This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

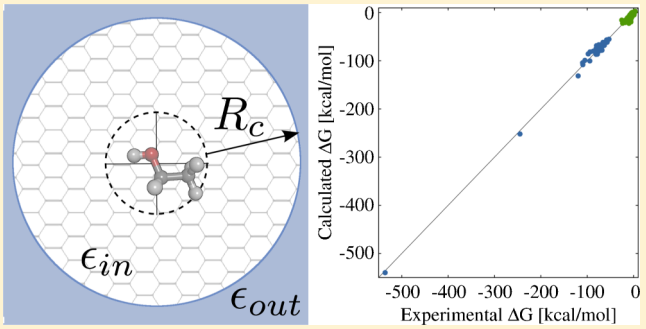
Hydration in Discrete Water (II): From Neutral to Charged Solutes

Piotr Setny*

Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

Supporting Information

ABSTRACT: In our previous work, we introduced a solvation model based on discrete solvent representation and demonstrated its ability to estimate hydration free energies for neutral solutes. Here, we present modifications extending the applicability of the model to charged solutes. They include improvements in the representation of the first hydration shell and systematic treatment of long-range interactions. While sharing computational efficiency of implicit solvent models, our approach avoids some of their important limitations, both in the context of electrostatic and nonpolar hydration effects: it naturally captures hydration asymmetry of opposite charges, it relies on solute description by standard all atom force fields instead of utilizing specialized sets of atomic parameters, it predicts solvent distribution in space without the need to geometrically define solvent accessible surface. By combining discrete solvent representation in vicinity of a solute with continuum description of long-range interactions, the model addresses two distinct aspects of biomolecular hydration: complex, short-range effects arising due to molecular nature of aqueous solvent, and bulk contributions. We demonstrate that the model provides good agreement with experimental results for an extensive set of roughly 700 diverse compounds, including neutral and charged solutes with hydration free energies ranging from +3.4 to −536 kcal/mol.



INTRODUCTION

Computational methods aiming at investigation of living matter at atomistic level need to take into account the influence of aqueous environment on biomolecules and their interactions.[1,2] One way of doing this is to perform simulations in which water molecules are explicitly modeled together with the system of interest. Such explicit solvent approaches provide most detailed insights; however, large computational effort makes their use prohibitive in many cases.[3] As an alternative one can estimate changes in hydration free energy—a free energy cost of transferring a solute from gas to aqueous phase—between the states of interest of a given system and include them into calculations.[4]

To date multiple approaches have been devised in this latter respect, starting from the evaluation of the type and area of hydrated molecular surface,[5,6] the use of continuum electrostatic calculations[7−10] together with less or more sophisticated estimates of nonpolar hydration free energies,[11−14] the application of the density functional theory of fluids,[15] to methods based on physicochemical descriptors.[16,17] Owing to computational efficiency, such implicit solvent methods can be used to account for the role of water in receptor−ligand binding, macromolecular docking, protein folding, and other problems,[18,19] which are currently hard to investigate with explicit solvent simulations.

Most implicit solvent approaches were introduced, parametrized and validated based on relatively simple, small molecules with experimentally determined hydration free energies.[20] This does not guarantee, however, their transferability to systems where peculiar solvent effects take place. For instance, water arrangement in protein interiors may vary from the absence in sterically hydratable cavities that remain dry due to hydrophobic nature of the surrounding residues,[21,22] to persistent residence of solitary water molecules embedded in hydrogen bond network within a macromolecular structure.[23] While of particular relevance for docking or protein folding studies, such effects are not captured by most implicit solvent approaches—in this respect, the development of 3D-RISM (three-dimensional reference interaction site model) based methods is of particular promise,[24] yet their routine application for high throughput calculations is still limited.

Another important issue is related to charged solutes such as atomic or molecular ions, including ubiquitous charged side chains in proteins or phosphate groups in nucleic acids. First, the representation of charged species in typical training sets for the estimation of hydration free energies is sparse due to experimental problems associated with determining thermodynamic data for extremely hydrophilic solutes.[25] As a result the accuracy of methods parametrization toward ionic solutes is limited, or such solutes are totally excluded from consideration. Second, the asymmetry of charge hydration (i.e., different hydration of positive and negative charges due to the asymmetry of water molecule)[26] is not captured by popular continuum electrostatic methods such as Poisson−Boltzmann

Received: February 28, 2015
Revised: April 20, 2015
Published: April 21, 2015

© 2015 American Chemical Society

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

or generalized Born approaches. In an attempt to address this issue, empirical corrections can be introduced to both formalisms,[27,28] yet their standard, unmodified versions are still used in the core of most implicit solvent calculations.

The complexity of hydration effects calls for methods suitable to represent solute-induced peculiarities in aqueous environment better than continuum models. A possible direction is to combine elements of explicit water description with methods allowing for simplified estimation of thermodynamic properties. In this spirit, a Langevin dipoles model[29] accounts for electrostatic effects by considering free energy of a grid of polarizable dipoles exposed to electric field generated by a solute. This approach explicitly captures water polarization (and its saturation in strong electric fields) as well as mutual interaction between solvent dipoles.

A more recent method, called semiexplicit assembly[30] makes use of tabularized free energy components (nonpolar and polar) of first hydration shell water determined in explicit solvent simulations for various model spheres. For an arbitrary solute, relevant free energy components are selected based on particular attributes sampled along its solvent accessible surface (curvature, electrostatic field, Lennard-Jones potential), and summed up to give the total hydration free energy. The method has been recently upgraded[31] to provide improved treatment of charged solutes by introducing solvent accessible surface that is iteratively adopted to account for closer solvent approach to solute regions producing strong electric field.

Pursuing the idea of combining explicit water model with simplified sampling of solvent configurations, we recently introduced a semiexplicit hydration model based on discrete solvent representation and mean field approach.[32] Most solvent properties are simplified to the extreme, but those particularly important for molecular hydration: water hydrogen bonding, as well as water-solute electrostatic and dispersion interactions are explicitly taken into account. While maintaining computational efficiency typical for fast implicit solvent approaches, the model addresses some of their important deficiencies. Model predictions include not only hydration free energies, but also solvent distribution in the presence of a solute. The predicted solvent distribution is not dependent on any (pre)defined solvent accessible surface construction, but adjusts self-consiestently to solute topography and physicochemical properties, thus capturing the aforementioned effects such as water expulsion from hydrophobic cavities or localization in hydrogen-bonded networks. Relying on atomistic force field description of water molecule, the model has build in asymmetry with respect to charge inversion. Furthermore, it is readily applicable to solutes described by standard force field parameters used for explicit solvent simulations.

