Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database

Guilherme Duarte Ramos Matos,[†] Daisy Y. Kyu,[‡] Hannes H. Loeffler,[§] John D. Chodera,[∥] Michael R. Shirts,[⊥] and David L. Mobley*,[†,‡]

[†]Department of Chemistry, University of California, Irvine, California 92697, United States
[‡]Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, United States
[§]Scientific Computing Department, STFC, Daresbury WA4 4AD, U.K.
[∥]Computational and Systems Biology Program, Sloan Kettering Institute, New York, New York 10065, United States
[⊥]Department of Chemical and Biological Engineering, University of Colorado, Boulder, Colorado 80309, United States

ⓢ *Supporting Information*

ABSTRACT: Solvation free energies can now be calculated precisely from molecular simulations, providing a valuable test of the energy functions underlying these simulations. Here we briefly review "alchemical" approaches for calculating the solvation free energies of small, neutral organic molecules from molecular simulations and illustrate these approaches by applying them to calculate aqueous solvation free energies (hydration free energies). These approaches use a nonphysical pathway to compute free energy differences from a simulation or set of simulations and appear to be a particularly robust and general-purpose approach for this task. We also present an update (version 0.5) to our FreeSolv database of experimental and calculated hydration free energies of neutral compounds and provide input files in formats for several simulation packages. This revision to FreeSolv provides calculated values generated with a single protocol and software version rather than the heterogeneous protocols used in the prior version of the database. We also further update the database to provide calculated enthalpies and entropies of hydration and some experimental enthalpies and entropies as well as electrostatic and nonpolar components of solvation free energies.
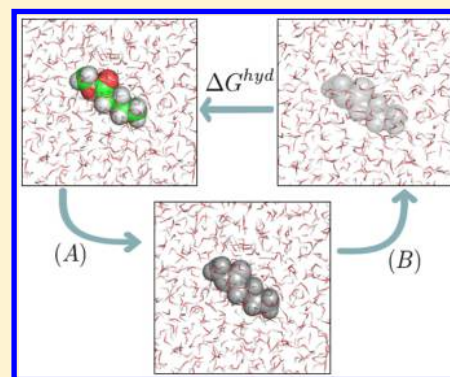
INTRODUCTION

The solvation free energy gives the free energy change associated with the transfer of a molecule between ideal gas and solvent at a certain temperature and pressure. While solvation free energies (ΔG^{solv}) in general and hydration free energies (ΔG^{hyd}, solvation in water) in particular might not seem to have far-reaching implications, in fact, researchers in diverse areas can benefit from their prediction because such solvation free energies are related to a broad range of physical properties such as infinite-dilution activity coefficients, Henry's law constants, solubilities, and distributions of chemical species between immiscible solvents or different phases.

Solvation free energies are differences in thermodynamic potentials that describe the relative populations of a chemical species in solution and the gas phase at equilibrium.[1,2] In the thermodynamic limit in the solvated phase and the ideal gas limit in the gas phase, ΔG^{solv} of component i is equal to $\mu_{i,\text{solv}} - \mu_{i,\text{gas}}$, the difference in the chemical potentials in the two phases. In the additional limit of one molecule of component i at infinite dilution, these become the infinite-dilution excess chemical potentials in the respective solvents.

Solvation free energies not only tell us how much a molecule prefers one phase over another but also can provide insight into how a solvent behaves in different environments. For example, water solvates molecules of opposite polarity differently because of its inherent asymmetry,[3] surfaces also have asymmetric effects on ion pairing that depend on the curvature of the surface,[4] and the molecular geometry and chemical environment affect hydrophobic solvation.[5] Although they can be difficult to measure experimentally, ΔG^{solv} and ΔG^{hyd} can be calculated to a precision of better than 0.4 kJ·mol^{-1}, even with a relatively modest investment of simulation time, for relatively diverse small neutral molecules[6] such as those seen in the FreeSolv database of hydration free energies[7] and in recent blind challenges such as the Statistical Assessment of the Modeling of Protein and Ligands (SAMPL) challenges. These challenges aim to improve the quality of predictive computational tools in drug design[1,6,8−21] and have leveraged solvation free energies to help drive improvements in modeling.

Since the solvation free energy of neutral compounds is an aggregate measure of many competing interactions and entropic effects that can span a wide energy range, comparison

of computed solvation free energies with experiment has proven to be an exacting test of force field quality that has been useful in revealing deficiencies in small-molecule force fields.[3,22,23] The relative ease with which solvation free energies can be calculated—as opposed to protein−ligand binding free energies, which are fraught with a variety of sampling issues— also makes them attractive for this purpose (but see the Supporting Information for how protonation state/tautomer challenges may apply here, as in protein−ligand binding). For instance, SAMPL has frequently (in SAMPL1 through SAMPL4) included blind predictions of hydration free energies in particular.[1,8−14] However, to our knowledge, no laboratories are currently measuring hydration free energies, leading the field to search for other simple physical properties that can be rapidly computed—such as relative solubilities,[24] distribution coefficients,[25] and solvation free energies in organic solvents[26]—as a tool to assess and improve small-molecule force fields. In computational chemistry, hydration free energies are of particular importance because they are frequently used in force field parametrization[26−29] and in the testing of free energy methods and force fields.[1,8−14,30−37] Furthermore, computed free energies are in some cases found to be accurate enough to highlight problems with experiments and assist in curation of experimental data.[13,38]

Solvation free energies are often calculated by alchemical free energy methods,[39] which simulate a series of nonphysical intermediates to compute the free energy of transferring a solute from solution to the gas phase (as here) or vice versa. This alchemical path provides an efficient way to move the solute from solution to the gas phase by perturbing its interactions in a nonphysical way. Since free energy changes are path-independent, this nonphysical process still yields the free energy change for transfer of the solute from solvent to gas.[39,40] The path is formed by constructing intermediate states with interactions that modulate between the end states of interest, with the variable λ parametrizing progress along the path. A particularly efficient set of intermediate states uses a two-step process, first turning off the van der Waals interactions using one parameter, λ_v, and then turning off the electrostatic interactions using a second parameter, λ_e. Here we compute the free energy change to transition between each pair of λ values, and the overall free energy change is the sum of these pairwise differences.

While other approaches have been used to calculate solvation free energies,[41] alchemical free energy calculations using explicit solvent have become a mainstream approach,[42,43] in part because of their formal rigor. Alternative approaches include implicit-solvent models,[34−37,44] which yield ΔG^{hyd} but do not take into consideration the solvent configuration around the solute, and Monte Carlo-based approaches using the Gibbs ensemble[45−51] and expanded ensemble,[52] though these are most commonly used for molecules that are particularly small and/or rigid.

HYDRATION AND SOLVATION FREE ENERGIES HAVE A RANGE OF APPLICATIONS

The activity coefficient γ_i of a solute species i can be calculated from ΔG^{solv} as follows:

$$\gamma_i = \exp\left(\frac{\mu_i^{excess}}{RT}\right) = \exp\left(\frac{\Delta G_i^{solv}}{RT}\right) \tag{1}$$

where μ_i^{excess} is the excess chemical potential of i and is equal to ΔG_i^{solv} in the ideal gas limit of the vapor phase, R is the universal gas constant, and T is the absolute temperature. For instance, solvation free energies are used to estimate infinite-dilution activity coefficients (γ_i^∞) in many solvents by using a single molecule of solute i.[53−59] Experimental results obtained from gas chromatography[60,61] can be compared to γ_i^∞ obtained from ΔG^{solv} to further test models and methodologies that use these free energy calculations.

Solubility prediction is another field where $\Delta G^{solv/hyd}$ prediction can have great value. One methodology computes the solubility free energy by computing both the sublimation free energy (from solid to gas) and hydration free energy (from gas to water).[62] Another way to predict molecular solid solubilities depends on excess chemical potential calculations. The chemical potential, μ, of a species is calculated at different concentrations to build the concentration-dependent chemical potential curve of solutions[63−66] in order to discover phase equilibrium conditions. Free energies of solvation in pure melts and pure amorphous matter have been used to find upper bounds for solubilities given that most druglike compounds have crystal polymorphs.[67−70] Relative solubilities of a given chemical species in different solvents can also be assessed with these calculations.[24,71] Henry's law solubility constants[72,73] and solubilities in supercritical fluids[74] can also be predicted using solvation free energies.

