JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Computational Prediction of Molecular Hydration Entropy with Hybrid Scaled Particle Theory and Free-Energy Perturbation Method

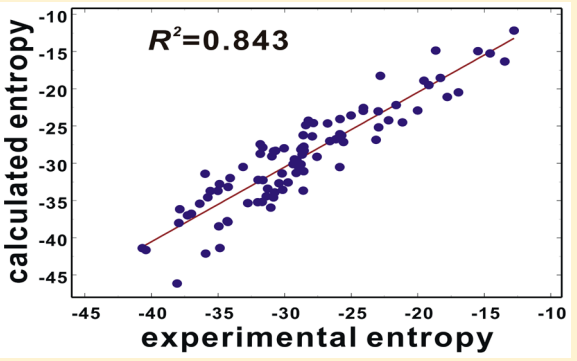
Hwanho Choi,[†,§] Hongsuk Kang,[‡,§] and Hwangseo Park*[,†]

[†]Department of Bioscience and Biotechnology, Sejong University, 209 Neungdong-ro, Kwangjin-gu, Seoul 143-747, Korea

[‡]Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, United States

Supporting Information

ABSTRACT: Despite the importance of the knowledge of molecular hydration entropy (ΔS_{hyd}) in chemical and biological processes, the exact calculation of ΔS_{hyd} is very difficult, because of the complexity in solute−water interactions. Although free-energy perturbation (FEP) methods have been employed quite widely in the literature, the poor convergent behavior of the van der Waals interaction term in the potential function limited the accuracy and robustness. In this study, we propose a new method for estimating ΔS_{hyd} by means of combining the FEP approach and the scaled particle theory (or information theory) to separately calculate the electrostatic solute−water interaction term (ΔS_{elec}) and the hydrophobic contribution approximated by the cavity formation entropy (ΔS_{cav}), respectively. Decomposition of ΔS_{hyd} into ΔS_{cav} and ΔS_{elec} terms is found to be very effective with a substantial accuracy enhancement in ΔS_{hyd} estimation, when compared to the conventional full FEP calculations. ΔS_{cav} appears to dominate over ΔS_{elec} in magnitude, even in the case of polar solutes, implying that the major contribution to the entropic cost for hydration comes from the formation of a solvent-excluded volume. Our hybrid scaled particle theory and FEP method is thus found to enhance the accuracy of ΔS_{hyd} prediction by effectively complementing the conventional full FEP method.



1. INTRODUCTION

Hydration free energy (ΔA_{hyd}) quantifies the free-energy change in the transfer of a solute from the gas phase to aqueous solution. Although ΔA_{hyd} serves as a fundamental thermodynamic quantity to characterize the behavior of a small molecule in a variety of chemical and biological systems, the experimental measurement of ΔA_{hyd} has lagged behind the need. To the best of our knowledge, the experimental ΔA_{hyd} values are available only for a few thousand, among the millions of organic compounds reported to date. Accordingly, several efficient computational methods to predict ΔA_{hyd} have been proposed on the basis of various theoretical frameworks, and assessed in Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) challenges.[1]

Although ΔA_{hyd} can be obtained straightforwardly from the solubility and vapor pressure data of a solute,[2] the latter is difficult to measure precisely with a standard apparatus, whereas the former can be measured with accuracy by determining the concentration of saturated solution. The rarity of experimental ΔA_{hyd} data can thus be attributed to the difficulty in measuring the vapor pressure of solute. Because the vapor pressure can be computed from the hydration entropy (ΔS_{hyd}) and vice versa,[3] the development of a reliable computational method to estimate ΔS_{hyd} seems to shed new light on the enrichment of the ΔA_{hyd} data for various solute molecules.

The knowledge of ΔS_{hyd} may also be informative for elucidating the thermodynamic aspects of the abnormal protein−ligand interactions in which the major contribution to the binding free energy comes from the entropic term without significant enthalpy−entropy compensation during the conformational changes for binding.[4] In contrast to the general notion that ligand binding to protein would be accompanied by the entropy loss, it was demonstrated that the binding entropy (ΔS_{bind}) could be positive during the formation of a protein−ligand complex,[5,6] and furthermore the $-T\Delta S_{bind}$ values were found to be more negative than ΔH_{bind} in some cases.[7−10] This indicates that the increase of entropy can serve as a driving force for binding of a ligand to the receptor protein. Therefore, the optimal activity of drug molecules can be achieved only by the compromise between the maximization of the protein−ligand interactions and the minimization of ΔS_{hyd} to maximize ΔS_{bind}, further necessitating the development of a reliable computational method to estimate the molecular ΔS_{hyd} values.

Despite the insufficient experimental data for validation, the development of a computational method for estimating ΔS_{hyd} has been pursued actively by several research groups since the pioneering work of Ben-Naim and Marcus on solvation thermodynamics.[11] The results of ideal solution theory-based

Received: April 7, 2015

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

calculations indicated that the entropic term associated with the first hydration shell would contribute dominantly to the solvation free energies of alkanes.[12] Langevin dipoles solvation model proved to be efficient in predicting the ΔS_{hyd} values of some small neutral molecules and monatomic ions.[13] Free-energy perturbation (FEP) and thermodynamic integration (TI) methods have been applied most frequently in estimating the molecular ΔS_{hyd} values through the thermodynamic decomposition of ΔA_{hyd}.[14−17] Recently, Gerogiokas et al. proposed "grid cell theory" to calculate ΔS_{hyd} by discretizing the cell theory methodology on a three-dimensional (3D) grid to resolve the density, enthalpy, and entropy of water molecules around the solute.[18] Through the graphical analyses of the spatially resolved components of ΔA_{hyd} obtained from only a single simulation, they were able to predict ΔS_{hyd} with comparable accuracy to FEP and TI-based methods. However, these methods seem to be insufficient to be useful in practical applications, because large deviations from experimental data were observed, even in the validation with only <30 solute molecules.

The present study aims to propose a new computational method to estimate the molecular ΔS_{hyd} values with high accuracy, based on the modified thermodynamic decomposition approach. The key feature of this method lies in that the solute−water interactions are decomposed into hydrophobic and electrostatic parts to calculate their respective contributions to ΔS_{hyd} within different theoretical frameworks. This decomposition is required because full FEP calculations have often been unsuccessful in reproducing the ΔS_{hyd} values of small solute molecules.[19,20] Our extended thermodynamic decomposition approach is thus anticipated to enhance the accuracy in ΔS_{hyd} prediction by complementing the conventional FEP methods.

2. THEORY AND COMPUTATIONAL METHODS

A solute molecule would dissolve into water when the resulting solute−water interactions are strong enough to surmount the entropic penalty for establishing the optimal configuration of hydration. The loss of entropy during hydration stems from the reordering of solute and water molecules to form a complex with the optimal intermolecular interactions. In this context, ΔS_{hyd} may be decomposed into van der Waals (ΔS_{vdW}) and electrostatic (ΔS_{elec}) contributions.

$$\Delta S_{hyd} = \Delta S_{vdW} + \Delta S_{elec} \tag{1}$$

In the usual FEP and TI methods, the ΔS_{vdW} and ΔS_{elec} terms are calculated simultaneously to obtain ΔS_{hyd} using Lennard-Jones and Coulomb potentials, respectively. However, it is often difficult to complete molecular dynamics (MD) simulations for FEP and TI procedures successfully, because of the appearance of singularity in the Lennard-Jones potential, since the potential function is scaled from the initial state to the final state. Although this singularity problem can be overcome by scaling the electrostatic and van der Waals interaction terms in the sequential manner, it is difficult to avoid the numerical instability in FEP calculations completely, because of the possibility of the appearance of unwanted minima in the nonbonded potential energy at intermediate λ values. This problem may arise especially when the attractive electrostatic interactions are counterbalanced insufficiently by van der Waals repulsions.[21] Because of such technical difficulty in coping with the van der Waals interaction term, it seems to be a reasonable choice to compute the ΔS_{vdW} term separately with a simpler method than the full FEP calculations.

