

Thermodynamic Decomposition of Solvation Free Energies with Particle Mesh Ewald and Long-Range Lennard-Jones Interactions in Grid Inhomogeneous Solvation Theory

Lieyang Chen, Anthony Cruz, Daniel R. Roe, Andrew C. Simmonett, Lauren Wickstrom, Nanjie Deng, and Tom Kurtzman*



Cite This: <https://doi.org/10.1021/acs.jctc.0c01185>



Read Online

ACCESS |



Metrics & More

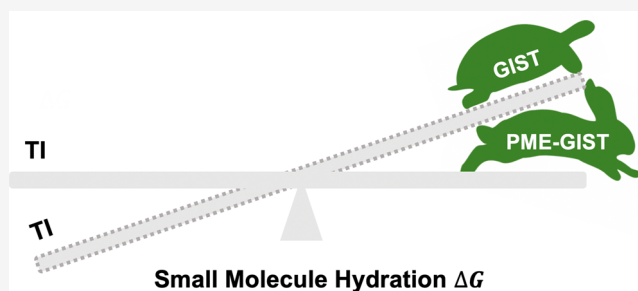


Article Recommendations



Supporting Information

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 nonbonded 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-state 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 second 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.



1. INTRODUCTION

Inhomogeneous solvation theory (IST)^{1–7} provides a statistical mechanical framework to calculate the thermodynamics of solvation from solvent molecular density distributions. Building upon IST, we developed WaterMap⁸ and GIST^{9,10} 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.^{11–13} 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, a couple of studies^{13–15} have used IST to estimate free energies of solvation for small molecules and amino acids. However, discrepancies between how GIST tools estimate energies and how molecular dynamics (MD) engines calculate them, particularly differences in the treatment of long-range interactions, have prevented precise comparison between GIST-calculated energies and the energies produced from

Received: November 11, 2020

MD free energy methods such as TI and FEP. Here, we address this by presenting a formalism by which PME-based electrostatic energies^{16,17} and long-range LJ energies¹⁸ 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^{19,20} 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 energies of 32 small molecules chosen from the FreeSolv data set.²¹ These small molecules range from hydrophobic to hydrophilic and have solvation free energies nearly evenly distributed over a range of 20 kcal/mol. For each solute, we ran 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 led 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 GIST²² in the MD analysis software CPPTRAJ,²³ 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.^{24–26} 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) are highly correlated with TI, we find that PME-GIST significantly improves the agreement. [In the text, we will use the term “GIST-2016” to refer to the original GIST implementation in 2016 and the term “PME-GIST” to refer to the current version that we developed in this study.] The first 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.

These 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 neutral small molecules were obtained from the PubChem database.²⁷ The atomic partial charges of the molecules were assigned using AM1-BCC in Antechamber,²⁸ and other force field parameters were parametrized by the GAFF2 force field.²⁹ The solutes were then solvated in a TIP3P³⁰ rectangular water box, where the box's edges are at least 15 Å from the solutes using tleap from AmberTool20²⁰ such that the solvated systems have between 1191 and 2004 water molecules. 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.^{31,32} 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 with a collision frequency of 2 ps^{−1}.³³ The MD simulations were performed using GPU-accelerated PMEMD^{34,35} from Amber20 with periodic boundary conditions. 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 simulations were conducted for 100 ns with a time step of 2 fs. System configurations were saved every 1000 steps for subsequent GIST analysis. All solute atoms were harmonically restrained to the initial positions and at the center of the systems by a 100 kcal/mol/Å² force constant throughout all simulation steps. The lengths of bonds involving hydrogen were constrained by the SHAKE algorithm³⁶ after minimization steps.

2.2. TI Calculations. The solvation free energy is calculated by Amber TI^{37,38} as

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

where $\lambda = 0$ represents the solvated state in which the solute fully interacts with the surrounding water molecules, and $\lambda = 1$ 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 2 ns production run 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 $\alpha_{\text{LJ}} = 0.5$ and $\beta_{\text{C}} = 12$.³⁹ TI simulations were run for each solute at three different temperatures (270, 300, and 330 K) with three runs at each temperature for which the velocities were randomly assigned prior to each 250 ps equilibration. Solvation entropies were then estimated using the finite-difference method,^{40,41} 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. 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 entire systems with a resolution of 0.125 Å³ for each voxel were

centered on the solutes with a bulk water number density of $0.03288 \text{ mol}/\text{\AA}^3$. In the PME-GIST method, the PME-based electrostatic energy calculation and long-range LJ correction were turned on, and a cutoff of $9 \text{ \AA}$ 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 around the solute, the GISTPP tool²² was used to define the voxels belonging to that region, and summing over the energy density, entropy density, and water density multiplied by the voxel volume over those voxels yielded the total energy, total entropy, and total water number for that region, respectively.

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

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

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

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

We also implemented a systematic correction to the GIST first order entropy calculation that accounts for the systematic bias in the nearest neighbor algorithm^{42–44} that estimates this term. The first order entropy term for a neat fluid should be zero; however, the nearest neighbor algorithm will yield a value for this term slightly greater than zero due to limited simulation length.¹⁵ To account for this systematic bias, we ran TIP3P bulk water simulations in the same NPT conditions and for the same duration as for other solute systems. Conducting GIST calculations on the neat water simulations yielded a systematic bias of $0.00334 \pm 0.00003 \text{ kcal/mol}$ per water in the first order entropy term for the TIP3P water model.