The latest SAMPL challenge, SAMPL5, included blind prediction of distribution coefficients between cyclohexane and water for 53 solutes.[32,33,75,76] Distribution and partition coefficients are important properties for toxicology and pharmacology because they play a major part in predicting absorption and distribution of a substance in different tissues.[77] Partition coefficients—which are the distribution coefficients of the neutral form of a compound—can be estimated from the difference between the solvation free energies of the neutral form of the chemical species in two different solvents,[21] as shown in eq 2:

$$\log_{10} P_{A\to B} = \frac{\Delta G^{solv,A} - \Delta G^{solv,B}}{RT \ln(10)} \tag{2}$$

where $\Delta G^{solv,A}$ and $\Delta G^{solv,B}$ are the solvation free energies of a molecule in solvents A and B, respectively. While in principle the calculation could be done by transferring the solute between phases, in many software implementations it is more straightforward to simply compute the solvation free energy in each phase separately, or the free energy of removing the solute from each phase. Thus, solvation free energy calculations have found relatively widespread application in calculating partition coefficients, including in SAMPL5.[15−21] Hydration free energies themselves are valuable quantities in drug design[43,78] and can be used to understand the impact of ligand desolvation on the binding process[79,80] or can be utilized as QSAR descriptors.[81]

THEORY AND PRACTICAL ASPECTS OF ALCHEMICAL CALCULATIONS

Solvation free energies can be calculated in various ways. In this paper we focus on alchemical free energy calculations, which have been one of the most consistently reliable methods in recent applications such as the SAMPL series of challenges.[1,8−14,25] Consider a pair of end states A and B and

their respective Hamiltonians $\mathcal{H}_A(\mathbf{q}, \mathbf{p}; \lambda)$ and $\mathcal{H}_B(\mathbf{q}, \mathbf{p}; \lambda)$. Then

$$\mathcal{H}(\mathbf{q}, \mathbf{p}; \lambda) = f(\lambda)\mathcal{H}_A(\mathbf{q}, \mathbf{p}; \lambda) + g(\lambda)\mathcal{H}_B(\mathbf{q}, \mathbf{p}; \lambda) \qquad (3)$$

where $f(\lambda)$ and $g(\lambda)$ are functions of λ used to mix the Hamiltonians, typically set such that $\mathcal{H} = \mathcal{H}_A$ at $\lambda = 0$ and $\mathcal{H} = \mathcal{H}_B$ at $\lambda = 1$, and $\mathbf{q}$ and $\mathbf{p}$ represent all of the positions and momenta of the system. With $\mathcal{H}(\mathbf{q}, \mathbf{p}; \lambda)$ we can calculate the free energy difference between A and B:

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad (4)$$

This method, called thermodynamic integration (TI),[82] is implemented in practice via a numerical quadrature approach after simulations are done at a discrete set of λ values. It performs similarly to more efficient methods when the integrand is smooth.[83−85] However, it can break down when the integrand is not smooth, and it can be difficult to capture numerical integration errors in resulting uncertainty estimates.

Exponential averaging (EXP), also known as free energy perturbation (FEP), was introduced by Zwanzig.[86] In this method, the free energy difference between two states A and B is given by

$$\Delta G = -\frac{1}{\beta} \ln \langle e^{-\beta[\mathcal{H}_B(\mathbf{q},\mathbf{p};\lambda) - \mathcal{H}_A(\mathbf{q},\mathbf{p};\lambda)]} \rangle_A \qquad (5)$$

where $\beta = (k_B T)^{-1}$. Although eq 5 is exact in the limit of large numbers of samples, EXP is inefficient and particularly sensitive to the tails of the relevant distributions, leading to unstable free energy estimates and other large biases when configurations sampled in one state are very unlikely to be found in the other state and vice versa. The probability that describes this likelihood is called the phase-space overlap between the two states. EXP convergence is far from ideal, requiring states to have sizable phase-space overlap with one another.[39,84,87] Thus, addition of intermediate states (with values of λ between 0 and 1) can improve the overlap dramatically and thus the quality of the final result.[88] Another issue is an asymmetric bias depending on which direction the free energy difference estimation is performed,[89,90] so other analysis methods are now preferred.[39] In the limit of adequate sampling, EXP converges to the same free energy value in both directions, but there are other ways to calculate free energies more efficiently.

An alternate method, Bennett's acceptance ratio (BAR), uses the information from both directions to derive the following relationship (which can and has been written in numerous ways):

$$\left\langle \frac{1}{1 + \frac{N_A}{N_B}e^{\beta \Delta \mathcal{H}_{BA}(\mathbf{q},\mathbf{p}) - \beta \Delta G}} \right\rangle_A$$
$$= \left\langle \frac{1}{1 + \frac{N_B}{N_A}e^{\beta \Delta \mathcal{H}_{AB}(\mathbf{q},\mathbf{p}) + \beta \Delta G}} \right\rangle_B \qquad (6)$$

where N_A and N_B are the numbers of statistically independent samples gathered from states A and B, respectively, and $\Delta \mathcal{H}_{BA}(\mathbf{q}, \mathbf{p}) = \mathcal{H}_B(\mathbf{q}, \mathbf{p}) - \mathcal{H}_A(\mathbf{q}, \mathbf{p}) = -\Delta \mathcal{H}_{AB}(\mathbf{q}, \mathbf{p})$ are the Hamiltonian differences between the states at a given point in phase space.

This expression minimizes the free energy variance[91] and makes BAR much more efficient than EXP.[89,90] Multistate Bennett acceptance ratio (MBAR) is an extension of BAR that considers the overlap between a given state and all of the others in the path between the end states.[92] BAR and MBAR perform similarly when the spacing between intermediate states is moderate[85] and therefore only neighboring states have phase-space overlap. The weighted histogram analysis method (WHAM)[93,94] is essentially an approximation to MBAR and thus also gives very similar results when carefully done with appropriately small bins. MBAR performs consistently well and indeed is perhaps the most consistently well-performing free energy estimator.[85] Thus, we recommend it as the analysis method of choice whenever possible. TI usually is more sensitive to the choice and number of intermediate states than BAR,[95] but it can perform as well as BAR and MBAR if the integrand is smooth.[83,84] EXP should generally be avoided because of its asymmetric bias and sensitivity to the tails of the distribution.[89]

Choice of Alchemical Pathway. Alchemical free energy calculations were given this name because the pathway involves unphysical changes to the atomic identities, such as to the interactions between components.[43,96,97] Solvation free energy calculations can use several different approaches to modulate the interactions. One approach, called *decoupling*, modulates only the interactions between the solute and its surroundings, retaining internal interactions (the approach we use here). An alternative approach, called *annihilation*, removes internal nonbonded interactions within the solute as well as those with the surroundings. Mixtures of the two approaches are also possible, such as annihilating internal electrostatic interactions while decoupling nonpolar interactions. Here, three main thermodynamic states are considered: a single, noninteracting molecule of the solute in a box of solvent; a solute molecule that interacts with its surroundings through nonpolar (dispersion and repulsion) forces; and a fully interacting system, in which solvent molecules interact with the solute molecule through both electrostatic and nonpolar (dispersion and exclusion) forces. Simulations are then conducted over a series of intermediates connecting these states, going through a phase that changes electrostatic interactions only and another phase that modifies van der Waals interactions only (Figure 1). Each of these intermediates has high configuration space overlap with at least neighboring states, allowing precise calculation of free energy differences.[98−101]

