Calculated Hydration Free Energies of Small Organic Molecules Using a Nonlinear Dielectric Continuum Model

Lars Sandberg, Rickard Casemyr, and Olle Edholm*

Theoretical Biophysics, Department of Physics, Royal Institute of Technology, Stockholm Centre for Physics, Astronomy and Biotechnology, SE-106 91 Stockholm, Sweden

Received: February 14, 2002; In Final Form: May 1, 2002

Prediction of solvation free energies is an important subject in fundamental natural science but also important to the pharmaceutical and food industry. A popular modeling approach is to treat the solution by an implicit solvent model. The solute molecule is rigid with a fixed effective charge distribution localized at the atomic nuclei positions. The hydration free energy is described by the van der Waals energy, the solute cavity formation energy in the water phase, and the change in electrostatic solute−solvent interaction energy. The dielectric continuum is generally assumed to be a simple medium, that is, linear, homogeneous, and isotropic. However, this approximation is quite severe and will give too hydrophilic solvation free energies. We show here that the simple medium approximation must be relaxed and nonlinearity must be taken into consideration. In strong electric fields, the solvent polarization becomes saturated and the dielectric no longer responds linearly in the applied field. This effect is well-described by the modified Langevin−Debye model. This nonlinear solvation model is used to study the hydration of 181 small organic molecules. Atomic charges and radii of the solute molecule are described by a standard classical force field. We apply the optimized potentials for liquid simulation all atom (OPLS-AA) force field, which is parametrized to reproduce both structural and thermodynamical data. This leads to a mean unsigned error of 0.6 kcal/mol, which is a 25% improvement compared to a simple medium approach. The nonlinear solvation model is further improved by introducing a few charge-scaling parameters for some functional groups that show a systematic deviation from their experimental data. This yields a mean unsigned error of 0.4 kcal/mol, which is only twice the experimental uncertainty. Hence, we conclude that nonlinear dielectric effects are indeed important to incorporate in implicit solvent models, even for neutral polar molecules.

1. Introduction

Theoretical models that predict solvation free energies of small organic molecules are of greatest significance in drug design, in biomolecular folding and binding, and in improving force field parameters used in computer simulations. There are two principal branches of modeling that treat the solvent either by an explicit solvent representation[1−6] or by an implicit solvent continuum approximation.[7−17] However, there also exist statistical approaches using quantitative structure−activity or −property relationships (QSAR/QSPR), which yield very good solvation predictions.[18] Another division is possible into models that do or do not include nonlinear dielectric response (NLR) effects. Explicit models inherently incorporate the nonlinearity, but implicit dielectric continuum models usually ignore this effect.[7−13,17]

Dielectric continuum theory is very popular to use in solvation modeling because it is computationally fast and considered to be an uncomplicated "down-to-earth" method. The most common approach is to treat the solvent as a *simple medium*, meaning that it is a linear, homogeneous, and isotropic medium. The relative permittivity of a simple medium is a constant. We will, in this work, relax the first assumption and treat the solvent as a nonlinear, homogeneous, and isotropic medium. The nonlinear dielectric response mainly consists of dielectric saturation and electrostriction. The dielectric saturation effect stems from the fact that when the thermally fluctuating dipoles of a dielectric medium are exposed to an electric field they tend to align in the field. If the field strength increases, the molecular dipoles become saturated and the medium orientational polarization has reached its upper limit. The relative permittivity is no longer a constant but becomes a function of the applied electric field. This phenomena is called normal saturation[19] and becomes significant in water for electric field strengths of about 10^7 V/m.[20,21] Even though this corresponds to very strong experimental fields, such field strengths are evident in the local vicinity of solvated ions, polar molecules, and charged groups in macromolecules and in biological bilayers. Moreover, if an electric field is applied at a constant pressure to a dielectric, the medium will feel an elastic strain. This phenomena is called electrostriction and will cause compression of the solvent liquid and therefore increase the density and relative permittivity. However, this effect is considered to be less important than the saturation effect.[19]

We have, in a recent work, concluded that the NLR contribution for monovalent alkali ions is less than 2.5 kcal/mol,[22] which corresponds to about 2% of the total hydration free energy. Hence, one might suspect that the NLR effect ought to be negligible for neutral molecules. However, we will show that the NLR effect is even more important for neutral polar molecules. The reason for this is that classical force field charge parameters are of the same magnitude as that of monovalent

* To whom correspondence should be addressed. Electronic mail: oed@theophys.kth.se.

10.1021/jp020434b CCC: $22.00
Published on Web 07/18/2002

ions (see, e.g., the optimized potentials for liquid simulation all atom force field, OPLS-AA[3,4,23−25]). Hence, we can expect to have NLR contributions similar in magnitude also for neutral polar molecules. Even though this effect is negligible for ions, which have a hydration free energy in the range of −120 to −60 kcal/mol, it will not be negligible for neutral small organic molecules, which are found to have hydration free energies between −10 and 5 kcal/mol. The electrostriction work was numerically calculated by the effective pressure model of Desnoyers, Verrall, and Conway[26] for the ions and was found to be at least 10 times smaller than the saturation contribution, that is, less than 0.25 kcal/mol.[22] Hence, we feel that it is safe to disregard the electrostriction term for small organic molecules.

Sitkoff, Sharp, and Honig have developed a continuum solvation model that solves the Poisson equation (finite difference Poisson−Boltzmann, FDPD, method) to calculate the electrostatic energy to which they add a nonpolar free energy term (cavity/van der Waals) equal to the surface tension coefficient, γ, times the solute's solvent accessible surface area.[8] This is a linear dielectric response (LR) model. The FDPB/γ model with a dielectric constant of 1 indicates that the calculated hydration free energies of polar molecules with the OPLS force field becomes too negative with a LR model. Most standard force fields are parametrized with an explicit solvent in a vacuum with a dielectric constant of 1. Sitkoff, Sharp, and Honig circumvent this by introducing their own force field called the PARSE parameter set with optimized atomic van der Waals radii and partial charges for the different functional groups. Furthermore, they introduce what they claim to be the solute electronic polarizability by using a dielectric constant of two inside the molecule. We find it doubtful to use an internal screening of two inside the solute molecule because the dielectric constant is a macroscopic averaged quantity. Even though the solute is distinct from the solvent, the latter still is dominant and ought to be the main contributing factor to the dielectric constant at each position in the solute−solvent system, at least for small solute molecules. This shows, however, that it is indeed possible to parametrize a LR model to give good hydration free energies, but at the cost of using a nonstandard, model-specific force field for the solute molecules.

The continuum treatment of solvent polarization has sometimes been combined with a semiempirical self-consistent field molecular orbital (SCF-MO) method to model the solute. The many comprehensive parametrizations of Cramer, Truhlar, and co-workers,[10−13] called the SM*x* models, use a large training set to fit the extensive set of adjustable parameters used in their solvation models. The electrostatic energy is calculated with the approximative generalized Born model, which still is a LR approach. The nonpolar free energy is estimated by a cavity surface tension approach involving atomic solvation parameters and atomic solvent accessible surface areas. This multiparameter model gives good correlation with available experimental data, at least for the over 200 solutes included in their fitting procedure (training set).

Our modeling philosophy is based on the use of an existing well-parametrized classical force field and further improvement of the physical solvation model by incorporating nonlinear solvation effects into the implicit solvent continuum model. We choose the OPLS-AA force field,[3,4,23] which has the advantage that it is parametrized for explicit solvent simulations to reproduce both thermodynamical and structural data. This is a standard force field without adjustable parameters. Furthermore, we aim at introducing as few additional solvation model parameters as possible. This corresponds to only four adjustable parameters in the original unmodified model and 13 in the modified optimized version. Our "training set" is also limited to only eight molecules in the unmodified version and 55 in the modified version. Because we have a total of 181 solute molecules, this leaves us with quite a large "test set" for our model. Because we apply so few model parameters, we may further improve our model by adding carefully selected (if) needed parameters in the future.

2. Theory and Methods

The hydration free energy of solute molecules at a certain temperature in aqueous solution is related to the partition coefficient between a water and a gas phase. We will assume that the solute molecule is represented by a rigid molecular structure and a fixed effective partial charge distribution, which is localized at the atomic nuclei positions of the molecule during the transfer. Rigid structures have been most useful for ab initio calculations and flexible structures do in most cases provide little additional important information.[12] We have used the CORINA 3D structure generator to construct the solute structures.[27] The structures were not relaxed in the applied force field but used in their original form.

