

Article

Identifying systematic force field errors using a 3D-RISM element counting correction

Lizet Casillas^{†,‡}, Vahe M. Grigorian[†] and Tyler Luchko^{*}

Department of Physics and Astronomy, California State University, Northridge, Northridge, CA 91330, USA

* Correspondence: tluchko@csun.edu

† These authors contributed equally to this work.

‡ Current address: Department of Earth, Planetary and Space Sciences, University of California, Los Angeles, Los Angeles, CA 90095, USA

Abstract: Hydration free energies of small molecules are commonly used as benchmarks for solvation models. However, errors in predicting hydration free energies are partially due to the force fields used and not just the solvation model. To address this, we have used the 3D reference interaction site model (3D-RISM) of molecular solvation and existing benchmark explicit solvent calculations with a simple element count correction (ECC) to identify problems with the non-bond parameters in the general Amber force field (GAFF). 3D-RISM was used to calculate hydration free energies of all 642 molecules in the FreeSolv database and a partial molar volume correction (PMVC), ECC and their combination (PMVECC) were applied to the results. The PMVECC produced a mean unsigned error of 1.01 ± 0.04 kcal/mol and root mean squared error of 1.44 ± 0.07 kcal/mol, better than the benchmark explicit solvent calculations from FreeSolv and required less than 15 s of computing time per molecule on a single CPU-core. Importantly, parameters for PMVECC showed systematic errors for molecules containing Cl, Br, I and P. Applying ECC to the explicit solvent hydration free energies found the same systematic errors. The results strongly suggest that some small adjustments to the Lennard-Jones parameters for GAFF will lead to improved hydration free energy calculations for all solvent models.

Keywords: solvation; hydration free energy; 3D-RISM; force field; volume correction; conformational sampling; generalized Born; Lennard-Jones; implicit solvent; partial molar volume

Citation: Casillas, L.; Grigorian, V.M.; Luchko, T. Identifying systematic force field errors using a 3D-RISM element counting correction. *Molecules* **2022**, *1*, 0. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2023 by the authors. Submitted to *Molecules* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hydration free energy and, more generally, solvation free energy of small molecules are important quantities for predicting physical properties of small molecules, such as hydrophobicity, solubility, and molecular binding. Physics-based molecular simulations are a powerful tool for such calculations and several approaches have been developed with various levels of detail and computational demands. [1] Amongst the most common approaches is to treat the solute with an all-atom, force-field-based model, which is then coupled with a solvent model and appropriate method to calculate the HFE. When coupled with an explicit all-atom solvent model, thermodynamic integration or free energy perturbation are commonly used to calculate the HFE. This approach is computationally expensive as molecular dynamics simulations are required for each solute, each of which requires adequate conformational sampling. Alternatively, implicit solvents, such as Poisson-Boltzmann (PB) [2] or generalized Born (GB) [3], may be used. These models typically provide significant computational savings, as the detail of the solvent model is significantly reduced and, if the solute is sufficiently rigid, conformational sampling is not required and only calculations in the aqueous-phase are often all that is needed. The 3D reference interaction site model (3D-RISM) [4,5] of molecular solvation is an implicit solvent of particular interest because it uses all-atom forces fields to model the solvent but utilizes statistical mechanics via an approximated Ornstein-Zernike (OZ) equation [6] to

calculate equilibrium distributions. As such, it produces complete solvation thermodynamic information in a single calculation and avoids thermodynamic integration or free energy perturbation for rigid solutes. However, the accuracy of calculated HFEs depends on both model of the solvent and solute, both of which may have systematic errors.

It is well known that 3D-RISM over estimates HFEs for small molecules and several post-calculation corrections have been developed. [7–11] These corrections have been shown to produce HFEs of similar quality to explicit solvent calculations and have been extended to correct entropy [12] as well as other solvents [13–19]. While other approaches exist, such as creating new bridge functionals [11] or using hard-sphere liquids as a reference [20], most have employed corrections based on the partial molar volume (PMV). Recently, a correction using van der Waals (vdW) volumes instead PMVs was proposed and improved agreement with experiment was demonstrated on a rigid subset of molecules from the FreeSolv database. [10,21] However, calculating the van der Waals volume was not based on physical principles but was fit for each element to provide maximum agreement with experiment. This correction has a free parameter for each element in the dataset, compared to only two free parameters used in the original PMV correction by Palmer et al. [9]. At the same time, the flexibility provided by this element-based approach may be of practical use in improving the solute force field employed when coupled with any solvation model.

In this paper, we simplify the vdW volume correction and combine it with PMV corrections in 3D-RISM and apply it to explicit solvent simulations. First, we show that the vdW correction can be reproduced by counting the number of atoms of each element in the solute rather than calculating a volume. This requires minimal computation, is trivial to implement and, we will show, performs as well as using the vdW or partial molar volume. We then extend the correction by first combining it with the Palmer et al. [9] PMV correction and, separately, use it to improve the HFEs from the explicit solvent calculations of Mobley et al. [22,23]. The observed improvements suggest that this correction may address underlying deficiencies in the solute force field.

We begin with reviewing the theoretical basis of the PMV and vdW volume corrections in Section 2. We then provide results for the various corrections used with both 3D-RISM and explicit solvent for both rigid and flexible subsets of the FreeSolv database in Section 3. In Section 4, we discuss our results in the context of other corrections. In Section 5, we provide details of the calculations.

2. Theory

Details of the 3D-RISM theory have been provided elsewhere. [24,25] Here we focus on corrections applied after a 3D-RISM calculation has been carried out and the solvation free energy has been calculated.

2.1. PMV correction

The primary source of error in 3D-RISM calculations appears to be due to extreme over estimation of pressure in the model, which is typically around 1000 times too high for water at ambient density and temperature [26]. This pressure artifact is likely due to the closure approximation employed when solving 3D-RISM, as pressure consistency approaches have been shown to improve both the pressure and excess chemical potential in OZ calculations [27]. Indeed, the closures most commonly used with 3D-RISM are the Kovalenko-Hirata (KH) [28], partial series expansion of order- n (PSE- n) [29], and the Kobryn-Guasrov-Kovalenko (KGK) closures [30], which are approximations of the hypernetted chain equation (HNC) [31], known to overestimate the pressure [32]. This was first addressed by Palmer et al. [9] as a simple linear correction,

$$\Delta G_{\text{PMVC}} = \Delta G_{\text{RISM}} + av + b, \quad (1)$$

where v is the PMV, and a and b are free parameters to be fit against experiment. In this form, a can be interpreted as the overestimation of the pressure and is negative when fit to experiment. However, a does not numerically correspond to the pressure calculated

from compressibility or the free energy routes. It has been suggested that a corresponds to the contact pressure [8,33] but this approach does not work for the KH closure [12] or non-spherical molecules [13].

