Published in final edited form as:
J Comput Chem. 2013 March 30; 34(8): 687–695. doi:10.1002/jcc.23181.

Parameterization of a Geometric Flow Implicit Solvation Model

Dennis G. Thomas[1], **Jaehun Chun**[2], **Zhan Chen**[3], **Guowei Wei**[4], and **Nathan A. Baker**[5]

Dennis G. Thomas: dennis.thomas@pnnl.gov; Jaehun Chun: jaehun.chun@pnnl.gov; Zhan Chen: chenzhan8@gmail.com; Guowei Wei: wei@math.msu.edu

[1]Pacific Northwest National Laboratory, East Lansing, MI, 48824, USA

[2]Pacific Northwest National Laboratory, East Lansing, MI, 48824, USA

[3]University of Minnesota, School of Mathematics, East Lansing, MI, 48824, USA

[4]Michigan State University, Department of Mathematics, East Lansing, MI, 48824, USA

Abstract

Implicit solvent models are popular for their high computational efficiency and simplicity over explicit solvent models and are extensively used for computing molecular solvation properties. The accuracy of implicit solvent models depends on the geometric description of the solute-solvent interface and the solvent dielectric profile that is defined near the surface of the solute molecule. Typically, it is assumed that the dielectric profile is spatially homogeneous in the bulk solvent medium and varies sharply across the solute-solvent interface. However, the specific form of this profile is often described by *ad hoc* geometric models rather than physical solute-solvent interactions. Hence, it is of significant interest to improve the accuracy of these implicit solvent models by more realistically defining the solute-solvent boundary within a continuum setting. Recently, a differential geometry-based geometric flow solvation model was developed, in which the polar and nonpolar free energies are coupled through a characteristic function that describes a smooth dielectric interface profile across the solvent–solute boundary in a thermodynamically self-consistent fashion. The main parameters of the model are the solute/solvent dielectric coefficients, solvent pressure on the solute, microscopic surface tension, solvent density, and molecular force-field parameters. In this work, we investigate how changes in the pressure, surface tension, solute dielectric coefficient, and choice of different force-field charge and radii parameters affect the prediction accuracy for hydration free energies of 17 small organic molecules based on the geometric flow solvation model. The results of our study provide insights on the parameterization, accuracy, and predictive power of this new implicit solvent model.

Introduction

Electrostatic interactions are ubiquitous and play a crucial role in polar or charged molecules such as water, aqueous ions, and biomolecules (proteins, nucleic acids, lipid bilayers, etc.) because of their long-range characteristics [1–4]. Since nearly all biomolecular interactions occur in aqueous environments, electrostatics are of central importance in understanding the biomolecular solvation which can, in turn, influence its structure, dynamics, and function. Several different levels of solvent detail are used in molecular simulations [4, 5]. Explicit solvent models treat the solvent in molecular or atomic detail but require extensive sampling to obtain thermodynamic or kinetic properties of interest. On the other hand, implicit solvent models simplify the solvent as a dielectric medium, which resolves the sampling problem at the cost of molecular detail and some accuracy. While each level of description has its

[5]To whom correspondence should be addressed. Pacific Northwest National Laboratory, PO Box 999, MSID K7-28, Richland, WA 99336, USA. 509-375-3997, nathan.baker@pnnl.gov.

strengths and weaknesses, implicit solvent models have become very popular and valuable for many biophysical studies due to their lower computational costs and reasonable accuracy for many calculated properties.

In implicit solvent models, the solvation energy is often assumed to consist of 'decoupled' polar and nonpolar contributions. However, recent computational studies on the solvation of organic molecules revealed the importance of coupling hydrophobic, dispersive, and electrostatic contributions to the solvation energy [6–8]. Dzubiella *et al.* [7, 8] proposed a theoretical formulation based on the minimization of the Gibbs solvation free energy with respect to a solvent volume exclusion function. Generally, in implicit solvent models, a biomolecule of interest and surrounding environments are separated by an arbitrary dielectric boundary which can be the molecular surface [9], the van der Waals surface [10], or other smoother surfaces [11, 12]; the solvation energy is very sensitive to the particular dielectric boundary of choice. However, unlike most implicit solvent models, the solvent-solute boundary is an output of Dzubiella *et al.* model to provide a more realistic interface and a self-consistent model for solvation. More recently, Chen *et al.* [6, 13] introduced a differential geometry based implicit solvation model in which a smooth solute-solvent boundary profile can be constructed by a generalized ("potential-driven") geometric flow equation. This model shares some similarities with the Dzubiella approach but includes a more detailed and accurate description of the nonpolar solvation free energy and a different treatment of the solute-solvent boundary.

The solute atom parameters can be obtained from empirical or *ab initio* based molecular force-fields. Other parameters may be obtained directly from experimental measurements. Typically, when applying implicit solvent models for computing solvation free energies of small molecules, it is often found necessary to optimize some of the parameter values to minimize the error between the computed (ΔG_{calc}) and experimental (ΔG_{expt}) solvation free energies [14]. Therefore, it is important to use appropriate measures of model validation on a new set of molecules to determine the domain of applicability, goodness-of-fit, robustness and predictability of the model.

In this work, we develop systematic parameterization of the geometric flow implicit solvent model with respect to changes in pressure, surface tension, solute dielectric coefficient, and the choice of force-field charges and radii. Recently, Nicholls *et al.* applied the Zap OpenEye implementation [14] of the Poisson-Boltzmann model to predict the hydration free energies of a set of 17 drug-like molecules (SAMPL0 data set) using AM1-BCC v1 charges and ZAP-9 radii parameterization. The ZAP-9 radii were derived by optimization of Bondi radii [15] for several different atom types to improve agreement between Zap Poisson-Boltzmann hydration free energy predictions and experimental values. In this work, we have used the same 17 molecule data set to optimize solvation hydration free energies from the geometric flow implicit solvent model with respect to changes in pressure and surface tension. It is well-known that the sensitivity/accuracy of implicit solvent models depends on the values selected for the force-field parameters. Therefore, we investigate the effect of different force-fields (AM1BCCv1/ZAP-9, AM1-BCCv1/Bondi, and OPLS-AA) on the optimal prediction of hydration energies. It should be noted that we only focus on parameters of the implicit solvent model (surface tension, dielectric coefficient, solvent pressure, etc.) and do not optimize the force-field charge or Lennard-Jones parameters.

Method

Problem formulation and governing equations

We consider a solute–solvent system, where the solute is an organic molecule and the solvent is water. We keep the atomic structure of the solute molecule fixed in the

calculations and represent the solvent as a dielectric continuum. The solute-solvent interface region, Ω_I, is defined as the region enclosed by the van der Waals surface of each solute atom and the solvent-accessible surface (SAS) of the solute molecule.

