Thermodynamic decomposition of solvation free energies with particle mesh Ewald and long-range Lennard-Jones interactions in Grid Inhomogeneous Solvation Theory

Lieyang Chen[1, 2], Anthony Cruz[1, 3] Daniel R. Roe[4], Andrew C. Simmonett[4], Lauren Wickstrom[5], Nanjie Deng[6], Tom Kurtzman[1, 2, 3] *

Correspondence to: Tom Kurtzman (E-mail: thomas.kurtzman@lehman.cuny.edu)

[1] Department of Chemistry, Lehman College, The City University of New York, 250 Bedford Park Boulevard West, Bronx, New York, 10468, USA
[2] Ph.D. Program in Biochemistry, The Graduate Center of The City University of New York, New York, New York, 10016, USA
[3] Ph.D. Program in Chemistry, The Graduate Center of The City University of New York, New York, New York, 10016, USA
[4] National Institutes of Health – National Heart, Lung and Blood Institute, Laboratory of Computational Biology, Bethesda, Maryland, 20892, USA
[5] Department of Science, Borough of Manhattan Community College, The City University of New York, New York, New York,10007, USA
[6] Department of Chemistry and Physical Sciences, Pace University, New York, New York, 10038, USA

Abstract:

Grid Inhomogeneous Solvation Theory (GIST) maps out solvation thermodynamic properties on a fine meshed grid and provides a statistical mechanical formalism for thermodynamic end-state calculations. However, differences in how long-range non-bonded interactions are calculated in molecular dynamics engines and in the current implementation of GIST have prevented precise comparisons between free energies estimated using GIST and those from other free energy methods such as thermodynamic integration (TI). Here, we address this by presenting PME-GIST, a formalism by which particle mesh Ewald (PME) based electrostatic energies and long-range Lennard-Jones (LJ) energies are decomposed and assigned to individual atoms and the corresponding voxels they occupy in a manner consistent with the GIST approach. PME-GIST yields potential energy calculations that are precisely consistent with modern simulation engines and performs these calculations at a dramatically faster speed than prior implementations. Here, we apply PME-GIST end-states analyses to 32 small molecules whose solvation free energies are close to evenly distributed from 2 kcal/mol to -17 kcal/mol and obtain solvation energies consistent with TI calculations (R^2 = 0.99, mean unsigned difference 0.8 kcal/mol). We also estimate the entropy contribution from the 2nd and higher order entropy terms that are truncated in GIST by the differences between entropies calculated in TI and GIST. With a simple correction for the high order entropy terms, PME-GIST obtains solvation free energies that are highly consistent with TI calculations (R^2 = 0.99, mean unsigned difference = 0.4 kcal/mol) and experimental results (R^2 = 0.88, mean unsigned difference = 1.4 kcal/mol). The precision of PME-GIST also enables us to show that the solvation free energy of small hydrophobic and hydrophilic molecules can be largely understood based on perturbations of the solvent in a region extending a few solvation shells from the solute. We have integrated PME-GIST into the open-source molecular dynamics analysis software CPPTRAJ.

Inhomogeneous fluid solvation theory (IST)[1–6] provides a statistical mechanical framework to calculate the thermodynamics of solvation from solvent molecular density distributions.[7,8] Building upon IST, we developed WaterMap [9,10] and GIST [11–13] to map out local water properties in high density spherical hydration sites and on a high resolution space-filling grid, respectively, while others have developed their own methods of applying IST[8,14,15]. A major motivation for the development of the grid-based implementation of IST, GIST, was the ability to account for all water molecules in the system being studied as opposed to just the portion that occupied high density regions as in WaterMap. In GIST, similar to an integration of hydrodynamic densities, a sum of density quantities over all the voxels of the system yields the corresponding system quantities. As GIST accounts for all the water in the system as well as the energetics of interaction between a solute and the water, sums over voxel quantities in the initial and final states can be used to calculate corresponding differences in thermodynamic state quantities such as energies and entropies for rigid solute systems. Indeed, several studies have used IST to estimate free energies of solvation for small molecules and amino acids.[15,16] However, discrepancies between how GIST tools estimate energies and how molecular dynamics (MD) engines calculate, particularly differences in the treatment of long-range interactions, have prevented precise comparison between GIST-calculated energies and the energies produced from MD free energy methods such as TI and FEP. Here, we address this by presenting a formalism by which PME-based electrostatic energies[17,18] and long-range LJ energies[19] are decomposed and assigned to individual atoms and the corresponding voxels they occupy in a manner consistent with the GIST formulation. This PME-GIST approach yields electrostatic and LJ energy calculations that are consistent with the modern simulation engine Amber-PMEMD[20] at a dramatically faster speed than GIST without PME. In addition, PME-GIST provides a more accurate solvation energy which, when combined with accurate approximations of the entropy contribution from high order entropy terms, leads to improved agreement in the solvation free energy compared with both TI and experimental measurements.

Here, we have run PME-GIST end states and TI calculations for the solvation free energy of 32 small molecules chosen from the FreeSolv dataset[21]. These small molecules range from hydrophobic to hydrophilic and have solvation free energies near evenly distributed over a range of 20 kcal/mol. For each solute, we run simulations at three different temperatures to extract the solvation entropies from the temperature dependence of the calculated free energies. Comparisons of these solvation entropies to those calculated using IST leads to a simple linear scaling correction to the IST entropies. While the IST formulation for system energies and entropies is formally exact, in practice, the N-body expansion for the entropy is truncated after the first order term for inhomogeneous systems. Most IST implementations, including the implementation of GIST in the MD analysis software CPPTRAJ[22], account only for the solute-water correlation entropy while neglecting the entropic contributions from correlations of two or more water molecules. Exceptions to this include explicit calculation of higher order correlations though these are not usually applied.[23,24] The differences between entropies calculated in TI and GIST are direct estimates of the higher order entropy terms that are not calculated in GIST.

Although the solvation energies calculated in the previous version of GIST implemented in CPPTRAJ (GIST-2016[1])[25] are highly correlated with TI, we find that PME-GIST significantly improves the agreement. The 1st order entropies in PME-GIST are also highly correlated with the solvation entropies calculated in TI (R^2=0.92). Interestingly, while the energies and entropies of PME-GIST are highly correlated with TI, the free energy is less so (R^2=0.75) however corrections to the IST entropy to account for the missing higher order terms leads to exceptionally high correlation between IST calculated free energies and TI (R^2=0.99). This leads us to conclude that in order to accurately calculate free energies with IST, the contribution of higher order terms must be taken into account.

We further investigate whether the solvation free energy of these small molecules can be well described by perturbations of the fluid proximal to the solute. We find that this is, indeed, the case for the full set of hydrophobic and hydrophilic molecules investigated in this study and that integrating out to approximately 10 Å is sufficient to highly correlate with the total solvation free energy.