2.2. vdW volume correction

Robert et al. [10] proposed a pressure correction and the vdW volume (PCvdW) of each solute site with the pressure from the free energy route,

$$\Delta G_{\text{PCvdW}} = \Delta G_{\text{RISM}} + P_{\text{FE}} V_{\text{vdW}} \quad (2)$$

$$= \Delta G_{\text{RISM}} + P_{\text{FE}} \sum_i^{N_{\text{atoms}}} V_i, \quad (3)$$

where V_i is the vdW volume of solute atom i , and P_{FE} is the pressure from the free energy route, though any pressure can be used in practice as it is a fixed value defined by the bulk liquid. In this case, V_i were calculated from vdW radii, by placing the solute on a grid with 0.05 Å spacing and assigning voxels to particular atoms. The radii were optimized against simulation or experiment and found to be close to Bondi values [34].

2.3. Element count correction

While the procedure of calculating the total vdW volume from the atomic volumes on a grid is relatively inexpensive, it can be estimated as the product of the total number, N_i , and average occupied volume, v_i , of each element, allowing a simple element count correction (ECC):

$$\Delta G_{\text{ECC}} = \Delta G_{\text{RISM}} + P_{\text{FE}} \sum_i^{N_{\text{elements}}} v_i N_i. \quad (4)$$

The justification for this approach is that the number of bonds for each element is consistent, as are the bond lengths and angles, so we expect little variation in the vdW volume of an element in different chemical environments. We note that because there is overlap between the atoms, v_i cannot be used to meaningfully calculate vdW radii. The above equation can also be expressed as

$$\Delta G_{\text{ECC}} = \Delta G_{\text{RISM}} + \sum_i^{N_{\text{elements}}} c_i N_i, \quad (5)$$

where c_i is a fit coefficient, which is simply an unspecified element-wise correction to the solvent model, solute force-field, or both. It is this interpretation that we will use in this paper and one that lends itself to being combined with PMVC, Equation (1), giving a partial molar volume with element count correction (PMVECC)

$$\Delta G_{\text{PMVECC}} = \Delta G_{\text{RISM}} + av + b + \sum_i^{N_{\text{elements}}} c_i N_i. \quad (6)$$

This correction can account for both systematic errors in 3D-RISM as well as issues with individual elements, though at the cost of using considerably more parameters than the PMVC.

3. Results

3.1. Identifying rigid and flexible molecules using molecular dynamics with GB solvent

All 3D-RISM calculations in this study use a single conformer of the solute. To assess and mitigate errors due to this approach, we classified the solutes in the FreeSolv database as ‘flexible’ or ‘rigid’ for separate analysis. In this case, we are interested in flexibility as it relates to the HFE; however, it is computationally demanding to obtain correctly sampled conformers. To address this, we carried out MD simulations using GB as the solvent, which

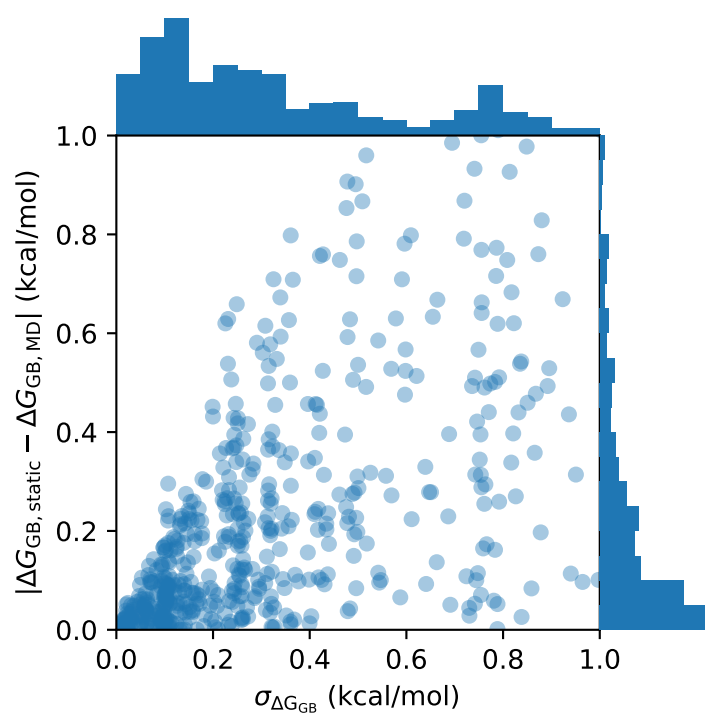


Figure 1. Categorizing rigid and flexible molecules from MD simulations. The standard deviation of the combined GB and surface area from MD simulations is given on the x -axis. The difference between E_{GB} calculated from just the first frame (static) and over the entire MD trajectory is given on the y -axis. Histograms for both quantities are given on their respective axes. For clarity, the full range of the data is not shown, which has maximum values of $\sigma_{\Delta G_{GB}} = 4.0$ kcal/mol and $|\Delta G_{GB,static} - \Delta G_{GB,MD}| = 7.4$ kcal/mol.

	PMVC	ECC	PMVECC	Explicit Solvent ECC
<i>a</i>	-0.15		-0.130(1)	
<i>b</i>	-0.04(1)		0.00(1)	
H		-1.199(1)	-0.225(5)	-0.098(1)
N		-1.573(6)	-0.392(7)	0.091(5)
C		-1.667(1)	-0.148(8)	0.114(1)
O		-1.277(3)	0.069(9)	0.088(3)
F		-2.082(4)	-0.05(1)	0.076(2)
Cl		-4.695(4)	-1.19(2)	-0.456(2)
Br		-5.544(7)	-1.06(2)	-0.412(6)
I		-6.27(1)	-0.79(3)	-0.25(1)
P		-1.03(3)	2.04(3)	2.93(3)
S		-3.18(1)	0.09(2)	0.32(1)

Table 1. Fit parameters for PMVC, ECC and PMVECC, averaged over all leave-one-out fits. Uncertainties in the last digit are given in parentheses and represent the standard deviation over all leave-one-out fits. Uncertainty for the *a* coefficient for PMVC is 8×10^{-5} kcal/mol/Å³. Coefficient *a* is in kcal/mol/Å³. All other values are in kcal/mol. See Section 5.5 for details of the fitting procedure.

allows for extensive sampling at modest computational cost and provides HFEs, ΔG_{GB} , for individual conformers from the simulation. Molecules that are rigid should have low standard deviation in ΔG_{GB} over the course of a simulation and this should be reflected in a small difference between ΔG_{GB} calculated from the entire simulation, $\Delta G_{GB,MD}$, and first frame only ($\Delta G_{GB,static}$).