2.4. End-State Calculation by GIST. In the end-state 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 which the rigid solute is fully decoupled from the solvent (eqs 5–7)

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

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

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

where ΔA_{solv} is the difference in the Helmholtz free energy between the two states, and ΔE_{solv} and $-T \Delta S_{\text{solv}}$ are the difference in energy and temperature times the entropy difference of two states, respectively. The prime in eq 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 nonbond interactions within the solute and between the solute and all of its periodic replicas and is calculated by the energy command in CPPTRAJ. The energies per neat water molecule ($-9.5353 \pm 0.0002 \text{ kcal/mol}$ and $-9.5422 \pm 0.0002 \text{ kcal/mol}$) are calculated by running GIST-2016 and PME-GIST, respectively, on 100 ns bulk water simulations in the same NPT conditions as the solute–water simulations. 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 as the $P \Delta V$ terms make negligible contributions to these systems.

3. RESULTS

3.1. Modifications to the Energy Calculation in GIST.

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

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

where the electrostatic energy of the cell can be written as

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

and the LJ energy for the system is

$$E_{\text{lj}} = \sum_n \sum_i \sum_j^N \frac{A_{(ij)}}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|^2} - \frac{B_{(ij)}}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|^6} \quad (10)$$

in which q_i and q_j are the atomic partial charges of two atoms i and j at positions $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively; N is the total particle number in the unit cell; and k_e is the Coulomb constant. $A_{(ij)}$ ($4\epsilon\sigma_{ij}^{12}$) and $B_{(ij)}$ ($4\epsilon\sigma_{ij}^6$) are the LJ parameters of two atoms; and $\mathbf{n}$ represents a vector between an atom in the unit cell and its image in a periodic cell. The sum over $\mathbf{n}$ 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.

3.1.1. Electrostatic Energy Calculation. In GIST-2016, the electrostatic energy of the unit cell is estimated by

$$E_{\text{elec}} \simeq \frac{1}{2} \sum_i \sum_j^N k_e \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|^*} \quad (11)$$

The asterisk indicates that 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.

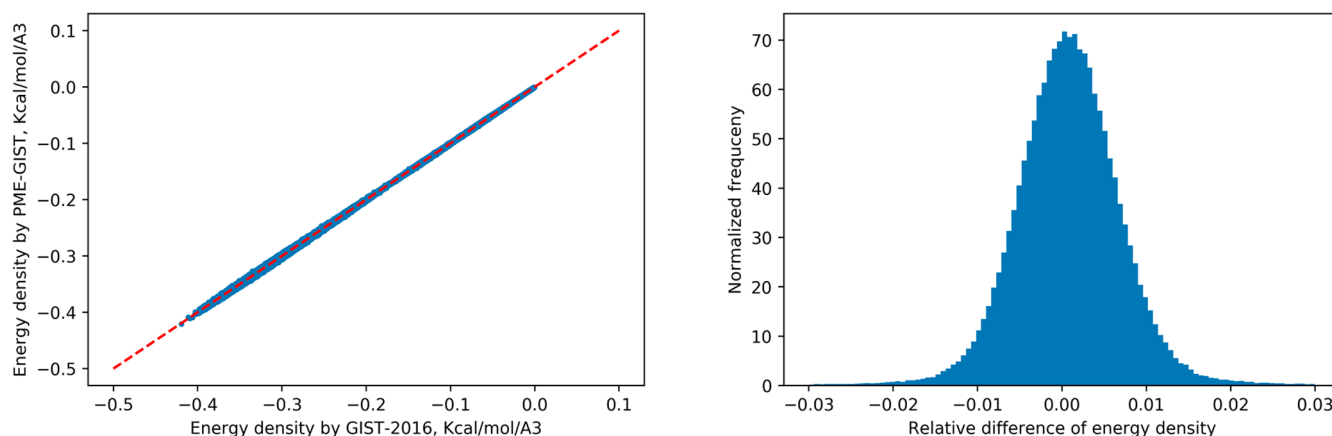


Figure 1. GIST and PME-GIST voxel energy densities (kcal/mol/Å³) for neat water. Left: Voxel energy densities calculated by GIST-2016 and PME-GIST. Each blue dot represents one of the 249,242 voxels in the system. Right: The relative differences in voxel energy densities calculated by GIST-2016 and PME-GIST. 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.

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^{16,17}).

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

$$E_{dir} = \frac{1}{2} k_e \sum_n \sum_i \sum_j \frac{q_i q_j \text{erfc}(\beta |r_j - r_i + nl|)}{|r_j - r_i + nl|} \quad (13)$$

$$E_{rec} = \frac{1}{2} \sum_i q_i \Psi_{rec}(r_i) \quad (14)$$

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

Here, erf() is the error function, erfc() is the complementary error function, β is the Ewald coefficient, and $\Psi_{rec}(r_i)$ is the reciprocal potential at position of r_i , which is calculated by using the helpME library.⁴⁵ 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 .

3.1.2. LJ Energy Calculation. In GIST-2016, the LJ energy of the unit cell is estimated by

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

As with the electrostatic energy (eq 11), the asterisk here denotes that the distance between two atoms is 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

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

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

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

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

in which ρ is the number density of all species in the system, and $\bar{B}$ 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 that 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. In GIST-2016, the assignment of each particle's energy to a voxel is straightforward given that both the electrostatic (eq 11) and LJ energies (eq 16) are pairwise decomposable, and the energy terms involving two atoms are evenly split and 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 in which the atom resides.

3.2. Comparison of Energies Calculated by GIST-2016, PME-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 300 K and 1 bar of pressure. The averages and standard deviations of the voxel energy densities calculated by PME-GIST and GIST-2016 are -0.2824 ± 0.0811 and -0.2822 ± 0.0810 kcal/mol/Å³, respectively. 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).