Our results suggest that including both the long-range interactions and the entropic contributions of higher order water configurational correlations enables the use of GIST as a robust tool to measure solvation free energy and its components both systematically and locally.

2. Methods

2.1 MD simulations

The 3D structures of the 32 small molecules were obtained from the PubChem database[26]. The atomic partial charges of the molecules were assigned using AM1-BCC in Antechamber[27] and other forcefield parameters were parameterized by the GAFF2 force field[28]. The solutes were then solvated in a TIP3P[29] rectangular water box, where the box's edges are at least 15 Å from the solutes using tleap from AmberTool20[20]. The system was minimized and then heated to 300 K in 20 ps under NVT conditions using Langevin temperature control with a time constant of 2 ps for the heat bath coupling[30]. The systems were then equilibrated for 250 ps at a constant temperature of 300 K and pressure of 1 bar maintained by the isotropic position scaling barostat[31] with a collision rate of 2 ps. The MD simulations were performed using GPU-accelerated PMEMD[32] from Amber20 [33] with periodic boundary. Electrostatic interactions were modeled by PME with 9 Å direct space cutoff. LJ interactions were fully accounted for up to a 9 Å cutoff and long-range interactions were treated by the default isotropic long-range correction. The NPT production MD runs were for 100 ns with the time step of 2 fs, the bond length involving hydrogen were constrained by SHAKE algorithm[22] . The system configurations stored every 2 ps for subsequent GIST analyses. Throughout all minimization, equilibration, and productions steps, all solute atoms were positionally restrained to the center of the systems by 100 kcal/mol/Å^2 force constant throughout the simulations.

2.2 TI calculations

The solvation free energy is calculated by Amber TI[20,34] as:

$$\Delta G_{solv} = -\int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle d\lambda \qquad (1)$$

where λ=0 represents the solvated state in which the solute fully interacts with the surrounding water molecules and λ=1 represents the state in which the solute is decoupled from the solvent. For the TI simulations, the systems were minimized, heated and equilibrated using the same protocol as for PME-GIST simulations except an additional 250 ps NPT equilibrium simulation performed before the production run

[1] In the text, we will use the term "GIST-2016" to refer to the original GIST implementation in 2016 and "PME-GIST" to refer to the current version that we developed in this study.

for every λ. The TI calculations were conducted over 15 λ states (0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.90, 0.95, 1.0) with a soft-core potential function using a one-step transformation with parameters of α_{LJ}=0.5 and β_C=12 [35]. TI simulations were run for each solute at three different temperatures (270 K, 300 K, and 330 K) with three runs at each temperature for which the velocities were randomly assigned prior to each 250ps equilibration. Solvation entropies were then estimated using the finite-difference method [36,37], in which the entropies were approximated by the slopes of solvation free energies over temperatures.

2.3 GIST calculations

Both GIST-2016 and PME-GIST are implemented in CPPTRAJ[22]. The GIST-2016 and PME-GIST calculations were performed on 100 ns production trajectories with frames stored every 2 ps for a total of 50,000 frames for each trajectory. Cubic GIST grids that covered the systems with a resolution of 0.125 Å^3 for each voxel were centered on the solutes with a bulk water number density of 0.03288 mol/Å^3. In the PME-GIST, the PME-based electrostatic energy calculation and long-range LJ correction were turned on and a cutoff of 9 Å was used for both the direct space electrostatic interactions and short-range LJ interactions. For each solute, GIST calculations were performed on three independent MD replicates.

In the output of the GIST calculations (GIST-2016 and PME-GIST), the energy density, entropy density and water density are printed for each voxel. The total energy, entropy and water number of the system then can be yielded by summing those densities multiplied by the voxel volume over the system (eqs. 2-4). For the total properties of a local region within a certain radius to the solute, GISTPP tool[25] was used to define the voxels belongs to that region and summing over the energy density, entropy density and water density multiplied by the voxel volume over those voxels will yielded the total energy, total entropy and total water number for that region, respectively.

$$E_{total} = \sum_k E_{dens} V_k \qquad (2)$$

$$S_{total} = \sum_k S_{dens} V_k \qquad (3)$$

$$N_{total} = \sum_k \rho_{dens} V_k \qquad (4)$$

For the entropy calculations, voxels that were not entirely in the simulation box were removed from the sum in equation 3. This occurs because the volume of the box fluctuates at constant pressure. As the water in this region is distal from the solute, its properties are bulk like and the contribution to the solvation entropy is zero.

We also implement a systematic correction to the GIST entropy for individual states which accounts for the systematic bias in the nearest neighbor algorithm for calculating the entropy of water molecules. In this study we account for this bias precisely and quantified it as 0.0033 kcal/mol per water molecule. For entropy of the entire system or a local region in the final state, the systematic bias is corrected by multiplying the total number of water molecules in the system or in the region by 0.0033 kcal/mol.

2.4 End states calculation by GIST

In the end states calculation, the solvation free energy for a rigid solute was estimated by the difference in free energy between a final state, in which a rigid solute fully interacts with the solvent; and an initial state, in the rigid solute is fully decoupled from the solvent (eqs. 5-7).

$$\Delta A_{solv} = \Delta E_{solv} - T\Delta S_{solv} \quad (5)$$

$$\Delta E_{solv} = E_{final} - E_{initial}' \quad (6)$$

$$\Delta S_{solv} = S_{final} - S_{initial} \quad (7)$$

Where ΔA_{solv} is the difference in the Helmholtz free energy between the two states, ΔE_{solv} and $-T\Delta S_{solv}$ are the difference in energy and temperature times the entropy difference of two states, respectively. The prime in equation 6 denotes the inclusion of a solute self-energy term in the initial state of the PME-GIST based solvation energy calculation. This term includes non-bond interactions within the solute and between the solute and all of its periodic replicas and is calculated by the energy command in CPPTRAJ[22]. The entropy calculations in GIST-2016 and PME-GIST are identical. In this study, we use ΔE_{solv} to estimate solvation enthalpy and ΔA_{solv} to estimate solvation Gibbs free energy given that in an NPT ensemble, the $P\Delta V$ term is negligible[15].

