Accurate Prediction of Absolute Acidity Constants in Water with a Polarizable Force Field: Substituted Phenols, Methanol, and Imidazole

George A. Kaminski*

Department of Chemistry, Central Michigan University, Mount Pleasant, Michigan 48859

Received: January 10, 2005

Validity of a force field with explicit treatment of electrostatic polarization in a form of inducible point dipoles for computing acidity constants was tested by calculating absolute pK_a values of substituted phenols, methanol, and imidazole in water with the molecular dynamics technique. The last two systems were selected as tyrosine and histidine side-chain analogues, respectively. The solvent was represented by an explicit polarizable water model. Similar calculations were also performed with a modified OPLS-AA nonpolarizable force field. The resulting pK_a values were compared with available experimental data. While the nonpolarizable force field yields errors of about 5 units in the absolute pK_a values for the phenols and methanol, the polarizable force field produces the acidity constant values within a ca. 0.8 units accuracy. For the case of imidazole, the fixed-charges force field was capable of reproducing the experimental value of pK_a (6.4 versus the experimental 7.0 units), but only at a cost of dramatically underestimating dimerization energy for the imidazolium−water complex. At the same time, the polarizable force field yields an even more accurate result of pK_a = 6.96 without any sacrifice of the accuracy in the dimerization energy. It has also been demonstrated that application of Ewald summation for the long-range electrostatics is important, and substitution of a simple cutoff procedure with Born correction for ions can lead to underestimation of absolute pK_a values by more than 5 units. The accuracy of the absolute acidity constants computed with the polarizable force field is very encouraging and opens road for further tests on more diverse organic molecules sets, as well as on proteins.

I. Introduction

Accurate estimation of acidity constants for organic and biological molecules is an important and challenging task for the computational chemistry community. This importance results from a wide variety of organic and biological applications, especially those related to determining pK_a values of protein and drug molecules.[1] The most popular technique involves fitting parameters of a linear free energy relationship to a large training set of molecules.[2] If the descriptors in the relationship are chosen correctly, this approach allows us to reproduce acidity constants for a molecule similar, but not identical, to those in the training set.

It is more attractive, however, to be able to predict pK_a values for a molecule based on purely physical simulations rather than on knowledge of properties of a set of similar compounds. On one hand, this should allow us to obtain results for any new molecule, even if there are no data for a similar one to fit parameters to. On the other hand, parameter fitting in this case could be less resource consuming. Finally, a number of other important molecular properties could be obtained in the course of such simulations, e.g., ionic salvation energies, molecular structure in solutions, etc. This is why a number of attempts have been made in this direction.[3] Nevertheless, a level of accuracy of ca. 0.8 pK_a units cannot be achieved without introducing extra correction terms or/and explicit fitting to an extensive training set. The reason for this is in the nature of the pK_a calculations. Let is consider deprotonation reaction for a molecule AH:

$$\mathrm{AH(aq)} \xrightarrow{\Delta G(\mathrm{aq})} \mathrm{A^-(aq) + H^+(aq)}$$

Then the acidity constant value for AH is defined as

$$K_a = [\mathrm{A^-}][\mathrm{H^+}]/[\mathrm{AH}] \qquad \mathrm{p}K_a = -\log(K_a) \tag{1}$$

$$\mathrm{p}K_a = \Delta G(\mathrm{aq})/(2.303RT) \tag{2}$$

To determine the pK_a experimentally, one needs to measure the concentrations of the protonated and deprotonated forms of the substance, [AH] and [A^-]. On the other hand, to find the value of the acidity constant computationally, it is usually necessary to compute the free energy change ΔG(aq), which includes a difference of two energies with large magnitudes—gas-phase deprotonation free energy for AH and difference in the hydration energies of AH and A^-. Each of these terms typically has a magnitude of hundreds of kcal/mol, and, to obtain the absolute pK_a within 0.8 units accuracy, it is necessary to compute the total ΔG(aq) value with the accuracy of ca. 1.1 kcal/mol. At the same time, a true thermodynamic sampling in the form of either Monte Carlo or molecular dynamics is desirable in computing the AH−A^- solvation energy change to avoid problems arising because of incomplete configuration space sampling. The true free energy change can then be computed with the statistical perturbation theory (SPT).[4] The most practical way to conduct such simulations is with an empirical force field, such as, for example, OPLS-AA,[5]

* E-mail: Kamin1ga@cmich.edu.

