–water correlations and further decreases the entropy. This result appears to be consistent with the room temperature results previously obtained in a study of methane hydration,<sup>64</sup> but less so with prior heuristic estimations in which the water–water term is assumed to increase, rather than decrease, the hydration entropy.<sup>52,67</sup>

Three-dimensional mapping of the second order entropy contribution in and around CB7 and FXa suggests that this local quantity depends not only on whether the nearby solute surface is polar or nonpolar, but also on overall shape of the nearby surface. In particular, the entryways to deep binding cavities have particularly low values of  $T\Delta S_{\text{ww,trans}}$ , indicating a high degree of correlation with other waters. This makes intuitive sense, as these are locations where the presence or absence of a water can strongly effect how other waters are able to pack and fill the nearby cavity, making for a large difference in the baseline and conditional translational entropy of water. We anticipate that this “entryway effect” will be important in many other systems as well, such as at the entryways of nanotubes and of other deep protein binding sites, like the active site gorge of the enzyme acetylcholinesterase.<sup>81</sup> Displacement of water from such locations, if the density is significant there, is expected to be thermodynamically favorable. Of course, it will be important to consider such second-order contributions in the context of other contributions to the thermodynamics, both entropic and enthalpic.

The present methodology opens new possibilities for exploring the thermodynamic properties of water at molecular surfaces and thus has potential to deepen our understanding of molecular recognition and to find practical application in various aspects of molecular design, including drug design. There is also room for methodological improvement. For this translational term, it will be of interest to seek enhanced convergence with respect to both grid spacing and simulation duration. In addition, it will be of interest to extend these studies to the entropic consequences of water–water orientational correlations, which may be substantial, due to the cooperative formation of various different hydrogen-bonded configurations of water molecules. The use of strong numerical methods, like nearest-neighbor entropy estimation, combined with growing computer power, should make for continued progress along these lines.

## APPENDIX

We examined the normalized second order translational entropy,  $TS_{\text{ww}}$ , of pure TIP3P water, computed using various grid spacings



with both the histogram (left) and nearest neighbor (right) methods. These calculations used a  $K$  grid of dimensions  $1.5 \text{ \AA} \times 1.5 \text{ \AA} \times 1.5 \text{ \AA}$  and  $L_k$  grids of dimensions  $16 \text{ \AA} \times 16 \text{ \AA} \times 16 \text{ \AA}$ , which corresponds to a minimum cutoff distance of  $8 \text{ \AA}$  for the effect of a water in voxel  $k$  on the surrounding water density profile. The graphs show that, for fine grid spacings, such as  $0.15 \text{ \AA}$ , the NN method has a clear advantage over the histogram method, in terms of the amount of sampling required for a given degree of convergence. Even at a grid spacing of  $0.25 \text{ \AA}$ , the histogram result is far from the reference RDF result after extensive sampling. Note that the apparent agreement of the histogram results with the reference RDF results for grid spacings of  $0.5$  and  $0.75 \text{ \AA}$  are misleading, due to the large range of the  $y$  axis; see Figure 3 in the main text.

## AUTHOR INFORMATION

### Corresponding Author

\*Voice: 858-822-0622. E-mail: [mgilson@ucsd.edu](mailto:mgilson@ucsd.edu).

### Notes

The authors declare the following competing financial interest(s): M.K.G. has an equity interest in and is a cofounder and scientific advisor of VeraChem LLC.

## ACKNOWLEDGMENTS

This publication was made possible in part by grant GM100946 from the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