In this report we present extensions of the method allowing for predictions of hydration free energies for charged solutes. Along with model simplification, leading to the reduction in the number of adjustable parameters from 6 to 5, they include the introduction of electrostatic reaction field, boundary corrections for long-range dispersion interactions, and improved treatment of the first hydration shell density. The model was reparametrized and tested based on an extensive set of 642 neutral[20] and 44 charged[25,29] solutes, giving quantitatively sound predictions for hydration free energies spanning the range from +3.4 to −536 kcal/mol.

As the preliminary version of the model was described in detail before,[32] in the following we briefly review key model assumptions and focus on the improvements introduced to allow calculations for charged solutes. Finally, we present the assessment of model efficacy on extended test sets.

THEORY AND METHODS

Basic Assumptions. The model is based on semiexplicit solvent representation, which is a combination of atomistic treatment of solute−water interactions with mean field solvent description. The space available for water molecules is discretized into three-dimensional body centered cubic (BCC) lattice with 1 Å distance between the nearest neighbors. Given the tetrahedral arrangement of the nearest neighbors in BCC lattice, it allows the placement of three atoms of SPC/E water model[33] (SPC/E - single point charge extended−one of popular atomistic water models used in explicit solvent molecular dynamics simulations, with OH distance of 1 Å, and HOH angle of 109.5°) in such a way that with an oxygen atom located at a given lattice point, two hydrogen atoms occupy two of its eight nearest neighbors (Figure 1). For each

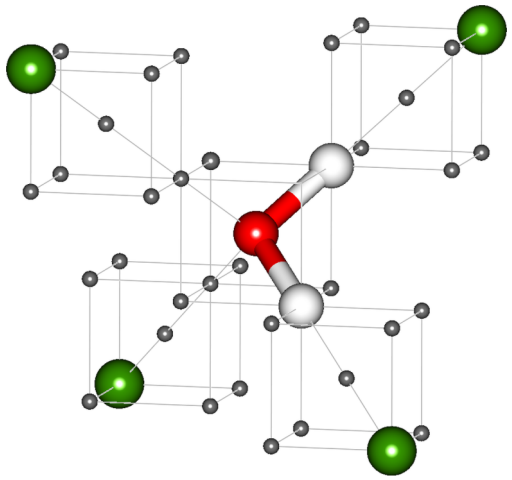


Figure 1. Section of BCC lattice with build in SPC/E water molecule (red, oxygen atom; white, hydrogen atoms). Green spheres represent lattice points considered for hydrogen bonding with the central water particle at its current orientation.

oxygen position, there are 12 unique combinations of hydrogen placements, corresponding to 12 possible orientations of a water molecule. A lattice point is considered as occupied by water (with density $\rho > 0$), if it holds an oxygen atom, or empty (with density $\rho = 0$) otherwise. In the absence of a solute, all lattice points are assumed to represent uniform density of bulk water at standard state $\rho_b = 0.0334$ Å^{-3} (997 kg/m^3).

A solute, considered in all atom representation, is treated as a source of external potential acting on the solvent lattice. For a given position of solute atoms, electrostatic and Lennard-Jones potentials are calculated at each lattice point using parameters taken from atomistic force field (GAFF force field[34] with AM1-BCC charges[35] is used for small organic molecules), adjusted for interaction with SPC/E water atoms by standard force field combination rules.

Effective Hamiltonian. The energy of a single solvent lattice point located at $\mathbf{r}$ is described by an effective Hamiltonian, which includes solute−solvent term, H_{UV}, and solvent−solvent term, H_{UU}, both ensemble-averaged over 12 possible orientations, θ, of an SPC/E water probe with oxygen atom placed at $\mathbf{r}$:

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

$$H_{eff}(\mathbf{r}, \{n\}) = -k_B T \ln\left(\sum_{\theta} e^{-\beta H_{UV}(\mathbf{r},\theta) - \beta H_{UU}(\mathbf{r},\theta,\{n\})}\right) \quad (1)$$

Here, $k_B T$, and β are the Boltzmann constant times temperature (T = 300 K is assumed), and its inverse respectively, and $\{n\}$ denotes the instantaneous distribution of occupied and empty solvent cells. The solute−solvent part is based on electrostatic, E_{el}, and Lennard-Jones, E_{LJ}, energy terms of SPC/E water molecule in respective solute potentials:

$$H_{UV}(\mathbf{r}, \theta) = E_{el}(\epsilon(\mathbf{r}), \mathbf{r}, \theta) + E_{LJ}(\mathbf{r}) \quad (2)$$

The electrostatic term includes position dependent dielectric constant, $\epsilon(\mathbf{r})$. It is either equal to 1 for water probes within the first hydration shell (the first hydration shell is constituted by occupied lattice points having at least one neighbor from which water was displaced by the solute), or adopts a constant value of ϵ_g (ϵ_g is one of five adjustable parameters of the model) for the rest of occupied lattice points. The E_{LJ} term does not depend on orientation, since the SPC/E water has only one Lennard-Jones interaction center at the oxygen atom.

The solvent−solvent term is based on the assumption that solvent lattice points interact with each other *only* through hydrogen bonds, whose creation depends on lattice occupancy. The number of hydrogen bonds maintained by a water probe at its given position and orientation is determined by the occupancies of four lattice points (Figure 1): two of them are three grid spacings away from the position of water probe oxygen in the direction of hydrogen atoms (which corresponds to a distance of 3 Å from oxygen center, and is well within the length limit for water−water hydrogen bond[36]), and two others are at the same distance, at the direction of oxygen lone pairs. Under these assumptions

$$H_{UU}(\mathbf{r}, \theta, \{n\}) = \frac{1}{2}\epsilon_{HB} N(\mathbf{r}, \theta, \{n\}) \quad (3)$$

were ϵ_{HB} is the assumed energy of a single hydrogen bond (a parameter of the model), and $N(\mathbf{r}, \theta, \{n\}) \in [0..4]$, is the number of occupied (i.e., with solvent density $\rho > 0$) lattice points at locations suitable for making a hydrogen bond with water probe at $(\mathbf{r}, \theta)$.

Bulk Corrections. In the original model, whose application was limited to neutral solutes, calculations were performed on a grid region sufficiently large to confine the influence of cutoff effects within the desired level of accuracy. In the case of charged species, such an approach is prohibitive due to long-range nature of electrostatic effects, motivating the introduction of boundary conditions.