10.1021/jp050156r CCC: $30.25
Published on Web 03/09/2005

AMBER,[6] MMFF,[7] etc. The problem is, however, that all these widely used force fields do not include electrostatic polarization in any explicit form. Therefore, one can expect that they might not perform adequately in simulations, in which a significant change of the electrostatic properties of the system involved takes place. Change from a neutral AH molecule to the A^- anion in aqueous solution and gas phase is certainly one of such cases. For example, it has been demonstrated in an earlier work that, although the OPLS-AA yields excellent liquid-state properties for organic molecules, the same force field overestimates magnitudes of gas-phase dimerization energies for some of the systems by more than 2 kcal/mol.[8] This is why this work was carried out with a polarizable empirical force field, and the superiority of this approach over using a nonpolarizable modified OPLS-AA force field proved to be possible to demonstrate.

The rest of the paper is organized as follows. Section II describes the methodology employed in these calculations. Results and discussion are given in section III. Finally, section IV contains conclusions and a brief description of the future directions of the presented research.

II. Method

A. Force Fields. We have tested performance of both a polarizable force field (PFF) and a modified fixed-charges OPLS-AA. The procedure for building the PFF has been described elsewhere.[8] Given below is a brief outline.

The total energy E_{total} is computed as a sum of the electrostatic term $E_{electrostatics}$, the nonelectrostatic van der Waals part E_{vdW}, bond stretching and angle bending energies E_{bond} and E_{angles}, and the torsional term $E_{torsion}$:

$$E_{total} = E_{electrostatics} + E_{vdW} + E_{bonds} + E_{angles} + E_{torsion} \quad (3)$$

The total electrostatic energy of the model results from interactions of fixed-magnitudes atomic charges, fixed-point dipoles, and inducible dipoles:

$$U(\{q_{ij}\},\{\mu_i\}) = \sum_i \left(\chi_i \cdot \mu_i + \frac{1}{2}\mu_i \cdot \alpha_i^{-1} \cdot \mu_i\right) + \frac{1}{2}\sum_{ij\neq kl} q_{ij} J_{ij,kl} q_{kl} + \sum_{ij,k} q_{ij} \mathbf{S}_{ij,k} \cdot \mu_k + \frac{1}{2}\sum_{i\neq j} \mu_i \cdot \mathbf{T}_{i,j} \cdot \mu_i \quad (4)$$

Here $J_{ij,kl}$ is a scalar coupling between bond-charge increments on sites i,j and k,l, q_{ij} and q_{kl}; $\mathbf{S}_{ij,k}$ is a vector coupling between a bond-charge increment on sites i,j and a dipole on site k, μ_k; a rank-two tensor coupling $\mathbf{T}_{i,j}$ describes interactions between dipoles on sites i and j. α_i is the polarizability of site i. Parameters χ_i describe "dipole affinity" of site i.

Following the formalism of the Coulomb interactions, we assume

$$J_{ij,kl} = \frac{1}{r_{ik}} - \frac{1}{r_{il}} - \frac{1}{r_{jk}} + \frac{1}{r_{jl}} \quad (5)$$

$$S_{ij,k} = \frac{r_{ik}}{r_{ik}^3} - \frac{r_{jk}}{r_{jk}^3} \quad (6)$$

$$T_{i,j} = \frac{1}{r_{ij}^3}\left(\mathbf{1} - 3\frac{r_{ik} r_{ik}}{r_{ij}^2}\right) \quad (7)$$

Inducible dipoles are placed on all the heavy atoms and on some polar hydrogens, as described in ref 8. Fixed charges q and permanent dipoles defined by the "dipole affinities" χ are present on all the atoms. "Virtual sites" representing electron lone pairs were placed on sp^2 and sp^3 oxygen atoms with the virtual site−oxygen distances set to 0.47 Å in the both cases. Dipole−dipole and charge-dipole screening is used to damp the polarization response when the perturbing site is at short distances. Each dipole and each charge has an individually set screening length. The corresponding length parameters for two atoms are added together to obtain the effective screening length for the pair interaction.

