



Published in final edited form as:

J Chem Eng Data. 2017 May 11; 62(5): 1559–1569. doi:10.1021/acs.jced.7b00104.

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, Department of Pharmaceutical Sciences, University of California, Irvine, Scientific Computing Department, STFC, UK, Computational and Systems Biology Program, Sloan Kettering Institute, Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, and Departments of Pharmaceutical Sciences and Chemistry, University of California, Irvine

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 by applying them to calculate aqueous solvation free energies (hydration free energies). These approaches use a non-physical 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.

Correspondence to: David L. Mobley.

[†]University of California, Irvine

[‡]University of California, Irvine

[¶]STFC Daresbury, Warrington, WA4 4AD, United Kingdom

[§]Memorial Sloan Kettering Cancer Center, New York

^{||}University of Colorado, Boulder

[⊥]University of California, Irvine

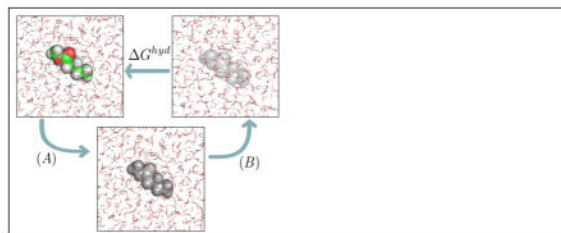
Supporting Information Available

The following files are available free of charge.

- SupportingInformation.pdf: document containing correlation plots between calculated free energies, enthalpies and entropies.
- GROMACS 4.6.7.mdp files: GROMACS input files containing all the details of the simulations.
- FreeSolv can be obtained free of charge at <http://github.com/mobleylab/FreeSolv> and, with additional GROMACS energy files analyzed in this manuscript, at www.escholarship.org/uc/item/30c9r5qj.

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

Graphical abstract



Introduction

Solvation free energies give 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 (G^{solv}) in general, and hydration free energies (G^{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 distribution of chemical species between immiscible solvents or different phases.

Solvation free energies are differences in thermodynamic potentials which describe the relative populations of a chemical species in solution and gas phase at equilibrium.^{1,2} In the thermodynamic limit in the solvated phase and the ideal gas limit in the gas phase, G^{solv} of component i is equal to $\mu_{i,solv} - \mu_{i,gas}$, the difference in 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 they also can provide insight into how solvent behaves in different environments. For example, water solvates molecules of opposite polarity differently, due to its inherent asymmetry,³ surfaces also have asymmetric effects on ion pairing which depend on the curvature of the surface,⁴ and molecular geometry and chemical environment affects hydrophobic solvation.⁵ Although they can be difficult to measure experimentally, G^{solv} and G^{hyd} can be calculated to a precision better than $0.4 \text{ kJ}\cdot\text{mol}^{-1}$, even with a relatively modest investment of simulation time, for relatively diverse small neutral molecules⁶ such as those seen in the FreeSolv database of hydration free energies⁷ 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 many kJ/mol, comparison of computed solvation free energies to 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 by 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^a. 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 solvents²⁶ – 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 parameterization^{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 non-physical intermediates to compute the free energy of transferring a solute from solution to 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 non-physical way. Since free energy is path-independent, this non-physical 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 parameterizing 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 another turning off the electrostatic interactions using a second 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 G^{hyd} but do not take into consideration 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 G^{solv} :

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

^aBut see the Supporting Information for how protonation state/tautomer challenges may apply here, as in protein-ligand binding.

where μ_i^{excess} is the excess chemical potential of i and is equal to ΔG_i^{solw} in the ideal gas limit of the vapor phase, R is the universal gas constant and T , 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 G^{solw} to further test models and methodologies that use these free energy calculations.

Solubility prediction is another field where $G^{sol/hyd}$ 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 drug-like compounds have crystal polymorphs.^{67–70} Relative solubilities of a given chemical species between different solvents can also be assessed with these calculations.^{24,71} Henry's law solubility constants^{72,73} and solubilities in supercritical fluids⁷⁴ 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 solvation free energies of the neutral form of the chemical species in two different solvents,²¹ as shown in equation 2:

$$\log_{10} P_{A \rightarrow B} = \frac{\Delta G^{solw,A} - \Delta G^{solw,B}}{RT \ln(10)} \quad (2)$$

where $G^{solw,A}$ and $G^{solw,B}$ 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 $\mathcal{H}_B(\mathbf{q}, \mathbf{p}; \lambda)$.

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

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 $\lambda = 1$; \mathbf{q} and \mathbf{p} represent all 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 :

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

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:

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

where $\beta = (k_B T)^{-1}$. Although equation 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 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 extrapolation 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, the 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):

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

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

This expression minimizes the free energy variance⁹¹ and makes BAR much more efficient than EXP.^{89,90} The Multistate Bennett acceptance ratio (MBAR) is an extension of BAR that considers the overlap between a given state and all the others in the path between the end states.⁹² BAR and MBAR perform similarly when the spacing between intermediate states is moderate⁸⁵ and therefore only neighboring states have phase-space overlap. 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 due to 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 modulating 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 non-bonded 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 non-polar interactions. Here, three main thermodynamic states are considered: a single, non-interacting molecule of the solute in a box of solvent; the 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 which changes electrostatic interactions only, and another phase which 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