Accordingly, the semiexplicit solvent region is restricted to a sphere of radius R, centered on solute's charge center (the geometric center of absolute atomic charges; mass center is used if all charges are zero). This radius is adopted to the actual solute size, by adding an offset, R_c, beyond the position of the most off-center solute atom (Figure 2). The region outside of the sphere is treated as a continuum with bulk water density and macroscopic dielectric constant, ϵ_{out}. The spherical shape of semiexplicit−continuum boundary, allows the use of analytical solutions for bulk corrections.

First, electrostatic correction, G_b^{el}, is evaluated as an energy of solute charges in the reaction field Φ, generated by the outer dielectric region. Such reaction field is obtained by the solution of Poisson equation for i solute charges, q_i, located at $\mathbf{r}_i$, inside a spherical cavity of radius R and dielectric permittivity ϵ_{in},

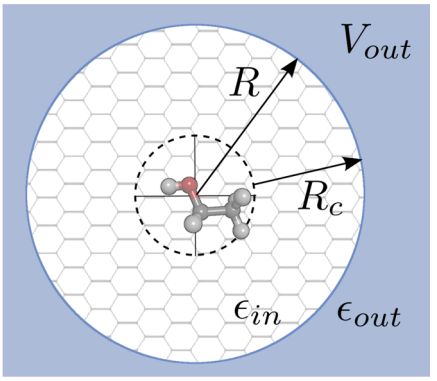


Figure 2. Schematic representation of spherical semiexplicit solvent region of radius R, extending by R_c from the position of most off-center solute atom, and an outer region, V_{out}, treated as a continuum with water dielectric constant ϵ_{out}.

centered at $\mathbf{r}$ = 0, and embedded in uniform medium of dielectric permittivity ϵ_{out}:[37,38]

$$\Phi(R, \mathbf{r}) = \sum_i q_i \sum_{n=0}^{\infty} \frac{1}{R}\left(\frac{r r_i}{R^2}\right)^n A_n P_n(\cos\gamma_i) \quad (4)$$

where $A_n = (((2n + 1)\ \epsilon_{in})/(n\epsilon_{in} + (n + 1)\epsilon_{out})) - 1$, γ_i is the angle between $\mathbf{r}$ and $\mathbf{r}_i$ (or 0, if $r r_i = 0$), and P_n are associated Legendre polynomials. For such reaction field, electrostatic bulk correction for a given cavity radius is[37,38]

$$G_b^{el}(R) = \frac{1}{2}\sum_i \frac{q_i}{4\pi\epsilon_0\epsilon_{in}}\Phi(R, \mathbf{r}_i) \quad (5)$$

where ϵ_0 is vacuum permittivity. For a single point charge located at r = 0, and assuming ϵ_{in} = 1, this expression reduces to the well-known Born formula for the free energy of transferring a charge from vacuum to the center of a spherical cavity of radius R inside uniform dielectric.[39] Fast convergence of $\Phi(R, \mathbf{r})$ calculation with the use of eq 4 is assured by the fact that all solute charges are at least a few Å away from the boundary of the semiexplicit solvent region, resulting in R considerably larger than any of r_i. In practice, including the first ten terms of the electrostatic potential expansion is enough to reach converged G_b^{el} values.

For hydration free energy calculations we assume ϵ_{out} = 78, as for bulk water at standard state. The value of internal dielectric permittivity, ϵ_{in}, is determined during the parametrization of the model, in order to ensure that the sum of electrostatic contributions due to semiexplicit solvent region and bulk correction does not depend on the actual radius of spherical boundary (Figure 3). Note that such construction implies that ϵ_{in} is not a freely adjustable parameter of the model, but is dependent on to the actual values of variable model parameters. The details of ϵ_{in} adjustment are given in the Supporting Information.

Bulk correction to dispersion interactions, G_b^{LJ}, is based on the assumption that the aqueous environment outside of semiexplicit solvent sphere can be treated as a uniform fluid with density ρ_b, interacting with solute atoms only thorough long-range, attractive part of the Lennard-Jones potential. The resulting interaction energy can be thus described by a volume integral over the infinite outer region or, following the application of Gauss theorem,[11] converted to surface integral over the spherical boundary:

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

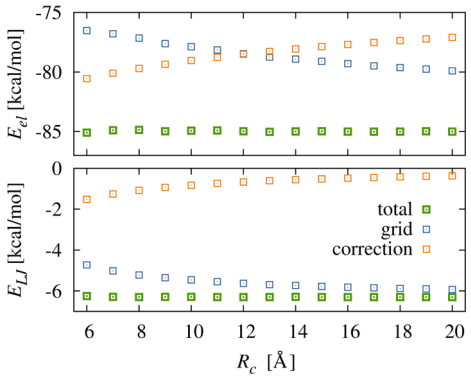


Figure 3. Electrostatic (E_{el}) and dispersion (E_{LJ}) contributions (green, total; blue, from semiexplicit region; orange, from bulk corrections) to hydration free energy for different sizes of semiexplicit grid region. Electrostatic contributions were evaluated for acetic acid (note that bulk correction values are shifted by −75 kcal/mol to fit into the plot) and dispersion contributions for benzene.

$$G_b^{LJ} = \rho_b \int_{v_{out}} U_{LJ}^{attr}(\mathbf{r})\, \mathrm{d}\mathbf{r} = \rho_b \oint_{\partial V_{out}} (\mathbf{A}\cdot\mathbf{n})\, \mathrm{d}S \tag{6}$$

Here, $U_{LJ}^{attr}(\mathbf{r})$ is a sum of attractive ($\sim r^{-6}$) Lennard-Jones contributions due to all solute atoms, $\mathbf{A} \sim \mathbf{r}/(3r^6)$ is a vector field, chosen such that $U_{LJ}^{attr} = \nabla\mathbf{A}$, and $\mathbf{n}$ is a surface normal vector. The calculation of $\mathbf{A}\cdot\mathbf{n}$ is particularly easy owing to the spherical symmetry of the boundary. The integral is performed numerically using three-dimensional spherical t-design with t = 240 points.[40] We found that in order to obtain total dispersion contribution to hydration free energy (i.e., the sum of grid contribution and bulk correction) that is independent of the actual size of semiexplicit solvent region, G_b^{LJ} needs to be evaluated for spherical boundary whose radius is shifted with respect to R by δR = 1.3 Å toward the solute. Such δR value was determined by trial and error and appears to be transferable to all investigated solutes.