The differential geometry-based geometric flow implicit solvent model of Chen *et al.* takes into account both the polar and nonpolar contributions to the solvation free energy of molecules [6]. In this model, the total free energy functional (G) is written as the sum of the polar free energy functional (G_p) and a nonpolar free energy functional (G_{np}); i.e., $G = G_p + G_{np}$. In the absence of mobile ions, the polar free energy functional for the solute-solvent system is given by the equation,

$$G_p = \int_\Omega \left\{ S \left[\rho_m \varphi - \frac{1}{2} \varepsilon_m |\nabla \varphi|^2 \right] + (1-S) \left[-\frac{1}{2} \varepsilon_s |\nabla \varphi|^2 \right] \right\} d\boldsymbol{r}, \quad (1)$$

where S represents a characteristic function that takes the value 1 in the solute region and the value 0 in the solvent region over the problem domain Ω. In the solute-solvent interface region, S values vary smoothly from 1 at the solute van der Waals surface to 0 in the bulk solvent. ρ_m is the charge distribution of the fixed solute molecule that is composed of M atoms and is defined as $\sum_{i=1}^{M} Q_i \, \delta(\boldsymbol{r} - \boldsymbol{x}_i)$, where Q_i is the partial charge of atom i, located at the center of the atom. $\boldsymbol{x}_i$ denotes the position vector of the center of atom i, and $\boldsymbol{r}$ is the position vector of any point in the solute-solvent system. ε_m and ε_s are the dielectric coefficients of the solute and solvent, respectively, and φ is the electrostatic potential at point $\boldsymbol{r}$.

The nonpolar free energy functional is given by,

$$G_{np} = \gamma(A) + p(V) + \rho_0 \int_{\Omega_s} U^{att} d\boldsymbol{r}. \quad (2)$$

where Ω_S is the solvent portion of the problem domain, γ is the microscopic surface tension, A is the surface area of the solute molecule, p is the hydrodynamic pressure, V is the volume of the solute molecule, ρ_0 is the solvent bulk density, and U^{att} is the attractive potential of the van der Waals dispersion interaction between the solute and the solvent.

Chen *et al.* [6] expressed A, V, and U^{att} in terms of the characteristic function, S, as shown below:

$$A = \int_\Omega \left\| \nabla S(\boldsymbol{r}) \right\| d\boldsymbol{r}, \quad (3)$$

$$V = \int_\Omega S(\boldsymbol{r}) d\boldsymbol{r}, \text{ and} \quad (4)$$

$$\rho_0 \int_{\Omega_s} U^{att} d\boldsymbol{r} = \rho_0 \int_\Omega (1 - S(\boldsymbol{r})) U^{att} d\boldsymbol{r}. \quad (5)$$

The polar and the nonpolar free energy functionals (Eqs. (1) and (2)) are coupled via the characteristic function S.

The attractive potential, U^{att}, is given by the attractive part of the 12-6 L-J interaction potential (U_{is}^{LJ}) between a solute atom at position $\boldsymbol{r}_i$ and a point $\boldsymbol{r}$ in the solvent domain. The general form of the L-J potential can be written as:

$$U_{is}^{LJ}=\frac{A_{is}}{(\left\|\boldsymbol{r}_i-\boldsymbol{r}\right\|)^{12}}-\frac{B_{is}}{(\left\|\boldsymbol{r}_i-\boldsymbol{r}\right\|)^{6}}. \quad (6)$$

Depending on how A_{is} and B_{is} are defined, the L-J potential can be expressed in different forms. For example, substituting for $A_{is} = 4\varepsilon_{is}\sigma_{is}^{12}$ and $B_{is} = 4\varepsilon_{is}\sigma_{is}^{6}$ in Eq. (6), will yield the following expression for U_{is}^{LJ}:

$$U_{is}^{LJ}=4\varepsilon_{is}\left(\frac{\sigma_{is}^{12}}{(\left\|\boldsymbol{r}_i-\boldsymbol{r}\right\|)^{12}}-\frac{\sigma_{is}^{6}}{(\left\|\boldsymbol{r}_i-\boldsymbol{r}\right\|)^{6}}\right) \quad \text{, where} \quad (7)$$

ε_{is} is the depth of the attractive well at $\left\|\boldsymbol{r}_i-\boldsymbol{r}\right\|=\left(2\frac{A_{is}}{B_{is}}\right)^{1/6}$. Eq. (7) is used in the OPLS-AA force-field to model the van der Waals interactions between the solvent molecules and solute atoms. In the OPLS-AA force-field, $\sigma_{is}=2\sqrt{\sigma_i\sigma_s}$, where σ_i and σ_s are the radii of the solute atom and the solvent molecule, respectively. $\varepsilon_{is}=\sqrt{\varepsilon_i\varepsilon_s}$, where ε_i and ε_s are the well depth parameters of the L-J potential between two solute atoms and between two solvent molecules, respectively.

Another form of the L-J potential can be written as,

$$U_{is}^{LJ}=\varepsilon_{is}\left(\frac{\sigma_{is}^{12}}{(\left\|\boldsymbol{r}_i-\boldsymbol{r}\right\|)^{12}}-2\frac{\sigma_{is}^{6}}{(\left\|\boldsymbol{r}_i-\boldsymbol{r}\right\|)^{6}}\right), \quad (8)$$

where, $A_{is} = \varepsilon_{is}\sigma_{is}^{12}$, $B_{is} = 2\varepsilon_{is}\sigma_{is}^{6}$ and $\sigma_{is} = \sigma_i + \sigma_{is}$.

In general, the L-J potential can be decomposed into attractive and repulsive parts following Weeks-Chandler-Andersen (WCA) theory [16] as shown below:

$$U_i^{(att,WCA)}(\boldsymbol{r_i},\boldsymbol{r})=\begin{cases} -\varepsilon_{is} & \left\|\boldsymbol{r_i}-\boldsymbol{r}\right\| < \sigma_{is} \\ U_{is}^{LJ} & \left\|\boldsymbol{r_i}-\boldsymbol{r}\right\| \geq \sigma_{is} \end{cases} \quad (9)$$

$$U_i^{(rep,WCA)}(\boldsymbol{r_i},\boldsymbol{r})=\begin{cases} U_i^{LJ}+\varepsilon_i & \left\|\boldsymbol{r_i}-\boldsymbol{r}\right\| < \sigma_{is} \\ 0 & \left\|\boldsymbol{r_i}-\boldsymbol{r}\right\| \geq \sigma_{is} \end{cases} \quad (10)$$

Following previous work [17–19], we have used Eq. (9) to describe the attractive potential for van der Waals dispersion interaction, i.e., $U^{att}=\sum_i U_i^{(att,WCA)}(\boldsymbol{r_i},\boldsymbol{r})$.

The solutions for the electrostatic potential (φ) and the characteristic function (S) are obtained by minimizing the free energy functional and solving the equations, $\delta G/\delta\varphi = 0$ and $\delta G/\delta S = 0$. In the absence of mobile ions, the equation $\delta G/\delta\varphi = 0$ gives rise to a generalized form of the Poisson equation for computing the electrostatic potential:

$$-\nabla \cdot (\varepsilon(S)\nabla\varphi) = S\rho_m, \quad (11)$$

where $\varepsilon(S)$ is the dielectric profile of the solute-solvent system, represented in terms of the characteristic function, S, as

$$\varepsilon(S) = S\varepsilon_m + (1-S)\varepsilon_s. \quad (12)$$

Based on this definition, the dielectric function takes the value ε_m in the solute region (S = 1) and the value ε_s (S = 0) in the solvent region (S = 0). For Eulerian formulations of the geometric flow problem [6], S varies smoothly across the solute-solvent interfaces. However, the problem can also be formulated in a Lagrangian manner [13] such that S has a step-like discontinuity at the solute-solvent boundary. The current results are obtained using the Eulerian form of the problem.