The values of the electrostatic parameters were fitted to reproduce quantum mechanically determined electrostatic potential around an isolated molecule (permanent charges and dipoles) and the same molecule with dipolar electrostatic probes placed at hydrogen-bonding and random positions (polarizabilities). B3LYP density functional theory (DFT) with the cc-pVTZ(-f) basis set of Dunning[9] was employed. The adequacy of the method has been shown in a previous work.[8] All the quantum mechanical calculations in this work were carried out with the Jaguar software suite.[10]

The overall nonelectrostatic pair potential form is

$$E_{vdW} = \sum_{i<j} A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6} + C_{ij}\exp(r_{ij}/\alpha_{ij}) \quad (8)$$

The A parameter is set so that the $1/r^{12}$ term is close to zero in the hydrogen bonding region but is large enough to prevent atoms from being positioned too close and thus penetrating the nonphysical region of the phase space. Fitting of the van der Waals parameters is done after the electrostatic part of the force field is ready. The repulsion parameters C and α are produced to ensure the correct gas-phase dimerization properties, as compared with the high-level ab initio results. The procedure is based on an extrapolation scheme utilizing LMP2/cc-pVTZ(-f) and LMP2/cc-pVQZ data and is described in refs 8 and 11. Because there are two parameters, both gas-phase dimerization energy and interatomic distances can be reproduced correctly for any reasonable values of parameters B. Combining rules are geometric for C and arithmetic for α. Then condensed-phase molecular dynamics simulations are carried out to compute pure liquid heats of vaporization and densities. The set with the B, which allows the best liquid-state results, is accepted as the PFF model.

Finally the bond stretching and angle bending terms have the standard harmonic functional form, and the torsional energy is described by a sum of Fourier series:

$$E_{bonds} = \sum_{bonds} K_r(r - r_{eq})^2 \quad (9)$$

$$E_{angles} = \sum_{angles} K_\Theta(\Theta - \Theta_{eq})^2 \quad (10)$$

$$E_{torsion} = 1/2 \sum_{dihedrals} V_1^i[1 + \cos(\phi_i)] + V_2^i[1 - \cos(2\phi_i)] + V_3^i[1 + \cos(3\phi_i)] \quad (11)$$

Here K_r and K_Θ represent the force constants; r, r_{eq}, Θ, and Θ_{eq} are actual and equilibrium values of bond lengths and angles; and ϕ are the dihedral angles. All the parameters in eqs 9−11 are taken directly from the OPLS-AA.[5]

For the nonpolarized calculations, the standard OPLS-AA functional form was used, with the $E_{electrostatics}$ and E_{vdW} terms in eq 3 replaced by the sum of the Coulomb and Lennard-Jones

contributions for pairwise intra- and intermolecular interactions:

$$E_{nb} = \sum_{i<j}[q_i q_j e^2/r_{ij} + 4\epsilon_{ij}(\sigma_{ij}^{12}/r_{ij}^{12} - \sigma_{ij}^{6}/r_{ij}^{6})]f_{ij} \quad (12)$$

Geometric combining rules for the Lennard-Jones coefficients were employed: $\sigma_{ij} = (\sigma_{ii}\sigma_{jj})^{1/2}$ and $\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2}$. The summation runs over all the pairs of atoms $i < j$ on molecules A and B or A and A for the intramolecular interactions. Moreover, in the latter case, the coefficient f_{ij} is equal to 0.0 for any *i-j* pairs connected by a valence bond (1-2 pairs) or a valence bond angle (1-3 pairs). $f_{ij} = 0.5$ for 1,4-interactions (atoms separated by exactly 3 bonds) and $f_{ij} = 1.0$ for all the other cases. Standard OPLS-AA parameters were used, except when unavailable. In the latter case, gas-phase quantum mechanical dimerization energies and interatomic distances were used for fitting of the missing values. Particular cases are listed in the Results section below.