Table 1. Ensemble Energy (kcal/mol) of the Entire Pure Water System Calculated by Different Methods^a

	ensemble energy	difference with Amber-PMEMD
GIST-2016	−8791.425 (0.243)	6.519
PME-GIST	−8797.925 (0.198)	0.019
Amber-PMEMD	−8797.944 (0.198)	0

^aThe 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. The standard errors are indicated in parentheses.

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 $N \log N$. 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: 6W63⁴⁶). The energy calculation in PME-GIST was sped up 27- to 65-fold depending on the system size (Table 2).

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.³⁷ In contrast, 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 several 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.^{13–15} Here, in order to investigate how the inclusion of long-range interactions affects GIST's accuracy in calculating solvation energy (ΔE_{solv}), 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. 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 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, 0.65, and 0.65 kcal/mol, respectively. The full set of TI-calculated solvation free energies, entropies, and energies for all 32 compounds can be found in Table S1.

To assess the accuracy of solvation energy (ΔE_{solv}) calculated by PME-GIST and GIST, we compared the solvation energies to those 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 was -3.8 ± 0.7 kcal/mol (unsigned average difference: 3.8 ± 0.7 kcal/mol); in contrast, the average difference decreased 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 were almost always more negative than those of TI, suggesting that GIST has a systematic error in its solvation energy calculation. The systematic error observed in GIST-2016 arises from the minimal image convention algorithm for calculating pairwise energies. This algorithm imposes a system size dependence on the energy calculation, as longer-range interactions are calculated as the system size increases. In PME-GIST, this system size dependence is removed by implementation of the PME-based energy calculation. Interestingly, the solvation energies calculated in GIST-2016 and PME-GIST are both highly correlated to those calculated by TI (Figure 4). The solvation energies for 32 molecules calculated by GIST and PME-GIST can be found in Table S2.

3.5. GIST End-State Solvation Entropy Calculations. The nonideal part of IST⁶ entropy for an infinitely dilute solvated rigid solute can be written as

$$S = S_{\text{sw}} + S_{\text{sww}} + O(3) + \dots \quad (21)$$

in which

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

and

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

Table 2. Computation Times (per Frame) of PME-GIST vs GIST-2016

system	# residues ^a	# atom ^b	GIST-2016 ^c	GPU-GIST ^d	PME-GIST ^c
6W63_monomer	306	50951	40.15 s	0.33 s	1.47 s (×27)
6W63_dimer	612	83566	146.25 s	1.10 s	2.23 s (×65)

^aNumber of protein residues. ^bTotal number of atoms in the system, which consists of protein and water molecules. ^cGIST-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. ^dGPU-GIST¹⁴ was performed on Nvidia GTX1080Ti (all cores). Energy calculations were conducted on 100 frames, and the times reported are the average calculation time per frame.

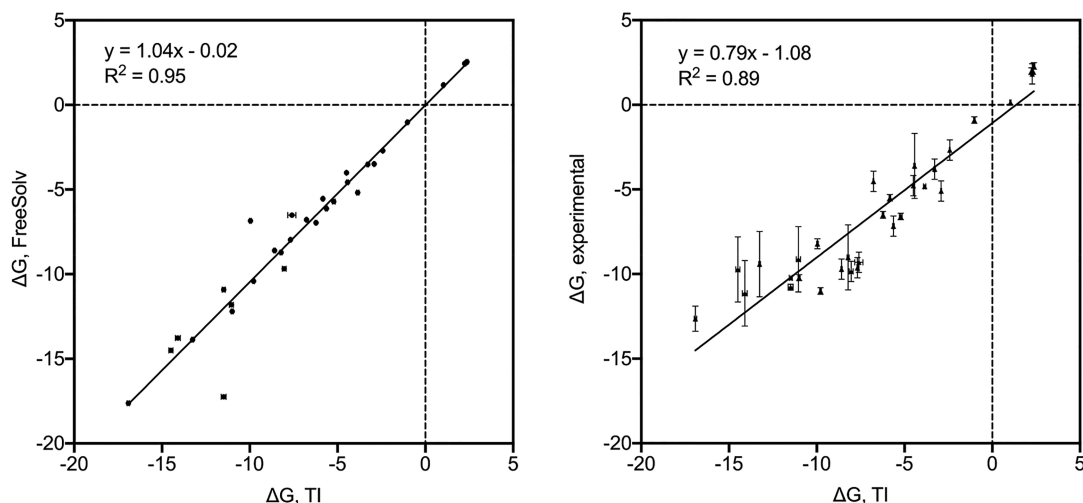


Figure 2. The solvation free energy (kcal/mol) of 32 molecules calculated by TI. Left: correlation of solvation free energies in FreeSolv vs calculated by TI; Right: correlation of solvation free energies reported in experimental data vs calculated by TI. Each dot represents the data for a molecule. The solid lines represent linear regression fits to the data for which the equations are shown. The experimental error bars were collected from the FreeSolv database, while the computational error bars are the standard errors from multiple simulations.

