#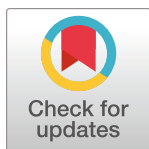 A molecular reconstruction approach to site-based 3D-RISM and comparison to GIST hydration thermodynamic maps in an enzyme active site

Crystal Nguyen[1], Takeshi Yamazaki[2¤], Andriy Kovalenko[3,4], David A. Case[5], Michael K. Gilson[1]*, Tom Kurtzman[6]*, Tyler Luchko[7]*

1 Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, United States of America, **2** Vancouver Prostate Centre, Vancouver, British Columbia, Canada, **3** National Institute for Nanotechnology, National Research Council of Canada, Edmonton, Alberta, Canada, **4** Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada, **5** Department of Chemistry, Lehman College, The City University of New York, Bronx, New York, United States of America, **6** Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey, United States of America, **7** Department of Physics and Astronomy, California State University, Northridge, California, United States of America

¤ Current address: 1QB Information Technologies (1QBit), Vancouver, British Columbia, Canada
* tluchko@csun.edu (TL); simpleliquid@gmail.com (TK); mgilson@ucsd.edu (MKG)

Abstract

Computed, high-resolution, spatial distributions of solvation energy and entropy can provide detailed information about the role of water in molecular recognition. While grid inhomogeneous solvation theory (GIST) provides rigorous, detailed thermodynamic information from explicit solvent molecular dynamics simulations, recent developments in the 3D reference interaction site model (3D-RISM) theory allow many of the same quantities to be calculated in a fraction of the time. However, 3D-RISM produces atomic-site, rather than molecular, density distributions, which are difficult to extract physical meaning from. To overcome this difficulty, we introduce a method to reconstruct molecular density distributions from atomic-site density distributions. Furthermore, we assess the quality of the resulting solvation thermodynamics density distributions by analyzing the binding site of coagulation Factor Xa with both GIST and 3D-RISM. We find good qualitative agreement between the methods for oxygen and hydrogen densities as well as direct solute-solvent energetic interactions. However, 3D-RISM predicts lower energetic and entropic penalties for moving water from the bulk to the binding site.

Introduction

Water is intimately involved in the interactions of small molecule drugs with their protein targets. Implicit solvent models offer one approach to gain insight into the role of water in binding, but they do not capture all the molecular detail thought to be relevant in the confined,

heterogeneous setting of a protein binding site. Explicit solvent models capture more detail, but do not directly provide insight into the role of water in a given binding site that can be used to guide ligand design. The last decade has seen a proliferation of computational tools and methods that address this gap by describing the hydration of protein binding sites and accounting in increasing detail for the role of water in the binding of small molecule drugs to their protein targets.

Two broad approaches, the Ornstein-Zernike (OZ) equation[1] and inhomogeneous fluid solvation theory (IST)[2–5], derive from the statistical mechanics of liquid state systems and have found application in academia and the pharmaceutical industry. Both OZ and IST are distribution function theories, in that they treat the fluid as a statistical distribution of particle or molecular densities. Although the formulation of each method is formally exact, computational implementation of these methods requires the use of approximations. The result is a mapping of local solvation thermodynamic properties, such as densities, energies and entropies.

In IST, explicit solvent Monte Carlo (MC) or molecular dynamics (MD) simulations are used to sample configurations from the Boltzmann distribution, and thermodynamic quantities are calculated from these samples. Although the molecular and atomic distribution functions generated from the simulations should asymptotically approach the correct results for the force field used, finite sampling means that IST results will have statistical uncertainties. Also, due to the limited sampling, the expression for entropy, which is represented in terms of single and multi-body molecular correlation functions in IST, must be truncated at low order (generally 1st, but occasionally 2nd order[6–9]), leading to incomplete evaluation of the entropy. The IST approach has been implemented in at least two ways, one in which water properties are evaluated only at highly occupied hydration sites, as in WaterMap[10,11] and STOW[12]; and another in which water properties are evaluated for all voxels of a three-dimensional grid in the region of interest, as in GIST[13,14] and SSTMap[15].

The 3-dimensional reference interaction site model (3D-RISM)[16] is a widely used approximation to the molecular OZ equation. Like GIST, this approach also utilizes a simulation force field and water model, but it does not require sampling from the Boltzmann distribution. Instead, distribution functions for the system of interest are generated from an interaction-site approximation to the OZ equation, which is an exact relationship for the model system expressed in terms of the direct and indirect correlations between pairs of particles. As a consequence, 3D-RISM is much faster computationally than IST approaches, and it does not have statistically uncertainties due to finite sampling. On the other hand, as with all theories based on the OZ equation, solving the 3D-RISM equation requires a second equation, known as the closure relationship[17], which is necessarily approximate. In particular, the full closure relation contains an infinite series of many-body correction functions, and it must, in practice, be truncated at low order. As a consequence, the atomic distribution functions provided by the method are approximations. Exact expressions for the local thermodynamic properties consistent with the approximations are then applied to the distribution functions to yield the relevant entropies, energies, and densities. As further considered below, these quantities may be mapped to a three-dimensional spatial grid, much as done by GIST and SSTMAP in the IST approach (above). The error associated with these thermodynamic quantities traces to any error in the force field used, and to the approximations used to compute the atomic distributions. In recent years, corrections to errors in thermodynamic quantities have been developed that are applied after the solvent density distributions have been calculated, leading to estimates of solvation free energies with accuracies as good as explicit solvent simulations[18–29].

Given the similarities between grid-based IST implementations and 3D-RISM, it is clearly of interest to compare the two approaches. However, there has until now been a major obstacle to making such a comparison. Whereas GIST yields molecular distribution functions, 3D-RISM provides separate spatial distributions of the hydrogen and oxygen atoms of the water molecules, and of their contributions to the thermodynamics. When the oxygen and hydrogen contributions to the thermodynamic density distributions are additively combined, the resulting function has large variations in amplitude over very small distances. This causes problems when attempting to estimate the free energy of a ligand displacing water from a binding site into bulk; a slight shift in the position of a ligand may produce very different predictions of the energetic cost of displacing the solvent. This problem has been addressed by either averaging the contributions of solvent over coarser grids[30] or collapsing them on to solvation sites[31], but both approaches sacrifice the detailed information contained in the density distributions.

Here we report a new method to combine the oxygen and hydrogen contributions from 3D-RISM, based on the internal structure of water, into molecular density distributions and directly compare them to GIST results. Both methods are applied to the same conformation of the same protein, Factor Xa, using matching grid geometries and essentially the same force field. The Methods section provides a brief review of GIST and 3D-RISM, then explains the molecular reconstruction approach for 3D-RISM, as well as a method for mapping the partial molar volume correction (above) to the 3D-RISM grids. The Result section analyzes and compares GIST and 3D-RISM results for the spatial distributions of water molecules, energy, and entropy, within the binding site of Factor Xa. The Discussion session summarizes key points and considerations directions for future analysis of these two methods.

Methods

Grid inhomogeneous solvation theory

Grid inhomogeneous solvation theory (GIST), a grid based implementation of Inhomogeneous Fluid Solvation Theory, maps out thermodynamics densities on a high-resolution grid[13,14]. Software implementing GIST is available in the SSTMap[15] and AmberTools cpptraj[32,33] software packages. Inhomogeneous fluid solvation theory and the GIST grid implementation of it are detailed elsewhere, but we give a brief outline here for the reader's convenience.

GIST utilizes system configurations sampled from the Boltzmann distribution, with molecular dynamics or Monte Carlo simulations, to compute spatial distributions of entropy and energy on a three-dimensional grid composed of cubic voxels. We focus here on the single-molecule (1st order) entropy contributions and the 1st and 2nd order energy contributions. The equation for 1st order approximation to the entropy is given by

$$S_{sw} = -\frac{k_b \rho^0}{8\pi^2} \int g(\mathbf{r}, \omega) \ln g(\mathbf{r}, \omega) d\omega d\mathbf{r} \qquad (1)$$

where k_b is the gas constant, and $g(\mathbf{r},\omega)$ is the solute-water pair distribution function (PDF); i.e., the probability density of water at position $\mathbf{r}$ with an orientation of ω. Using the relationship:

$$g(\mathbf{r}, \omega) = g(\omega|\mathbf{r})g(\mathbf{r}) \qquad (2)$$

and the approximation that the orientational distribution is independent of the position within each cubic grid point, or voxel, the entropy of each voxel k with volume V_k can be separated

into translational and orientational terms.

$$\Delta S_{\text{sw},k}^{\text{trans}} = -k_b \rho^0 \int_{V_k} g(\mathbf{r}) \ln g(\mathbf{r}) d\mathbf{r} \tag{3}$$

$$\Delta S_{\text{sw},k}^{\text{orient}} = \rho^0 \int_{V_k} g(\mathbf{r}) S^\omega(\mathbf{r}) d\mathbf{r} \tag{4}$$

$$S^\omega(\mathbf{r}) = \frac{-k_b}{8\pi^2} \int d\mathbf{r} \int g(\omega|\mathbf{r}) \ln g(\omega|\mathbf{r}) d\omega \tag{5}$$

Here $g(\mathbf{r}, \omega)$ is the density of water at position $\mathbf{r}$, $g(\mathbf{r}|\omega)$ is the conditional density probability of finding a water with an orientation ω given a water at position $\mathbf{r}$, and the subscript "sw" denotes the solute-solvent correlations. All the density probabilities in these equations are normalized by the bulk water density, ρ^0, such as $g(\mathbf{r}) = \rho(\mathbf{r})/\rho^0$. In practice, the translational and orientational integrals may be calculated by histogram and/or nearest neighbor method [13,34,35].