The most straightforward way to switch between states is the linear pathway

$$\mathcal{H}(\mathbf{q}, \mathbf{p}; \lambda) = (1 - \lambda)\mathcal{H}_A(\mathbf{q}, \mathbf{p}) + \lambda \mathcal{H}_B(\mathbf{q}, \mathbf{p}) \qquad (7)$$

but this pathway is in general problematic for solvation of all but the smallest molecules. This is the case because repulsive forces are often handled by a $1/r^{12}$ term (such as in the Lennard-Jones functional form), which leads to nonintegrable singularities in $\langle \partial \mathcal{H}/\partial \lambda \rangle$ at terminal λ values due to sudden changes in the potential at small r. This is not a problem that is specific to TI; rather, this issue can still result in numerical instabilities or large errors in calculated free energies even with other analysis approaches.[43,102,103] Thus, more complicated λ pathways are required, such as soft-core potentials, which should in general be used to avoid such numerical problems.[98,102,103] A common soft-core form for the Lennard-Jones potential between two particles i and j is
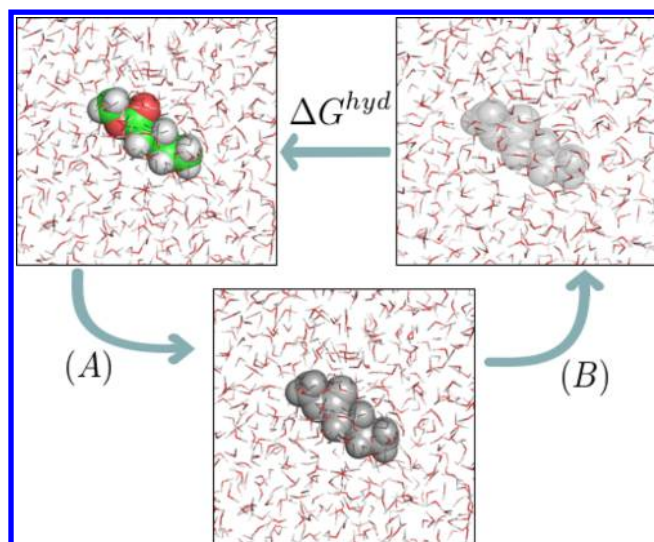
Figure 1. Thermodynamic cycle used to calculate hydration free energies (or, more generally, solvation free energies). In (A), we have states in which charge−charge interactions between the solute and its environment are progressively turned off. In (B), dispersion interactions between the solute and water are progressively turned off. Colored atoms (green for carbon, red for oxygen, white for hydrogen) have electrostatic and nonpolar interactions with the environment; gray atoms retain only nonpolar interactions; and transparent atoms have no interactions with their environment (and thus represent the solute in vacuum).

$$U_{ij}^{LJ}(r_{ij}, \lambda) = \lambda^n 4\epsilon_{ij}\left(\frac{1}{[\alpha(1-\lambda)^m + (r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\alpha(1-\lambda)^m + (r_{ij}/\sigma_{ij})^6}\right)$$

(8)

where ϵ_{ij} and σ_{ij} are the Lennard-Jones well-depth and length scale parameters, respectively, and α is a positive constant that should typically be set to 0.5.[103,104] The exponents m and n are most efficient at $n = 1$ and $m = 1$, but other values have been used too.[100,103−105] Improvements have been achieved by using new soft-core functions that ease the problem with additional minima within the formulation of the original soft-core potential[106] and alternate potentials that construct nearly optimal paths for alchemical simulations.[107] Linear basis functions can be used as an alternative to soft-core potentials that approaches the minimum variance possible over all pair potentials;[101,108] these can also enhance the efficiency of alchemical calculations.

The use of soft-core potentials promotes better convergence in many circumstances and provides much lower variance free energy estimates given a fixed amount of simulation time.[98,100,102,104,107] Thus, their use is highly recommended for successful free energy calculations. Without soft-core potentials, convergence is much more difficult or nearly impossible to achieve in many types of solvation free energy calculations.

Considerations for Successful Alchemical Calculations. The accuracy of these calculations is affected by at least three factors:[109,110] Is our sampling representative and adequate? Is the free energy estimator good enough? Is the force field adequate for the system? Have critical chemical effects been omitted from the simulation, such as protonation

state or tautomer effects? For solvation free energies of small molecules in solvents with relatively fast dynamics, such as water, sampling is typically adequate with a few nanoseconds of dynamics per λ window (at least for relatively rigid solutes), and the free energy estimators above are robust when applied carefully.

However, in the design of new studies, it is still important to choose robustly performing estimators and ensure adequate sampling. As discussed above, we recommend MBAR as the best and most reliable general-purpose estimator.[85] Sampling remains a critical issue,[109,111] both as the solute size and flexibility grow and as the solvent dynamics or environment becomes heterogeneous, for example, for solvation free energies in octanol, which can form local clusters of hydrophilic and hydrophobic sites,[21] or in mixed solvents.[25]

■ UPDATE OF FREESOLV, THE FREE COMMUNITY SOLVATION FREE ENERGY DATABASE

About FreeSolv. FreeSolv[7] is a hydration free energy database for neutral compounds (for additional discussion of why we focus on neutral compounds, see section I in the Supporting Information (SI)) that contains experimental and calculated hydration free energy values, SMILES strings, PubChem compound IDs, IUPAC names, and now (as of version 0.5, presented in this work) calculated enthalpies and entropies of hydration for 643 small organic molecules. The molecular weights for compounds in FreeSolv range from 16.06 Da (methane) to 498.88 Da (decachlorobiphenyl). The number of rotatable bonds runs as high as 12, but most of the compounds are largely rigid. Since experimental and calculated hydration free energies, ΔG^{hyd}, can be computed quite precisely for quantitative comparison, FreeSolv can provide information for force field development[26−29] and can assist the testing of new solvation free energy methods.[112,113] One example of the use of hydration free energies as target physical properties to fit in force field development is the Automated Force Field Topology Builder and Repository (ATB).[114−116] ATB is an online platform based in large part on FreeSolv and provides similar information, though with force field parameters of the GROMOS family. However, the database is not available in an easily downloadable public format and is only accessible via web queries. ATB partly relies on data taken from previous works by Mobley and collaborators,[117] which are included in FreeSolv.[7]

While calculated hydration free energies for all of the compounds have been available in FreeSolv since the database was constructed,[7] previous values had been calculated with somewhat heterogeneous protocols in a variety of different studies spread over roughly 10 years.[2,6,11,13,23,42,117,118] In this work, we have updated FreeSolv by repeating all of the calculations using a single protocol, now also computing enthalpies and entropies of hydration.

Method Details. We obtained FreeSolv's calculated hydration free energies using alchemical free energy calculations, connecting the end states (corresponding to the solute in vacuum and in solution) via a λ path with 20 intermediate states (full details are provided in the SI). The first five states correspond to changes in electrostatic interactions, while the last 15 modify the Lennard-Jones terms in the potential. This separation allows electrostatic interactions to be changed linearly and soft-core potentials to be used only when changing nonpolar interactions.[101] The box size does not affect the results of solvation free energy calculations as long as good

practices, which recommend that the box edges be at least twice the Lennard-Jones cutoff distance, are followed.[119] We ran 5 ns of Langevin dynamics per state with 2 fs time steps in GROMACS 4.6.7[120−125] at 298.15 K. Van der Waals interactions were neglected beyond a smoothly switched cutoff of 1.2 nm. Different cutoff radii are commonly used, but one should be aware that the choice of cutoff can affect calculated solvation free energies. However, long-range dispersion corrections can be employed (as here) to remove the cutoff dependence of the calculated free energies.[126] (It is worth noting that in the case of heterogeneous systems, such as for binding free energy calculations, it may be necessary to use reweighting techniques instead).[126] Our choice of soft core is the so-called 1−1−6 form (m and n equal to 1 in eq 8), which leads to statistical uncertainties of approximately the same size as the uncertainties from simulations using optimized-path soft cores.[104] The pressure was maintained at 1 atm using the Parrinello−Rahman barostat.[127] Enthalpy and entropy decomposition required 60 ns Langevin dynamics simulations, with 2 fs time steps at 298.15 K and 1 atm in water and in vacuo for each molecule in the database. These long simulations were necessary to reduce the error bars on the computed enthalpies to levels around 2.9 kJ·mol^{-1}, roughly the level of typical thermal energy ($1k_BT$), as done in, for example, host−guest binding calculations.[128] We used the default Langevin dynamics friction coefficients implemented in GROMACS (see the SI). The size of the friction coefficient affects only equilibration and correlation times and should not affect the calculated hydration free energies and enthalpies. In order to obtain consistent results, we used simulation boxes with 1309 water molecules and one solute molecule. The same system parameters and water model were used as in the free energy calculations. Full details can be found in the SI.