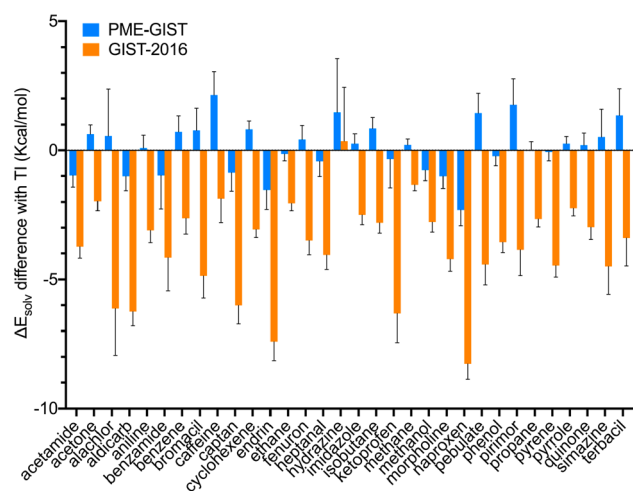


Figure 3. Solvation energy calculated by GIST-2016 or PME-GIST compared to TI. The names of the 32 molecules are displayed on the x-axis. Bars represent mean values, and error bars represent standard errors in the differences between the GIST calculations and TI. Both PME-GIST and GIST-2016 were performed on the entire system for each solute.

where k_B is Boltzmann's constant, ρ^0 is the number density of bulk water, Ω is the total volume of the orientational space (for asymmetric molecules such as water molecules, $\Omega = 8\pi^2$); and r and ω are the position and orientation variables of the water molecule, respectively. $g_{sw}(r, \omega)$ is the solute–water correlation function, and $g_{sww}(r, r', \omega, \omega')$ is the solute–water–water correlation. Due to both computational expense and difficulties with sparse sampling in calculating the second order (S_{sww}) and high order entropy terms, only the first order entropy term, which accounts for the solute–water correlation entropy, is calculated in GIST (eq 22).

In this study, we break down the solvation entropy into two terms

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

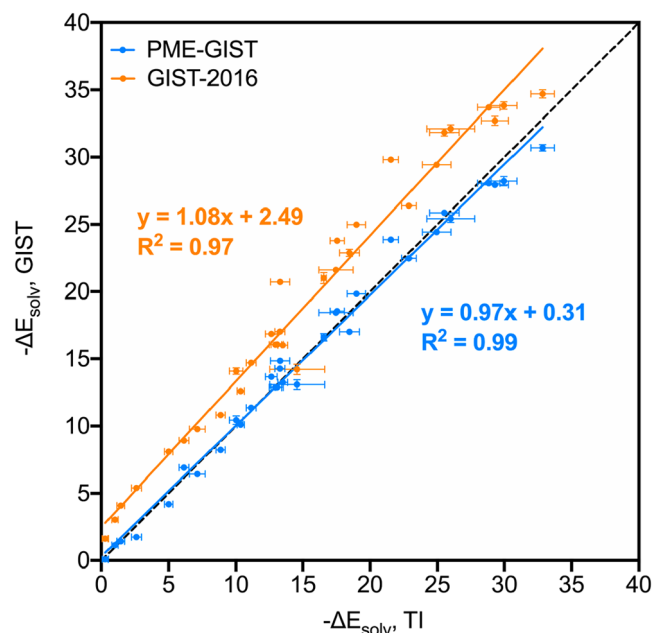


Figure 4. Correlation of solvation energy (ΔE_{solv}) calculated by PME-GIST and GIST-2016 with TI. The average standard error of ΔE_{solv} calculated by PME-GIST and GIST-2016 is 0.18 and 0.19 kcal/mol, respectively. The black dashed diagonal line corresponds to perfect agreement between GIST and TI. 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. The error bars reflect the standard error of the calculations.

where ΔS_{sw} is the solute–water entropy change of two states, and solute–water entropy equals 0 in the initial state. $\Delta O(2)$ is the difference in second 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 $\Delta O(2)$).

The correlation between GIST-calculated entropies and those computed from TI is shown in Figure 5. While the

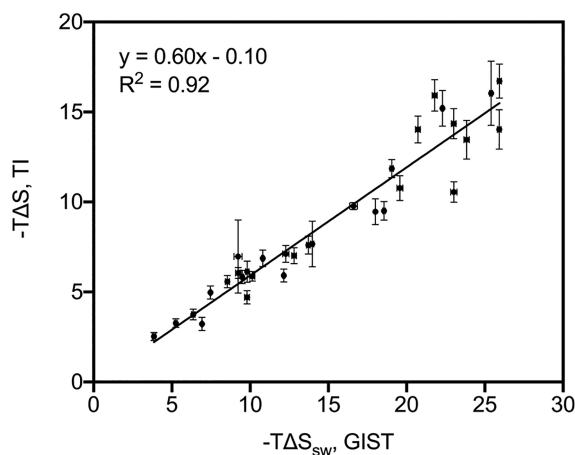


Figure 5. The correlation between solvation entropy and solute–water entropy. Each dot represents data for a molecule, and the solid line is the linear regression line whose equation is written on the graph. The average standard errors of $T\Delta S_{sw}$ and $T\Delta S$ are 0.13 and 0.65 kcal/mol, respectively.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

curves are highly correlated ($R^2 = 0.92$), the slope, 0.60, is significantly different than 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) (Figure 7). While the simple scaling factor (−0.4) we report here works well in this study of small molecules, for macromolecules such as proteins with more complex surface topographies, the relationship between solute–water entropy and higher order entropy terms deserves further investigation.