3 Results

3.1 Modifications to the energy calculation in GIST

The total energy of the central unit cell in a periodic system is a sum the electrostatic energy and LJ energy.

$$E_{total} = E_{elec} + E_{lj} \quad (8)$$

Where the electrostatic energy of the cell can be written as:

$$E_{elec} = \frac{1}{2}\sum_{\boldsymbol{n}}\sum_{i}^{N}\sum_{j}^{N'} k_e \frac{q_i q_j}{|\boldsymbol{r}_j - \boldsymbol{r}_i + \boldsymbol{n}|} \quad (9)$$

And the LJ energy for the system is:

$$E_{lj} = \sum_{\boldsymbol{n}}\sum_{i}^{N}\sum_{j}^{N'} \frac{A_{(i,j)}}{|\boldsymbol{r}_j - \boldsymbol{r}_i + \boldsymbol{n}|^{12}} - \frac{B_{(i,j)}}{|\boldsymbol{r}_j - \boldsymbol{r}_i + \boldsymbol{n}|^{6}} \quad (10)$$

in which q_i and q_j are the atomic partial charges of two atoms at position $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$ respectively. N is the total particle number in the unit cell, k_e is the coulomb constant. $A_{(i,j)}$ ($4\varepsilon\sigma_{ij}^{12}$) and $B_{(i,j)}$ ($4\varepsilon\sigma_{ij}^{6}$) are the LJ parameters of two atoms. $\boldsymbol{n}$ represents a vector between an atom in the unit cell and its image in a periodic cell. The sum over $\boldsymbol{n}$ is over the full set of vectors that move a particle from the unit cell to each of its periodic images. The prime on the second N (in eq.9 and eq.10) indicates that in the unit cell, j is not within 1-4 interactions of i.

In GIST-2016, both the electrostatic and LJ energies include only the interactions between pairs of particles with the shortest distance (using the minimum image convention). In this study, we modify the way in which the energies are calculated by using PME-based electrostatic energy calculations and long-range LJ corrections that account for all long-range interactions including the interactions between the particles in the central unit cell and all periodic images of those particles.

3.11 Electrostatic energy calculation

In GIST-2016, the electrostatic energy of the unit cell is estimated by

$$E_{elec} \simeq \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N'} k_e \frac{q_i q_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^{*}} \quad (11)$$

Of note, the asterisk indicates the distance between two atoms is calculated by minimum image convention; therefore, the electrostatic interactions longer than half the length of the unit cell's diagonal are not included in electrostatic energy in GIST-2016.

In PME, the total electrostatic energy (eq. 9) of the unit cell is broken down into a direct term, a reciprocal term, and a correction term (eqs. 12-15 [18]).

$$E_{elec} = E_{dir} + E_{rec} + E_{corr} \quad (12)$$

$$E_{dir} = \frac{1}{2} k_e \sum_{\boldsymbol{n}} \sum_{i}^{N} \sum_{j}^{N'} \frac{q_i q_j erfc(\beta |\boldsymbol{r_j} - \boldsymbol{r_i} + \boldsymbol{n}|)}{|\boldsymbol{r_j} - \boldsymbol{r_i} + \boldsymbol{n}|} \quad (13)$$

$$E_{rec} = \frac{1}{2} \sum_{i}^{N} q_i \Psi_{rec}(\boldsymbol{r_i}) \quad (14)$$

$$E_{corr} = -\frac{1}{2} k_e \sum_{(i,j) \in M} \frac{q_i q_j erf(\beta |\boldsymbol{r_i} - \boldsymbol{r_j}|)}{|\boldsymbol{r_i} - \boldsymbol{r_j}|} - \frac{k_e \beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \quad (15)$$

Here, *erf()* is the error function, *erfc()* is the complementary error function, β is the Ewald coefficient, and $\Psi_{rec}(r)$ is the reciprocal potential at position of $\boldsymbol{r_i}$. which is calculated by using helPME library[38]. The first term on the right-hand side of eq. 15 is the correction for intramolecular interactions when i and j are within 1-4 interactions in molecule M.

3.12 LJ energy calculation

In the GIST-2016, the LJ energy of the unit cell is estimated by:

$$E_{lj} \simeq \sum_{i}^{N} \sum_{j}^{N'} \frac{A_{(i,j)}}{|\boldsymbol{r_j} - \boldsymbol{r_i}|^{*12}} - \frac{B_{(i,j)}}{|\boldsymbol{r_j} - \boldsymbol{r_i}|^{*6}} \quad (16)$$

As with the electrostatic energy (eq. 11), the asterisk here denotes that the distance between two atoms are calculated by the minimum image convention; therefore, interactions longer than half the length of the unit cell's diagonal are not included in the LJ energy in GIST-2016.

In PME-GIST, the total LJ energy is broken down into short-range and long-range correction terms (eq. 17). The short-range term accounts for the LJ interactions within a defined distance cutoff and the long-range correction term is an estimation of the LJ interactions beyond this cutoff.

$$E_{lj} = E_{lj_short} + E_{lj_corr} \quad (17)$$

$$E_{lj_short} = \sum_{i}^{N} \sum_{j}^{N'} \frac{A_{(i,j)}}{|\boldsymbol{r_j} - \boldsymbol{r_i}|^{12}} - \frac{B_{(i,j)}}{|\boldsymbol{r_j} - \boldsymbol{r_i}|^{6}} \quad (18)$$

$$E_{lj_corr} = -\frac{2}{3} \pi N \rho < B > r_{cut}^{-3} \quad (19)$$

$$< B >= \frac{2}{N(N-1)} \sum_{i}^{N} \sum_{j>i}^{N} B_{(i,j)} \quad (20)$$

In which, ρ is the number density of all species in the system and $<B>$ is the weighted average B parameter of all particle interaction pairs in the system.

GIST calculates the ensemble energy density of each voxel in the unit cell. To do this, for every frame, the energies of each atom are calculated and assigned to the corresponding voxels where the atoms occupy (for water molecules, all the energy is assigned to the voxel in which the oxygen atom resides). Averaging the voxel energies over all frames of the trajectory gives the estimate of the reported voxel energy. The energy densities of each voxel are the voxel energies divided by the voxel volume[11,12,23]. In GIST-2016, the assignment of every particle's energy to each voxel is straightforward given both the electrostatic (eq.11) and LJ energy (eq.16) are pairwise decomposable and the energy terms involve two atoms are evenly split then assigned

to the two corresponding voxels. In PME-GIST, the energy terms involving two atoms (eq.13 and the first term on the right-hand side of eq.15) are evenly divided between the two atoms and then assigned to the voxels that those atoms occupy. The terms that have only one atom index (the reciprocal term and second term on right-hand side of eq. 15) are simply assigned to voxel where the atom resides.

3.2 Comparison of energies calculated by GIST and Amber-PMEMD

Figure 1 shows a comparison between energy densities computed using PME-GIST and GIST-2016 for a neat system of 922 water molecules at a temperature of 300K and 1 bar of pressure. The average and standard deviation of the voxel energy densities calculated by PME-GIST and GIST-2016 are -0.2824 ± 0.0811 and -0.2822 ± 0.0810 kcal/mol$Å^3$. The relative differences in the energy densities for the neat system computed by PME-GIST and GIST-2016 are within 2% of each other for the majority of voxels (98%) and the average unsigned relative difference is 0.54%. We also compare the GIST-2016 and PME-GIST calculated voxel water energies in these solute-water systems, for the 32 solute-water systems we study here, the average unsigned relative difference in voxel water energy calculated by GIST-2016 and PME-GIST is 0.45%.