The sign of ΔS_{vdW} is negative for most organic solutes, because of the immiscibility between their hydrophobic parts and water. Therefore, ΔS_{vdW} can be approximated as the loss of entropy associated with the reordering of water molecules to form a cavity (ΔS_{cav}) in which a solute molecule can be accommodated. Then, ΔS_{hyd} can be expressed in the following form:

$$\Delta S_{hyd} = \Delta S_{cav} + \Delta S_{elec} \tag{2}$$

Some implicit solvent models, such as the solvent-accessible surface area (SASA) method, seems to be appropriate for the calculation of molecular hydration free energy.[22] It would be possible to calculate the S_{cav} term within the SASA framework upon the availability of a proper method for calculating the first derivative of hydration free energy, with respect to temperature. In the viewpoint that the S_{cav} term can be obtained straightforwardly from the relation between the hydration free energy and temperature, two theoretical frameworks can be applied to obtain the molecular ΔS_{cav} values. We first employed the scaled-particle theory (SPT) developed for describing the radial distribution of hard sphere mixtures in the liquid phase.[23,24] Because SPT proved to be appropriate for describing the solute−solvent interactions for hydrophobic solutes,[25,26] it is anticipated to be useful for estimating the ΔS_{vdW} values of organic solute molecules.

Basically SPT aims to calculate the chemical potential ($\Delta\mu$) for the insertion of a cavity into the hard sphere fluid.[27] Since the formation of a cavity for a hydrophobic solute in water may be approximated as an entropic process, because of the weak solute−water interactions, ΔS_{cav} can be expressed as follows:

$$\frac{\Delta S_{cav}}{k_B} = -\ln(1-\varphi) + 3\left(\frac{\varphi}{1-\varphi}\right)R + \left[\left(\frac{3\varphi}{1-\varphi}\right) + 4.5\left(\frac{\varphi}{1-\varphi}\right)\right]R^2 + \left(\frac{\varphi P}{\rho k_B T}\right)R^3 \tag{3}$$

Here, φ and ρ represent the volume fraction of water molecules, with respect to the total volume (0.371), and the number density of water (0.333×10^{24} m^{-3}), respectively. R denotes the ratio of the radius of solute to that of the water molecule (r_t/r_w) while P stands for the pressure of water (2.306 kPa). Because r_w is known to be 1.4 Å,[28] the knowledge of r_t is the only requisite for calculating the ΔS_{cav} value of a solute molecule. The r_t parameter of each solute was set equal to the radius of the water-accessible volume (WAV) that was assumed to have spherical symmetry. Within this approximation, r_t was calculated in a straightforward manner from the WAV value that could be set equal to $(4/3)\pi r_t^3$.

The accuracy of the present hybrid SPT/FEP method would be limited because a hard-sphere model had to be used for solutes in SPT calculations. Prior to calculating the ΔS_{cav} values within the SPT framework, we carried out MD simulations with the full molecular structures of solutes and explicit water molecules to calculate the WAV values of all solute molecules as accurately as possible. These MD-based calculations would have the effect of reflecting the conformational diversity of solute molecules, which was shown to be a significant contributor to hydration entropy.[29] The disadvantage of conventional SPT can thus be complemented in part by

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

statistical sampling of various conformations of solute molecules in aqueous solution. The WAV value of each solute molecule was then calculated with the analytical method proposed by Petitjean,[30] in which it was given by the union of the van der Waals volumes of all of the solute atoms.

As the alternative to SPT, the information theory (IT) can also be applied to calculate ΔS_{cav}, as suggested by Hummer et al.[31] In this theoretical framework, the probability (p_n) that n water molecules are found in a cavity is defined as follows:

$$p_n = \exp(\omega_0 + \omega_1 n + \omega_2 n^2) \quad (4)$$

Here, ω_0, ω_1, and ω_2 parameters can be determined by solving Lagrange equations with the pair correlation function ($g(r)$) between a pair of oxygen atoms in water. MD simulations were performed with the NAMD program (Version 2.8)[32] in the calculation of $g(r)$. To select the number of water molecules to describe the bulk solvent, we compared the $g(r)$ profiles obtained with 233 and 1981 TIP3P[33] waters used as the solvent model for 2-propanol. As shown in Figure 1, $g(r)$ remains

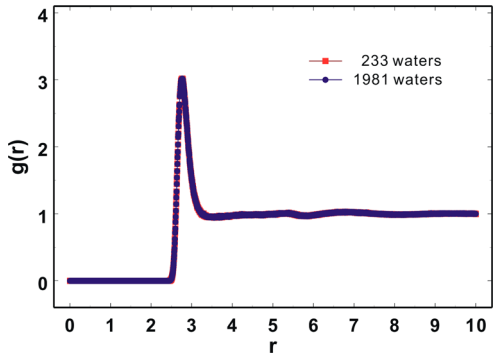


Figure 1. Comparison of the pair correlation functions calculated via 1 ns MD simulations with 233 and 1981 water molecules.

almost intact with the increase in the number of water molecules from 233 to 1981. Therefore, to save the computational burden, we carried out MD simulation with 233 TIP3P waters to calculate the ΔS_{cav} values of the solute molecules.

After the equilibration dynamics for 1 ns, water configurations were sampled for 1 ns production dynamics in NVT ensemble to obtain $g(r)$. ω_0, ω_1, and ω_2 parameters were then determined by numerically solving the Lagrange equations with a multidimensional root-finding algorithm.[34] Finally, the ΔS_{cav} value of each solute molecule was calculated with the negative logarithm of p_0.

$$\Delta S_{cav} = -k_B \ln p_0 \quad (5)$$

As mentioned above, it would be reasonable to separately calculate the entropic cost for van der Waals interactions between the solute and water with SPT or IT. Then, FEP simulations can be performed for the purpose of calculating the ΔS_{elec} term only by scaling the Coulomb potential with the unstable Lennard-Jones potential being kept intact throughout the simulations. This was accomplished using a scaling parameter (λ_{elec}) that played the role of creating and annihilating the partial charges of solute and water atoms in the FEP simulations of solute hydration. The electrostatic part of ΔA_{hyd} calculated with FEP was then used to obtain ΔS_{elec}, which can be expressed as the temperature derivative of ΔA_{hyd}. As widely adopted in the literature,[35−39] we used the finite difference method to calculate the ΔS_{elec} value of each solute molecule.

Because entropy can be given by the derivative of Helmholtz free energy (A), with respect to temperature at constant volume, the entropy change (ΔS) associated with the solute−water interactions is written as follows, using the finite difference method.

$$\Delta S = -\frac{\Delta A(T + \Delta T) - \Delta A(T - \Delta T)}{2\Delta T} \quad (6)$$

If one neglects the van der Waals term in the potential energy function, ΔA at a given temperature is the difference of Helmholtz free energies between the charged and uncharged states of the system. In this case, ΔS_{elec} can be given by eq 7:

$$\Delta S_{elec} = -\{[A_q(T + \Delta T) - A_0(T + \Delta T)] - [A_q(T - \Delta T) - A_0(T - \Delta T)]\}/(2\Delta T) \quad (7)$$

Here, A_q and A_0 are the Helmholtz free energies of the charged and uncharged systems, respectively. It is thus necessary to compute the differences between A_q and A_0 at temperatures $T + \Delta T$ and $T - \Delta T$ to obtain ΔS_{elec}, which were accomplished successfully with the constrained FEP simulations.