The construction of bulk corrections described above, makes the results independent of the size of semiexplicit solvent region (Figure 3), as long as the R_c offset is large enough to contain at least a continuous shell of occupied lattice points around the solute. Obviously, smaller R_c values speed up the calculations, but at the same time closer proximity of boundary surface to solute atoms may increase inaccuracies of bulk corrections. We observed consistent results for R_c as small as 5 Å, but for model parametrization and all tests described below we used R_c = 6 Å. In the case of neutral compounds, such margin of semiexplicit solvent lattice typically results in G_b^{LJ} accounting for roughly 15% of total dispersion contribution to hydration free energy, and G_b^{el} which is negligible. Opposite proportions are observed for charged compounds, whose net charge is responsible for strong reaction field resulting in noticeable G_b^{el}, while van der Waals interactions are dominated by large solute−solvent repulsion captured by the first, semi explicit hydration shell.

Solvent Distribution. In a simplified, mean field solvent treatment, its average density at a given point, $\rho(\mathbf{r})$, is governed by the local excess chemical potential, $\mu(\mathbf{r})$, by the relation $\rho(\mathbf{r}) \sim \mathrm{e}^{-\mu(\mathbf{r})/k_BT}$, which implies that solvent accumulates in regions of favorable (i.e., low) excess chemical potential.[41] In the case of semiexplicit solvent lattice, water density is assigned to a given lattice point depending on its effective Hamiltonian (eq 1), which, for the proposed model, is equivalent to local excess chemical potential. It is assumed that for pure water, in the absence of external solute fields, $H_{\mathrm{eff}} \equiv H_b$, where H_b represents bulk excess chemical potential, and is an adjustable parameter of the model. In the preliminary model version, developed for neutral compounds that typically induce only moderate variations in H_{eff} with respect to H_b in solvent occupied areas, good estimates of hydration free energies were obtained with the assumption of only binary solvent density, that is $\rho \equiv \rho_b$ in all occupied regions, and ρ = 0 in regions where solvent is displaced by solutes.[32]

Such approach proved unsatisfactory, however, when applied to charged solutes. A set of model parameters providing hydration free energies in good agreement with experiment for neutral compounds, resulted in underestimated predictions for ions (Figure 4). The observed discrepancies were particularly

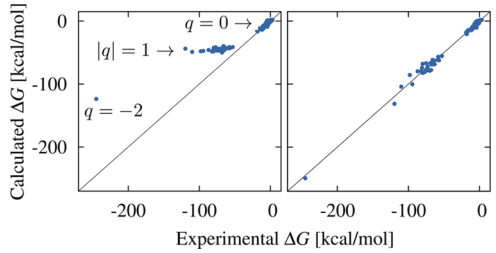


Figure 4. Results for the training set before (left) and after (right) the introduction of density scaling for the first hydration shell, illustrating dramatic improvement for charged ($q \neq 0$) solutes.

pronounced for small monovalent, or doubly charged ions, suggesting their link to the strength of electric field and its influence on the compactness of hydration shell. In order to account for the effect of strong solute−solvent attraction, we introduced the following coupling between H_{eff} and solvent density:

$$\rho(\mathbf{r}) = f(H_{eff}(\mathbf{r}, \{n\})) = \begin{cases} \rho_b(1 + a(H_b - H_{eff})), & \text{if } 1 + a(H_b - H_{eff}) \geq \eta \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Here η is a "drying threshold" (a fraction of bulk density, below which a given cell is considered as no longer occupied), and a > 0 is a density scaling factor, which are another two parameters of the model.

Under those assumptions solvent distribution in the presence of a solute is determined in the following way: having calculated solute potentials for all lattice points, we start with uniform solvent density $\rho = \rho_b$ for the whole lattice. Then, H_{eff} is calculated for each lattice point, and solvent density is changed according to eq 7. The last two steps are iterated until stationary distribution of occupied points is reached. Finally, solute hydration free energy, ΔG_s, is evaluated using the following formula:

$$\Delta G_s = \sum_{\{\mathbf{r}\}} v_b(\rho(\mathbf{r})H_{eff}(\mathbf{r}) - \rho_b H_b) + G_b^{el} + G_b^{LJ} \tag{8}$$

Here, the summation extends over all occupied lattice points, $\{\mathbf{r}\}$, v_b is a volume per BCC lattice point.

Convergence. Owing to discrete nature of the solvent lattice, ΔG_s values for a given, fixed solute conformation

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

depend on its orientation on the lattice. In order to limit the lattice-related bias, N independent runs with random solute placements are performed, and ΔG_s values are Boltzmann averaged to give the final estimate of hydration free energy. The convergence of such procedure depends on chemical complexity of the solute and the magnitude of its hydration free energy. For neutral solutes, standard deviation of hydration free energy estimates obtained upon Boltzmann averaging is typically confined to <0.25 kcal/mol within $N \sim 100$ runs, whereas $N \sim 2000$ is required to reach the same level of convergence for charged molecules (Figure 5). It is worth stressing, that owing

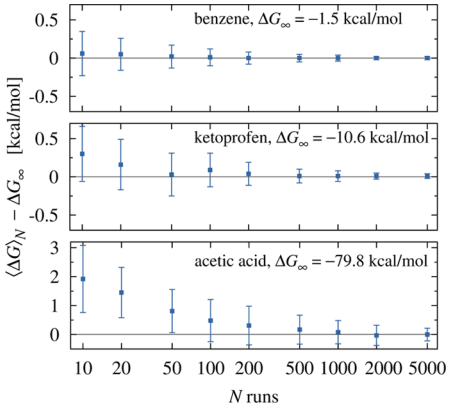


Figure 5. Convergence of ΔG_s estimates as a function of the number of random solute placements, N, considered for Boltzmann averaging. Error bars correspond to one standard deviation of 100 independent ΔG_s calculations for a given N.

to the simplicity of the model, 100 independent runs for an average, drug-like molecule take ~0.4 s on a single processor core (see Supporting Information for detailed analysis). Unless otherwise stated, the results reported here were obtained with $N = 100$ and $N = 500$ independent runs for neutral and charged solutes, respectively.