The equation $\delta G/\delta S = 0$ is applied for the solute-solvent interface region and it gives rise to a nonlinear partial differential equation for the characteristic function, S, shown below:

$$-\nabla \cdot \left(\gamma \frac{\nabla S}{\|\nabla S\|}\right) + p - \rho_0 U^{att} + \rho_m \varphi - \frac{1}{2}\varepsilon_m |\nabla\varphi|^2 + \frac{1}{2}\varepsilon_s |\nabla\varphi|^2 = 0 \quad (13)$$

Chen *et al.* [6] have shown that the solution to Eq. (13) can be obtained by solving the following parabolic partial differential equation,

$$\frac{\partial S}{\partial t} = \gamma \|\nabla S\| \left[\nabla \cdot \left(\frac{\nabla S}{\|\nabla S\|}\right) + \frac{V}{\gamma}\right], \quad (14)$$

where V is the generalized flow "potential", written as,

$$V = -p + \rho_0 U^{att} - \rho_m \varphi + \frac{\varepsilon_m}{2}|\nabla\varphi|^2 - \frac{\varepsilon_s}{2}|\nabla\varphi|^2. \quad (15)$$

Eq. (14) is called the generalized geometric flow equation (GGFE) and it is coupled with the Poisson equation (Eq. (11)) via the characteristic function S.

Molecular structure parameterization

The geometric flow implicit solvent model, similar to other implicit solvent approaches, needs the coordinates, radii, partial charges, and L-J parameters of the solute atoms. This information for all the 17 molecules were taken from SAMPL0 dataset provided in the paper of Nichols *et al.* [14]. The atomic radii, charges, and well depth parameters for the L-J potential depend on the force-fields used: AM1-BCCv1/ZAP-9, AM1-BCCv1/Bondi, and OPLS-AA. These parameters are given in the Supporting Information. Note that, when specific atomic charges and radii were unavailable for a test molecule, we approximated those values as described in the Supporting Information.

We used single conformations for each test molecule. The importance of multiple conformations in implicit solvent model solvation free energy prediction has been the subject of debate. Past work by Mobley et al [20] has shown that implicit solvent energies can vary significantly with specific solute conformation. On the other hand, recent work by Nicholls et al [21, 22] suggest that the inclusion of multiple conformations does not

significantly influence the agreement between calculated and experimental solvation free energies. This paper focuses on influence of continuum model parameters (solute dielectric coefficient, solvent pressure, and surface tension); future work will focus on the impact of conformational heterogeneity on the model accuracy.

Calculation procedure

We computed the solvation free energies of the 17 molecules of SAMPL0 molecule set for a wide range of pressure and surface tension values at various solute dielectric coefficients using atomic partial charges and radii derived from 3 different force-fields. The force-fields are (1) AM1-BCCv1 charges with ZAP-9 radii (AM1-BCCv1/ZAP-9), (2) AM1-BCCv1 charges with Bondi radii (AM1-BCCv1/Bondi), and (3) OPLS-AA [23].

For each force-field and solute dielectric coefficient value, we calculated the root mean square error (χ) between the experimental (ΔG_{expt}) and calculated (ΔG_{calc}) solvation free energies for the entire 17 molecule set in the pressure-surface tension (p-γ) parameter space. A grid-based parameter optimization was performed over a range that generally included p values from 10^{-4} to approximately 0.1 kcal/(mol·Å^3) and γ values from 10^{-5} to approximately 0.1 kcal/(mol·Å^2). The exact ranges of values and search increments used to explore the p-γ parameter space for each force-field are summarized in Supplementary Information Table S1 for $\varepsilon_s = 1$. We denote the minimum value of χ in the p-γ parameter space as χ_{min}. Therefore, the (p,γ) point at which $\chi = \chi_{min}$, corresponds to the optimal values of pressure, p_{opt}, and surface tension, γ_{opt}.

While the pressure and microscopic surface tension are associated with repulsive interactions in nonpolar free energy functional, the van der Waals (vdW) dispersion interaction between solute and solvent is attractive, which may change χ_{min}, p_{opt}, and γ_{opt}. To determine the effect of the van der Waals dispersion on the values of χ_{min}, p_{opt}, and γ_{opt}, we performed calculations with and without the inclusion of the vdW term (Eq. (5)) for each force-field. We computed the vdW dispersion term in AM1-BCCv1/ZAP-9 and AM1-BCCv1/Bondi calculations based on the L-J potential, defined by Eq. (8), where the well-depth parameter value was calculated based on the formula given by Chen *et al.* [6], i.e., $\varepsilon_{is}=\left(\left(\frac{\sigma_i+\sigma_s}{\sigma_i}\right)^{12}-2\left(\frac{\sigma_i+\sigma_s}{\sigma_i}\right)^{6}\right)^{-1}$. To compute the vdW dispersion term in OPLS-AA calculations, we used the L-J potential defined by Eq. (7), where the well-depth parameter between the two oxygen water molecules (based on SPC water model [24]), $\varepsilon_s = 0.1554$ kcal/mol. The corresponding solvent radius used for these calculations was $\sigma_s = 1.5828$ Å, which is different from the much smaller solvent radius value (0.65 Å) used by Chen *et al.* [6]. Chen *et al.* noted that it is necessary to increase the atomic radii to get better agreement with experiment results. Specifically, they rescaled all the atomic radii values of the ZAP-9 parameters by an arbitrary multiplicative factor of 1.1. In our calculations, we do not rescale the atomic radii; our goal in this work is not to optimize the solute radii parameters but to determine how the optimal solvent model parameters (the χ_{min}, p_{opt} and γ_{opt} values) vary depending on the choice of different force-field parameter sets at different values of the solute dielectric coefficient.

For all of the calculations, we fixed the solvent bulk density ρ_0 at 0.0334 Å^{-3}. The closest distance between the box boundary (along each axis) and the surface of each molecule was defined as R_o, and set at 3.8 Å. The equations were solved using the second-order central finite difference scheme discussed in Chen *et al.*'s paper [6]. The grid spacing for the solver, h, was set at 0.25 Å. The time-stepping value used to solve Eq. (14) was calculated using the formula, $h^2/4.5$. Our sensitivity studies on R_o and h, based on AM1-BCCv1/ZAP-9, demonstrated that higher-resolution calculations (e.g., a three-fold increase in R_o and a 40%

decrease in h) only change the results less than 3–5%, which ensures both setting values are reasonable. We chose the solvent dielectric value ε_s as 80 and varied the solute dielectric value ε_m from 1 to 4. As expected and shown in Figure 1, the sensitivity of the results to the internal dielectric ε_m is high.

In addition to calculating the χ_{min}, p_{opt}, and γ_{opt} values for each force-field and ε_m value, we applied the F-test method [25] to compute the 95% joint confidence intervals around the optimal point (p_{opt}, γ_{opt}) in the p-γ parameter space. The F-test method is applied as follows: If $W(p,\gamma)$ represents the sum of the square of the error between the calculated and experimental free energies of over n molecules as a function of p and γ, then the confidence region, based on the F-distribution at 100(1-α)% confidence level, is given by the set of all (p,γ) points satisfying the condition:

$$\frac{W(p,\gamma)-W(p_{opt},\gamma_{opt})}{W(p_{opt},\gamma_{opt})} \leq F^{\alpha}_{d,n-d}. \quad (16)$$

In equation (16), d denotes the number of parameters to be optimized. In this work, $n = 17$, $d = 2$, and $\alpha = 0.05$. Therefore, from the F-distribution, we have $F^{0.05}_{2,15}=0.49067$ [26].