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

but this pathway is in general problematic for solvation of all but the smallest molecules. This is because repulsive forces are often handled by a $1/r^{12}$ term (such as in the Lennard-Jones functional form) which leads to non-integrable singularities in $\langle H/\lambda \rangle$ at terminal λ values due to sudden changes in the potential at small r . This is not a problem which 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 Lennard-Jones potential between two particles i and j is:

$$U_{ij}^{LJ}(r_{ij}, \lambda) = \lambda^n 4\epsilon_{ij} \left(\frac{1}{[\alpha(1-\lambda)^m + (r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\alpha(1-\lambda)^m + (r_{ij}/\sigma_{ij})^6} \right) \quad (8)$$

where ϵ_{ij} and σ_{ij} are the Lennard-Jones well-depth and lengthscale parameters, respectively, and α is a positive constant which 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 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 near 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? Are there critical chemical effects 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, when designing 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 grows and as solvent dynamics or environment become 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

We updated FreeSolv, the free community solvation free energy database

About FreeSolv

FreeSolv⁷ is a hydration free energy database for neutral^b compounds 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 of 643 small organic molecules. The molecular weights for compounds in FreeSolv range from 16.06 Daltons (methane) to 498.88 Daltons (decachloro-biphenyl). The number of rotatable bonds runs as high as 12, but most compounds are largely rigid. Since experimental and calculated hydration free energies, G^{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 from Mobley and collaborators,¹¹⁷ which are included in FreeSolv.⁷

While calculated hydration free energies for all 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 in SI). The first five states corresponded to changes in electrostatic interactions, while the last 15 modified 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 non-polar interactions.¹⁰¹ Box size does not affect the result of solvation free energy calculations as long as good practices, which recommend that box edges be at least twice the Lennard-Jones cutoff distance, are followed.¹¹⁹ We ran 5 nanoseconds of Langevin dynamics per state with 2 femtosecond time steps in GROMACS 4.6.7^{120–125} at 298.15K. Van der Waals interactions

^bFor additional discussion of why we focus on neutral compounds, see the Supporting Information Section I.

were neglected beyond a smoothly switched cutoff of 1.2 nm. Different cut-off radii are commonly used, but one should be aware the choice of cut-off can affect calculated solvation free energies. However, long range dispersion corrections can be employed (as here) to remove the cutoff-dependence of calculated free energies.¹²⁶ (However, 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 (m and n equal to 1 in equation 8) which leads to statistical uncertainties approximately of the same size as uncertainties from simulations using optimized path soft-cores.¹⁰⁴ Pressure was maintained at 1 atm by the Parrinello-Rahman barostat.¹²⁷ Enthalpy and entropy decomposition required 60 nanosecond Langevin dynamics simulations, with two femtosecond timesteps 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 error bars on the computed enthalpies to levels around $2.9 \text{ kJ} \cdot \text{mol}^{-1}$, roughly the level of typical thermal energy ($1 k_B T$) as done in, for example, host-guest binding calculations.¹²⁸ We used the default Langevin dynamics' friction coefficients implemented in GROMACS (see SI). The size of friction coefficient only affects equilibration and correlation times, but should not affect the calculated hydration free energies and enthalpies. In order to obtain consistent results, we used simulation boxes with 1, 309 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 supporting materials.

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} then charged mol2 files were written out. The AMBER Antechamber package (as distributed with AMBER14) was then used to assign parameters from the GAFF²⁰ 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 which performs the setup and re-generates all 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 v0.51, which extends v0.5 by making the small molecule parameter sets available additionally in CHARMM, DESMOND, and LAMMPS formats. Additional details can be found in the supporting material and in the FreeSolv README files.

FreeSolv hydration free energy results

Computed hydration free energies are compared with experiment in figure 2.

In the calculations described in this study, we found an average error of $1.3 \pm 0.3 \text{ kJ} \cdot \text{mol}^{-1}$, RMS error of $6.4 \pm 0.3 \text{ kJ} \cdot \text{mol}^{-1}$, average absolute error of $4.7 \pm 0.2 \text{ kJ} \cdot \text{mol}^{-1}$, Kendall τ of 0.80 ± 0.01 , and Pearson R of 0.933 ± 0.008 , comparable to those in the original FreeSolv set,⁷ though some individual compounds have reasonably significant discrepancies (see SI). This level of accuracy is consistent with what is often seen from classical fixed-charge force fields, which typically yield RMS errors around 4–8 kJ/mol in computed hydration free

energies.⁴³ We have previously used this data to address force field issues on hydroxyl groups,²⁸ and also to highlight functional groups in the set which 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 (G^E) and the free energy of decoupling the nonpolar interactions in water (G^{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 due to 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 (H^{hyd}) and entropies of hydration (S^{hyd}), and have added these to the database. Enthalpies of transfer, due to 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 to 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 which have 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:

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

Here, $\langle U_{solution} \rangle$ is the internal energy of the solution (containing the solute); $\langle U_{water} \rangle$ is the internal energy of a box of the same number of water molecules (under the same conditions) without the solute; and $\langle U_{vacuo} \rangle$ is the internal energy of the solute molecule alone in vacuum. We have neglected the pressure-volume contribution to the enthalpies, $P \cdot V$, since for solutes of this size, the contribution is much smaller than our typical uncertainties of $\approx 2.9 \text{ kJ} \cdot \text{mol}^{-1}$;¹⁴¹ at larger 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; for a review of these contributions, see.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:

$$\Delta S = \frac{\Delta H - \Delta G}{T} \quad (10)$$

with G and H calculated as described previously. 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 the Pearson R between the calculated H^{hyd} and G^{hyd} respectively were 0.76 ± 0.02 and 0.943 ± 0.005 (see supplementary information).

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. Calculated and experimental enthalpies have a Kendall τ of 0.77 ± 0.05 , and a Pearson R of 0.87 ± 0.03 (see SI). These values indicate that the computed hydration free energies are relatively predictive of experimental values, though there is also clear room for improvement. 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 studies¹⁴¹) that enthalpies provide additional information on the thermodynamics and constraints on the force field (though as noted above, additional enthalpy corrections may be needed¹⁴³). While H^{hyd} and S^{hyd} can act as additional constraints for force field parameters, one of them can always be calculated from the other and the corresponding G^{hyd} , meaning that it is not worthwhile to use all three values as constraints simultaneously. That is, H^{hyd} and S^{hyd} are always highly anti-correlated because of how they are calculated. More details can be found in the supporting information.