Data Sets and Solutes Preparation. In order to parametrize the model and evaluate its performance, we considered an extensive set of 686 small organic molecules with experimentally determined hydration free energies. It included both neutral and charged compounds, with hydration free energies ranging from 3.4 to −536 kcal/mol. The neutral part of the set comprised 642 compounds with curated experimental data and matching molecular structures provided by Mobley and Guthrie.[20] Charged compounds included molecular ions with experimental data assembled in the work of Florián and Warshel,[29] and monatomic ions with hydration free energies and LJ parameters adjusted for the SPC/E water model taken from the extensive study by Reif and Hünenberger (M_E parameter set).[25] All molecules (except from monatomic ions) were parametrized using GAFF force field[34] with AM1-BCC[35] partial charges. In order to determine conformations for compounds with rotatable bonds, such compounds were subjected to explicit solvent molecular dynamics simulations using the GROMACS[42] package. A standard protocol involved placing a molecule of interest in a rhombic dodecahedron simulation box of SPC/E water molecules extending at least 12 Å from solute atoms, and performing 1 ns of production run in NpT ensemble and under periodic boundary conditions, preceded by 100 ps thermalization and 100 ps equilibration. Electrostatic interactions were treated using particle mesh Ewald method with 10 Å cutoff in real space, and LJ potential was truncated at 10 Å. Temperature was maintained at 300 K with a Nose−Hoover thermostat, and pressure was set to 1 bar with Parrinello−Rahman barostat, using standard GROMACS implementations of both methods. During production runs, solute structures were saved every 5 ps, and than clustered according to all atom RMSD using gromos algorithm with 1 Å cutoff, as implemented in GROMACS. A middle structure of the most populated cluster was used for calculations with the model.

Parametrisation. The model has five adjustable parameters: ϵ_{HB}, the energy of hydrogen bond in water; H_b, bulk water free energy; η, a "drying threshold"; a, a density scaling factor; and ϵ_g, the effective dielectric constant of semiexplicit solvent region. Their values were determined based on extensive search of parameter space, with the goal to minimize root mean standard error (RMSE) of ΔG_s predictions with respect to experimental values. A training set, containing 119 rigid, neural organic molecules (molecules with no rotatable bonds were selected in order to limit the dependence of hydration free energies on solute conformations), and 33 randomly selected charged compounds was used. The estimation of hydration free energies for the entire training set takes ~6 s on a single processor with 8-cores, allowing a complete scan of 10^5 combinations within a few days. Three rounds of nested searches with progressively finer spacing in parameter space were conducted, leading to a final set of parameters: $\epsilon_{HB} = -2.65$ kcal/mol, $H_b = -9.94$ kcal/mol, $\eta = 0.28$, $a = 0.112$ (kcal/mol)$^{-1}$, and $\epsilon_g = 7.0$.

Poisson−Boltzmann Calculations. Hydration free energies obtained with the model were compared with predictions based on implicit solvent calculations. They were carried out with the adaptive Poisson−Boltzmann solver (APBS),[43] using zero bulk ionic strength, and a solvent dielectric of 78. The rest of physical parameters was either adjusted to give the best results or taken as default, as described in the Results. A grid spacing of 0.25 Å was used, and electrostatic contribution to hydration free energies was obtained as a difference between electrostatic energies in solvent medium and vacuum.

RESULTS

Hydration Free Energies. Model performance with respect to reproduction of experimental hydration free energies was evaluated based on 686 structures. The set comprised 642 neutral compounds with hydration free energies ranging from 3.4 to −25.4 kcal/mol,[20] and 44 ions with charges from +1 to −3 e, and hydration free energies from −53 to −536 kcal/mol.[25,44] Out of this set, 119 rigid neutral and 33 randomly selected charged compounds were used for model parametrization, and the rest served as a test set.

The overall results indicate that the model fairly well reproduces experimental hydration free energies, covering the range of almost 3 orders of magnitude (Figure 6). As expected, predictions for neutral compounds (Table 1) are generally much better (RMSE of 1.79 kcal/mol) than for charged compounds (RMSE of 5.56 kcal/mol). In the latter case, however, one must also take into account large experimental uncertainties, typically in the order of several kcal/mol. Importantly, modifications introduced to handle charged molecules also slightly improved model performance with respect to neutral solutes: the original model version[32] achieves RMSE of 1.94 kcal/mol on the entire neutral set.

Given the fact that the model depends on five adjustable parameters whose values were determined based on a test set

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

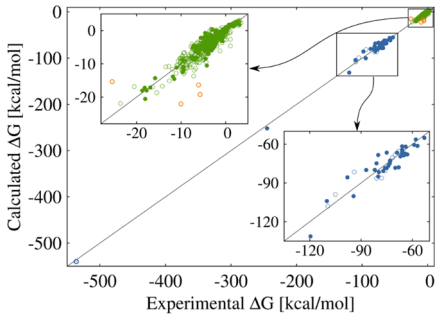


Figure 6. Experimental vs calculated hydration free energies for the entire set, comprising 642 neutral (green) and 44 charged (blue) compounds. The members of the training and test sets are marked with filled and empty circles, respectively. Four outliers among neutral compounds are marked with orange.

Table 1. Prediction Errors (in kcal/mol) with Respect to Experiment for the Training and the Test Sets, and Their Subgroups[a]

set	RMSE	⟨|Err.|⟩	⟨Err.⟩	$RMSE_{sim}$
all (686)	2.25	1.44	−0.13	
training (153)	2.95	1.85	−0.37	
test (533)	2.00	1.33	−0.05	
neutral (642)	1.79	1.23	−0.01	1.27
training (119)	1.51	1.10	−0.30	1.33
test (523)	1.85	1.26	−0.06	1.26
charged (44)	5.56	4.43	−0.30	
training (33)	5.57	4.47	−0.51	
test (11)	5.56	4.35	0.32	

[a]RMSE = root mean square error, ⟨|Err.|⟩ = an average absolute error, ⟨Err.⟩ = an average error, and $RMSE_{sim}$ = RMSE with respect to FEP calculations (where available).

comprising 153 molecules, the risk of overfitting should be low. Indeed, the results obtained for the test set are of similar quality as for the training set. When isolated outliers are removed from the test set (four neutral compounds—see below for the discussion regarding outliers), RMSE values for the neutral and charged parts of the test set are at similar levels as those for the corresponding parts of the training set: 1.56 and 5.56 kcal/mol respectively (Table 1).