The 1$^{\text{st}}$ order (solute-solvent) and 2$^{\text{nd}}$ order (solvent-solvent) energies are given by:

$$\Delta E_{\text{sw}}(\mathbf{r}_k) = \int_{V_k} \Delta E_{\text{sw}}(\mathbf{r}) d\mathbf{r} \tag{6}$$

$$\Delta E_{\text{sw}}(\mathbf{r}) = \frac{1}{8\pi^2} \int g_{\text{sw}}(\omega|\mathbf{r}) U_{\text{sw}}(\mathbf{r}, \omega) d\omega \tag{7}$$

$$\Delta E_{\text{ww}}(\mathbf{r}_k) = \rho^0 \int_{V_k} g_{\text{sw}}(\mathbf{r}) \Delta E_{\text{ww}}(\mathbf{r}) d\mathbf{r} \tag{8}$$

$$\begin{aligned}
\Delta E_{\text{ww}}(\mathbf{r}) = \left(\frac{1}{8\pi^2}\right)^2 \rho^0 \int g_{\text{sw}}(\omega|\mathbf{r}) \\
[g_{\text{sw}}(\mathbf{r}', \omega') g_{\text{ww}}(\mathbf{r}, \omega, \mathbf{r}', \omega') - g^0_{\text{ww}}(\mathbf{r}, \omega, \mathbf{r}', \omega')] \\
U_{\text{ww}}(\mathbf{r}, \omega, \mathbf{r}', \omega') d\omega d\mathbf{r}' d\omega
\end{aligned} \tag{9}$$

Here $g_{\text{sw}}(\mathbf{r}) = g(\mathbf{r})$, $g_{\text{ww}}(\mathbf{r},\omega,\mathbf{r}',\omega')$ is the pairwise distribution function between two water molecules with coordinates $(\mathbf{r},\omega)$ and $(\mathbf{r}',\omega')$ in the solute frame of reference, $g^0_{\text{ww}}(\mathbf{r},\omega,\mathbf{r}',\omega')$ is the corresponding distribution in bulk water, $U_{\text{sw}}(\mathbf{r},\omega)$ is the solute-solvent interaction potential, and $U_{\text{ww}}(\mathbf{r},\omega,\mathbf{r}',\omega')$ is the solvent-solvent potential.

The GIST software computes these energy and entropy terms for each voxel on the grid and produces data files for visualization and further analysis. When comparing GIST with 3D-RISM, it is important to recognize that GIST assigns properties (energy and entropy) to a voxel based on the water molecules whose oxygen atoms are located in it. Thus, for example, the energy of a voxel includes the interactions of the hydrogens of these water molecules even though the hydrogens are not located in the same voxel. In contrast, 3D-RISM by default assigns voxel energies on an atomic, rather than a molecular basis. Thus, for example, if a voxel contains a water oxygen but not the water's hydrogens, only the oxygen's interactions are assigned to it.

3D-RISM

Details of 3D-RISM[16,36–38], its implementation in AmberTools[39,40], and analytic temperature derivatives[41,42], in particular, three-dimensional generalization of the analytical derivatives[43] can be found elsewhere. Here we briefly summarize the 3D-RISM theory and the use of temperature derivatives to calculate solvation energy and entropy contributions to the total solvation free energy.

The 3D-RISM equation is

$$h_\alpha(\mathbf{r}) = \sum_\gamma \int c_\gamma(\mathbf{r} - \mathbf{r}')\chi_{\alpha\gamma}(r')d\mathbf{r}' \tag{10}$$

where $h_\alpha(\mathbf{r})$ is the 3D total correlation function (TCF) of solvent site α around the solute and $c_\alpha(\mathbf{r})$ is the 3D direct correlation function (DCF).

$$\chi_{\alpha\gamma}(r) = \omega_{\alpha\gamma}(r) + \omega_{\alpha\gamma}(r)\rho_\gamma h_{\alpha\gamma}(r) \tag{11}$$

is the site-site bulk solvent susceptibility obtained from dielectrically consistent RISM (DRISM), ρ_γ is the site number-density,

$$\omega_{\alpha\gamma}(\mathbf{r}) = \sum_{\lambda=1}^{M_\gamma} \delta_{\alpha\gamma}\delta(\mathbf{r}) + \frac{(1 - \delta_{\alpha\gamma})\delta(r - L_{\alpha\lambda})}{4\pi L_{\alpha\lambda}^2} \tag{12}$$

is the intramolecular correlation function, M_γ is the multiplicity of site γ ($M_{\mathrm{o}} = 1$ and $M_{\mathrm{H}} = 2$), $L_{\alpha\gamma}$ is the distance constraint between site α and the λ position of site γ of the rigid solvent molecule, and $\delta_{\alpha\gamma}$ and $\delta(\mathbf{r})$ are the Kronecker and Dirac delta functions.

As $h_\alpha(\mathbf{r})$ and $c_\alpha(\mathbf{r})$ are unknown functions, a closure equation is required in order to obtain iterative solutions. We will make use of the partial series expansion of order-n (PSE-n)[44] approximate closure equation,

$$g_\alpha(\mathbf{r}) = \begin{cases} \exp(t_\alpha^*(\mathbf{r})) & t_\alpha^*(\mathbf{r}) < 0 \\ \sum_{i=0}^n \dfrac{(t_\alpha^*(\mathbf{r}))^i}{i!} & t_\alpha^*(\mathbf{r}) \geq 0 \end{cases} \tag{13}$$

With

$$t^*_\alpha(\mathbf{r}) = -\beta u_\alpha(\mathbf{r}) + h_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r}). \tag{14}$$

Here $\beta = k_b T$, where T is the temperature and k_b is Boltzmann's constant, $u_\alpha(\mathbf{r})$ is the potential energy function between solvent site α and the solute, and $g_\alpha(\mathbf{r}) = h_\alpha(\mathbf{r})+1$ is the water-solute PDF. When $n = 1$ we obtain the Kovalenko-Hirata (KH) closure[37] and when $n = \infty$ the hypernetted-chain equation (HNC)[45] is recovered. For a given closure it is possible to obtain the excess chemical potential, or solvation free energy, of the solute. For PSE-n this is

$$\Delta\mu^{\mathrm{PSE}-n} = \sum_\gamma \int \Delta\mu_\gamma^{\mathrm{PSE}-n}{}_\alpha(\mathbf{r})d\mathbf{r} \tag{15}$$

$$\Delta\mu^{\mathrm{PSE}-n}{}_\gamma(\mathbf{r}) = k_b T\rho_\gamma \frac{h_\gamma^2(\mathbf{r})}{2} - c_\gamma(\mathbf{r}) - \frac{h_\gamma(\mathbf{r})c_\gamma(\mathbf{r})}{2} - \frac{(t^*_\gamma(\mathbf{r}))^{n+1}}{(n+1)!}\Theta(h_\gamma(\mathbf{r})) \tag{16}$$

where Θ is the Heaviside function and $\Delta\mu^{\mathrm{PSE}-n}{}_\gamma(\mathbf{r})$ is the local excess chemical potential density. For any closure with a path independent expression for the solvation free energy[41], it is

possible to decompose Eq (16) into energy, E, and entropy, S, contributions,

$$\Delta\mu(\mathbf{r}) = \Delta E_{T,V}(\mathbf{r}) - T\Delta S_{T,V}(\mathbf{r}). \tag{17}$$

Omitting the 'T,V' notation and taking the isochoric temperature derivative, $\delta_T \equiv T(\partial/\partial T)_\rho$, of the solvation free energy, we have for the PSE-n closure

$$
\begin{aligned}
-T\Delta S_\gamma^{\mathrm{PSE}-n}(\mathbf{r}) &= T\left(\frac{\partial\Delta\mu_\gamma^{\mathrm{PSE}-n}(\mathbf{r})}{\partial}\right) = \delta\Delta\mu_\gamma^{\mathrm{PSE}-n}(\mathbf{r}) \\
&= \Delta\mu_\gamma^{\mathrm{PSE}-n}(\mathbf{r}) + k_b T\rho_\gamma \left\{ h_\gamma(\mathbf{r})\delta_T h_\gamma(\mathbf{r}) - \delta_T c_\gamma(\mathbf{r})\frac{(t_\gamma^*(\mathbf{r}))^n}{n!} \right. \\
&\qquad - \frac{1}{2}[[\delta_T h_\gamma(\mathbf{r})]c_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r})\delta_T c_\gamma(\mathbf{r})] \\
&\qquad \left. - \frac{(t_\gamma^*(\mathbf{r}))^n}{n!}\delta_T t_\gamma^*(\mathbf{r})\Theta(h_\gamma(\mathbf{r})) \right\}
\end{aligned}
\tag{18}
$$

Independent of the closure, the solvation energy density is then

$$\Delta E_\gamma(\mathbf{r}) = \Delta\mu_\gamma(\mathbf{r}) - \delta_T\Delta\mu_\gamma(\mathbf{r}). \tag{19}$$

As with GIST, the solvation energy can be broken down into solute-water and water-water contributions. The solute-water energy density is given by

$$\Delta E_{\mathrm{sw},\gamma}(\mathbf{r}) = \rho_\gamma g_\gamma(\mathbf{r})u_\gamma(\mathbf{r}) \tag{20}$$

and the water-water energy density is then

$$\Delta E_{\mathrm{ww},\gamma}(\mathbf{r}) = \Delta E_\gamma(\mathbf{r}) - \Delta E_{\mathrm{sw},\gamma}(\mathbf{r}). \tag{21}$$

However, since 3D-RISM does not explicitly consider the orientation of the water molecules, it is not possible to separate translational and orientational contributions.