## REFERENCES

- (1) Kataev, E. A.; Müller, C. Recent Advances in Molecular Recognition in Water: Artificial Receptors and Supramolecular Catalysis. *Tetrahedron* **2014**, *70* (2), 137–167.
- (2) Levy, Y.; Onuchic, J. N. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35* (1), 389–415.
- (3) García-Sosa, A. T. Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-Bound, Bridging Water Molecules and Their Effects and Consequences on Molecular Design Strategies. *J. Chem. Inf. Model.* **2013**, *53* (6), 1388–1405.
- (4) Ladbury, J. Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and Its Potential Application to Drug Design. *Chem. Biol.* **1996**, *3* (12), 973–980.
- (5) Bellissent-Funel, M.-C. *Hydration Processes in Biology: Theoretical and Experimental Approaches*; IOS Press, 1999.
- (6) Mancera, R. Molecular Modeling of Hydration in Drug Design. *Curr. Opin. Drug Discovery Dev.* **2007**, *10* (3), 275–280.
- (7) de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The Role of Water Molecules in Computational Drug Design. *Curr. Top. Med. Chem.* **2010**, *10* (1), 55–66.
- (8) Li, Z.; Lazaridis, T. Water at Biomolecular Binding Interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9* (5), 573.
- (9) Poornima, C. S.; Dean, P. M. Hydration in Drug Design. 2. Influence of Local Site Surface Shape on Water Binding. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 513–520.
- (10) Baron, R.; Setny, P.; Andrew McCammon, J. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* **2010**, *132* (34), 12091–12097.
- (11) Thirumalai, D.; Reddy, G.; Straub, J. E. Role of Water in Protein Aggregation and Amyloid Polymorphism. *Acc. Chem. Res.* **2012**, *45* (1), 83–92.
- (12) Cuzzini, P.; Fornabai, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free Energy of Ligand Binding to Protein: Evaluation of the Contribution of Water Molecules by Computational Methods. *Curr. Med. Chem.* **2004**, *11* (23), 3093–3118.
- (13) Gilson, M. K.; Honig, B. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies and Conformational Analysis. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7–18.
- (14) Gilson, M. K.; Honig, B. The Inclusion of Electrostatic Hydration Energies in Molecular Mechanics Calculations. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 5–20.
- (15) Sitkoff, D.; Sharp, K.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98* (7), 1978–1988.
- (16) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.
- (17) Dominy, B. N.; Brooks, C. L., III Development of a Generalized Born Model Parametrization for Proteins and Nucleic Acids. *J. Phys. Chem. B* **1999**, *103*, 3765–3773.
- (18) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002**, *23* (5), 517–529.
- (19) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Genet.* **2004**, *55* (2), 383–394.
- (20) Gallicchio, E.; Paris, K.; Levy, R. M. The AGBNP2 Implicit Solvation Model. *J. Chem. Theory Comput.* **2009**, *5* (9), 2544–2564.
- (21) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical Electrostatics for Biomolecules: Beyond the Generalized Born Approximation. *J. Chem. Phys.* **2006**, *124*, 124902.
- (22) Berendsen, H.; Postma, J.; van Gunsteren, W.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Reidel Publishing Co.: Dordrecht, pp 331–342.
- (23) 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.* **1983**, *79* (2), 926.
- (24) Jorgensen, W. L.; Madura, J. D. Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water. *Mol. Phys.* **1985**, *56* (6), 1381–1392.
- (25) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
- (26) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933–5947.
- (27) Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T.; Horn, H. W. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665–9678.
- (28) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123* (23), 234505–234505-12.
- (29) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863–3871.
- (30) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (3), 808–813.
- (31) Englert, L.; Biela, A.; Zayed, M.; Heine, A.; Hangauer, D.; Klebe, G. Displacement of Disordered Water Molecules from Hydrophobic Pocket Creates Enthalpic Signature: Binding of Phosphonamidate to the S1'-Pocket of Thermolysin. *Biochim. Biophys. Acta, Gen. Subj.* **2010**, *1800* (11), 1192–1202.
- (32) Setny, P.; Baron, R.; McCammon, J. A. How Can Hydrophobic Association Be Enthalpy Driven? *J. Chem. Theory Comput.* **2010**, *6* (9), 2866–2871.
- (33) Biela, A.; Nasief, N. N.; Betz, M.; Heine, A.; Hangauer, D.; Klebe, G. Dissecting the Hydrophobic Effect on the Molecular Level: The Role of Water, Enthalpy, and Entropy in Ligand Binding to Thermolysin. *Angew. Chem., Int. Ed.* **2013**, *52* (6), 1822–1828.