The hydration free energy is described by a sum of the change in electrostatic solute−solvent interaction energy, the van der Waals energy in the water phase, and the solute cavity formation energy in the water phase, that is,

$$\Delta G_{\mathrm{hyd}} \doteqdot \Delta G_{\mathrm{e}} + \Delta G_{\mathrm{vdW}} + \Delta G_{\mathrm{cav}} \tag{1}$$

We use the standard states of 1 mol/L ideal gas for the gas phase and 1 mol/L ideal aqueous solution for the liquid phase. We apply the OPLS-AA force field[3,4] for the solute partial charges and Lennard-Jones parameters and the TIP4P water model for the solvent Lennard-Jones parameters. The Lennard-Jones parameters for the commonly applied water models are very similar, for example, see SPC, TIP3P, TIP4P, and TIP5P parameters.[24,25] The advantage of the OPLS-AA force field is, as stated before, that it is parametrized to reproduce both thermodynamical and structural data.

2.1. The Electrostatic Interation Energy. All electrostatic expressions presented in this work are represented in the Système International d'unités (SI) except for the energy unit for which we will use kcal/mol (1 kcal/mol is equal to 4.184 kJ/mol). The solvent is represented in the dipole approximation, but the solute is described classically by a localized charge distribution at the atomic nuclei positions. The solute atoms have an atomic radius defined by the Lennard-Jones parameters of the OPLS-AA force field. Hence, $r_i = 0.5\sqrt{\sigma_i \sigma_{\mathrm{w}}}$, where σ_i is the Lennard-Jones parameter of solute atom i and σ_{w} is the solvent Lennard-Jones parameter. The solute volume, V_{s}, is calculated from the solute atomic radii, and the solvent volume, V, is defined as the region exterior to the solute.

The electrostatic energy of the bulk solvent in the electric field, E, of the solute is

$$W_{\mathrm{e}} = \int_V \mathrm{d}^3 r \int_0^E \epsilon_0 \epsilon(E) E \, \mathrm{d}E \tag{2}$$

In the case of linear dielectric response, the static permittivity function, $\epsilon(E)$, is a constant equal to the dielectric constant, ϵ_{r}, of the medium. This is true for a dielectric system to which a weak electric field is applied. However, in the case of strong electric fields, the building blocks of the dielectric medium, that is, the molecular dipoles of the solvent, become saturated and nonlinear dielectric response effects will emerge. The dielectric

TABLE 1: The Hydration Model Parameters Used in the Vacuum/Water Transfer Calculations, See Also Sandberg and Edholm[22]

temperature	$T = 298$ K
water density	$n = 3.33 \times 10^{28}$ m^{-3}
electronic permittivity	$\epsilon_\infty = 1.77$
water dielectric constant	$\epsilon_r = 78.5$
water vacuum dipole moment	$\mu = 1.85$ D
Onsager correction factor	$C_F = 0.0695$
Kirkwood correlation factor	$g = 2.80$
SPT coefficients	$K_0 = 10.3$ kcal mol^{-1}
	$K_1 = -99.5$ kcal mol^{-1} nm^{-1}
	$K_2 = 363.9$ kcal mol^{-1} nm^{-2}
	$K_3 = 26.7$ kcal mol^{-1} nm^{-3}

constant becomes a function of the electric field called the static permittivity function, $\epsilon(E)$, and is here modeled by the modified Langevin−Debye (mLD) model:[22]

$$F - \frac{n\mu g(\epsilon_\infty + 2)}{9\epsilon_0} L(\mu\beta C_F F) = \frac{\epsilon_\infty + 2}{3} E \tag{3}$$

$$\frac{\epsilon(E) - 1}{\epsilon(E) + 2} = \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} + \frac{n\beta\mu^2 g C_F}{3\epsilon_0} L'(\mu\beta C_F F) \tag{4}$$

where F is the local field, n is the solvent molecular number density, μ is the solvent molecular vacuum dipole moment, g is the Kirkwood correlation factor, ϵ_∞ is the square of the solvent refractive index, $\beta = 1/(k_B T)$, $C_F = 3\epsilon_r(\epsilon_\infty + 2)/((\epsilon_r + 2)(2\epsilon_r + \epsilon_\infty))$ is the Onsager correction factor, and $L(x) = \coth(x) - 1/x$ is the Langevin function with its derivative $L'(x) = -\operatorname{csch}^2(x) + 1/x^2$. For the derivation and the details of the modified Langevin−Debye model, we refer to Sandberg and Edholm.[22] The parameter values for water are found in Table 1. For a given electric field, E, the local field, F, obtained in a self-consistent way by eq 3, is inserted in eq 4 yielding the static relative permittivity, $\epsilon(E)$. The problem that remains is to determine the electric field strength, $E(\mathbf{r})$, in the solvent in which the solute is embedded.

The solute−solvent system has, in general, no symmetry that can be used to simplify the problem. However, the electric field fulfills Maxwell's macroscopic equations

$$\nabla\cdot\mathbf{D} = \rho(\mathbf{r}) \tag{5}$$

$$\nabla \times \mathbf{E} = 0 \tag{6}$$

where the displacement field, $\mathbf{D}$, is related to $\mathbf{E}$ and the polarization, $\mathbf{P}$, by the constitutive relation

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}(\mathbf{E}) \tag{7}$$

The electric field can, in principle, be solved from these three equations. However, this system of first-order vector partial differential equations (PDE) can be formulated into one second-order scalar PDE. According to Maxwell's eq 6, the electric field is irrotational, and hence, there exists a potential, ϕ, such that $\mathbf{E} = -\nabla\phi$. If we use this in eqs 7 and 5, we find the following scalar potential PDE:

$$-\epsilon_0\nabla^2\phi + \nabla\cdot\mathbf{P}(-\nabla\phi) = \rho(\mathbf{r}) \tag{8}$$

Because $\mathbf{P}(-\mathbf{E}) = -\mathbf{P}(\mathbf{E})$ for isotropic liquids, we may series expand the polarization into

$$\mathbf{P}(\mathbf{E}) = \epsilon_0(\chi_e^{(1)}\mathbf{E} + \chi_e^{(3)}E^2\mathbf{E} + \mathcal{O}(E^4\mathbf{E})) \tag{9}$$

and find that the polarization is always parallel with the electric field, $\mathbf{E}$. Therefore, we can write $\mathbf{P} = P\mathbf{e}$, where $\mathbf{e}$ is a unit vector parallel with $\mathbf{E}$. In weak fields, the first series coefficient, $\chi_e^{(1)}$, is the weak-field electric susceptibility and $\chi_e^{(1)} \equiv \epsilon_r - 1$. The polarization P consists, according to the modified Langevin−Debye model, of an orientation and a deformation term:[22]

$$P = n\mu g L(\mu\beta C_F F) + n\alpha F = n\mu g L(\mu\beta C_F F) + 3\epsilon_0\frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}F \tag{10}$$

where we have used the Lorentz−Lorenz relation in the final step to substitute the molecular polarizability, α. The electronic contribution is assumed to show linear response, but the orientational polarization of the solvent molecules is described by a nonlinear response model via the Langevin function. The local field F is again obtained self-consistently by eq 3.

The above theory yields a continuum nonlinear dielectric response model for the solvent. In the weak-field limit where $E \to 0$, the theory turns into a linear dielectric response model with the dielectric constant ϵ_r. This linear dielectric response analogue is easily obtained by only keeping the first term in eq 9 with $\chi_e^{(1)} \equiv \epsilon_r - 1$, which together with the constitutive relation eq 7 and Maxwell's first eq 5 results in the ordinary Poisson equation

$$\nabla^2\phi = -\frac{\rho(\mathbf{r})}{\epsilon_0\epsilon_r} \tag{11}$$

To calculate the spatial dependence of the electrostatic potential, which will yield the electric field, we apply the finite difference method (FDM) and numerically solve

$$\nabla^2\phi = -\frac{\rho'(\mathbf{r})}{\epsilon_0} \tag{12}$$

where $\rho' \doteqdot \rho - \nabla\cdot\mathbf{P}$. The polarization P is given by eq 10, and thus, the polarization model is a nonlinear dielectric response model. The solute molecule is placed in both a vacuum and water. We use a simple cubic finite difference grid to describe the solute/solvent system with a grid spacing, $a = 0.05$ nm. This grid spacing gives that an explicit water molecule of the solution is covered by about 240 grid points. The grid size is determined by the solute molecular dimensions to which a solvent layer of 1.2 nm is added. Outside this layer, the solvent is assumed to show linear dielectric response. We apply Coulombic boundary conditions and set $\phi = 0$ initially at all inner node points. The partial charges, which electrostatically describe the solute molecule, are spread on nearest neighboring grid points. We assume that the system is homogeneous and that there is no solute/solvent boundary. The relative permittivity at each grid point is equal to the solvent permittivity. However, the water solvent becomes saturated close to the charges, which will lead to a low permittivity at grid points inside the solute volume, V_s. The finite difference system is then iterated by the successive overrelaxation method (SOR),[28]

$$\phi_{i,j,k}^{(n+1)} = (1 - \omega)\phi_{i,j,k}^{(n)} + \frac{\omega}{6}(\phi_{i-1,j,k}^{(n+1)} + \phi_{i,j-1,k}^{(n+1)} + \phi_{i,j,k-1}^{(n+1)} + \phi_{i+1,j,k}^{(n)} + \phi_{i,j+1,k}^{(n)} + \phi_{i,j,k+1}^{(n)}) - \frac{\omega}{6}a^2\rho'_{i,j,k} \tag{13}$$

where the superscript represents the iteration step index. The overrelaxation parameter is set to $\omega(\text{vacuum}) = 3/2$ and $\omega(\text{water})$