Comparison with MD. One may consider three main sources of inaccuracies in model predictions: (a) general limitations of force field description, (b) simplification of solvent treatment, and (c) the lack of sampling of solute internal degrees of freedom. The fidelity of force field description can be assessed by comparing experimental hydration free energies with the results obtained by free energy perturbation (FEP) in explicit solvent—in principle, the most accurate computational approach available—using the same force field as in the model. FEP results for neutral compounds in the test set[45] show RMSE with respect to experiment of 1.50 kcal/mol, which is moderately better than 1.85 kcal/mol obtained by the model. It is worth noting, however, that RMSE of each method with respect to experiment is larger than RMSE between them (Table 1). Furthermore, four clear outliers (Figure 6, orange points) are poorly handled by both methods, likely indicating a common force field issue: their removal reduces the RMSE to 1.56 and 1.37 kcal/mol for the model and FEP results, respectively. Three of those outliers have their hydration free energies grossly overestimated (i.e., predicted as too favorable) both by the model and FEP calculations. All of them are organothiophosphates containing a common structural motif with a double bond between phosphorus and sulfur atoms, whose parametrization, in particular the assignment of partial charges, may pose a challenge to a semiempirical quantum mechanical method such as AM1-BCC used here. The fourth outlier, with underestimated ΔG_s prediction, is D-glucose. It contains five hydroxyl groups, whose standard parametrization by GAFF/AM1-BCC method was recently shown to produce nearly 1.5 kcal/mol systematic underestimation of hydration free energy per group.[46] Indeed, simply adding the respective summary correction to the predicted result reduces its error with respect to experiment by 75%.

One should not conclude, however, that force field limitation is the major source of inaccuracies in model predictions, just like it is in properly sampled FEP. A pairwise comparison of model and FEP predictions for all neutral molecules, results in RMSE of 1.27 kcal/mol. Although the model was parametrized with the goal to reproduce experimental rather than FEP results, it indicates that simplification in solvent description also significantly contributes to prediction errors but, fortunately, is not additive with force-field related issues. On the contrary, the lack of solute conformational sampling seems to have only moderate effect, as is indicated by the lack of RMSE difference with respect to FEP results for rigid (training test) versus flexible (test set) compounds (Table 1). Such relatively little effect of including solute conformational sampling on *average* error of hydration free energy predictions is consistent with previous observations for implicit solvent models.[47]

Comparison with PBSA. With its semiexplicit solvent treatment the model is roughly 10^4 times faster than FEP method (fraction of a second—see Supporting Information—vs hours per prediction), and as such should compete with fast, implicit solvent approaches. We chose to compare the results with a reference, PBSA method, which is based on Poisson–Boltzmann calculations for the evaluation of polar contributions, and surface area based term for the estimation of nonpolar contributions to hydration free energy. Two kinds of PBSA calculations were considered, both with the use of APBS program[43] and atomic radii as described in the Theory and Methods:

1 PBSAo, in which adjustable parameters, such as ϵ_{in}^{PB} = solute internal dielectric constant, R_{el} = solvent probe radius for defining electrostatic boundary, γ = surface tension coefficient, and R_{np} = solvent probe radius for surface area estimation, were optimized to provide the lowest RMSE with respect to experimental hydration free energies. Initially, we optimized PBSA parameters based on the same training set of 153 molecules as the one used for model parametrization. This approach, gave PBSA parametrization that performed worse than the semi explicit model in all considered categories (training and test sets, and their subsets as in Table 1). In order to provide an upper bound for the performance of PBSA method in its implementation considered here, we decided to use the *entire* set of 686 molecules for parameter optimization, which resulted in ϵ_{in}^{PB} = 1.75, R_{el} = 1.40 Å, γ = 0.003 kcal/mol/Å^2, and R_{np} = 0.30 Å,
2 PBSAs, in which standard APBS parameters for small solutes were used: R_{el} = R_{np} = 1.4 Å, ϵ_{in}^{PB} = 1, γ = 0.002 kcal/mol/Å^2.

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

In both cases the performance of the PBSA method evaluated over the entire set was worse than that of the semi explicit model (Figure 7). The main reason is apparently worse

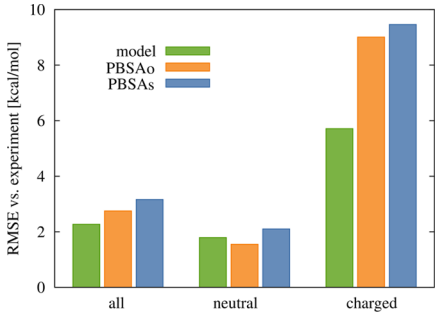


Figure 7. Root mean square errors for predicted hydration free energies vs experimental data for the model, PBSA approach with optimized parameters (PBSAo), and PBSA approach with standard parameters (PBSAs).

reproduction of experimental hydration free energies for charged solutes. Notably, even an attempt to optimize PBSA parameters solely for the subset of charged molecules did not significantly improve predictions for this group, resulting in RMSE of 8.92 kcal/mol (also at the cost of much worse performance for neutral compounds) versus RMSE of 9.01 kcal/mol achieved with PBSAo parameters, as reported in Figure 7. Furthermore, the exclusion of atomic ions from optimization, thus leaving only charged solutes sharing common atom types with the members of neutral set, resulted in RMSE of 8.38 kcal/mol, indicating that a single set of intrinsic atomic radii provides only limited accuracy when charged groups are present. An important factor here may be the indifference of implicit solvent description to solute charge inversion, as well as nonlocal character of hydration effects induced by charged groups, affecting also the transferability of parameters used to describe neighboring solute regions.[26]

In turn, PBSA predictions for all 642 neutral solutes were at similar level of accuracy to those obtained with the proposed model, with RMSE of 1.55 and 2.10 kcal/mol for PBSAo and PBSAs calculations, respectively. Interestingly, three out of four outliers among semi explicit model predictions (all three organothiophosphates mentioned above), had also the largest prediction errors in PBSA approach, albeit with not as large magnitudes as in the case of the semi explicit model: their removal from the neutral set reduced its RMSE by 0.24 and 0.11 kcal/mol for the model and PBSAo calculations respectively (detailed results are provided in the Supporting Information).