3.6. 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 estimated well by considering only the voxels proximal to the solute, we estimate the contribution by integrating the energies and entropies over only voxels that are within various distances of the heavy atoms of each of the solutes. This is illustrated for the benzene–water system as a representative example (Figure 8 and Figure 9). As shown in Figure 8, the largest perturbation of surrounding water molecules occurs within 5 Å of the solvated benzene molecule, while the calculated solvation energies and entropies flatten after the region is extended to approximately 10 Å from the solute.⁴⁷ A similar pattern is observed for all solutes (see the SI) which is consistent with similar observations in a study using the density-functional approach.⁴⁸ We can also estimate the free energy of solvation

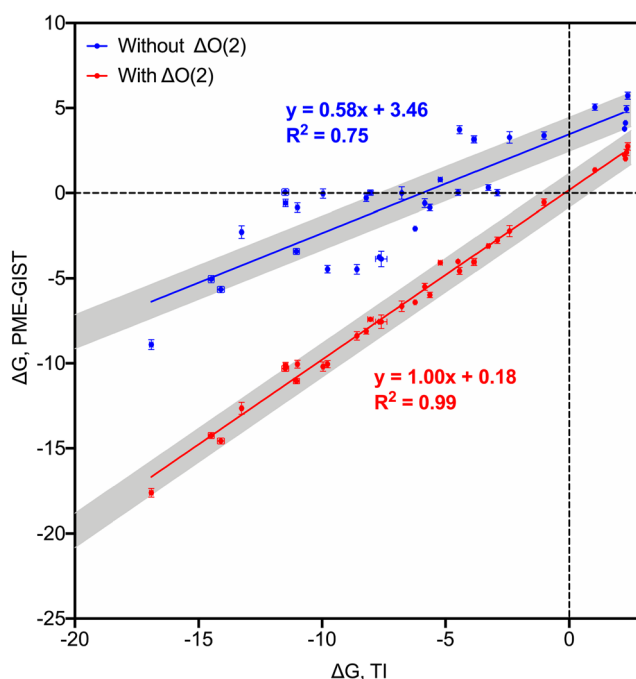


Figure 6. Solvation free energies calculated by PME-GIST, with or without $\Delta O(2)$ correction, compared to TI. The mean standard errors of solvation free energy calculated by PME-GIST and TI are 0.26 and 0.06 kcal/mol, respectively. The mean unsigned difference in solvation free energies calculated by PME-GIST and TI is 0.4 kcal/mol (with $\Delta O(2)$ correction) and 6.2 kcal/mol (without $\Delta O(2)$ correction). Each dot represents the data for a molecule, and the solid lines represent the linear regression lines whose equations are written on the graph. The area shaded gray represents values within ± 1 kcal/mol of the regression lines. Error bars reflect the standard error of the calculations.

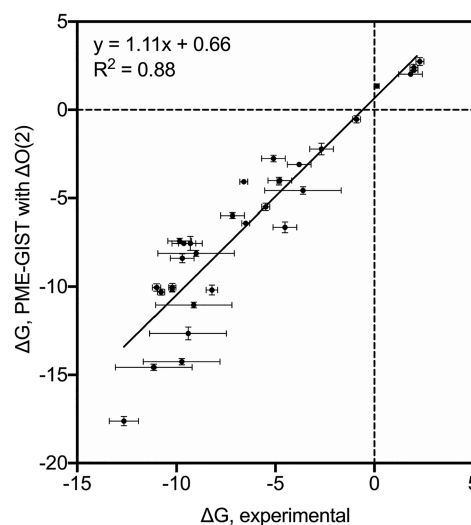


Figure 7. Solvation free energies calculated with PME-GIST, with $\Delta O(2)$ correction, compared to experimental results. Each dot represents the data for a molecule, and the solid line is the linear regression line whose equation is written on the graph. The error bars reflect the standard error of the calculations or the uncertainties of the experimental values collected from the FreeSolv database.

by integrating only over voxels that are within a certain distance of the solute heavy atoms instead of over the entire system. When the integration region extends to include all

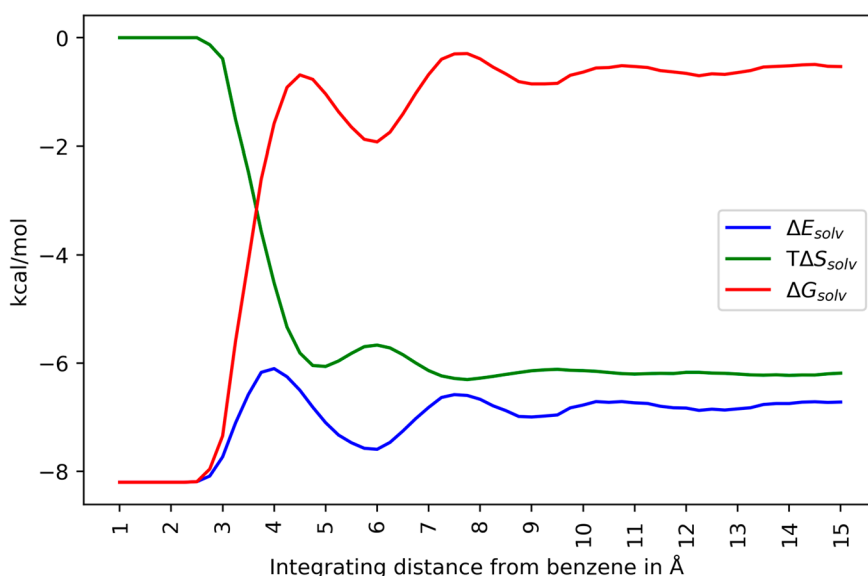


Figure 8. Perturbation of water molecules in regions of increasing distance around benzene. The solvation energies, entropies, and free energies were calculated by integrating the voxels within an increasing radius (1 Å–15 Å) from the heavy atoms of benzene. The initial value (about −8 kcal/mol) reflects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