Figure 1 shows the distribution of ΔG_{GB} standard deviations, $\sigma_{\Delta G_{GB}}$, over the course of the simulations and $|\Delta G_{GB,static} - \Delta G_{GB,MD}|$. Almost all molecules with a low standard deviation, $\sigma_{\Delta G_{GB}} \leq 0.4$ kcal/mol, show a difference in ΔG_{GB} between single snapshots and MD of no more than 0.2 kcal/mol. Though we do observe that some molecules have a poor initial configuration and $|\Delta G_{GB,static} - \Delta G_{GB,MD}|$ is larger than 0.2 kcal/mol despite low $\sigma_{\Delta G_{GB}}$. This demonstrates, as expected, that there is little difference between using a rigid structure and a full MD simulation if the molecules display low HFE variance. As a result we define the 313 molecules with $\sigma_{\Delta G_{GB}} \leq 0.4$ kcal/mol and $|\Delta G_{GB,static} - \Delta G_{GB,MD}| \leq 0.2$ kcal/mol as rigid and the remaining 328 molecules as flexible for the rest of the study. The full list of molecules with $\Delta G_{GB,MD}$, $\Delta G_{GB,static}$, $\sigma_{\Delta G_{GB}}$ and rigid/flexible classification is available in Spreadsheet S1.

3.2. Fitting PMVC, ECC and PMVECC parameters

After identifying rigid and flexible molecules, 3D-RISM calculations were run on single conformers of all molecules. The PMVC, ECC and PMVECC were fit to all the data for 3D-RISM. Only ECC was fit to explicit solvent MD calculations from the FreeSolv database [22] (Table 1). The values of the fit parameters for PMVC (Table 1) are quite close to prior results [9,12]. Many of the molecules in the FreeSolv database were also used in prior fits, but this shows that the fit is not sensitive to the molecules used.

When fit without the PMVC, the ECC parameters are all negative (Table 1). We expect these parameters to be negative and correlated with the size of the elements because uncorrected 3D-RISM overestimates the HFE in a way that can be largely addressed with the PMVC. The largest corrections are for F, Cl, Br, I and S. With the exception of S, these are all halides and the correction scales roughly with their size, so the correction is largely a pressure correction. The correction for S is large compared to oxygen and may indicate some discrepancy in the parameters.

The PMVECC fit shows only small changes in the PMVC parameters, *a* and *b*, compared to the PMVC-only fit, while the element specific parameters differ greatly from the ECC-only fit (Table 1). This demonstrates that the PMVC accounts for the major errors

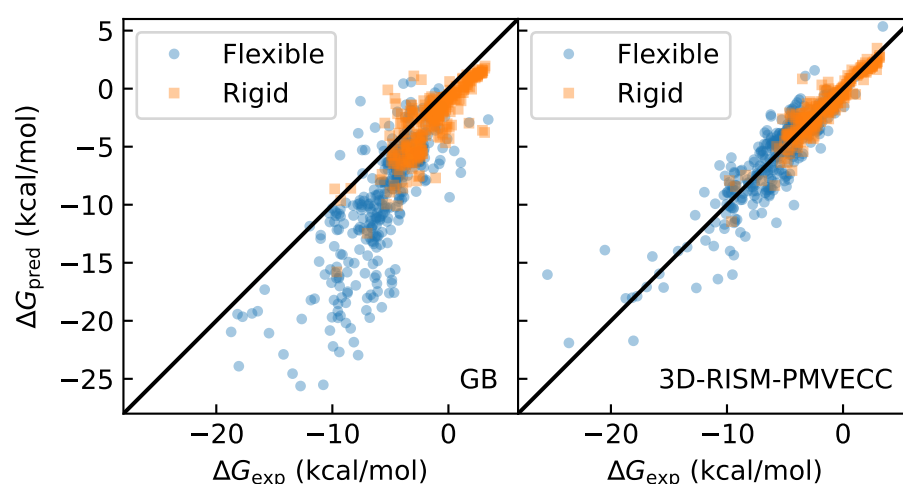


Figure 2. HFEs from a single (original conformation), rigid and flexible datasets for GB and 3D-RISM with PMVECC.

due to over estimating the pressure, while the ECC can account for discrepancies in how the force field parameters interact with the solvent. For O, F and S, the fit parameters are positive and negative values close to zero, suggesting that the force field parameters are well tuned. Elements H, N, and C all have slightly negative values, suggesting that the interaction between water and these atoms could be improved. The remaining elements (Cl, Br, I, and P) have larger coefficients, suggesting larger adjustments are required. Of these, P requires a much larger correction, indicating that this is a particular problem for the force field parameterization.

Applying ECC to the explicit solvent calculations provides a similar picture to the PMVECC fit for 3D-RISM. Clearly, no pressure correction is required for the explicit solvent calculations and we observe that the element coefficients are strongly correlated with those for 3D-RISM-PMVECC ($R^2 = 0.91$) (Table 1). Values for O and F are again close to zero, though S now has a larger coefficient. In contrast to PMVECC, values for H and N are close to zero and the values for C has the opposite sign. Elements Cl, Br and I have relatively large, negative coefficients that, while smaller than those of 3D-RISM-PMVECC, suggesting that both explicit solvent and 3D-RISM would benefit from the same force-field parameter adjustments. We also see that the correction for P is positive and even larger, suggesting that this element is a particular problem.

3.3. Quality of fit

To assess the quality of the different fits, we compared the corrected HFEs to results from experiment in Table 2 and Figure 3 and assess the flexible and rigid molecules separately as well as combined. Though the two datasets are nearly equal in size (313 rigid and 328 flexible), the rigid molecules are clustered in a much smaller range of HFE values (Figure 2). This is important because, for the same amount of relative error, the Pearson correlation coefficient, R^2 will be larger when computed over a smaller domain. Despite this, R^2 is larger for the rigid molecules compared to the flexible ones due to the much smaller mean-unsigned errors (MUE), root-mean-squared errors (RMSE) and maximum errors for the rigid molecules. We also observe that the maximum error is much larger for the flexible molecules, even after the ECC was applied. Together, these highlight the importance of sampling, even for small molecules, and suggest that the explicit solvent calculations may benefit from additional conformational sampling.