Components of hydration enthalpies

We also partitioned the hydration enthalpy, H , into two components: a solvent interaction term and a conformational change term, ΔH_{int}^{hyd} and ΔH_{conf}^{hyd} , respectively, in order to understand how much the solvation enthalpy is influenced by the solute conformation, and how much 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:

$$\Delta H_{int}^{hyd} = \langle U_{solv} \rangle_s - \langle U_{vac} \rangle_s - \langle U_{water} \rangle_w \quad (11)$$

where $\langle U_{solv} \rangle_s$ is the average potential energy over the original solvated trajectory, $\langle U_{vac} \rangle_s$ is the average potential energy of the solute molecule in the solvated trajectory after removing its water molecules, and $\langle U_{water} \rangle_w$ is the average potential of a box of pure water containing the same number of water molecules under the same conditions. ΔH_{int}^{hyd} thus corresponds to the change in solvation enthalpy due to transferring 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 is no conformational change 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 by taking the change in solute internal energy on going from gas to water, which we can evaluate as follows:

$$\Delta H_{conf}^{hyd} = \langle U_{vac} \rangle_s - \langle U_{vac} \rangle_v \quad (12)$$

where $\langle U_{vac} \rangle_v$ is the potential energy of the solute molecule in vacuum evaluated from the trajectory run in vacuum, and $\langle U_{vac} \rangle_s$ is the potential energy of the solute molecule in vacuum evaluated from the trajectory run in solvent (after stripping the solvent molecules).

ΔH_{conf}^{hyd} thus gives the enthalpy change due to solute conformational changes on 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 on transfer to solvent, ΔH_{conf}^{hyd} will be zero. It can trivially be verified that these components still sum to the total enthalpy change:

$$\Delta H^{hyd} = \Delta H_{solv}^{hyd} + \Delta H_{conf}^{hyd} \quad (13)$$

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 which provides some insight into changes undergone by the solute and environment. For example, solutes which undergo significant conformational changes on solvation may tend to have a large change in the conformational component of the hydration enthalpy (fig. 3). This happens because solutes that make hydrogen bonds with water or have strong internal electrostatic interactions in the gas phase can assume conformations that were energetically unfavorable *in vacuo* when solvated.

Conclusions

Solvation free energies have been the subject of considerable scientific interest for many years because they are related to a 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, BAR and MBAR 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, in partly because of the large bias introduced.

We also introduced an update to FreeSolv⁷ (v0.5), a database of calculated and experimental hydration free energies, enthalpies and entropies. 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 is 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.

Calculated free energies show reasonable agreement with experimental values (fig. 2) with an RMS error around $6 \text{ kJ} \cdot \text{mol}^{-1}$ and an average error close to $1 \text{ kJ} \cdot \text{mol}^{-1}$. 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 updateable, 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 is not abundant, certainly at least some data is available that is 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 needs to be tracked back to its original source material rather than literature compilations of data which 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 development and testing of new methods. We also hope that FreeSolv's new features help serve the scientific community, and provide a valuable resource the community will help extend.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

DLM and GDRM appreciate the financial support from the National Science Foundation (CHE 1352608), and computing support from the UCI GreenPlanet cluster, supported in part by NSF Grant CHE-0840513. GDRM appreciates support from the Brazilian agency CAPES - Science without Borders program (BEX 3932-13-3). JDC acknowledges partial support from NIH grant P30 CA008748. HHL is supported through an EPSRC provided SLA, funding the core support of CCPBioSim. CCPBioSim is the Collaborative Computational Project for Biomolecular Simulation funded by EPSRC grants EP/J010588/1 and EP/M022609/1. We particularly appreciate Kyle Beauchamp (Counsyl, South San Francisco, CA) and Lee-Ping Wang (UC Davis) for input on curation of the FreeSolv database. We also thank Gaetano Calabrò and Caitlin Bannan for their support and assistance, and sharing of knowledge.

DLM is a member of the Scientific Advisory Board for OpenEye Scientific Software. JDC is a member of the Scientific Advisory Board for Schrödinger, LLC.