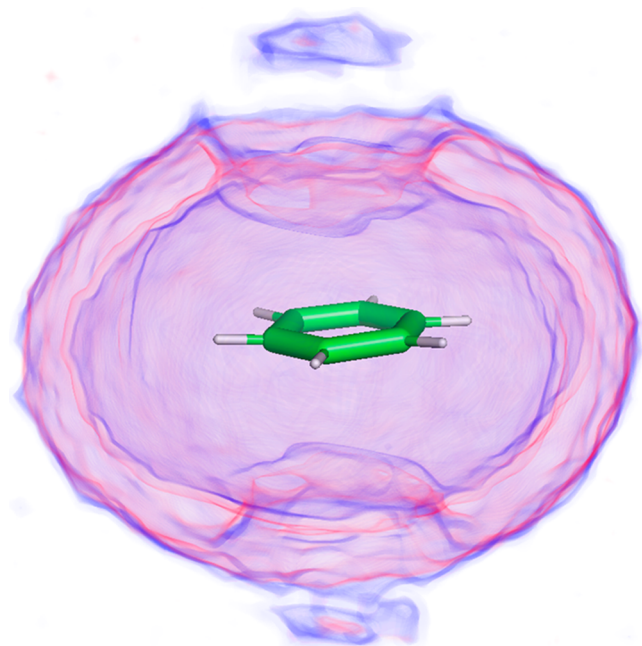


Figure 9. A GIST energy and entropy map around a benzene molecule. The benzene molecule is represented by the CPK model at the center. Regions where the water entropies are very unfavorable (entropy density is lower than -0.007 kcal/mol/Å³) are colored in red. Regions where the water energies are very favorable (energy density is lower than -0.39 kcal/mol/Å³) are colored in blue.

voxels within 10 Å of the solute heavy atoms, the correlation between the local PME-GIST free energies and TI-calculated free energies has an R^2 of 0.99 for all 32 molecules with a slope of 0.98 and a mean unsigned error of 0.37 (Table S4).

3.7. 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

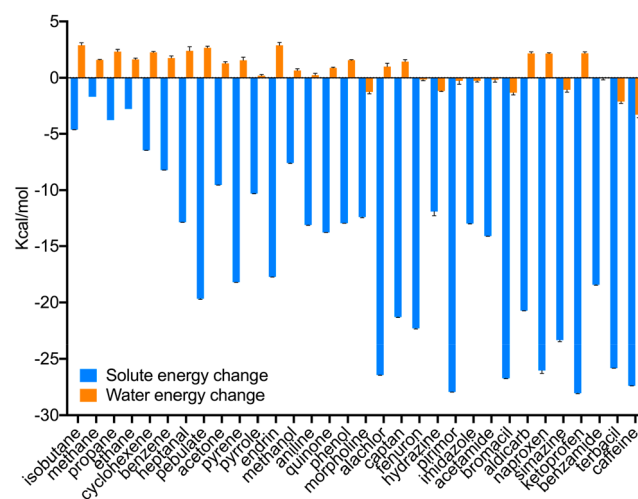


Figure 10. The decomposition of ΔE_{solv} into the energy changes of solute and water molecules. The molecules are ranked from hydrophobic to hydrophilic by the experimental free solvation energy. Bars represent mean values, and the error bars correspond to standard errors.

molecules constitute 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).

4. DISCUSSION OF HIGHER ORDER ENTROPY SCALING

The negative correlation between ΔS_{sw} and $\Delta O(2)$ suggests that the entropy change from the first order entropy term is

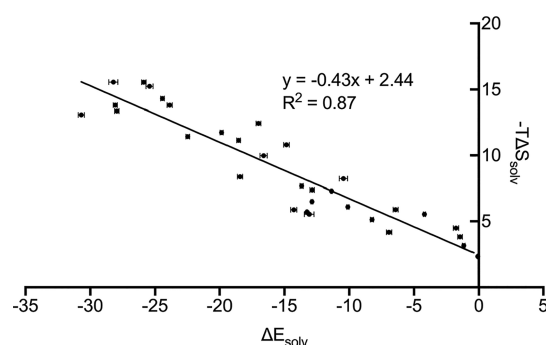


Figure 11. The energy–entropy compensation in solvation of small molecules. Each dot represents the data for a molecule, and the solid line is the linear regression fit to this data whose equation is written on the graph. The solvation energy and entropy values were calculated by PME-GIST with the $\Delta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

offset to a significant degree by an opposing change in the higher order entropy terms. A negative correlation between first order entropy and high order entropy terms was reported in other studies,^{15,49} but its physical significance was not discussed. We consider that this observation is largely due to two behaviors of water molecules. First, the more strongly a water molecule is held rigid by the field of the solute, the less sensitive its position and orientation are to the fluctuations of neighboring water molecules. At one limit, in neat water there is no solute, and the IST entropy of a water molecule is -14.05 cal/(mol·K),⁵⁰ which, in IST, is entirely accounted for in the second and higher order IST terms, as the first order IST term is zero. In the other limit, a water molecule is held entirely rigid by the solute and is insensitive to the fluctuations of neighboring water molecules. In this limit, all of the IST entropy would be accounted for in the first order terms, and the higher order terms would be zero. In this case, a rigid water molecule would have the entropy of that in an ideal crystal water which, in IST, is -30.75 cal/(mol·K).⁵¹ A simple linear fit between these two limits yields a negative correlation between the first order term and the second and higher order terms with a slope of -0.46 (Figure 12), which is close to -0.40 we report in this study.