Input files for version 0.5 of FreeSolv were constructed from scratch from the isomeric SMILES strings for the compounds, which are deposited in the database. From these SMILES strings, we used the OpenEye Python toolkits[129−131] to generate molecular structures and assign AM1-BCC partial charges,[132,133] and then charged mol2 files were written out. The AMBER Antechamber package (as distributed with AMBER14) was then used to to assign parameters from the GAFF[20] small-molecule force field (version 1.7), and these were then converted to GROMACS format and solvated with the TIP3P water model.[134] The script that performs the setup and regenerates all of the input and molecular structure files in the database is available in the scripts directory of FreeSolv and provides full details. Following the calculations, MBAR hydration free energies were obtained using `alchemical-analysis.py` (github.com/mobleylab/alchemical-analysis).[97] Here we also introduce FreeSolv version 0.51, which extends version 0.5 by making the small-molecule parameter sets available additionally in CHARMM, DESMOND, and LAMMPS formats. Additional details can be found in the SI and in the FreeSolv README files.

FreeSolv Hydration Free Energy Results. Computed hydration free energies are compared with experimental values in Figure 2. In the calculations described in this study, we found an average error of 1.3 ± 0.3 kJ·mol^{-1}, a root-mean-square (RMS) error of 6.4 ± 0.3 kJ·mol^{-1}, an average absolute error of 4.7 ± 0.2 kJ·mol^{-1}, a Kendall τ value of 0.80 ± 0.01, and a Pearson R value of 0.933 ± 0.008, comparable to those in the original FreeSolv set,[7] though some individual compounds have reasonably significant discrepancies (see the SI). This level of
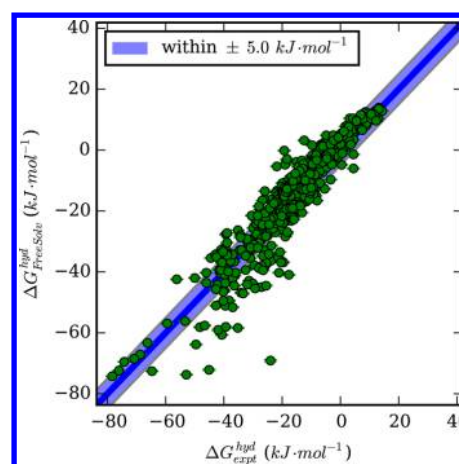

Figure 2. Calculated vs experimental hydration free energies for the compounds in FreeSolv version 0.5. Calculated values are on the vertical axis and experimental on the horizontal.

accuracy is consistent with what is often seen from classical fixed-charge force fields, which typically yield RMS errors of around 4−8 kJ/mol in computed hydration free energies.[43] We have previously used these data to address force field issues on hydroxyl groups[28] and also to highlight functional groups in the set that pose particular challenges.[42] Full details about which compounds have systematic errors, along with the functional groups represented in each compound, are present in the FreeSolv database itself.

In addition to experimental and calculated values, FreeSolv now includes the free energy of decoupling the solute−solvent electrostatic interactions (ΔG^q) and the free energy of decoupling the nonpolar interactions in water (ΔG^{vdW}) (available at github.com/mobleylab/FreeSolv). These quantities have been used for various purposes, including to assist in the study, development, and testing of implicit-solvent models.[135,136] However, it is important to remember that these components come from our particular decomposition of the free energy[137−140] and are not state functions; other decompositions are possible, so considerable care needs to be taken in interpreting these components. For example, annihilation rather than decoupling of Coulomb interactions would result in somewhat different decompositions because of electrostatics-induced conformational differences while van der Waals interactions are being decoupled.

Hydration Enthalpy Calculations. In addition to hydration free energies, we have also computed enthalpies (ΔH^{hyd}) and entropies of hydration (ΔS^{hyd}) and have added these to the database. Enthalpies of transfer, because of their larger dynamic range and lack of compensating entropic effects, are generally more sensitive to force field parameters than free energies[128,141,142] and thus can be sensitive probes of force field accuracy, providing an additional point of comparison with experiments. While only a few hydration enthalpies are available experimentally, there are a sufficient number to note that significant discrepancies between experiment and computed values exist for some compounds (Figure S2 and Table S1). We find that compounds having accurate hydration free energies do not necessarily have accurate hydration enthalpies and vice versa; for example, the calculated hydration free energy of benzene is within error of the experimental value, but the enthalpy is off by approximately 12 kJ/mol. In contrast, the hydration free energy of cyclohexanol is off by more than 5 kJ/

mol but the enthalpy is within error of the experimental value. Thus, clearly these quantities yield different information.

To compute hydration enthalpies, we used a difference in potential energies between a water box solvating the compound and a neat water box with the compound removed to vacuum:

$$\Delta H^{\text{hyd}} = \langle U_{\text{solution}} \rangle - (\langle U_{\text{water}} \rangle + \langle U_{\text{vacuo}} \rangle) \tag{9}$$

where $\langle U_{\text{solution}} \rangle$ is the internal energy of the solution (containing the solute), $\langle U_{\text{water}} \rangle$ is the internal energy of a box containing the same number of water molecules (under the same conditions) without the solute, and $\langle U_{\text{vacuo}} \rangle$ is the internal energy of the solute molecule alone in vacuum. We have neglected the pressure−volume contribution to the enthalpy, $P\Delta V$, since for solutes of this size the contribution is much smaller than our typical uncertainty of ∼2.9 kJ·mol^{-1};[141] at higher pressures or for larger solutes than in this set, this term could become significant. Notably, this scheme also omits other contributions that may be relevant in making direct comparison with experimental enthalpies of hydration, including contributions from the cost of polarizing the molecule from vacuum to solvated phase charges (relevant to fixed-charge force fields), corrections to the vibrational modes due to the quantum-chemical nature of real solutes, nonideality of the gas phase, and the fact that the simulation of the liquid is carried out at atmospheric pressure rather than at the vapor pressure of the gas phase; a review of these contributions is provided in ref 143. We note that other groups have also omitted these contributions, which still await a thorough assessment of relative magnitude for small-molecule hydration enthalpies.[141]

Hydration entropies are calculated via the equation:

$$\Delta S = \frac{\Delta H - \Delta G}{T} \tag{10}$$

with ΔG and ΔH calculated as described previously. The calculated hydration enthalpies exhibit some correlation with calculated hydration free energies, but the correlation is not perfect, indicating that enthalpies can indeed provide additional constraints on the force field.[128] The Kendall τ and Pearson R values between the calculated ΔH^{hyd} and ΔG^{hyd} were 0.76 ± 0.02 and 0.943 ± 0.005, respectively (see the SI).

Our conclusion that enthalpies can provide an additional constraint on the force field is further supported by comparison to experimental data. Specifically, 11 experimental hydration enthalpies and entropies from ORCHYD, a database of experimental hydration properties,[144] were added to FreeSolv. The calculated and experimental enthalpies have a Kendall τ value of 0.77 ± 0.05 and a Pearson R value of 0.87 ± 0.03 (see the SI). These values indicate that the computed hydration free energies are relatively predictive of experimental values, though there is also clear room for improvement. The calculated hydration enthalpies and their experimental counterparts show significant differences that are not observed in the plot of experimental versus calculated free energies of the same 11 compounds, suggesting (as in previous studies[141]) that enthalpies provide additional information on the thermodynamics and constraints on the force field (though as noted above, additional enthalpy corrections may be needed[143]). While ΔH^{hyd} and ΔS^{hyd} can act as additional constraints for force field parameters, one of them can always be calculated from the other and the corresponding ΔG^{hyd}, meaning that it is not possible to use all three values as constraints simultaneously. That is, ΔH^{hyd} and ΔS^{hyd} are always highly anticorrelated because of how they are calculated. More details can be found in the SI.