For all groupings of molecules, a similar picture emerges for the quality of the different corrections applied to 3D-RISM. The PMVC has lower MUE and RMSE compared to ECC, while the R^2 values are comparable. Unsurprisingly, PMVECC has the lowest

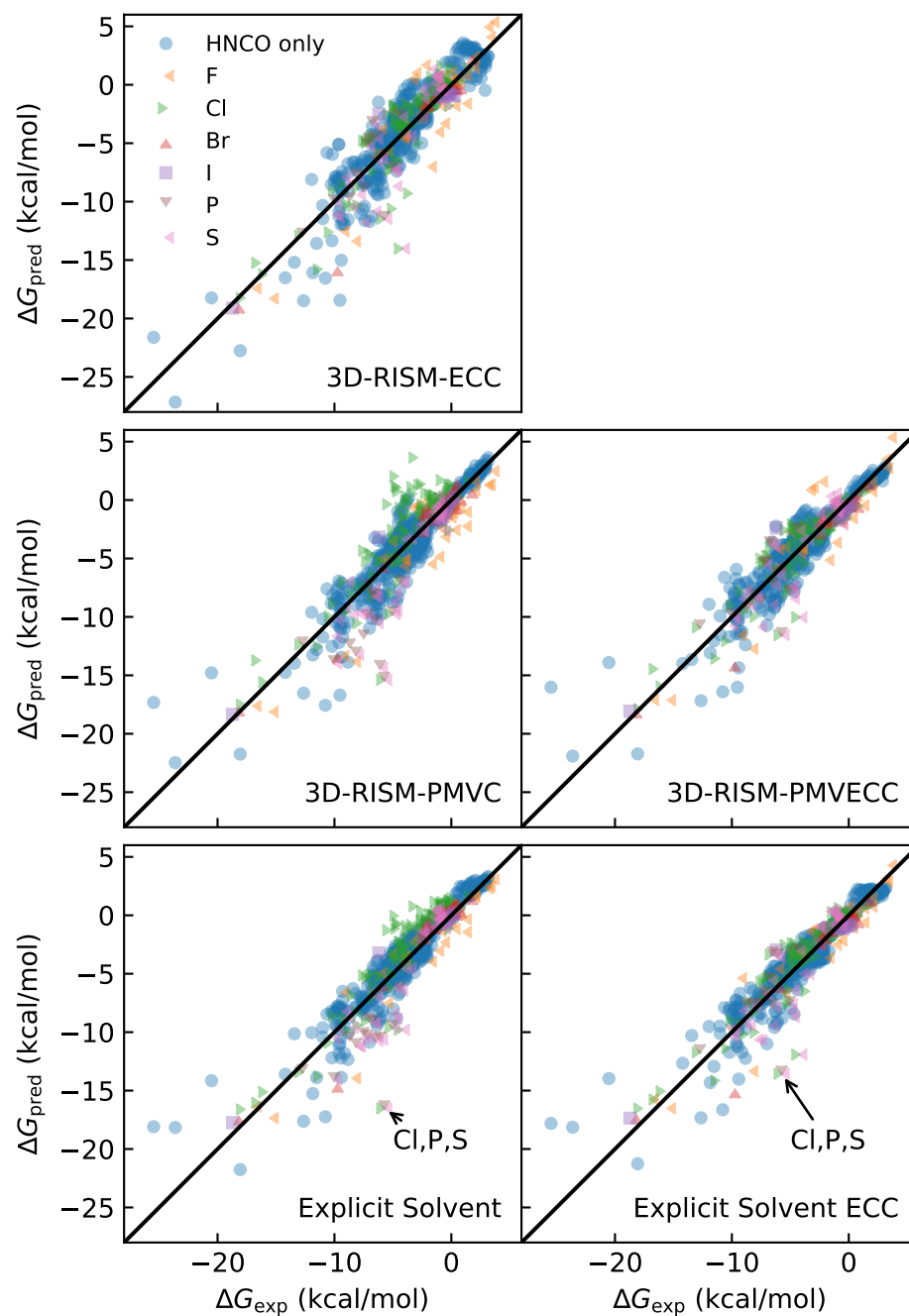


Figure 3. HFEs for 3D-RISM-PMVC, 3D-RISM-ECC, 3D-RISM-PMVECC, and explicit solvent using parameters from Table 1. Leave-out data was used for all plots, except for uncorrected explicit solvent calculations, which are from Ref. [22]. Molecules containing combinations of F, Cl, Br, P and S atoms are plotted with multiple symbols (e.g., see labeled molecule in the bottom row). See Section 5.5 for details of the fitting procedure.

	Slope	MUE	MSE	RMSE	R^2	Max Error
Rigid						
3D-RISM-PMVC	0.93(4)	0.86(6)	-0.29(7)	1.3(1)	0.75(4)	6.6
3D-RISM-ECC	0.92(4)	1.02(6)	-0.51(8)	1.37(8)	0.76(3)	5.9
3D-RISM-PMVECC	0.92(2)	0.61(3)	0.05(5)	0.83(6)	0.89(2)	4.4
Explicit solvent	0.96(3)	0.85(4)	-0.59(6)	1.11(6)	0.86(2)	4.6
Explicit solvent-ECC	0.91(2)	0.66(3)	-0.14(5)	0.86(4)	0.88(1)	3.1
Flexible						
3D-RISM-PMVC	0.98(4)	1.53(8)	0.2(1)	2.1(1)	0.75(3)	9.6
3D-RISM-ECC	1.07(4)	1.56(7)	0.0(1)	2.1(1)	0.78(3)	9.8
3D-RISM-PMVECC	0.95(5)	1.35(6)	-0.04(9)	1.8(1)	0.79(3)	9.4
Explicit solvent	0.97(4)	1.34(7)	-0.09(0)	1.8(1)	0.79(3)	10.8
Explicit solvent-ECC	0.91(4)	1.17(6)	-0.13(9)	1.7(1)	0.81(3)	7.8
Total						
3D-RISM-PMVC	1.01(3)	1.22(5)	0.00(7)	1.77(9)	0.83(2)	9.6
3D-RISM-ECC	1.06(2)	1.32(5)	-0.21(7)	1.80(8)	0.84(2)	9.8
3D-RISM-PMVECC	0.96(3)	1.01(4)	0.00(6)	1.44(7)	0.87(1)	9.4
Explicit solvent	1.02(3)	1.11(4)	-0.32(6)	1.53(8)	0.87(1)	10.8
Explicit solvent-ECC	0.94(2)	0.94(4)	-0.13(5)	1.35(8)	0.88(1)	7.8

Table 2. Hydration free energies calculated with 3D-RISM and explicit solvent [22] with PMVC, ECC and PMVECC corrections using parameters from Table 1. Leave-out data was used to calculate statistics, except for uncorrected explicit solvent calculations, which used data from Ref. [22] with the same bootstrap procedure. All values are given in kcal/mol. Uncertainties in the last digit are given in parentheses and represent the standard error in the mean. See Section 5.5 for details of the fitting procedure.

errors and highest R^2 of all three corrections. In fact, PMVECC outperforms uncorrected explicit solvent calculation and provides similar accuracy and correlation as explicit solvent calculations with the ECC (Table 2).