B. General Protocol for the pK_a Calculations. Let us consider the following closed thermodynamic cycle:

$$\begin{array}{ccccc} & & \Delta G(\text{gas}) & & \\ & \text{AH(gas)} & \rightarrow & \text{A}^-\text{(gas)} + \text{H}^+\text{(gas)} & \\ \Delta G_{sol}(\text{AH}) \downarrow & & & \downarrow \Delta G_{sol}(\text{A}^-) \quad \downarrow \Delta G_{sol}(\text{H}^+) & \\ & \text{AH(aq)} & \rightarrow & \text{A}^-\text{(aq)} + \text{H}^+\text{(aq)} & \\ & & \Delta G(\text{aq}) & & \end{array}$$

One can write

$$\Delta G(\text{gas}) + \Delta G_{sol}(\text{A}^-) + \Delta G_{sol}(\text{H}^+) - \Delta G(\text{aq}) - \Delta G_{sol}(\text{AH}) = 0 \quad (13a)$$

$$\Delta G(\text{aq}) = \Delta G(\text{gas}) + \Delta G_{sol}(\text{H}^+) + \Delta G_{sol}(\text{A}^-) - \Delta G_{sol}(\text{AH}) \quad (13b)$$

and

$$\Delta G(\text{aq}) = \Delta G(\text{gas}) + \Delta G_{sol}(\text{H}^+) + \Delta G(\text{aq, AH}\rightarrow\text{A}^-) - \Delta G(\text{gas, AH}\rightarrow\text{A}^-) \quad (14)$$

It has been shown previously that quantum mechanical calculations of the ΔG(gas) term usually lack the accuracy level needed for the 0.8 pK_a units accuracy mentioned above, even if a high-level theory is employed.[12] This is why the author of this article has decided to concentrate on the accuracy of the model in describing the liquid-state effect of the deprotonation, and thus on correctly predicting the ΔG(aq, AH→A$^-$) and ΔG(gas, AH→A$^-$) terms in eq 14. The gas-phase bond breaking energies ΔG(gas) were adopted directly from the available experimental data with no further adjustments.[12,13] Furthermore, as the proton free energy of solvation $\Delta G_{sol}(\text{H}^+)$ has the same value for any molecular system studied in this work or to be studied in the future, a carefully determined literature result[13] of −269.0 kcal/mol was used.

Both the aqueous solution ΔG(aq, AH→A$^-$) and gas-phase ΔG(gas, AH→A$^-$) deprotonation free energies were determined with the statistical perturbation theory using the molecular dynamics module in the SIM software.[14] The molecules being deprotonated and the solvent water molecules were completely flexible, with no geometric constraints. All the molecular dynamics simulations were carried out with the target temperature of 298.15 K and at 1 atm. constant pressure, except where explicitly noted otherwise. Gas-phase simulations were run with 8×10^{-4} ps steps. A total of 80 ps of initial equilibration was followed by 80 ps of the actual free energy perturbation. In the liquid phase, time steps of 2×10^{-4} to 8×10^{-4} ps were used, with at least 10^5 steps of equilibration followed by at least 2×10^5 steps of the actual perturbation. A cubic cell of 209 water molecules (211 for imidazole) was included in each of the solvated simulations, with the periodic boundary conditions applied and the Ewald summation technique used to account for the interactions beyond the cell boundaries.[15] In the case of the OPLS-AA methanol deprotonation calculations, simulations with spherical cutoffs for the interactions were also carried out to assess the importance of using the Ewald summation procedure. These calculations were done with the Monte Carlo technique using BOSS program.[16] A total of 505 water molecules was employed in these simulations. The TIP5P[17] fixed-charges water model was employed in this and all the other fixed-charges simulations. The whole deprotonation process was divided into 20 smaller parts. For each of them, both gas-phase and solvated Monte Carlo runs included 2×10^6 configurations of equilibration and 4×10^6 configurations of averaging. The cutoff distances were 8.5 Å for the solvent−solvent and 10 Å for solute−solvent interactions.