The estimate of nonpolar contribution to hydration free energy based on surface area (SA) used in PBSA approach, $\Delta G_{np}^{SA} = \gamma \mathrm{SA}(R_{np})$, turned out to have rather limited predictive value. Although, the optimized γ value obtained for PBSAo calculations is even larger than the one used in the standard setting (0.003 vs 0.002 kcal/mol/Å^2, respectively), the influence of the entire SA term on the actual RMSE for all compounds is low: setting $\gamma = 0$ (thus, completely neglecting the SA term), and optimizing only ϵ_{in}^{PB} and R_{el}, results in an increase of RMSE by only 0.02 kcal/mol. Furthermore, optimization of PBSA parameters performed only on the training set, results in $\gamma = 0$, indicating low transferability of γ values. Little importance of the SA term is more directly illustrated by the comparison with FEP results for hydration free energies of solutes without partial charges (Figure 8, lower plot), which results in rather poor correlation (Pearson's R coefficient of 0.25). Such result has

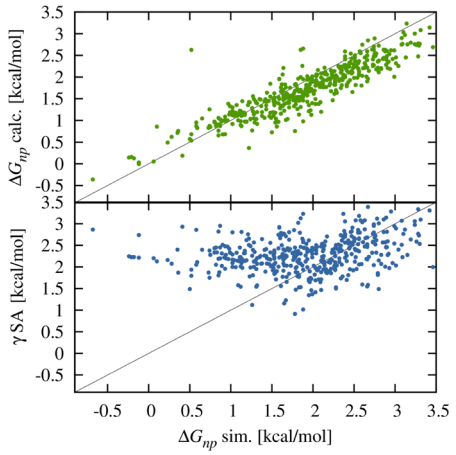


Figure 8. Comparison of nonpolar contribution to hydration free energy obtained by FEP explicit solvent simulations (ΔG_{np}sim.),[45] the model (upper plot), and (PB)SA approach (lower plot).

been already observed,[45,48] and puts into question the usefulness of adding surface area based term to Poisson–Boltzmann and generalized Born electrostatic models, at least in the context of hydration free energy predictions for small molecules. In contrast, the proposed model reproduces ΔG_{np} relatively well, with RMSE of 0.42 kcal/mol, and Pearson's correlation coefficient of 0.84 (Figure 8, upper plot), even though it was not directly parametrized to do so.

SAMPL4 Set. Periodically organized statistical assessment of modeling of proteins and ligands (SAMPL) challenge gives the opportunity to compare diverse methods for hydration free energy predictions on common, "blind" test sets. In this respect, a recent version of the challenge, included a set of 47 small, neutral molecules, covering a relatively wide range of hydration free energies of nearly 20 kcal/mol.[49]

Although the new version of semiexplicit model was still in development at the time when SAPML4 was organized and, hence, did not participate in the challenge, we provide our predictions in order to allow the comparison with other approaches. Obviously, it may be considered as not entirely fair, as all the results were already known to us, still we did not alter any model parameters nor modified our standard protocol of solutes preparation (see Theory and Methods) in order to influence the results. For solutes parametrization we used GAFF/AM1-BCC parametrization provided by Mobley and Guthrie in their newest database of experimental and calculated hydration free energies.[20]

The semiexplicit model achieved RMSE of 1.73 kcal/mol, and an average unsigned error of 1.45 kcal/mol, which places it between 16th and 17th of 49 submissions (45% of which represented alchemical MD simulations).[49] Interestingly, RMSE between model predictions and the results of FEP calculations using the same force field parameters[20] is only 1.02 kcal/mol, again indicating that the model and MD simulations give the results closer to each other than any of the two to experiment (RMSE of the cited FEP calculations with respect to experiment is 1.42 kcal/mol). Strikingly, PBSAo calculations for the SAMPL4 set (with the same parameters as used above, GAFF/AM1-BCC ligand parametrization, and conformations as used for semiexplicit model calculations) result in RMSE

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

with respect to experiment of only 1.12 kcal/mol—even slightly better than the top performers in the original blind challenge. Whether it is just a fortuitous statistical fluctuation needs to be evaluated on a larger and more diverse data set. Importantly, however, given only moderate success in parametrizing PBSA method for ions as described above, it illustrates the fact that an approach capable of delivering excellent results for neutral compounds may be not readily transferable to charged species. It raises an important question regarding the performance of such methods in the context of protein structures, where charged side chains can be buried or exposed to solvent during the process under study.

CONCLUSIONS

We presented improvements to the previously introduced semi explicit solvent model: systematic treatment of boundary corrections and scalable solvent density. Together they extend the applicability of the method to atomic and molecular ions, and also slightly improve the performance on neutral solutes. At the same time they speed up calculations by reducing the extent of solvent grid necessary to capture long-range interactions.

The performance of the model with respect to neutral molecules remains at comparable level to PBSA approach, while the results for charged solutes are significantly better. Importantly, the model relies on solute description by standard atomistic force fields, routinely used for explicit solvent simulations, and does not require any specific parametrization for charged solutes. In this respect, the model, having only five adjustable parameters, represents different philosophy than implicit solvent models based on continuum electrostatic calculations, typically utilizing specialized sets of atomic radii of often limited transferability.

In its current form, the model bridges two important aspects of biomolecular hydration. First is water behavior in direct vicinity of a solute. Here, molecular nature of aqueous solvent comes into play, determining the unique structure of the first hydration shell and its energetic response to solute fields. In this regard, the semiexplicit nature of the model allows for self-consistent evolution of solvent distribution in space, captures the interplay between water mutual hydrogen bonding and orientational preference toward the solute (including the saturation of water polarization in strong electric fields), as well as asymmetric hydration of opposite charges (see Supporting Information for discussion of model performance in this respect). The second aspect concerns contributions to hydration free energies resulting from long-range interactions. Here, the inclusion of analytic corrections based on, well validated in the bulk-like limit, continuous description of aqueous medium provides for physically sound description. Importantly, we demonstrated that once the first hydration shell is captured within the semiexplicit solvent region, an estimate of total hydration free energy does not depend on the actual radius of grid-continuum boundary, indicating that semi explicit solvent lattice provides a reasonable model for the free energy of bulk water dispersion and electrostatic interactions with external fields.

The ultimate area of application for simplified solvent models is, however, not the reproduction of hydration free energies of small solutes considered here, but rather the treatment of their interactions with macromolecular biological receptors, or the modeling of just macromolecules and their assemblies. In the case of large solutes with boundaries having complex topography and physicochemical character, capturing all peculiarities of the first hydration shell may be of particular importance. We believe that models being aware of molecular character of aqueous solvent, and relying on possibly general force field solute(s) description are a promising way of addressing such challenge. The efforts to evaluate the performance of our model in the context of biological macromolecules are currently underway.

ASSOCIATED CONTENT

Supporting Information

Notes on the parametrization of electrostatic bulk correction, issue of charge asymmetry, computational time, and detailed list of hydration free energies for compounds considered in the study. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b01982.

AUTHOR INFORMATION

Corresponding Author

*(P.S.) E-mail: p.setny@cent.uw.edu.pl.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Foundation for Polish Science grant Homing-Plus 2012-6/13.