$= {}^1/_{20}$ (i.e., underrelaxation). The iteration continues until the energetic contribution of five successive iterative steps falls below 0.1% in water and 0.01% in a vacuum, similar to about 80 and 120 iteration steps per molecule. This corresponds to a CPU time on the order of minutes on a personal computer or work station. The solution yields the electrostatic potential at every grid point.

The electrostatic energy of the solvent (the solute grid points excluded) inside a sphere with diameter equal to the grid size is calculated by eq 2 using $\mathbf{E} = -\nabla\phi$. Bulk corrections to the solvation free energy originating from the solvent outside the sphere are calculated using the Born equation,

$$\Delta G^{\text{Born}} = -\frac{Q^2}{8\pi\epsilon_0 r_{\text{sphere}}}\left(1 - \frac{1}{\epsilon_r}\right) \tag{14}$$

where Q is the solute charge. The electrostatic contribution to the hydration free energy is

$$\Delta G_e = W_e(\text{water}) - W_e(\text{vacuum}) \tag{15}$$

2.2. The van der Waals Interaction Energy. For the present, we assume that the solvent consists of a uniform number density of molecules, so the radial distribution function is approximated by a step function, which leads to

$$\Delta G_{\text{vdW}} = \int_V \mathrm{d}^3 r \sum_{i=1}^{N} n u_i(\mathbf{r}) \tag{16}$$

where V is the solvent, N is the number of atoms of the solute, and $u_i(\mathbf{r})$ is the potential energy. The latter is described by an effective (12,6) Lennard-Jones pair potential term

$$u_i(\mathbf{r}) = 4\epsilon_{iw}\left(\left(\frac{\sigma_{iw}}{r}\right)^{12} - \left(\frac{\sigma_{iw}}{r}\right)^{6}\right) \quad \text{and} \quad r = |\mathbf{r} - \mathbf{r}_i| \tag{17}$$

with atomic OPLS-AA parameters $\{\epsilon_i, \sigma_i\}_{i=1}^{N}$, solvent TIP4P water parameters $\{\epsilon_w, \sigma_w\}$, and OPLS standard combining rules, $\epsilon_{iw} = \sqrt{\epsilon_i \epsilon_w}$ and $\sigma_{iw} = \sqrt{\sigma_i \sigma_w}$.

2.3. The Cavity Formation Energy. According to scaled particle theory (SPT),[29,30,31] the cavity formation energy can be expanded into

$$\Delta G_{\text{cav}} = K_0 + K_1 r + K_2 r^2 + K_3 r^3 \tag{18}$$

where r is the solute cavity radius, which is assumed to be larger than the solvent molecular radius. SPT gives the theoretical expressions for the four model coefficients $\{K_0, K_1, K_2, K_3\}$. We will, however, treat these coefficients as model parameters. These parameters will depend on the selected force field and are adjusted by fitting the calculated *n*-alkane (methane to octane) hydration free energies in a least-squares fit to the experimental data (see Figure 1). The hydration free energy for these alkanes is largely determined by balancing the van der Waals free energy contribution (force-field dependent) with the cavity formation free energy. The cavity radius is calculated from the solute volume, V_s, in a spherical approximation, that is, $r = \sqrt[3]{3V_s/4\pi}$. The solute volume is determined by a Monte Carlo method approach using the solute atomic positions and radii.

3. Results and Discussion

We have created a database of 181 small organic molecules ranging in size from 3 to 29 atoms, containing mostly mono- and bifunctional compounds. Experimental hydration free energies are mainly taken from Li et al.,[13] but some data were also found elsewhere.[5,18,32,33] The uncertainty of the experimental partition coefficients increases with magnitude, and the detection limit is exceeded when $|\Delta G| \gtrsim 12$ kcal/mol.[1] Cramer, Truhlar, and co-workers estimate the uncertainty in experimental values to be about 0.2 kcal/mol.[13]

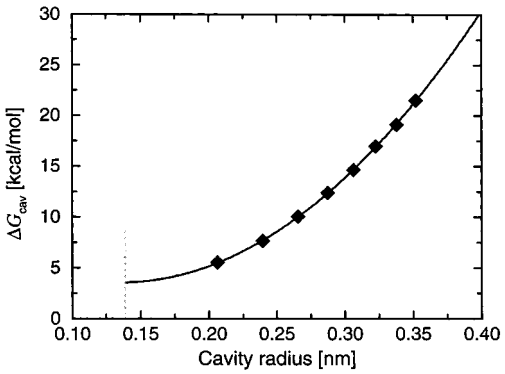

Figure 1. The cavity formation free energy data of the *n*-alkanes (methane to octane) is fitted by the method of least squares to the experimental data as a third-order polynomial in the cavity radius. The scaled particle theory coefficients are tabulated in Table 1.

To estimate the electrostatic discretization uncertainty due to the grid, we made two independent calculations for each solute. In the second calculation, the grid origin was translated by one-half of a grid spacing along each coordinate. The outcome closest to the experimental value was chosen, and the difference between the two calculations was taken as a measure of the electrostatic mean error. This error is less than the experimental uncertainty when $|\Delta G_e| \leq 4.0$ kcal/mol. We conclude that the average electrostatic grid uncertainty is less than 10% of the calculated electrostatic energy.

Our purpose is to examine how nonlinear effects influence calculated hydration free energies of small organic molecules. Hence, we apply both a linear response model (LR) and the nonlinear response model (NLR), called the modified Langevin−Debye model. The molecule, represented by a fixed effective partial charge distribution over the rigid structure of the atomic nuclei positions, is embedded in the dielectric continuum with a dielectric constant in the LR model. However, in the NLR model, the dielectric medium may become saturated in the field of the solute molecule.

First, we examined the 31 solute molecules in our data set that are well-parametrized in the OPLS-AA force field. The result is shown in Figure 2 and in Table 2. The calculated hydration free energy shows both positive and negative mean signed error (MSE). For positive errors, both models agree, but for negative MSE, the NLR model always gives a less negative and better estimate. The difference may be as large as 2.6 kcal/mol (TIP4P water). The average MSE LR−NLR difference is 23.7% for compounds with a LR negative MSE. Figure 3 shows the electricostatic potential field in space around hydrated phenol in aqueous solution. Note further the saturation effect, which is expressed in the contour plots of the relative permittivity. It extends several angstroms away from the hydroxyl group location. Hence, the nonlinear dielectric response is an important effect to include in the continuum model.

Second, we calculated the hydration free energies of all 181 solute molecules in our data set. The result is presented in Table 3 and in Figure 4. The original NLR model improves the model accuracy with 20−25% in the mean unsigned error (MUE) and root-mean-square displacement (rmsd). Although already the

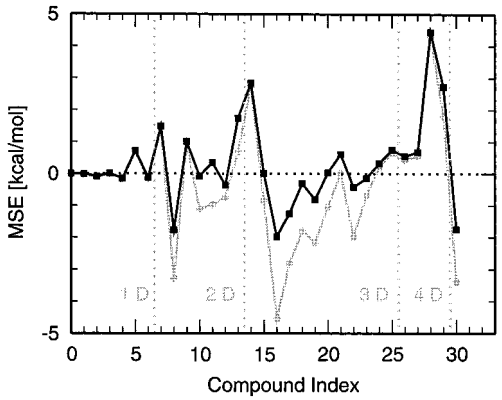

Figure 2. The 31 compounds that are well-parametrized in the OPLS-AA force field[3,4,23,,24,25] and represented in our solute data set. The black squares are the calculated MSEs of the NLR model, and the gray circles correspond to the calculated MSE of the LR model. The compounds are sorted by their calculated dipole moment. Note that when the MSE is positive the two models give similar results but when the MSE is negative the NLR model always yields a better calculated hydration free energy.