Components of Hydration Enthalpies. We also partitioned the hydration enthalpy, ΔH, into two components: a solvent interaction term, $\Delta H_{\text{int}}^{\text{hyd}}$, and a conformational change term, $\Delta H_{\text{conf}}^{\text{hyd}}$, in order to understand how much the solvation enthalpy is influenced by the solute conformation and how much the solute conformation is modulated by solvation. We obtained the solvent interaction component by taking the average energy of the solute in water and subtracting off the solute internal energy and the energy of a corresponding box of pure water, leaving only the enthalpy change due to changing solute−solvent interactions and solvent reorganization:

$$\Delta H_{\text{int}}^{\text{hyd}} = \langle U_{\text{solv}} \rangle_s - \langle U_{\text{vac}} \rangle_s - \langle U_{\text{water}} \rangle_w \tag{11}$$

where $\langle U_{\text{solv}} \rangle_s$ is the average potential energy over the original solvated trajectory, $\langle U_{\text{vac}} \rangle_s$ is the average potential energy of the solute molecule in the solvated trajectory after removal of its water molecules, and $\langle U_{\text{water}} \rangle_w$ is the average potential of a box of pure water containing the same number of water molecules under the same conditions. $\Delta H_{\text{int}}^{\text{hyd}}$ thus corresponds to the change in solvation enthalpy due to transfer of a solute molecule from vacuum to water with a *fixed set of configurations* (as given by the solvated trajectory), i.e., it treats the solute as if there were no conformational change upon going from gas to water, so it includes only changes in solvent structure and solute−solvent interactions.

The conformational change component of the enthalpy is obtained as the change in solute internal energy upon going from gas to water, which we can evaluate as follows:

$$\Delta H_{\text{conf}}^{\text{hyd}} = \langle U_{\text{vac}} \rangle_s - \langle U_{\text{vac}} \rangle_v \tag{12}$$

where $\langle U_{\text{vac}} \rangle_v$ is the potential energy of the solute molecule in vacuum evaluated from the trajectory run in vacuum and $\langle U_{\text{vac}} \rangle_s$ is the potential energy of the solute molecule in vacuum evaluated from the trajectory run in solvent (after stripping away the solvent molecules). $\Delta H_{\text{conf}}^{\text{hyd}}$ thus gives the enthalpy change due to solute conformational changes upon solvation; these occur because interactions with water can stabilize configurations that are not common in vacuum. If a compound's distribution of configurations is unchanged upon transfer to the solvent, $\Delta H_{\text{conf}}^{\text{hyd}}$ will be zero. It can trivially be verified that these components still sum to the total enthalpy change:

$$\Delta H^{\text{hyd}} = \Delta H_{\text{solv}}^{\text{hyd}} + \Delta H_{\text{conf}}^{\text{hyd}} \tag{13}$$

These components, while certainly not a unique decomposition of the total enthalpy, do provide a way to intuitively understand one important set of contributions to the enthalpy of hydration in a way that provides some insight into changes undergone by the solute and environment. For example, solutes that undergo significant conformational changes upon solvation may tend to have a large change in the conformational component of the hydration enthalpy (Figure 3). This happens because solutes that make hydrogen bonds with water or have strong internal electrostatic interactions in the gas phase can assume conformations when solvated that were energetically unfavorable in vacuo.

■ CONCLUSIONS

Solvation free energies have been the subject of considerable scientific interest for many years because they are related to a
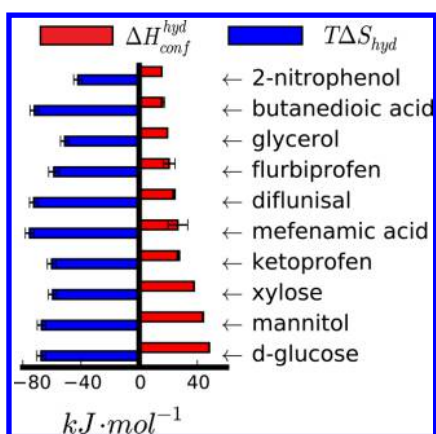
Figure 3. Conformational enthalpies and associated entropies of compounds with highest ΔH_{conf}^{hyd}. Error bars represent the standard error.

large number of physical properties. Here we have provided a short review of alchemical methods for computing solvation free energies of small organic molecules and discussed their application to hydration free energies. Solvation free energies for such molecules can be calculated precisely and effectively using alchemical free energy calculations, as described here. In our experience, the BAR and MBAR free energy estimators require less tuning to work well, while TI requires special care to get the gradients right in rapidly varying regions and introduces unknown integration error. Thus, we recommend MBAR as our preferred general-purpose method, even though TI can in principle also work well. EXP should be avoided in general, partly because of the large bias introduced.

We have also introduced an update to FreeSolv (version 0.5), a database of calculated and experimental hydration free energies, enthalpies, and entropies.[7] The database was designed to be easily incorporated into automated workflows: we provide IUPAC names, PubChem compound IDs, and SMILES strings as well as topology and coordinate files, but additional data are welcome. Additionally, we provide calculated and experimental free energy values that can be used to assist method and force field development. Unfortunately, experimental hydration enthalpies and entropies are not available for every compound.

The calculated free energies show reasonable agreement with experimental values (Figure 2), with an RMS error of around 6 kJ·mol^{-1} and an average error close to 1 kJ·mol^{-1}. With the aid of ORCHYD,[144] we were able to extend FreeSolv to contain experimental hydration enthalpies for a few (11) compounds for the first time. We observe significant errors for hydration enthalpies that are much larger than those for hydration free energies, so further investigation will be needed. This result also suggests that enthalpies can be used as additional constraints in force field development.

Our intention is that FreeSolv serve as an updatable, extensible community resource. While it already covers a large number of molecules, we would be delighted to include input files and calculated values from other force fields and/or methods so it can further serve as a benchmark of methods, simulation packages, and so on. Additionally, while hydration free energy data are not abundant, certainly at least some data are available that are not presently included in FreeSolv, so community contributions of experimental data with references will be appreciated. Additional curation of the experimental data already present is likely needed—for example, much of the

experimental data still need to be tracked back to their original source material rather than literature compilations of data that are currently cited. FreeSolv is available on GitHub at http://github.com/mobleylab/FreeSolv, and contributions are welcomed there.

We believe that this update of FreeSolv will assist future efforts in force field development and the development and testing of new methods. We also hope that FreeSolv's new features will help serve the scientific community and provide a valuable resource the community will help extend.

■ ASSOCIATED CONTENT

⑤ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jced.7b00104.

> Additional discussion; correlation plots between calculated free energies, enthalpies, and entropies; and simulation details (PDF)
>
> GROMACS input files containing all of the details of the simulations (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dmobley@uci.edu. Phone: 949-824-6383.

ORCID ⓘ

David L. Mobley: 0000-0002-1083-5533

Notes

The authors declare the following competing financial interest(s): D.L.M. is a member of the Scientific Advisory Board for OpenEye Scientific Software. J.D.C. is a member of the Scientific Advisory Board for Schrodinger, LLC.

FreeSolv can be obtained free of charge at http://github.com/mobleylab/FreeSolv and, with additional GROMACS energy files analyzed in this review, at www.escholarship.org/uc/item/30c9r5qj.

■ REFERENCES

(1) Guthrie, J. P. A Blind Challenge for Computational Solvation free Energies: Introduction and Overview. *J. Phys. Chem. B* **2009**, *113*, 4501−4507.

(2) Klimovich, P. V.; Mobley, D. L. Predicting Hydration free Energies Using All-atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 307−316.

(3) Mobley, D. L.; Barber, A. E., II; Fennell, C. J.; Dill, K. A. Charge Asymmetries in Hydration of Polar Solutes. *J. Phys. Chem. B* **2008**, *112*, 2405−2414.