While these values are relatively small on a per voxel basis, when the energies are summed over all voxels to compute the total energy for the system, the difference is significant (6.5 kcal/mol). Importantly, PME-GIST system energies are consistent with Amber-PMEMD with a total difference of 0.019 kcal/mol (0.0002%) for the system of 922 water molecules (**Table 1**).

	Ensemble energy	Difference to Amber-PMEMD
GIST-2016	-8791.425	6.519
PME-GIST	-8797.925	0.019
Amber-PMEMD	-8797.944	0

Table 1. Ensemble energy (Kcal/mol) of the entire pure water system calculated by different methods. The system comprises 922 water molecules, The GIST calculations analyzed the results of a 100 ns trajectory (50,000 frames) generated by the production phase of a neat water MD simulation.

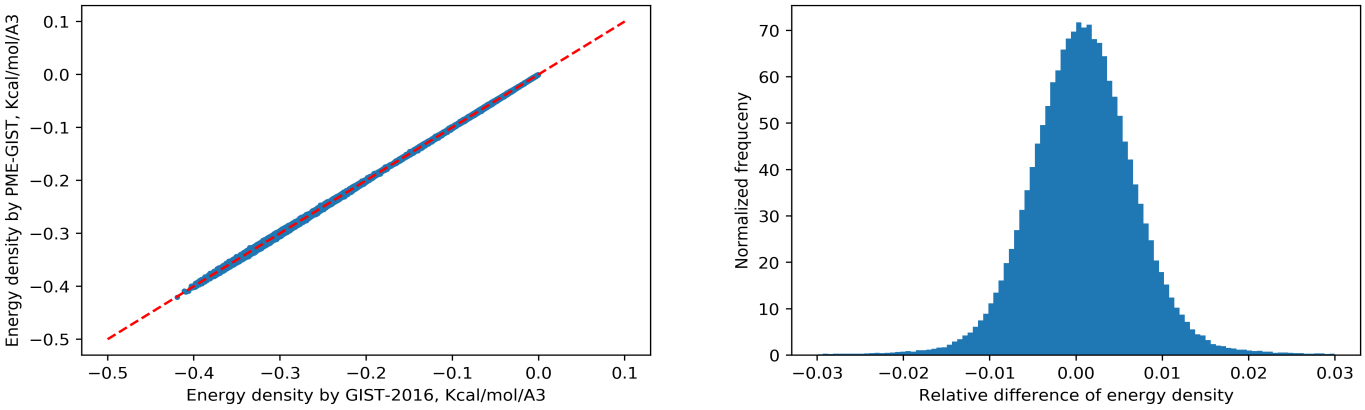

Figure 1. GIST and PME-GIST voxel energy densities (Kcal/mol/$Å^3$) for neat water. Left: Voxel energy densities calculated by GIST-2016 and PME-GIST. Each blue dot represents one of the 249,242 voxels in the system. Right: The relative differences in voxel energy densities calculated by GIST-2016 and PME-GIST. The relative difference was obtained by taking the difference between the voxel energy densities calculated by PME-GIST and GIST and dividing by the GIST-calculated voxel energy.

3.3 Speed of PME-GIST

In GIST-2016, the potential energy of atom i is computed by looping over the electrostatic and LJ interactions between atom i and all other atoms j in the system. This is a time consuming N^2 loop which resulted in the energy calculations taking approximately 90% of the total GIST computation time. In contrast, the computational complexity of the PME energy calculations in PME-GIST is $NlogN$. This significantly speeds up the GIST energy computation time. Here, we tested the speed of PME-GIST on a solvated monomer and dimer of SARS Cov-2 main protease (PDB ID: 6W63 [39]). The energy calculation in PME-GIST was sped up 27- to 65-fold depending on the system size.

System	6W63 mono	6W63 dimer
# Residues[a]	306	612
# Atoms[b]	50951	83566
GIST-2016[c]	40.15 s	146.25 s
GPU-GIST[d]	0.33 s	1.10 s
PME-GIST[c]	1.47 s [27]	2.23 s [65]

Table 2. Computation times (per frame) of PME-GIST vs. GIST-2016.
a) Number of protein residues
b) Total number of atoms in the system
c) GIST-2016 and PME-GIST were performed on a single Intel i7-5820 core, the number in the bracket shows the speedup of PME-GIST to GIST-2016.
d) GPU-GIST[16] was performed on Nvidia GTX1080Ti (all cores).

Energy calculations were conducted on 100 frames, and the times shown are the average times per frame.

3.4 Small molecule solvation free energies calculated by TI

Solvation free energies are typically computed by alchemical free energy methods, which require simulations over a series of nonphysical intermediate states.[21,40,41] In contract, IST can yield estimates of both the solvation energy and entropy and differences in end-states from a single simulation of each end-state. Although serval studies have applied IST and GIST for solvation free energy calculations, long-range interactions have not been included in these works, and high order entropies have been either approximated or neglected completely.[15,16,40] Here, in order to investigate how the inclusion of long-range interactions affect GIST's accuracy in calculating solvation energy ($\Delta Esolv$), we computed solvation energies using GIST-2016 and PME-GIST and then compared them to the solvation energies obtained by TI.

We first used TI to calculate the free energy, entropy, and energy of solvation of 32 small molecules. The molecules were chosen to cover a wide range of solvation free energies referenced by the FreeSolv database.[21] As shown in **Figure 2**, the solvation free energies we obtained from TI were consistent with the values reported in the FreeSolv database (R^2 = 0.95) and correlated comparably as well as FreeSolv to the experimental values (R^2 = 0.90). For the entire set of molecules, the average standard errors in solvation free energy, $T\Delta S$ and energy are 0.06 kcal/mol, 0.65 kcal/mol and 0.65 kcal/mol. The full set of TI calculated solvation free energies, entropies, and energies for all 32 compounds can be found in **Table S1.**