Results

We calculated the root-mean-squared hydration energy error χ for a range of solute dielectric coefficient, pressure, and microscopic surface tension values using each force field parameter set, with and without the inclusion of the vdW term in the equations. A detailed analysis of each force field at selected values of solute dielectric coefficient is presented in Supporting Information. The χ_{min}, p_{opt}, and γ_{opt} values for all the force-fields at the different solute dielectric coefficient values are summarized in Table 1. Figure 1 shows the dependence of χ_{min} on dielectric value. Among the 3 force-fields, AM1-BCCv1/ZAP-9 and OPLS-AA (with vdW) give the lowest χ_{min} values (1.82 and 1.89 kcal/mol, respectively), while the lowest χ_{min} value of AM1-BCCv1/Bondi force-field is 2.25 kcal/mol. Contour plots of χ values in the 95% confidence region of the optimal pressure and surface tension values are given in the Supporting Information figures S7–10.

Figure 2 and Figure 3 show a comparison between the calculated and experimental solvation free energies of the 17 organic molecules for AM1-BCCv1/ZAP-9, AM1-BCCv1/Bondi and OPLS-AA at the optimal value of the solute dielectric coefficient ($\varepsilon_m = 1.5$ for AM1-BCCv1/ZAP-9 and AM1-BCCv1/Bondi models, 2.0 for OPLS-AA without vdW and 3.0 for OPLS-AA with vdW). The results for the calculated solvation free energies are tabulated in the Supporting Information. The effect of vdW dispersion appears negligible for AM1-BCCv1-based force-fields but becomes appreciable for OPLS-AA. As shown in Table 1 and Figure 1, the vdW dispersion effects are clearly evident for OPLS-AA; for example at $\varepsilon_m = 2.0$, the vdW term decreases the χ_{min} value by 0.55 kcal/mol while increasing the γ_{opt} value from 0.00002 to 0.1 kcal/(mol·Å^2) and p_{opt} value from 0.0001 to 0.015 kcal/(mol·Å^2).

Discussion

Comparison of results to previous work

Recently, Nicholls *et al.* applied the Zap OpenEye implementation [11] of the Poisson-Boltzmann model, along with a surface area-based nonpolar term, to predict the hydration free energies of a set of 17 drug-like molecules (SAMPL0 data set) using AM1-BCC v1 charges and ZAP-9 radii parameterization [14]. In their calculations, the nonpolar contribution to the free energy was defined by the surface area term with a surface tension

value of 0.0064 kcal/(mol·A^2) at ε_m = 1.0. They found that the root mean square error between the calculated and experimental free energies based on ZAP-9 radii (χ = 1.87 kcal/mol) was smaller than that obtained using Bondi radii (χ = 2.57 kcal/mol).

Like the current work, Chen *et al.* also applied the geometric flow implicit solvent model described above to compute the hydration free energies of the same 17 molecule set [6]. To apply the model for computing the solvation free energies of molecules, one needs to know what values to use for the different parameters in the equations. These parameters are (1) the partial charge of each solute atom, Q_i; (2) solute atom radius, R_i; (3) solute dielectric coefficient, ε_m; (4) solvent dielectric coefficient, ε_s; (5) microscopic surface tension, γ; (6) solvent pressure, p; (7) the well-depth parameter, ε_{is}; (8) solvent radius, σ_s; and, (9) the solute atom radius defined in the L-J potential, σ_i. In their study, Chen *et al.* used the AM1-BCCv1 charges and radii values based on the ZAP-9 parameterization for the solute atoms. They obtained the coordinate, charge and radii parameters of the molecules from the structure files found in the supplementary information provided by Nicholls *et al.* [14]. However, for better agreement with experimental results, Chen *et al.* arbitrarily increased the ZAP-9 radii values by 10%. Additionally, they considered surface tension, γ, as the only fitting parameter in their model based on assumption that $\frac{p}{\gamma}=0.2$ and $\frac{\rho_0}{\gamma}=2$. For the L-J parameters, they chose 0.65 Å as a good fitting value for solvent radius, σ_s. For simplicity, they set the well depth parameter, $\varepsilon_{is}=\left(\left(\frac{\sigma_i+\sigma_s}{\sigma_i}\right)^{12}-2\left(\frac{\sigma_i+\sigma_s}{\sigma_i}\right)^{6}\right)^{-1}$. And, the solute dielectric coefficient was fixed at 1.0 in all their calculations. Based on this parameterization scheme, the χ computed from the AM1-BCCv1/ZAP-9 model was 1.76 kcal/mol and the corresponding γ value was fitted to be 0.0065 kcal/(mol·Å^2).

Our work differs from Chen *et al.* [6] in several key aspects. First of all, Chen *et al.* fixed the ratios for $\frac{p}{\gamma}$ and $\frac{\rho_0}{\gamma}$. From a physical standpoint, it is more appropriate to fix the bulk solvent (number) density at 0.0334 Å^{-3} (equivalent to 1 g/cm^3), and consider both pressure and microscopic surface tension as independent fitting parameters of the model as done in the present work. Secondly, the fitted parameter values were obtained based on scaling the solute atomic radii by 10%; therefore, we are comparing distinctly different parameter sets. Based on our parameterization scheme, the χ_{min} value computed from the AM1-BCCv1/ZAP-9 model at ε_m = 1.0 is 2.17 kcal/mol and the corresponding γ_{opt} value is 0.02 kcal/(mol·Å^2) with negligible contribution to the free energy from the pressure-volume term (see Supporting Information).

Problematic small molecules

Figure S5 in the Supporting Information shows that there are a subset of molecules which consistently show large errors with all the force-fields at ε_m = 1.0: N-dimethyl-p-methoxybenzamide, N,N-4-trimethylbenzamide, glycerol triacetate, and imidazole, as also seen in Figure S6. As mentioned earlier, Nichols *et al.* [14] developed the ZAP-9 radii by adjusting the Bondi radii values of carbon, hydrogen, carbonyl oxygen, secondary/tertiary nitrogen, sulfur, fluorine, chlorine, and iodine. As part of this optimization, the radii values for the secondary/tertiary nitrogen were reduced from 1.55 Å (Bondi) to 1.4 Å (ZAP-9). Nichols *et al.* noted that, although the ZAP-9 radii parameterization improves the overall agreement with experimental hydration energies, it can gives rise to larger differences (errors) between the computed and experimental free energies of specific molecules, such as the benzamides and esters. They attributed the errors for the benzamides to reduction of the tertiary nitrogen radius from 1.55 Å (Bondi) to 1.4 Å (ZAP-9). Our results with the geometric flow implicit solvent model also show large errors for the benzamides and

imidazole; however with reversed trends. In particular, AM1-BCCv1/ZAP-9 performs better than AM1-BCCv1/Bondi for the benzamides; however, the error for imidazole is at least 5 times larger with AM1-BCCv1/ZAP-9 and OPLS-AA compared to that with AM1-BCCv1/Bondi. The differences between our results and those of Nicholls *et al.* are not surprising and are likely due to the very different nature of the underlying implicit solvent methodology.