To select a proper ΔT value for obtaining the first derivative of ΔA via the finite difference method, we calculated the ΔS_{elec} values with varying ΔT (10, 20, and 30 K), using 2-propanol as the probe molecule. In this test case, the statistical errors of the calculated ΔS_{elec} values appeared to become larger as ΔT decreased from 30 K to 20 K and 10 K. This validation result is consistent with the previous computational finding that the finite difference method can be valid near the room temperature with ΔT = 30 K.[17] Therefore, ΔT was set equal to 30 K in this study.

Because ΔA_{hyd} refers to the change in free energy for a solute molecule to leave the gas phase and to enter the liquid water, a thermodynamic cycle may be useful to calculate the ΔA_{hyd} values at varying temperatures, which is illustrated in Figure 2.

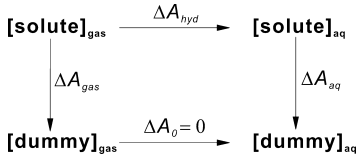


Figure 2. Thermodynamic cycle for calculating the hydration free energies of solute molecules. Note that $\Delta A_0 = 0$ because solute is not involved in the process.

Because A is thermodynamically a state function, the sum of all free-energy changes in the cycle must be zero. Therefore, ΔA_{hyd} can be expressed as follows:

$$\Delta A_{hyd} = \Delta A_{gas} - \Delta A_{aq} \quad (8)$$

ΔA_{aq} and ΔA_{gas} correspond to the difference between A_q and A_0 values of the solute in aqueous solution and that in the gas phase, respectively. We obtained the ΔA_{hyd} values of all the solute molecules through the FEP calculations of ΔA_{aq} and ΔA_{gas}, based on the independent MD simulations in water and in the gas phase, respectively.

To determine the ΔA_{aq} and ΔA_{gas} values, the solute molecule with atomic charges (state 0) was changed gradually to the dummy molecule in which all the atoms were uncharged (state

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

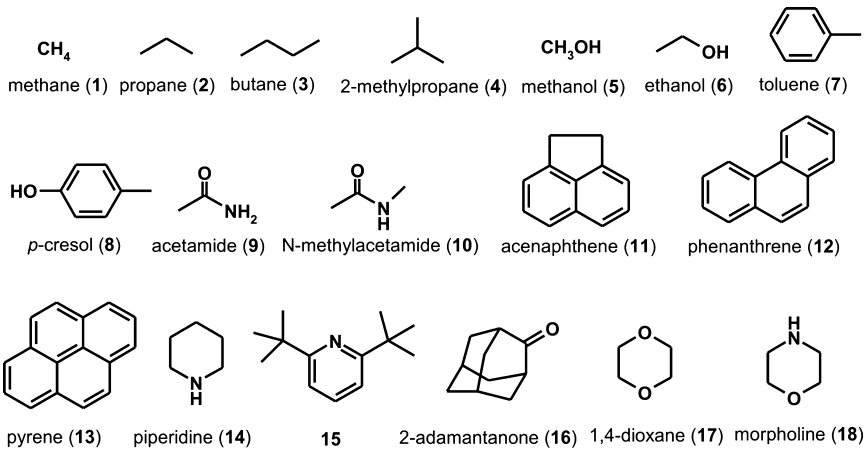


Figure 3. Structures of the molecules used to compare the performances of hybrid SPT/FEP method with those of the conventional full FEP simulations. IUPAC name of **15** is 4-(1,1-dimethylethyl)pyridine.

1) by perturbing the Hamiltonian of the system. This transformation could be accomplished by scaling the electrostatic interaction potential with the change of a state variable (λ_{elec}) that maps the initial and the final states onto 0 and 1, respectively. ΔA_{aq} and ΔA_{gas} values for the change from the initial state to the final state were then calculated by summing the incremental free-energy changes over the several windows of λ_{elec}, which were varied from 0 to 1 with the discrete and uniformly spaced intervals.[40]

$$\Delta A(0 \rightarrow 1) = -RT \sum_i \ln \left\langle \exp\left(-\frac{V_{\lambda(i+1)} - V_{\lambda(i)}}{RT}\right)\right\rangle_{\lambda(i)} \quad (9)$$

Here V_i is the potential energy function for the representative state i, and $\langle\cdots\rangle_i$ denotes the ensemble average over the samples of the enclosed quantity in state i.

FEP calculations to compute ΔA_{aq} and ΔA_{gas} values were performed through MD simulations of the solutes in aqueous solution and in the gas phase, respectively. In the former case, each solute molecule was immersed in a periodic box with dimensions of 29.1 Å × 29.1 Å × 29.1 Å, which was followed by the addition of 824 TIP3P water molecules in the simulation box, to be consistent with the density of water at 298 K. For all the solute molecules under investigation, the entire water−solute system exhibited a good convergent behavior at 268 K, as well as at 328 K, after ∼500 ps of simulation time, because it included only a small molecule and water molecules.

After the equilibration dynamics for 1 ns, we carried out 18 ns of perturbation using the NAMD program (Version 2.8)[32] at 268 and 328 K with varying λ_{elec} values. Each perturbation consisted of nine windows with 1 ns of equilibration and 1 ns of data collection. CGenFF[41] parameter sets were used for all the solute molecules and CHARMM[42] parameter sets for TIP3P water. Atomic charges were assigned using the extended bond-charge increment scheme to be able to capture short- and medium-range inductive effects. All MD-based FEP calculations were conducted with periodic boundary conditions in the NVT ensemble and with the nonbond-interaction cutoff radius of 12 Å. We used the time step of 1 fs to integrate the Langevin equations of motion with the friction coefficient of 5 ps^{-1}. MD simulations started from λ_{elec} = 0, which corresponds to the fully charged system. Both intramolecular and intermolecular electrostatic interactions were gradually made weaker by increasing λ_{elec} until it reaches 1, at which point all atomic charges of the solute molecule vanish. Because only the electrostatic part of ΔA_{hyd} (ΔA_{elec}) needed to be calculated in the FEP simulations, the Coulomb potential was scaled by the parameter 1 − λ_{elec}, whereas the Lennard-Jones term remained unchanged during the entire course of simulations. The configurations were sampled every 10 time steps to calculate the variation of electrostatic interaction energies with increasing the λ_{elec} value by 0.125. At each λ_{elec} value, we took the ensemble average of total energies to obtain the incremental hydration free energy ($\Delta A_{elec}(\lambda_{elec})$) as implemented in the NAMD program. All of these $\Delta A_{elec}(\lambda_{elec})$ values were then summed over the λ_{elec} values to obtain the ΔA_{elec} values at 268 and 328 K, which were, in turn, combined to calculate ΔS_{elec}, using eq 7. A double-wide sampling was performed for all of the charge transformations, and the final results were obtained by taking the averages from the backward and forward simulations.

To compare the performances of our hybrid SPT (or IT) and FEP method with the conventional full FEP simulations, we also calculated the ΔS_{hyd} values of 18 representative solute molecules shown in Figure 3. This test set included methane, propane, butane, 2-methylpropane, toluene, methanol, ethanol, *p*-cresol, acetamide, and *N*-methylacetamide, which had been widely studied as the simplified structural models for side-chain and backbone groups of proteins. Some large hydrophobic molecules such as acenaphthene, phenanthrene, pyrene, 4-(1,1-dimethylethyl)pyridine, and 2-adamantanone were also included in the comparative analysis, because SPT and IT may become less effective in estimating the ΔS_{cav} values with the increase in the van der Waals volume. Despite the extensive search of literature and publicly accessible chemical databases, we found that the experimental hydration entropy data were unavailable for the solutes with molecular weights of >300 amu, compared to the abundance of hydration free-energy data. Therefore, the validation set was further enriched with piperidine, 1,4-dioxane, and morpholine, because many drug molecules include the heterocyclic moieties. Although some of the calculated ΔS_{hyd} values were available in the literature, we recalculated all of them, because of the strong dependence of FEP results for ΔS_{hyd} on force fields and water models.[43]

Full FEP calculations were carried out by scaling both van der Waals and electrostatic interaction terms with the two state variables λ_{vdW} and λ_{elec}, respectively. More specifically, ΔA_{hyd} values of the test molecules were obtained by switching off the

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

electrostatic and van der Waals interaction potentials in the sequential way to warrant the stability of the entire system including the solute and water molecules. We first defined 21 windows to perform the FEP simulation of each solute with the λ values ranging from 0 to 1 ($\Delta\lambda = 0.05$). To further reduce the potential statistical errors in FEP calculations, the two λ states were added near the physical end states ($\lambda = 0.01$ and $\lambda = 0.99$) as recommended in the literature.[44] Following the suggestion by Pohorille et al.,[45] we combined the results of forward and backward simulations to obtain the ΔA_{hyd} values of all solute molecules rather than using the single unidirectional simulation results.