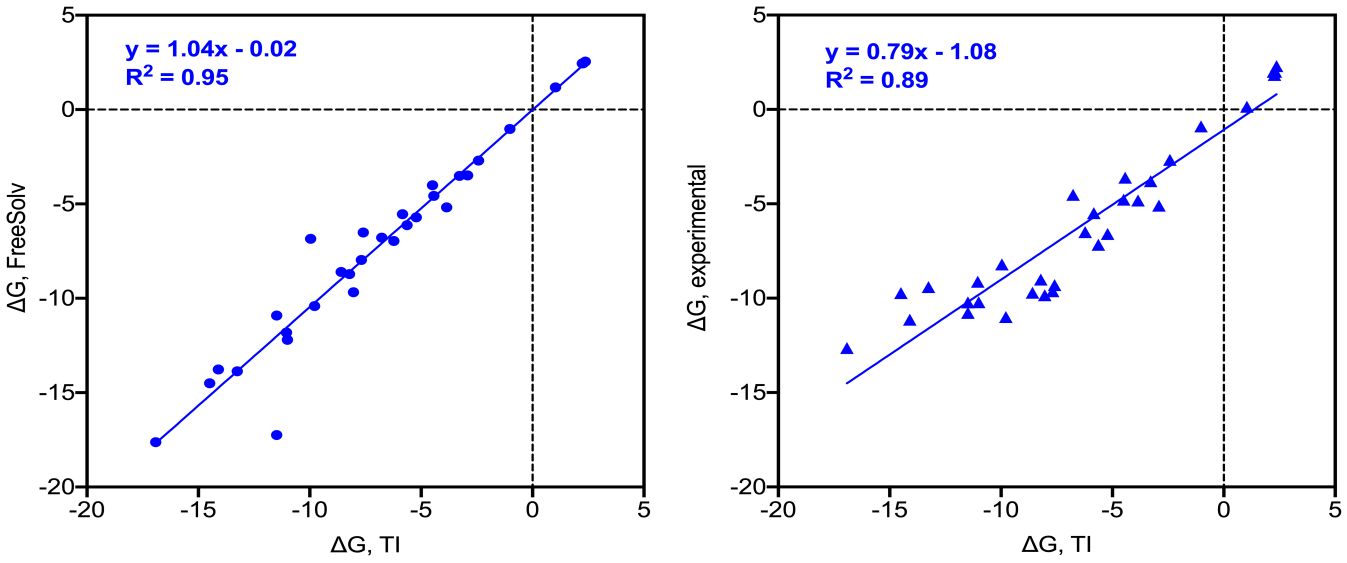

Figure 2. The solvation free energy (kcal/mol) of 32 molecules calculated by TI. Left: correlation of solvation free energies calculated by TI vs. values in FreeSolv; Right: correlation of solvation free energies calculated by TI vs. experimental values. Each dot represents the data for a molecule. The solid lines represent linear regression fits to the data for which the equations are shown.

To assess the accuracy of solvation energy (ΔE_{solv}) calculated by PME-GIST and GIST, we compared the solvation energies to that obtained by TI. As shown in **Figure 3**, PME-GIST gave solvation energy values closer to those of TI than GIST did. The average difference between solvation energies calculated by GIST-2016 and TI is -3.8 ± 0.7 kcal/mol (unsigned average difference: 3.8 ± 0.7 Kcal/mol), in contrast, the average difference decreases to 0.1 ± 0.7 kcal/mol (unsigned average difference: 0.8 ± 0.7 kcal/mol) in PME-GIST vs TI. the ΔH values calculated by GIST are almost always more negative than those of TI, suggesting that GIST has a systematic error in its solvation energy calculation; PME-GIST corrected this error. Interestingly, the solvation energies calculated in GIST-2016 and PME-GIST are both highly correlated to that calculated by TI. (**Figure 4**). The solvation energies for 32 molecules calculated by GIST and PME-GIST can be found in **Table S2**.

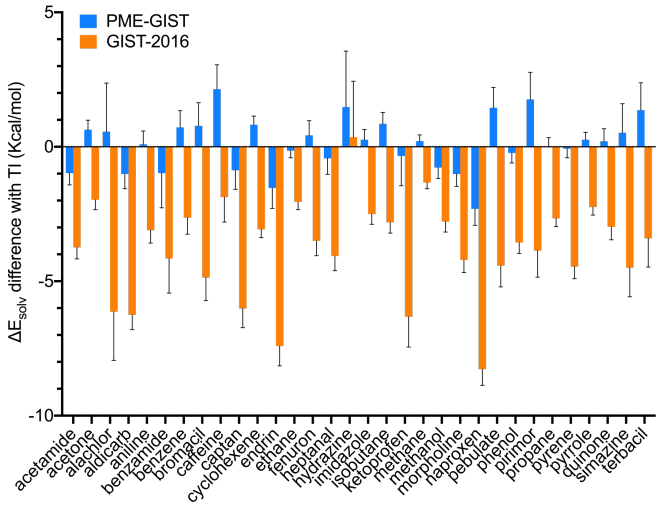

Figure 3. Solvation energy calculated by GIST-2016, PME-GIST compared to TI. The names of the 32 molecules are displayed on the x-axis. Bars represent mean values, and error bars represent standard errors.

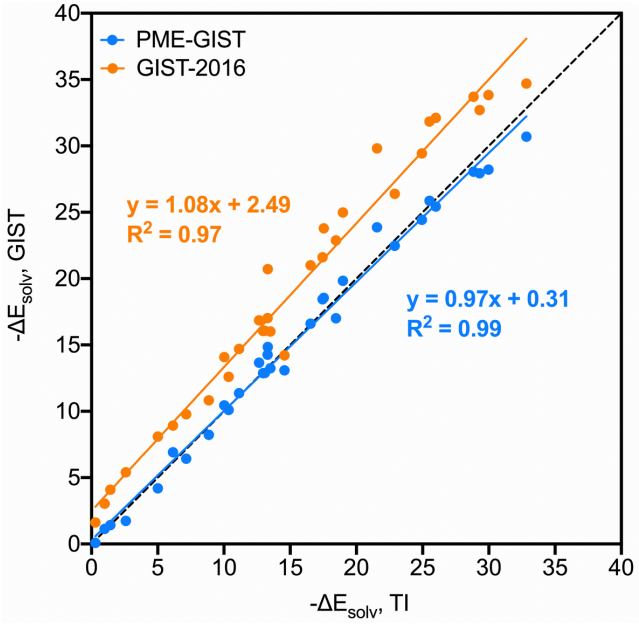

Figure 4. Correlation of solvation energy (ΔE_{solv}) calculated by PME-GIST, GIST-2016 with TI. The average standard error of ΔE_{solv} calculated by PME-GIST and GIST-2016 is 0.18 kcal/mol and 0.19 kcal/mol, respectively. The black dashed diagonal line corresponds to perfect agreement between the GIST and TI. Solvation energies calculated by PME-GIST and GIST-2016 are shown in blue and orange, respectively. Each dot represents data for a molecule. The solid lines represent the linear regression fit for each data set with equations shown on the graph.