Comparison of Figure 3 and Figure S6 shows that the errors for the benzamides are still significant at the optimal solute dielectric coefficients determined for each force-field. However for imidazole, AM1-BCCv1/ZAP-9 performs better than AM1-BCCv1/Bondi and OPLS-AA at the respective optimal ε_m value. In particular, the magnitude of the error for imidazole reduces from 3.58 kcal/mol at $\varepsilon_m = 1.0$ to 0.05 kcal/mol at $\varepsilon_m = 1.5$ (optimal value) for AM1-BCCv1/ZAP-9 force-field while the error increases from 0.62 to 2.04 for AM1-BCCv1/Bondi force-field.

Our calculations at $\varepsilon_m = 1.0$ also exhibited large errors for glycerol triacetate, irrespective of force-field. The study of Nichols *et al.* [14] showed that esters are often problematic for implicit solvent hydration free energy prediction because of their strong hydrophilic nature; glycerol triacetate has three ester groups with a highly asymmetric structure. However, the errors for glycerol-triacetate are significantly reduced at the optimal value of the solute dielectric coefficient determined for each force-field, as seen in Figures 2 and 3, and from the tables given in the Supporting Information.

Assessing the importance of dispersion interactions

In general, the inclusion of attractive contributions to the nonpolar free energy defined by the van der Waals (vdW) dispersion is expected to improve implicit solvent models [17, 18, 27]. The inclusion of the dispersive term made no difference to the AM1-BCCv1 parameter systems, most likely because the well-depth parameter used for the L-J potential is not adequately parameterized for the two force-fields. This represents an area of future research into further optimization of the AM1-BCCv1 parameter systems. On the other hand, the inclusion of the dispersive term improved the OPLS-AA results, as expected given the availability of appropriate L-J well depth information from the force field.

Role of pressure-volume work in small molecule solvation

Table 1 indicates that the pressure-volume contribution to the nonpolar free energy for these small molecules is negligible compared to the surface area contribution, regardless of the inclusion of the vdW term for AM1-BCCv1/ZAP-9 and AM1-BCCv1/Bondi models at $\varepsilon_m = 1.0$. This directly supports the use of SASA (Solvent Accessible Surface Area) only type of nonpolar solvation models, where the pressure term is assumed to be negligible. However, it can be noted from Table 1 that the pressure-volume contribution to the non-polar free energy cannot be generally assumed to be negligible compared to the surface area contribution for ε_m values greater than 1.

Wagoner and Baker [18] have shown that the contribution from both SASA and pressure terms can provide significantly better agreement between implicit and explicit model nonpolar solvation forces. However, there are several differences between the Wagoner-Baker calculations and the present work. First, Wagoner and Baker used a traditional nonpolar model with a heuristic spline-based surface definition rather than the self-consistent geometric flow model used here. The current model provides a much improved definition of the solvent-solute interface that is coupled to thermodynamics of the model. Second, due to the size of biomolecular system under investigation, Wagoner and Baker focused on simulated solvation *forces* rather than experimental solvation *energies* to parameterize their model. Therefore, as pointed out in their paper [18], the conclusions are

dependent on the energy of the molecular dynamics force fields used for their explicit solvent simulations.

Relationship between pressure and surface tension

As illustrated in Supporting Information using the results obtained at $\varepsilon_m = 1.0$, the pressure appears to be inversely correlated with the microscopic surface tension, especially at high pressures. Such correlation may be understood by a simple qualitative argument based on interfacial energy and equation of states. Consider the transfer of a solute molecule (denoted as the phase '1') from vacuum (denoted as the phase '2') to water (denoted as the phase '3'). Interfacial energy associated with this transfer (ΔE) can be represented by $A(\gamma_{13} - \gamma_{12})$, where A is the surface area of the molecule, γ_{13} is an interfacial tension between the solute molecule and water, and γ_{12} is an interfacial tension between the solute molecule and vacuum. Applying the Dupré equation [28], one obtains:

$$\gamma_{13}=\frac{1}{2}W_{11}+\frac{1}{2}W_{33}-W_{13}, \quad (16)$$

where W_{11} (or W_{33}) is known as the work of cohesion for the phase '1' (or '3'), meaning the work done to separate a unit area from contact to infinity for phase '1' (or '3'). W_{13} is known as the work of adhesion denoting the work done to separate a unit area of two phases '1' and '3' from contact to infinity. Similarly, one obtains a corresponding relation for γ_{12}. Recognizing the surface tension as $\gamma_{13} - \gamma_{12}$ [28], the surface tension can be represented by:

$$\gamma=\frac{1}{2}(W_{33}-W_{22})+(W_{12}-W_{13}). \quad (17)$$

Note that all W_{ij} are positive owing to attractive van der Waals interactions. Since no energy involved due to no molecules in the phase '2', Eq. (17) can be simplified as:

$$\gamma=\frac{1}{2}W_{33}-W_{13}. \quad (18)$$

Following the microscopic perspective, W_{33} is associated with attractive interactions between water molecules, and W_{11} is correlated with attractive interactions between the water molecule and the solute molecule. Eq. (18) implies that an increase in the surface tension corresponds to an increase in the attractive interaction between water molecules, in comparison to that between water molecule and the solute. A connection between the attractive interactions and hydrodynamic pressure (p) in the phase '3' (i.e., aqueous phase) can be obtained from an equation of state. As the first approximation, one can use the van der Waals equation of state [29]:

$$\left(p+\frac{a}{V^2}\right)(V-b)=RT, \quad (19)$$

where V is the molar volume of the phase '3', R is the gas constant, and T is the temperature. Here, a is a coefficient reflecting the attractive interaction between water molecules (proportional to the attractive interaction) and b is related to a strong short-range repulsive interaction between water molecules. Eq. (19) states that hydrodynamic pressure decreases with increasing a (i.e., attractive interaction between water molecules). Therefore, it can be understood that increasing hydrodynamic pressure can lead to the decrease in microscopic surface tension under the equilibrium solvation process, which is generally shown in our calculations.

Surface tension value comparisons

At $\varepsilon_m = 1.0$, the γ_{opt} values obtained with AM1-BCCv1/ZAP-9, AM1-BCCv1/Bondi, OPLS-AA (without vdW dispersion), and OPLS-AA (with vdW dispersion) are 0.02, 0.03, 0.02, and 0.14 kcal/(mol·Å^2), respectively. These values obtained without the vdW dispersion term compare well to the microscopic surface tension (0.025 kcal/(mol·Å^2)) for hydrocarbons in water [30]. Furthermore, these values (obtained with and without vdW dispersion term), are comparable to popular microscopic surface tension definitions: 0.03–0.125 kcal/(mol·Å^2) [30–33]. However, these values are higher than that used by Sitkoff *et al.* [34] (0.005 kcal/(mol Å^2), Wagoner and Baker [18] (0.003–0.005 kcal/(mol·Å^2)), and Gallicchio *et al.* [19] (0.009 kcal/(mol·Å^2)), which were all obtained based on heuristic solute-solvent dielectric interface definitions.