TABLE 2. Molecules that Are Well-Parametrized in the OPLS-AA Force Field[3,4,23,24,25] and for Which the Calculated Dipole Moment Is Larger Than 1 D

name	ΔG_{hyd}^{expt} [a]	μ_{gas}^{expt} [b] [D]	μ^{calcd} [D]	MSE[c] LR	MSE[c] NLR
water SPC	−6.31	1.85	2.27	−2.81	−1.25
water TIP3P	−6.31	1.85	2.29	−2.16	−0.81
water TIP4P	−6.31	1.85	2.29	−4.54	−1.98
water TIP5P	−6.31	1.85	2.29	−1.80	−0.31
methanol	−5.11	1.70	2.37	−1.04	0.03
ethanol	−5.01	1.69	2.44	−0.67	−0.14
1-propanol	−4.83	1.68	2.44	−2.00	−0.42
tert-butyl alcohol	−4.51		2.43	0.02	0.61
phenol	−6.62	1.22	2.17	−0.85	0.01
ammonia	−4.29	1.47	1.86	−1.11	−0.08
methylamine	−4.56	1.31	1.92	−0.97	0.35
dimethylamine	−4.29	1.01	2.00	0.68	1.74
trimethylamine	−3.23	0.61	2.03	2.74	2.84
ethanamide	−9.71	3.76	4.15	−3.40	−1.75
N-methylethanamide	−10.08		3.81	1.81	2.72
N,*N*-dimethylethanamide	−8.55		3.50	4.31	4.44
acetic acid	−6.70	1.70	1.78	−3.29	−1.77
ethanal	−3.50	2.75	2.78	0.19	0.32
propanal	−3.44	2.52	2.78	0.64	0.75
propanone	−3.85	2.88	3.04	0.43	0.55
butanone	−3.64	2.78	3.04	0.56	0.68
dimethyl ether	−1.92	1.30	1.82	0.96	1.01
diethyl ether	−1.76	1.15	1.65	1.40	1.47
ethanethiol	−1.30	1.58	1.97	−0.74	−0.37
average MSE[d]				−0.85	0.06
rmsd[d]				1.85	1.13

[a] Free energies are reported in kcal/mol. [b] Experimental gas-phase dipole moments from ref 34. [c] MSE = $\Delta G_{hyd}^{calcd} - \Delta G_{hyd}^{expt}$. [d] Average MSE and rmsd not including trimethylamine and *N*,*N*-dimethylethanamide. These two compounds have similar large positive MSE, which will drown the remaining data. Hence, they will not contribute to the understanding of how the LR and NLR model differ.

LR shows a good correlation with experimental data, the linear regression slope is much improved by introducing nonlinearity into the implicit solvent model. Note further the change in MSE from −0.3 to 0.0 kcal/mol. The difference in the calculated hydration free energy between the LR and NLR models exceeds the experimental uncertainty limit for 30% of the 181 solute molecules in the data set. For 15% of the solutes, the difference exceeds 1 kcal/mol. This shows the importance of incorporating

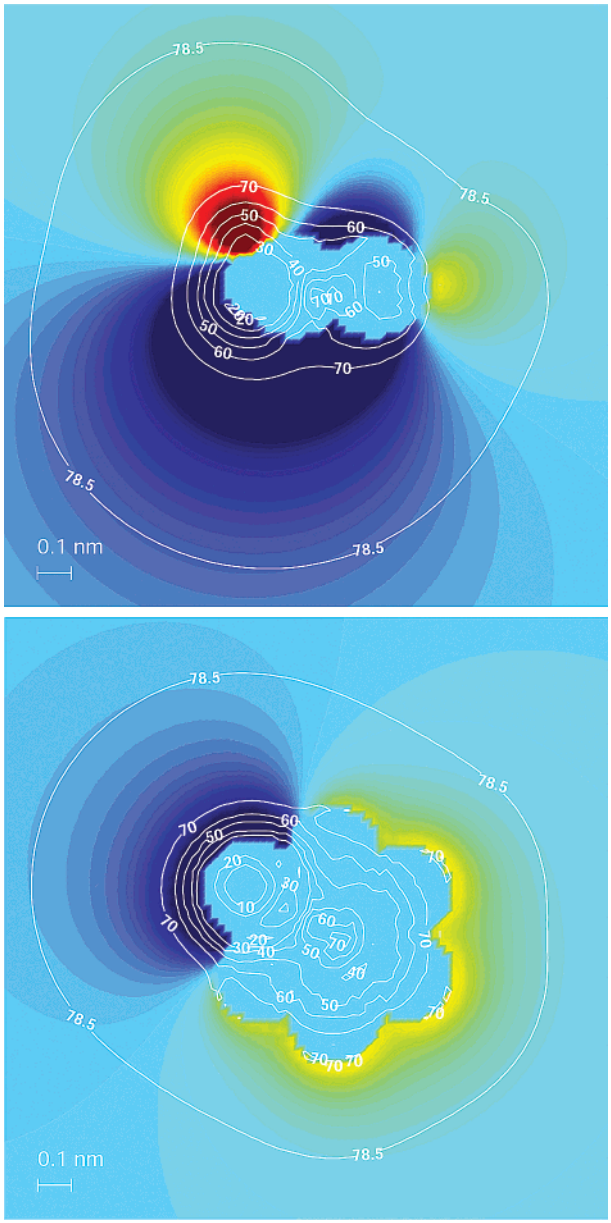

Figure 3. Hydration of a phenol molecule in water seen from the side (top) and from above (bottom). The colors represent the electrostatic potential in space where the color interval goes from blue (negative potential) to red (positive potential). The contours depict the spatial dependence of the relative permittivity of water. The flat blue silhouette of the benzene ring structure is clearly visible (see bottom figure) with the hydroxyl group hydrogen pointing up from the plane of the benzene ring (see top figure). Please note that the permittivity of the water near the hydroxyl group is saturated several angstroms away from the solute molecule (see the contour plots).

the nonlinear dielectric response effect into the dielectric continuum model.

If the different functional groups are examined separately, we find that some groups deviate more than others, see Table 4. This led us to introduce new solvation model parameters to scale the functional group charge distribution for alkynes, ethers, amines, aldehydes, ketones, carboxylic acids, nitroalkanes, nitriles, and chloroalkanes (cf. the PARSE parameter set[8]). We adjusted these model parameters to obtain an optimized NLR solvation model by minimizing the alkyl family rmsd for each functional group [$CH_3(CH_2)_n$−X where $n = 0, 1, ..., N_X$ and X is the functional group]. The needed charge-scaling factors were less than 10% for most groups (see Table 4), but for alkynes

TABLE 3: Statistical Data of the Different Model Approaches with the OPLS-AA Force Field Nonbonded Parameters[a]

	LR	original NLR	optimized NLR
MSE [kcal/mol]	−0.29	0.04	0.00
MUE [kcal/mol]	0.79	0.64	0.42
rmsd [kcal/mol]	1.09	0.90	0.67
corr coeff	0.956	0.955	0.974
slope	1.12	1.00	0.99
y-intercept [kcal/mol]	−0.04	0.04	−0.03

[a] The data set includes 174 solutes. The TIP5P water model is selected. 1,2,3-Propanetriol and the three molecules of the amide-group are excluded (see the text). The correlation coefficient, slope, and *y*-intercept refer to a linear regression fit of the calculated hydration free energies to experimental data.

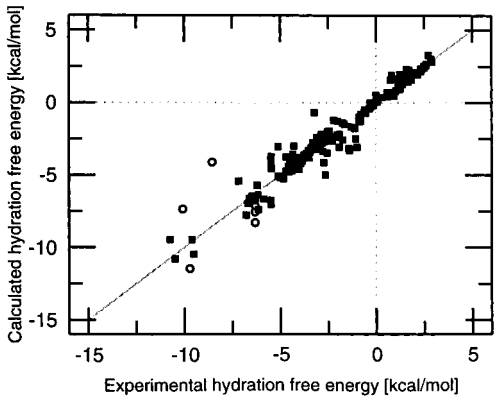

Figure 4. The 174 solute molecules (■) used in the statistics and the remaining seven molecules in the database set (○) (1,2,3-propanetriol falls outside the axis interval).