3.52 GIST end states solvation entropy calculations

The non-ideal part of IST[6,19] entropy for an infinitely dilute solvated rigid solute can be written as:

$$S = S_{sw} + S_{sww} + O(3) + \ldots \qquad (21)$$

In which:

$$S_{sw} = -\frac{k_B \rho^0}{\Omega} \int g_{sw}(\boldsymbol{r},\omega) \ln g_{sw}(\boldsymbol{r},\omega) d\boldsymbol{r} dw \qquad (22)$$

$$S_{sww} = -\frac{1}{2} k_B \frac{\rho^0}{\Omega^2} \int g_{sw}(\boldsymbol{r},\omega) g_{sw}(\boldsymbol{r}',\omega') \{g_{sww}(r,r',\omega,\omega') \ln g_{sww}(r,r',\omega,\omega')\} d\boldsymbol{r} d\boldsymbol{r}' d\omega d\omega' \qquad (23)$$

Where k_B is Boltzmann's constant, ρ^0 is number density of bulk water, Ω is the total volume of the orientational space (for a water molecule, $\Omega = 8\pi^2$); $\boldsymbol{r}$ and ω are the position and orientation variables of the water molecule, respectively. $g_{sw}(\boldsymbol{r},\omega)$ is the solute-water corelation funtion and $g_{sww}(r,r',\omega,\omega')$ is the solute-water-water correlation. Due to both computational expense and difficulities with sparse sampling in calculating the 2nd order (S_{sww}) and high order entropy terms, only the 1st order entropy term which accounts for the soute-water correlation entropy is calculated in GIST (eq. 22).

In this study, we break the solvation entropy into:

$$\Delta S_{solv} = \Delta S_{sw} + \Delta O(2) \qquad (24)$$

where, ΔS_{sw} is the solute-water entropy change of two states and solute-water entropy equals to 0 in the initial state. $\Delta O_{(2)}$ is difference of 2nd and higher order entropy terms between the final state and initial state. The solute-water entropy values for 32 small molecules can be found in **Table S3**. Subtracting the solute-water entropy calculated by GIST from the solvation entropy calculated by TI gives a direct estimate of the solvation entropy contributed by high order entropy terms (thus $\Delta O(2)$).

Figure 5 shows the correlation between GIST calculated entropies and those computed from TI. While the curves are highly correlated ($R^2 = 0.92$), the slope, 0.60, is significantly different from 1. A simple approximation for the higher order terms

($\Delta O(2) = -0.4\ \Delta S_{sw}$) changes the slope to be close to 1. With this simple correction for $\Delta O(2)$, as shown in **Figure 6**, the solvation free energies calculated by PME-GIST are highly consistent with those calculated by TI (R^2 = 0.99, mean unsigned difference = 0.4 kcal/mol) and with experimental results (R^2 = 0.88 and mean unsigned difference = 1.4 kcal/mol) (**Figure7**).

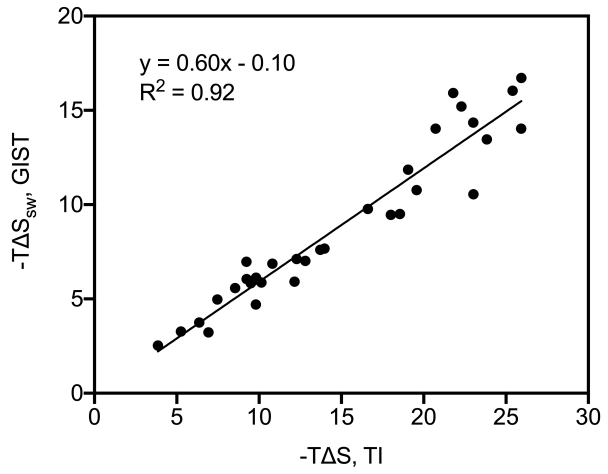

Figure 5. The correlation between solute-water entropy and solvation entropy. Each dot represents a molecule, and the solid line is the linear regression line whose equation is written on the graph. The average standard errors of $T\Delta S_{sw}$ and $T\Delta S$ are 0.13 kcal/mol and 0.65 kcal/mol, respectively.

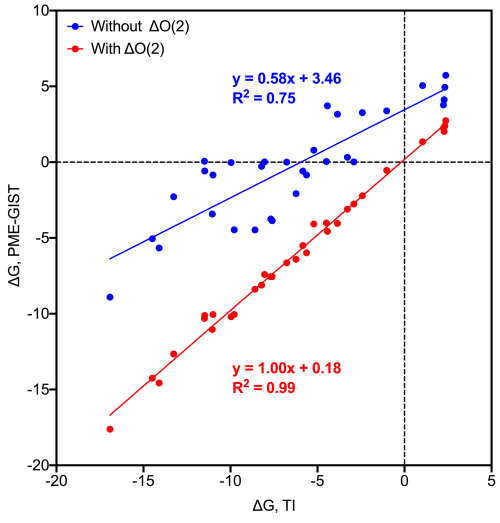

Figure 6. Solvation free energies calculated by PME-GIST, with or without $\Delta O(2)$ correction, compared to TI. The mean standard errors of solvation free energy calculated by PME-GIST and TI are 0.26 kcal/mol and 0.06 kcal/mol, respectively. The mean unsigned difference in solvation free energies calculated by PME-GIST and TI are 0.4 kcal/mol (with $\Delta O(2)$ correction) and 6.2 kcal/mol (without $\Delta O(2)$ correction). Each dot represents the data of a molecule, and the solid lines represent the linear regression lines whose equations are written on the graph.

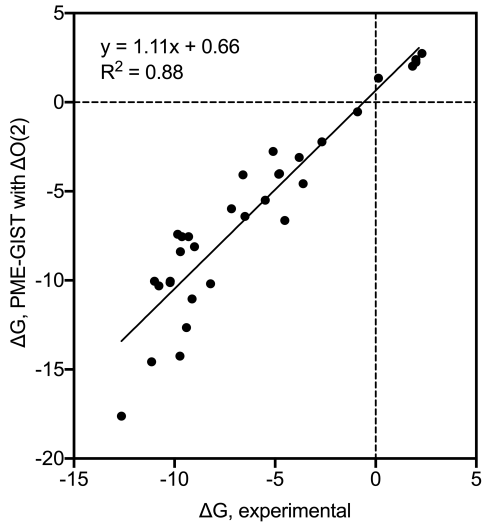

Figure 7. Solvation free energies calculated with PME-GIST (with $\Delta O(2)$ correction) compared to experimental results. Each dot represents the data of a molecule, and the solid line is the linear regression line whose equation is written on the graph.

3.7 The perturbation of water molecules upon solvation of small molecules

In contrast to TI and FEP methods, which calculate the free energy change for the entire system, PME-GIST provides information on the spatial distribution of the free energy contributions to the molecular solvation. It is often convenient to investigate the thermodynamics of a region such as the binding site of a protein. Here, to investigate whether the total solvation free energy can be well estimated by considering only the voxels proximal to the solute, we estimate the

contribution by integrating the energies and entropies over the voxels within various distances from the heavy atoms of each of the solutes. This is illustrated for the benzene-water system in **Figure 8**. In this figure, the calculated solvation energies and entropies flatten once a region extending within approximately 10 Å of the solute[44]. The perturbated region can also be visualized around the benzene molecule (**Figure 9**).

