JCTC Journal of Chemical Theory and Computation

Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations

David L. Mobley,*,[†] Christopher I. Bayly,[§] Matthew D. Cooper,[§] Michael R. Shirts,[ll] and Ken A. Dill[‡]

Department of Chemistry, University of New Orleans, New Orleans, Louisiana 70148
Department of Pharmaceutical Chemistry, University of California at San Francisco,
San Francisco, California 94158 Merck-Frosst Canada Ltd., 16711 TransCanada
Highway, Kirkland, Quebec, Canada H9H 3L1, and Department of Chemical Engineering,
University of Virginia, P.O. Box 400741, Charlottesville, Virginia 22904-4741

Abstract: Using molecular dynamics free energy simulations with TIP3P explicit solvent, we compute the hydration free energies of 504 neutral small organic molecules and compare them to experiments. We find, first, good general agreement between the simulations and the experiments, with an rms error of 1.24 kcal/mol over the whole set (i.e., about 2 kT) and a correlation coefficient of 0.89. Second, we use an automated procedure to identify systematic errors for some classes of compounds and suggest some improvements to the force field. We find that alkyne hydration free energies are particularly poorly predicted due to problems with a Lennard-Jones well depth and find that an alternate choice for this well depth largely rectifies the situation. Third, we study the nonpolar component of hydration free energies—that is, the part that is not due to electrostatics. While we find that repulsive and attractive components of the nonpolar part both scale roughly with surface area (or volume) of the solute, the total nonpolar free energy does not scale with the solute surface area or volume, because it is a small difference between large components and is dominated by the deviations from the trend. While the methods used here are not new, this is a more extensive test than previous explicit solvent studies, and the size of the test set allows identification of systematic problems with force field parameters for particular classes of compounds. We believe that the computed free energies and components will be valuable to others in the future development of force fields and solvation models.

I. Introduction

Aqueous solvation (hydration) of molecules is important for much of chemistry and biochemistry. Many experimental hydration free energies are available, providing a wonderful opportunity for testing force fields and computational treatments of solvation.

There have been a number of extensive tests of hydration free energies computed using continuum representations of water and static solute conformations.[1−4] One recent study extended this by sampling ensembles of solute conformations using classical molecular dynamics and using these to compute hydration free energies.[5] Continuum representations of solvent, however, have known limitations,[6,7] and explicit treatment of solvent provides a "gold standard" for molecular simulations. Early explicit solvent hydration free energy studies were limited by computational cost to a few compounds and, more recently, by the availability of

---

* Corresponding author e-mail: dmobley@gmail.com.
† University of New Orleans.
§ Merck-Frosst Canada Ltd.
ll University of Virginia.
‡ University of California at San Francisco.

Small Molecule Hydration Free Energies

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **351**

parameters for small molecules. Thus a variety of studies have looked at hydration free energies of amino acid side chain analogs in explicit solvent (for example, refs 8−11), but few have studied a more diverse set.

With recent computational and methodological developments, both of these hurdles—computational cost and parameters—are now at least partially surmountable. Hydration free energy calculations can now be conducted more efficiently,[8,12] and computers are faster. Recent developments also make possible semiautomatic parametrization of small molecules, using general small molecule force fields like the general Amber force field (GAFF)[13] and parameter assignment tools like Antechamber.[14] Thus, two recent studies have examined hydration free energies of a total of roughly 60 small molecules in explicit solvent.[4,12]

Here, we perform a much more extensive test of explicit solvent modeling, on a test set of 504 molecules previously used for implicit solvent hydration free energy calculations[5]—more than 10 times larger than the largest previous explicit solvent tests.[12] Because this test is so extensive, we believe it provides a good benchmark for the best results that can currently be expected from molecular dynamics models of hydration. We also hope that others will find this compilation of computational and experimental results useful for analysis and force field parametrization efforts.

II. Simulation Methods

A. General Simulation Parameters. In this work, we use alchemical free energy calculations to compute hydration free energies in explicit solvent for 504 small molecules, using the compound set from a previous implicit solvent study.[5] Simulation protocols were similar to those used in previous explicit solvent studies.[4,12] Hydration free energies were computed using the Bennett acceptance ratio (BAR).[15] A brief summary of the methods follows, and we note the deviations from the previous studies.[4,12]

Here, starting molecular conformations were the same as those for the previous implicit solvent study,[5] except that here a single starting conformation was used for each molecule (rather than 5) due to computational limitations relating to the size of the set. Simulations were performed in GROMACS 3.3.1[16,17] using the GAFF small molecule parameters[13] as assigned by Antechamber[14] (as in the implicit solvent study).[5] Here, AM1-BCC[18,19] partial charges were assigned using the Merck-Frosst implementation of AM1-BCC.

This data set contains several nitro-containing compounds which did not have improper torsions for the nitro-ring system in the GAFF parameter set, specifically improper torsions for GAFF types ca-o-no-o and c3-o-no-o. We added these using generic GAFF values (that is, the values used for the majority of the improper torsions in GAFF)—a barrier height of 2.2 kcal/mol, a phase shift of 180°, and a periodicity of 2.