References

1. Guthrie JP. A Blind Challenge for Computational Solvation free Energies: Introduction and Overview. *J Phys Chem B*. 2009; 113:4501–4507. [PubMed: 19338360]
2. Klimovich PV, Mobley DL. Predicting Hydration free Energies Using All-atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J Comput-Aided Mol Des*. 2010; 24:307–316. [PubMed: 20372973]
3. Mobley DL, Baker JR, Barber AE, Fennell CJ, Dill KA. Charge Asymmetries in Hydration of Polar Solutes. *J Phys Chem B*. 2008; 112:2405–2414. [PubMed: 18251538]
4. Chorny I, Dill KA, Jacobson MP. Surfaces Affect ion Pairing. *J Phys Chem B*. 2005; 109:24056–24060. [PubMed: 16375397]
5. Harris RC, Pettitt BM. Effects of Geometry and Chemistry on Hydrophobic Solvation. *Proc Natl Acad Sci U S A*. 2014; 111:14681–14686. [PubMed: 25258413]
6. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS. Predicting Small-molecule Solvation free Energies: an Informal Blind test for Computational Chemistry. *J Med Chem*. 2008; 51:769–779. [PubMed: 18215013]
7. Mobley DL, Guthrie JP. FreeSolv: A Database of Experimental and Calculated Hydration free Energies, with Input Files. *J Comput-Aided Mol Des*. 2014; 28:711–720. [PubMed: 24928188]
8. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ. The SAMPL2 Blind Prediction Challenge: Introduction and Overview. *J Comput-Aided Mol Des*. 2010; 24:259–279. [PubMed: 20455007]
9. Geballe MT, Guthrie JP. The SAMPL3 Blind Prediction Challenge: Transfer Energy Overview. *J Comput-Aided Mol Des*. 2012; 26:489–496. [PubMed: 22476552]
10. Gallicchio E, Levy RM. Prediction of SAMPL3 Host-guest Affinities with the Binding Energy Distribution Analysis Method (bedam). *J Comput-Aided Mol Des*. 2012; 26:505–516. [PubMed: 22354755]
11. Mobley DL, Liu S, Cerutti DS, Swope WC, Rice JE. Alchemical Prediction of Hydration free Energies for sampl. *J Comput-Aided Mol Des*. 2012; 26:551–562. [PubMed: 22198475]
12. Gallicchio E, Deng N, He P, Wickstrom L, Perryman AL, Santiago DN, Forli S, Olson AJ, Levy RM. Virtual Screening of Integrase Inhibitors by Large Scale Binding free Energy Calculations: the SAMPL4 Challenge. *J Comput-Aided Mol Des*. 2014; 28:475–490. [PubMed: 24504704]
13. Mobley DL, Wymer KL, Lim NM, Guthrie JP. Blind Prediction of Solvation free Energies from the SAMPL4 Challenge. *J Comput-Aided Mol Des*. 2014; 28:135–150. [PubMed: 24615156]
14. Muddana HS, Sapra NV, Fenley AT, Gilson MK. The SAMPL4 Hydration Challenge: Evaluation of Partial Charge sets with Explicit-water Molecular Dynamics Simulations. *J Comput-Aided Mol Des*. 2014; 28:277–287. [PubMed: 24477800]

15. Michel J, Orsi M, Essex JW. Prediction of Partition Coefficients by Multiscale Hybrid Atomic-Level/Coarse-Grain simulations. *J Phys Chem B*. 2008; 112:657–660. [PubMed: 18163606]
16. Genheden S. Predicting Partition Coefficients with a Simple All-Atom/Coarse-Grained Hybrid model. *J Chem Theory Comput*. 2016; 12:297–304. [PubMed: 26592737]
17. Garrido NM, Queimada AJ, Jorge M, Macedo EA, Economou IG. 1-Octanol/Water Partition Coefficients of n-Alkanes from Molecular Simulations of Absolute Solvation Free energies. *J Chem Theory Comput*. 2009; 5:2436–2446. [PubMed: 26616624]
18. Garrido NM, Economou IG, Queimada AJ, Jorge M, Macedo EA. 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 SI. Comparison of two Simulation Methods to Compute Solvation free Energies and Partition Coefficients. *J Comp Chem*. 2013; 34:284–293. [PubMed: 23109246]
20. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and Testing of a General Amber Force Field. *J Comp Chem*. 2004; 25:1157–1174. [PubMed: 15116359]
21. Bannan CC, Calabrò G, Kyu DY, Mobley DL. Calculating Partition Coefficients of Small Molecules in Octanol/Water and cyclohexane/water. *J Chem Theory Comput*. 2016; 12:4015–4024. [PubMed: 27434695]
22. Shirts MR, Pitera JW, Swope WC, Pande VS. 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 DL, Dumont E, Chodera JD, Dill KA. 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. [PubMed: 17291029]
24. Liu S, Cao S, Hoang K, Young KL, Paluch AS, Mobley DL. Using MD Simulations To Calculate How Solvents Modulate solubility. *J Chem Theory Comput*. 2016; 12:1930–1941. [PubMed: 26878198]
25. Bannan CC, Burley KH, Chiu M, Shirts MR, Gilson MK, Mobley DL. Blind Prediction of Cyclohexane-water Distribution Coefficients from the SAMPL5 Challenge. *J Comput-Aided Mol Des*. 2016; 30:927–944. [PubMed: 27677750]
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. [PubMed: 26010106]
27. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A Biomolecular Force Field Based on the free Enthalpy of Hydration and Solvation: The GROMOS Force-field Parameter sets 53A5 and 53A6. *J Comp Chem*. 2004; 25:1656–1676. [PubMed: 15264259]
28. Fennell CJ, Wymer KL, Mobley DL. 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. [PubMed: 24702668]
29. Jämbeck JPM, Lyubartsev AP. 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. [PubMed: 24684585]
30. Bosisio S, Mey ASJS, Michel J. Blinded Predictions of Host-guest Standard free Energies of Binding in the SAMPL5 Challenge. *J Comput-Aided Mol Des*. 2016:1–10.
31. Harwood DB, Peters CJ, Siepmann JI. 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*. 2016:1–9.
33. Rustenburg AS, Dancer J, Lin B, Feng JA, Ortwine DF, Mobley DL, Chodera JD. Measuring Experimental Cyclohexane-water Distribution Coefficients for the SAMPL5 Challenge. *bioRxiv*. 2016:063081.
34. Nicholls A, Honig B. A Rapid Finite Difference Algorithm, Utilizing Successive Over-relaxation to Solve the Poisson–Boltzmann Equation. *J Comp Chem*. 1991; 12:435–445.
35. Grant JA, Pickup BT, Nicholls A. A Smooth Permittivity Function for Poisson–Boltzmann Solvation Methods. *J Comp Chem*. 2001; 22:608–640.