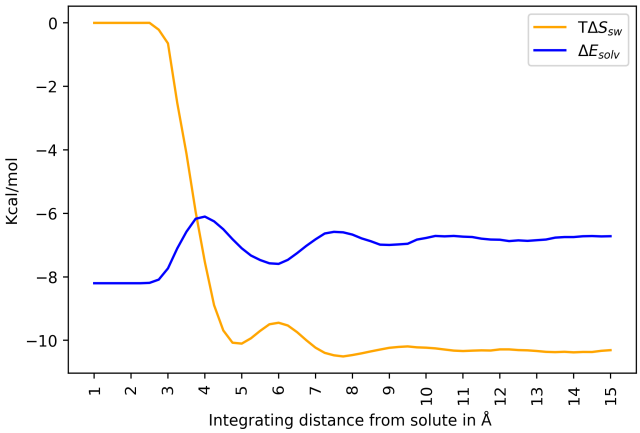

Figure 8. Perturbation of water molecules in regions of increasing distance around the solute. The solvation enthalpies and solute-water entropies were calculated by integrating the voxels within an increasing radius (1 Å – 15 Å) from the heavy atoms of benzene. The initial value (about -8 kcal/mol) reflects the energy change on benzene upon solvation; both curves are initially flat because there are no water molecules within ~3 Å of the benzene molecule.

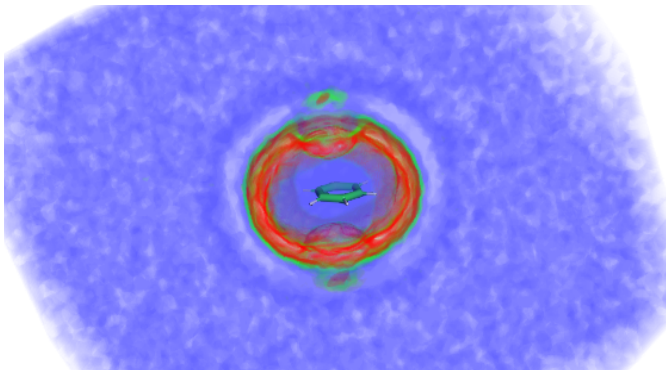

Figure 9. A GIST map of the solvation entropy densities around benzene. The benzene is represented by the CPK model at the center of the box. Voxels with unfavorable entropies are colored in red (more unfavorable) and green (less unfavorable). The blue voxels are those with entropies close to the bulk value.

3.9 The decomposition of solvation free energy

GIST calculates energy and entropy terms separately in a single simulation; this can provide additional insight into the solvation process. For solvation energy, given that PME-GIST decomposes the system energy to the level of each atom, the respective energy changes of solute atoms and water molecules can also be examined. Interestingly, as shown in **Figure 10**, favorable interactions between the solutes and water molecules constitutes the main contribution to the solvation energies of the 32 small molecules. We also observe compensation between energy and entropy for each molecule (**Figure 11**).

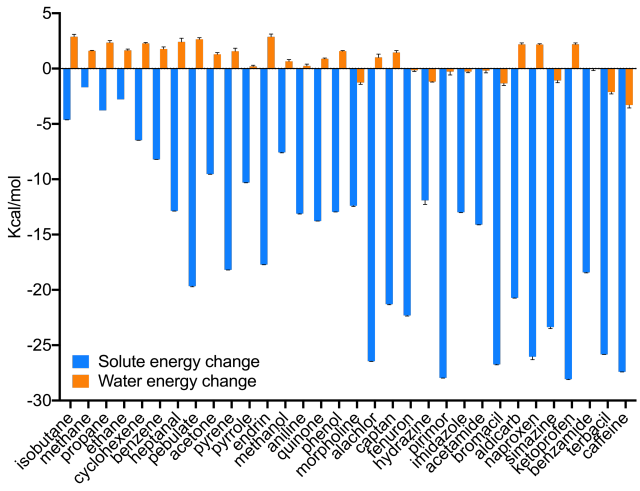

Figure 10. The decomposition of ΔE_{solv} to the energy changes of solute and water molecules. The molecules are ranked from hydrophobic to hydrophilic by the experimental free solvation energy. Bars represent mean values, and the error bars correspond to standard errors.

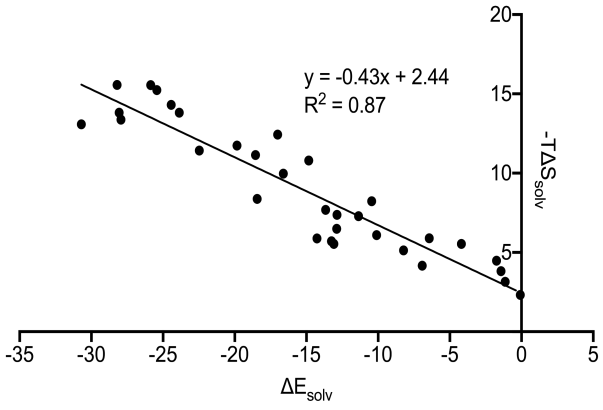

Figure 11. The energy-entropy compensation in solvation of small molecules. Each dot represents the data for a molecule, and the solid line is the linear regression fit to this data whose equation is written on the graph. The solvation energy and entropy values are calculated by PME-GIST with the $\Delta O(2)$ correction added.

4 Conclusions

In this study, we developed the methodology and implemented PME-based electrostatic treatment and long-range LJ corrections into GIST. This increased the accuracy of GIST energetic calculations to fully agree with those of modern MD engines. With its precise energy calculations, we applied PME-GIST end states analyses to measure the solvation energies of small molecules and obtained results highly consistent with the alchemical TI method without the need to simulate many intermediate states. We also examined quantitatively the entropy contributions from higher order terms that are truncated in GIST, by comparing the solute-water entropy from GIST with the solvation entropy from TI calculations. While the GIST solute-water entropy is highly correlated with the total solvation entropy from TI across a wide range of small molecules, the higher order entropy is found to account for about 40 % of the total solvation entropy for these small molecule solutes. Using a simple correction for higher order entropy together with the enhanced accuracy in the energy calculation the PME-GIST end-states calculations can yield solvation free energies in agreement with those calculated by the TI method to within 0.2 kcal/mol. Furthermore, we also investigated the perturbation of water molecules upon solvation by varying the GIST integration region surrounding the solutes. The convergence of solvation energy and entropy at approximately 10 Å from the solutes suggests that the perturbation to water molecules extends to about three shells and the water molecules beyond this distance make negligible contributions to the solvation free energy. Our results also support the local approximations that are often used in displaced solvent functionals[10,45–47]. The utility of PME-GIST end-state method in calculating solvation free energies as demonstrated here could potentially extend the application of IST to free energy calculations of more complex scenarios such as drug-target binding.

Keywords: GIST, Solvation Free Energy, Inhomogeneous Solvation Theory, Thermodynamic Integration

Acknowledgments

We thank Drs. Emilio Gallicchio for his insightful comments regarding this project. This work was supported by NIH R01-GM100946.