The second behavior comes from the fact that water molecules are most correlated with their water neighbors in the first shell. A study by Lazaridis et al. showed that 90% of neat water entropy comes from the correlations between neighbors

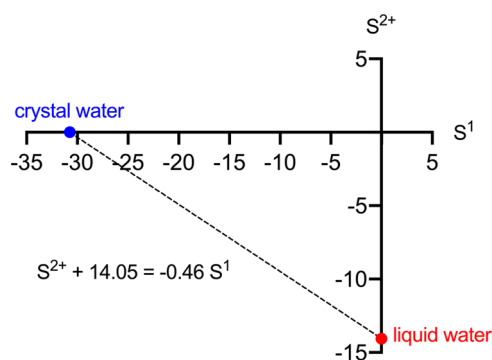


Figure 12. The entropy of a water molecule in the crystal and liquid states. S^1 denotes the first order entropy term, and S^{2+} denotes the second and higher order entropy terms. “ $S^{2+} + 14.05$ ” is equivalent to $\Delta O(2)$, and S^1 is equivalent to ΔS_{sw} in eq 24.

in the first shell.⁵² If a water molecule has no neighbors, then the high order IST terms can be approximated as zero, whereas if a fluid has the bulk number of neighbors, then the second and higher order entropy terms make the full contribution. Here, we also propose that the second and higher order entropy terms of a water molecule would linearly scale by the number of its water neighbors in the first solvation shell. The average number of water neighbors for water molecules within 5 Å of the solutes is about 90% of the water neighbor number of a bulk water molecule (Table S5). When the contributions of the second and higher order entropy terms are scaled by this value, the slope of the negative correlation is estimated to be -0.41 , which is quite close to the value estimated in this study (-0.4). Again, our study of small molecules only covers a limited first order entropy range. Whether this negative linear correlation between the first order entropy term and high order entropy terms can be applied broadly, such as to water molecules on the more complex surfaces of macromolecules, requires further investigation.

5. CONCLUSIONS

In this study, we developed the methodology for decomposing PME-based electrostatic treatment and long-range LJ corrections and implemented 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-state 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. A simple correction for higher order entropy ($\Delta O(2) = -0.4\Delta S_{sw}$) together with the enhanced accuracy in the energy calculation allows the PME-GIST end-state calculations to 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 of 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.^{53–55} The utility of the 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. The code for PME-GIST is available on the CPPTRAJ GitHub.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c01185>.

Table S1: solvation free energies of 32 small molecules calculated by TI at three different temperatures with three replicates (a , b , c); entropies calculated by finite-difference method, SEM reported for solvation free energies (ΔG), temperature multiplied solvation en-

tropies ($T\Delta S$), and enthalpies (ΔH) at 300 K for each molecule; and R^2 - coefficient of determination of ΔG over temperature (T) (XLSX)

Table S2: end-state energy of each solute. Solute intramolecular energies were calculated by CPPTRAJ energy command. Calculations were performed on three MD replicates (MD1, MD2, MD3). Average solvation energy (ΔE_{solv}) and SEM for each solute are shown at the bottom of the table. (XLSX)

Table S3, solute–water entropy calculated by GIST. Calculations were conducted on three MD replicates and averaged solute–water entropy, and SEM for each solute are shown at the bottom of the table. (XLSX)

Table S4: correlations between local approximations of PME-GIST free energies and TI free energy calculations at different integrating distances. Radius is the integrating distance around solute heavy atoms. Slope and R^2 values are for the linear regression line for locally integrated solvation free energies with solvation free energies calculated by TI for all 32 molecules. Mean unsigned differences (kcal/mol) are between the solvation free energies calculated by PME-GIST (integrating regions of different radii) and those calculated by TI (accounting for entire system) (XLSX)

Table S5: average first shell water neighbor number for water molecules in regions within certain distances of solute heavy atoms (XLSX)

energy, entropy, and free energy perturbation upon solvation of remaining 31 molecules (PDF)

AUTHOR INFORMATION

Corresponding Author

Tom Kurtzman – Department of Chemistry, Lehman College, The City University of New York, Bronx, New York 10468, United States; Ph.D. Program in Biochemistry and Ph.D. Program in Chemistry, The Graduate Center of The City University of New York, New York 10016, United States; orcid.org/0000-0003-0900-772X; Email: thomas.kurtzman@lehman.cuny.edu

Authors

Lieyang Chen – Department of Chemistry, Lehman College, The City University of New York, Bronx, New York 10468, United States; Ph.D. Program in Biochemistry, The Graduate Center of The City University of New York, New York 10016, United States; orcid.org/0000-0002-1583-8039

Anthony Cruz – Department of Chemistry, Lehman College, The City University of New York, Bronx, New York 10468, United States; Ph.D. Program in Chemistry, The Graduate Center of The City University of New York, New York 10016, United States; orcid.org/0000-0002-2904-7684

Daniel R. Roe – Laboratory of Computational Biology, National Institutes of Health – National Heart, Lung and Blood Institute, Bethesda, Maryland 20892, United States

Andrew C. Simmonett – Laboratory of Computational Biology, National Institutes of Health – National Heart, Lung and Blood Institute, Bethesda, Maryland 20892, United States; orcid.org/0000-0002-5921-9272

Lauren Wickstrom – Department of Science, Borough of Manhattan Community College, The City University of New York, New York 10007, United States; orcid.org/0000-0003-4055-3235

Nanjie Deng – Department of Chemistry and Physical Sciences, Pace University, New York 10038, United States; orcid.org/0000-0001-8805-3526

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.0c01185>

Notes

The authors declare no competing financial interest.

The 3D molecule structure files and raw dx files generated by GIST-2016 and PME-GIST are provided as additional information on Github and can be accessed at https://github.com/KurtzmanLab/Small_Molecule_Solvation.