To improve the poor convergent behavior of the van der Waals interaction term in the full FEP calculations, we used the soft-core potential (U_{vdW}^{SC}) shown in eq 10,[46,47] instead of the original Lennard-Jones potential.

$$U_{vdW}^{SC}(r, \lambda_{vdW}) = 4\lambda_{vdW}\varepsilon\left\{\left[\frac{\sigma^6}{\alpha(1-\lambda_{vdW})^2\sigma^6 + r^6}\right]^2 - \frac{\sigma^6}{\alpha(1-\lambda_{vdW})^2\sigma^6 + r^6}\right\} \quad (10)$$

Here, α is set equal to 0.5 to control the magnitude of the soft-core term. This soft-core potential has a functional dependence on λ_{vdW} and reduces to the original Lennard-Jones potential when $\lambda_{vdW} = 1$. Because of the presence of the soft-core term in both repulsive and attractive parts, the problematic singularity can be avoided in a large part in FEP simulations.

To establish a method for estimating ΔS_{hyd}, it was required to prepare a reference data set comprising various organic molecules and their experimental ΔS_{hyd} values. Therefore, we built a chemical library of 90 organic molecules (21 hydrocarbons and 69 polar molecules) for which the experimental hydration enthalpies were available in the Minnesota Solvation Database and Physical/Chemical Property Database,[48] and the corresponding experimental heat capacity data could be extracted from the *CRC Handbook of Chemistry and Physics*.[49] The enthalpy and heat capacity data were then combined to generate the experimental ΔS_{hyd} values that served as the yardsticks to validate the corresponding computational results. The standard states of the reference data set were normalized to be consistent with the definition of hydration entropy by Ben-Naim and Marcus.[11]

3. RESULTS AND DISCUSSION

With respect to the calculation of ΔS_{cav}, SPT and IT could be applied, because both proved to be effective in estimating the entropic cost for creating a void volume in which a solute molecule can be accommodated. To select the more appropriate theoretical tool, we compared the results for the ΔS_{cav} values of 90 solute molecules calculated with SPT and IT. As shown in Figure 4, the ΔS_{cav} calculation results of SPT are in almost complete agreement with those of IT with the associated squared linear correlation coefficient (R^2) of 0.999. Furthermore, the slope and intercept of the regression line amount to 1.005 and 0.111, respectively. This excellent agreement indicates almost the same performance of SPT as IT in estimating the ΔS_{cav} value of a solute. Considering the small difference between SPT and IT results, we will hereafter refer to the average of the two ΔS_{cav} values for each solute molecule calculated with SPT and IT as its ΔS_{cav} value.

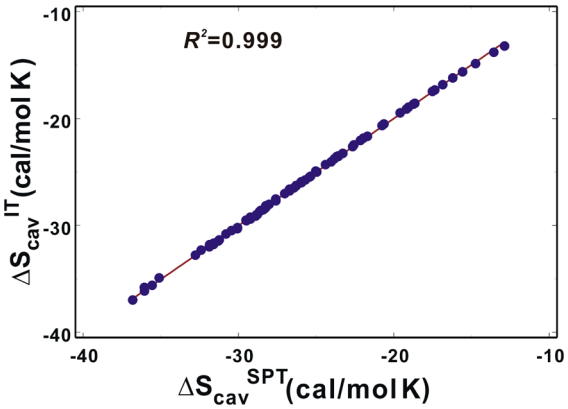


Figure 4. Correlation diagram for the cavity formation entropies (ΔS_{cav}) of 90 organic molecules calculated with scaled particle theory (SPT) and information theory (IT). The slope and intercept of the fitting are 1.005 and 0.111, respectively.

With respect to the high correlation between the ΔS_{cav} values calculated with SPT and IT, we note that all 90 solute molecules are small, with molecular weights of <200 amu. For this reason, their van der Waals radii fall within 3.5 Å, which has the effect of limiting the number of water molecules in the cavity formed by a solute molecule to one or two. In this case, the probability of finding no water molecule in the cavity (p_0) proved to be very similar in SPT and IT.[50] Since ΔS_{cav} is given by the negative logarithm of p_0 (eq 5), the small cavity volume of solutes can be invoked to explain the coincidence of the ΔS_{cav} results obtained with SPT and IT.

Table 1 lists the calculated ΔS_{cav}, ΔS_{elec}, and ΔS_{hyd} values, in comparison with the corresponding experimental hydration entropy (ΔS_{hyd}^{X}) values for all 90 solute molecules under investigation. ΔS_{cav} has a tendency to become more negative with the increase in the molecular weight of solutes. This is not surprising because the larger a solute molecule is, the more it will make the water molecules lose the degrees of freedom needed to construct a void volume. More interestingly, ΔS_{hyd} values appear to be dominated in magnitude by ΔS_{cav}, rather than ΔS_{elec}. For example, the magnitude of the latter is only 11.3% of that of the former, on average, for all of the solute molecules. Such a minor contribution of electrostatic solute−water interactions to hydration entropy was also observed in thermodynamic decomposition analysis of hydration free energies of alkanes and amines.[17,18] The dominance of the ΔS_{cav} term in ΔS_{hyd} is also consistent with the previous computational finding that the loss of entropy during the hydration of a solute could be attributed in a large part to the formation of a solvent-excluded volume with only a minor contribution of the electrostatic solute−water interactions.[51,52] Thus, our computational results confirm that the entropic penalty for hydration stems from the repulsive hydrophobic solute−water interactions rather than the electrostatic ones.

Note that only the repulsive van der Waals interactions were considered in this study to calculate ΔS_{cav}, with the attractive part being neglected, for the sake of simplicity. To address the validity of this approximation, we compared the calculated ΔS_{cav} values of 21 hydrocarbons for which the electrostatic interactions with water would be negligibly small with the corresponding experimental ΔS_{hyd} values. As shown in Figure 5, the ΔS_{cav} values compare reasonably well with experimental ΔS_{hyd} data with the associated R^2 and root-mean-square error (RMSE) values of 0.842 and 2.65 cal mol^{-1} K^{-1}, respectively.