TABLE 4: Optimization Data of the Alkyl Functional Group Families, $CH_3(CH_2)_n-X$, Used in the Optimized NLR Model[a]

functional group (X)	N_X	original rmsd	α	optimized rmsd
alkane	8	0.047	1.00	
cycloalkane	5	0.829	1.00	
alkene	6	0.207	1.00	
alkyne	7	0.474	1.25	0.181
arene	3	0.200	1.00	
alcohol	8	0.313	1.00	
ether	3	1.284	1.25	0.387
amine	6	0.864	1.05	0.471
aldehyde	8	0.673	1.10	0.229
ketone	7	0.807	1.10	0.193
carboxylic acid	5	1.795	0.90	0.572
ester	11	0.391	1.00	
nitro	3	0.710	0.90	0.234
nitrile	3	0.840	0.95	0.269

[a] N_X is the number of family members, and α is the charge-scaling factor (OPLS-AA force field), which minimizes the rmsd of the family. The scaling was performed at a step rate of 0.05. If α = 1.00, no scaling was performed.

and ethers, it was as large as 25%. However, the alkynes were found to have a too low condensed-phase dipole moment using the OPLS-AA nonbonded parameters compared with the experimental vacuum dipole moment. The calculated dipole moments are 0.64 D for propyne to 1-hexyne, and the experimental values are as follows:[34] propyne 0.78 D, 1-butyne 0.80 D, 1-pentyne 0.84 D, and 1-hexyne 0.83 D. To obtain the experimental vacuum dipole moments in the condensed solvent phase, these have to be multiplied with a factor of about 1.25. This is exactly the same value as the scaling factor that optimizes the *n*-alkyl family, see Table 4. The mean unsigned error (MUE) for the solutes of the optimized NLR solvation model is 0.4,

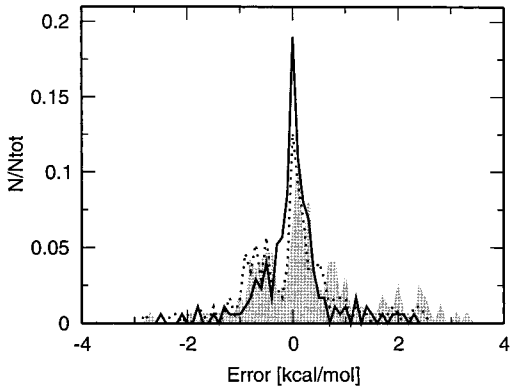

Figure 5. The error equal to $\Delta G_{hyd}^{expt} - \Delta G_{hyd}^{calcd}$. The number of solute molecules in the data set is $N_{tot} = 174$. The filled graph corresponds to the LR model, the dotted line is the original NLR model, and the solid line is the optimized NLR model. The latter two lines have the shape of a centered normal distribution. However, the center of the LR model is shifted to the right. This is also seen in the calculated mean signed error, which is not zero, see Table 3.

that is, only twice the experimental uncertainty. The MUE and rmsd are almost cut in half by the optimized NLR relative the LR solvation model. The improvement is clearly seen when the error distribution is plotted for the different models, see Figure 5. Hence, we conclude that nonlinearity is indeed important and should be included in solvation models of small neutral organic molecules in water. The small scaling factors that are needed may arise from the use of rigid molecular structures, which ought to be relaxed in the applied force field. The calculated, optimized and experimental hydration free energies for all 181 solutes are summarized in Table 5.

Even though the hydration free energies of unbranched and branched alkanes are well-reproduced, cycloalkanes prefer the gas phase too strongly. This is also evident for cycloalcohols, which become much better when the cycloalkane error is compensated, that is, cyclopentanol −4.4, cyclohexanol −5.3, and cycloheptanol −5.4 kcal/mol. The experimental values are −5.5, −5.5, and −5.5 kcal/mol. The problem is related to the Lennard-Jones parameters for cycloalkanes because the electrostatic free energy term is almost negligible.

The calculated hydration free energies of alcohols agree well with available experimental data. However, for the polyfunctional compounds of diols and triols, that is, 1,2-ethanediol and 1,2,3-propanetriol, in the present database, the agreement starts to deviate more and more. The partial charge distribution in the OPLS-AA force field shows an increase for the hydroxyl group from −0.683 for a monoalcohol oxygen to −0.70 for a diol and −0.73 for a triol oxygen, but this does not fully explain this behavior. The reason that the hydration free energy becomes too negative is probably due to missing intramolecular hydrogen bonds.[1,35] If gas-phase optimized structures are used, the calculated hydration free energies are much improved. The hydrogen-bonded formations have the lowest energy in both the gas and aqueous phase according to ab initio density functional theory calculations with the Becke−Lee−Yang−Parr hybrid functional and the 6-31G* basis set (DFT-B3LYP/6-31G*). Moreover, the PCM method[36,37] with scaled 1.2 Bondi radii[38] was used in the water phase. Thus, the solute molecular conformation is important for compounds that are able to form internal hydrogen bonds. This effect results in a change of about +5 kcal/mol for 1,2-ethanediol and about +9 kcal/mol for 1,2,3-propanetriol in the calculated electrostatic free energy. The experimental value for 1,2-ethanediol is −9.6 kcal/mol,[35,39] which differs from the value −7.7 kcal/mol[32,40] found in some

TABLE 5: Calculated Electrostatic (Optimized NLR Value), van der Waals, and Cavity Formation Free Energy, Optimized NLR Hydration Free Energy, Experimental Hydration Free Energy,[5,8,13,18,32,33] and LR Hydration Free Energy (kcal/mol) of All 181 Solutes Represented in the Created Data Base of Small Organic Compounds