The γ_{opt} values obtained with AM1-BCCv1/ZAP-9, AM1-BCCv1/Bondi, OPLS-AA (without vdW dispersion), and OPLS-AA (with vdW dispersion) at the respective optimal ε_m value are 0.0001, 0.01, 0.00002, and 0.08 kcal/(mol·Å^2), respectively. These values are much smaller compared to the values obtained at $\varepsilon_m = 1.0$. Among these values, the value obtained using OPLS-AA with vdW dispersion (0.08 kcal/(mol·Å^2)) compares better to the popular microscopic surface tension definitions: 0.03–0.125 kcal/(mol·Å^2) [30–33].

Conclusions

In this work, we explored the pressure-surface tension and solute dielectric coefficient parameter space to determine the optimal values of pressure and surface tension at which the root mean square error between the calculated and experimental free energies (χ) takes the minimum value (χ_{min}) for a set of 17 molecules of the SAMPL0 dataset. We performed the calculations at various solute dielectric coefficients using 3 different force-fields: AM1-BCCv1/ZAP-9, AM1-BCCv1/Bondi, and OPLS-AA. We showed that the optimal values of surface tension and pressure and the associated χ_{min} values vary depending on the solute dielectric coefficient, force-field charge and radii parameters, with AM1-BCCv1/ZAP-9 providing the best overall agreement with experiment.

In order to get better agreement between calculated and experimental hydration free energies using this model, it will be necessary to further improve the model parameters and behavior. One possible area of improvement is in the use of more sophisticated charge distributions which offer atomic polarizability or quantum-based charge distributions. Work by Schnieders *et al.* [35] describes ways in which electronic polarizability can be incorporated into implicit solvent models to provide a more sophisticated description of molecular charge distributions and dielectric response. Recently, a quantum mechanical density functional theory has been integrated into the geometric flow implicit solvent model, thus laying the foundation for more realistic solute charge density modeling [36]. Another possible area of improvement is through revised Lennard-Jones radii and well depth parameters. As discussed earlier, the AM1-BCCv1 has the best overall performance but does not benefit from the inclusion of van der Waals interactions. The current approach could provide a framework for further refining radii and optimizing well depths in the current geometric flow model.

An important advance brought forth by the geometric flow approach to implicit solvation modeling is that it avoids making *ad hoc* assumptions about the solute-solvent boundary but rather computes it in a thermodynamically consistent fashion. This reduces the number of variables and sources of potential error that can be introduced into these types of models. The results of this work provide useful insights on the parameterization of the geometric flow model for improving accuracy and predictive power and point to the next steps for continued improvement of implicit solvent modeling.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge the NIH for financial support of this work through grants R01 GM069702 and R01 GM090208 as well as Anthony Nicholls, David Mobley, and Vijay Pande for helpful comments and suggestions.

References

1. Davis M, McCammon A. Electrostatics in biomolecular structure and dynamics. Chemical Reviews. 1990; 90(3):509–521.
2. Warshel A, Papazyan A. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. Current opinion in structural biology. 1998; 8(2):211–217. [PubMed: 9631295]
3. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. Current opinion in structural biology. 2000; 10(2):153–159. [PubMed: 10753808]
4. Roux B, Simonson T. Implicit solvent models. Biophysical chemistry. 1999; 78(1–2):1–20. [PubMed: 17030302]
5. Baker, N. Biomolecular Applications of Poisson-Boltzmann Methods. In: Lipkowitz, K.; Larter, R.; Cundari, T., editors. Reviews in Computational Chemistry. John Wiley & Sons, Inc; 2005. p. 349-379.
6. Chen Z, Baker N, Wei GW. Differential geometry based solvation model I: Eulerian formulation. Journal of Computational Physics. 2010; 229(22):8231–8258. [PubMed: 20938489]
7. Dzubiella J, Swanson JMJ, McCammon JA. Coupling nonpolar and polar solvation free energies in implicit solvent models. The Journal of Chemical Physics. 2006; 124(8):084905. [PubMed: 16512740]
8. Dzubiella J, Swanson JMJ, McCammon JA. Coupling Hydrophobicity, Dispersion, and Electrostatics in Continuum Solvent Models. Physical Review Letters. 2006; 96(8):087802. [PubMed: 16606226]
9. Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. Science. 1983; 221(4612): 709–713. [PubMed: 6879170]
10. Tjong H, Zhou HX. On the Dielectric Boundary in Poisson-Boltzmann Calculations. Journal of Chemical Theory and Computation. 2008; 4(3):507–514. [PubMed: 23304097]
11. Grant A, Pickup B, Nicholls A. A smooth permittivity function for Poisson-Boltzmann solvation methods. Journal of computational chemistry. 2001; 22(6):608–640.
12. Im W, Beglov D, Roux B. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Computer Physics Communications. 1998; 111(1–3):59–75.
13. Chen Z, Baker N, Wei GW. Differential geometry based solvation model II: Lagrangian formulation. Journal of mathematical biology. 2011; 63(6):1139–1200. [PubMed: 21279359]
14. Nicholls A, et al. Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. Journal of Medicinal Chemistry. 2008; 51(4):769–779. [PubMed: 18215013]
15. Bondi A. van der Waals Volumes and Radii. The Journal of Physical Chemistry. 1964; 68(3):441–451.
16. Weeks J, Chandler D, Andersen H. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. The Journal of Chemical Physics. 1971; 54(12):5237–5247.
17. Levy R, et al. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. Journal of the American Chemical Society. 2003; 125(31):9523–9530. [PubMed: 12889983]
18. Wagoner J, Baker N. Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. Proceedings of the National Academy of Sciences. 2006; 103(22):8331–8336.

19. Gallicchio E, Kubo MM, Levy RM. Enthalpy-Entropy and Cavity Decomposition of Alkane Hydration Free Energies: Numerical Results and Implications for Theories of Hydrophobic Solvation. The Journal of Physical Chemistry B. 2000; 104(26):6271–6285.
20. Mobley DL, Dill KA, Chodera JD. Treating Entropy and Conformational Changes in Implicit Solvent Simulations of Small Molecules. Journal of Physical Chemistry B. 2008; 112(3):938–946.
21. Nicholls A, Wlodek S, Grant J. SAMPL2 and continuum modeling. Journal of Computer-Aided Molecular Design. 2010; 24(4):293–306. [PubMed: 20372975]
22. Nicholls A, Wlodek S, Grant JA. The SAMP1 Solvation Challenge: Further Lessons Regarding the Pitfalls of Parametrization. Journal of Physical Chemistry B. 2009; 113(14):4521–4532.
23. Jorgensen W, Maxwell D, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. Journal of the American Chemical Society. 1996; 118(45):11225–11236.
24. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. The Journal of Physical Chemistry. 1987; 91(24):6269–6271.
25. Vugrin, KW., et al. Confidence Region Estimation Techniques for Nonlinear Regression: Three Case Studies. Sandia National Laboratory; 2005.
26. Box, GEP.; Hunter, WG.; Hunter, JS. Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. John Wiley & Sons; 1978.
27. Su Y, Gallicchio E. The non-polar solvent potential of mean force for the dimerization of alanine dipeptide: the role of solute-solvent van der Waals interactions. Biophysical chemistry. 2004; 109(2):251–260. [PubMed: 15110943]
28. Israelachvili, J. Intermolecular & Surface Forces. San Diego: Academic Press; 1991.
29. Hirschfelder, JO.; Curtiss, CF.; Bird, RB. Molecular Theory of Gases and Liquids. New York: Wiley; 1954.
30. Sharp KA, et al. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. Science. 1991; 252(5002):106–109. [PubMed: 2011744]
31. Chothia C. Hydrophobic bonding and accessible surface area in proteins. Nature. 1974; 248(5446): 338–339. [PubMed: 4819639]
32. Eisenberg D, McLachlan A. Solvation energy in protein folding and binding. Nature. 1986; 319(6050):199–203. [PubMed: 3945310]
33. Elcock A, et al. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin1. Journal of Molecular Biology. 1999; 291(1):149–162. [PubMed: 10438612]
34. Sitkoff D, Sharp K, Honig B. Correlating solvation free energies and surface tensions of hydrocarbon solutes. Biophysical chemistry. 1994; 51(2–3):397–409. [PubMed: 7919044]
35. Schnieders M, et al. Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. The Journal of chemical physics. 2007; 126(12):124114. [PubMed: 17411115]
36. Chen Z, Wei G. Differential geometry based solvation model. III. Quantum formulation. The Journal of Chemical Physics. 2011; 135(19):194108. [PubMed: 22112067]