ACKNOWLEDGMENTS

We thank Professors Emilio Gallicchio and Michael K. Gilson for their insightful comments regarding this manuscript. 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*, 1365–1370.
- (2) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102*, 3542–3550.
- (3) Wallace, D. C. Statistical Theory for the Entropy of a Liquid. *Phys. Rev. A: At., Mol., Opt. Phys.* **1989**, *39*, 4843–4847.
- (4) Baranyai, A.; Evans, D. J. Direct Entropy Calculation from Computer Simulation of Liquids. *Phys. Rev. A: At., Mol., Opt. Phys.* **1989**, *40*, 3817–3822.
- (5) Morita, T.; Hiroike, K. A New Approach to the Theory of Classical Fluids. III-General Treatment of Classical Systems. *Prog. Theor. Phys.* **1961**, *25*, 537–578.
- (6) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102*, 3531–3541.
- (7) Raveché, H. J. Entropy and Molecular Correlation Functions in Open Systems. I. Derivation. *J. Chem. Phys.* **1971**, *55*, 2242–2250.
- (8) 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*, 808–813.
- (9) 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*, 2769–2780.
- (10) 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*, 973–980.
- (11) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. *Methods Mol. Biol.* **2012**, *819*, 393–404.
- (12) Raman, E. P.; MacKerell, A. D. Rapid Estimation of Hydration Thermodynamics of Macromolecular Regions. *J. Chem. Phys.* **2013**, *139*, No. 055105.
- (13) 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*, 8232–8244.
- (14) 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*, 5872–5882.
- (15) Huggins, D. J. Estimating Translational and Orientational Entropies Using the K-Nearest Neighbors Algorithm. *J. Chem. Theory Comput.* **2014**, *10*, 3617–3625.

- (16) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (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*, 10089–10092.
- (18) 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*, 13052–13063.
- (19) 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. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (20) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Kollman, D. M. Y.; Kollman, P. A. *Amber 2020*; 2020.
- (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*, 1559–1569.
- (22) 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*, 2029–2037.
- (23) 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*, 3084–3095.
- (24) Velez-Vega, C.; McKay, D. J. J.; Kurtzman, T.; Aravamuthan, V.; Pearlstein, R. A.; Duca, J. S. Estimation of Solvation Entropy and Enthalpy via Analysis of Water Oxygen-Hydrogen Correlations. *J. Chem. Theory Comput.* **2015**, *11*, 5090–5102.
- (25) 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*, 414–429.
- (26) Lazaridis, T. Solvent Reorganization Energy and Entropy in Hydrophobic Hydration. *J. Phys. Chem. B* **2000**, *104*, 4964–4979.
- (27) PubChem. <https://pubchem.ncbi.nlm.nih.gov/> (accessed 2021-01-20).
- (28) 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*, 1623–1641.
- (29) 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*, 1157–1174.
- (30) 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*, 926–935.
- (31) Grest, G. S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *33*, 3628–3631.
- (32) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- (33) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (34) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise - A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.
- (35) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- (36) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (37) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*. 2nd ed.; Academic Press: 2002.
- (38) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (39) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011**, *32*, 3253–3263.
- (40) Levy, R. M.; Gallicchio, E. Computer Simulations with Explicit: Recent Progress in the Thermodynamic Decomposition of Free Energies and in Modeling Electrostatic Effects. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531–567.
- (41) 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*, 9539–9548.
- (42) Hnizdo, V.; Darian, E. V. A.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. Nearest-Neighbor Nonparametric Method for Estimating the Configurational Entropy of Complex Molecules. *J. Comput. Chem.* **2007**, *28*, 655–668.
- (43) Hnizdo, V.; Jun, T.; Killian, B. J.; Gilson, M. K. Efficient Calculation of Configurational Entropy from Molecular Simulations by Combining the Mutual-Information Expansion and Nearest-Neighbor Methods. *J. Comput. Chem.* **2008**, *29*, 1605–1614.
- (44) Singh, H.; Misra, N.; Hnizdo, V.; Fedorowicz, A.; Demchuk, E. Nearest Neighbor Estimates of Entropy. *Am. J. Math. Manag. Sci.* **2003**, *23*, 301–321.
- (45) Simmonett, A. C.; Burns, L. A.; Roe, D. R.; Brooks, B. R. *helpPME*. <https://github.com/andysim/helpme> (accessed 2020-10-16).
- (46) Mesecar, A. D. 6W63. <https://www.rcsb.org/structure/6w63> (accessed 2020-10-09).
- (47) 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. *J. Chem. Phys.* **1994**, *100*, 3334–3345.
- (48) Ishii, Y.; Yamamoto, N.; Matubayasi, N.; Zhang, B. W.; Cui, D.; Levy, R. M. Spatially-Decomposed Free Energy of Solvation Based on the Endpoint Density-Functional Method. *J. Chem. Theory Comput.* **2019**, *15*, 2896–2912.
- (49) Raman, E. P.; Mackerell, A. D. Spatial Analysis and Quantification of the Thermodynamic Driving Forces in Protein-Ligand Binding: Binding Site Variability. *J. Am. Chem. Soc.* **2015**, *137*, 2608–2621.
- (50) Wagner, W.; Pruß, A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–478.
- (51) Dunitz, J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science (Washington, DC, U. S.)* **1994**, *264*, 670.
- (52) Lazaridis, T.; Karplus, M. Orientational Correlations and Entropy in Liquid Water. *J. Chem. Phys.* **1996**, *105*, 4294–4316.
- (53) Uehara, S.; Tanaka, S. AutoDock-GIST: Incorporating Thermodynamics of Active-Site Water into Scoring Function for Accurate Protein-Ligand Docking. *Molecules* **2016**, *21*, 1604.
- (54) 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*, 1409–1423.
- (55) Balias, 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*, E6839–E6846.