no.	name	ΔG_e	ΔG_{vdw}	ΔG_{cav}	ΔG_{hyd}^{NLR}	ΔG_{expt}	ΔG_{hyd}^{LR}	no.	name	ΔG_e	ΔG_{vdw}	ΔG_{cav}	ΔG_{hyd}^{NLR}	ΔG_{expt}	ΔG_{hyd}^{LR}
1	methane[a]	−0.02	−3.54	5.48	1.91	2.00	1.90	**63**	water TIP4P	−10.27	−1.76	3.74	−8.29	−6.31	−10.85
2	ethane[a]	−0.05	−5.84	7.72	1.83	1.83	1.80	**64**	water TIP5P	−8.58	−1.76	3.73	−6.62	−6.31	−8.11
3	propane[a]	−0.06	−8.00	10.02	1.95	1.96	1.92	**65**	methanol	−6.29	−4.79	6.00	−5.08	−5.11	−6.15
4	*n*-butane[a]	−0.10	−10.15	12.34	2.09	2.08	2.05	**66**	ethanol	−6.39	−7.04	8.27	−5.15	−5.01	−5.68
5	*n*-pentane[a]	−0.11	−12.21	14.66	2.33	2.33	2.30	**67**	1-propanol	−6.64	−9.19	10.58	−5.25	−4.83	−6.83
6	*n*-hexane[a]	−0.15	−14.32	16.95	2.48	2.49	2.43	**68**	1-butanol	−6.32	−11.32	12.90	−4.73	−4.72	−6.01
7	*n*-heptane[a]	−0.16	−16.39	19.20	2.66	2.62	2.60	**69**	*tert*-butyl alcohol	−5.63	−11.14	12.87	−3.90	−4.51	−4.49
8	*n*-octane[a]	−0.15	−18.47	21.44	2.83	2.89	2.75	**70**	1-pentanol	−6.44	−13.33	15.22	−4.55	−4.47	−5.16
9	2-methylpropane	−0.09	−10.05	12.31	2.17	2.32	2.14	**71**	2-pentanol	−5.83	−13.28	15.19	−3.92	−4.39	−4.82
10	2,2-dimethyl-propane	−0.06	−11.98	14.56	2.53	2.50	2.49	**72**	3-pentanol	−5.45	−13.27	15.17	−3.55	−4.35	−4.44
11	2-methylbutane	−0.09	−12.06	14.47	2.32	2.38	2.30	**73**	1-hexanol	−6.32	−15.56	17.50	−4.38	−4.36	−5.17
12	2,2-dimethylbutane	−0.08	−13.90	16.55	2.57	2.60	2.51	**74**	1-heptanol	−6.50	−17.59	19.75	−4.34	−4.24	−6.47
13	2-methylpentane	−0.12	−14.11	16.76	2.54	2.52	2.51	**75**	1-octanol	−6.60	−19.68	21.97	−4.31	−4.09	−5.90
14	3-methylpentane	−0.13	−14.09	16.61	2.40	2.51	2.38	**76**	2-methoxyethanol	−8.75	−10.54	11.52	−7.76	−6.77	−8.45
15	2,4-dimethyl-pentane	−0.11	−15.91	18.83	2.81	2.89	2.76	**77**	cyclopentanol	−5.27	−12.33	13.64	−3.95	−5.49	−4.33
16	2,2,4-trimethyl-pentane	−0.10	−17.60	20.70	3.00	2.85	2.96	**78**	cyclohexanol	−6.08	−14.27	15.81	−4.54	−5.47	−6.15
17	2,2,5-trimethyl-hexane	−0.14	−19.66	23.08	3.27	2.72	3.23	**79**	cycloheptanol	−5.84	−16.10	17.66	−4.29	−5.49	−5.07
18	cyclopropane	−0.13	−7.07	8.76	1.56	0.75	1.54	**80**	1,2-ethanediol (gauche, int. H-bond)	−8.76	−8.25	7.54	−9.47	−9.60	−10.79
19	cyclopentane	−0.19	−11.26	13.09	1.65	1.20	1.61	**81**	1,2,3-propanetriol (all gauche, int. H-bonds)	−16.80	−11.37	11.63	−16.54	−9.22	−20.11
20	cyclohexane	−0.11	−13.20	15.26	1.95	1.23	1.92	**82**	phenol	−6.61	−12.13	12.14	−6.61	−6.62	−7.47
21	cycloheptane	−0.13	−15.10	17.12	1.89	0.80	1.84	**83**	*o*-cresol	−6.93	−14.18	14.45	−6.66	−5.87	−7.34
22	cyclooctane	−0.20	−16.92	18.77	1.64	0.84	1.60	**84**	*m*-cresol	−6.99	−14.22	14.47	−6.73	−5.49	−7.42
23	methylcyclopentane	−0.18	−13.18	15.36	2.00	1.60	1.97	**85**	*p*-cresol	−7.63	−14.24	14.48	−7.39	−6.14	−8.58
24	methylcyclo-hexane	−0.14	−15.13	17.51	2.24	1.71	2.21	**86**	dimethyl ether[b,c]	−3.55	−7.23	8.60	−2.17	−1.92	−0.96
25	*cis*-1,2-dimethyl-cyclohexane	−0.11	−16.88	19.28	2.29	1.58	2.25	**87**	diethyl ether[b,c]	−3.18	−11.50	13.30	−1.37	−1.76	−0.36
26	ethene	−0.52	−4.96	6.47	0.98	1.27	0.95	**88**	methyl-*n*-propyl ether[b,c]	−3.25	−11.48	13.28	−1.46	−1.66	−0.43
27	propene	−0.41	−7.24	8.79	1.14	1.27	1.10	**89**	methyl-isopropyl ether[c]	−3.09	−11.38	13.19	−1.27	−2.01	−0.28
28	2-methylpropene	−0.40	−9.41	11.08	1.27	1.16	1.23	**90**	*tert*-butyl-methyl ether[c]	−3.10	−13.28	15.15	−1.23	−2.21	−0.29
29	1-butene	−0.46	−9.40	11.10	1.24	1.38	1.19	**91**	1,2-dimethoxy-ethane[c]	−6.60	−12.85	14.22	−5.23	−4.84	−3.00
30	3-methyl-1-butene	−0.45	−11.40	13.40	1.55	1.82	1.50	**92**	ammonia	−6.16	−2.09	3.88	−4.37	−4.29	−5.40
31	1-pentene	−0.48	−11.49	13.40	1.43	1.66	1.39	**93**	methylamine[b,c]	−6.44	−4.46	6.12	−4.78	−4.56	−5.53
32	2-pentene	−0.42	−11.60	13.53	1.51	1.34	1.47	**94**	dimethylamine[c]	−5.02	−6.80	8.83	−2.99	−4.29	−3.61
33	1-hexene	−0.50	−13.58	15.72	1.64	1.68	1.57	**95**	trimethylamine[c]	−3.19	−9.00	11.51	−0.68	−3.23	−0.49
34	1-octene	−0.54	−17.75	20.24	1.94	2.16	1.86	**96**	ethylamine[b,c]	−6.02	−6.88	8.40	−4.50	−4.50	−5.31
35	1,3-butadiene	−0.91	−8.56	9.88	0.41	0.61	0.35	**97**	propylamine[b,c]	−6.39	−9.03	10.72	−4.70	−4.39	−3.90
36	1,4-pentadiene	−0.89	−10.79	12.14	0.46	0.93	0.37	**98**	butylamine[b,c]	−6.00	−11.13	13.04	−4.08	−4.29	−4.02
37	1,5-hexadiene	−0.86	−12.86	14.48	0.75	1.00	0.68	**99**	pentylamine[b,c]	−5.92	−13.26	15.35	−3.82	−4.10	−3.34
38	1,6-heptadiene	−0.90	−14.96	16.77	0.91	1.23	0.83	**100**	hexylamine[b,c]	−6.08	−15.35	17.63	−3.79	−4.03	−4.87
39	cyclopentene	−0.57	−10.60	11.87	0.70	0.56	0.66	**101**	aniline[c]	−7.02	−12.33	12.32	−7.03	−5.49	−7.99
40	ethyne	−1.27	−3.54	5.31	0.51	−0.01	0.44	**102**	pyridine[c]	−3.64	−11.08	10.97	−3.75	−4.70	−3.65
41	propyne[b,c]	−1.19	−6.38	7.57	−0.00	−0.31	0.35	**103**	pyrrolidine[c]	−5.25	−10.43	11.96	−3.72	−5.48	−4.14
42	1-butyne[b,c]	−1.23	−8.74	9.91	−0.06	−0.16	0.29	**104**	piperidine[c]	−4.73	−12.42	14.12	−3.03	−5.11	−3.10
43	1-pentyne[b,c]	−1.29	−10.84	12.23	0.09	0.01	0.46	**105**	morpholine[c]	−6.46	−11.72	12.76	−5.42	−7.17	−5.70
44	1-hexyne[b,c]	−1.28	−12.89	14.53	0.36	0.29	0.72	**106**	ethanal[b,c]	−4.43	−6.97	7.57	−3.83	−3.50	−3.20
45	1-heptyne[b,c]	−1.31	−15.03	16.83	0.50	0.62	0.85	**107**	propanal[b,c]	−4.16	−9.13	9.89	−3.41	−3.44	−2.80
46	1-octyne[b,c]	−1.30	−17.11	19.09	0.68	0.69	1.02	**108**	butanal[b,c]	−4.20	−11.21	12.18	−3.23	−3.18	−2.63
47	1-nonyne[b,c]	−1.32	−19.17	21.33	0.84	1.04	1.17	**109**	pentanal[b,c]	−4.03	−13.28	14.49	−2.82	−3.03	−2.25
48	benzene	−1.65	−11.03	11.62	−1.05	−0.87	−1.13	**110**	hexanal[b,c]	−4.16	−15.41	16.79	−2.78	−2.81	−2.21
49	toluene	−1.88	−13.08	13.96	−1.00	−0.89	−1.08	**111**	heptanal[b,c]	−4.25	−17.47	19.05	−2.68	−2.67	−2.09
50	ethylbenzene	−2.00	−15.08	16.28	−0.80	−0.80	−0.89	**112**	octanal[b,c]	−4.22	−17.47	19.05	−2.64	−2.29	−2.06
51	2-propylbenzene	−1.78	−16.93	18.40	−0.31	−0.31	−0.41	**113**	nonanal[b,c]	−4.23	−21.64	23.49	−2.38	−2.08	−1.86
52	*m*-xylene	−2.22	−15.12	16.27	−1.06	−0.84	−1.16	**114**	benzaldehyde	−4.20	−14.21	13.80	−4.61	−4.02	−4.78
53	*o*-xylene	−2.38	−15.05	16.15	−1.28	−0.90	−1.38	**115**	acetophenone	−4.36	−16.36	16.04	−4.67	−4.58	−4.86
54	*p*-xylene	−2.37	−15.16	16.27	−1.26	−0.81	−1.35	**116**	3-hydroxybenz-aldehyde	−9.48	−15.32	14.32	−10.47	−9.51	−11.48
55	1,2,4-trimethyl-benzene	−2.64	−17.10	18.43	−1.31	−0.86	−1.41	**117**	4-hydroxybenz-aldehyde	−9.77	−15.33	14.32	−10.79	−10.48	−12.87
56	biphenyl	−2.58	−20.06	20.08	−2.56	−2.63	−2.70	**118**	propanone[b,c]	−4.62	−9.28	9.88	−4.02	−3.85	−3.39
57	naphthalene	−2.49	−16.78	16.75	−2.52	−2.39	−2.63	**119**	butanone[b,c]	−4.47	−11.41	12.19	−3.69	−3.64	−3.08
58	anthracene	−3.11	−22.48	21.82	−3.76	−4.23	−3.90	**120**	2-pentanone[b,c]	−4.29	−13.50	14.50	−3.29	−3.53	−2.73
59	phenanthrene	−3.15	−22.45	21.65	−3.95	−3.95	−4.08	**121**	3-pentanone[c]	−4.14	−13.50	14.50	−3.14	−3.41	−2.59
60	pyrene	−3.28	−24.68	23.56	−4.40	−4.57	−4.52	**122**	2-hexanone[b,c]	−4.29	−15.59	16.80	−3.07	−3.29	−2.54
61	water SPC	−9.53	−1.77	3.75	−7.56	−6.31	−9.12	**123**	2-heptanone[b,c]	−4.31	−17.69	19.07	−2.94	−3.04	−2.40
62	water TIP3P	−9.12	−1.74	3.74	−7.12	−6.31	−8.47	**124**	4-heptanone[c]	−3.79	−17.64	19.05	−2.38	−2.93	−1.90