A large part of the success of the PMVECC is in correcting outliers. For example, clusters of outliers containing Cl, P and S atoms can be observed for 3D-RISM-PMVC and uncorrected explicit solvent in Figure 3, where HFEs are overestimated for Cl containing molecules and underestimated for P and S containing molecules. Accordingly, the PMVECC and explicit solvent ECC coefficients for are negative for Cl and positive for P, and S (Table 1), correcting the worst outliers. The ECC correction for 3D-RISM does mitigate the errors for many of these molecules, but the lack of an overall pressure correction limits its effectiveness. We also note that PMVECC effectively deals with molecules that contain multiple Cl, Br, I, P and S atoms. For example, of the 14 P containing molecules, 10 contain S, 5 contain Cl and 3 contain both S and Cl. Similarly, the largest cluster of Cl outliers contain five or more Cl atoms per molecule. To illustrate the magnitude of the correction, the Cl coefficient in the PMVECC contributes -5.94 kcal/mol to the HFE for a molecule with five Cl atoms.

4. Discussion

4.1. Dealing with conformational sampling

Insufficient conformational sampling is a common source of error when MD simulations are used [35,36], yet it is common to use single conformers of small ligands when testing or parameterizing implicit solvent models, especially for 3D-RISM (e.g., [7,9,12,14–19]). Care is required because a single conformer approach neglects the solute adopting different conformations in solvent than in vacuum, conformers that not dominant in either state, or molecules that have multiple conformers. Furthermore, the structure provided by FreeSolv or any software package may or may not be in the dominant conformation for either state. While, improving sampling for 3D-RISM is an import topic, it is beyond the

scope of this study. Therefore, we wish to differentiate between rigid and flexible molecules to better assess parameter fitting and the quality of the corrections.

Solute conformational sampling, of course, can be accounted for using free energy perturbation. For example, Mobley et al. [35] connected simulations in vacuum and GB solvents using the Bennett acceptance ratio [37]. We have previously employed a similar approach with 3D-RISM [13]; however, despite work on using 3D-RISM with MD [38–43], this approach is still computationally demanding. An alternative is to sample in a state that will overlap the conformational space in 3D-RISM and exponential averaging [44] to calculate the free energy difference. GB could be used for such an approach, but the overlap between GB and 3D-RISM conformational space is unknown and possibly insufficient to obtain accurate results.

Rather than addressing conformational sampling directly, we follow the approach of Robert et al. [10] by partitioning the data set into flexible and rigid groups. However, our approach differs in how we categorize molecules as rigid or flexible. Robert et al. [10] compared the SFEs calculated by rigid Monte Carlo in explicit solvent to the value computed with MD in the FreeSolv database. Those that differed by < 0.1 kcal/mol were considered rigid. However, their approach required HFEs to be calculated in explicit solvent for both completely fixed conformation and flexible solutes – an expensive calculation we would like to avoid. Furthermore, if the right conformer is selected, the HFE of the rigid molecule could be within the threshold when compared to the flexible solute HFE, even if the molecule is quite flexible. In contrast, our approach used MD with a GB solvent and defined molecules to be rigid if $\sigma_{\Delta G_{GB}} \leq 0.4$ kcal/mol and the difference between the first frame and the HFE for the trajectory calculated by the Bennett acceptance ratio is < 0.2 kcal/mol. The first criterion is a measure of how flexibility directly contributes to the HFE by utilizing the excellent computational and sampling efficiency of GB [45], while the second criterion removes molecules with poor initial conformers. In all, we define 313 molecules as rigid, compared to 288 and 520 molecules defined as rigid by Robert et al. [10] and Luukkonen et al. [21] respectively. A full list of molecules with assigned categories is available in Spreadsheet S1. The categorization of molecules as rigid or flexible was only used to assess the quality of the corrections and not to fit the parameters as the small range of HFEs for rigid from experiment could bias the results. Furthermore, the error due to lack of conformational sampling should be randomly distributed, so there should be no systematic effect on fitting the corrections.

From our analysis we see that there are larger errors for flexible molecules than for rigid ones, both for 3D-RISM and explicit solvent calculations. Two potential explanations could account for this. A constant relative error for all HFEs would lead to an increased absolute error for molecules with large magnitude HFEs. Flexible molecules have a wider distribution of HFE values and would be expected to have larger absolute errors on average. Alternatively, flexible molecule conformations may be under sampled, even in explicit solvent MD. We can test this by considering the relative error,

$$\epsilon = \left| \frac{\Delta G_{\text{pred}} - \Delta G_{\text{exp}}}{\Delta G_{\text{exp}}} \right|,$$

for rigid and flexible molecules. To ensure the relative error is well behaved, we avoid dividing by a small number by only consider molecules with $\Delta G_{\text{exp}} < -5$ kcal/mol. We find that the relative error for flexible molecules in explicit solvent is slightly lower than that for rigid molecules (22% for rigid and 20% for flexible), but in 3D-RISM-PMVECC flexible molecules have a larger relative error (16% for rigid and 19% for flexible). Thus, the explicit solvent MD calculations likely have adequate sampling, whereas 3D-RISM calculations do not. Improving solute conformational sampling in 3D-RISM calculations is an important next step for increasing accuracy, even for small molecules.

4.2. Accuracy and computational efficiency of 3D-RISM-PMVECC

Of the several PMV-based HFE corrections that have been previously developed [7–9], we found that the original correction proposed by Palmer et al. [9] performed the best until now. [12] By itself, we find that the ECC is of similar accuracy to PMVC, despite including many more parameters. Overall, the correlation is improved and systematic errors for specific elements are removed, though there is a slight increase in the error. Levesque and co-workers report similar statistics using their original method that employed calculating the vdW volume instead of counting atoms. [10,21] PMVECC shows a significant improvement in both the error and the correlation to experiment. In fact, it statistically outperforms uncorrected explicit solvent. With recent improvements to our 3D-RISM implementation [46], the average calculation time per molecule on a single CPU-core was 14.7 s and the maximum time was 38.1 s.

Though simple to implement, PMVECC does require sufficient experimental data to fit 12 parameters, potentially limiting its practical use. In addition, all or nearly all of the atoms in the FreeSolv database are solvent accessible and it is likely that the PMVECC correction is not transferrable to molecules with many buried atoms. Future improvements to calculating HFEs will likely come from new closure approximations instead [11,27,47].