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. ΔS_{cav}, ΔS_{elec}, and ΔS_{hyd} Values of 90 Solute Molecules Calculated at 298 K, in Comparison with the Corresponding Experimental Hydration Entropy (ΔS_{hyd}^{X})

compound	Entropy Values (cal mol^{-1} K^{-1})				compound	Entropy Values (cal mol^{-1} K^{-1})			
	ΔS_{cav}	ΔS_{elec}	ΔS_{hyd}	ΔS_{hyd}^{X}		ΔS_{cav}	ΔS_{elec}	ΔS_{hyd}	ΔS_{hyd}^{X}
methane	−13.67	0.21	−13.46	−16.33	3-methylpyridine	−26.29	−2.28	−28.56	−27.76
ethane	−17.34	0.39	−16.94	−20.47	4-methylpyridine	−26.30	−2.39	−28.69	−27.95
propane	−20.57	0.57	−20.00	−22.91	3-ethylpyridine	−28.91	−2.91	−31.82	−27.44
butane	−23.55	0.42	−23.13	−26.85	4-ethylpyridine	−28.87	−1.95	−30.81	−28.38
2-methylpropane	−23.53	0.59	−22.93	−25.17	3,4-dimethylpyridine	−28.88	−2.77	−31.64	−27.86
pentane	−26.27	0.41	−25.86	−30.51	3,5-dimethylpyridine	−28.89	−2.74	−31.63	−32.26
2,2-dimethylpropane	−26.27	0.70	−25.57	−27.14	4-*tert*-butylpyridine	−33.37	−2.60	−35.97	−31.41
hexane	−28.91	0.31	−28.60	−33.67	acetonitrile	−17.48	2.01	−15.47	−14.94
heptane	−31.29	0.26	−31.03	−35.95	propanenitrile	−20.70	1.17	−19.53	−18.91
ethene	−19.14	0.49	−18.65	−14.86	2-butanone	−24.04	−3.81	−27.84	−24.60
1-propene	−16.20	1.64	−14.56	−15.25	2-pentanone	−26.66	−3.37	−30.03	−27.98
cyclopropane	−19.53	1.75	−17.78	−21.09	3-pentanone	−26.73	−3.95	−30.68	−28.31
benzene	−24.35	0.27	−24.08	−22.57	3-methyl-2-butanone	−26.66	−5.36	−32.01	−35.21
methylbenzene	−27.02	1.17	−25.85	−26.09	2-hexanone	−29.31	−3.81	−33.11	−30.49
ethylbenzene	−29.51	0.42	−29.08	−29.59	4-methyl-2-pentanone	−29.22	−4.87	−34.09	−31.98
m-xylene	−29.52	0.75	−28.77	−28.14	2-heptanone	−31.68	−4.68	−36.37	−35.42
p-xylene	−29.52	0.75	−28.77	−30.13	4-heptanone	−31.63	−5.37	−37.00	−36.76
propylbenzene	−31.94	0.68	−31.26	−33.40	2-nonanone	−36.08	−4.33	−40.41	−41.63
(1-methylethyl)benzene	−31.82	−0.16	−31.98	−32.22	5-nonanone	−35.92	−4.76	−40.68	−41.38
9H-fluorene	−35.57	0.69	−34.88	−32.78	propanoic acid	−22.51	−2.50	−25.01	−23.58
anthracene	−36.89	1.32	−35.57	−33.72	butanoic acid	−25.38	−2.55	−27.92	−26.38
methanol	−15.59	−2.70	−18.29	−18.53	diethyl ether	−24.96	−4.16	−29.12	−31.29
ethanol	−18.99	−3.97	−22.96	−23.02	1,2-dimethoxyethane	−26.46	−3.72	−30.18	−31.35
1-propanol	−22.09	−4.50	−26.59	−27.01	tetrahydrofuran	−23.27	−2.44	−25.71	−26.25
2-propanol	−22.09	−5.47	−27.56	−29.12	tetrahydropyran	−25.96	−5.86	−31.82	−28.72
1-butanol	−24.96	−5.18	−30.14	−33.55	2-methyltetrahydrofuran	−26.01	−3.34	−29.35	−30.11
2-methyl-1-propanol	−24.98	−5.44	−30.42	−32.67	2,5-dimethyltetrahydrofuran	−28.57	−4.19	−32.76	−35.34
2-butanol	−24.96	−5.86	−30.83	−34.58	1,3-dioxolane	−21.89	−0.91	−22.80	−18.26
2-methyl-2-propanol	−24.97	−5.76	−30.74	−33.90	methyl acetate	−22.63	−1.46	−24.09	−22.95
1-pentanol	−27.63	−3.74	−31.37	−34.43	ethyl acetate	−25.44	−3.15	−28.60	−26.23
3-pentanol	−27.55	−6.68	−34.24	−33.17	methylbutanoate	−28.02	−2.93	−30.95	−29.06
2-methyl-2-butanol	−27.64	−6.68	−34.31	−37.76	*N*-butylacetate	−30.46	−4.55	−35.01	−33.69
1-hexanol	−30.18	−4.75	−34.93	−38.45	ethylpentanoate	−32.77	−4.51	−37.28	−36.98
3-hexanol	−30.12	−5.81	−35.93	−42.11	2-methoxyethanol	−23.61	−4.93	−28.54	−28.30
4-heptanol	−32.35	−5.72	−38.07	−46.13	2-propoxyethanol	−29.01	−6.75	−35.76	−34.56
1,2-ethanediol	−20.57	−5.59	−26.15	−26.77	2-butoxyethanol	−31.37	−6.57	−37.94	−38.01
cyclopentanol	−25.92	−5.76	−31.68	−35.18	3-hydroxybenzaldehyde	−28.67	0.27	−28.40	−24.87
cyclohexanol	−28.48	−5.78	−34.26	−37.87	3-hydroxybenzonitrile	−28.39	0.18	−28.21	−24.27
cycloheptanol	−30.80	−4.05	−34.85	−41.37	fluoromethane	−14.79	2.03	−12.76	−12.17
phenol	−25.72	−0.13	−25.84	−24.06	2,2,2-trifluoroethanol	−21.67	−0.36	−22.03	−24.51
o-cresol	−28.27	−0.66	−28.93	−30.54	formic acid	−16.13	−3.82	−19.96	−19.49
p-cresol	−28.20	−0.48	−28.68	−28.83	*N*,*N*-dimethylformamide	−18.68	−3.51	−22.19	−24.23
4-*tert*-butylphenol	−35.00	−2.87	−37.87	−36.16	acetamide	−20.09	−1.53	−21.62	−22.17
pyrroidine	−23.79	−5.95	−29.74	−32.57	*N*-methylacetamide	−23.19	−5.38	−28.57	−31.06
pyridine	−23.57	−3.21	−26.77	−24.65	tetrahydro-2-furanmethanol	−27.32	−1.92	−29.25	−29.47

The slope and intercept of the linear correlation diagram amount to 0.914 and −1.501, respectively. These results support the reasonableness of the assumption that the attractive van der Waals interactions between the hydrophobic solute and water would have little effect on ΔS_{cav}.

The largest differences between ΔS_{cav} and ΔS_{elec} are observed in hydrocarbons: the average absolute value of the latter amount to only 2.6% of that of the former for 21 hydrocarbons from methane to anthracene. Such a low absolute ΔS_{elec} value for hydrocarbons is actually not surprising, because their individual atoms have insignificant partial charges due to the similarity of electronegativities between carbon and hydrogen. It is also noteworthy that, in contrast to the solutes with polar moieties, most hydrocarbons have positive ΔS_{elec} values, which implies the gain of entropy through the solute−water electrostatic interactions. This may be understood in the context that the configurational degrees of freedom of hydrocarbons would increase in water, because of the lack of a stable conformation. Judging from the highly negative ΔS_{cav} values (Table 1), however, such a gain of entropy seems to be negligible when compared to the loss of entropy in the formation of a cavity. Thus, the difficulty in the formation of a void volume may be responsible for the immiscibility of hydrocarbons in water.