Molecular reconstruction

As noted above, in 3D-RISM, the contributions to the thermodynamic quantities from a single water molecule's hydrogens and oxygen may fall into different grid voxels, whereas in GIST, all interactions of the water molecule are associated with the voxel occupied by the oxygen atom. We now describe a method of generating a molecular mapping of properties, similar to that of GIST, from the grids provided by 3D-RISM.

Consider an arbitrary thermodynamic quantity $A(\mathbf{r})$, such as $\Delta E_{\mathrm{sw}}(\mathbf{r})$ or $-T\Delta S(\mathbf{r})$, composed of the sum of contributions from all solvent sites, e.g., $A(\mathbf{r}) = \sum_\alpha A_\alpha(\mathbf{r})$. A molecular distribution can be approximately reconstructed around a central site, α, e.g., oxygen, using the structural information in the intramolecular correlation function, Eq (12),

$$A(\mathbf{r}) \approx A_\alpha(\mathbf{r}) + g_\alpha(\mathbf{r})\sum_{\gamma\neq\alpha}\omega_{\alpha\gamma}(\mathbf{r}) * A_\gamma(\mathbf{r}). \tag{22}$$

Here, the convolution, $*$, between $\omega_{\alpha\gamma}(\mathbf{r})$ and $A_\gamma(\mathbf{r})$ has the effect of placing $A_\gamma(\mathbf{r})$ a distance $L_{\alpha\gamma}$ away in all directions. This is then weighted by the number density distribution of site α, $g_\alpha(\mathbf{r})$. The final expression is the superposition of the thermodynamic quantity density distribution due to site α and that of the other sites mapped onto the number density distribution of α.

While it is possible to select any solvent site as α, for water, it is natural to use oxygen; e.g.,

$$A(\mathbf{r}) \approx A_{\mathrm{O}}(\mathbf{r}) + g_{\mathrm{O}}(\mathbf{r})(\omega_{\mathrm{OH}}(\mathbf{r}) * A_{\mathrm{H}}(\mathbf{r})). \tag{23}$$

Because this reconstruction uses only the local information contained in $\omega_{\alpha\gamma}(\mathbf{r})$ it does not properly include non-local contributions due to the DCF. For example, contributions from the excluded volume of the solute, where the DCF contributes most, are set to zero for the contributions from the γ sites (e.g., hydrogen) due to the weighting by the PDF. For consistency, contributions in the excluded volume due to the α site (e.g., oxygen) are also set to zero. Contributions to $A(\mathbf{r})$ from the excluded volume region are therefore neglected, and contributes to errors in $\Delta E_{ww}(\mathbf{r})$, $\Delta E(\mathbf{r})$ and $-T \Delta S(\mathbf{r})$ but not $\Delta E_{sw}(\mathbf{r})$. However, we find that when these quantities are corrected (see below) the contribution from the excluded volume is small, making the approximation valid.

Partial molar volume corrections

When integrated over all space, Eqs (16), (18), and (19) provide the excess chemical potential, energy, and entropy, respectively. However, these values have significant errors when compared to values from experiment. Several corrections have been developed to mitigate these issues for the excess chemical potential[18–21]. Recently, we demonstrated that several of these methods provide good agreement with excess energies and entropies from experiment as well[24].

While these corrections are known to work well for the integrated quantities of Eqs (17), (18) and (19), we wish to test if they also have an impact on the spatial distribution of these quantities. To test this, we apply the Universal Correction with temperature dependence (UCT)[24] to the thermodynamic density distributions,

$$\Delta E^{\mathrm{UCT}}(\mathbf{r}) = \left(\sum_{\gamma} \Delta E_{\gamma}^{\mathrm{PSE}-n}(\mathbf{r}) \right) - a(\nu(\mathbf{r}) - \delta_T \nu(\mathbf{r})) - T a_1 \nu(\mathbf{r}) + b_0 \tag{24}$$

$$-T\Delta S^{\mathrm{UCT}}(\mathbf{r}) = \left(\sum_{\gamma} -T\Delta S_{\gamma}^{\mathrm{PSE}-n}(\mathbf{r}) \right) + a_1 T\nu(\mathbf{r}) + a\delta_T \nu(\mathbf{r}) + b_1 T \tag{25}$$

where 'RISM' identifies the uncorrected quantities from Eqs (18) and (19),

$$\nu = k_b T \chi_T \left(1 - \sum_{\gamma} \int \rho_{\gamma} c_{\gamma}(\mathbf{r}) d\mathbf{r} \right) \tag{26}$$

is the partial molar volume, and χ_T is the isothermal compressibility calculated from dielectrically consistent RISM (DRISM)[46]. Four parameters, a_0, a_1, b_0, and b_1, are fit to experiment are related by

$$
\begin{aligned}
a &= a_1 T + a_0 \\
b &= b_1 T + b_0.
\end{aligned}
\tag{27}
$$

As will be discussed below we do not attempt to perform a molecular reconstruction on Eqs (24) and (25) because the corrections provide no observable benefit outside the excluded volume of the solute and the corrections make the contribution from the excluded volume close to zero.

GIST analysis

The GIST calculations used a 20.5 x 20.5 x 22.5 Å grid composed of cubic voxels with side lengths 0.5 Angstroms, and centered in the binding site of the protein Factor Xa. GIST results were computed based on a 100 ns MD simulation of Factor Xa, as now detailed.

The crystal structure of Factor Xa (pdbid: 1FJS) was prepared with the ligand Z34 removed in AmberTools software tleap using ff99SB[47]. TIP3P[48] water molecules were added to form a solvent medium with a minimum distance of 10 Å from the protein surface to the edge of the water box. Four disulfide bonds were connected. All crystal ions (1 Ca^{2+} and 1 Cl^-) were kept, resulting in a +2 net charge for the system. The simulation box had dimensions of 66.5 x 72.2 x 60.9 Å with a total of 29338 atoms and 8557 water molecules.

All molecular dynamics simulations were run in Amber14 using pmemd.cuda on a single GPU. Computer time was 52 hours for 100ns simulation with a time step of 2 fs. A 9 Å cutoff was applied to all non-bonded interactions while PME was used to account for long-range interactions. SHAKE was used to constrain all covalent bonds involving hydrogen atoms during the simulations. Cartesian restraint is employed throughout the simulation with a force constant of 100 kcal/mol/$Å^2$.

The system was first relaxed by using two minimization procedures. Each was done with the steepest descent algorithm for the first 1500 time steps followed by the conjugate gradient method with a maximum of 2000 time steps. The first minimization was performed on the water alone by keeping all the protein's atoms restrained. The second allowed the water and the protein's heavy atoms to relax. The minimized system was heated gradually, at constant volume, with an increment of 50K for 20ps from zero to 300K. The temperature was regulated by the use of Langevin dynamics with a collision frequency of 2.0 ps^{-1}. An equilibration simulation of 10 ns was run by keeping the temperature at 300 K and regulating the pressure at 1 atm, via isotropic position scaling with a pressure relaxation time of 0.5 ps. An additional NVT equilibration process keeping the temperature and volume regulated was conducted for 5 ns. An NVT production run of 100 ns was then carried out, while simulated configurations were stored every 1 ps. To study convergence of the GIST results, we performed four additional simulations of 100 ns starting from the same minimized configuration. The last simulation stored configurations every 0.2 ps.