4.3. Force field parameters

A potential immediate use of the PMVECC or ECC is to correct solute force field parameters. While errors have been correlated with various solute properties, such as functional groups [21,22,48], in the absence of conformational sampling, the only force field parameters that matter are the partial charges and the Lennard-Jones parameters. Different methods of assigning partial charges may yield better results but alternate approaches have not yet improved on AM1-BCC for GAFF [49]. A more promising approach, in our opinion is to optimize Lennard-Jones parameters, which play a significant role in controlling the distance between the charge sites on the solute and solvent. With this in mind, Table 1 provides a clear prescription that Lennard-Jones radii should be increased for P and S and decreased for Cl, Br and I in the GAFF [50] used in the FreeSolv database. Generally, 3D-RISM would benefit from greater changes than for explicit solvent but the trend is the same in both cases.

5. Materials and Methods

5.1. Structure preparation

For all calculations, we used the Amber parameter and coordinate files provide with the FreeSolv v0.52 dataset. [22,51] The FreeSolv authors used GAFF for the solute force field [50] with AM1-BCC charges [52,53].

5.2. GB HFE

From the initial coordinates provided with FreeSolv, we ran 100 ns of MD simulation using both GB with surface area implicit solvent (igb=2, gbsa=1) and vacuum (igb=6) [3,54,55] environments in the sander MD engine from AmberTools 2017 [56]. For all simulations, a 1 fs time step was used, temperature was held to 298.15 K using a Langevin thermostat with $\gamma = 5 \text{ ps}^{-1}$, and conformations were saved every 10,000 steps. The resulting trajectories were then post-processed in sander (imin=5) using the GB with surface area implicit solvent (igb=2, gbsa=1) and in vacuum to obtain the potential energy of each conformation in aqueous and gas phases. HFEs were then calculated from these potential energies using pyMBAR 3.1.1. [57,58]

5.3. 1D-RISM

Bulk properties of water at 298.15 K and a concentration of 55.4 M were calculated with *rism1d* from AmberTools 2021 [43,59]. The coincident extend simple point charge model (cSPC/E) was used to model water. [43,60] The dielectrically consistent RISM (DRISM) equations [61,62] were solved with a dielectric constant of 78.497 to a residual tolerance of

10^{-12} on 16,384 point grid, with a grid spacing of 0.025 Å. Convergence was accelerated with the modified inversion of iterative subspace (MDIIS) method. [63]

5.4. 3D-RISM Calculations

rism3d.snglpnt from AmberTools 2021 [43,59] was used to calculate the HFE and PMV using the Amber parameter and coordinate files provided with the FreeSolv dataset for each solute and bulk water properties from rism1d. The 3D-RISM equations were solved to a residual tolerance of 10^{-4} on a grid with spacing of 0.3 Å, accelerated by MDIIS. Lennard-Jones cutoffs with a relative tolerance of 10^{-4} were used to determine the size of the grid and analytic corrections were applied. [46] Reciprocal-space long-range asymptotics were calculated with a relative tolerance of 10^{-5} .

5.5. Parameter fitting

Parameters for the PMVC, ECC and PMVECC were fit with leave-one-out cross validation as follows. HFEs calculated by 3D-RISM and explicit solvent molecular dynamics [22] for the 642 molecules in the FreeSolv database were collected in a Pandas dataframe [64,65] and ordinary least squares was applied to produce 642 fits of the parameters to the data using the statsmodels package [64]. One molecule was left out for each fit in turn. Values in Table 1 are the mean of all fits and the uncertainties are the standard errors in the mean. Statistics in Table 2 were computed from the leave-out data from generating the parameters Table 1 and uncertainties are the standard errors in the mean for 1000 rounds of bootstrap analysis. Coefficients in Table 1 and all of the leave-out data used for Table 2 and Figure 3 can be reproduced with Script S1 in the Supplementary Materials. The corrected leave-out HFEs can also be found in Spreadsheet S1.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/molecules1010000/s1>, Spreadsheet S1: HFEs from all methods (kcal/mol), elements counts, and PMVs (Å³) for each molecule in the FreeSolv database. Script S1: Python script to calculate correction parameters in Table 1 and corrected HFEs used in Table 2 and Figure 3.

Author Contributions: Conceptualization, T.L.; methodology, L.C., V.G., T.L.; software, L.C., V.G.; validation, L.C., V.G.; formal analysis, L.C., V.G., T.L.; investigation, L.C., V.G., T.L.; resources, T.L.; data curation, L.C., V.G., T.L.; writing—original draft preparation, L.C., V.G., T.L.; writing—review and editing, L.C., V.G., T.L.; visualization, L.C., V.G., T.L.; supervision, T.L.; project administration, T.L.; funding acquisition, T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the National Science Foundation (NSF) under Grants CHE-2102668 and CHE-2018427.

Data Availability Statement: The data presented in this study are available in supplementary material.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. E. Skyner, R.; L. McDonagh, J.; R. Groom, C.; Mourik, T.v.; O. Mitchell, J.B. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Physical Chemistry Chemical Physics* **2015**, *17*, 6174–6191. Publisher: Royal Society of Chemistry, doi:10.1039/C5CP00288E.
2. Sharp, K.A.; Honig, B. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *The Journal of Physical Chemistry* **1990**, *94*, 7684–7692. doi:10.1021/j100382a068.
3. Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society* **1990**, *112*, 6127–6129. doi:10.1021/ja00172a038.
4. Kovalenko, A.; Hirata, F. Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model. *The Journal of Chemical Physics* **1999**, *110*, 10095–10112. doi:10.1063/1.478883.

5. Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *The Journal of Physical Chemistry B* **1997**, *101*, 7821–7826. Publisher: American Chemical Society, doi:10.1021/jp971083h. 363
6. Ornstein, L.S.; Zernike, F. Accidental deviations of density and opalescence at the critical point of a single substance. *Proc. Akad. Sci.(Amsterdam)*, 1914, Vol. 17, p. 793. 364
7. Truchon, J.F.; Pettitt, B.M.; Labute, P. A Cavity Corrected 3D-RISM Functional for Accurate Solvation Free Energies. *Journal of Chemical Theory and Computation* **2014**, *10*, 934–941. doi:10.1021/ct4009359. 365
8. Sergiievskiy, V.P.; Jeanmairet, G.; Levesque, M.; Borgis, D. Fast Computation of Solvation Free Energies with Molecular Density Functional Theory: Thermodynamic-Ensemble Partial Molar Volume Corrections. *The Journal of Physical Chemistry Letters* **2014**, *5*, 1935–1942. Publisher: American Chemical Society, doi:10.1021/jz500428s. 366
9. Palmer, D.S.; Frolov, A.I.; Ratkova, E.L.; Fedorov, M.V. Towards a universal method for calculating hydration free energies: a 3D reference interaction site model with partial molar volume correction. *Journal of Physics: Condensed Matter* **2010**, *22*, 492101. doi:10.1088/0953-8984/22/49/492101. 367
10. Robert, A.; Luukkonen, S.; Levesque, M. Pressure correction for solvation theories. *The Journal of Chemical Physics* **2020**, *152*, 191103. Publisher: American Institute of Physics, doi:10.1063/5.0002029. 368
11. Borgis, D.; Luukkonen, S.; Belloni, L.; Jeanmairet, G. Accurate prediction of hydration free energies and solvation structures using molecular density functional theory with a simple bridge functional. *The Journal of Chemical Physics* **2021**, *155*, 024117. Publisher: American Institute of Physics, doi:10.1063/5.0057506. 369
12. Johnson, J.; Case, D.A.; Yamazaki, T.; Gusarov, S.; Kovalenko, A.; Luchko, T. Small molecule hydration energy and entropy from 3D-RISM. *Journal of Physics: Condensed Matter* **2016**, *28*, 344002. doi:10.1088/0953-8984/28/34/344002. 370
13. Luchko, T.; Blinov, N.; Limon, G.C.; Joyce, K.P.; Kovalenko, A. SAMPL5: 3D-RISM partition coefficient calculations with partial molar volume corrections and solute conformational sampling. *Journal of Computer-Aided Molecular Design* **2016**, *30*, 1115–1127. PubMed PMID: 27585474, doi:10.1007/s10822-016-9947-7. 371
14. Roy, D.; Kovalenko, A. A molecular solvation theory simulation of liquid alkyl esters of acetic acid with the 3D Reference Interaction Site Model. *Journal of Molecular Liquids* **2021**, *344*, 117763. doi:10.1016/j.molliq.2021.117763. 372
15. Roy, D.; Kovalenko, A. Application of the Approximate 3D-RISM Molecular Solvation Theory to Acetonitrile as Solvent. *The Journal of Physical Chemistry B* **2020**. Publisher: American Chemical Society, doi:10.1021/acs.jpcc.0c03230. 373
16. Roy, D.; Kovalenko, A. Performance of 3D-RISM-KH in Predicting Hydration Free Energy: Effect of Solute Parameters. *The Journal of Physical Chemistry A* **2019**, *123*, 4087–4093. Publisher: American Chemical Society, doi:10.1021/acs.jpca.9b01623. 374
17. Roy, D.; Kovalenko, A. Benchmarking Free Energy Calculations in Liquid Aliphatic Ketone Solvents Using the 3D-RISM-KH Molecular Solvation Theory. *J* **2021**, *4*, 604–613. doi:10.3390/j4040044. 375
18. Roy, D.; Kovalenko, A. Application of the 3D-RISM-KH molecular solvation theory for DMSO as solvent. *Journal of Computer-Aided Molecular Design* **2019**. doi:10.1007/s10822-019-00238-4. 376
19. Roy, D.; Blinov, N.; Kovalenko, A. Predicting Accurate Solvation Free Energy in n-Octanol Using 3D-RISM-KH Molecular Theory of Solvation: Making Right Choices. *The Journal of Physical Chemistry B* **2017**, *121*, 9268–9273. Publisher: American Chemical Society, doi:10.1021/acs.jpcc.7b06375. 377
20. Sumi, T.; Mitsutake, A.; Maruyama, Y. A solvation-free-energy functional: A reference-modified density functional formulation. *Journal of Computational Chemistry* **2015**, *36*, 1359–1369. doi:10.1002/jcc.23942. 378
21. Luukkonen, S.; Belloni, L.; Borgis, D.; Levesque, M. Predicting Hydration Free Energies of the FreeSolv Database of Drug-like Molecules with Molecular Density Functional Theory. *Journal of Chemical Information and Modeling* **2020**, *60*, 3558–3565. doi:10.1021/acs.jcim.0c00526. 379
22. Mobley, D.L.; Guthrie, J.P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 711–720. doi:10.1007/s10822-014-9747-x. 380
23. Duarte Ramos Matos, G.; Kyu, D.Y.; Loeffler, H.H.; Chodera, J.D.; Shirts, M.R.; Mobley, D.L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *Journal of Chemical & Engineering Data* **2017**, *62*, 1559–1569. doi:10.1021/acs.jced.7b00104. 381
24. Kovalenko, A. Three-dimensional RISM Theory for Molecular Liquids and Solid-Liquid Interfaces. In *Molecular Theory of Solvation*; Hirata, F., Ed.; Number 24 in Understanding Chemical Reactivity, Springer Netherlands, 2004; pp. 169–275. 382
25. Luchko, T.; Joung, I.S.; Case, D.A. Chapter 4 Integral Equation Theory of Biomolecules and Electrolytes. In *Innovations in Biomolecular Modeling and Simulations*; 2012. 383
26. Joung, I.S.; Luchko, T.; Case, D.A. Simple electrolyte solutions: Comparison of DRISM and molecular dynamics results for alkali halide solutions. *The Journal of Chemical Physics* **2013**, *138*, 044103. doi:10.1063/1.4775743. 384
27. Tsednee, T.; Luchko, T. Closure for the Ornstein-Zernike equation with pressure and free energy consistency. *Physical Review E* **2019**, *99*, 032130. doi:10.1103/PhysRevE.99.032130. 385
28. Kovalenko, A.; Hirata, F. Potential of Mean Force between Two Molecular Ions in a Polar Molecular Solvent: A Study by the Three-Dimensional Reference Interaction Site Model. *The Journal of Physical Chemistry B* **1999**, *103*, 7942–7957. Publisher: American Chemical Society, doi:10.1021/jp991300+. 386
29. Kast, S.M.; Kloss, T. Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. *The Journal of Chemical Physics* **2008**, *129*, 236101. doi:10.1063/1.3041709. 387