36. Knight JL, Brooks CL. Surveying Implicit Solvent Models for Estimating Small Molecule Absolute Hydration free Energies. *J Comp Chem*. 2011; 32:2909–2923. [PubMed: 21735452]
37. Yang PK. Modifying Poisson Equation for Near-solute Dielectric Polarization and Solvation free Energy. *Chemical Physics*. 2016; 472:229–240.
38. Guthrie JP. SAMPL4, a blind challenge for computational solvation free energies: the compounds considered. *J Comput-Aided Mol Des*. 2014; 28:151–168. [PubMed: 24706106]
39. Shirts MR, Mobley DL, Chodera JD. Chapter 4 Alchemical Free Energy Calculations: Ready for Prime time? *Annu Rep Comput Chem*. 2007; 3:41–59.
40. Shirts MR. *Methods Mol Biol*. 2012; 819:425–467. [PubMed: 22183551]
41. Skyner RE, McDonagh JL, Groom CR, van Mourik T, Mitchell JBO. 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. [PubMed: 25660403]
42. Mobley DL, Bayly CI, Cooper MD, Dill KA. Predictions of Hydration free Energies from All-atom Molecular Dynamics Simulations. *J Phys Chem B*. 2009; 113:4533–4537. [PubMed: 19271713]
43. Shirts MR, Mobley DL, Brown SP. Free-energy Calculations in Structure-based drug Design. *Drug Design*. 2010:61–86.
44. Cramer CJ, Truhlar DG. A Universal Approach to Solvation modeling. *Acc Chem Res*. 2008; 41:760–768. [PubMed: 18512970]
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 JI. Partitioning of Alkane and Alcohol Solutes between Water and (Dry or Wet) 1-Octanol. *J Am Chem Soc*. 2000; 122:6464–6467.
47. Anderson, KE., Siepmann, JI. *Dev Appl Solubility*. IUPAC; Oxford: 2007. p. 171-187.
48. Martin MG, Siepmann JI. Calculating Gibbs free Energies of Transfer from Gibbs Ensemble Monte Carlo Simulations. *Theor Chem Acc*. 1998; 99:347–350.
49. McGrath MJ, Kuo IFW, Ngouana WBF, Ghogomu JN, Mundy CJ, Marenich AV, Cramer CJ, Truhlar DG, Siepmann JI. 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. [PubMed: 23831584]
50. Huang YL, 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 JI. 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 AP, Jacobsson SP, 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 JT. Estimation of Infinite Dilution Activity Coefficients in Aqueous Mixtures via Molecular Simulation. *Fluid Phase Equilib*. 1998; 153:45–61.
54. da Silva EF. Use of free Energy Simulations to Predict Infinite Dilution Activity Coefficients. *Fluid Phase Equilib*. 2004; 221:15–24.
55. Tong C, Clegg SL, Seinfeld JH. Comparison of Activity Coefficient Models for Atmospheric Aerosols Containing Mixtures of Electrolytes, Organics, and Water. *Atmos Environ*. 2008; 42:5459–5482.
56. Gerber RP, Soares RdP. Prediction of Infinite-Dilution Activity Coefficients Using UNIFAC and COSMO-SAC variants. *Ind Eng Chem Res*. 2010; 49:7488–7496.
57. Ge ML, Deng XM, Zhang LH, Chen JY, Xiong JM, Li WH. Activity Coefficients at Infinite Dilution of Organic Solutes in the Ionic Liquid 1-butyl-3-methylimidazolium Methyl Sulfate. *J Chem Thermodynamics*. 2014; 77:7–13.
58. Ge ML, Lu CY, Liu XY, Li XB, Chen JY, Xiong JM. Activity Coefficients at Infinite Dilution of Alkanes, Alkenes, Alkyl Benzenes in Dimethylphosphate Based Ionic Liquids Using Gas-liquid Chromatography. *J Chem Thermodynamics*. 2015; 91:279–285.

59. Ge ML, Zhang Q, Li SN, Li YJ, Zhang XZ, 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 Thermodynamics*. 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 MT, 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 MJ, 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. [PubMed: 22582032]
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. [PubMed: 17212500]
65. Aragonés JL, Sanz E, Vega C. Solubility of NaCl in Water by Molecular Simulation Revisited. *J Chem Phys*. 2012; 136:244508. [PubMed: 22755588]
66. Benavides AL, Aragonés JL, Vega C. Consensus on the Solubility of NaCl in Water from Computer Simulations Using the Chemical Potential Route. *J Chem Phys*. 2016; 144:124504. [PubMed: 27036458]
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. [PubMed: 17266351]
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. [PubMed: 17266352]
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. [PubMed: 17550285]
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 Comp Chem*. 2009; 30:1859–1871. [PubMed: 19115279]
71. Paluch AS, Parameswaran S, Liu S, Kolavennu A, Mobley DL. 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. [PubMed: 25637996]
72. Ferguson AL, Debenedetti PG, Panagiotopoulos AZ. Solubility and Molecular Conformations of n-Alkane Chains in Water. *J Phys Chem B*. 2009; 113:6405–6414. [PubMed: 19361179]
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 JJ, Karimi-Sabet J, Robert MA. Solvation free Energy and Solubility of Acetaminophen and Ibuprofen in Supercritical Carbon Dioxide: Impact of the Solvent Model. *J of Supercritical Fluids*. 2016; 109:166–176.
75. Genheden S, Essex JW. All-atom/coarse-grained Hybrid Predictions of Distribution Coefficients in sampl5. *J Comput-Aided Mol Des*. 2016; 12:297–304.
76. Chung KC, 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. [PubMed: 27448686]
77. Bodor, N., Buchwald, P. *Retrometabolic Drug Design and Targeting*. John Wiley & Sons, Inc; 2012. p. 9-38.
78. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free Energy Methods for drug Discovery: Progress and Challenges. *Curr Opin Struct Biol*. 2011; 21:150–160. [PubMed: 21349700]