3D-RISM analysis

3D-RISM calculations were carried out on the protein reference structure from the MD simulation with all water molecules removed. cTIP3P water[39] was used as a close analogue to the TIP3P water used in the MD simulations. cTIP3P differs by adding a Lennard-Jones potential energy term for the hydrogens, which is necessary for the application of RISM theory; it ensures that water hydrogens do not catastrophically overlap with neighboring water oxygens. Bulk properties of cTIP3P at 298 K with a dielectric of 78.497 and a density of 55.345 M were calculated for 3D-RISM using DRISM[46]. The DRISM solution was obtained on a grid of 16,384 grid points, with a spacing of 0.025 Å, to a target residual tolerance 10^{-12} using the modified direct inversion of the iterative subspace (MDIIS)[49] to accelerate convergence. 3D-RISM and DRISM calculations were carried out in AmberTools 16[39,50,51], with 3D-RISM modified to produce 3D solvation thermodynamics maps. The 3D-RISM solutions were computed on a grid with a 0.5 Å grid spacing and a minimum buffer of 24 Å between the solute and the outer edges of the grid. The *center = 3* option, which rounds the position of the grid points within the reference frame, was used to align the 3D-RISM and GIST grids. For visualization purposes, the 3D-RISM grid was truncated to match that used for the GIST analysis. The 3D-RISM equation was solved to a tolerance of 10^{-6} with the PSE-3 closure, using
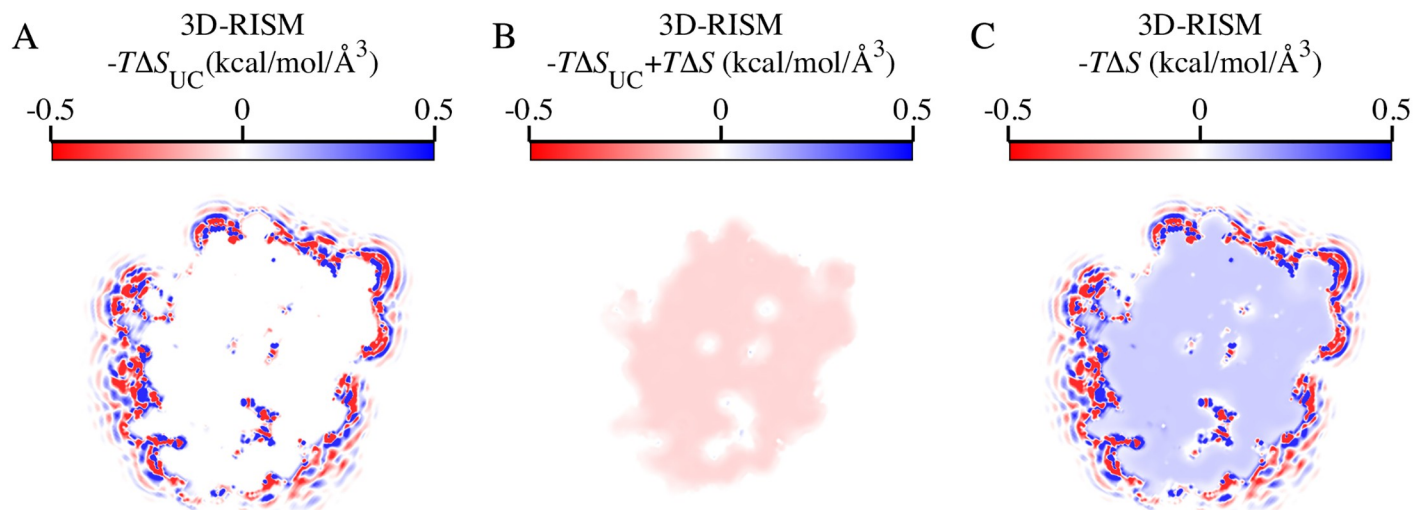
Fig 1. Cross section of the additively combined oxygen and hydrogen site solute-water energy density distributions about Factor Xa. Calculation from (A) 3D-RISM-PSE-3-UCT, (C) 3D-RISM-PSE-3, and (B) the difference.

previous solutions obtained with the KH and PSE-2 closures, and accelerated by the (MDIIS) solver. Solute-solvent interaction potentials were calculated with an infinite cut off. The calculation took about 30 min on a quad core 3.5 GHz Haswell Intel Core i5 processor.

To test the relevance of PMV corrections to thermodynamic density distributions, we applied the UCT with parameters of $a_1 = 0.0327564$, $a_0 = -0.000507492$, $b_1 = -3.26166$, and $b_0 = 0.0100166$, based on our unpublished results for cTIP3P using the same protocol as Ref.[24]. We have previously found for small molecules[24] that these corrections can provide solvation energies and entropies that agree well with experiment. In particular, the corrections primarily affect the solvation entropies and provided only a small improvement in solvation energies. A key feature of these corrections is that they are applied after the solvent density distribution has been calculated with 3D-RISM. They affect only the integrand of Eqs (15), (18) and (19), and do not modify the site density distributions $g_\alpha(\mathbf{r})$ or $c_\alpha(\mathbf{r})$. The corrected integrand for thermodynamic properties, such as solvation entropy (Fig 1), can still be visualized in 3D.

While these corrections provide large improvements to the solvation entropy, the region of space they affect is restricted to the excluded volume of the solute, as is shown by the difference (Fig 1B) between the corrected (Fig 1A) and uncorrected (Fig 1C) distributions. There is negligible impact on the distribution of solvation energy or entropy outside of this core region, which is the focus of this paper. Since the thermodynamic distributions in the binding pocket of the protein are not modified, we present uncorrected output from 3D-RISM-PSE-3 throughout Results.

Results

This section presents a comparison of molecular and thermodynamic density distributions generated by 3D-RISM and MD/GIST. First, we compare the PDFs for both oxygen and hydrogen. Next, we illustrate the benefits of the molecular reconstruction for 3D-RISM in the context of the solute-water energy density distributions. Finally, we compare the two methods' results for the water-water energy, total energy and total entropy density distributions.

It is worth emphasizing that well-converged GIST results are based on MD simulations that capture the true distribution functions for the choice of system and force field, while the faster

but approximate 3D-RISM method approximates these distribution functions. As a consequence, the GIST results for number densities and energy densities may be viewed as a reference data against which 3D-RISM may be evaluated. However, the GIST expression for the entropy is generally truncated after the single-body first term (Eq (5); but see Lazaridis[2]), so entropic contributions due to water-water correlations are neglected. In contrast, although 3D-RISM uses approximate distribution functions, it computes the full entropy, including both single-body and multi-body contributions for these functions. Based on these considerations, neither GIST nor 3D-RISM should be regarded as a reference touchstone for the hydration entropies.

Oxygen and hydrogen distributions

Number density distributions are the basic thermodynamic density distribution for GIST and 3D-RISM. Fig 2 shows the number density distributions relative to bulk generated by MD, which are used for GIST and 3D-RISM calculations. While there is overall agreement on high-density locations of oxygen and hydrogen, there are some cases where 3D-RISM and MD disagree on the position or existence of oxygen sites. In some cases, the position of a maximum is shifted. For example, the oxygen site found above location 4 in the MD distribution is moved up to the edge of the binding pocket in 3D-RISM. This is also the case, at location 7, where MD predicts a large oxygen density that is shifted much lower in the 3D-RISM distribution. Meanwhile, to the left of location 4, 3D-RISM predicts an oxygen site that is not observed in the MD distribution.

Overall, 3D-RISM has a less defined solvent distribution with lower, wider density maxima than found in the MD distributions. These results are consistent with previous comparisons of

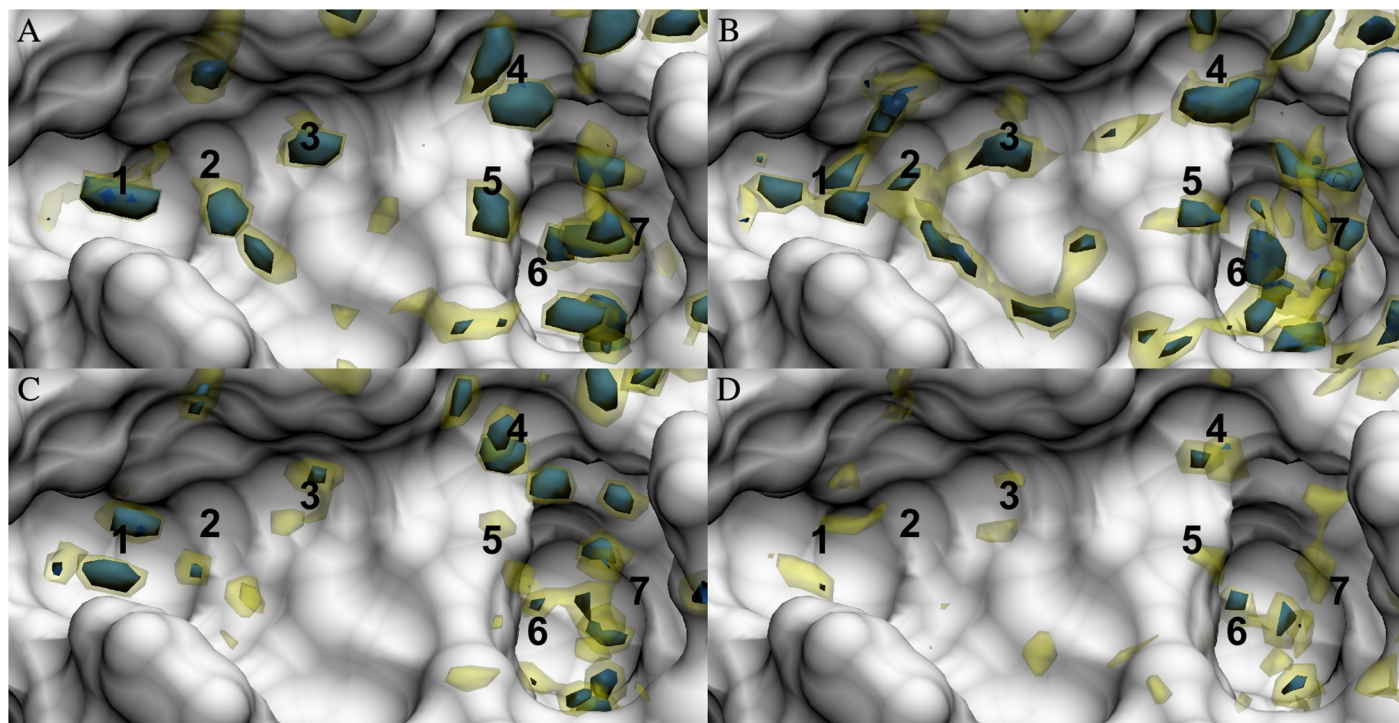**Fig 2. Solute-water pair distribution functions in the FXa binding site.** Oxygen pair distribution functions as calculated by (A) MD and (B) 3D-RISM. Hydrogen pair distribution functions as calculated by (C) MD and (D) 3D-RISM. Isosurfaces are shown for 4 (yellow) and 8 (blue) times bulk density. The FXa solvent accessible surface is shown in white. Numbers identify density regions that are discussed in the text.