30. Kobryn, A.E.; Gusarov, S.; Kovalenko, A. A closure relation to molecular theory of solvation for macromolecules. *Journal of Physics: Condensed Matter* **2016**, *28*, 404003. doi:10.1088/0953-8984/28/40/404003.
31. Morita, T.; Hiroike, K. A New Approach to the Theory of Classical Fluids. I. *Progress of Theoretical Physics* **1960**, *23*, 1003–1027. doi:10.1143/PTP.23.1003.
32. Verlet, L.; Levesque, D. On the theory of classical fluids II. *Physica* **1962**, *28*, 1124–1142. doi:10.1016/0031-8914(62)90058-7.
33. Sergiievskiy, V.; Jeanmairet, G.; Levesque, M.; Borgis, D. Solvation free-energy pressure corrections in the three dimensional reference interaction site model. *The Journal of Chemical Physics* **2015**, *143*, 184116. doi:10.1063/1.4935065.
34. Bondi, A. van der Waals Volumes and Radii. *The Journal of Physical Chemistry* **1964**, *68*, 441–451. doi:10.1021/j100785a001.
35. Mobley, D.L.; Dill, K.A.; Chodera, J.D. Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small Molecules. *The Journal of Physical Chemistry B* **2008**, *112*, 938–946. doi:10.1021/jp0764384.
36. Knapp, B.; Ospina, L.; Deane, C.M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *Journal of Chemical Theory and Computation* **2018**, *14*, 6127–6138. doi:10.1021/acs.jctc.8b00391.
37. Bennett, C.H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, *22*, 245–268. doi:10.1016/0021-9991(76)90078-4.
38. Omelyan, I.; Kovalenko, A. Enhanced solvation force extrapolation for speeding up molecular dynamics simulations of complex biochemical liquids. *The Journal of Chemical Physics* **2019**, *151*, 214102. doi:10.1063/1.5126410.
39. Omelyan, I.; Kovalenko, A. MTS-MD of Biomolecules Steered with 3D-RISM-KH Mean Solvation Forces Accelerated with Generalized Solvation Force Extrapolation. *Journal of Chemical Theory and Computation* **2015**, *11*, 1875–1895. doi:10.1021/ct5010438.
40. Omelyan, I.; Kovalenko, A. Multiple time step molecular dynamics in the optimized isokinetic ensemble steered with the molecular theory of solvation: Accelerating with advanced extrapolation of effective solvation forces. *The Journal of Chemical Physics* **2013**, *139*, 244106. doi:10.1063/1.4848716.
41. Omelyan, I.; Kovalenko, A. Generalised canonical–isokinetic ensemble: speeding up multiscale molecular dynamics and coupling with 3D molecular theory of solvation. *Molecular Simulation* **2013**, *39*, 25–48. doi:10.1080/08927022.2012.700486.
42. Miyata, T.; Hirata, F. Combination of molecular dynamics method and 3D-RISM theory for conformational sampling of large flexible molecules in solution. *Journal of Computational Chemistry* **2008**, *29*, 871–882. doi:10.1002/jcc.20844.
43. Luchko, T.; Gusarov, S.; Roe, D.R.; Simmerling, C.; Case, D.A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *Journal of Chemical Theory and Computation* **2010**, *6*, 607–624. doi:10.1021/ct900460m.
44. Zwanzig, R.W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426. doi:10.1063/1.1740409.
45. Anandakrishnan, R.; Drozdetski, A.; Walker, R.C.; Onufriev, A.V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophysical Journal* **2015**, *108*, 1153–1164. doi:10.1016/j.bpj.2014.12.047.
46. Wilson, L.; Krasny, R.; Luchko, T. Accelerating the 3D reference interaction site model theory of molecular solvation with treecode summation and cut-offs. *Journal of Computational Chemistry* **2022**, *43*, 1251–1270. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.26889>.
47. Miyata, T.; Hikasa, Y. Sigma enlarging bridge correction of three dimensional Ornstein–Zernike theory for solvation free energy of polyatomic solutes immersed in Lennard–Jones monatomic solvent. *AIP Advances* **2022**, *12*, 085206. Publisher: American Institute of Physics, doi:10.1063/5.0102003.
48. 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. *Journal of Chemical Theory and Computation* **2009**, *5*, 350–358. doi:10.1021/ct800409d.
49. Riquelme, M.; Lara, A.; Mobley, D.L.; Verstraelen, T.; Matamala, A.R.; Vöhringer-Martinez, E. Hydration Free Energies in the FreeSolv Database Calculated with Polarized Iterative Hirshfeld Charges. *Journal of Chemical Information and Modeling* **2018**, *58*, 1779–1797. Publisher: American Chemical Society, doi:10.1021/acs.jcim.8b00180.
50. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174. doi:10.1002/jcc.20035.
51. Mobley, D.L.; Shirts, M.; Lim, N.; Chodera, J.; Beauchamp, K.; Lee-Ping. MobleyLab/FreeSolv: Version 0.52, 2018. doi:10.5281/zenodo.1161245.
52. Jakalian, A.; Bush, B.L.; Jack, D.B.; Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry* **2000**, *21*, 132–146. doi:10.1002/(SICI)1096-987X(20000130)21:2<132::AID-JCC5>3.0.CO;2-P.
53. Jakalian, A.; Jack, D.B.; Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **2002**, *23*, 1623–1641. doi:10.1002/jcc.10128.
54. Onufriev, A.; Bashford, D.; Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 383–394. doi:10.1002/prot.20033.
55. Weiser, J.; Shenkin, P.S.; Still, W.C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry* **1999**, *20*, 217–230. doi:10.1002/(SICI)1096-987X(19990130)20:2<217::AID-JCC4>3.0.CO;2-A.
56. Case, D.A.; Cerutti, D.S.; Cheatham, T.E.; III.; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Greene, D.; et al. AMBER 2017, 2017.

57. Shirts, M.R.; Chodera, J.D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*, 124105. doi:10.1063/1.2978177. 479
58. Shirts, M.; Beauchamp, K.; Naden, L.; Chodera, J.; Rodríguez-Guerra, J.; Martiniani, S.; Stern, C.; Henry, M.; Fass, J.; Gowers, R.; et al. choderalab/pymbar: 3.1.1, 2022. doi:10.5281/zenodo.7383197. 480
59. Case, D.A.; Aktulga, H.M.; Belfon, K.; Ben-Shalom, I.Y.; Brozell, S.R.; Cerutti, D.S.; Cheatham, T.E.I.; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; et al. *Amber 2021*; University of California, San Francisco, 2021. 481
60. Berendsen, H.J.C.; Grigera, J.R.; Straatsma, T.P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* **1987**, *91*, 6269–6271. doi:10.1021/j100308a038. 482
61. Perkyns, J.; Pettitt, B.M. A site–site theory for finite concentration saline solutions. *The Journal of chemical physics* **1992**, *97*, 7656–7666. 483
62. Perkyns, J.S.; Montgomery Pettitt, B. A dielectrically consistent interaction site theory for solvent–electrolyte mixtures. *Chemical Physics Letters* **1992**, *190*, 626–630. doi:10.1016/0009-2614(92)85201-K. 484
63. Kovalenko, A.; Ten-no, S.; Hirata, F. Solution of three-dimensional reference interaction site model and hypernetted chain equations for simple point charge water by modified method of direct inversion in iterative subspace. *Journal of Computational Chemistry* **1999**, *20*, 928–936. doi:10.1002/(SICI)1096-987X(19990715)20:9<928::AID-JCC4>3.0.CO;2-X. 485
64. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python; , 2010; pp. 92–96. doi:10.25080/Majora-92bf1922-011. 486
65. McKinney, W. Data Structures for Statistical Computing in Python; , 2010; pp. 56–61. doi:10.25080/Majora-92bf1922-00a. 487