TABLE 5 (Continued)

no.	name	ΔG_e	ΔG_{vdw}	ΔG_{cav}	ΔG_{hyd}^{NLR}	ΔG_{expt}	ΔG_{hyd}^{LR}	no.	name	ΔG_e	ΔG_{vdw}	ΔG_{cav}	ΔG_{hyd}^{NLR}	ΔG_{expt}	ΔG_{hyd}^{LR}
125	2-octanone[b,c]	−4.22	−19.76	21.29	−2.68	−2.88	−2.18	**153**	2-nitropropane[c]	−3.53	−11.71	12.39	−2.85	−3.18	−3.83
126	2-nonanone[b,c]	−4.23	−21.80	23.50	−2.53	−2.49	−2.01	**154**	1-nitrobutane[b,c]	−3.87	−13.83	14.69	−3.01	−3.08	−4.10
127	5-nonanone[c]	−4.14	−21.75	23.49	−2.41	−2.67	−1.90	**155**	nitrobenzene[c]	−3.22	−14.95	14.02	−4.15	−4.17	−4.90
128	3,3-dimethylbutanone[c]	−4.04	−15.15	16.58	−2.61	−2.89	−2.07	**156**	4-nitrophenol[c]	−8.11	−16.05	14.70	−9.46	−10.74	−13.53
129	acetic acid[b,c]	−6.74	−8.23	7.99	−6.97	−6.70	−9.47	**157**	*o*-nitrotoluene[c]	−3.17	−16.89	16.57	−3.50	−3.63	−4.19
130	propionic acid[b,c]	−6.35	−10.39	10.30	−6.44	−6.47	−8.97	**158**	*m*-nitrotoluene[c]	−3.29	−16.99	16.47	−3.81	−3.50	−4.52
131	butyric acid[b,c]	−6.92	−12.46	12.62	−6.76	−6.36	−9.43	**159**	acetonitrile[b,c]	−5.82	−5.40	7.01	−4.20	−3.94	−4.98
132	pentanoic acid[b,c]	−6.73	−14.57	14.93	−6.38	−6.16	−8.93	**160**	propionitrile[b,c]	−5.57	−7.76	9.33	−3.99	−3.90	−4.76
133	hexanoic acid[b,c]	−6.31	−16.61	17.22	−5.70	−6.21	−7.88	**161**	butyronitrile[b,c]	−5.48	−9.88	11.66	−3.70	−3.69	−4.47
134	methylformate	−2.81	−9.03	8.46	−3.38	−2.78	−3.47	**162**	benzonitrile[c]	−3.22	−14.18	13.42	−3.98	−4.10	−4.45
135	methyl acetate	−3.11	−10.41	10.78	−2.74	−3.32	−2.86	**163**	chloromethane[b,c]	−1.45	−5.68	6.42	−0.71	−0.56	−1.08
136	methylpropionate	−2.82	−13.28	13.10	−3.00	−2.93	−3.10	**164**	chloroethane[b,c]	−1.50	−7.89	8.71	−0.68	−0.63	−1.06
137	methylbutyrate	−2.83	−15.39	15.40	−2.82	−2.83	−2.93	**165**	1-chloropropane[b,c]	−1.42	−10.04	11.04	−0.42	−0.36	−0.79
138	methylpentanoate	−2.82	−17.43	17.68	−2.57	−2.57	−2.68	**166**	2-chloropropane[c]	−1.27	−9.99	11.03	−0.23	−0.25	−0.56
139	methylhexanoate	−2.86	−19.50	19.93	−2.42	−2.49	−2.54	**167**	1-chlorobutane[b,c]	−1.42	−12.15	13.36	−0.21	−0.14	−0.59
140	methyloctanoate	−2.88	−23.69	24.35	−2.22	−2.04	−2.35	**168**	2-chlorobutane[c]	−1.22	−12.07	13.33	0.04	0.07	−0.26
141	ethylmethanoate	−4.48	−11.36	10.85	−4.99	−2.65	−5.14	**169**	1-chloropentane[b,c]	−1.43	−14.22	15.67	0.01	−0.07	−0.37
142	ethyl acetate	−2.86	−12.69	13.16	−2.39	−3.10	−2.50	**170**	2-chloropentane[c]	−1.22	−14.09	15.64	0.33	0.07	0.02
143	propylethanoate	−2.82	−14.85	15.48	−2.18	−2.86	−2.31	**171**	3-chloropentane[c]	−1.12	−14.10	15.60	0.38	0.04	0.10
144	butylethanoate	−2.85	−16.90	17.76	−1.99	−2.55	−2.10	**172**	dichloromethane[c]	−2.79	−7.77	7.40	−3.16	−1.42	−3.89
145	pentylethanoate	−2.92	−19.02	20.01	−1.94	−2.45	−2.08	**173**	1,2-dichloroethane[c]	−2.35	−9.93	9.72	−2.56	−1.75	−3.16
146	ethanamide	−11.13	−8.47	8.14	−11.46	−9.71	−13.11	**174**	1,3-dichloropropane[c]	−3.08	−12.04	12.05	−3.07	−1.92	−3.88
147	*N*-methylethanamide	−7.20	−11.14	10.98	−7.36	−10.08	−8.27	**176**	chloroform[c]	−0.99	−10.05	8.55	−2.49	−1.07	−2.76
148	*N*,*N*-dimethylethanamide	−4.22	−13.51	13.62	−4.11	−8.55	−4.24	**177**	chlorobenzene[c]	−1.37	−13.03	12.63	−1.77	−1.12	−2.15
149	ethanethiol	−2.66	−7.84	8.83	−1.67	−1.30	−2.04	**178**	1,2-dichlorobenzene[c]	−1.94	−14.98	13.63	−3.28	−1.38	−3.80
150	thiophenol	−3.30	−12.95	12.76	−3.48	−2.55	−3.81	**179**	1,3-dichlorobenzene[c]	−1.72	−15.01	13.64	−3.10	−0.99	−3.55
151	nitroethane[b,c]	−4.10	−9.60	10.09	−3.62	−3.76	−4.73	**180**	1,4-dichlorobenzene[c]	−1.66	−15.04	13.64	−3.06	−1.01	−3.51
152	1-nitropropane[b,c]	−3.86	−11.68	12.41	−3.13	−3.38	−4.25	**181**	2,2′-dichlorobiphenyl[c]	−2.61	−23.57	22.04	−4.14	−2.73	−4.85

[a] The *n*-alkanes used in the fitting of the cavity formation parameters. [b] The alkyl family members used in the optimization of the charge-scaling factors for some functional groups (see Table 4). [c] Molecules that have been charge scaled in accordance with their functional group charge-scaling factor.

investigations.[8,18] We will use the former value in our compilation. The 1,2,3-propanetriol hydration free energy value of −9.2 kcal/mol[32,41] differs from the value −8.50 that is found in the work by Viswanadhan et al.[18] However, both values are less than the hydration free energy of 1,2-ethanediol, which we find strange, and because Butler and co-workers obtained −7.7 kcal/mol for 1,2-ethanediol,[40] we suspect the experimental value of 1,2,3-propanetriol[41] to be too small as well. The experimental value may be estimated by looking at 1-propanol to which we add two intramolecular hydrogen-bonded hydroxyl groups (2 × ethanol → 1,2-ethanediol). This yields a hydration free energy for 1,2,3-propanetriol of about −14 kcal/mol. Hence, we believe the "true" hydration free energy for 1,2,3-propanetriol is found somewhere between −11 and −14 kcal/mol. This estimated value is at the experimental partitioning dectection limit and therefore probably difficult to measure. That is why 1,2,3-propanetriol is excluded in the statistical analysis of our data set, see Table 3.