MD and 3D-RISM[52–54]. The difference can be clearly seen at locations 1, 2, and 3 in Fig 2. Here, MD presents highly localized oxygen sites while 3D-RISM generally has smaller high-density regions (8 times bulk density) but larger low-density regions (4 times bulk density). Low-density channels between the oxygen sites connect the high-density sites for both 3D-RISM and MD, but are only visible in MD at around 2 times bulk density (not shown).

For both MD and 3D-RISM, hydrogen distributions are less localized than the oxygen distributions with few peaks above 8 times bulk density. As with the oxygen distributions, MD has more density maxima than 3D-RISM and these have higher densities but there are no channels in either 3D-RISM or MD connecting the maxima. Both methods do locate some well-defined hydrogen sites, indicating highly oriented water molecules. At location 3 an oxygen site has a pair of well-defined hydrogen sites on either side, suggesting that this is the location of well-ordered and localized water molecule. Location 6 shows similar behavior deep within the binding pocket.

Energy distributions

Solute-water energy density distributions. Using Eqs (16), (18) and (19), it is possible to transform the oxygen and hydrogen site densities generated from 3D-RISM (Fig 2) into 3D density distributions of the solvation energy, entropy, and free energy. As noted in the Introduction and Methods, the site-site nature of 3D-RISM makes it difficult to interpret the direct visualization of thermodynamic density distributions. Fig 3A shows this for the solute-water energy density distribution, generated from Eq (20). On the right, we observe many positive and negative regions of the distribution in close proximity to each other. It is possible, when
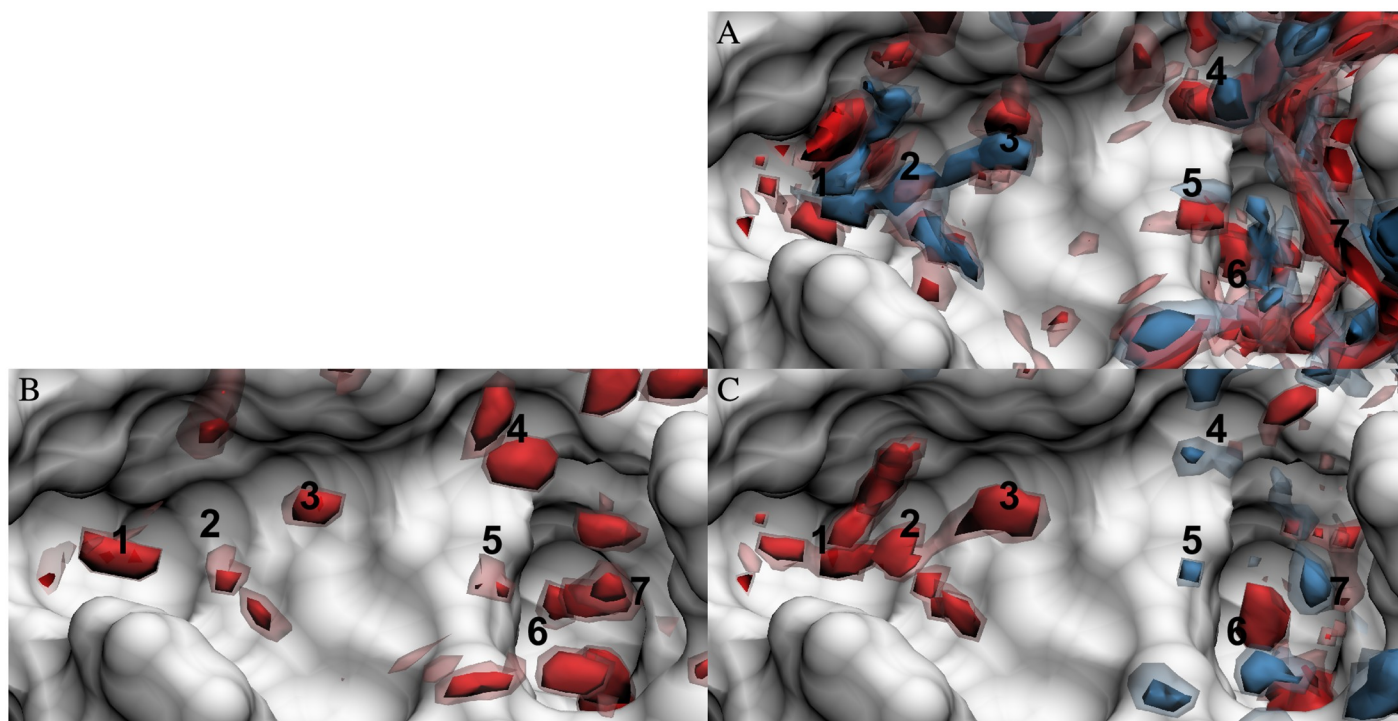
Fig 3. Solute-water energy density distributions in the FXa binding site. Energy density distribution were calculated using (A) 3D-RISM with additively combined oxygen and hydrogen site solute-water energy density distributions, (B) GIST, and (C) 3D-RISM with molecular reconstruction. Isosurfaces are shown for -2 (dark red), -1 (light red), 1 (light blue), and 2 (dark blue) kcal/mol/Å3. The FXa solvent accessible surface is shown in white. Numbers identify density regions that are discussed in the text.

comparing to Fig 2, to identify contributions from hydrogen and oxygen separately. For example, at location 3, the contribution from the oxygen can be seen clearly in blue (positive energy) while the two localized hydrogen density maxima can be identified in red (negative energy). However, it is not possible to visually determine if a water molecule at this location would have positive or negative energy.

Applying Eqs (22) to (19) effectively combines the oxygen and hydrogen contributions and presents a molecular picture (Fig 3C), analogous to GIST (Fig 3B). Looking again at location 3, we see that the negative energy solute-hydrogen energy outweighs the positive solute-oxygen energy, leading to a favorable overall interaction between solute and water. This is the same result provided by GIST. Throughout the binding cavity, we see that the molecular reconstruction removes most of the unfavorable blue regions, leaving primarily favorable red regions. Because molecular thermodynamic densities are easier to interpret, we will only discuss the molecular reconstruction thermodynamic densities from this point on.

The resulting molecular solute-water energy density distribution from 3D-RISM qualitatively reproduces many of the features of the GIST density distribution. As with the oxygen density distributions, extrema from the two distributions significantly overlap. Both methods also show that the weakest energetic interactions are in the middle of the binding pocket, although there is still significant water density in this region.

There are several notable differences as well. Around locations 2 and 3, 3D-RISM solute-water energy density distributions are broader than those of GIST, though they still overlap well. The energy density peak from 3D-RISM at location 1 is also similar to GIST, except that the main peak is split in two. GIST also has a density above location 1, near the edge of the pocket, that is shifted down in 3D-RISM. Location 6 is similar except that 3D-RISM predicts a much broader minimum than GIST. In contrast, at locations 4 and 5, 3D-RISM and GIST disagree on the nature of the interaction: 3D-RISM predicts positive solute-water energies while GIST predicts negative energies. Pressure artifacts of the PSE-n closures are the most likely reason why 3D-RISM finds these locations to be energetically unfavorable, in contrast to the reference GIST result. The geometry of the binding cavity and the positions of neighboring water molecules will cause density maxima to form at these locations. Since there are no particularly strong electrostatic interactions between the solute and the water at these locations, the van der Waals forces will dominate. The lack of repulsive terms in the closures causes the solvent sites to be placed too close to the solute and the potential energy is then too high (see Discussion for details). While locations 4 and 5 clearly demonstrate this behavior, there are a few other locations in the binding pocket where similar disagreement occurs. Location 7 shows a different type of disagreement in that the significant negative solute-water energy found at this location is shifted down and is of much lower magnitude in the 3D-RISM distribution. This is consistent with the location of the oxygen density presented in Fig 2. While the water density is not well localized even in GIST, it seems to make a significant contribution to the solute-water energy.

Water-water energy density distributions. Water-water energy density distributions (Fig 4) indicate the change in energy of the current distribution relative to the bulk liquid, excluding interactions with the protein. These interactions should be unfavorable, in general, as water molecules in the binding pocket will not form the same hydrogen bond network as found in bulk water. This is the case for GIST, where all water-water energies are positive.