79. Deng Y, Roux B. Computations of Standard Binding Free Energies with Molecular Dynamics simulations. *J Phys Chem B*. 2009; 113:2234–2246. [PubMed: 19146384]
80. Michel J, Essex JW. Prediction of Protein-ligand Binding Affinity by free Energy Simulations: Assumptions, Pitfalls and Expectations. *J Comput-Aided Mol Des*. 2010; 24:639–658. [PubMed: 20509041]
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 JG. Statistical Mechanics of Fluid Mixtures. *J Chem Phys*. 1935; 3:300–313.
83. Ytreberg FM, Swendsen RH, Zuckerman DM. Comparison of free Energy Methods for Molecular Systems. *J Chem Phys*. 2006; 125:184114. [PubMed: 17115745]
84. Shirts MR, Pande VS. 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. [PubMed: 15847516]
85. Paliwal H, Shirts MR. 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. [PubMed: 26598357]
86. Zwanzig R. High-Temperature Equation of State by a Perturbation Method. I. Non-polar gases. *J Chem Phys*. 1954; 22:1420–1426.
87. Lu N, Singh JK, Kofke DA. Appropriate Methods to Combine Forward and Reverse Free-energy Perturbation Averages. *J Chem Phys*. 2003; 118:2977–2984.
88. Wu D, Kofke DA. Phase-space Overlap Measures. II. Design and Implementation of Staging Methods for Free-energy Calculations. *J Chem Phys*. 2005; 123:084109–1–084109–10. [PubMed: 16164284]
89. Wu D, Kofke DA. 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 DA. Phase-space Overlap Measures. I. Fail-safe bias Detection in free Energies Calculated by Molecular Simulation. *J Chem Phys*. 2005; 123:054103. [PubMed: 16108627]
91. Bennett CH. Efficient Estimation of free Energy Differences from Monte Carlo data. *J Comp Phys*. 1976; 22:245–268.
92. Shirts MR, Chodera JD. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J Chem Phys*. 2008; 129:124105. [PubMed: 19045004]
93. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. *Phys Rev Lett*. 1989; 63:1195–1198. [PubMed: 10040500]
94. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comp 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 Comp Chem*. 2013; 34:1024–1034. [PubMed: 23335287]
96. Tembre BL, McCammon J. Ligand-receptor interactions. *Comput & Chem*. 1984; 8:281–283.
97. Klimovich PV, Shirts MR, Mobley DL. Guidelines for the Analysis of free Energy Calculations. *J Comput-Aided Mol Des*. 2015; 29:397–411. [PubMed: 25808134]
98. Zacharias M, Straatsma TP, McCammon JA. 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 DL, Case DA. Nonlinear Scaling Schemes for Lennard-Jones Interactions in free Energy Calculations. *J Chem Phys*. 2007; 127:214108. [PubMed: 18067350]
101. Naden LN, Shirts MR. 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. [PubMed: 26575553]

102. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF. 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 DA. Soft-core Potentials in Thermodynamic Integration: Comparing one- and Two-step Transformations. *J Comp Chem.* 2011; 32:3253–3263. [PubMed: 21953558]
104. Pham TT, Shirts MR. Identifying low Variance Pathways for free Energy Calculations of Molecular Transformations in Solution Phase. *J Chem Phys.* 2011; 135:034114. [PubMed: 21786994]
105. Shirts MR, Pande VS. Solvation free Energies of Amino acid side Chain Analogs for Common Molecular Mechanics Water Models. *J Chem Phys.* 2005; 122:134508. [PubMed: 15847482]
106. Gapsys V, Seeliger D, de Groot BL. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy calculations. *J Chem Theory Comput.* 2012; 8:2373–2382. [PubMed: 26588970]
107. Pham TT, Shirts MR. Optimal Pairwise and Non-pairwise Alchemical Pathways for free Energy Calculations of Molecular Transformation in Solution Phase. *J Chem Phys.* 2012; 136:124120. [PubMed: 22462848]
108. Naden LN, Pham TT, Shirts MR. Linear Basis Function Approach to Efficient Alchemical free Energy Calculations. 1. Removal of Uncharged Atomic Sites. *J Chem Theory Comput.* 2014; 10:1128–1149. [PubMed: 26580188]
109. Christ CD, Fox T. Accuracy Assessment and Automation of free Energy Calculations for drug Design. *J Chem Inf Model.* 2014; 54:108–120. [PubMed: 24256082]
110. Hansen N, Van Gunsteren WF. Practical Aspects of Free-energy Calculations: A Review. *J Chem Theory Comput.* 2014; 10:2632–2647. [PubMed: 26586503]
111. Mobley DL. Let's get Honest About Sampling. *J Comput-Aided Mol Des.* 2012; 26:93–95. [PubMed: 22113833]
112. Wang B, Wei GW. Parameter Optimization in Differential Geometry Based Solvation Models. *J Chem Phys.* 2015; 143:134119. [PubMed: 26450304]
113. Nessler IJ, Litman JM, Schnieders MJ. 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. [PubMed: 27524378]
114. Koziara KB, Stroet M, Malde AK, Mark AE. 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. [PubMed: 24477799]
115. Canzar S, El-Kebir M, Pool R, Elbassioni K, Malde AK, Mark AE, Geerke DP, Stougie L, Klau GW. Charge Group Partitioning in Biomolecular Simulation. *J Comp Biol.* 2013; 20:188–198.
116. Malde AK, Zuo L, Breeze M, Stroet M, Poger D, Nair PC, Oostenbrink C, Mark AE. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J Chem Theory Comput.* 2011; 7:4026–4037. [PubMed: 26598349]
117. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic simulations. *J Chem Theory Comput.* 2009; 5:350–358. [PubMed: 20150953]
118. Mobley DL, Dill KA, Chodera JD. Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small molecules. *J Phys Chem B.* 2008; 112:938–946. [PubMed: 18171044]
119. Parameswaran S, Mobley DL. Box size Effects are Negligible for Solvation free Energies of Neutral Solutes. *J Comput-Aided Mol Des.* 2014; 28:825–829. [PubMed: 24976043]
120. Berendsen HJC, Van Der Spoel D, van Drunen R. GROMACS: A Message-passing Parallel Molecular Dynamics Implementation. *Comp Phys Comm.* 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 AE, Berendsen HJC. GROMACS: Fast, Flexible, and Free. *J Comp Chem.* 2005; 26:1701–1718. [PubMed: 16211538]