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

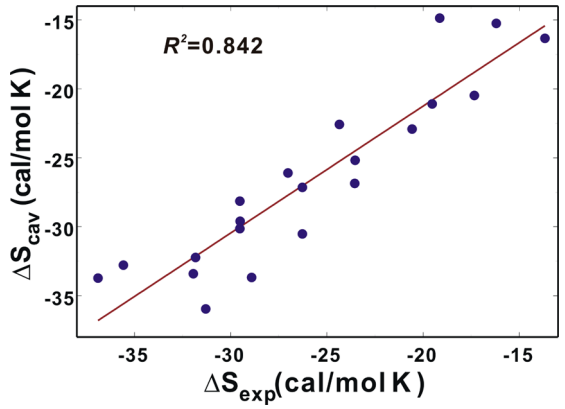


Figure 5. Linear correlation diagram between the experimental hydration entropies and the calculated ΔS_{cav} values of 21 hydrocarbons. The slope and intercept of the fitting are 0.914 and −1.501, respectively.

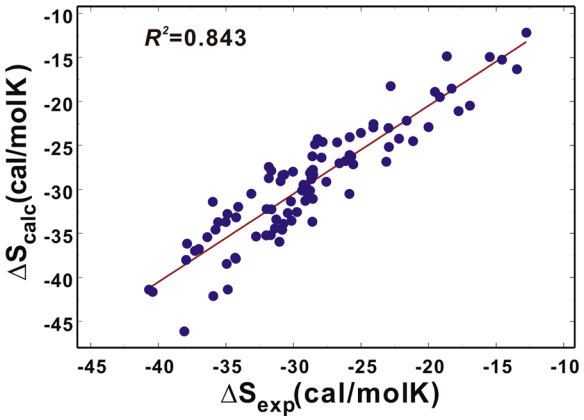


Figure 6. Linear correlation diagram between the experimental and calculated hydration entropies of 90 compounds at 298 K. The slope and intercept of the fitting are 1.011 and −0.148, respectively.

Most solute molecules with polar group(s) seem to have negative ΔS_{elec} values, except for acetonitrile, propanenitrile, 3-hydroxybenzaldehyde, and 3-hydroxybenzonitrile (see Table 1), which indicates a loss of configurational entropy in the electrostatic complexation between the solute and the water molecules. Because the hydration of a hydrophilic solute is thermodynamically favorable, the entropic penalties both for the formation of a cavity and for the establishment of electrostatic interactions seem to be compensated with extra energetic stabilizations of the polar groups in water. ΔS_{elec} values of 3-hydroxybenzaldehyde and 3-hydroxybenzonitrile are likely to be overestimated in FEP calculations, because they are positive (albeit quite small), despite the possession of two polar groups. More-serious overestimations of ΔS_{elec} are observed in acetonitrile and propanenitrile (see Table 1). This is quite unexpected, because the polar −C≡N moiety can establish a hydrogen bond with water in the ordered form. Therefore, the good agreement between the computational and experimental ΔS_{hyd} values of both nitrile compounds is most likely to be achieved by error cancellation in the ΔS_{cav} and ΔS_{elec} results.

In Table 1, it is also worth noting that the polar solutes with a phenolic −OH moiety have very small negative ΔS_{elec} values. The average of the ΔS_{elec} values for phenol, *o*-cresol, and *p*-cresol is −0.40 cal mol^{-1} K^{-1}, while that of the other polar solutes with at least one hydrogen-bonding group amounts to −3.64 cal mol^{-1} K^{-1}. This may be understood in terms of the asymmetric distribution of negative and positive charges in water molecule, which makes the polar groups hydrated differently according to the sign and magnitude of atomic charge. In this regard, the solute atoms with large negative charges were found to be hydrated in preference to those with positive charges.[53] Because it is peculiar to the phenolic oxygen that one of its lone electron pairs tends to delocalize into the adjacent phenyl ring, its low negative ΔS_{elec} value is consistent with the weakening of electrostatic interactions with water molecules caused by the decrease of the negative charge on the phenolic oxygen.

Next, we compare the ΔS_{hyd} values of all 90 solute molecules given by ΔS_{cav} plus ΔS_{elec} values calculated with SPT (or IT) and MD-based FEP simulations, respectively, to the corresponding experimental ones. The linear correlation diagram between the experimental and calculated results is shown in Figure 6. Overall, the estimated hydration entropies compare reasonably well with the experimental ones with the associated RMSE of 2.68 cal mol^{-1} K^{-1}, which corresponds to 9.2% of the average of the absolute experimental values. In this fitting, the R^2 value amounts to 0.843, with slope and intercept values of 1.011 and −0.148, respectively. The present method for estimating ΔS_{hyd} is found to outperform the Langevin dipoles solvation model, in terms of both R^2 and RMSE values.[13] The ΔS_{hyd} results in this study are also in better agreement with the experimental ones than those obtained by thermodynamic decomposition analysis of the ΔA_{hyd} values calculated from full FEP simulations.[14,17] Furthermore, our hybrid SPT/FEP hydration model shows higher performance in estimating ΔS_{hyd} than the grid cell theory in which the R^2 and RMSE values for 12 solutes amounted to 5.089 cal mol^{-1} K^{-1} and 0.675, respectively.[18] These comparisons imply that the combination of SPT (or IT) and electrostatic perturbation may serve as a new efficient computational tool for estimating the hydration entropies of organic molecules.

The largest differences between the experimental and calculated ΔS_{hyd} values are observed in a few polar solutes, including 4-heptanol, cycloheptanol, and 3-hexanol: the deviations are >6.0 cal mol^{-1} K^{-1}. The present computational method also shows a poor performance in estimating the ΔS_{hyd} values of alkane compounds with 5−7 carbons (pentane, hexane, and heptane). with the difference between experimental and calculated results being >4.5 cal mol^{-1} K^{-1}. For the better prediction of ΔS_{hyd}, therefore, some modifications seem to be required both in the SPT/IT approach, to compute ΔS_{cav}, and in FEP calculation of ΔS_{elec}. In case of *N*-methylacetamide that has been actively studied as a simplified structural model for protein backbone,[54] our method yields an error of only 8.0% in the estimation of ΔS_{hyd}, compared to 30.4% in full FEP simulation results of Wan et al.[37]

The overall accuracy of the calculated ΔS_{hyd} values of hydrocarbons seems to be similar to that of polar solutes with the associated RMSE values of 2.80 and 2.64 cal mol^{-1} K^{-1}, respectively. In the case of hydrocarbons, significant errors may be accumulated in ΔS_{cav}, because of the imperfect conformational sampling in MD simulations, as well as the ambiguity in choosing a representative structure. Such difficulties in the structural characterization of hydrocarbons basically stems from a large number of conformational degrees of freedom afforded by the immiscibility with water molecules. On the other hand, the inaccuracy of ΔS_{hyd} values of some polar molecules may be attributed to the difficulty in unraveling the effects of solute−

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

water interactions on the configurational change of water molecules. Because polar solutes interact directly with the water molecules in the first hydration shell, they induce the larger configurational change than hydrophobic solutes. Indeed, the ordered first hydration shell waters were found to fractionate into not only a more highly ordered but also into a more disordered component in the presence of polar solutes.[55]

Despite the inherence of uncertainty in ΔS_{cav} and ΔS_{elec} values, the hybrid SPT/FEP method seems to significantly outperform the full FEP method, because of the better description of the solute−water van der Waals interaction term, which makes a much larger contribution to ΔS_{hyd} than the electrostatic counterpart. To address this hypothesis, we compare the deviations of the ΔS_{hyd} values of 18 solute molecules calculated with the hybrid SPT/FEP and full FEP methods, with respect to the experimental ones. In the full FEP calculations of ΔS_{hyd}, we used a method that was identical to that adopted in the hybrid SPT/FEP approach to warrant a fair comparison. The ΔS_{hyd} values calculated in both forward and backward FEP simulations are presented in the Supporting Information. The standard deviations for ΔS_{hyd} values range from 0.18 cal mol^{-1} K^{-1} to 5.02 cal mol^{-1} K^{-1}, which corresponded to 0.3%−8.7%, with respect to the actual ΔS_{hyd} values. As shown in Table 2, the hybrid SPT/FEP method seems to predict the ΔS_{hyd} values of all of the solute molecules with even higher accuracy than the full FEP method. The average of deviations amounts to only 2.01 cal mol^{-1} K^{-1} in the former, compared to 20.42 cal mol^{-1} K^{-1} in the latter. These comparisons confirm the effectiveness of the separate calculations of van der Waals and electrostatic interaction terms with SPT (or IT) and FEP methods, respectively, in the estimation of molecular hydration entropy.