While 3D-RISM has points of agreement (locations 6 and 7) with the reference GIST results, 3D-RISM disagrees significantly with GIST in some regions, providing often neutral or favorable, rather than unfavorable, water-water interactions relative to bulk (locations 1–5). This is consistent with the overall impression that 3D-RISM water-water energies are more negative than those from GIST. A possible explanation for this is that bulk water from DRISM

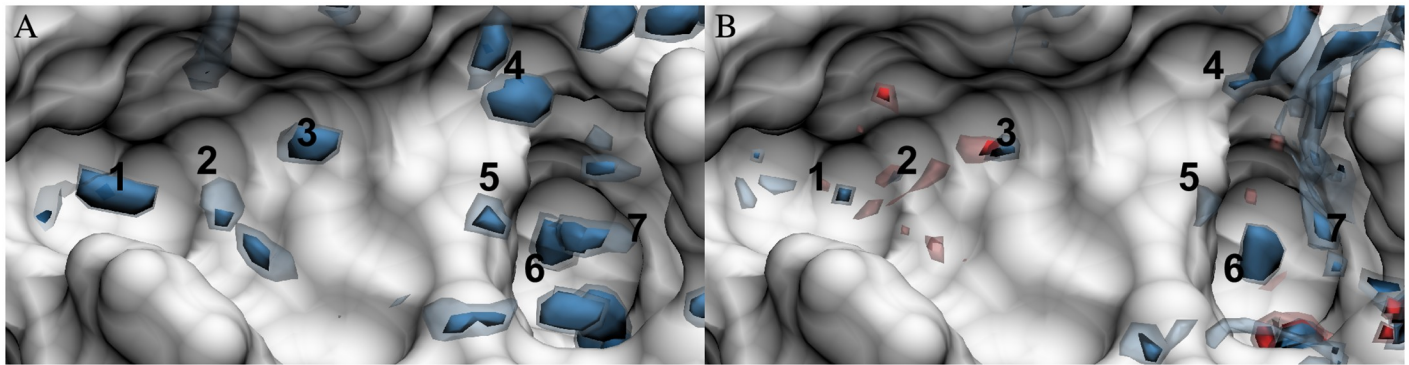**Fig 4. Water-water energy density distributions in the FXa binding site.** (A) GIST and (B) 3D-RISM with molecular reconstruction. Coloring and numbering are as in Fig 3.

has less structure and fewer hydrogen bonds are broken when water moves from the bulk to the binding site. Thus, the loss of favorable interactions with other water molecules is less for 3D-RISM than it is for GIST. It is possible that the structure of the binding site even creates a template for improved water-water interactions within RISM theory. The lack of structure in DRISM bulk water is attributed to the well-known artifacts found in the PSE-n family of closures (see Discussion for details).

Total water energy density distributions. When water-water and solute-water density distributions are combined to obtain total water energy density distributions, key differences already observed become magnified (Fig 5). The following analysis of specific locations illustrates the types of differences observed.

Locations 2 and 3 exemplify one type of deviation of 3D-RISM from the reference GIST results. Here, 3D-RISM reports a large area of favorable energies for the water, whereas GIST reports bulk-like energies. While both methods provide similar estimates of the solute-water energy, 3D-RISM underestimates the loss of the bulk water interactions, causing the total energy to be too negative. In contrast, at locations 4 and 5, 3D-RISM predicts positive total water energies whereas GIST predicts neutral or negative energies. While 3D-RISM again discounts the loss of water-water interactions, it also predicts repulsive interactions with the protein at these locations. Deep in the binding pocket, location 6, both GIST and 3D-RISM qualitatively agree on the density of oxygen with well-ordered hydrogens (Fig 2), the water-
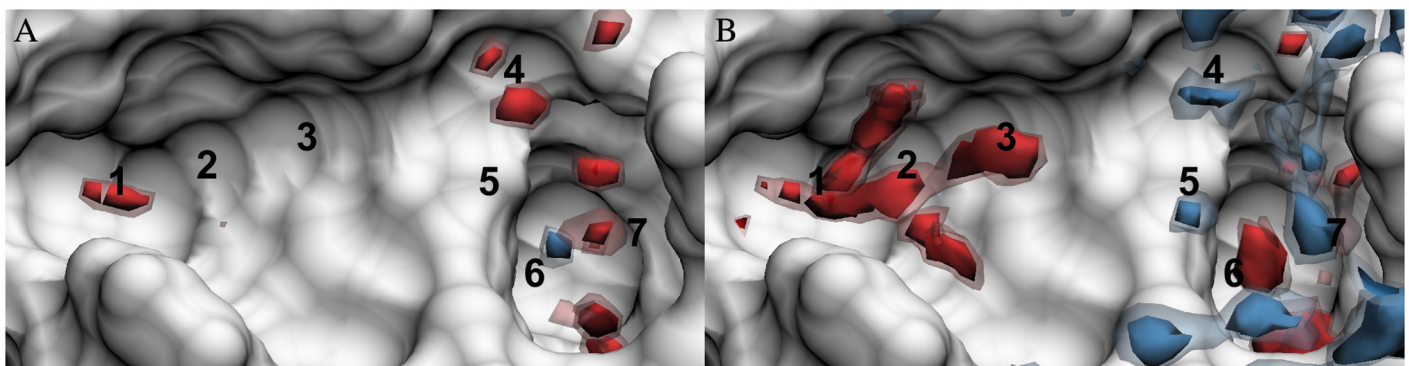

Fig 5. Total water energy density distributions in the FXa binding site. (A) GIST and (B) 3D-RISM with molecular reconstruction. Coloring and numbering are as in Fig 3.
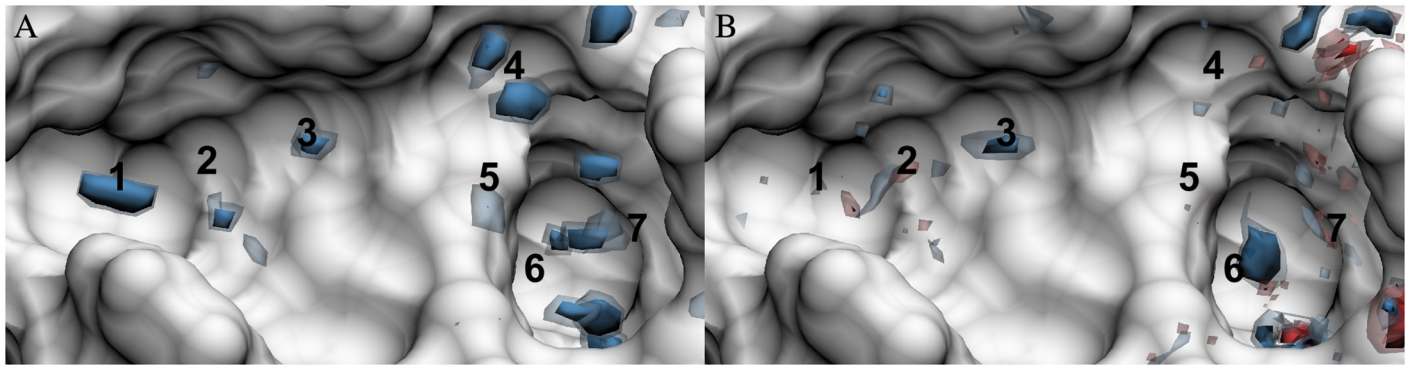
Fig 6. Total water entropy (-TΔS) density distributions in the FXa binding site. (A) GIST and (B) 3D-RISM with molecular reconstruction. Isosurfaces are shown for -2 (dark red), -1 (light red), 1 (light blue), and 2 (dark blue) kcal/mol/Å3. The FXa solvent accessible surface is shown in white. Numbers identify regions that are discussed in the text.

water energy (Fig 4), and the solute-water energy (Fig 3). However, the difference between the two energies results in a negative total energy from 3D-RISM while that of GIST is positive. Finally, at location 7, 3D-RISM does not assign a significant solvent density, and thus does not have the high energy density seen in GIST at this location.

Entropy density distributions

As noted above, both GIST and 3D-RISM provide only approximate results for the entropy, and each makes different approximations. On one hand, GIST provides entropy under the assumption of only low-order correlations, and thus, cannot provide the full entropy, even in the limit of infinite simulation time. On the other hand, 3D-RISM accounts for all multi-body correlations, but its results are based on approximate distribution functions, due to the necessity of using an approximate closure relation. As a consequence, although it is useful to compare the two methods, neither can be relied upon as a reliable reference method for the other.

The density distribution of $-T\Delta S$ from GIST (Fig 6A) resembles the distribution of the water-water energy density from GIST (Fig 4A). In particular, both have peaks in similar locations, due largely to the fact that both thermodynamic densities are weighted by number density; and both are consistently unfavorable. Indeed, a consequence of the truncation of the entropy expansion to single-body interactions Eq (5) is that only unfavorable entropies, relative to bulk, are possible. It is thus worth noting that the neglected higher-order terms could, at least in principle, contribute enough to generate entropically favorable regions. For 3D-RISM, the entropy density distribution is smoother and, unlike GIST's, includes some favorable regions (Fig 6B). Locations 3 and 6 are particularly well-peaked sites of unfavorable entropy. Overall, though, 3D-RISM has much lower entropy penalties for water to leave bulk and move to the surface of the protein.

Discussion

Overview

This study compares particle density and thermodynamic density distributions from 3D-RISM, analyzed with a new molecular reconstruction technique, and GIST, for essentially identical solutes, force fields, and grids. This is an excellent test of the number and energy density results from 3D-RISM, because GIST provides these distributions without approximation, except for the small numerical uncertainty associated with finite MD sampling in this case. It

is also of interest to compare the entropy distributions, although it is important to note here that the GIST calculations omit the effects of water-water correlations, whereas these are included in 3D-RISM, so GIST cannot be viewed as a rigorous reference method for this term. Overall, we find good qualitative agreement between the number densities from the two methods, as they agree on the locations of most density peaks in the protein-water pair distribution function. However, 3D-RISM tends to produce less sharply peaked distributions than MD, which is a commonly observed difference for PSE-n and related closures[39,52–54]. The solute-water energy density distributions are also in general good agreement between the two methods, but the water-water energies deviate more. Indeed, GIST consistently shows only unfavorable contributions from this term, because water at the protein surface has fewer stabilizing interactions with other waters than does water in bulk; but 3D-RISM gives favorable water-water energies, relative to bulk at some locations. A similar picture emerges for the entropy terms, which, again, are consistently unfavorable in GIST, but occasionally favorable in 3D-RISM. As a consequence, 3D-RISM may tend to assign less thermodynamic benefit to displacing surface water from a binding site than does GIST.