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. [PubMed: 26620784]
124. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a High-throughput and Highly Parallel open Source Molecular Simulation Toolkit. *Bioinformatics (Oxford, England).* 2013; 29:845–854.
125. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, 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 MR, Mobley DL, Chodera JD, Pande VS. Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations. *J Phys Chem B.* 2007; 111:13052–13063. [PubMed: 17949030]
127. Parrinello M, Rahman A. Crystal Structure and Pair Potentials: A Molecular-Dynamics study. *Phys Rev Lett.* 1980; 45:1196–1199.
128. Henriksen NM, Fenley AT, Gilson MK. Computational Calorimetry: High-Precision Calculation of Host–Guest Binding thermodynamics. *J Chem Theory Comput.* 2015; 11:4377–4394. [PubMed: 26523125]
129. OpenEye Scientific Software, I. oechem. 2010.
130. OpenEye Scientific Software, I. QUACPAC 1702.
131. Hawkins PCD, 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. [PubMed: 23082786]
132. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: I. method. *J Comp Chem.* 2000; 21:132–146.
133. Jakalian A, Jack DB, Bayly CI. Fast, Efficient Generation of High-quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J Comp Chem.* 2002; 23:1623–1641. [PubMed: 12395429]
134. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys.* 1983; 79:926–935.
135. Gallicchio E, Levy RM. AGBNP: an Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-resolution Modeling. *J Comp Chem.* 2004; 25:479–499. [PubMed: 14735568]
136. Gallicchio E, Paris K, Levy RM. The AGBNP2 Implicit Solvation model. *J Chem Theory Comput.* 2009; 5:2544–2564. [PubMed: 20419084]
137. Boresch S, Archontis G, Karplus M. Free energy simulations: The meaning of the individual contributions from a component analysis. *Proteins: Struct, Funct Bioinf.* 1994; 20:25–33.
138. Zacharias M, Straatsma TP. Path Dependence of Free Energy Components in Thermodynamic Integration. *Mol Simu.* 1995; 14:417–423.
139. Dill KA. Additivity Principles in Biochemistry. *J Biol Chem.* 1997; 272:701–704. [PubMed: 8995351]
140. Levy RM, 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.* 1997; 49:701–704.
141. Hess B, van der Vegt NFA. 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. [PubMed: 16942107]
142. Fenley AT, Muddana HS, Gilson MK. 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. [PubMed: 23150595]
143. Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J Chem Phys.* 2004; 120:9665–9678. [PubMed: 15267980]

144. Plyasunova NV, Plyasunov AV, Shock EL. Database of Thermodynamic Properties for Aqueous Organic compounds. *Int J Thermophys.* 2004; 25:351–360.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

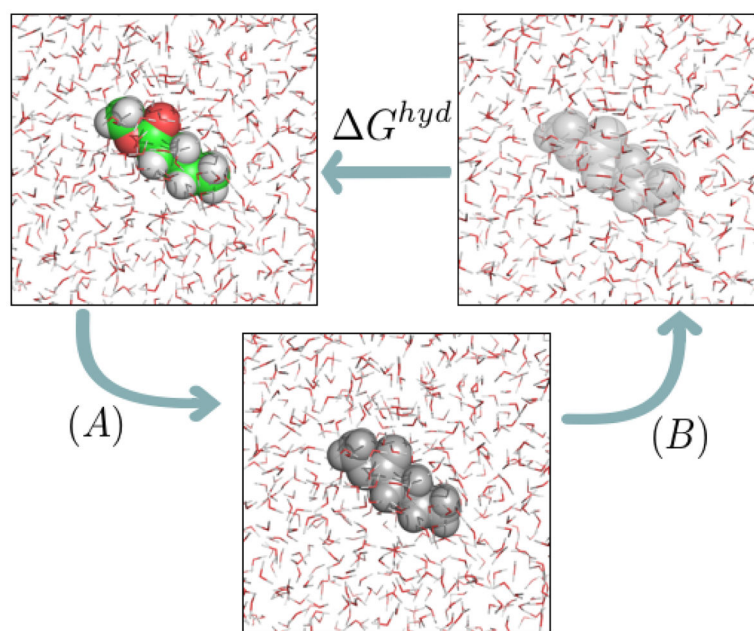


Figure 1. Thermodynamic cycle used to calculate hydration free energies (or, more generally, solvation free energies). In (A), we have states in which charge-charge interactions between the solute and its environment are progressively turned off. In (B) dispersion interactions between solute and water 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).