Table 2. Comparison of the Discrepancies of the ΔS_{hyd} Values Obtained with Hybrid SPT/FEP and Those with Full FEP Method, with Respect to the Experimental Ones[a]

	Deviation from Experimental Data (cal mol^{-1} K^{-1})			Deviation from Experimental Data (cal mol^{-1} K^{-1})	
solute	full FEP	hybrid SPT/FEP	solute	full FEP	hybrid SPT/FEP
1	5.59	2.98	**10**	18.70	2.40
2	10.13	2.86	**11**	32.41	1.23
3	15.05	3.71	**12**	30.76	4.44
4	30.11	2.18	**13**	32.37	1.74
5	14.21	0.45	**14**	7.50	3.21
6	13.99	0.05	**15**	29.04	0.06
7	37.21	0.26	**16**	26.13	4.37
8	26.71	0.01	**17**	16.39	5.37
9	19.30	0.51	**18**	1.95	0.42

[a]See Figure 3 for the identification of solute molecules.

Because the end-point catastrophe is not observed explicitly for 18 representative solute molecules, the large errors in the ΔS_{hyd} values calculated with the full FEP method may be attributed to the fact that the current NAMD force fields are not optimized adequately for ΔS_{hyd} calculations. Therefore, the development of the new potential parameters to enhance the computational accuracy would be prerequisite for the full FEP calculations to be useful for ΔS_{hyd} prediction. Because the imperfection of the van der Waals interaction term proves to be responsible for the large error of ΔS_{hyd} estimation in this study, the reoptimization of force fields should be focused on the depths of the potential energy wells and the equilibrium separations between the solute and solvent atoms.

Finally, we turn to the comparison of the ΔG_{hyd} data obtained by combining the calculated ΔS_{hyd} and ΔH_{hyd} values with the corresponding experimental ΔG_{hyd} data. This comparative analysis was performed to validate the usefulness of the ΔS_{hyd} values calculated with the hybrid SPT/FEP method for the enrichment of ΔG_{hyd} data. To compute the ΔH_{hyd} values of 90 molecules listed in Table 1, we employed the solvent-contact model suggested by Stouten et al.[56] Within the framework of the solvent−contact model, ΔH_{hyd} for a solute molecule can be expressed as follows.

$$\Delta H_{hyd} = \sum_{i}^{atoms} S_i \left[O_i^{max} - \sum_{j \neq i}^{atoms} V_j e^{r_{ij}^2/(2\sigma^2)} \right] \quad (11)$$

Here, the Gaussian-type envelope function, with respect to the interatomic distance between solute atoms (r_{ij}), and a constant (σ = 3.5 Å) are employed to define the occupied volume to which the approach of solvent molecules is restricted. S_i, O_i^{max}, and V_j represent the atomic hydration energy per unit volume, the maximum atomic occupancy, and the atomic fragmental volume, respectively. These three atomic parameters for all the atom types available in the dataset were optimized with a standard genetic algorithm, using the training set, which is comprised of 45 molecules. All details for the optimized atomic parameters and the calculated ΔH_{hyd} values of 90 solute molecules under investigation were provided in the Supporting Information (Tables S7 and S8).

The calculated ΔH_{hyd} values were then combined with the ΔS_{hyd} data obtained with the hybrid SPT/FEP method to estimate the molecular ΔG_{hyd} values, which were compared with the corresponding experimental ΔG_{hyd} results. As shown in Figure 7, the calculated ΔG_{hyd} values seem to be in good agreement with the experimental values, with associated R^2 and RMSE values of 0.855 and 1.15 kcal/mol, respectively. The slope and intercept of the linear correlation diagram amount to 0.978 and 0.173, respectively. Judging from such a reasonably good fit, the combination of ΔH_{hyd} and ΔS_{hyd} values calculated with the solvent−contact model and the hybrid SPT/FEP method, respectively, seems to be useful for estimating the ΔG_{hyd} values of small solute molecules.

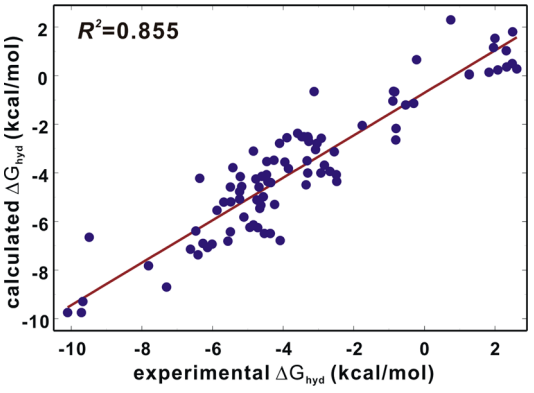


Figure 7. Linear correlation diagram between the experimental ΔG_{hyd} values and those obtained by combining the calculated ΔS_{hyd} and ΔH_{hyd} values.

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Despite the significant improvement of accuracy in predicting the hydration entropies of organic molecules, we expect that the present hybrid SPT/FEP method can be further modified in several aspects. First, the decomposition of the hydration entropy into a cavity term and an electrostatic term may be a rough approximation, because of the complexity in solute−water interactions. This is likely to be the case for large solute molecules whose structures in water are determined by a compromise between van der Waals and electrostatic interactions. Therefore, a change of the theoretical framework may be required for accurate estimation of the hydration entropies of large solute molecules.

The use of Lennard-Jones potential in SPT calculations instead of the simple hard-sphere model is most likely to lead to the better estimation of the molecular ΔS_{cav} values, because of the better description of solute−water interactions. In this case, however, the calculation of ΔS_{cav} seems to be more difficult than in the hard-sphere model, because of the lack of the analytical expression for ΔS_{cav}. Therefore, extensive atomistic simulations are needed to calculate the various contributions to ΔS_{cav} in the mixtures of Lennard-Jones particles.

It should also be taken into consideration that only the nonbond electrostatic interactions are evaluated in FEP simulations, despite the inherence of covalent character in strong hydrogen bonds. Therefore, the ΔS_{elec} values of the solutes that can form strong hydrogen bond(s) with water are likely to be less negative than the actual values. This problem is expected to be resolved, in large part, by the use of a hybrid quantum mechanics/molecular mechanics (QM/MM) method, instead of a classical MD simulation in FEP calculations, because the former can model the electron-transfer interactions between hydrogen-bond donor and acceptor groups. Finally, the use of high-quality atomic charges would also enhance the ΔS_{elec} calculation results, because only the electrostatic solute−water interactions are considered in the FEP simulations. Therefore, the atomic charge calculations must be performed using the sophisticated quantum chemical methods at the post-Hartree−Fock level, such as configuration interaction, Møller−Plesset perturbation, and coupled cluster theories. In this case, it should be kept in mind that that the accuracy of our hybrid SPT/FEP method for the estimation of S_{hyd} may be limited by decomposing the potential function whose parameters were optimized simultaneously. Therefore, to enhance the accuracy of S_{hyd} prediction, all of the potential parameters in SPT and FEP terms must be reoptimized in such a way to minimize the difference between experimental and calculated S_{hyd} values. Our future studies on the hydration entropy will focus on the further improvement of the accuracy under consideration of the above-mentioned points.