Accounting for the differences between GIST and 3D-RISM

A number of factors may contribute to the differences seen between GIST and 3D-RISM results. Naturally, dissimilarities in the density distributions produced by 3D-RISM and MD lead to differences between their thermodynamic distributions. 3D-RISM with the PSE-n family of closures is known to produce high density regions that are broader and lower peaked than those produced by MD[52,54]. These characteristics are observed in all of the density and thermodynamic distributions presented here. Small differences in the water models used and approximations in the molecular reconstruction procedure also contribute the differences observed between 3D-RISM and GIST results, though the amount is difficult to quantify.

Differences in the 3D-RISM density distributions from MD are partially due to well-known problems with the PSE-n family of closures. PSE-n closures are partial series expansions of the HNC closure. HNC (PSE-∞), in turn, approximates the full closure in terms of simple two-body correlation functions and misses the infinite series of many-body correlation functions contained in the full closure. These missing correlation functions are mainly repulsive, and their omission allows over-solvation, in which solvent sites form too close to each other or the solute[55], and results in seemly collapsed or distorted solvation structure.

When over-solvation occurs, hydrogen or oxygen sites approach the solute too closely and there is an increase in the magnitude of steric (positive) and electrostatic (negative) potential energies at the solute-water interface. Steric clashes can result in regions of positive solute-water energy density (Fig 3, location 4) while electrostatic interactions result in regions that are too negative (Fig 3, locations 1–3). This has previously been observed in other systems, such as 3D-RISM's over representation of sodium ions around DNA phosphate backbones compared to explicit solvent molecular dynamics[54].

Over-solvation directly leads to global artifacts, such as excessively high pressure[53], which can be treated by PMV corrections. PMV corrections have been used successfully to mitigate the shortcomings of the PSE-n closures but we have found that these corrections have negligible impact outside the core region of the solute and are, therefore, unnecessary. This is due to the limited range of the direct correlation function, from which the PMV is calculated. As a result, these corrections only correct the PV work of introducing a solute into the solvent and do not significantly modify the details of the thermodynamics density distributions outside of the excluded volume of the solute.

The cTIP3P water model used with 3D-RISM, which is similar, but not identical, to the TIP3P model used in MD, also affects the structure of the solute-water pair distribution function. The most significant difference is the addition of a Lennard-Jones potential to the hydrogen sites, which is required to prevent catastrophic overlaps of hydrogen and water in RISM theory. In cTIP3P the Lennard-Jones diameters of hydrogen, σ_H, is determined by $\sigma_H = \sigma_O\text{-}2L_{OH} = 1.2363$ Å and enters 3D-RISM theory both through $\chi_{\alpha\gamma}(\mathbf{r})$ in Eq (10) and $u_\alpha(\mathbf{r})$ in Eq (13). Unlike smaller values sometimes used with three-point water models, it gives the appropriate contact distance for hydrogen bonding. This greatly improves radial distribution functions of bulk water for OH and HH, particularly in improving the position and height of the first peak, but slight broadening of peaks in the OO distribution. cTIP3P, cSPC/E, and similar models have been widely used for the calculation of bulk water hydration thermodynamics with much success[18–21,26,39,42,53,56]. It is also worth noting that, through DRISM, we impose a dielectric constant of 78.497 on the water model to match experiment, compared to a value of 94 for TIP3P[57]. We expect this water model to give the best agreement with experiment when used with 3D-RISM; but it is not optimized to agree with explicit solvent MD.

Another factor that contributes to the differences found between GIST and 3D-RISM is the molecular reconstruction method we have developed. The molecular reconstruction remaps the hydrogen contributions to the thermodynamic density distributions of the parent oxygen using only intramolecular bond distances. This works well for $h_\alpha(\mathbf{r})$ and the protein-water energy density as a result, which are local quantities. However, the water-water energy, Eq (21), and the total entropy, Eq (18), depend on non-local contributions from $c_\alpha(\mathbf{r})$, $h_\alpha(\mathbf{r})\,c_\alpha(\mathbf{r})$, and their temperature derivatives. While $c_\alpha(\mathbf{r})$ contributes mainly to the core region of the protein, $h_\alpha(\mathbf{r})\,c_\alpha(\mathbf{r})$ can influence the first solvation shell. Alternate reconstructions, that better account for $h_\alpha(\mathbf{r})\,c_\alpha(\mathbf{r})$, may work better.

Implications and directions

Both 3D-RISM and GIST provide detailed thermodynamic density distributions of solvent in binding cavities that can be used to guide the development of new pharmaceutical compounds. For systems like Factor Xa, where relevant water configurations can be easily sampled due to the solvent-exposed character of the binding site, GIST provides rigorous results at modest computational cost. Indeed, beyond the use of the force field, the only approximations in GIST are those due to sampling and truncation of the entropy calculation. On the other hand, exhaustive sampling is not possible in all systems. For example, in some systems water molecules may have frustrated dynamics or exchange with buried solvent pockets. In addition, if one wishes to examine solvation for multiple receptor conformations, the cost of running multiple MD simulations for GIST may become prohibitive. In such cases, the low computational cost of 3D-RISM can be of great benefit despite the approximations inherent in the theory. In future work, we will evaluate the potential benefits of both methods in the context of drug design.

Author Contributions

Conceptualization: Michael K. Gilson, Tom Kurtzman, Tyler Luchko.

Formal analysis: Crystal Nguyen, Takeshi Yamazaki, Tyler Luchko.

Funding acquisition: David A. Case, Michael K. Gilson, Tom Kurtzman, Tyler Luchko.

Investigation: Crystal Nguyen, Tyler Luchko.

Methodology: Andriy Kovalenko, David A. Case, Michael K. Gilson, Tom Kurtzman, Tyler Luchko.

Software: Crystal Nguyen, Takeshi Yamazaki, Tyler Luchko.

Visualization: Tyler Luchko.

Writing – original draft: Crystal Nguyen, Michael K. Gilson, Tom Kurtzman, Tyler Luchko.

Writing – review & editing: Michael K. Gilson, Tom Kurtzman.

References

1. Ornstein LS, Zernike F. Accidental deviations of density and opalescence at the critical point of a single substance. Proc Akad Sci(Amsterdam). 1914. p. 793.

2. Lazaridis T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. J Phys Chem B. 1998; 102: 3531–3541. https://doi.org/10.1021/jp9723574

3. Lazaridis T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. J Phys Chem B. 1998; 102: 3542–3550. https://doi.org/10.1021/jp972358w

4. Green HS. The Molecular Theory of Fluids. Amsterdam: North-Holland; 1952.

5. Nettleton RE, Green MS. Expression in Terms of Molecular Distribution Functions for the Entropy Density in an Infinite System. J Chem Phys. 1958; 29: 1365–1370. https://doi.org/10.1063/1.1744724

6. Lazaridis T, Karplus M. Orientational correlations and entropy in liquid water. J Chem Phys. 1996; 105: 4294–4316. https://doi.org/10.1063/1.472247

7. Nguyen CN, Kurtzman T, Gilson MK. Spatial Decomposition of Translational Water–Water Correlation Entropy in Binding Pockets. J Chem Theory Comput. 2016; 12: 414–429. https://doi.org/10.1021/acs.jctc.5b00939 PMID: 26636620

8. Wang L, Abel R, Friesner RA, Berne BJ. 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: 1462–1473. https://doi.org/10.1021/ct900078k PMID: 19851475

9. Zielkiewicz J. Structural properties of water: Comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. J Chem Phys. 2005; 123: 104501. https://doi.org/10.1063/1.2018637 PMID: 16178604

10. Young T, Abel R, Kim B, Berne BJ, Friesner RA. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. Proc Natl Acad Sci. 2007; 104: 808–813. https://doi.org/10.1073/pnas.0610202104 PMID: 17204562

11. Abel R, Young T, Farid R, Berne BJ, Friesner RA. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. J Am Chem Soc. 2008; 130: 2817–2831. https://doi.org/10.1021/ja0771033 PMID: 18266362

12. Li Z, Lazaridis T. Computing the Thermodynamic Contributions of Interfacial Water. In: Baron R, editor. Computational Drug Discovery and Design. Springer New York; 2012. pp. 393–404. https://doi.org/10.1007/978-1-61779-465-0_24 PMID: 22183549

13. Nguyen C, Gilson MK, Young T. Structure and Thermodynamics of Molecular Hydration via Grid Inhomogeneous Solvation Theory. ArXiv11084876 Phys Q-Bio. 2011; http://arxiv.org/abs/1108.4876

14. Nguyen CN, Young TK, Gilson MK. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. J Chem Phys. 2012; 137: 044101. https://doi.org/10.1063/1.4733951 PMID: 22852591

15. Haider K, Cruz A, Ramsey S, Gilson MK, Kurtzman T. Solvation Structure and Thermodynamic Mapping (SSTMap): An open-source, flexible package for the analysis of water in molecular dynamics trajectories.