- (34) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10* (7), 2769–2780.
- (35) Frank, H. S.; Evans, M. W. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.* **1945**, *13* (11), 507.
- (36) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959**, *14*, 1–63.
- (37) Eisenberg, D.; Kauzmann, W. *The Structure and Properties of Water*, 1st ed.; Oxford University Press: New York, 1969.
- (38) Green, H. S. *The Molecular Theory of Fluids*; North-Holland Publishing Company: Amsterdam, 1952.
- (39) Nettleton, R. E.; Green, M. S. Expression in Terms of Molecular Distribution Functions for the Entropy Density in an Infinite System. *J. Chem. Phys.* **1958**, *29* (6), 1365.
- (40) Wallace, D. C. On the Role of Density Fluctuations in the Entropy of a Fluid. *J. Chem. Phys.* **1987**, *87* (4), 2282.
- (41) Baranyai, A.; Evans, D. J. Direct Entropy Calculation from Computer Simulation of Liquids. *Phys. Rev. A: At, Mol, Opt. Phys.* **1989**, *40* (7), 3817.
- (42) Morita, T.; Hiroike, K. A New Approach to the Theory of Classical Fluids. III. *Prog. Theor. Phys.* **1961**, *25* (4), 537–578.
- (43) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102* (18), 3531–3541.
- (44) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102* (18), 3542–3550.
- (45) 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.* **2008**, *130* (9), 2817–2831.
- (46) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. *Methods Mol. Biol.* **2012**, *819*, 393–404.
- (47) Huggins, D. J. Application of Inhomogeneous Fluid Solvation Theory to Model the Distribution and Thermodynamics of Water Molecules around Biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14* (43), 15106.
- (48) Huggins, D. J. Benchmarking the Thermodynamic Analysis of Water Molecules around a Model Beta Sheet. *J. Comput. Chem.* **2012**, *33* (15), 1383–1392.
- (49) Haider, K.; Huggins, D. J. Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules. *J. Chem. Inf. Model.* **2013**, *53* (10), 2571–2586.
- (50) Nguyen, C.; Gilson, M. K.; Young, T. Structure and Thermodynamics of Molecular Hydration via Grid Inhomogeneous Solvation Theory. *arXiv:1108.4876*, **2011**.
- (51) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137* (4), 044101.
- (52) Raman, E. P.; MacKerell, A. D., Jr Rapid Estimation of Hydration Thermodynamics of Macromolecular Regions. *J. Chem. Phys.* **2013**, *139* (5), 055105.
- (53) Abel, R.; Wang, L.; Friesner, R. A.; Berne, B. J. A Displaced-Solvent Functional Analysis of Model Hydrophobic Enclosures. *J. Chem. Theory Comput.* **2010**, *6* (9), 2924–2934.
- (54) Lazaridis, T.; Karplus, M. Orientational Correlations and Entropy in Liquid Water. *J. Chem. Phys.* **1996**, *105* (10), 4294.
- (55) Zielkiewicz, J. Structural Properties of Water: Comparison of the SPC, SPCE, TIP4P, and TIPSP Models of Water. *J. Chem. Phys.* **2005**, *123* (10), 104501.
- (56) Attard, P.; Jepps, O. G.; Marcelja, S. Information Content of Signals Using Correlation Function Expansions of the Entropy. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1997**, *56*, 4052.
- (57) Matsuda, H. Physical Nature of Higher-Order Mutual Information: Intrinsic Correlations and Frustration. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, *62* (3), 3096.
- (58) Killian, B. J.; Yundenfreund Kravitz, J.; Gilson, M. K. Extraction of Configurational Entropy from Molecular Simulations via an Expansion Approximation. *J. Chem. Phys.* **2007**, *127*, 024107.
- (59) Killian, B. J.; Kravitz, J. Y.; Somani, S.; Dasgupta, P.; Pang, Y.-P.; Gilson, M. K. Configurational Entropy in Protein-Peptide Binding. Computational Study of TSG101 UEV Domain with an HIV-Derived PTAP Motif. *J. Mol. Biol.* **2009**, *389*, 315.
- (60) Numata, J.; Knapp, E.-W. Balanced and Bias-Corrected Computation of Conformational Entropy Differences for Molecular Trajectories. *J. Chem. Theory Comput.* **2012**, *8* (4), 1235–1245.
- (61) Fenley, A. T.; Muddana, H. S.; Gilson, M. K. Entropy–enthalpy Transduction Caused by Conformational Shifts Can Obscure the Forces Driving Protein–ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (49), 20006–20011.
- (62) Huggins, D. J. Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophys. J.* **2015**, *108* (4), 928–936.
- (63) Velez-Vega, C.; McKay, D. J. J.; Kurtzman, T.; Aravamuthan, V.; Pearlstein, R. A.; Duca, J. S. Estimation of Solvation Entropy and Enthalpy via Analysis of Water Oxygen–Hydrogen Correlations. *J. Chem. Theory Comput.* **2015**, *11*, 5090.
- (64) Lazaridis, T. Solvent Reorganization Energy and Entropy in Hydrophobic Hydration. *J. Phys. Chem. B* **2000**, *104* (20), 4964–4979.
- (65) Huggins, D. J.; Payne, M. C. Assessing the Accuracy of Inhomogeneous Fluid Solvation Theory in Predicting Hydration Free Energies of Simple Solutes. *J. Phys. Chem. B* **2013**, *117* (27), 8232–8244.
- (66) Li, Z.; Lazaridis, T. Thermodynamics of Buried Water Clusters at a Protein-Ligand Binding Interface. *J. Phys. Chem. B* **2006**, *110* (3), 1464–1475.
- (67) Huggins, D. J. Estimating Translational and Orientational Entropies Using the  $k$ -Nearest Neighbors Algorithm. *J. Chem. Theory Comput.* **2014**, *10* (9), 3617–3625.
- (68) Singh, H.; Misra, S.; Hnizdo, V.; Fedorowicz, A.; Demchuk, E. Nearest Neighbor Estimates of Entropy. *Am. J. Math. Management Sci.* **2003**, *23*, 301–321.
- (69) Zhou, H.; Gilson, M. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109* (9), 4092–4107.
- (70) Arya, S.; Mount, D. M.; Netanyahu, N. S.; Silverman, R.; Wu, A. An Optimal Algorithm for Approximate Nearest Neighbor Searching. In *Proceedings of the fifth annual ACM-SIAM symposium on Discrete algorithms*; SODA '94, Society for Industrial and Applied Mathematics: Philadelphia, PA, 1994; pp 573–582.
- (71) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Hayik, S.; Roitberg, A. E.; Seabra, G.; Swails, J. M.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California, San Francisco, 2012.
- (72) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.
- (73) Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
- (74) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge-Restraints for Deriving Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (75) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Goetz, H. G. A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.;

Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 2015*; University of California: San Francisco, CA, 2015.

(76) Adler, M.; Davey, D. D.; Phillips, G. B.; Kim, S. H.; Jancarik, J.; Rumennik, G.; Light, D. R.; Whitlow, M. Preparation, Characterization, and the Crystal Structure of the Inhibitor ZK-807834 (CI-1031) Complexed with Factor Xa. *Biochemistry* **2000**, 39 (41), 12534–12542.

(77) *AMBER 11*; *AmberTools 1.5*; University of California: San Francisco, CA, 2010.

(78) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, 65 (3), 712–725.

(79) Wang, L.; Abel, R.; Friesner, R. A.; Berne, B. J. Thermodynamic Properties of Liquid Water: An Application of a Nonparametric Approach to Computing the Entropy of a Neat Fluid. *J. Chem. Theory Comput.* **2009**, 5 (6), 1462–1473.

(80) Humphrey, W.; Dalke, A.; Schulten, K. VMD -Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, 14, 33–38.

(81) Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. Atomic Structure of Acetylcholinesterase from Torpedo Californica: A Prototypic Acetylcholine-Binding Protein. *Science* **1991**, 253, 872–879.