The experimental hydration free energy changes of methylation going from ammonia to trimethylamine are $[NH_3 \rightarrow NH_2CH_3] = 0.3$, $[NH_2CH_3 \rightarrow NH(CH_3)_2] = -0.3$, and $[NH(CH_3)_2 \rightarrow N(CH_3)_3] = -1.1$ kcal/mol. However, several different solvation model approaches, both implicit and explicit ones, have failed to reproduce this series (see, e.g., refs 1, 2, and 42 and references therein). The dimethylamine and trimethylamine solutes are predicted to be less hydrophilic than what is experimentally expected. It has been proposed that this is due to an incorrect description of short-range hydrogen-bonding effects by methods that use electrostatic models.[42] The OPLS-AA force field has recently been reparametrized for amines.[23] However, the optimized NLR model follows the general trend and yields 0.4, −1.8, and −2.3 kcal/mol. Nevertheless, we may conclude that this effect is not due to the neglect of nonlinear dielectric response in the implicit continuum model.

Esters are rather well-predicted. However, formate and methanoate groups have the charge on the single hydrogen set to zero in accordance with the OPLS-AA force field when the hydrogen is replaced with a "neutral" methyl group.

The amide functional group includes only three family members in the present version of our database set. Moreover, these three solutes represent methylation of the ethaneamide molecule, which makes it difficult to try to optimize the amide group charge representation. In short, the amide group family consists of too few and too diverse members to improve any parametrization.

We have tried to scale the partial charges of the chloroalkanes to get better agreement with experimental values. However, the calculated values start to deviate for polyfunctional chloroalkanes. The problem may well reside in the Lennard-Jones parameters for the chloride atom in the OPLS-AA force field.

The consistent change in solvation free energy due to the insertion of a $-CH_2-$ group is apparent in the experimental data.[11] We examine the free energy difference between the butyl and octyl pairs of molecules (calcd, exptl) [kcal/mol]: *n*-butane and *n*-octane (−0.7, −0.8), 1-butene and 1-octene (−0.7, −0.8), 1-butyne and 1-octyne (−0.7, −0.9), 1-butanol and 1-octanol (−0.4, −0.6), butanal and octanal (−0.6, −0.9), butanone and 2-octanone (−1.0, −0.8), and methylbutanoate and methyloctanoate[4] (−0.6, −0.8).

Our solvation model may be criticized for the use of a fixed effective charge distribution of the solute in both the vacuum and the water phases. This simplifies the physical problem because we do not have to calculate the change in the internal solute energy during the transfer. The solute molecule will probably become polarized by the solvent when it is hydrated, and the partial charge distribution will be redistributed within the molecule. This can be calculated by combined QM/continuum models, for example, the polarizable continuum model (PCM) of Miertus, Scrocco, and Tomasi.[36,37,43] This is a linear dielectric response model in which the solute molecule forms a cavity in the solvent treated as a continuum dielectric medium. However, the cavity model is adjusted because the

applied standard atomic radii are reparametrized (scaled by a factor larger than one), which will quench the saturation effect for neutral molecules. The reaction field from the solvent acts back on the solute, which becomes polarized. However, because we have shown the importance of incorporating nonlinear dielectric response of the solvent, the reliability of such calculated ab initio distributions must be questioned due to the dielectric saturation. Moreover, condensed-phase dipole moments cannot be experimentally measured directly. Therefore, we prefer to use a standard classical force field like OPLS-AA, which is parametrized to reproduce both structural and thermodynamical data. Moreover, if nonlinear dielectric response effects are to be included in the polarization model, it must be noted that reaction field models cannot fully be utilized for a nonlinear dielectric in an inhomogeneous field.[19,22,44,45] These ab initio QM problems must be adressed before one can assert the superiority of QM polarization models over fixed effective charge distribution models.

4. Conclusions

Implicit solvent models used for predicting the hydration free energies of small organic molecules often represent the solvent by a simple medium. Such models have been very popular to use in the past decade. However, the electric field originating from the solute molecule can locally become quite strong, which sometimes gives rise to the nonlinear dielectric response effect called normal saturation. We have in a recent work developed a method called the modified Langevin−Debye model that is suited for such situations.[22] This is evident for neutral polar molecules for which the saturation contribution can be as large as 2.5 kcal/mol. The improvement in average rmsd is almost 40% from 1.9 kcal/mol for the linear response model to 1.1 kcal/mol for the nonlinear response analogue for molecules explicitly parametrized in a standard classical force field. The nonlinear response will drastically improve the continuum description of the implicit solvent model. Therefore, we conclude that nonlinear dielectric response effects are vital to incorporate in the implicit solvent model, although previous studies have shown that it is possible to parametrize a linear response model and thereby, with the use of a nonstandard force field, to give satisfactory estimates.

We apply the standard OPLS-AA force field, which is parametrized for small molecules in an explicit aqueous solution. Moreover, it is parametrized to reproduce not only structural but also thermodynamical data. We have constructed a data set containing 181 small organic molecules. If the LR model is compared with the NLR model for the hydration of these molecules, we find an improvement in the average error of about 25%. If the different functional groups are examined separately, some groups are found to deviate more than others. Hence, we introduce some new charge-scaling parameters that are fitted to minimize the rmsd of the alkyl family for each functional group. This optimized NLR model is about twice as accurate to predict hydration free energies as the LR model with an average MUE of 0.4 kcal/mol, which is just twice the experimental uncertainty. This model result might also be of interest as a starting point for people who parametrize and improve classical force fields.

Acknowledgment. We thank Ulf Norinder at AstraZeneca R&D in Södertälje, Sweden, for providing some of the molecules included in our data set. Tore Brinck was kind to perform the ab initio calculations on 1,2-ethanediol and 1,2,3-propanetriol. We also thank W. L. Jorgensen for providing a list of OPLS-AA parameters to our group.

References and Notes

(1) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.
(2) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 10282.
(3) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077.
(4) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(5) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878.
(6) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531.
(7) Still, W. C.; Tempczyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(8) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(9) Simonson, T.; Brünger, A. T. *J. Phys. Chem.* **1994**, *98*, 4683.
(10) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.
(11) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.
(12) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117.
(13) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9.
(14) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(15) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1.
(16) Du, Q.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **2000**, *104*, 796.
(17) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(18) Viswanadhan, V. N.; Ghose, A. K.; Singh, U. C.; Wendoloski, J. J. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 405.
(19) Böttcher, C. J. F. *Theory of Electric Polarization*, 2nd ed.; Elsevier: Amsterdam, 1973; Vol. I, Chapter 7.
(20) Malsch, J. *Phys. Z.* **1928**, *29*, 770.
(21) Kołodziej, H. A.; Jones, G. P.; Davies, M. *J. Chem. Soc., Faraday Trans. 2* **1975**, *71*, 269.
(22) Sandberg, L.; Edholm, O. *J. Chem. Phys.* **2002**, *116*, 2936.
(23) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827.
(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.
(25) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.
(26) Desnoyers, J. E.; Verrall, R. E.; Conway, B. E. *J. Chem. Phys.* **1965**, *43*, 243.
(27) Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1030.
(28) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992.
(29) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. *J. Chem. Phys.* **1959**, *31*, 369.
(30) Reiss, H.; Frisch, H. L.; Helfand, E.; Lebowitz, J. L. *J. Chem. Phys.* **1959**, *32*, 119.
(31) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717.
(32) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(33) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(34) Lide, D. R., Ed. *CRC Handbook of Chemistry and Physics*, 80th ed.; CRC Press: Boca Raton, FL, 1999.
(35) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1994**, *116*, 3892.
(36) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(37) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.
(38) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(39) Suleiman, D.; Eckert, C. A. *J. Chem. Eng. Data* **1994**, *39*, 692.
(40) Butler, J. A. V.; Ramchandani, C. N. *J. Chem. Soc.* **1935**, 952.
(41) Butler, J. A. V.; Reid, W. S. N. *J. Chem. Soc.* **1936**, 1171.
(42) Marten, B.; Kyungsun, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.
(43) Tomasi, J.; Cammi, R.; Mennucci, B. *Int. J. Quantum Chem.* **1999**, *75*, 783.
(44) Bordewijk, P. *J. Chem. Phys.* **1973**, *58*, 1220.
(45) Brito, P.; Grosse, C.; Halloy, C. *Am. J. Phys.* **1986**, *54*, 1014.