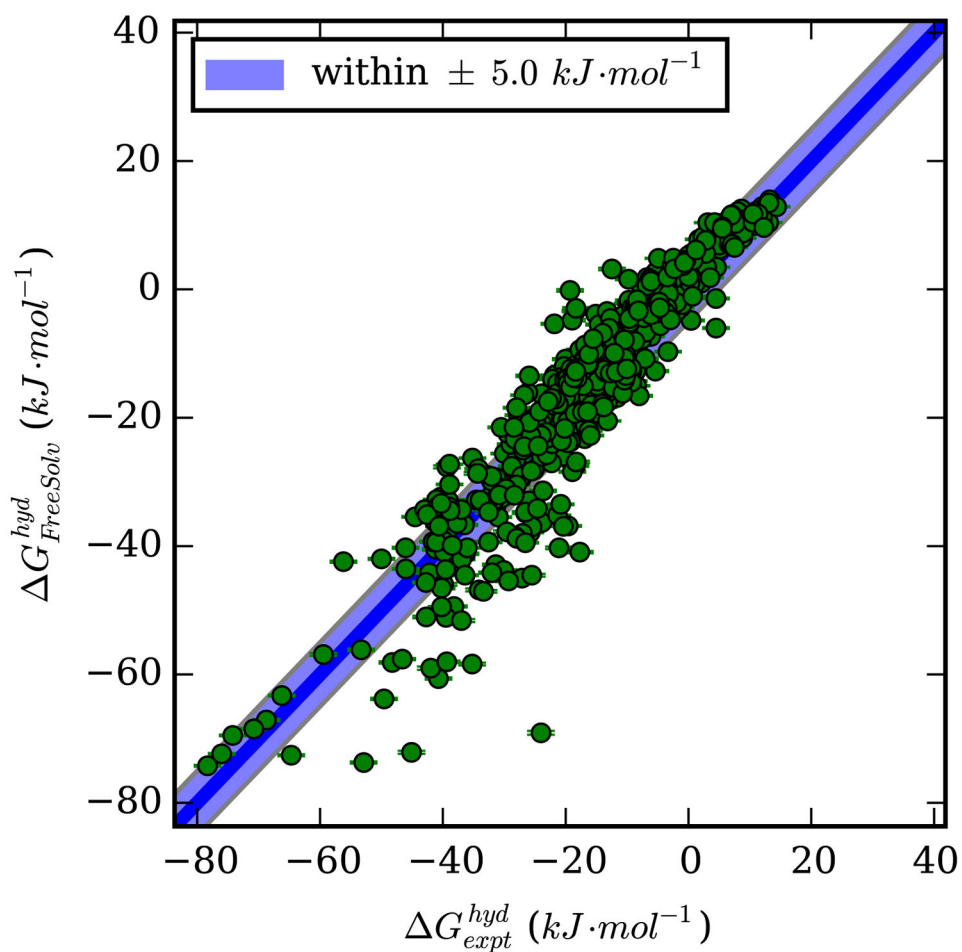


Figure 2. Calculated versus experimental hydration free energies for the compounds in Free-Solv version 0.5. Calculated values are on the vertical axis and experimental on the horizontal.

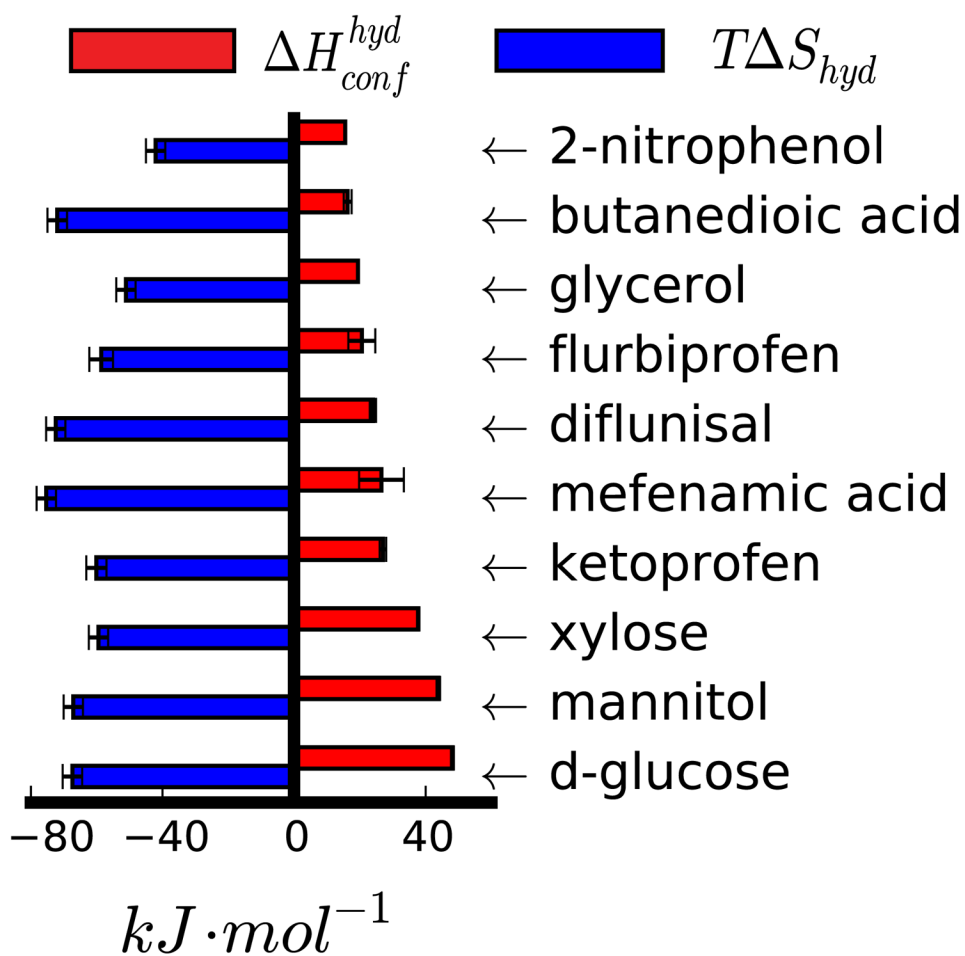


Figure 3.

Conformational enthalpies and associated entropies of compounds with highest ΔH_{conf}^{hyd} . Error bars represent the standard error.