(4) Chorny, I.; Dill, K. A.; Jacobson, M. P. Surfaces Affect ion Pairing. *J. Phys. Chem. B* **2005**, *109*, 24056−24060.

(5) Harris, R. C.; Pettitt, B. M. Effects of Geometry and Chemistry on Hydrophobic Solvation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 14681−14686.

(6) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small-molecule Solvation free Energies: an Informal Blind test for Computational Chemistry. *J. Med. Chem.* **2008**, *51*, 769−779.

(7) Mobley, D. L.; Guthrie, J. P. FreeSolv: A Database of Experimental and Calculated Hydration free Energies, with Input Files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711−720.

(8) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 Blind Prediction Challenge: Introduction and Overview. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 259−279.

(9) Geballe, M. T.; Guthrie, J. P. The SAMPL3 Blind Prediction Challenge: Transfer Energy Overview. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 489−496.

(10) Gallicchio, E.; Levy, R. M. Prediction of SAMPL3 Host-guest Affinities with the Binding Energy Distribution Analysis Method (bedam). *J. Comput.-Aided Mol. Des.* **2012**, *26*, 505−516.

(11) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical Prediction of Hydration free Energies for sampl. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 551−562.

(12) Gallicchio, E.; Deng, N.; He, P.; Wickstrom, L.; Perryman, A. L.; Santiago, D. N.; Forli, S.; Olson, A. J.; Levy, R. M. Virtual Screening of Integrase Inhibitors by Large Scale Binding free Energy Calculations: the SAMPL4 Challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 475−490.

(13) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. Blind Prediction of Solvation free Energies from the SAMPL4 Challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 135−150.

(14) Muddana, H. S.; Sapra, N. V.; Fenley, A. T.; Gilson, M. K. The SAMPL4 Hydration Challenge: Evaluation of Partial Charge sets with Explicit-water Molecular Dynamics Simulations. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 277−287.

(15) Michel, J.; Orsi, M.; Essex, J. W. Prediction of Partition Coefficients by Multiscale Hybrid Atomic-Level/Coarse-Grain simulations. *J. Phys. Chem. B* **2008**, *112*, 657−660.

(16) Genheden, S. Predicting Partition Coefficients with a Simple All-Atom/Coarse-Grained Hybrid model. *J. Chem. Theory Comput.* **2016**, *12*, 297−304.

(17) Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. 1-Octanol/Water Partition Coefficients of n-Alkanes from Molecular Simulations of Absolute Solvation Free energies. *J. Chem. Theory Comput.* **2009**, *5*, 2436−2446.

(18) Garrido, N. M.; Economou, I. G.; Queimada, A. J.; Jorge, M.; Macedo, E. A. Prediction of the n-Hexane/Water and 1-Octanol/Water Partition Coefficients for Environmentally Relevant Compounds Using Molecular Simulation. *AIChE J.* **2012**, *58*, 1929−1938.

(19) Yang, L.; Ahmed, A.; Sandler, S. I. Comparison of two Simulation Methods to Compute Solvation free Energies and Partition Coefficients. *J. Comput. Chem.* **2013**, *34*, 284−293.

(20) 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.

(21) Bannan, C. C.; Calabrò, G.; Kyu, D. Y.; Mobley, D. L. Calculating Partition Coefficients of Small Molecules in Octanol/Water and cyclohexane/water. *J. Chem. Theory Comput.* **2016**, *12*, 4015−4024.

(22) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise free Energy Calculations of Amino acid side Chain Analogs:

Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(23) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-charge Force Fields: Small-molecule Hydration free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242−2254.

(24) Liu, S.; Cao, S.; Hoang, K.; Young, K. L.; Paluch, A. S.; Mobley, D. L. Using MD Simulations To Calculate How Solvents Modulate solubility. *J. Chem. Theory Comput.* **2016**, *12*, 1930−1941.

(25) Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. Blind Prediction of Cyclohexane-water Distribution Coefficients from the SAMPL5 Challenge. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 927−944.

(26) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of solvation. *J. Chem. Inf. Model.* **2015**, *55*, 1192−1201.

(27) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the free Enthalpy of Hydration and Solvation: The GROMOS Force-field Parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.

(28) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. A Fixed-charge Model for Alcohol Polarization in the Condensed Phase, and its role in Small Molecule Hydration. *J. Phys. Chem. B* **2014**, *118*, 6438−6446.

(29) Jämbeck, J. P. M.; Lyubartsev, A. P. Update to the General Amber Force Field for Small Solutes with an Emphasis on Free Energies of hydration. *J. Phys. Chem. B* **2014**, *118*, 3793−3804.

(30) Bosisio, S.; Mey, A. S. J. S.; Michel, J. Blinded Predictions of Host−Guest Standard Free Energies of Binding in the SAMPL5 Challenge. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 61−70.

(31) Harwood, D. B.; Peters, C. J.; Siepmann, J. I. A Monte Carlo Simulation Study of the Liquid−liquid Equilibria for Binary Dodecane/ethanol and Ternary Dodecane/ethanol/water Mixtures. *Fluid Phase Equilib.* **2016**, *407*, 269−279.

(32) Klamt, A.; Eckert, F.; Reinisch, J.; Wichmann, K. Prediction of Cyclohexane−Water Distribution Coefficients with COSMO-RS on the SAMPL5 data set. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 959−967.

(33) Rustenburg, A. S.; Dancer, J.; Lin, B.; Feng, J. A.; Ortwine, D. F.; Mobley, D. L.; Chodera, J. D. Measuring Experimental Cyclohexane−Water Distribution Coefficients for the SAMPL5 Challenge. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 945.

(34) Nicholls, A.; Honig, B. A Rapid Finite Difference Algorithm, Utilizing Successive Over-relaxation to Solve the Poisson−Boltzmann Equation. *J. Comput. Chem.* **1991**, *12*, 435−445.

(35) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function for Poisson−Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22*, 608−640.

(36) Knight, J. L.; Brooks, C. L. Surveying Implicit Solvent Models for Estimating Small Molecule Absolute Hydration free Energies. *J. Comput. Chem.* **2011**, *32*, 2909−2923.

(37) Yang, P.-K. Modifying Poisson Equation for Near-solute Dielectric Polarization and Solvation free Energy. *Chem. Phys.* **2016**, *472*, 229−240.

(38) Guthrie, J. P. SAMPL4, a blind challenge for computational solvation free energies: the compounds considered. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 151−168.

(39) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Chapter 4 Alchemical Free Energy Calculations: Ready for Prime time? *Annu. Rep. Comput. Chem.* **2007**, *3*, 41−59.

(40) Shirts, M. R. *Methods Mol. Biol.* **2012**, *819*, 425−467.

(41) Skyner, R. E.; McDonagh, J. L.; Groom, C. R.; van Mourik, T.; Mitchell, J. B. O. A Review of Methods for the Calculation of Solution Free Energies and the Modelling of Systems in Solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174−6191.

(42) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. Predictions of Hydration free Energies from All-atom Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 4533−4537.

(43) Shirts, M. R.; Mobley, D. L.; Brown, S. P. Free-energy Calculations in Structure-based drug Design. *Drug Design* **2010**, *61−86*.

(44) Cramer, C. J.; Truhlar, D. G. A Universal Approach to Solvation modeling. *Acc. Chem. Res.* **2008**, *41*, 760−768.

(45) Panagiotopoulos, A.; Quirke, N.; Stapleton, M.; Tildesley, D. Phase equilibria by simulation in the Gibbs ensemble. *Mol. Phys.* **1988**, *63*, 527−545.

(46) Chen, B.; Siepmann, J. I. Partitioning of Alkane and Alcohol Solutes between Water and (Dry or Wet) 1-Octanol. *J. Am. Chem. Soc.* **2000**, *122*, 6464−6467.

(47) Anderson, K. E.; Siepmann, J. I. Molecular Simulation Approaches to Solubility. In *Development and Applications in Solubility*; Letcher, T. M., Ed.; Royal Society of Chemistry: Cambridge, U.K., 2007; pp 171−187.