16. Kovalenko A, Hirata F. Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model. J Chem Phys. 1999; 110: 10095–10112. https://doi.org/10.1063/1.478883

17. Hansen J-P, McDonald IR. Theory of Simple Liquids. Academic Press; 2006.

18. Palmer DS, Frolov AI, Ratkova EL, Fedorov MV. Towards a universal method for calculating hydration free energies: a 3D reference interaction site model with partial molar volume correction. J Phys Condens Matter. 2010; 22: 492101. https://doi.org/10.1088/0953-8984/22/49/492101 PMID: 21406779

19. Truchon J-F, Pettitt BM, Labute P. A Cavity Corrected 3D-RISM Functional for Accurate Solvation Free Energies. J Chem Theory Comput. 2014; 10: 934–941. https://doi.org/10.1021/ct4009359 PMID: 24634616

20. Misin M, Fedorov MV, Palmer DS. Communication: Accurate hydration free energies at a wide range of temperatures from 3D-RISM. J Chem Phys. 2015; 142: 091105. https://doi.org/10.1063/1.4914315 PMID: 25747054

21. Sergiievskyi V, Jeanmairet G, Levesque M, Borgis D. Solvation free-energy pressure corrections in the three dimensional reference interaction site model. J Chem Phys. 2015; 143: 184116. https://doi.org/10.1063/1.4935065 PMID: 26567655

22. Ratkova EL, Chuev GN, Sergiievskyi VP, Fedorov MV. An Accurate Prediction of Hydration Free Energies by Combination of Molecular Integral Equations Theory with Structural Descriptors. J Phys Chem B. 2010; 114: 12068–12079. https://doi.org/10.1021/jp103955r PMID: 20804181

23. Palmer DS, Sergiievskyi VP, Jensen F, Fedorov MV. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. J Chem Phys. 2010; 133: 044104. https://doi.org/10.1063/1.3458798 PMID: 20687630

24. Johnson J, Case DA, Yamazaki T, Gusarov S, Kovalenko A, Luchko T. Small molecule hydration energy and entropy from 3D-RISM. J Phys Condens Matter. 2016; 28: 344002. https://doi.org/10.1088/0953-8984/28/34/344002 PMID: 27367817

25. Tielker N, Tomazic D, Heil J, Kloss T, Ehrhart S, Güssregen S, et al. The SAMPL5 challenge for embedded-cluster integral equation theory: solvation free energies, aqueous pKa, and cyclohexane–water log D. J Comput Aided Mol Des. 2016; 1–10. https://doi.org/10.1007/s10822-016-9939-7 PMID: 27554666

26. Luchko T, Blinov N, Limon GC, Joyce KP, Kovalenko A. SAMPL5: 3D-RISM partition coefficient calculations with partial molar volume corrections and solute conformational sampling. J Comput Aided Mol Des. 2016; 1–13. https://doi.org/10.1007/s10822-016-9947-7 PMID: 27585474

27. Misin M, Fedorov MV, Palmer DS. Hydration Free Energies of Molecular Ions from Theory and Simulation. J Phys Chem B. 2016; 120: 975–983. https://doi.org/10.1021/acs.jpcb.5b10809 PMID: 26756333

28. Misin M, Palmer DS, Fedorov MV. Predicting Solvation Free Energies Using Parameter-Free Solvent Models. J Phys Chem B. 2016; 120: 5724–5731. https://doi.org/10.1021/acs.jpcb.6b05352 PMID: 27270044

29. Huang W, Blinov N, Kovalenko A. Octanol–Water Partition Coefficient from 3D-RISM-KH Molecular Theory of Solvation with Partial Molar Volume Correction. J Phys Chem B. 2015; 119: 5588–5597. https://doi.org/10.1021/acs.jpcb.5b01291 PMID: 25844645

30. Güssregen S, Matter H, Hessler G, Lionta E, Heil J, Kast SM. Thermodynamic Characterization of Hydration Sites from Integral Equation-Derived Free Energy Densities: Application to Protein Binding Sites and Ligand Series. J Chem Inf Model. 2017; 57: 1652–1666. https://doi.org/10.1021/acs.jcim.6b00765 PMID: 28565907

31. Sindhikara DJ, Hirata F. Analysis of Biomolecular Solvation Sites by 3D-RISM Theory. J Phys Chem B. 2013; 117: 6718–6723. https://doi.org/10.1021/jp4046116 PMID: 23675899

32. Ramsey S, Nguyen C, Salomon-Ferrer R, Walker RC, Gilson MK, Kurtzman T. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. J Comput Chem. 2016; 37: 2029–2037. https://doi.org/10.1002/jcc.24417 PMID: 27317094

33. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput. 2013; 9: 3084–3095. https://doi.org/10.1021/ct400341p PMID: 26583988

34. Huggins DJ. Estimating Translational and Orientational Entropies Using the k-Nearest Neighbors Algorithm. J Chem Theory Comput. 2014; 10: 3617–3625. https://doi.org/10.1021/ct500415g PMID: 26588506

35. Singh H, Misra N, Hnizdo V, Fedorowicz A, Demchuk E. Nearest Neighbor Estimates of Entropy. Am J Math Manag Sci. 2003; 23: 301–321. https://doi.org/10.1080/01966324.2003.10737616

36. Kovalenko A. Three-dimensional Rism Theory for Molecular Liquids and Solid-Liquid Interfaces. In: Hirata F, editor. Molecular Theory of Solvation. Springer Netherlands; 2004. pp. 169–275. http://link.springer.com/chapter/10.1007/1-4020-2590-4_4

37. Kovalenko A, Hirata F. Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. J Chem Phys. 2000; 112: 10391–10402. https://doi.org/10.1063/1.481676

38. Beglov D, Roux B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. J Phys Chem B. 1997; 101: 7821–7826. https://doi.org/10.1021/jp971083h

39. Luchko T, Gusarov S, Roe DR, Simmerling C, Case DA, Tuszynski J, et al. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. J Chem Theory Comput. 2010; 6: 607–624. https://doi.org/10.1021/ct900460m PMID: 20440377

40. Kovalenko A. Multiscale Modeling of Solvation. In: Breitkopf C, Swider-Lyons K, editors. Springer Handbook of Electrochemical Energy. Springer Berlin Heidelberg; 2017. pp. 95–139. https://doi.org/10.1007/978-3-662-46657-5_5

41. Yu H-A, Roux B, Karplus M. Solvation thermodynamics: An approach from analytic temperature derivatives. J Chem Phys. 1990; 92: 5020–5033. https://doi.org/10.1063/1.458538

42. Johnson J, Luchko T, Yamazaki T, Kovalenko A, Gusarov S, Case DA. Small Molecule Solvation Energy and Entropy from 3D-RISM. In progress.;

43. Yamazaki T, Kovalenko A. Spatial Decomposition Analysis of the Thermodynamics of Cyclodextrin Complexation. J Chem Theory Comput. 2009; 5: 1723–1730. https://doi.org/10.1021/ct9000729 PMID: 26609997

44. Kast SM, Kloss T. Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. J Chem Phys. 2008; 129: 236101. https://doi.org/10.1063/1.3041709 PMID: 19102559

45. Morita T. Theory of Classical Fluids: Hyper-Netted Chain Approximation, I Formulation for a One-Component System. Prog Theor Phys. 1958; 20: 920–938. https://doi.org/10.1143/PTP.20.920

46. Perkyns J, Pettitt BM. A site–site theory for finite concentration saline solutions. J Chem Phys. 1992; 97: 7656–7666.

47. 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 Bioinforma. 2006; 65: 712–725. https://doi.org/10.1002/prot.21123 PMID: 16981200

48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79: 926–935. https://doi.org/10.1063/1.445869

49. Kovalenko A, Ten-no S, Hirata F. Solution of three-dimensional reference interaction site model and hypernetted chain equations for simple point charge water by modified method of direct inversion in iterative subspace. J Comput Chem. 1999; 20: 928–936. https://doi.org/10.1002/(SICI)1096-987X(19990715)20:9<928::AID-JCC4>3.0.CO;2-X

50. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, et al. AMBER 12. University of California, San Francisco; 2012.

51. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci. 2013; 3: 198–210. https://doi.org/10.1002/wcms.1121

52. Stumpe MC, Blinov N, Wishart D, Kovalenko A, Pande VS. Calculation of Local Water Densities in Biological Systems: A Comparison of Molecular Dynamics Simulations and the 3D-RISM-KH Molecular Theory of Solvation. J Phys Chem B. 2011; 115: 319–328. https://doi.org/10.1021/jp102587q PMID: 21174421

53. Joung IS, Luchko T, Case DA. Simple electrolyte solutions: Comparison of DRISM and molecular dynamics results for alkali halide solutions. J Chem Phys. 2013; 138: 044103. https://doi.org/10.1063/1.4775743 PMID: 23387564

54. Giambaşu GM, Luchko T, Herschlag D, York DM, Case DA. Ion Counting from Explicit-Solvent Simulations and 3D-RISM. Biophys J. 2014; 106: 883–894. https://doi.org/10.1016/j.bpj.2014.01.021 PMID: 24559991

55. Kovalenko A, Hirata F. Hydration free energy of hydrophobic solutes studied by a reference interaction site model with a repulsive bridge correction and a thermodynamic perturbation method. J Chem Phys. 2000; 113: 2793–2805. https://doi.org/10.1063/1.1305885

56. Fujita T, Yamamoto T. Assessing the accuracy of integral equation theories for nano-sized hydrophobic solutes in water. J Chem Phys. 2017; 147: 014110. https://doi.org/10.1063/1.4990502 PMID: 28688395

57. Vega C, Abascal JLF. Simulating water with rigid non-polarizable models: a general perspective. Phys Chem Chem Phys. 2011; 13: 19663–19688. https://doi.org/10.1039/c1cp22168j PMID: 21927736