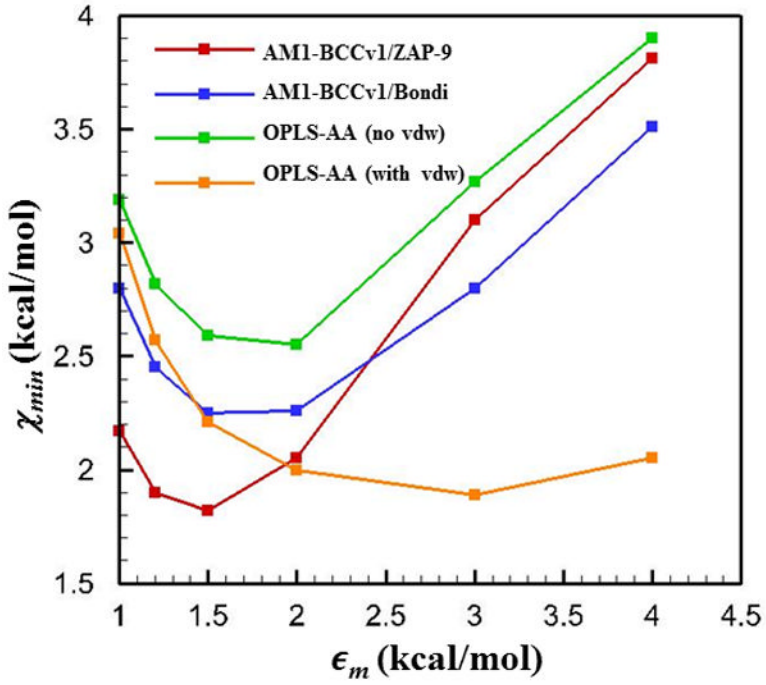

Figure 1.

The minimum root-mean-squared error $\chi_{\min}$ versus solute dielectric coefficient, ε_m obtained from the different force-field calculations: AM1-BCCv1/ZAP-9 (with or without vdw), AM1-BCCv1/Bondi (with or without vdw), OPLS-AA (without vdw), and OPLS-AA (with vdw).

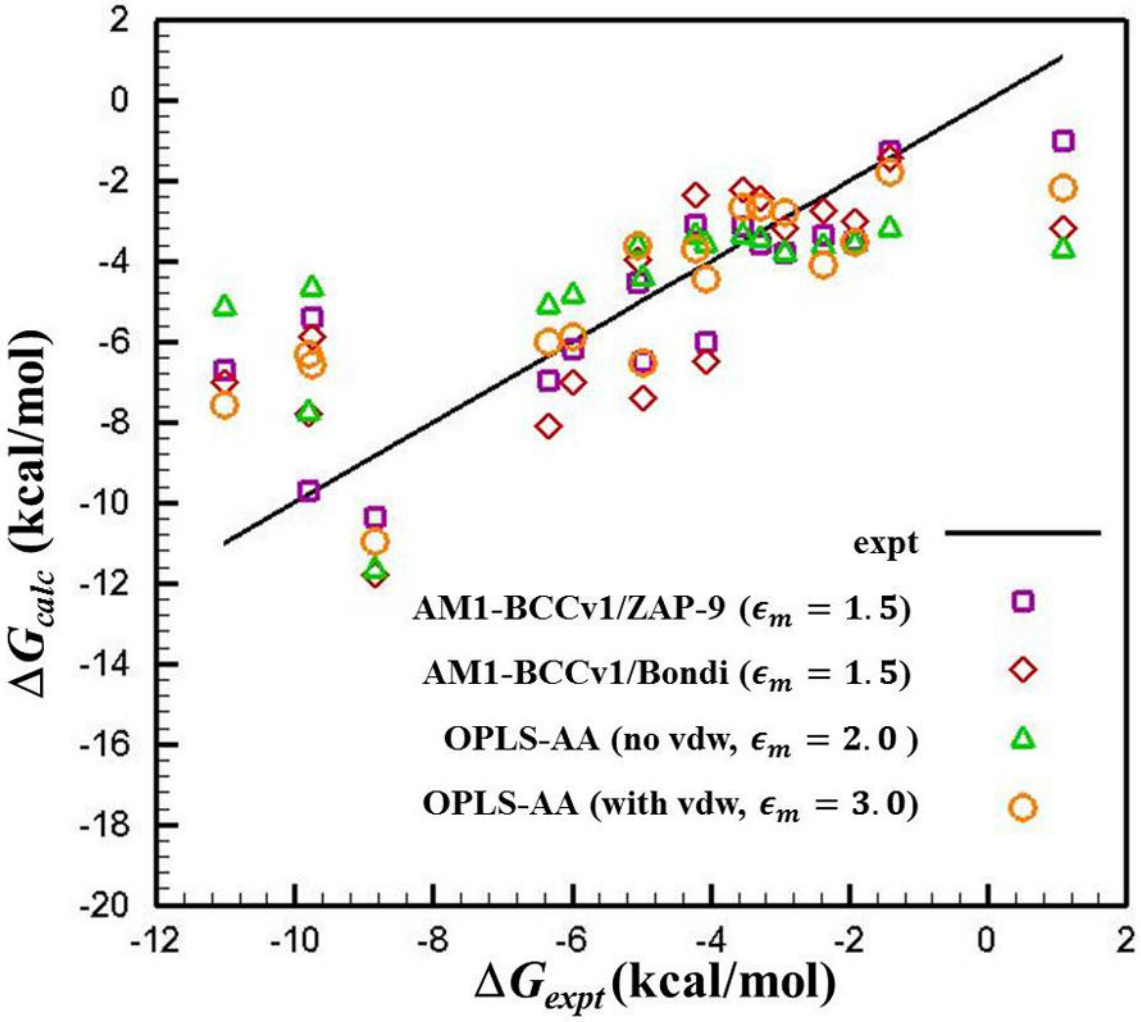

Figure 2.
Comparison of calculated and experimental hydration free energies for the 17 molecules, at the optimal values of pressure, surface tension and solute dielectric coefficient, obtained using AM1-BCCv1/ZAP-9, AM1-BCCv1/Bondi, and OPLS-AA.