REFERENCES

(1) Levy, Y.; Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389−415.

(2) Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **2008**, *108*, 74−108.

(3) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic ingredients of free energy calculations: a review. *J. Comput. Chem.* **2010**, *31*, 1569−1582.

(4) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1−20.

(5) Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086−3090.

(6) Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P. A. Solvation model based on weighted solvent accessible surface area. *J. Phys. Chem. B* **2001**, *105*, 5055−5067.

(7) Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117−129.

(8) Still, W.; Tempczyk, A. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(9) Sharp, K. A.; Honig, B. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.* **1990**, *94*, 7684−7692.

(10) Cramer, C. J.; Truhlar, D. General parameterized SCF model for free energies of solvation in aqueous solution. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.

(11) Floris, F.; Tomasi, J. Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation. *J. Comput. Chem.* **1989**, *10*, 616−627.

(12) Gallicchio, E.; Levy, R. M. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **2004**, *25*, 479−99.

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978

(13) Wagoner, J. a.; Baker, N. A. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331−6.
(14) Corbeil, C. R.; Sulea, T.; Purisima, E. O. Rapid prediction of solvation free energy. 2. the first-shell hydration (FISH) continuum model. *J. Chem. Theory Comput.* **2010**, *6*, 1622−1637.
(15) Beglov, D.; Roux, B. Solvation of complex molecules in a polar liquid: an integral equation theory. *J. Chem. Phys.* **1996**, *104*, 8678−8689.
(16) Bernazzani, L.; Duce, C.; Micheli, A.; Mollica, V.; Tiné, M. R. Quantitative structure - property relationship (QSPR) prediction of solvation gibbs energy of bifunctional compounds by recursive neural networks. *J. Chem. Eng. Data* **2010**, *55*, 5425−5428.
(17) Choi, H.; Kang, H.; Park, H. New solvation free energy function comprising intermolecular solvation and intramolecular self-solvation terms. *J. Cheminform.* **2013**, *5*, 8.
(18) Feig, M.; Brooks, C. L. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217−24.
(19) Baker, N. A. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137−43.
(20) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput. Aided. Mol. Des.* **2014**, *28*, 711−20.
(21) Setny, P.; Geller, M. Water properties inside nanoscopic hydrophobic pocket studied by computer simulations. *J. Chem. Phys.* **2006**, *125*, 144717.
(22) Berne, B. J.; Weeks, J. D.; Zhou, R. Dewetting and hydrophobic interaction in physical and biological systems. *Annu. Rev. Phys. Chem.* **2009**, *60*, 85−103.
(23) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573−81.
(24) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins Struct. Funct. Bioinforma.* **2007**, *66*, 804−813.
(25) Reif, M. M.; Hünenberger, P. H. Computation of methodology-independent single-ion solvation properties from molecular simulations. IV. Optimized Lennard-Jones interaction parameter sets for the alkali and halide ions in water. *J. Chem. Phys.* **2011**, *134*, 144104.
(26) Mobley, D. L.; Baker, J. R.; Barber, A. E.; Fennell, C. J.; Dill, K. A. Charge asymmetries in hydration of polar solutes. *J. Phys. Chem. B* **2008**, *112*, 2405−14.
(27) Purisima, E. O.; Sulea, T. Restoring charge asymmetry in continuum electrostatics calculations of hydration free energies. *J. Phys. Chem. B* **2009**, *113*, 8206−9.
(28) Mukhopadhyay, A.; Aguilar, B. H.; Tolokh, I. S.; Onufriev, A. V. Introducing charge hydration asymmetry into the generalized Born model. *J. Chem. Theory Comput.* **2014**, *10*, 1788−1794.
(29) Florián, J.; Warshel, A. Langevin dipoles model for ab initio calculations of chemical processes in solution: parametrization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution. *J. Phys. Chem. B* **1997**, *101*, 5583−5595.
(30) Fennell, C. J.; Kehoe, C. W.; Dill, K. A. Modeling aqueous solvation with semi-explicit assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3234−9.
(31) Li, L.; Fennell, C. J.; Dill, K. A. Field-SEA: a model for computing the solvation free energies of nonpolar, polar, and charged solutes in water. *J. Phys. Chem. B* **2014**, *118*, 6431−6437.
(32) Setny, P.; Zacharias, M. Hydration in discrete water. A mean field, cellular automata based approach to calculating hydration free energies. *J. Phys. Chem. B* **2010**, *114*, 8667−75.
(33) Berendsen, H. J. C.; R, G. J.; Straatsma, T. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(35) Jakalian, A.; Bush, B. Fast, efficient generation of high-quality atomic charges. AM1BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(36) Kumar, R.; Schmidt, J. R.; Skinner, J. J. Hydrogen bonding definitions and dynamics in liquid water. *J. Chem. Phys.* **2007**, *126*, 204107.
(37) Kirkwood, J. G. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* **1934**, *2*, 351−361.
(38) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **1995**, *102*, 5451.
(39) Born, M. Volume and heat of hydration of ions. *Z. Phys.* **1920**, *1*, 45.
(40) Hardin, R.; Sloane, N. McLarens improved snub cube and other new spherical designs in three dimensions. *Discret. Comput. Geom.* **1996**, *15*, 429−441.
(41) Hansen, J. P.; Mcdonald, I. R. *Theory of simple liquids*; Elsevier: Amsterdam, 2006.
(42) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(43) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(44) Florián, J.; Warshel, A. Calculations of hydration entropies of hydrophobic, polar, and ionic solutes in the framework of the Langevin dipoles solvation model. *J. Phys. Chem. B* **1999**, *103*, 10282−10288.
(45) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(46) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. A fixed-charge model for alcohol polarization in the condensed phase, and its role in small molecule hydration. *J. Phys. Chem. B* **2014**, *118*, 6438−6446.
(47) Mobley, D. L.; Dill, K. A.; Chodera, J. D. Treating entropy and conformational changes in implicit solvent simulations of small molecules. *J. Phys. Chem. B* **2008**, *112*, 938−46.
(48) Rizzo, R. C.; Aynechi, T.; Case, D. a.; Kuntz, I. D. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of nonpolar contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128−139.
(49) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. Blind prediction of solvation free energies from the SAMPL4 challenge. *J. Comput. Aided. Mol. Des.* **2014**, *28*, 135−50.

DOI: 10.1021/acs.jpcb.5b01982
J. Phys. Chem. B 2015, 119, 5970−5978