After setup in Antechamber and Leap, small molecule parameters were converted to GROMACS topology and coordinate files using a Perl conversion script developed previously.[20] Small molecules were then solvated using GROMACS utilities in a dodecahedral water box with at least 1.2 nm from the solute to the nearest box edge using the TIP3P model of water.[21] The number of water molecules varied depending on the solute size. Simulations were performed separately at a variety of different alchemical intermediate λ values, with the number of λ values and the amount of equilibration as described previously.[12] Production simulations were 5 ns in length at each λ value, and free energies and uncertainties were computed as described previously.[4,12] Uncertainties were computed using the block bootstrap procedure described previously. Cutoffs and simulation parameters were as described previously except that the real-space electrostatic cutoff was 10 Å rather than 9 Å.

We computed the electrostatic and nonpolar components of solvation. The electrostatic component was computed as the free energy of turning on the solute partial charges in water, less the free energy of the same transformation in vacuum. The nonpolar component was the free energy of turning on the Lennard-Jones interactions between the uncharged solute and water, as in previous studies.[4,12] Alternative definitions of the nonpolar component are possible.[44]

B. Analysis of the Nonpolar Component. In implicit solvent models, the nonpolar component of solvation is often assumed to correlate with the surface area and/or the volume based on theoretical arguments relating to cavity creation cost.[22−26] To explore this we computed the solvent accessible surface area and volume for all of the solutes considered here using GROMACS tool g_sas with a probe radius of 1.4 nm.

We also further dissected the nonpolar part (due the Lennard-Jones interactions) into repulsive and attractive components using the Weeks-Chandler-Andersen (WCA) separation.[27] To do this, we implemented the WCA separation in a modified version of GROMACS 3.3.1.[45]

In our main study, we simply computed the total nonpolar component and retained the trajectories. The attractive component for each solute was then obtained by applying the WCA separation to stored trajectories of the fully interacting solute and reprocessing these simulations with the attractive interactions turned off to re-evaluate the energies. We computed the free energy for turning off the attractive interactions using exponential averaging (the Zwanzig relation)[28] and standard error analysis. This assumes that phase-space overlap is good between the ensemble where the solute has attractive interactions with water and that where it does not. Error analysis should tell us if this is not the case. We further tested this by recomputing the attractive contribution using simulations at a series of separate λ values (where λ modifies only the attractive interactions) for selected solutes (phenol, *p*-xylene, pyridine, and toluene) and found that computed free energies were within uncertainty of the values computed using exponential averaging, indicating overlap was sufficient.

With these attractive components, we then obtained repulsive components by subtracting the attractive component from the total nonpolar component. This probably results in slightly larger uncertainties in computed repulsive compo-

nents than would have resulted from computing the repulsive component separately, but it also saves a large amount of computer time since we had already computed the total nonpolar component, and the repulsive portion of the calculation is the most difficult to converge.

C. Identification of Systematic Errors. Some functional groups may lead to systematic errors, resulting in errors which are larger for some types of molecules than for other types. Alternatively, there might be no systematic errors. We seek an approach to easily identify systematic errors and prioritize functional groups which have the largest errors.

We make a list of compounds and sort it by rms error, from largest error to smallest error. Following a method that is often used to determine *enrichment factors* for drug discovery, we look at the cumulative distribution function (CDF) for each functional group—the probability of compounds with that functional group having a ranked rms error up to rank x. Those functional groups that are systematically wrong will tend to cluster at high rms error and will result in a rapid rise in the CDF versus x. This can be assessed easily by computing the area under the CDF, biased by a weighting function to give the most weight to high rms errors. Here, we do this using the recently developed BEDROC metric,[29] which evaluates the integral of the CDF multiplied by an exponentially decaying weighting factor and then rescales this to run from 0 to 1. Chemical groups which occur most often in compounds with high rms errors will have larger BEDROC values, while chemical groups which have more random errors will have smaller BEDROC values (the expected BEDROC value for a uniform distribution can be computed analytically).[29] Chemical groups that only occur in compounds with low rms errors have the smallest BEDROC values. In Section III we report BEDROC values for a variety of chemical groups and atom types. Uncertainties were computed using the standard deviation of the mean for 40 iterations of a bootstrap procedure where BEDROC values for each chemical group are recomputed using a new list of compounds made up of a random selection of compounds from the original list.

Here, BEDROC values were computed using a weighting factor of $\alpha = 1.0$. This value was obtained empirically by experimenting with different α values to see what gave the best ability to recognize functional groups which differ substantially from random. If α is too large, the weighting is too strong, and only compounds at the very highest rms errors matter. If α is too small, making BEDROC equivalent to the ROC metric, the weighting of the early part of the curve is too weak, also apparently reducing the ability to recognize systematic errors. $\alpha = 1.0$ was a good compromise.

To avoid having to assign functional groups to all of the compounds in the test set by hand, we used the program Checkmol,[30] which automatically assigns chemical groups to molecules. We used MDL molfiles generated by OpenEye's OEChem toolkit as input. This resulted in an extremely large set of chemical groups, so we retained only those chemical groups which occurred in at least 5 molecules. We also combined some small groups. For example, we made a single group of amines, containing all types of amines. We also did the same for amides, ethers, esters, thiols, acids, and
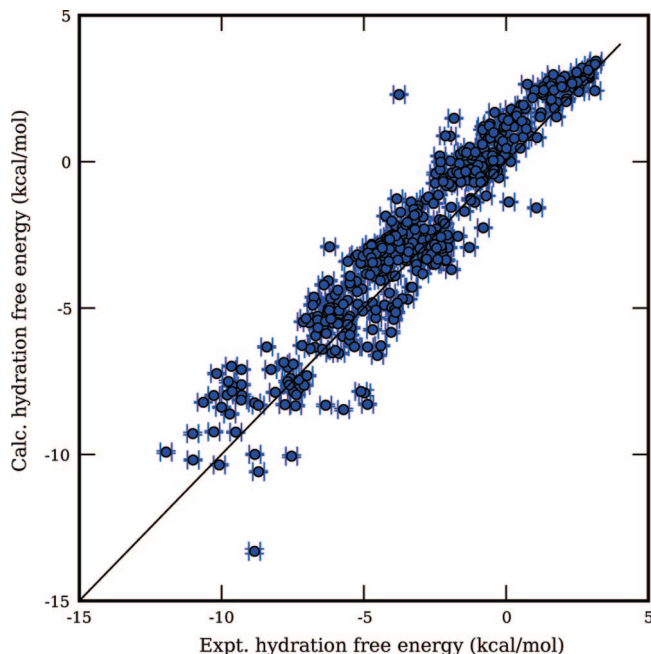

Figure 1. Calculated hydration free energies versus experiment. Shown are the calculated hydration free energies versus experiment for the full test set. The diagonal line is $x = y$. Vertical error bars denote computed uncertainties, and horizontal error bars are a conservative estimate.

alcohols. We also manually created a "hypervalent S" group and included the appropriate compounds in this group. The resulting list of molecules assigned to chemical groups was used to generate BEDROC values for these chemical groups.

We also tried using Student's *t*-test to look for systematic errors to supplement the BEDROC approach. We used our own implementations of the *t*-test and SciPy's implementation of the incomplete beta function for computing the significance. Results from this are discussed below.

III. Results and Discussion

A. The Mean Error Relative to Experiment Is Less than 1 kcal/mol. Here we evaluate the agreement between computed hydration free energies and the experimental values for the full test set. A previous study on the same 504 small-molecule test set compared the accuracy of several different implicit solvent models[5] using molecular dynamics free energy calculations. rms errors ranged from 2.014 ± 0.008 kcal/mol to 2.433 ± 0.002 kcal/mol depending on the implicit solvent model, with correlation coefficients (r^2) from 0.685 ± 0.001 to 0.774 ± 0.001. In all four solvent models tested, the computed hydration free energies were systematically too negative relative to experiments (the solutes preferred the water phase too much in the simulations), so the mean error was negative (-0.65 ± 0.09 to -1.1 ± 0.1).

Here, using explicit TIP3P water, we find an rms error of 1.26 ± 0.01 kcal/mol, with a correlation coefficient of 0.889 ± 0.006 and a mean error of 0.676 ± 0.002 (Figure 1). Hence, on average, explicit solvent simulations give significantly better agreement with experiments than our earlier implicit solvent study, consistent with an earlier comparison.[4] Interestingly, the systematic errors of explicit and implicit

Small Molecule Hydration Free Energies

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **353**

solvent studies are in opposite directions. In explicit solvent, the hydration free energies here are systematically too positive. These differences are likely due to the solvent models rather than the force field parameters, since the solute parameters are very similar in the two cases. Systematic errors in other explicit solvent models tended to be in the same direction as the explicit solvent deviation here,[8] so perhaps limitations of the water model are playing a role. Another potential source of such differences is the neglect, in implicit solvent models, of asymmetries in the response of water to solutes of different polarities.[7] Another origin of differences could be the nonpolar term in the implicit models. That is, the term $\gamma \times A$ (where A is the surface area) in implicit solvent models involves an adjustable parameter which can change the errors. A table of the full results from this study is available in the Supporting Information.

B. Improving the Alkyne Lennard-Jones Parameters and Identifying Other Systematic Errors. Are there systematic errors in the force field parameters for molecules in our test set? We found that the computed hydration free energies for alkynes were systematically much too positive (Figure 1 and the Supporting Information). There were six alkynes in the set, and the mean error was 1.92 ± 0.21 kcal/mol. All of the computed hydration free energies were actually around 2 kcal/mol, while experimental values are close to zero. For all of the alkynes, the electrostatic component of hydration is quite small (-0.8 to -0.9 kcal/mol), since these molecules are largely nonpolar. We reasoned that errors in alkyne parameters are thus not likely to be in the electrostatic terms, nor are the errors expected to come from the bonded parameters (bond stretching, angle bending, etc.), which should affect hydration free energies only weakly. Hence, we focused on the alkyne Lennard-Jones parameters. In GAFF, the alkyne carbon Lennard-Jones parameters are identical to those for all carbons except selected sp2 carbons (the 'c2' atom type) and are taken directly from comparable carbons in older AMBER force fields.[13] We were particularly concerned about the parameters for the GAFF "c1" atom type, for the triple bonded carbons in alkynes. These apparently originated with the work of Howard et al., where they "were obtained by analogy to the Weiner et al. and Cornell et al. force fields".[31] In that work, those Lennard-Jones parameters were taken to be the same as for the other carbons.

Many AMBER Lennard-Jones parameters were originally taken from the OPLS force field, so we examined the OPLS choices for triple bonded carbons. It turned out that OPLS uses several different atom types for alkyne carbons, originating from simulations of linear and substituted alkynes,[32-34] and some of these have much stronger dispersion interactions than those for the GAFF c1 type, which is intuitively reasonable. It seemed likely that missing dispersion interactions could account for at least part of the error we were seeing for alkynes, thus we examined modifying the Lennard-Jones well-depth for alkynes in GAFF.

We sought to avoid adding additional atom types to GAFF, but OPLS has several different carbon well-depths for alkynes, depending on whether the carbon is terminal ($\varepsilon = 0.086$ kcal/mol), nonterminal with an attached atom having

two or three hydrogens ($\varepsilon = 0.210$ kcal/mol), nonterminal with an attached atom having one hydrogen ($\varepsilon = 0.135$ kcal/mol), or nonterminal with an attached phenyl or other atom having no hydrogens ($\varepsilon = 0.100$ kcal/mol).[32-34] To avoid adding additional atom types to GAFF, we needed to pick just one of these, so we chose the one which gave the most accurate hydration free energies when used for all alkyne triple bonded carbons. This was $\varepsilon = 0.210$ kcal/mol. The original GAFF well depth was $\varepsilon = 0.086$ kcal/mol.

Using this new ε value for triple-bonded carbons, the computed hydration free energies for alkynes are much closer to zero (although still slightly positive); now the mean error is 0.49 ± 0.07 kcal/mol, down from 1.92 ± 0.21 kcal/mol initially. Increasing the well depth further could reduce this somewhat more, but this might cause other inconsistencies within the force field. Nevertheless, the systematic error here on alkynes is compelling, and we recommend that future GAFF studies use a well depth of $\varepsilon = 0.210$ kcal/mol for triple bonded carbons (GAFF types c1, cg, and ch).[46]

The alkynes also provide an example of how the BEDROC metric works for identifying systematic errors. Before the adjustment of the well depth for alkynes, the BEDROC value (with $\alpha = 1$) for alkynes was 0.90 ± 0.02 (compared to 0.49 for a random distribution with this α),[47] indicating that alkynes were systematically wrong. After the fix, the BEDROC value was 0.26 ± 0.05, indicating that alkynes now actually are considerably better than other typical compounds (Figure 2). Although our correction of ε was done without regard for the carbonitriles, the change results in a decrease in BEDROC for the carbonitriles from 0.86 ± 0.05 to 0.73 ± 0.06 (compared to 0.49 for uniform). So carbonitriles are now improved too but still have substantial systematic errors. With this change, the overall rms error decreases slightly to 1.24 ± 0.01 kcal/mol, and the correlation coefficient remains essentially the same (0.891 ± 0.006). In all that follows we report values computed with the new well depth.

We believe that the approach utilized here (looking for compounds that are over-represented at the highest rms errors) is a general and useful strategy for identifying systematic flaws in the energy functions used for molecular modeling simulations and prioritize reparameterization efforts. Functional groups which tend to cause significant errors should occur frequently at the high-rms error end of the set, while functional groups which are not necessarily linked to the errors should be roughly randomly distributed over the test set. For example, one would intuitively expect that whether a compound is aromatic or not will have little to do with whether it is systematically mispredicted. Indeed, aromatic compounds have a BEDROC value of 0.48 ± 0.03, roughly randomly distributed (Figure 2). BEDROC values by functional group for the set are shown in Table 1. These BEDROC values show that cyclic hydrocarbons, alkynes (with the fix), alkanes, aldehydes, and ketones are now particularly well predicted. On the other hand, there appear to be systematic errors for alcohols, alkyl bromides, and carbonitriles.

We also tried another approach for identifying systematic errors involving Student's t-test, which compares the means
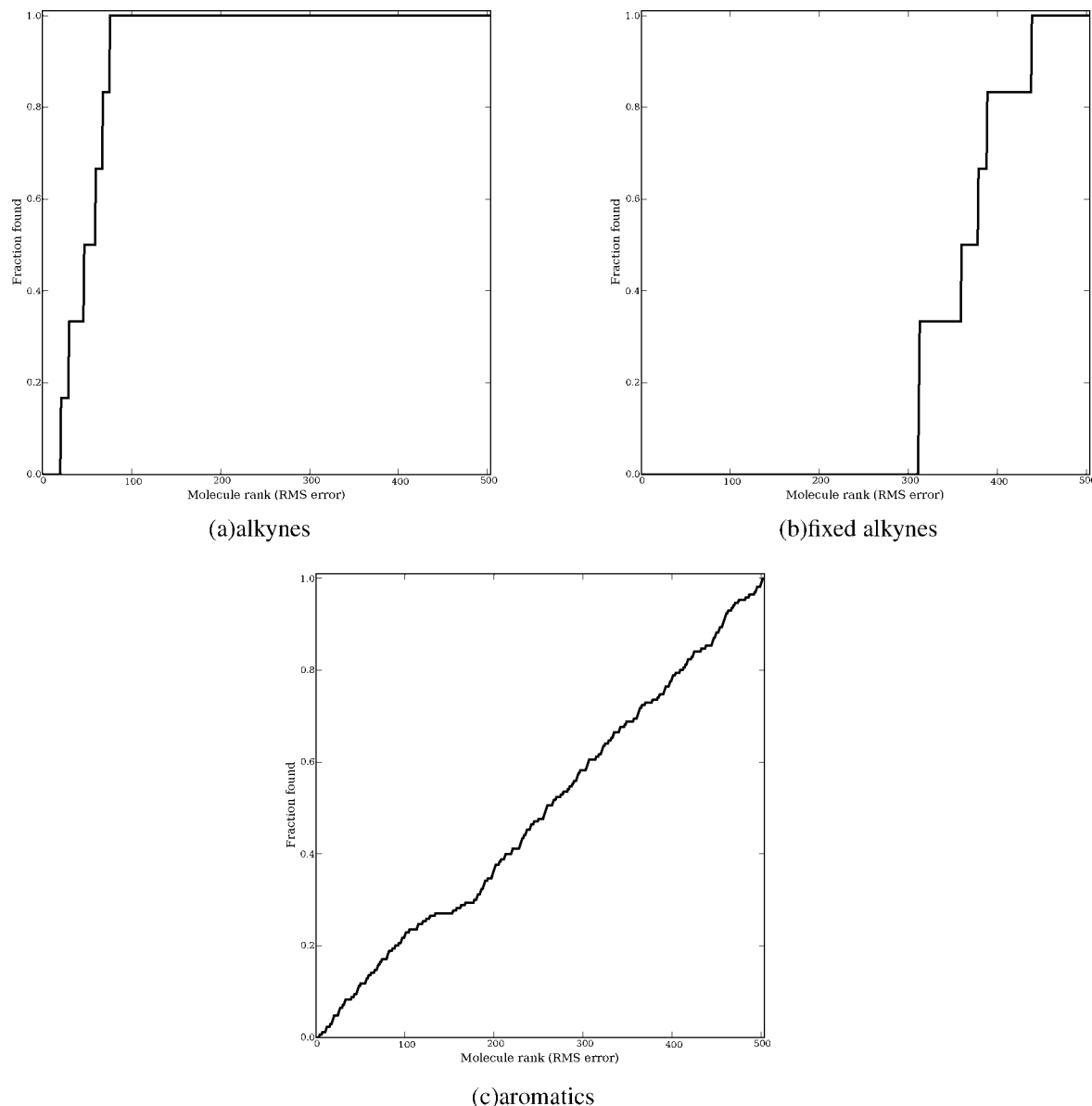
(a)alkynes

(b)fixed alkynes

(c)aromatics

Figure 2. CDFs for selected functional groups versus error. Shown are cumulative distribution functions for finding compounds with particular functional groups at a given ranked error. Compounds found far to the left have very large errors; compounds far to the right have very small errors. An ideal random distribution of errors would give rise to a linear rise in the CDF. CDFs are shown for (a) alkynes before fixing the Lennard-Jones well-depth, (b) alkynes after fixing the Lennard-Jones well-depth, and (c) aromatics.

of two distributions and provides a measure of the significance of any difference in the means. We applied this approach in two different ways:

(1) We compared the mean experimental value for each functional group with the mean calculated value for each functional group (Supporting Information, Table 5). This proved not to be particularly useful, as these means are significantly different for almost every functional group. This is not surprising given the fact that the mean error across the entire test set is 0.676 ± 0.002, so most computed values (in all functional groups) are too positive. This does show that results could be improved across the entire set by addressing this systematic offset, but it does not provide any insight into which functional groups are particularly problematic.

(2) We compared the error for the compounds in each functional group with the error for the entire set (Table 2). This shows which functional groups have a significantly *different* performance than the overall set, though this performance could be better or worse. We also show the mean error for each functional group in Table 2; functional groups with mean errors around 0.676 kcal/mol are typical, while those with larger mean errors are worse than average, and those with smaller mean errors are better than average. The *t*-test tells us which of these differences are significant, and many are. This appears to be a useful analysis that complements the BEDROC analysis. The advantage of the BEDROC analysis is that it tells us which functional groups have the worst errors, while this analysis can tell us which functional groups have the most significant errors.

Small Molecule Hydration Free Energies

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **355**

Table 1. BEDROC Values by Functional Group for the Different Functional Groups Represented in the Test Set, Compared to What Would Be Expected for the Same Number of Compounds Distributed Randomly Across the Test Set[a]

functional group	number	BEDROC
acid	73	0.48 ± 0.03
alcohol	38	0.76 ± 0.03
aldehyde	20	0.22 ± 0.04
alkanes	28	0.16 ± 0.03
alkene	35	0.55 ± 0.04
alkyl bromide	17	0.72 ± 0.08
alkyl chloride	31	0.61 ± 0.05
alkyl iodide	9	0.44 ± 0.06
alkyne	6	0.26 ± 0.04
amine	44	0.47 ± 0.04
aromatic compound	170	0.48 ± 0.03
aryl chloride	20	0.54 ± 0.05
carbonitrile	12	0.73 ± 0.07
cyclic hydrocarbon	8	0.14 ± 0.03
ester	8	0.46 ± 0.11
ether	42	0.60 ± 0.04
halogen derivative	22	0.58 ± 0.07
heterocyclic compound	48	0.60 ± 0.04
hypervalent S	5	0.62 ± 0.20
ketone	25	0.26 ± 0.06
nitro compound	17	0.63 ± 0.08
other	29	0.62 ± 0.06
phenol or hydroxyhetarene	33	0.60 ± 0.05
thiol	5	0.46 ± 0.04

[a] Functional groups with high BEDROC values (relative to the value for random, roughly 0.5 here) are overrepresented in compounds with high RMS errors.

Table 2. Statistics from Applying Student's *t*-Test to the Difference between the Calculated and Experimental Means by Functional Group[a]

functional group	number	t-value	significance	mean error
acid	73	−7.43	4e-13	−0.34
alcohol	38	3.62	0.0003	1.29
aldehyde	20	−3.04	0.003	−0.07
alkanes	28	−1.69	0.09	0.31
alkene	35	2.34	0.02	1.07
alkyl bromide	17	3.31	0.001	1.50
alkyl chloride	31	2.31	0.02	1.09
alkyl iodide	9	0.59	0.6	0.86
alkyne	6	−0.38	0.7	0.49
amine	44	−0.65	0.5	0.55
aromatic compound	170	−1.05	0.3	0.55
aryl chloride	20	1.65	0.1	1.04
carbonitrile	12	3.22	0.001	1.63
cyclic hydrocarbon	8	−1.18	0.2	0.21
ester	8	−1.69	0.09	0.02
ether	42	2.18	0.03	1.01
halogen derivative	22	0.32	0.8	0.73
heterocyclic compound	48	2.38	0.02	1.02
hypervalent S	5	−4.55	7e-06	−1.50
ketone	25	−2.77	0.006	0.05
nitro compound	17	1.86	0.06	1.13
other	29	−0.48	0.6	0.55
phenol or hydroxyhetarene	33	2.72	0.007	1.16
thiol	5	0.51	0.6	0.89

[a] Shown are the number of compounds in each functional group, the calculated *t* value, the computed significance (probability that *t* could be this large or larger by chance), and the mean error for this group (in kcal/mol). The overall mean error is 0.676 ± 0.002 kcal/mol, so groups with mean errors smaller than this may be significantly better than average (until the mean error becomes negative), while those with mean errors larger than this may be significantly worse.

The study done here uses one particular charge model. Charge model may affect which compounds are particularly poorly predicted, though in two recent tests, the compounds which were poorly predicted tended to be poorly predicted by most charge models.[4,35] Still, our analysis here does not in general point to a specific source of error. Errors may be due to the charge model, Lennard-Jones parameters, or bonded parameters, some combination, or even due to the water model. In the case of the alkynes, we can be fairly confident that the source of error is the Lennard-Jones parameters for the reasons noted above. But for the other cases noted here, further work will be required to determine the source of error.

C. The Total Nonpolar Component Does Not Correlate with Surface Area or Volume. We examined the nonpolar components (the nonelectrostatic component of the hydration free energy) for our data set. The total nonpolar contribution to the solvation free energy is typically assumed to correlate with surface area or volume in implicit solvent models. Yet we find that there is essentially no correlation. Plots of nonpolar components versus surface area and volume are shown in Figure 3. The correlation of the nonpolar component with surface area is $r^2 = 0.019 \pm 0.001$, and that with volume is $r^2 = 0.011 \pm 0.001$. The molecules in this test set are small enough that surface area and volume are highly correlated ($r^2 = 0.991 \pm 0.001$).

We further dissect the nonpolar component using the WCA separation of the Lennard-Jones potential energy (and thus the nonpolar component) into attractive and repulsive parts. The potential is split based on the sign of the force, as discussed in the Methods section. We find that both the

attractive and repulsive components individually correlate strongly with surface area and volume (repulsive: $r^2 = 0.964 \pm 0.002$ with surface area, $r^2 = 0.952 \pm 0.002$ with volume; attractive: $r^2 = 0.945 \pm 0.002$ with surface area, $r^2 = 0.946 \pm 0.002$ with volume; Figure 4), and it is only the total (the small difference of the two large individual components) that does not correlate well with surface area or volume. This is in accord with previous work on a smaller set of compounds.[36] Essentially, the total nonpolar component is the sum of two anticorrelated quantities, and so the total ends up being dominated by the scatter in these quantities. It is interesting to note that the minimum in the Lennard-Jones potential is precisely where these two forces, the attractive and repulsive components, are very well balanced, so it is perhaps not surprising that the attractive and repulsive components correlate so well.

The observed poor correlation, and the importance of attractive interactions, is consistent with several previous studies which have found that the nonpolar component of solvation does not correlate well with surface area.[36−39]

Why is the correlation with surface area so poor? In Figure 3, it is apparent that compounds containing only carbon and hydrogen have a nonpolar component that is less favorable to solvation than molecules of an equal size which additionally contain nitrogen and/or oxygen. The likely reason for this is that nitrogen and oxygen atoms tend to have stronger attractive dispersion interactions with their environment than
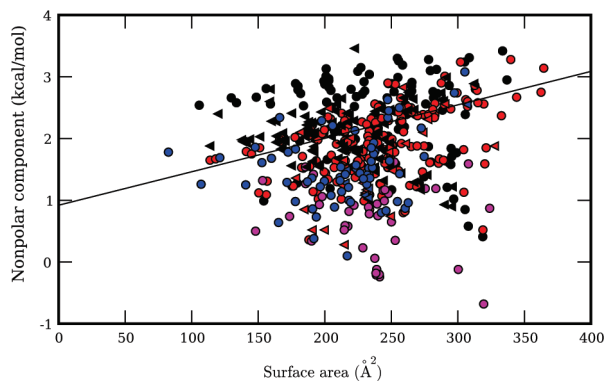
(a)Nonpolar component versus surface area
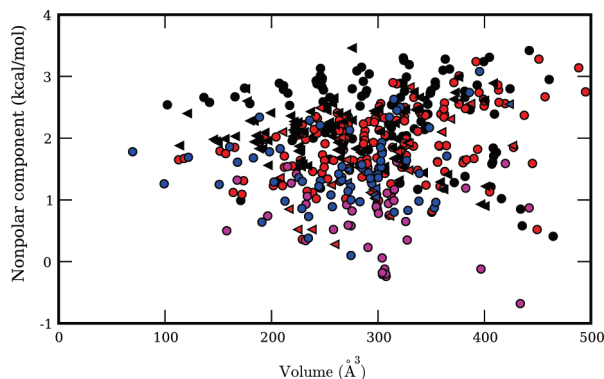
(b)Nonpolar component versus volume
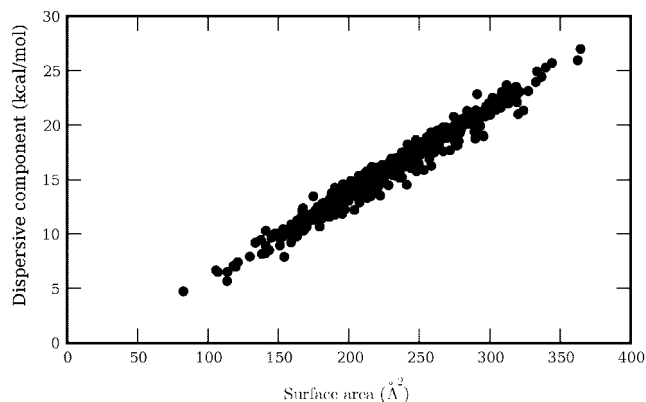
Figure 3. Nonpolar components versus solvent accessible surface area and volume. Shown are the calculated nonpolar component of the hydration free energies versus solvent accessible surface area and volume for the compounds in the set. Carbon and hydrogen containing compounds are black, those with oxygen additionally are red, those with nitrogen additionally are blue, and those with nitrogen and oxygen both are magenta. Compounds with diamond symbols contain other elements in addition to C, H, N, and O. In the surface area plot, the line is a typical implicit solvent nonpolar component estimate of $G_{np} = (0.00542 \cdot SA + 0.92)$ kcal/mol[1].

(a)Repulsive part of nonpolar component versus surface area

(b)Attractive part of nonpolar component versus surface area

Figure 4. Repulsive and attractive parts of the nonpolar component versus surface area. Shown are the repulsive (a) and attractive (b) parts of the nonpolar component, as calculated using the WCA separation, plotted versus the solvent accessible surface area for solutes in the test set. Similar plots comparing the repulsive and attractive components to volume are given in the Supporting Information.

do carbon and hydrogen. Several other studies have noted that dispersion interactions play an important role in nonpolar solvation.[36−39] Even interior solute atoms contribute to these attractive interactions in an important way.[40] Other factors may also contribute to the poor correlation with surface area. For example, geometric effects may play an important role as well.
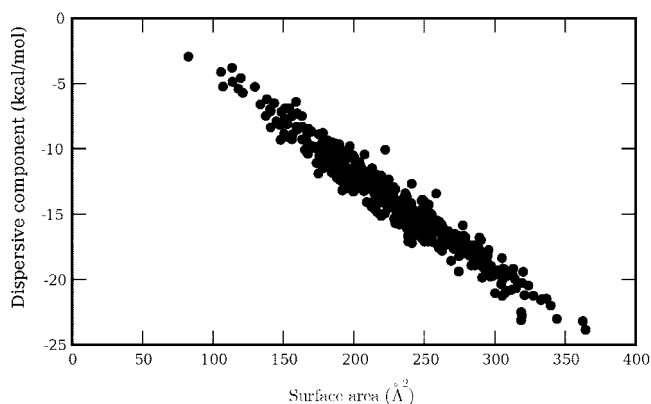
Overall, our results strongly support the growing consensus that implicit solvent models should move beyond the simple surface area model for treatment of the nonpolar component, perhaps at least to include a treatment of dispersion interactions. A number of alternate models have already been proposed.[36,38,39,41,42]

IV. Conclusions

We used molecular dynamics simulations in explicit TIP3P water to compute the hydration free energies for a set of 504 neutral compounds. We compared the results with experimental data in the most extensive such test in explicit solvent to date. We find a good correlation (r^2 of 0.891 ±

0.06) and an rms error of 1.24 ± 0.01 kcal/mol or roughly 2 kT. We believe this is representative of the accuracy that can be expected from the best current physical models for hydration free energies. It may be possible to develop new models which can do somewhat better, though we expect that it may be very hard to increase accuracies past 1 kT. A key finding is that these explicit solvent free energies are considerably more accurate than the corresponding implicit solvent values for the same data set.

At the same time, many of the molecules in this test set are relatively small and simple compared to typical druglike molecules, which may be highly polyfunctional. Recent work suggests that overall performance of the approach applied here may be significantly worse in tests where the compounds involved are more polar and polyfunctional.[4,35] This may suggest we need much more hydration free energy data on more polyfunctional, druglike molecules in order to refine our force fields.

Here, we also propose a way to identify systematic errors in force field parameters for particular functional groups. We do this using the BEDROC method.[29] Using this approach,

Small Molecule Hydration Free Energies

J. Chem. Theory Comput., Vol. 5, No. 2, 2009 **357**

we were able to fix a systematic problem with alkyne Lennard-Jones parameters. We also identified several other classes of compounds which appear to have systematic errors, and for which further force field development should be done. Having a method to systematically identify problematic compound classes provides good opportunities for force field improvements.

In addition, we studied the nonpolar component of the hydration free energy for the compounds in the test set. We find that while the large repulsion and attraction terms both correlate well with the size (area or volume) of the solute, the total nonpolar component, which is a small difference between these two quantities, does not. This strongly suggests that implicit solvent models need to move away from treating the nonpolar component as simply dependent on the surface area. The data additionally suggest that new models must address the nonlinear behavior arising from the delicate balance of repulsive and attractive nonpolar terms. Furthermore, implicit solvent models that have been parametrized to match experimental hydration free energies using a simple surface area-based nonpolar term may need to be reparameterized.

Here, the real value of this study is not the methods presented—the methods were used in previous work. Rather, it is the extensive nature of the test, which provides the opportunity to actually identify systematic errors in the force field descriptions of particular functional groups. It also provides guidance into what compounds are likely to be particularly challenging to study computationally with current force fields.

Because we believe the real value of this study is these results, we have deposited the full set of computed free energies, components, starting molecular structures, and parameters for this work in the Supporting Information. We hope that others find this experimental data set and the computational results to be useful in future studies of solvation and for force field development.

Supporting Information Available: Coordinate files (mol2) with AM1-BCC partial charges for the small molecules in the test set used here; computed hydration free energies, electrostatic and nonpolar components, and the experimental values; AMBER parameter and coordinate files for all of the molecules in the test set; plots of attractive and repulsive components versus solute volume; a table mapping the names used for the files to IUPAC names; a table of computed solvent accessible surface area and volume for each solute; and results from Student's *t*-test comparing the mean experimental and calculated values for each functional group.This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(2) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A *J. Phys. Chem. B* **2002**, *106*, 11009–11015.

(3) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532–6542.

(4) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769–778.

(5) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2008**, *111*, 938–946.

(6) Chorny, I.; Dill, K. A.; Jacobson, M. P. *J. Phys. Chem. B* **2005**, *109*, 24056–24060.

(7) Mobley, D. L.; Barber, A. E., II; Fennell, C. J.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 2405–2414.

(8) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(9) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616–17626.

(10) Deng, Y.; Roux, B. *J. Chem. Phys.* **2004**, *108*, 16567–16576.

(11) Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002**, *23*, 548–553.

(12) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242–2254.

(13) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(14) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *26*, 247260.

(15) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(16) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(17) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(18) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.

(19) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(20) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys.* **2006**, *125*, 084902.

(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(22) Stillinger, F. H. *J. Solution Chem.* **1973**, *2*, 141–158.

(23) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717–726.

(24) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.

(25) Chothia, C. *Nature* **1974**, *248*, 338.

(26) Reynolds, J. A.; Gilbert, D. B.; Tanford, C. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2925.

(27) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237–5247.

(28) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(29) Truchon, J.-F.; Bayly, C. I. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.

(30) Haider, N. Checkmol. http://merian.pch.univie.ac.at/ nhaider/cheminf/cmmm.html (accessed July 20, 2007).

(31) Howard, A. E.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 243–261.

(32) Jorgensen, W. L., personal communication, 2007.

(33) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665–6670.

(34) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689–1700.

(35) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. *J. Phys. Chem. B* Accepted for publication.

(36) Tan, C.; Tan, Y.-H.; Luo, R. *J. Phys. Chem. B* **2007**, *111*, 12263–12274.

(37) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *J. Am. Chem. Soc.* **1999**, *121*, 9243–9244.

(38) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271–6285.

(39) Wagoner, J. A.; Baker, N. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331–8336.

(40) Pitera, J. W.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 3163–3164.

(41) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517–529.

(42) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479–499.

(43) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761.

(44) Here, the alchemical pathway used turns off all solute partial charges, meaning that the nonpolar component is calculated without solute intramolecular electrostatic interactions. An alternative pathway[8,43] involves turning off only intermolecular electrostatic interactions between the solute and its environment, while maintaining intramolecular electrostatic interactions. While the two pathways must give equivalent hydration free energies, the breakdown into electrostatic and nonpolar components will be slightly different, since the conformational ensemble sampled during the nonpolar component of the calculation will be altered by the presence (or lack thereof) of intramolecular electrostatic interactions.

(45) Other separation schemes and probe radii are possible, as are other surface definitions, but our main conclusions here should not depend significantly on these factors, as suggested by the work of Tan et al.[36]

(46) We do not believe it is necessary to perform additional tests to examine other alkyne properties (such as pure liquid properties) before making this recommendation for two reasons. First, the AMBER force field does not typically use these properties to inform the parameterization process, so including them would be a deviation from AMBER parameterization strategies. Second, the AMBER force field and GAFF claim (in the force field files) to use the OPLS Lennard-Jones parameters for alkynes. The suggested modification simply makes this claim true and brings AMBER/GAFF back into conformity with OPLS.

(47) BEDROC values for a random distribution actually depend on the number of compounds being considered relative to the total. But here the BEDROC value for the random distribution is 0.49 for all of the sizes of our chemical groups except for aromatics, where it is 0.50. To simplify our tables, then, we simply compare all BEDROC values to 0.49.