(48) Martin, M. G.; Siepmann, J. I. Calculating Gibbs free Energies of Transfer from Gibbs Ensemble Monte Carlo Simulations. *Theor. Chem. Acc.* **1998**, *99*, 347−350.

(49) McGrath, M. J.; Kuo, I.-F. W.; Ngouana W., B. F.; Ghogomu, J. N.; Mundy, C. J.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.; Siepmann, J. I. Calculation of the Gibbs free Energy of Solvation and Dissociation of HCl in Water via Monte Carlo Simulations and Continuum Solvation Models. *Phys. Chem. Chem. Phys.* **2013**, *15*, 13578−13585.

(50) Huang, Y.-L.; Miroshnichenko, S.; Hasse, H.; Vrabec, J. Henry's Law Constant from Molecular Simulation: A Systematic Study of 95 Systems. *Int. J. Thermophys.* **2009**, *30*, 1791.

(51) Zhang, L.; Siepmann, J. I. Direct calculation of Henry's law constants from Gibbs ensemble Monte Carlo simulations: nitrogen, oxygen, carbon dioxide and methane in ethanol. *Theor. Chem. Acc.* **2006**, *115*, 391−397.

(52) Lyubartsev, A. P.; Jacobsson, S. P.; Sundholm, G.; Laaksonen, A. Solubility of Organic Compounds in Water/Octanol Systems. A Expanded Ensemble Molecular Dynamics Simulation Study of log P Parameters. *J. Phys. Chem. B* **2001**, *105*, 7775−7782.

(53) Slusher, J. T. Estimation of Infinite Dilution Activity Coefficients in Aqueous Mixtures via Molecular Simulation. *Fluid Phase Equilib.* **1998**, *153*, 45−61.

(54) da Silva, E. F. Use of free Energy Simulations to Predict Infinite Dilution Activity Coefficients. *Fluid Phase Equilib.* **2004**, *221*, 15−24.

(55) Tong, C.; Clegg, S. L.; Seinfeld, J. H. Comparison of Activity Coefficient Models for Atmospheric Aerosols Containing Mixtures of Electrolytes, Organics, and Water. *Atmos. Environ.* **2008**, *42*, 5459−5482.

(56) Gerber, R. P.; Soares, R. d. P. Prediction of Infinite-Dilution Activity Coefficients Using UNIFAC and COSMO-SAC variants. *Ind. Eng. Chem. Res.* **2010**, *49*, 7488−7496.

(57) Ge, M.-L.; Deng, X.-M.; Zhang, L.-H.; Chen, J.-Y.; Xiong, J.-M.; Li, W.-H. Activity Coefficients at Infinite Dilution of Organic Solutes in the Ionic Liquid 1-butyl-3-methylimidazolium Methyl Sulfate. *J. Chem. Thermodyn.* **2014**, *77*, 7−13.

(58) Ge, M.-L.; Lu, C.-Y.; Liu, X.-Y.; Li, X.-B.; Chen, J.-Y.; Xiong, J.-M. Activity Coefficients at Infinite Dilution of Alkanes, Alkenes, Alkyl Benzenes in Dimethylphosphate Based Ionic Liquids Using Gas−liquid Chromatography. *J. Chem. Thermodyn.* **2015**, *91*, 279−285.

(59) Ge, M.-L.; Zhang, Q.; Li, S.-N.; Li, Y.-J.; Zhang, X.-Z.; Mu, Z. Thermodynamics and Activity Coefficients at Infinite Dilution for Organic Solutes in the Ionic Liquid 1-hexyl-2,3-dimethylimidazolium Bis(trifluoromethylsulfonyl)imide. *J. Chem. Thermodyn.* **2016**, *102*, 303−309.

(60) Krummen, M.; Gruber, D.; Gmehling, J. Measurement of Activity Coefficients at Infinite Dilution in Solvent Mixtures Using the Dilutor technique. *Ind. Eng. Chem. Res.* **2000**, *39*, 2114−2123.

(61) Martínez, R.; Sanz, M. T.; Beltrán, S.; Corcuera, E. Activity Coefficients at Infinite Dilution of Volatile Compounds in Water: Effect of Temperature and Salt concentration. *J. Chem. Eng. Data* **2012**, *57*, 1480−1485.

(62) Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L.; Yang, W.; Ren, P. The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force field. *J. Chem. Theory Comput.* **2012**, *8*, 1721−1736.

(63) Ferrario, M.; Ciccotti, G.; Spohr, E.; Cartailler, T.; Turq, P. Solubility of KF in Water by Molecular Dynamics Using the Kirkwood Integration Method. *J. Chem. Phys.* **2002**, *117*, 4947−4953.

(64) Sanz, E.; Vega, C. Solubility of KF and NaCl in Water by Molecular Simulation. *J. Chem. Phys.* **2007**, *126*, 014507.

(65) Aragones, J. L.; Sanz, E.; Vega, C. Solubility of NaCl in Water by Molecular Simulation Revisited. *J. Chem. Phys.* **2012**, *136*, 244508.

(66) Benavides, A. L.; Aragones, J. L.; Vega, C. Consensus on the Solubility of NaCl in Water from Computer Simulations Using the Chemical Potential Route. *J. Chem. Phys.* **2016**, *144*, 124504.

(67) Westergren, J.; Lindfors, L.; Höglund, T.; Lüder, K.; Nordholm, S.; Kjellander, R. In Silico Prediction of drug Solubility: 1. Free Energy of Hydration. *J. Phys. Chem. B* **2007**, *111*, 1872−1882.

(68) Lüder, K.; Lindfors, L.; Westergren, J.; Nordholm, S.; Kjellander, R. In Silico Prediction of drug Solubility: 2. Free Energy of Solvation in pure Melts. *J. Phys. Chem. B* **2007**, *111*, 1883−1892.

(69) Lüder, K.; Lindfors, L.; Westergren, J.; Nordholm, S.; Kjellander, R. In Silico Prediction of drug Solubility. 3. Free Energy of Solvation in pure Amorphous Matter. *J. Phys. Chem. B* **2007**, *111*, 7303−7311.

(70) Lüder, K.; Lindfors, L.; Westergren, J.; Nordholm, S.; Persson, R.; Pedersen, M. In Silico Prediction of drug Solubility: 4. Will Simple Potentials Suffice? *J. Comput. Chem.* **2009**, *30*, 1859−1871.

(71) Paluch, A. S.; Parameswaran, S.; Liu, S.; Kolavennu, A.; Mobley, D. L. Predicting the Excess Solubility of Acetanilide, Acetaminophen, Phenacetin, Benzocaine, and Caffeine in Binary Water/ethanol Mixtures via Molecular Simulation. *J. Chem. Phys.* **2015**, *142*, 044508.

(72) Ferguson, A. L.; Debenedetti, P. G.; Panagiotopoulos, A. Z. Solubility and Molecular Conformations of n-Alkane Chains in Water. *J. Phys. Chem. B* **2009**, *113*, 6405−6414.

(73) Olsen, R.; Kvamme, B.; Kuznetsova, T. Free Energy of Solvation and Henry's law Solubility Constants for Mono-, di- and Tri-ethylene Glycol in Water and Methane. *Fluid Phase Equilib.* **2016**, *418*, 152−159.

(74) Noroozi, J.; Ghotbi, C.; Sardroodi, J. J.; Karimi-Sabet, J.; Robert, M. A. Solvation free Energy and Solubility of Acetaminophen and Ibuprofen in Supercritical Carbon Dioxide: Impact of the Solvent Model. *J. Supercrit. Fluids* **2016**, *109*, 166−176.

(75) Genheden, S.; Essex, J. W. All-Atom/Coarse-Grained Hybrid Predictions of Distribution Coefficients in sampl5. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 969−976.

(76) Chung, K.-C.; Park, H. Extended Solvent-contact Model Approach to Blind SAMPL5 Prediction Challenge for the Distribution Coefficients of Drug-like Molecules. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1019−1033.

(77) Bodor, N.; Buchwald, P. *Retrometabolic Drug Design and Targeting*; John Wiley & Sons: Hoboken, NJ, 2012; pp 9−38.