References

(1) 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), 1365–1370.

(2) Wallace, D. C. Statistical Theory for the Entropy of a Liquid. *Phys. Rev. A* **1989**, *39* (9), 4843–4847.

(3) Baranyai, A.; Evans, D. J. Direct Entropy Calculation from Computer Simulation of Liquids. *Phys. Rev. A* **1989**, *40* (7), 3817–3822.

(4) Morita, T.; Hiroike, K. *A New Approach to the Theory of Classical Fluids. III-General Treatment of Classical Systems*; 1961; Vol. 25.

(5) Raveché, H. J. Entropy and Molecular Correlation Functions in Open Systems. I. Derivation. *J. Chem. Phys.* **1971**, *55* (5), 2242–2250.

(6) Lazaridis, T. *Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory*; 1998.

(7) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102* (18), 3542–3550.

(8) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. *Methods Mol. Biol.* **2012**, *819*, 393–404.

(9) 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.

(10) 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.

(11) 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), 973–980.

(12) Nguyen, C.; Gilson, M. K.; Young, T. Structure and Thermodynamics of Molecular Hydration via Grid Inhomogeneous Solvation Theory. **2011**.

(13) 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.

(14) Raman, E. P.; MacKerell, A. D. Rapid Estimation of Hydration Thermodynamics of Macromolecular Regions. *J. Chem. Phys.* **2013**, *139* (5).

(15) 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.

(16) Kraml, J.; Kamenik, A. S.; Waibl, F.; Schauperl, M.; Liedl, K. R. Solvation Free Energy as a Measure of Hydrophobicity: Application to Serine Protease Binding Interfaces. *J. Chem. Theory Comput* **2019**, *15*, 38.

(17) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N⋅Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 5648.

(18) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *A Smooth Particle Mesh Ewald Method*; 1995.

(19) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. Accurate and Efficient Corrections for Missing Dispersion

Interactions in Molecular Simulations. *J. Phys. Chem. B* **2007**, *111* (45), 13052–13063.
(20) DA, C.; TE, C.; T, D.; H, G.; R, L.; KM, M.; A, O.; C, S.; B, W.; RJ, W. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16).
(21) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data* **2017**, *62*.
(22) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095.
(23) Nguyen, C. N.; Kurtzman, T.; Gilson, M. K. Spatial Decomposition of Translational Water-Water Correlation Entropy in Binding Pockets. *J. Chem. Theory Comput.* **2016**, *12* (1), 414–429.
(24) Lazaridis, T. Solvent Reorganization Energy and Entropy in Hydrophobic Hydration. *J. Phys. Chem. B* **2000**, *104* (20), 4964–4979.
(25) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation Thermodynamic Mapping of Molecular Surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37* (21), 2029–2037.
(26) Ihlenfeldt, W. D.; Bolton, E. E.; Bryant, S. H. The PubChem Chemical Structure Sketcher. *J. Cheminform.* **2009**, *1* (1), 20.
(27) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23* (16), 1623–1641.
(28) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
(29) 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), 926–935.
(30) Basconi, J. E.; Shirts, M. R. Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9* (7), 2887–2899.
(31) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2019**, *1* (1).
(32) Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and Flexible Gpu Accelerated Binding Free Energy Calculations within the Amber Molecular Dynamics Package. *J. Comput. Chem.* **2018**.
(33) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry*. J Comput Chem December 2005, pp 1668–1688.
(34) Vukovic, S.; Brennan, P. E.; Huggins, D. J. Exploring the Role of Water in Molecular Recognition: Predicting Protein Ligandability Using a Combinatorial Search of Surface Hydration Sites. *J. Phys. Condens. Matter* **2016**, *28* (34).
(35) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One-and Two-Step Transformations. *J. Comput. Chem.* **2011**, *32* (15), 3253–3263.
(36) Levy, R. M.; Gallicchio, E. COMPUTER SIMULATIONS WITH EXPLICIT SOLVENT: Recent Progress in the Thermodynamic Decomposition of Free Energies and in Modeling Electrostatic

Effects. *Annu. Rev. Phys. Chem.* **1998**, *49* (1), 531–567.
(37) Wan, S.; Stole, R. H.; Karplus, M. Calculation of the Aqueous Solvation Energy and Entropy, as Well as Free Energy, of Simple Polar Solutes. *J. Chem. Phys.* **2004**, *121* (19), 9539–9548.
(38) Simmonett, A. C.; Burns, L. A.; Roe, D. R.; Brooks, B. R. andysim/helpme: helPME: an efficient library for particle mesh Ewald https://github.com/andysim/helpme (accessed Oct 16, 2020).
(39) RCSB PDB - 6W63: Structure of COVID-19 main protease bound to potent broad-spectrum non-covalent inhibitor X77 https://www.rcsb.org/structure/6w63 (accessed Oct 9, 2020).
(40) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Chapter 4 Alchemical Free Energy Calculations: Ready for Prime Time? *Annual Reports in Computational Chemistry*. Elsevier BV January 1, 2007, pp 41–59.
(41) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *Journal of Chemical Physics*. American Institute of Physics December 21, 2012, p 230901.
(42) Kraml, J.; Hofer, F.; Sophia Kamenik, A.; Waibl, F.; Kahler, U.; Schauperl, M.; R. Liedl, K. Solvation Thermodynamics in Different Solvents – Water-Chloroform Partition Coefficients from Grid Inhomogeneous Solvation Theory. *J. Chem. Inf. Model.* **2020**, *0* (ja).
(43) Huggins, D. J. Estimating Translational and Orientational Entropies Using the K - Nearest Neighbors Algorithm. *J. Chem. Theory Comput.* **2014**.
(44) Lee, S. H.; Rossky, P. J. A Comparison of the Structure and Dynamics of Liquid Water at Hydrophobic and Hydrophilic Surfaces - A Molecular Dynamics Simulation Study. *The Journal of Chemical Physics*. American Institute of PhysicsAIP February 15, 1994, pp 3334–3345.
(45) Hüfner-Wulsdorf, T.; Klebe, G. Protein-Ligand Complex Solvation Thermodynamics: Development, Parameterization, and Testing of GIST-Based Solvent Functionals. *J. Chem. Inf. Model.* **2020**, *60* (3), 1409–1423.
(46) Uehara, S.; Tanaka, S. AutoDock-GIST: Incorporating Thermodynamics of Active-Site Water into Scoring Function for Accurate Protein-Ligand Docking. *Molecules* **2016**, *21* (11).
(47) Balius, T. E.; Fischer, M.; Stein, R. M.; Adler, T. B.; Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T.; Shoichet, B. K. Testing Inhomogeneous Solvation Theory in Structure-Based Ligand Discovery. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (33), E6839–E6846.