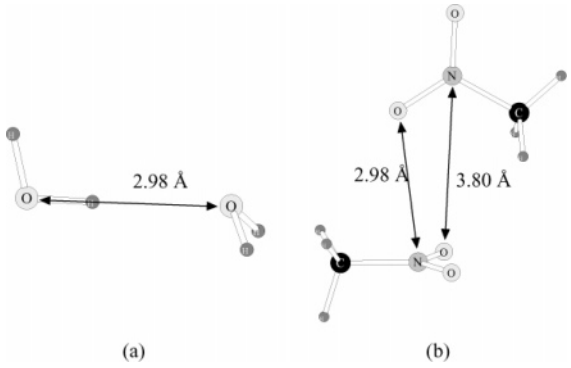

Figure 1. Lowest energy gas-phase dimers of water (a) and nitromethane (b) molecules with the ab initio distances between selected heavy atoms shown.

Results of all the new modified OPLS and PFF model parametrizations (including the PFF water model) are presented in the next section.

III. Results and Discussion

A. Potential Function Parameters. First, PFF parameters were produced for water, CH_3Cl and CH_3NO_2. The water model was used to represent the explicitly simulated solvent. The chloro- and nitromethane parameters were used to model the substituent groups in phenols. The parameters fitting was carried out as described above and in ref 8. Both gas-phase dimerization energies and interatomic distances and liquid-state heats of vaporization and densities were reproduced for water and CH_3NO_2. Parametrization of CH_3Cl relied on the liquid-state properties only as no hydrogen-bonded CH_3Cl dimer can be formed.

The lowest energy ab initio dimer structures are shown in Figure 1. The geometries of these and other dimers in this paper were optimized at the LMP2/cc-pVTZ(-f) level of theory. Table 1 presents comparison of the results of liquid- and gas-phase simulations for water, chloromethane, and nitromethane, carried out with the PFF, as compared to the experimental or quantum mechanical data. Experimental results are from ref 18. Ab initio energies computed by extrapolating LMP2/cc-pVTZ and LMP2/cc-pVQZ as described in ref 11. It can be seen from the results in Table 1 that the gas-phase dimerization energies and distances

TABLE 1: Gas-Phase and Liquid-State Properties of the H_2O, CH_3Cl, and CH_3NO_2 Models[a]

	gas phase				liquid state			
	dimer energy		distance		ΔH_{vap}		density	
system	PFF	exptl/ ab initio	PFF	exptl/ ab initio	PFF	exptl/ ab initio	PFF	exptl/ ab initio
H_2O	5.54	5.44	2.88	2.98	10.54	10.51	0.992	0.997
CH_3Cl					5.61	5.12[a]	0.968	1.005[a]
CH_3NO_2	5.47	5.68	2.94/ 3.79	2.98/ 3.80	9.40	9.15	1.139	1.137

[a] Energies in kcal/mol, distances in Å, and densities in g/cm^3. Data at 298.15 K for Water and Nitromethane and 249 K Chloromethane.

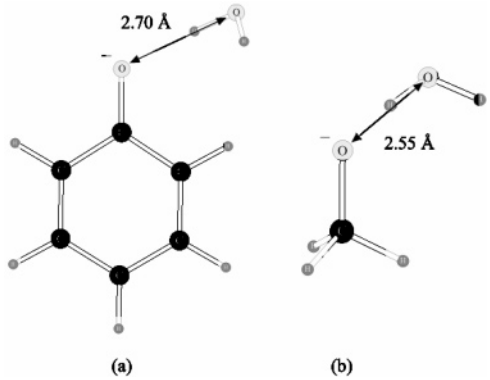

Figure 2. Lowest energy gas-phase dimers of water with $C_6H_5O^-$ (a) and CH_3O^- (b) molecules with the ab initio distances between selected heavy atoms shown.