4. CONCLUSIONS

Molecular hydration entropies could be predicted with reasonable accuracy by means of the hybrid scaled-particle theory/free-energy perturbation (SPT/FEP) method to separately calculate the cavity formation entropy (ΔS_{cav}) as the hydrophobic contribution and electrostatic solute−water interaction term (ΔS_{elec}). Because of the effective partitioning of the two entropic components, the prediction accuracy of ΔS_{hyd} was enhanced to a substantial extent, when compared to the conventional full FEP calculations. This improvement was made possible by successfully overcoming the difficulty in calculating ΔS_{vdW} in full FEP calculations. SPT and IT exhibited almost the same performance in the estimation of ΔS_{cav}. ΔS_{cav} was found to dominate over ΔS_{elec} in magnitude not only for hydrocarbons but also for polar solutes. This implies that most entropic costs for the hydration of a solute is spent for the formation of solvent-excluded volume, rather than for the direct solute−water interactions. The entropic features of hydration found in this study seem to serve as useful information in drug discovery. In order to optimize the binding affinity of a ligand for the target protein, for example, the ligand structure should be modified in such a way to maximize the strength of attractive protein−ligand interactions and simultaneously to maximize the entropic penalty for cavity formation in water to minimize the desolvation cost for protein−ligand binding.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00325.

Tables listing various entropy and hydration free-energy values of solute and reference molecules, as well as optimized atomic fragmental volume, maximum atomic occupancy, and atomic solvation parameters for various atom types (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +82-2-3408-3766. Fax: +82-2-3408-3334. E-mail: hspark@sejong.ac.kr.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Author Contributions

[§]These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by Ministry of Education, Science and Technology (No. NRF-2011-0022858). We also thank Kwanghyewon Medical Foundation for the support of this study.

■ REFERENCES

(1) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. *J. Comput.-Aided Mol. Des.* **2014**, *28* (3), 135−150.

(2) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119* (3), 1661−1670.

(3) Hagen, D. E.; Suck Salk, S. H. *Chem. Phys.* **1986**, *102* (3), 459−466.

(4) Ahmad, M.; Helms, V.; Lengauer, T.; Kalinina, O. V. *J. Chem. Theory Comput.* **2015**, *11* (4), 1410−1418.

(5) Reynolds, C. H.; Holloway, M. K. *ACS Med. Chem. Lett.* **2011**, *2* (6), 433−437.

(6) Olsson, T. S.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. *J. Mol. Biol.* **2008**, *384* (4), 1002−1017.

(7) Kavanagh, K. L.; Guo, K.; Dunford, J. E.; Wu, X.; Knapp, S.; Ebetino, F. H.; Rogers, M. J.; Russell, R. G. G.; Oppermann, U. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (20), 7829−7834.

(8) Kaul, M.; Barbieri, C. M.; Pilch, D. S. *J. Mol. Biol.* **2005**, *346* (1), 119−134.

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(9) Carbonell, T.; Freire, E. *Biochemistry* **2005**, *44* (35), 11741−11748.
(10) Yanchunas, J., Jr.; Langley, D. R.; Tao, L.; Rose, R. E.; Friborg, J.; Colonno, R. J.; Doyle, M. L. *Antimicrob. Agents Chemother.* **2005**, *49* (9), 3825−3832.
(11) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81* (4), 2016−2027.
(12) Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1994**, *98* (15), 4141−4147.
(13) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103* (46), 10282−10288.
(14) Wyczalkowski, M. A.; Vitalis, A.; Pappu, R. V. *J. Phys. Chem. B* **2010**, *114* (24), 8166−8180.
(15) Jedlovszky, P.; Idrissi, A. *J. Chem. Phys.* **2008**, *129* (16), 164501.
(16) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104* (26), 6271−6285.
(17) Kubo, M. M.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* **1997**, *101* (49), 10527−10534.
(18) Gerogiokas, G.; Calabro, G.; Henchman, R. H.; Southey, M. W. Y.; Law, R. J.; Michel, J. *J. Chem. Theory Comput.* **2014**, *10* (1), 35−38.
(19) Pitera, J. W.; van Gunsteren, W. F. *Mol. Simul.* **2002**, *28* (1), 45−65.
(20) Blondel, A. *J. Comput. Chem.* **2004**, *25* (7), 985−993.
(21) Genheden, S.; Kongsted, J.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2010**, *6* (11), 3558−3568.
(22) Gapsys, V.; Seeliger, D.; de Groot, B. L. *J. Chem. Theory Comput.* **2012**, *8* (7), 2373−2382.
(23) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. *J. Chem. Phys.* **1959**, *31* (2), 369−380.
(24) Pierotti, R. A. *Chem. Rev.* **1976**, *76* (6), 717−726.
(25) Ashbaugh, H. S.; Pratt, L. R. *Rev. Mod. Phys.* **2006**, *78* (1), 159−178.
(26) Floris, F. M.; Selmi, M.; Tani, A.; Tomasi, J. *J. Chem. Phys.* **1997**, *107* (16), 6353−6365.
(27) Lebowitz, J. L.; Helfand, E.; Praestgaard, E. *J. Chem. Phys.* **1965**, *43* (3), 774−779.
(28) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6* (1), 151−176.
(29) Mobley, D. L.; Dill, K. A.; Chodera, J. D. *J. Phys. Chem. B* **2008**, *112* (3), 938−946.
(30) Petitjean, M. *J. Comput. Chem.* **1994**, *15* (5), 507−523.
(31) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (17), 8951−8955.
(32) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(34) Powell, M. J. D. A Hybrid Method for Nonlinear Equations. In *Numerical Methods for Nonlinear Algebraic Equations*; Rabinowitz, P., Ed.; Gordon and Breach: New York, 1970; pp 87−114.
(35) Yu, H. A.; Karplus, M. *J. Chem. Phys.* **1988**, *89* (4), 2366−2379.
(36) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531−567.
(37) Wan, S.; Stote, R. H.; Karplus, M. *J. Chem. Phys.* **2004**, *121* (19), 9539−9548.
(38) Smith, D. E.; Haymet, A. D. J. *J. Chem. Phys.* **1993**, *98* (8), 6445−6454.
(39) Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. *J. Chem. Phys.* **2004**, *120* (6), 2652−2661.
(40) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22* (8), 1420−1426.
(41) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. *J. Comput. Chem.* **2010**, *31* (4), 671−690.
(42) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30* (10), 1545−1615.
(43) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110* (35), 17616−17626.
(44) Bruckner, S.; Boresch, S. *J. Comput. Chem.* **2011**, *32* (7), 1303−1319.
(45) Pohorille, A.; Jarzynski, A.; Chipot, C. *J. Phys. Chem. B* **2010**, *114* (32), 10235−10253.
(46) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100* (12), 9025−9031.
(47) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222* (6), 529−539.
(48) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10* (8), 563−595.
(49) Haynes, W. M. *CRC Handbook of Chemistry and Physics*; CRC Press: Boca Raton, FL, 2012.
(50) Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (6), 8800−8803.
(51) Madan, B.; Lee, B. *Biophys. Chem.* **1994**, *51* (2−3), 279−289.
(52) Durell, S. R.; Wallqvist, A. *Biophys. J.* **1996**, *71* (4), 1695−1706.
(53) Mobley, D. L.; Baker, J. R.; Barber, A. E., II; Fennell, C. J.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112* (8), 2405−2414.
(54) Graziano, G. *J. Phys. Soc. Jpn.* **2000**, *69* (11), 3720−3725.
(55) Godec, A.; Smith, J. C.; Merzel, F. *Phys. Rev. Lett.* **2011**, *107* (26), 267801.
(56) Stouten, P. F. W.; Frömmel, C.; Nakamura, H.; Sander, C. *Mol. Simul.* **1993**, *10*, 97−120.

DOI: 10.1021/acs.jctc.5b00325
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX