Published in final edited form as:

J Phys Chem B. 2016 August 25; 120(33): 8102–8114. doi:10.1021/acs.jpcb.6b00399.

Hydration Properties and Solvent Effects for All-Atom Solutes in Polarizable Coarse-Grained Water

Xin Cindy Yan, Julian Tirado-Rives, and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Abstract

Due to the importance of water in chemical and biological systems, a coarse-grained representation of the solvent can greatly simplify the description of the system while retaining key thermodynamic properties of the medium. A multiscale solvation model that couples all-atom solutes and polarizable Martini coarse-grained water (AAX/CGS) is developed to reproduce free energies of hydration of organic solutes. Using Monte Carlo/free energy perturbation (MC/FEP) calculations, results from multiscale and all-atom simulations are compared. Improved accuracy is obtained with the AAX/CGS approach for hydrophobic and sulfur- or halogen-containing solutes, but larger deviations are found for polar solute molecules where hydrogen bonding is featured. Furthermore, solvent effects on conformational and tautomeric equilibria of AA solutes were investigated using AA, CG, and GB/SA solvent models. It is found that the CG solvent model can reproduce well the medium effects from experiment and AA simulations; however, the GB/SA solvent model fails in some cases. A 7 to 30-fold reduction in computational cost is found for the present AAX/CGS multiscale simulations compared to the AA alternative.

Graphical Abstract

*Corresponding Author: william.jorgensen@yale.edu.

Notes

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

Computed bulk properties of polarizable Martini CG water using Monte Carlo and molecular dynamics simulations are tabulated. Plots of computed radial distribution functions for the polarizable Martini CG water are also provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Keywords

solvent effects; free energies of hydration; polarizable Martini coarse-grained water; multiscale simulations

Introduction

Coarse-graining (CG) techniques greatly expand the spatial and temporal scales of molecular modeling by reducing full atomic information into lower resolution models.^{1–5} Nevertheless, some key interactions and structural details may be sacrificed as a result of simplifications of the CG models. Multiscale simulations strike a balance between speed and accuracy, by allowing atomistic details in a subdomain where specific interactions are vital, and a CG description for the rest of the system.^{6,7} In particular, the development of multiscale models that combine all-atom (AA) solutes and coarse-grained (CG) water is valuable given the large amount of water in typical biochemical systems. However, the coupling of AA solutes with CG solvent is non-trivial, as an adequate balance of intra- and inter-resolution interactions is required.⁸ Previously, several coupling approaches have been developed, including force matching,⁹ virtual CG site,^{10,11} and direct coupling using mixing rules for nonbonded parameters.^{12–15}

Since a faithful description of solvation is central to any reliable solvent model, it is important to evaluate an AAX/CGS multiscale model based on solvation properties. In a multiscale simulation study using virtual sites,¹¹ the potentials of mean force (PMFs) between pairs of AA butane and dialanine molecules were computed in CG and AA solvents and compared. It was found that the effect of solvent on solute association was quantitatively captured by the CG model. Another study¹⁰ featuring a direct electrostatic coupling scheme, however, discovered that PMFs of charged and polar amino acid side-chain analogues in CG solvents differ significantly from the corresponding AA results. Subsequently, Riniker *et al.*¹⁶ carried out MD simulations of four AA proteins solvated in CG water and concluded that

averaged properties of the proteins were better preserved in the multiscale simulations than in implicit solvent or vacuum. They also pointed out that the inability for CG solvent to form hydrogen bonds resulted in an artificial increase in the occurrence of side chain – side chain hydrogen bonds. In a recent study, Orsi *et al.* found the mean unsigned error (MUE) for the hydration free energies of 14 amino acid side-chain analogues to be 1.22 kcal/mol with the ELBA CG water model and a direct coupling approach.¹⁷ In addition, using the GROMOS polarizable CG water model, Kuhn *et al.* reported an MUE of 1.96 kcal/mol for 12 side-chain analogues, with less negative calculated free energy values for strongly polar side chains and more negative values for non-polar ones.¹⁸

It is widely known that solvents can impose a wide range of effects on solutes via specific and non-specific interactions, thus affecting many phenomena including chemical equilibria, electronic spectra, and reaction rates.¹⁹ In comparison to the numerous past studies on solvent effects for atomistic^{20–22} and continuum solvent models,^{23,24} the current understanding of the nature of CG solvation in multiscale simulations is still very limited. The present study systematically investigates hydration properties and solvent effects for an AAX/CGS multiscale model. The model consists of AA solute molecules described by the OPLS-AA force field^{25,26} and the polarizable Martini CG water model.²⁷ The mixed-resolution solute-solvent interactions are parameterized to best reproduce experimental hydration free energies of 42 small, neutral organic molecules,²⁸ and the optimal model is validated on an additional set of 28 small molecules. Solvents effects of CG water on torsional, *E/Z* conformational, and tautomeric equilibria of AA solutes are then studied and compared against the performance obtained using TIP4P²⁹ and Generalized Born/Surface Area (GB/SA) continuum solvent models.³⁰ Lastly, the differences between CG and AA solvent, solvation thermodynamics, as well as the computational efficiency of multiscale simulations are discussed. Through quantitative determination of hydration free energies and solvation properties of small organic molecules, this study offers insights into the merits and challenges of the CG solvent model.

Computational Details

The AAX/CGS Multiscale Model

The total Hamiltonian of the multiscale system was partitioned into a sum of intra- and inter-molecular interaction terms within each resolution domain, plus a mixed-resolution intermolecular interaction term (eq 1). The functional form and

$$H_{tot} = H_{intra}^{AA} + H_{inter}^{AA} + H_{intra}^{CG} + H_{inter}^{CG} + H_{inter}^{CG/AA}. \quad (1)$$

parameters of the OPLS-AA force field^{25,26} with 1.14*CM1A^{31,32} or 1.20*CM5^{28,33,34} partial atomic charges were used to describe solute interactions. Martini force field 2.0³⁵ and the polarizable Martini CG water model²⁷ were used for intra- and inter-molecular CG interactions. Each CG particle represents four water molecules and consists of three interaction sites. A positive and a negative charged site (± 0.46 e) are attached at a fixed distance (1.4 Å) to the central Lennard-Jones (LJ) site. The charged sites have angular

freedom, which makes the model polarizable. A dielectric constant $\epsilon_r = 2.5$ is used to dampen all electrostatic interactions with the polarizable Martini model. The correct incorporation of the CG water within our Metropolis Monte Carlo simulations was confirmed by reproduction of various bulk liquid properties (Table S1, Figure S1).

The mixed-resolution AAX/CGS interaction consists of electrostatic and a LJ terms. The Coulomb potential in eq 2 was used to describe the solvent-solute electrostatic interactions,

$$H_{elec}^{CG/AA} = \sum_{i,j} \frac{1}{4\pi\epsilon_{mix}} \frac{q_i q_j}{r_{ij}}, \quad (2)$$

where a constant dielectric permittivity ϵ_{mix} ($1 < \epsilon_{mix} < 2.5$) was adopted to represent the dielectric screening effect of the CG water on an AA solute. Due to the separation of the LJ and charge sites in the polarizable Martini CG water model, numerical instability was observed in the simulations when one of the CG charged particles overlapped with solute atoms due to highly attractive electrostatic interaction but insufficient LJ repulsion. This artifact usually occurs for solute molecules with polar hydrogen atoms and has been observed in a previous study.¹⁴ To counteract the overly attractive interaction while keeping the number of additional parameters minimal, the repulsive LJ term was scaled by a constant scaling factor c ($c > 1$) to increase the short range repulsion (eq 3). The well depths and radii of the LJ potential were determined by geometric combination rules (eq 4). Then the optimal scaling factor c and dielectric constant ϵ_{mix} were parameterized to minimize errors in computed hydration free energies of 42 small organic molecules.

$$H_{vdW}^{CG/AA} = \sum_{i,j} 4\epsilon_{ij} \left[c \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (3)$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \quad \sigma_{ij} = \sqrt{\sigma_i \sigma_j} \quad (4)$$

Free Energies of Hydration

Metropolis Monte Carlo sampling in conjunction with Free Energy Perturbations (MC/FEP) calculations were carried out following standard procedures using a modified BOSS program to compute the absolute free energies of hydration for AA solute molecules in AAX/CGS multiscale simulations.^{36–38} For each aqueous phase simulation, a cubic cell consisting of a single AA solute molecule and 500 CG particles (2000 water molecules) was generated, and the simulations were conducted at 25 °C and 1 atm in the isothermal-isobaric (NPT) ensemble. Solvent-solvent and solute-solvent cutoff distances for nonbonded interactions were set at 12 Å based on all non-hydrogen atom pairs. As in the standard Martini CG simulation protocol,³⁵ the intermolecular interaction potential between CG particles was feathered to zero in the range of 0 – 12 Å for the electrostatic term and 9 – 12

Å for the LJ term. The solute-solvent nonbonded interactions were quadratically feathered to zero over the last 0.5 Å. Solute and volume moves were attempted every 100 and 3125 configurations, respectively. Ranges for translations and rotations of ± 0.10 Å and $\pm 6.0^\circ$ for solute and ± 0.22 Å and $\pm 15.0^\circ$ for CG water were chosen to produce MC acceptance rates of 30 – 50%.

Full molecular annihilations of solute molecules were achieved in two successive steps: first atomic charges were neutralized, then the LJ parameters were reduced to zero, as well as shrinking all the covalent bonds to 0.3 Å.²⁸ For each annihilation phase, 25-double-wide-sampling (25-DWS) windows with 5M/5M (equilibration/averaging) configurations for gas phase simulations and 15M/30M for multiscale simulations were used. The overall uncertainties of calculated absolute free energies of hydration are generally below 0.15 kcal/mol; typical experimental uncertainties are 0.3–0.5 kcal/mol.^{39–44}

Solvent Effects

Monte Carlo statistical mechanics simulations for AAX/CGS systems were carried out to obtain equilibrium distributions of solute conformations. 50M/80M MC configurations were used in aqueous phase simulations, and the setting of parameters was the same as in the MC/FEP calculations. In MC simulations of AA systems, a single solute with 1.14*CM1A charges was solvated by 512 TIP4P water in a cubic cell with periodic boundary conditions. The number of trial configurations was increased to 50M/150M, and simulation conditions and cutoff distance were identical to the AAX/CGS systems, except the solvent-solvent nonbonded interactions were quadratically feathered to zero over the last 0.5 Å. Ranges for translational and rotational moves of ± 0.15 Å and $\pm 15.0^\circ$ were adopted for TIP4P water. Preferential sampling of the solute, together with attempted flipping of key dihedral angles at constant intervals was used to enhance barrier crossings and conformation sampling. Gas-phase simulations were carried out using 5M/10M configurations. Simulations in the GB/SA continuum solvent environment were carried out following the standard protocols implemented in the BOSS program using 5M/10M configurations. 1.07*CM1A atomic charges were adopted for solute molecules, as previously recommended.³⁰

Free energy profiles for rotation about a chosen dihedral angle were computed by MC/FEP calculations using the dihpmpf procedure implemented in the BOSS program.⁴⁵ Specifically, all internal degrees of freedom for solutes were sampled with the exception of the chosen dihedral angle, which was perturbed from 0° to 180° or 360° , if no symmetry is present, in a cubic periodic system with CG or TIP4P water as described above. The calculations were done with 18- (or 36- if no symmetry is present) DWS windows in torsional angle increments of 10° from the reference geometry. 25-DWS MC/FEP calculations were carried out to obtain relative free energies of hydration of *E/Z* conformers by annihilating each individual conformer and calculating the difference. Estimates of free energies of hydration in the GB/SA continuum solvent were made following the standard protocol as in the previous study by Terhorst *et al.*⁴⁶ For the study of hydration effects on tautomerism, MC/FEP calculations were performed to gradually perturb one tautomer into another in 10-DWS windows.

Analysis of Solvation Thermodynamics

The absolute hydration entropy at temperature T is related to the free energy of hydration by eq 5. According to the finite-difference method, the relationship can be approximated using the free energies of hydration at two different temperatures, $T + \Delta T$ and $T - \Delta T$ (eq 6).⁴⁷ Note that a fundamental assumption behind the

$$\Delta S_{hyd}(T) = - \left(\frac{\partial \Delta G_{hyd}}{\partial T} \right)_{P,N} \quad (5)$$

$$-T \Delta S_{hyd}(T) \approx T \frac{\Delta G_{hyd}(T + \Delta T) - \Delta G_{hyd}(T - \Delta T)}{2\Delta T} \quad (6)$$

approximation is that the heat capacity is constant within $\pm \Delta T$ of the target temperature T . For aqueous solutions, this assumption normally holds true near room temperature with ΔT no greater than 50° .⁴⁸ Moreover, the statistical uncertainty for the entropy is inversely proportional to ΔT as given in the error propagation formula in eq 7. Previous study⁴⁹ suggested the use of $\Delta T \approx 30^\circ$ to yield an uncertainty of *ca.* 1 kcal/mol in $T \Delta S_{hyd}$, when the free energy uncertainty is *ca.* 0.1 kcal/mol. The enthalpy change H_{hyd} is reported as the difference between G_{hyd} and $-T \Delta S_{hyd}$, and the associate error is calculated as the sum of the statistical errors of the two quantities.

$$\sigma_{T \Delta S_{hyd}(T)} = \frac{T}{2\Delta T} [\sigma_{\Delta G_{hyd}(T + \Delta T)} + \sigma_{\Delta G_{hyd}(T - \Delta T)}] \quad (7)$$

In the current study, MC/FEP calculations were performed at 1 atm and three different temperatures: -5 , 25 , and 55°C . The annihilation calculations followed standard protocols described earlier, except that 50M/100M configurations per FEP window were used for simulations in TIP4P water, and 30M/50M configurations in the CG water.

Results

Parameterization of the AAX/CGS Multiscale Model

The initial focus was to develop the AAX/CGS multiscale model by calibrating the mixed-resolution interaction $H_{CG/AA}^{inter}$ based on experimental hydration free energies of 42 small, neutral organic solute molecules with diverse functionality. Absolute free energies of hydration were computed for multiscale systems, in which the OPLS-AA force field with 1.14*CM1A or 1.20*CM5 partial atomic charges were used for AA solute molecules, and the polarizable Martini CG water model was used for the solvent (CM1A/polCG and CM5/polCG). The scaling factor c for the LJ repulsive term and the dielectric constant ϵ_{mix} in

$H_{CG/AA}^{inter}$ were optimized to minimize the mean unsigned error (MUE) in the computed free energies of hydration.

For the CM1A/polCG model, the smallest MUE, 1.24 kcal/mol, is obtained with $c = 1.80$ and $\epsilon_{mix} = 1.27$. The root mean square error (RMSE) is 1.66 kcal/mol. These results show modest degradation compared to those determined previously by AA simulations using the TIP4P water model (CM1A/TIP4P), where the MUE and RMSE are 1.17 and 1.46 kcal/mol (Table 1).²⁸ The correlation of experimental and computed free energies of hydration is shown in Figure 1. Although in both models the data points are evenly distributed around the diagonal line designating the ideal correlation, the correlation coefficient (R^2) of the least squares fitting for the multiscale model ($R^2 = 0.77$) is poorer than for the AA model ($R^2 = 0.87$). Closer examination reveals a few specific outliers with the multiscale model. The hydration free energies of alkylamines (methylamine, dimethylamine) in the multiscale simulations magnify the discrepancies with the CM1A model, being too positive by 3–4 kcal/mol. As for nitro compounds, large errors are predominantly a ramification of the CM1A charge model, with which the dipole moments are far too large.²⁸

In a similar process, the inter-resolution coupling parameters for solutes with 1.20*CM5 partial charges were determined yielding $c = 1.80$ and $\epsilon_{mix} = 1.30$. This multiscale model gives a minimum MUE of 1.12 kcal/mol, lower than the MUE of 1.37 kcal/mol in the AA simulations. The RMSE values of the two models are close (Table 1). The larger MUE in the CM5/TIP4P model can be explained by the presence of a greater number of data points in the upper diagonal region (underestimation of the hydration free energies) of the correlation plot than the multiscale model (Figure 1). Despite this, the AA results give better linearity in the least squares fitting ($R^2 = 0.95$ vs. 0.84). A few large deviations for the multiscale simulations still arise, especially for amines and methanol.

Validation of the AAX/CGS Model

To further validate the assigned coupling parameters in the multiscale model, the hydration free energies of an additional 28 small molecules were computed and compared to those from experiment and AA simulations. As shown in Table 2, the MUEs for 1.14*CM1A/polCG and 1.20*CM5/polCG models are 1.61 and 1.58 kcal/mol, comparable to the MUEs in AA systems (1.61 and 1.88 kcal/mol). The MUE of 1.88 kcal/mol for 1.20*CM5 solutes in AA simulations is larger than expected.³⁴ Close inspection shows the largest errors come from sulfur and halogen containing compounds, which are known to be problematic with the CM5 partial charge model. Figure 2 provides further comparison of the multiscale and AA results. As in the parameterization section, the multiscale simulations yield marginally larger variation as also reflected in the RMSE values. Overall, the results confirm the accuracy of the multiscale model, and that average errors of 1.1 – 1.6 kcal/mol can be expected for the hydration free energies of small molecules using either CM1A or CM5 charges. If nitro compounds are not being modeled, CM1A charges are an expedient choice as they are computed within BOSS, while access to CM5 charges requires a call to an external *ab initio* program.^{28,33,34} The errors are reduced to 0.7 kcal/mol using OPLS-AA partial charges for the solutes.³²

Solvent Effects in the AAX/CGS Systems

The influence of the solvent is fundamental in controlling the conformation and reactivity of solutes in condensed phases. The effects occur in multiple ways, including specific (e.g., hydrogen bonding) and non-specific (e.g., electrostatics, dispersion) solute-solvent interactions, and entropic contributions (e.g., cavity formation and rearrangement of solvent structure). The present study investigated solvent effects for polarizable Martini CG water, TIP4P, and GB/SA continuum solvent models on torsional, *E/Z* conformational, and tautomeric equilibria.

Solvent Effects on Torsional Profiles

The effects of solvent on simple conformational equilibria have been well appreciated in past experimental^{19,50} and computational²⁰ studies. Computational investigations were carried out here for four solutes, 1,2-dichloroethane, 1,2-dichloropropane, dimethoxymethane and 2-methoxytetrahydropyran, for which significant conformational changes upon solvation are well known.^{51–53} For each case, MC simulations were performed in the gas phase, TIP4P water, polarizable Martini CG water, and GB/SA continuum solvent. Free energy profiles were computed *via* MC/FEP calculations, where the chosen dihedral angles were perturbed in steps of 10°.

Figure 3 illustrates the equilibrium distributions from direct sampling and the computed free energy profiles from the FEP calculations for the key dihedral angles of the solute molecules using the different solvent models. For 1,2-dichloroethane, the probability distribution of the dihedral angle Cl-C-C-Cl (Figure 3a) exhibits very different conformational preference in the gas phase and explicit water. In the gas phase, the *anti* conformation is dominant, as the electrostatic repulsion between the two chlorine atoms is minimized. In the aqueous phase, the conformational equilibrium shifts in favor of the more polar *gauche* form in both TIP4P and polarizable CG water. This *gauche* enhancement corroborates many past experimental measurements^{51,54} and theoretical results.^{20,54} However, the effect is underestimated with the GB/SA continuum solvent. In Figure 3b, the free energy profiles from the FEP calculations are in line with the dihedral angle distributions. The computed free energy difference in the gas phase ($G = G_{gauche} - G_{trans}$) is 1.13 kcal/mol, which is similar to the values found in photoelectron spectroscopy (1.0 ± 0.2 kcal/mol)⁵⁵ and electron diffraction studies (1.05 ± 0.10 kcal/mol).⁵⁶ Hydration shifts the equilibrium by 1.59 and 2.05 kcal/mol in TIP4P and CG water, close to the experimental estimates of 1.22 ± 0.13 ⁵⁴ and 1.26 ± 0.13 ⁵¹ kcal/mol from Raman scattering studies.

For 1,2-dichloropropane, there are three conformations, *gauche*[−], *anti*, and *gauche*⁺, as illustrated in Figure 4. As shown in Figures 3c and 3d, the results from the TIP4P and CG water simulations are nearly identical, and differ greatly from the gas-phase findings. Specifically, the *anti* conformation, which has a negligible dipole moment, is again the most favored in the gas phase, followed by a slight preference for the *gauche*[−] over the *gauche*⁺ conformation. This trend agrees well with experimental observations.⁵² In aqueous solution, the *gauche*[−] conformation accounts for the largest population due to its larger dipole moment, which allows it to have more favorable electrostatic interactions with polar solvent molecules. The same preference is observed in the free energy profiles (Figure 3d). While

the *gauche* enhancement for 1,2-dichloropropane is represented well in explicit water, it is again not captured adequately in the GB/SA continuum solvent.

In some systems, solvation has been found to mitigate or overcome the anomeric effect,⁵⁷ which denotes the tendency of heteroatomic substituents adjacent to a heteroatom within a tetrahydropyran ring to prefer the axial conformation over the sterically more favorable equatorial one.⁵⁸ In a more general sense, the anomeric effect is also used to describe the preference of *gauche* over an *anti* conformation in a segment R-X-A-Y (A = C, P, S; Y = N, O, F; and X is a lone-pair containing atom).⁵⁹ The current study examined the impact of solvent on two molecules that exhibit anomeric effects, dimethoxymethane (DMM) and 2-methoxytetrahydropyran (MTP).

For DMM, dihedral angle distributions and free energy profiles were determined for two geometries. In one a Me-O-C-O dihedral angle was restricted to be *gauche* and in the other *anti*, similar to the case with α - and β -anomers, respectively. The results of the α -like case are illustrated in Figure 3e and 3f. In the gas phase, the distribution of the flexible Me-O-C-O dihedral has a dominant peak around the *gauche* (60°) conformation, which corresponds to an overall *gauche*⁻, *gauche*⁻ conformation and the global energy minimum. In contrast, two additional minor conformations at *ca.* 170° (*gauche*⁻, *anti*) and 250° (*gauche*⁻, *gauche*⁺) emerge in the TIP4P and polarizable CG water simulations. Examination of the free energy profiles from the explicit solvent simulations also finds the two additional shallow minima. For the continuum GB/SA solvent, there is also a small bump for the 170° conformation, but it is smaller than in the explicit solvent simulations. For the β -like case, the corresponding results are in Figures 3g and 3h. The anomeric effect is clearly governing the conformational preference in the gas phase, with two peaks for the equivalent *gauche* forms near 60° and 300°. An additional small peak emerges at *ca.* 180° (*anti*, *anti*) in the TIP4P and polarizable CG water simulations, as this conformer has the largest dipole moment. The free energy profiles in Figure 3h show that the barriers separating the *anti*, *gauche* and *anti*, *anti* conformations decrease from *ca.* 4.3 kcal/mol in the gas phase to *ca.* 2.0 kcal/mol in both explicit solvent simulations, and the *anti*, *anti* valley becomes more pronounced. These observations corroborate previous findings from *ab initio* calculations on the DMM system.⁵⁹ Again, the solvent effects obtained with the GB/SA continuum model are diminished.

Lastly, the anomeric effect for a pyranose derivative, 2-methoxytetrahydropyran (MTP), was studied by interconverting the axial (*gauche*, *gauche*) and equatorial (*anti*, *gauche*) anomers (Figure 4) *via* MC/FEP calculations. While the axial conformer is favored in the gas phase, which is in line with the anomeric preference,⁵³ the results in the aqueous-phase simulations uniformly predict that the equatorial conformer is better solvated than the axial one, specifically by 1.15 kcal/mol in TIP4P water and 1.80 kcal/mol in the polarizable CG water. These compare favorably with prior computational studies on MTP, which give hydration free energy differences of 1.4 and 2.1 kcal/mol in water, using an *ab initio* reaction field method⁶⁰ and MC/FEP calculations.⁵⁷ The prediction from the GB/SA continuum solvent model is smaller, with a hydration energy difference of 0.57 kcal/mol favoring the equatorial conformer. In addition, the free energy profiles for the Me-O-C-O torsional angle were determined separately for the axial and equatorial anomers (Figures 3i and 3j). The free energy minima are found at the *gauche* position (*ca.* 60 and 300°) for both anomers

reflecting the exoanomeric effect, regardless of the solvent representation. Although the presence of solvent lowers the rotational barriers for both anomers, solvation does not alter the preferred rotamer. Overall, in agreement with previous findings on the MTP system,^{23,53,57} the current study shows that the aqueous medium shifts the conformational equilibrium in favor of the equatorial form, thus lessening the anomeric effect. Notably, the results from the CG solvent representation resemble well those using TIP4P water, while the solvent shifts using the GB/SA continuum model are too small.

Solvent Effects on *E/Z* Conformational Equilibria

Solvent effects on *E/Z* conformational equilibria for 10 pairs of organic molecules were also examined using MC/FEP calculations. The free energy difference for each *E/Z* conformer pair, in which the conformers differ by a ca. 180° rotation about a single bond, were determined in the gas phase and in TIP4P water, CG water, and the GB/SA solvent model.

As reported in Table 3, the accord for the computed hydration shifts are generally good and the shifts do not change the preferred conformer except for furfural. For acetic acid and methyl acetate, the *Z* conformers are much preferred in the gas phase, with G_{gas} being ca. 10 kcal/mol. While hydration shifts the equilibria towards the more polar *E* conformers, the *Z* preferences remain large. For *N*-methylacetamide (NMA) and *N*-methylformamide (NMF), the experimental differences in the hydration free energies are near zero, in line with the small change in the dipole moments for the *E* – *Z* interconversions ($\mu = 0.32$ and 0.34 D, respectively).^{40,41,43,46} However, all solvation models predict that *Z*-NMA will be better hydrated by over 1 kcal/mol, indicating possible inaccuracies with the CM1A charge model. The G_{hyd} determined for NMF by AAX/CGS and TIP4P models are close to the experiment, whereas the GB/SA continuum model gives larger deviation. For 1,3-dimethylurea and 1,3-dimethylthiourea, hydration does not shift the conformational preferences, which is in the order of $(Z,Z) > (E,Z) \gg (E,E)$, and the numerical accord between the three solvent models is very good. According to the computations, the only occurrence of an *E/Z* conformational interconversion upon hydration is for furfural, where the *Z* conformer is predicted to be better hydrated than the *E* form by 1.90, 2.77, and 2.41 kcal/mol with the GB/SA, TIP4P and polarizable CG solvent models, respectively. These solvent effect are in accord with the experimental trend, in which G_{hyd} is 1.55 and ca. 2 kcal/mol in going from the gas phase to acetone and DMSO.⁶¹ The conformational change is largely driven by the marked increase in the dipole moment ($\mu(E) = 3.92$ D and $\mu(Z) = 4.82$ D).

Overall, the results demonstrate that the solvent models at three different resolutions are mostly in good agreement with each other and can adequately represent the effect of hydration on *E/Z* equilibria.

Solvent Effects on a Tautomeric Equilibrium

The presence of solvent can also affect tautomeric equilibria, where the more polar tautomer should be progressively more favored in solvents of increasing dielectric constants.²¹ The current study revisited the prototypical 2-hydroxypyridine/2-pyridone equilibrium. MC/FEP

calculations were performed in the gas phase, TIP4P water, and polarizable CG water, by perturbing 2-pyridone to 2-hydroxypyridine using 1.14*CM1A atomic charges.

In gas phase *ab initio* and experimental studies, 2-hydroxypyridine is the lower-energy tautomer usually by 0.5 – 1.5 kcal/mol.^{62–64} Some computed free energy differences from high-end methods are in kcal/mol 1.18 (CBS/QB3), 0.94 (G4), and 1.70 (M06–2X/aug-cc-pVTZ).⁶² In the aqueous phase, the present simulations in TIP4P and polarizable CG water both find that the pyridone form is significantly better hydrated. Specifically, $G_{\text{hyd}}(\text{pyridine} - \text{pyridone}) = 2.89 \pm 0.11$ kcal/mol in TIP4P water and 3.38 ± 0.11 kcal/mol in CG water. Combined with the gas-phase data, both AA and CG results favor the pyridone tautomer in water by ca. 2 kcal/mol. A previously reported QM/MM result for the difference in TIP4P water using the AM1/OPLS/CM1A method is 1.83 ± 0.22 ,⁶⁴ and the experimental result deduced from pK_a values is around 4.0 kcal/mol,^{64–67} all in favor of the pyridone tautomer. Given that the experimental measurements might be subject to uncertainties due to self-association of the solute molecules,⁶⁸ the computed shifts with both methods for this equilibrium are acceptable. Their consistency again supports the notion that the CG model responds in a similar manner to the AA one in many contexts.

Further Results and Discussion

Performance of the AAX/CGS Multiscale Model on Small Molecule Hydration

As pointed out in the results section, MUEs for small molecule hydration free energies in multiscale and AA simulations are comparable in magnitude. Despite this, the two solvent models do exhibit dissimilar features on a case-by-case basis. The following analyses take a closer look at the hydration properties of each model.

Analysis of Error Distribution

The histogram in Figure 5 gives the distribution of unsigned errors (UEs) for the hydration free energies of the 70 organic molecules examined in the training and validation sets. The UEs from the multiscale simulations span a wider range than from the AA simulations. While the AAX/CGS simulations yield a greater number of predictions with UEs less than 0.5 kcal/mol, there are also strong outliers with UEs greater than 3 kcal/mol.

Analysis of the Solvation Free Energy Errors for Compound Classes

To compare the hydration features of the multiscale and AA model in detail, MUEs for various compound classes were separately determined (Table 4). Both AA and polarizable Martini CG solvent models yield lower MUEs for aromatic than non-aromatic compounds. For the CM5 charge model, the results for aromatic compounds and hydrocarbons in the multiscale simulations are in better agreement with experiment, indicating that the CG water model can characterize hydrophobic solvation well.

On the contrary, the multiscale model falters for polar solute molecules containing N or O, such as amines, alcohols, ethers, and amides. The larger deviations likely result from the lack of specific hydrogen bonding interactions in the CG solvent. In comparison to the results for alkylamines, good accuracy is obtained for the three aniline compounds (Table 1),

which are expected to be less sensitive to hydrogen bonding, with MUEs of 0.71 (CM1A) and 0.67 (CM5) kcal/mol in CG water, compared to 1.87 (CM1A) and 0.44 (CM5) kcal/mol in TIP4P water. However, for nitrile and carbonyl compounds, the MUEs determined in the multiscale simulations are comparable or smaller than in the AA simulations. Noticing the absence of polar hydrogen atoms in these compounds, this suggests that the current AAX/CGS model is particularly challenged in accounting for hydrogen bonding with polar hydrogen atoms. Consistently, the errors generally decline along the series, aniline, *N*-methylaniline, *N,N*-dimethylaniline, with errors of 1.32, 0.52, 0.29 kcal/mol for CM1A, and 1.10, 1.22, 0.38 kcal/mol for CM5. Similar observation is also made for alkylamines, although the errors are greater there. For nitro compounds, the large MUEs are dominated by the noted problems with the CM1A charges.

The compounds were also categorized according to hydrogen-bond (HB)-forming capacity. In agreement with analysis above, the MUEs from the multiscale simulations are greater than from the AA simulations for solutes with strong HB-forming tendency for both charge models. For compounds that have weak or no HB-forming capacity, the accuracy of the solvation models is reversed. When the compounds are further partitioned into classes of hydrogen-bond donor (HBD), hydrogen-bond acceptor (HBA) and hydrogen-bond donor and acceptor (HBDA), the multiscale model gives significantly greater errors for HBDA solutes, and marginally larger errors for HBA solutes. 3-Methylindole was the only molecule assigned as purely HBD; the CG solvent yields somewhat lower errors in this case, which may be dominated by better performance for the hydrocarbon part of the molecule. Overall, the analysis confirms that the accuracy of the multiscale model is closely related to the extent of HB-formation in the solvation process such that errors accumulate for polar solutes with multiple HB donating and accepting sites.

In contrast, the AAX/CGS multiscale model outperforms the AA model for the hydration of sulfur and halogen-containing compounds. For both CM1A and CM5 charge models, the hydration free energies of these compounds are mostly too positive in AA simulations. Dimethyl sulfoxide is the exception; its overly favorable hydration with AA solvent is much improved with the CG model. It is generally acknowledged that sulfur and halogen-containing compounds are difficult to model, due to the large size, electrostatic anisotropy, and high polarizability associated with sulfur and halogen atoms.^{28,69,70}

Solvation Thermodynamics

Thermodynamic analyses were carried out to obtain a more in-depth view of the AAX/CGS multiscale model. By dissecting the free energy of hydration into entropic and enthalpic terms, the analysis is informative in understanding the character and variations in solvation.^{47,49,71} Table 5 tabulates the hydration free energies, entropies, and enthalpies for propane, toluene, and methylamine, and methanol. These four molecules were selected for their diversity and because they exhibit large discrepancies between different solvent models, or between the computed results and experiment.⁷²

For the hydrophobic solute molecules, propane and toluene, although the hydration free energies predicted by the AAX/CGS solvation model are close to experiment, there is substantial variation in the entropic and enthalpic contributions. The experimental results

show that the small net hydration free energies reflect significant entropy-enthalpy compensation. However, with the multiscale solvation model, both the entropy and enthalpy terms are much smaller in magnitude. The CG model does not capture the essential character of hydrophobic solvation, which features a substantially attractive enthalpic component that is offset by the entropic cost of cavity formation and ordering of water molecules in the first solvent shell. A similar enthalpy-entropy discrepancy has also been observed in a previous study.¹⁸ In comparison, the results from the AA simulations are consistent with experiment, although quantitatively the entropy and enthalpy terms are too small in magnitude for propane and too large for toluene. For methylamine, the AA simulations nicely reproduce the experimental thermodynamic data. The large error (3.4 kcal/mol) for the free energy of hydration with the CG model is found to result primarily from an insufficiently favorable hydration enthalpy. Finally, for methanol, the free energies of hydration from both solvation models are nearly identical; the 2–3 kcal/mol discrepancy with the experimental result again arises from too positive enthalpies of hydration.

In short, the above analysis reveals details about a solvation model beyond the results of free energy calculations. Although good accuracy for hydration free energies is achieved with the multiscale simulations for hydrophobic solutes, the CG solvent does not reflect the proper compensation of entropy and enthalpy terms. For polar solutes, the major problem with the CG model is confirmed to be insufficiently favorable hydration enthalpies, which is attributable to the inadequate description of hydrogen bonding. Overall, the errors associated with entropy and enthalpy terms in multiscale simulations are related to the underestimate of favorable enthalpic interactions and underestimate of entropic penalties from the associated flattening of the potential energy surfaces.

Efficiency of the Multiscale Model

To assess computational efficiency, a benchmark study was performed to compare the performance of the AA and multiscale model. MC/FEP calculations were carried out to determine the absolute free energies of hydration following the standard protocol described in the methods section using 1.14*CM1A charges for the solutes. Four molecules (acetamide, acetophenone, DMSO, and toluene) in combination with two solvent box sizes: one with 250 polarizable Martini CG water (polCG), equivalent to 1000 TIP4P water; and the other with 500 polarizable Martini CG water, equivalent to 2000 TIP4P water, were used in the benchmark. The number of trial configurations used in the MC sampling was adjusted to yield similar statistical uncertainty for the hydration free energy. As reported in Table 6, the variations among hydration free energies determined with different sizes of solvent boxes are within the range of the statistical uncertainties. The AAX/CGS multiscale solvation model has a substantial computational efficiency advantage over the AA model, achieving a speedup factor of *ca.* 7–10 for the 250 polCG vs. 1000 TIP4P matchup, and *ca.* 20–30 for a 500 polCG vs. 2000 TIP4P comparison. These results are consistent with the expected *ca.* N^2 dependence of the computational time for an *N* particle system, which anticipates a *ca.* 16-fold advantage for the CG model. Thus, even greater reductions in computational time can be obtained for larger systems such as in modeling biomolecular systems in water.

Conclusion

The present study developed a multiscale model that allows direct coupling of AA solute molecules with the polarizable Martini CG water model by a parameterization based on free energies of hydration for 42 small organic solutes. From MC/FEP calculations, the mean unsigned errors (MUEs) determined in the multiscale simulations was 1.24 kcal/mol with the OPLS-AA/1.14*CM1A solute model and 1.12 kcal/mol with OPLS-AA/1.20*CM5 alternative, comparable to MUEs of 1.17 and 1.37 kcal/mol determined in simulations with the AA TIP4P water model. Further validation for hydration free energies of 28 additional solute molecules confirmed that the current multiscale model yields average errors similar to those from the AA approach.

Nevertheless, additional error and thermodynamics analyses did reveal some important discrepancies between the two solvent models. Though the multiscale simulations achieve higher accuracy for hydration free energies of hydrophobic compounds, they do not reproduce the experimentally observed enthalpy-entropy compensation effect. Poorer performance is also found with the CG model for polar solutes such as amines and alcohols owing to insufficiently favorable enthalpies of hydration. In addition, the representation of solvent effects using TIP4P water, polarizable Martini CG water and GB/SA continuum solvent models was examined for torsional, *E/Z* conformational, and tautomeric equilibria of multiple OPLS-AA/CM1A solutes. For these examples, the multiscale and AA models showed similar accuracy in reproducing the observed effects of solvation on equilibrium changes. These cases are controlled more by general electrostatic interactions that are captured by the CG model rather than by specific hydrogen bonding interactions. Not surprisingly, the GB/SA continuum solvent model performed less well in many cases with a pattern of underestimation of the effects of hydration.

Finally, FEP benchmarks demonstrated that the AAX/CGS multiscale simulations are about 7 to 30-fold more efficient than the AA counterparts for systems of the present size. There are tradeoffs when the resolution of the solvent is reduced; however, the overall performance of the current AAX/CGS model showed modest degradation compared to the AA alternative. A greater point of concern may be the average errors of 1 – 2 kcal/mol for free energies of hydration for small molecules from both models. This reflects continuing difficulties in the underlying force fields especially the assignment of atomic charges for arbitrary molecules. Much remains to be improved for modeling organic and biomolecular systems in solution. The investigation of alternative models for the aqueous environment is an important aspect of these efforts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Gratitude is expressed to the National Institutes of Health (GM32136) for support; and to Jonah Vilseck, Leela Dodda, Daniel Cole, and John Faver for helpful discussions. This publication is dedicated to Prof. J. Andrew McCammon in celebration of his 70th birthday.

References

1. Voth, GA. Coarse-Graining of Condensed Phase and Biomolecular Systems. CRC Press; FL: 2009.
2. Ingólfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, Marrink SJ. The Power of Coarse Graining in Biomolecular Simulations. *WIREs Comput Mol Sci*. 2013; 4:225–248.
3. Brini E, Algaer EA, Ganguly P, Li C, Rodriguez-Ropero F, van der Vegt NFA. Systematic Coarse-Graining Methods for Soft Matter Simulations - A Review. *Soft Matter*. 2013; 9:2108–2119.
4. Noid WG. Perspective: Coarse-grained Models for Biomolecular Systems. *J Chem Phys*. 2013; 139:090901. [PubMed: 24028092]
5. Marrink SJ, Tieleman DP. Perspective on the Martini Model. *Chem Soc Rev*. 2013; 42:6801–6822. [PubMed: 23708257]
6. Ayton GS, Noid WG, Voth GA. Multiscale Modeling of Biomolecular Systems: in Serial and in Parallel. *Curr Opin Struct Biol*. 2007; 17:192–198. [PubMed: 17383173]
7. Nielsen SO, Buló RE, Moore PB, Ensing B. Recent Progress in Adaptive Multiscale Molecular Dynamics Simulations of Soft Matter. *Phys Chem Chem Phys*. 2010; 12:12401–12414. [PubMed: 20734007]
8. Riniker S, Allison JR, van Gunsteren WF. On Developing Coarse-Grained Models for Biomolecular Simulation: A Review. *Phys Chem Chem Phys*. 2012; 14:12423–12430. [PubMed: 22678152]
9. Shi Q, Izvekov S, Voth GA. Mixed Atomistic and Coarse-Grained Molecular Dynamics: Simulation of a Membrane-Bound Ion Channel. *J Phys Chem B*. 2006; 110:15045–15048. [PubMed: 16884212]
10. Wassenaar TA, Ingólfsson HI, Prieß M, Marrink SJ, Schäfer LV. Mixing MARTINI: Electrostatic Coupling in Hybrid Atomistic-Coarse-Grained Biomolecular Simulations. *J Phys Chem B*. 2013; 117:3516–3530. [PubMed: 23406326]
11. Rzeplia AJ, Louhivuori M, Peter C, Marrink SJ. Hybrid Simulations: Combining Atomistic and Coarse-Grained Force Fields using Virtual Sites. *Phys Chem Chem Phys*. 2011; 13:10437–10448. [PubMed: 21494747]
12. Michel J, Orsi M, Essex JW. Prediction of Partition Coefficients by Multiscale Hybrid Atomic-Level/Coarse-Grain Simulations. *J Phys Chem B*. 2008; 112:657–660. [PubMed: 18163606]
13. Orsi M, Sanderson WE, Essex JW. Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. *J Phys Chem B*. 2009; 113:12019–12029. [PubMed: 19663489]
14. Riniker S, van Gunsteren WF. Mixing Coarse-Grained and Fine-Grained Water in Molecular Dynamics Simulations of a Single System. *J Chem Phys*. 2012; 137:044120. [PubMed: 22852610]
15. Huang W, Riniker S, van Gunsteren WF. Rapid Sampling of Folding Equilibria of β -Peptides in Methanol using a Supramolecular Solvent Model. *J Chem Theory Comput*. 2014; 10:2213–2223. [PubMed: 26580745]
16. Riniker S, Eichenberger AP, van Gunsteren WF. Solvating Atomic Level Fine-Grained Proteins in Supra-Molecular Level Coarse-Grained Water for Molecular Dynamics Simulations. *Eur Biophys J*. 2012; 41:647–661. [PubMed: 22797564]
17. Orsi M, Ding W, Palaiokostas M. Direct Mixing of Atomistic Solutes and Coarse-Grained Water. *J Chem Theory Comput*. 2014; 10:4684–4693. [PubMed: 26588159]
18. Kuhn AB, Gopal SM, Schäfer LV. On Using Atomistic Solvent Layers in Hybrid All-Atom/Coarse-Grained Molecular Dynamics Simulations. *J Chem Theory Comput*. 2015; 11:4460–4472. [PubMed: 26575936]
19. Reichardt, C., Welton, T. Solvents and Solvent Effects in Organic Chemistry. Wiley-VCH; Weinheim, Germany: 2010.
20. Jorgensen WL. Theoretical Studies of Medium Effects on Conformational Equilibria. *J Phys Chem*. 1983; 87:5304–5314.
21. Orozco M, Luque FJ. Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. *Chem Rev*. 2000; 100:4187–4225. [PubMed: 11749344]
22. Gao J. Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc Chem Res*. 1996; 29:298–305.

23. Cramer CJ, Truhlar DG. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem Rev.* 1999; 99:2161–2200. [PubMed: 11849023]
24. Bashford D, Case DA. Generalized Born Models of Macromolecular Solvation Effects. *Annu Rev Phys Chem.* 2000; 51:129–152. [PubMed: 11031278]
25. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc.* 1996; 118:11225–11236.
26. Jorgensen WL, Tirado-Rives J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc Natl Acad Sci U S A.* 2005; 102:6665–6670. [PubMed: 15870211]
27. Yesylevskyy SO, Schäfer LV, Sengupta D, Marrink SJ. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput Biol.* 2010; 6:e1000810. [PubMed: 20548957]
28. Vilseck JZ, Tirado-Rives J, Jorgensen WL. Evaluation of CM5 Charges for Condensed-Phase Modeling. *J Chem Theory Comput.* 2014; 10:2802–2812. [PubMed: 25061445]
29. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys.* 1983; 79:926–935.
30. Jorgensen WL, Ulmschneider JP, Tirado-Rives J. Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J Phys Chem B.* 2004; 108:16264–16270.
31. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG. Class IV Charge Models: A New Semiempirical Approach in Quantum Chemistry. *J Comput-Aided Mol Des.* 1995; 9:87–110. [PubMed: 7751872]
32. Udier-Blagovi M, Morales de Tirado P, Pearlman SA, Jorgensen WL. Accuracy of Free Energies of Hydration using CM1 and CM3 Atomic Charges. *J Comput Chem.* 2004; 25:1322–1332. [PubMed: 15185325]
33. Marenich AV, Jerome SV, Cramer CJ, Truhlar DG. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J Chem Theory Comput.* 2012; 8:527–541. [PubMed: 26596602]
34. Dodda LS, Vilseck JZ, Cutrona KJ, Jorgensen WL. Evaluation of CM5 Charges for Non-Aqueous Condensed Phase Modeling. *J Chem Theory Comput.* 2015; 11:4273–4282. [PubMed: 26575922]
35. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J Phys Chem B.* 2007; 111:7812–7824. [PubMed: 17569554]
36. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys.* 1953; 21:1087–1092.
37. Jorgensen WL, Tirado-Rives J. Molecular Modeling of Organic and Biomolecular Systems Using *BOSS* and *MCPRO*. *J Comput Chem.* 2005; 26:1689–1700. [PubMed: 16200637]
38. Jorgensen WL, Thomas LL. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J Chem Theory Comput.* 2008; 4:869–876. [PubMed: 19936324]
39. Abraham MH, Andonian-Haftvan J, Whiting GS, Leo A, Taft RS. Hydrogen-Bonding. 34. The Factors that Influence the Solubility of Gases and Vapors in Water at 298 K, and a new Method for its Determination. *J Chem Soc Perkin Trans.* 1994; 2:1777–1791.
40. Wolfenden R. Interaction of Peptide-Bond with Solvent Water: A Vapor Phase Analysis. *Biochemistry.* 1978; 17:201–204. [PubMed: 618544]
41. Radzicka A, Pedersen L, Wolfenden R. Influences of Solvent Water on Protein Folding: Free-Energies of Solvation of *cis* and *trans* Peptides are Nearly Identical. *Biochemistry.* 1988; 27:4538–4541. [PubMed: 3166998]
42. Juaristi, E., Cuevas, G. *The Anomeric Effect*. CRC Press; Boca Raton: 1994.
43. Shivakumar D, Williams J, Wu YJ, Damm W, Shelley J, Sherman W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J Chem Theory Comput.* 2010; 6:1509–1519. [PubMed: 26615687]
44. Mobley, DL. *Experimental and Calculated Small Molecule Hydration Free Energies*. UC Irvine: Department of Pharmaceutical Sciences, UCI; 2013.
45. Jorgensen WL, Buckner JK. Use of Statistical Perturbation Theory for Computing Solvent Effects on Molecular Conformation: Butane in Water. *J Phys Chem.* 1987; 91:6083–6085.

46. Terhorst JP, Jorgensen WL. E/Z Energetics for Molecular Modeling and Design. *J Chem Theory Comput.* 2010; 6:2762–2769. [PubMed: 20871784]
47. Peter C, Oostenbrink C, van Dorp A, van Gunsteren WF. Estimating Entropies from Molecular Dynamics Simulations. *J Chem Phys.* 2004; 120:2652–2661. [PubMed: 15268408]
48. Privalov PL, Gill SJ. Stability of Protein Structure and Hydrophobic Interaction. *Adv Protein Chem.* 1988; 39:191–234. [PubMed: 3072868]
49. Kubo MM, Gallicchio E, Levy RM. Thermodynamic Decomposition of Hydration Free Energies by Computer Simulation: Application to Amines, Oxides, and Sulfides. *J Phys Chem B.* 1997; 101:10527–10534.
50. Abraham, RJ., Bretschneider, E. Medium Effects on Rotational and Conformational Equilibria. In *Internal Rotation in Molecules.* Wiley-Interscience; New York: 1974. p. 481-584.
51. Kato M, Abe I, Taniguchi Y. Raman Study of the *trans-gauche* Conformational Equilibrium of 1,2-Dichloroethane in Water: Experimental Evidence for the Hydrophobic Effect. *J Chem Phys.* 1999; 110:11982–11986.
52. Schei SH, Stølevik R. Conformation and Molecular Structure of 1,2-Dichloropropane and 1,2-Dibromopropane as Determined by Gas-Phase Electron Diffraction and Molecular-Mechanics Calculation. *J Mol Struct.* 1985; 128:171–180.
53. Wiberg KB, Marquez M. The Energy Components of the Anomeric Effect for 2-Methoxytetrahydropyran. An Experimental Comparison of the Gas-Phase and Solutions. *J Am Chem Soc.* 1994; 116:2197–2198.
54. Cappelli C, Corni S, Tomasi J. Solvent Effects on *trans/gauche* Conformational Equilibria of Substituted Chloroethanes: a Polarizable Continuum Model Study. *J Phys Chem A.* 2001; 105:10807–10815.
55. Gan TH, Peel JB, Willett GD. Photoelectron Spectra of the *gauche* and *trans* Conformers of 1,2-Dichloroethane. *J Chem Soc Faraday Trans 2.* 1977; 73:965–972.
56. Kveseth K. Conformational Analysis. 1. The Temperature Effect on the Structure and Composition of Rotational Conformers of 1,2-Dichloroethane as Studied by Gas Electron Diffraction. *Acta Chem Scand.* 1974; 28:482–490.
57. Jorgensen WL, Morales de Tirado PI, Severance DL. Monte-Carlo Results for the Effect of Solvation on the Anomeric Equilibrium for 2-Methoxytetrahydropyran. *J Am Chem Soc.* 1994; 116:2199–2200.
58. McNaught, AD., Wilkinson, A. IUPAC Compendium of Chemical Terminology (the “Gold Book”). Blackwell Scientific Publications; Oxford: 1997.
59. Marcos ES, Pappalardo RR, Chiara JL, Domene MC, Martinez JM, Parrondo RM. Role of Geometrical Relaxation in Solution of Simple Molecules Exhibiting Anomeric Effects. *J Mol Struct: THEOCHEM.* 1996; 371:245–256.
60. Montagnani R, Tomasi J. The Influence of the Solvent on the Conformational Energy Differences due to the Anomeric Effect. *Int J Quantum Chem.* 1991; 39:851–870.
61. Foresman JB, Keith TA, Wiberg KB, Snoonian J, Frisch MJ. Solvent Effects. 5. Influence of Cavity Shape, Truncation of Electrostatics, and Electron Correlation *ab initio* Reaction Field Calculations. *J Phys Chem.* 1996; 100:16098–16104.
62. McCann BW, McFarland S, Acevedo O. Benchmarking Continuum Solvent Models for Keto–Enol Tautomerizations. *J Phys Chem A.* 2015; 119:8724–8733. [PubMed: 26202336]
63. Nowak MJ, Lapinski L, Fulara J, Les A, Adamowicz L. Matrix Isolation IR Spectroscopy of Tautomeric Systems and its Theoretical Interpretation. 2-Hydroxypyridine/2(1H)-Pyridinone. *J Phys Chem.* 1992; 96:1562–1569.
64. Kaminski GA, Jorgensen WL. A Quantum Mechanical and Molecular Mechanical Method Based on CM1A Charges: Applications to Solvent Effects on Organic Equilibria and Reactions. *J Phys Chem B.* 1998; 102:1787–1796.
65. Cook MJ, Katritzky AR, Linda P, Tack RD. Aromaticity and Tautomerism. Part I. The Aromatic Resonance Energy of 2-Pyridone and the Related Thione, Methide, and Imine. *J Chem Soc, Perkin Trans 2.* 1972:1295–1301.
66. Mason SF. The Tautomerism of N-Heteroaromatic Hydroxy-Compounds. 3. Ionization Constants. *J Chem Soc.* 1958:674–685.

67. Albert A, Phillips JN. Ionization Constants of Heterocyclic Substances. 2. Hydroxy-Derivatives of Nitrogenous Six-Membered Ring Compounds. *J Chem Soc.* 1956:1294–1304.
68. Beak P, Covington JB, White JM. Quantitative Model of Solvent Effects on Hydroxypyridine-Pyridone and Mercaptopyridine-Thiopyridone Equilibria: Correlation with Reaction Field and Hydrogen-Bonding Effects. *J Org Chem.* 1980; 45:1347–1353.
69. Mobley DL, Bayly CI, Cooper MD, Dill KA. Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations. *J Phys Chem B.* 2009; 113:4533–4537. [PubMed: 19271713]
70. Chamberlin AC, Cramer CJ, Truhlar DG. Extension of a Temperature-Dependent Aqueous Solvation Model to Compounds Containing Nitrogen, Fluorine, Chlorine, Bromine, and Sulfur. *J Phys Chem B.* 2008; 112:3024–3039. [PubMed: 18281971]
71. Yu HA, Karplus M. A Thermodynamics Analysis of Solvation. *J Chem Phys.* 1988; 89:2366–2379.
72. Ben-Naim A, Marcus Y. Solvation Thermodynamics of Nonionic Solutes. *J Chem Phys.* 1984; 81:2016–2027.

Figure 1. Correlation between experimental and computed free energies of hydration (kcal/mol) using 1.14*CM1A (left) and 1.20*CM5 (right) charge models. The results computed with CG and TIP4P water are shown in blue and green, respectively. The dashed red line shows the ideal $y = x$ correlation; the solid lines indicate the best linear least-squares fit of the computed data.

Figure 2.

Correlation between experimental and computed free energies of hydration G_{hyd} (kcal/mol) for the 28 test solutes using 1.14*CM1A (left) and 1.20*CM5 (right) charge models. The results from multiscale and AA simulations are shown in blue and green, respectively. The dashed red line shows the ideal $y = x$ correlation.

Figure 3.

Computed probability distributions (a, c, e, g) and free energy profiles (b, d, f, h, i, j) for dihedral angle ϕ . a,b: 1,2-dichloroethane, $\phi = \text{Cl-C-C-Cl}$; c,d: 1,2-dichloropropane, $\phi = \text{Cl-C-C-Me}$; e,f: α -dimethoxymethane, $\phi = \text{C-O-C-O}$; g, h: β -dimethoxymethane, $\phi = \text{C-O-C-O}$; i: (axial)-2-methoxytetrahydropyran, $\phi = \text{Me-O-C-O}$; j: (equatorial)-2-methoxytetrahydropyran, $\phi = \text{Me-O-C-O}$. $S(\phi)$ denotes the probability distribution for the dihedral angle. $\Delta G(\phi)$ is the cumulative free energy (kcal/mol) relative to $G(\phi = 0^\circ)$.

Figure 4.
Major conformations of (a) 1,2-dichloropropane and (b) 2-methoxytetrahydropyran.

Figure 5. Histogram of the distribution of unsigned errors (kcal/mol) for computed hydration free energies of 70 solute molecules using the polarizable CG and TIP4P water as solvent.

Computed Absolute Free Energies of Hydration ΔG_{hyd} (kcal/mol) in Multiscale and AA Simulations Using 1.14*CM1A and 1.20*CM5 Charge Models.^a

Table 1

solute	1.14*CM1A/polCG	1.14*CM1A/TIP4P ^b	1.20*CM5/polCG	1.20*CM5/TIP4P ^c	exptl. ^d
acetic acid	-8.48	-5.81	-8.63	-5.54	-6.7
acetophenone	-4.62	-4.54	-4.19	-3.39	-4.58
acetamide	-10.18	-10.14	-9.74	-10.28	-9.71 ^e
acetone	-2.19	-1.91	-2.46	-2.59	-3.81
aniline	-6.81	-7.30	-4.39	-4.78	-5.49
anisole	-2.98	-2.81	-2.34	-1.40	-2.46
N-methylaniline	-4.17	-6.79	-3.47	-3.30	-4.69
N,N-dimethylaniline	-3.74	-5.15	-3.07	-2.08	-3.45
trifluorotoluene	-1.34	0.17	-0.90	1.65	-0.25
benzene	-0.97	-1.14	-0.42	0.37	-0.86
benzamide	-11.64	-13.26	-10.33	-10.93	-11.01
benzonitrile	-4.60	-3.41	-5.25	-4.34	-4.21
methyl benzoate	-4.87	-4.31	-4.10	-2.71	-3.93
dimethylamine	-0.54	-3.20	0.29	-1.61	-4.3
N,N-dimethylacetamide	-6.10	-8.18	-7.02	-7.35	-8.55 ^e
dimethyl sulfoxide	-9.86	-12.88	-9.36	-11.87	-10.11
methyl acetate	-2.72	-3.27	-2.30	-2.58	-3.32
methyl formate	-2.32	-3.62	-1.58	-2.61	-2.78
ethane	1.71	2.92	1.70	2.89	1.83
dimethyl ether	-0.54	-0.90	-0.82	-1.67	-1.91
dimethyl sulfide	0.04	1.18	0.13	1.20	-1.61
acetaldehyde	-1.59	-1.93	-1.82	-2.41	-3.5
methyl chloride	0.56	0.87	0.59	0.84	-0.55
acetonitrile	-2.85	-2.59	-3.70	-3.87	-3.89
methanol	-2.51	-3.09	-2.01	-3.39	-5.1
methanethiol	0.60	0.56	0.61	0.53	-1.24
methylamine	-1.36	-5.12	0.28	-2.53	-4.56
nitroethane	-7.16	-6.35	-4.49	-2.65	-3.71

solute	1.14*CM1A/polCG	1.14*CM1A/TIP4p ^b	1.20*CM5/polCG	1.20*CM5/TIP4p ^c	expt. ^d
(<i>E</i>)- <i>N</i> -methylacetamide	-6.43	-8.34	-7.58	-8.45	-10.08 ^e
(<i>Z</i>)- <i>N</i> -methylacetamide	-8.07	-9.62	-8.97	-9.50	-10.08 ^e
bromobenzene	-1.60	-1.27	-1.10	0.25	-1.46
chlorobenzene	-1.45	-0.48	-0.81	0.99	-1.12
fluorobenzene	-0.74	0.01	-0.37	0.89	-0.8
nitrobenzene	-8.02	-8.11	-4.96	-2.83	-4.12
phenol	-6.95	-5.65	-5.62	-4.04	-6.62
thiophenol	-2.01	-2.32	-1.28	-0.46	-2.55
propane	1.34	3.30	1.34	3.27	1.96
propene	1.72	2.76	1.75	2.90	1.32
pyridine	-2.49	-3.08	-2.39	-2.89	-4.7
α -methylstyrene	-1.79	-0.98	-1.07	0.97	-1.24
thioanisole	-4.05	-2.16	-3.34	-0.27	-2.73
toluene	-1.47	-1.21	-0.76	0.48	-0.89
MUE	1.24	1.17	1.12	1.37	
RMSE	1.66	1.46	1.54	1.53	

^a Result from MC/FEP calculations. The statistical uncertainties for G_{hyd} are below 0.15 kcal/mol.

^b Ref 28.

^c Ref 34.

^d Ref 39.

^e Ref 40, 41.

Computed Absolute Free Energies of Hydration G_{hyd} (kcal/mol) in Multiscale and AA Simulations for AA Solutes in the Validation Set.^a

Table 2

solute	1.14*CM1A/polCG	1.14*CM1A/TIP4p ^b	1.20*CM5/polCG	1.20*CM5/TIP4P	exptl. ^c
2-propen-1-ol	-2.61	-2.61	-1.85	-2.14	-5.03
azetidine	-1.77	-4.62	-0.52	-1.90	-5.56 ^d
benzaldehyde	-4.16	-4.46	-3.57	-3.07	-4.02
2,2-dimethylbutane	-0.08	4.00	-0.12	3.89	2.51
butanoic acid	-7.61	-5.68	-7.64	-4.97	-6.36
cyclohexane	-1.01	2.54	-1.02	2.47	1.23
1,2-dichloroethane	-1.01	0.09 ^g	-0.49	0.47	-1.79 ^f
diethyl disulfide	-1.89	0.87 ^g	-2.05	0.47	-1.43 ^f
1,4-dioxane	-3.14	-4.43	-2.93	-4.16	-5.06
dimethyl disulfide	-1.16	0.86 ^g	-1.23	0.40	-1.54 ^f
propionitrile	-2.72	-2.21	-3.46	-3.60	-3.85
1,1,1-trichloroethane	-0.40	1.66 ^g	-0.08	1.94	-0.19 ^f
ethyl methyl sulfide	-0.48	1.24 ^g	-0.44	1.06	-1.50 ^f
hydrogen sulfide	2.39	1.43	2.43	1.42	-0.70 ^d
3-methylindole	-8.82	-9.04	-6.78	-4.69	-5.91
morpholine	-2.74	-5.86	-2.43	-5.15	-7.17
<i>o</i> -xylene	-2.17	-1.09	-1.45	0.86	-0.90
2-methylpyrazine	-3.42	-3.07	-2.05	-3.41	-5.51
4-bromophenol	-7.87	-5.74	-7.07	-4.50	-7.12
<i>Z</i> -2-pentene	0.82	2.59	0.89	2.83	1.31
1-bromopropane	-0.37	0.95 ^g	-0.74	0.80	-0.56 ^f
1-chloropropane	-0.33	1.53	-0.14	1.59	-0.33
2,2,2-trifluoroethanol	-5.38	-1.27	-4.00	-0.75	-4.31
thiophene	-0.69	-0.73 ^g	-0.13	0.67	-1.40 ^f
triethylamine	-1.11	-1.40	-0.49	-0.24	-3.22
trimethyl amine	-0.42	-1.00 ^g	-0.13	-0.78	-3.24 ^f

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

solute	1.14*CM1A/polCG	1.14*CM1A/TIP4P ^b	1.20*CM5/polCG	1.20*CM5/TIP4P	exptl. ^c
(E)-N-methylformamide	-7.37	-9.02	-8.81	-9.73	-10.00 ^e
(Z)-N-methylformamide	-7.85	-9.67	-8.66	-8.78	-10.00 ^d
MUE	1.61	1.61	1.58	1.88	
RMSE	1.99	1.80	2.11	2.05	

^aResults from MC/FEP calculations. The statistical uncertainties for G_{hyd} are below 0.15 kcal/mol.

^bRef 28.

^cRef 39.

^dRef 43.

^eRef 40, 41.

^fRef 44.

^gDetermined in this study.

Table 3

Computed $E - Z$ Free Energy Differences^a (kcal/mol) in the Gas Phase and in GB/SA, TIP4P and Polarizable CG Aqueous Solutions at 25 °C.

$E - Z$	G_{hyd} GB/SA ^b	G_{hyd} TIP4P ^c	G_{hyd} polCG ^e	G_{gas} 1.07 ^b	G_{gas} 1.14 ^e	G_{aq} GB/SA	G_{aq} TIP4P	G_{aq} polCG
acetic acid	-2.39 ^d	-3.43	-4.74	7.81	8.67	5.42	5.24	3.93
<i>N</i> -methylacetamide	1.43 ^d	1.28 ^f	1.64	0.76	0.48	2.19	1.76	2.12
<i>N</i> -methylformamide	1.80	0.65 ^f	0.48	1.21	1.12	3.01	1.77	1.60
methyl acetate	-0.60 ^d	-2.13	-2.68	12.10	12.75	11.50	10.62	10.07
(<i>EE</i> - <i>EZ</i>)1,3-dimethylurea	0.65	-0.51	0.25	8.25	7.80	8.90	7.29	8.05
(<i>EZ</i> - <i>ZZ</i>)1,3-dimethylurea	1.49	1.04	1.57	1.86	1.22	3.35	2.26	2.79
(<i>EE</i> - <i>EZ</i>)1,3-dimethylthiourea	0.99	0.27	0.43	8.60	8.22	9.59	8.49	8.65
(<i>EZ</i> - <i>ZZ</i>)1,3-dimethylthiourea	1.28 ^d	1.13	1.40	2.21	1.72	3.49	2.85	3.12
(<i>EZ</i> - <i>ZZ</i>)dimethyl carbonate	0.26 ^d	0.29	-0.75	9.94	10.59	10.20	10.88	9.84
furfural	1.90 ^e	2.77	2.41	-0.69	-1.05	1.21	1.72	1.36

^aReported as $G = \alpha(E) - \alpha(Z)$.

^b1.07*CM1A charge model was used for solutes.

^c1.14*CM1A charge model was used for solutes.

^dRef 46.

^eRef 30.

^fRef 28.

Table 4

Mean Unsigned Errors for Compound Classes (kcal/mol).

compound classes	CM1A/polCG	CM1A/TIP4P	CM5/polCG	CM5/TIP4P
non-aromatic (44)	1.70	1.54	1.65	1.60
aromatic (26)	0.91	1.08	0.78	1.55
C,H containing (10)	0.90	0.90	0.78	1.47
C,H,O containing (15)	1.19	0.95	1.34	1.23
alcohols/ethers (7)	1.94	1.24	2.19	1.63
carbonyls (9)	0.97	0.75	1.05	1.01
C,H,N containing (14)	1.97	1.66	2.23	1.63
amines (12)	2.45	1.73	2.87	2.04
nitriles (3)	0.85	1.25	0.54	0.13
C,H,N,O containing (10)	2.58	1.45	1.48	0.99
amides (7)	2.00	0.94	1.20	0.79
nitros (2)	3.67	3.32	0.81	1.17
sulfur containing (10)	1.12	1.84	1.26	2.15
thiols/sulfides (8)	1.29	1.87	1.32	2.21
sulfoxide (1)	0.25	2.77	0.75	1.76
halogen containing (11)	0.60	1.47	0.48	2.19
HB forming (41)	1.78	1.40	1.63	1.41
weak HB (11)	1.12	1.85	1.14	2.24
non-HB (18)	0.65	1.06	0.66	1.62
HBD (1)	2.91	3.13	0.87	1.22
HBA (21)	1.49	1.37	1.27	1.13
HBDA (19)	2.04	1.34	2.07	1.72

Table 5
Hydration Free Energies, Entropies, and Enthalpies (kcal/mol) at 25 °C Using 1.14*CM1A Charges.^a

solute	propane	toluene	methylamine	methanol
G_{hyd}	polCG	-1.51±0.11	-1.16±0.07	-2.64±0.07
	TIP4P	-1.41±0.13	-5.33±0.08	-2.94±0.07
	exptl. ^b	-0.88	-4.57	-5.10
$-T S_{\text{hyd}}$	polCG	0.44±0.55	4.49±0.48	5.45±0.46
	TIP4P	4.48±0.67	5.94±0.57	4.28±0.54
	exptl. ^b	6.79	5.70	5.16
H_{hyd}	polCG	0.83±0.61	-5.66±0.55	-8.08±0.53
	TIP4P	-1.49±0.74	-11.27±0.65	-7.22±0.61
	exptl. ^b	-4.83	-10.27	-10.25

^aThe calculated G_{hyd} values are slightly different from the corresponding ones listed in Table 1, because this analysis was carried out for a separate set of simulations.

^bAll experimental values are from ref 72.

Table 6
Computation Times for MC/FEP Calculations with the AAX/CGS and AA Solvation Models.

solute	solvent	# solvent mol.	# config/M ^a	CPU time/hr ^b	G_{hyd}^c
acetamide	polCG	250	15/20	25	-10.48±0.13
	polCG	500	15/30	42	-10.18±0.15
	TIP4P	1000	50/100	174	-10.13±0.11
acetophenone	TIP4P	2000	100/200	870	-10.08±0.10
	polCG	250	15/20	27	-5.03±0.15
	polCG	500	15/30	43	-4.62±0.13
DMSO	TIP4P	1000	50/100	199	-3.62±0.14
	TIP4P	2000	100/200	896	-3.76±0.13
	polCG	250	15/20	25	-9.90±0.11
	polCG	500	15/30	40	-9.86±0.13
	TIP4P	1000	50/100	274	-13.15±0.12
toluene	TIP4P	2000	100/200	1309	-12.93±0.10
	polCG	250	15/20	26	-1.07±0.13
	polCG	500	15/30	42	-1.47±0.11
	TIP4P	1000	50/100	292	-0.82±0.13
	TIP4P	2000	100/200	852	-1.18±0.10

^aNumber of trial configurations per FEP window. 25-DWS windows were used.

^bTotal CPU time for the MC/FEP calculation on an Intel Core2 Quad 3.3 GHz processor.

^cIn kcal/mol.