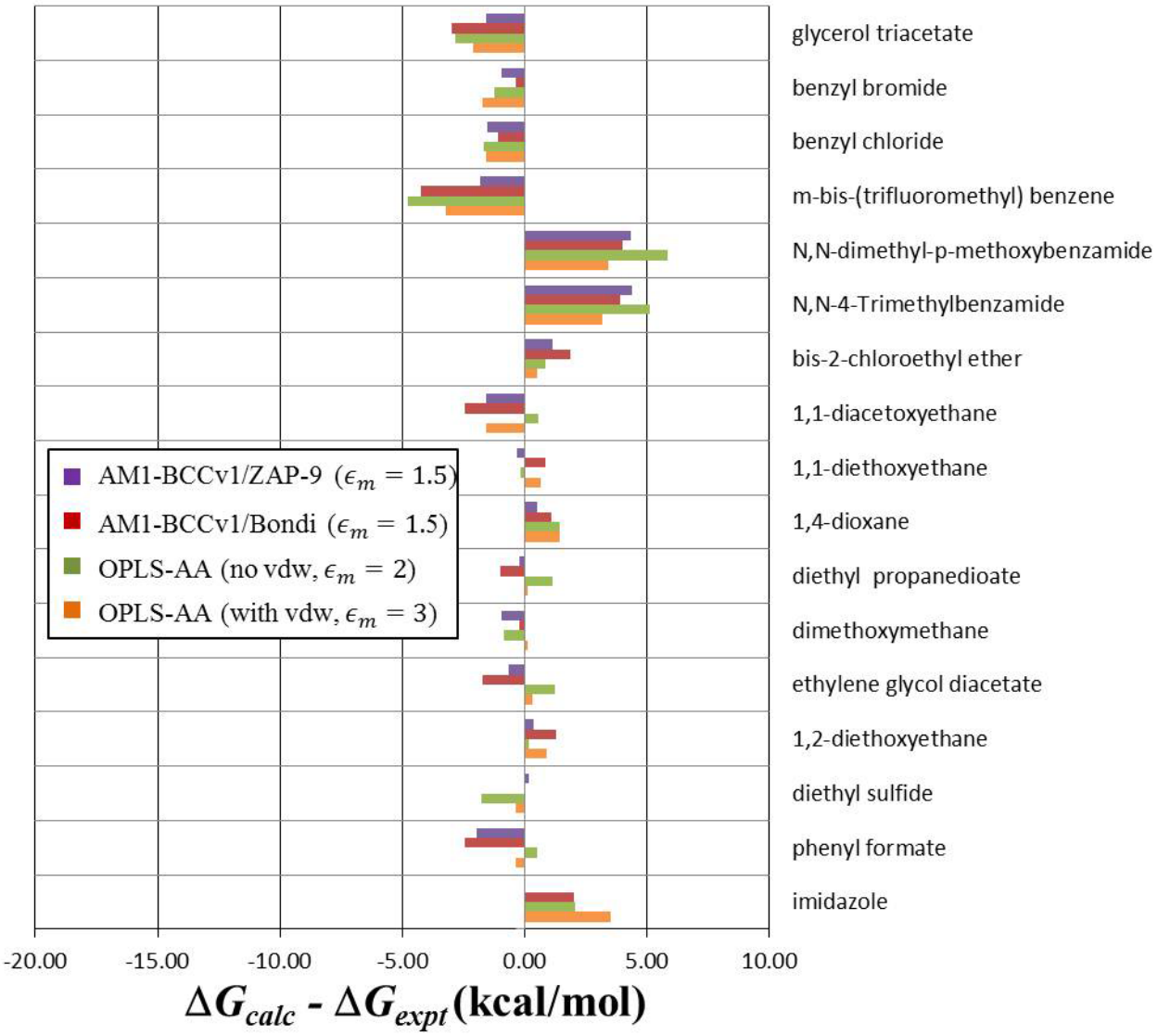

Figure 3.

Comparison of errors in the hydration free energies ($\Delta G_{calc} - \Delta G_{expt}$) of each molecule, obtained from the different force-field calculations at the optimal values of pressure, surface tension, and solute dielectric coefficient.

Table 1

The optimal values of pressure (p_{opt}) and surface tension (γ_{opt}) at $\chi = \chi_{min}$ for the three force-field parameter sets at solute dielectric coefficients (ε_m), 1, 1.2, 1.5, 2, 3, and 4.

Force-field	vdW	ε_m	p_{opt} (kcal/mol Å^3)	γ_{opt} (kcal/mol Å^2)	χ_{min}
AM1-BCCv1/ZAP-9	No	1	0.0002	0.02	2.17
AM1-BCCv1/ZAP-9	No	1.2	0.004	0.01	1.9
AM1-BCCv1/ZAP-9	No	1.5	0.008	0.0001	1.82
AM1-BCCv1/ZAP-9	No	2	0.0001	0.00001	2.05
AM1-BCCv1/ZAP-9	No	3	0.0001	0.00001	3.1
AM1-BCCv1/ZAP-9	No	4	0.0001	0.00001	3.81
AM1-BCCv1/ZAP-9	Yes	1	0.0009	0.02	2.17
AM1-BCCv1/ZAP-9	Yes	1.2	0.004	0.01	1.91
AM1-BCCv1/ZAP-9	Yes	1.5	0.009	0.00001	1.83
AM1-BCCv1/ZAP-9	Yes	2	0.0001	0.00001	2.05
AM1-BCCv1/ZAP-9	Yes	3	0.0001	0.00001	3.07
AM1-BCCv1/ZAP-9	Yes	4	0.0001	0.00001	3.78
AM1-BCCv1/Bondi	No	1	0.001	0.03	2.8
AM1-BCCv1/Bondi	No	1.2	0.0002	0.02	2.45
AM1-BCCv1/Bondi	No	1.5	0.0001	0.01	2.25
AM1-BCCv1/Bondi	No	2	0.0002	0.002	2.26
AM1-BCCv1/Bondi	No	3	0.0001	0.00001	2.8
AM1-BCCv1/Bondi	No	4	0.0001	0.00001	3.51
AM1-BCCv1/Bondi	Yes	1	0.001	0.03	2.8
AM1-BCCv1/Bondi	Yes	1.2	0.0001	0.02	2.45
AM1-BCCv1/Bondi	Yes	1.5	0.0001	0.01	2.26
AM1-BCCv1/Bondi	Yes	2	0.0001	0.002	2.26
AM1-BCCv1/Bondi	Yes	3	0.0001	0.00001	2.78
AM1-BCCv1/Bondi	Yes	4	0.0001	0.00001	3.48
OPLSAA	No	1	0.001	0.02	3.19
OPLSAA	No	1.2	0.0001	0.015	2.82

Force-field	vdW	ε_m	p_{opt} (kcal/mol Å^3)	γ_{opt} (kcal/mol Å^2)	χ_{min}
OPLSAA	No	1.5	0.0001	0.005	2.59
OPLSAA	No	2	0.0001	0.00002	2.55
OPLSAA	No	3	0.0001	0.00001	3.27
OPLSAA	No	4	0.0001	0.00001	3.9
OPLSAA	Yes	1	0.0001	0.14	3.04
OPLSAA	Yes	1.2	0.0003	0.13	2.57
OPLSAA	Yes	1.5	0.035	0.095	2.21
OPLSAA	Yes	2	0.015	0.1	2
OPLSAA	Yes	3	0.03	0.08	1.89
OPLSAA	Yes	4	0.03	0.075	2.05