(78) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free Energy Methods for drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.

(79) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.

(80) Michel, J.; Essex, J. W. Prediction of Protein-ligand Binding Affinity by free Energy Simulations: Assumptions, Pitfalls and Expectations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639−658.

(81) Zafar, A.; Reynisson, J. Hydration Free Energy as a Molecular Descriptor in Drug Design: A Feasibility study. *Mol. Inf.* **2016**, *35*, 207−214.

(82) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.

(83) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. Comparison of free Energy Methods for Molecular Systems. *J. Chem. Phys.* **2006**, *125*, 184114.

(84) Shirts, M. R.; Pande, V. S. Comparison of Efficiency and bias of free Energies Computed by Exponential Averaging, the Bennett Acceptance Ratio, and Thermodynamic Integration. *J. Chem. Phys.* **2005**, *122*, 144107.

(85) Paliwal, H.; Shirts, M. R. A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy methods. *J. Chem. Theory Comput.* **2011**, *7*, 4115−4134.

(86) Zwanzig, R. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(87) Lu, N.; Singh, J. K.; Kofke, D. A. Appropriate Methods to Combine Forward and Reverse Free-energy Perturbation Averages. *J. Chem. Phys.* **2003**, *118*, 2977−2984.

(88) Wu, D.; Kofke, D. A. Phase-space Overlap Measures. II. Design and Implementation of Staging Methods for Free-energy Calculations. *J. Chem. Phys.* **2005**, *123*, 084109.

(89) Wu, D.; Kofke, D. A. Asymmetric bias in Free-energy Perturbation Measurements Using two Hamiltonian-based Models. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2004**, *70*, 066702.

(90) Wu, D.; Kofke, D. A. Phase-space Overlap Measures. I. Fail-safe bias Detection in free Energies Calculated by Molecular Simulation. *J. Chem. Phys.* **2005**, *123*, 054103.

(91) Bennett, C. H. Efficient Estimation of free Energy Differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.

(92) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.

(93) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.

(94) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(95) de Ruiter, A.; Boresch, S.; Oostenbrink, C. Comparison of thermodynamic integration and Bennett acceptance ratio for calculating relative protein-ligand binding free energies. *J. Comput. Chem.* **2013**, *34*, 1024−1034.

(96) Tembre, B. L.; McCammon, J. Ligand-receptor interactions. *Comput. Chem.* **1984**, *8*, 281−283.

(97) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of free Energy Calculations. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 397−411.

(98) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted Scaling, a new Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*, 9025−9031.

(99) Wang, J.; Hou, T.; Xu, X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum models. *Curr. Comput.-Aided Drug Des.* **2006**, *2*, 287−306.

(100) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear Scaling Schemes for Lennard-Jones Interactions in free Energy Calculations. *J. Chem. Phys.* **2007**, *127*, 214108.

(101) Naden, L. N.; Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical free Energy Calculations. 2. Inserting and Deleting Particles with Coulombic Interactions. *J. Chem. Theory Comput.* **2015**, *11*, 2536−2549.

(102) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding Singularities and Numerical Instabilities in free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.

(103) 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.

(104) Pham, T. T.; Shirts, M. R. Identifying low Variance Pathways for free Energy Calculations of Molecular Transformations in Solution Phase. *J. Chem. Phys.* **2011**, *135*, 034114.

(105) Shirts, M. R.; Pande, V. S. Solvation free Energies of Amino acid side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.

(106) Gapsys, V.; Seeliger, D.; de Groot, B. L. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2373−2382.

(107) Pham, T. T.; Shirts, M. R. Optimal Pairwise and Non-pairwise Alchemical Pathways for free Energy Calculations of Molecular Transformation in Solution Phase. *J. Chem. Phys.* **2012**, *136*, 124120.

(108) Naden, L. N.; Pham, T. T.; Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical free Energy Calculations. 1. Removal of Uncharged Atomic Sites. *J. Chem. Theory Comput.* **2014**, *10*, 1128−1149.

(109) Christ, C. D.; Fox, T. Accuracy Assessment and Automation of free Energy Calculations for drug Design. *J. Chem. Inf. Model.* **2014**, *54*, 108−120.

(110) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632−2647.

(111) Mobley, D. L. Let's get Honest About Sampling. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 93−95.

(112) Wang, B.; Wei, G. W. Parameter Optimization in Differential Geometry Based Solvation Models. *J. Chem. Phys.* **2015**, *143*, 134119.

(113) Nessler, I. J.; Litman, J. M.; Schnieders, M. J. Toward Polarizable AMOEBA Thermodynamics at Fixed Charge Efficiency Using a dual Force Field Approach: Application to Organic Crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 30313−30322.

(114) Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 221−233.

(115) Canzar, S.; El-Kebir, M.; Pool, R.; Elbassioni, K.; Malde, A. K.; Mark, A. E.; Geerke, D. P.; Stougie, L.; Klau, G. W. Charge Group Partitioning in Biomolecular Simulation. *J. Comput. Biol.* **2013**, *20*, 188−198.

(116) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026−4037.

(117) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.

(118) Mobley, D. L.; Dill, K. A.; Chodera, J. D. Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small molecules. *J. Phys. Chem. B* **2008**, *112*, 938−946.

(119) Parameswaran, S.; Mobley, D. L. Box size Effects are Negligible for Solvation free Energies of Neutral Solutes. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 825−829.

(120) Berendsen, H. J. C.; Van der Spoel, D.; van Drunen, R. GROMACS: A Message-passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(121) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306−317.

(122) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(123) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(124) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a High-throughput and Highly Parallel open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(125) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(126) 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.

(127) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics study. *Phys. Rev. Lett.* **1980**, *45*, 1196−1199.

(128) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational Calorimetry: High-Precision Calculation of Host−Guest Binding thermodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 4377−4394.

(129) *OEChem Toolkit*; OpenEye Scientific Software: Santa Fe, NM, 2010; http://www.eyesopen.com/.

(130) *QUACPAC 1.7.0.2*; OpenEye Scientific Software: Santa Fe, NM; http://www.eyesopen.com/.

(131) Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the data set and the Analysis of Failures. - PubMed - ncbi. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.

(132) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: I. method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(133) 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.

(134) 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.

(135) Gallicchio, E.; Levy, R. M. AGBNP: an Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-resolution Modeling. *J. Comput. Chem.* **2004**, *25*, 479−499.

(136) Gallicchio, E.; Paris, K.; Levy, R. M. The AGBNP2 Implicit Solvation model. *J. Chem. Theory Comput.* **2009**, *5*, 2544−2564.

(137) Boresch, S.; Archontis, G.; Karplus, M. Free energy simulations: The meaning of the individual contributions from a component analysis. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 25−33.

(138) Zacharias, M.; Straatsma, T. P. Path Dependence of Free Energy Components in Thermodynamic Integration. *Mol. Simul.* **1995**, *14*, 417−423.

(139) Dill, K. A. Additivity Principles in Biochemistry. *J. Biol. Chem.* **1997**, *272*, 701−704.

(140) Levy, R. M.; Gallicchio, E. COMPUTER SIMULATIONS WITH EXPLICIT SOLVENT: Recent Progress in the Thermodynamic Decomposition of Free Energies and in Modeling Electrostatic Effects. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531−567.

(141) Hess, B.; van der Vegt, N. F. A. Hydration Thermodynamic Properties of Amino Acid Analogues: A Systematic Comparison of Biomolecular Force Fields and Water models. *J. Phys. Chem. B* **2006**, *110*, 17616−17626.

(142) Fenley, A. T.; Muddana, H. S.; Gilson, M. K. Entropy-enthalpy Transduction Caused by Conformational Shifts can Obscure the Forces Driving Protein-ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20006−20011.

(143) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.

(144) Plyasunova, N. V.; Plyasunov, A. V.; Shock, E. L. Database of Thermodynamic Properties for Aqueous Organic compounds. *Int. J. Thermophys.* **2004**, *25*, 351−360.