TABLE 2: Gas-Phase Dimerization Energies (kcal/mol) and O···O Distances (Å) for $C_6H_5O^-$ and CH_3O^- Complexes with Water

	energy			distance		
system	ab initio	PFF	OPLS	ab initio	PFF	OPLS
$C_6H_5O^-$	10.69	10.64	10.62	2.70	2.72	2.68
CH_3O^-	19.32	19.40	18.59	2.55	2.59	2.71

are within no more than ca. 0.1 kcal/mol and 0.1 Å from the target, and the liquid-state deviations are within ca. 0.5 kcal/mol and 5% of the densities. This level of accuracy is the same as for the PFF development reported in ref 8, and therefore, the developed parameters are compatible with the phenol model reported there. Moreover, to further assess this compatibility, we have computed gas-phase dimerization energy and O···O distance for a complex of water with phenol. The PFF model yielded 5.05 kcal/mol and 2.95 Å, respectively. The experimental numbers are 5.48−5.60 kcal/mol and 2.88−2.93 Å.[19] Thus, even though we did not carry out any explicit phenol−water parameters fitting, the energy and interatomic distance is within ca. 0.5 kcal/mol and 0.07 Å form the available experimental data. This confirms compatibility of the parameters and indirectly witnesses the ability of the PFF to respond correctly to changes in the electrostatic environment.

The next step was to produce parameters for the anionic systems containing O^-, namely, $C_6H_5O^-$ and CH_3O^-. The former system required parameters to be produced with both the PFF and modified OPLS-AA approaches, whereas the latter already exists for the standard OPLS-AA case.[16] No condensed-state fitting was performed for this part of the parameters development. We computed dimerization energies and O···O distances for $C_6H_5O^-$−H_2O and CH_3O^-−H_2O complexes only. Structures of the complexes are presented in Figure 2, and the energetic results are shown in Table 2.

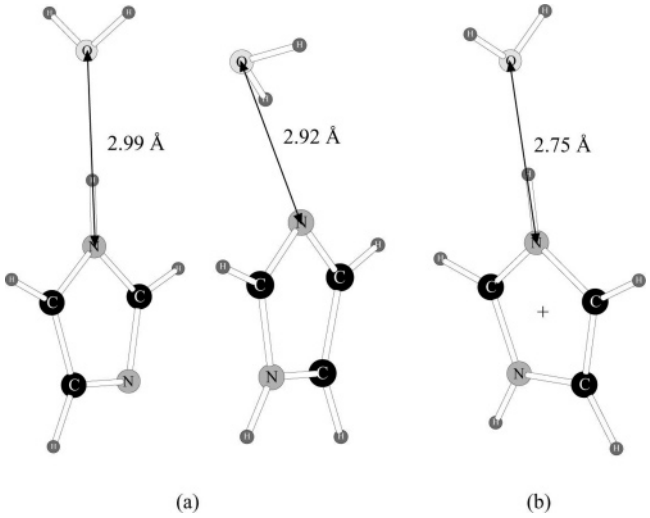

Figure 3. Lowest energy gas-phase dimers of water with $C_3H_4N_2$ (a) and $C_3H_5N_2^+$ (b) molecules with the quantum mechanical distances between selected heavy atoms shown.

TABLE 3: Gas-Phase Dimerization Energies (kcal/mol) and N···O Distances (Å) for $C_3H_4N_2$ and $C_3H_5N_2^+$ Complexes with Water

	energy			distance		
system	ab initio	PFF	OPLS[a]	ab initio	PFF	OPLS[a]
$C_3H_3NNH\cdots OH_2$	6.11	5.90	6.38	2.99	3.00	3.38
$C_3H_4NN\cdots HOH$	6.59	6.39	5.16	2.92	3.02	2.89
$C_3H_4NNH^+\cdots OH_2$	18.14[b]	17.97	15.82	2.75[b]	2.64	2.69

[a] OPLS parameters for the imidazolium are the OPLS-AA parameters for the histidine side-chain, from ref 16. [b] Reference 21.

It can be seen from Table 2 that both the PFF and modified OPLS results for the water−phenoxide dimer have energies within 0.07 kcal/mol and distances within 0.02 Å from their ab initio counterparts. Although the PFF accuracy for the CH_3O^-−water dimer is ca. 0.17 kcal/mol and 0.04 Å, the OPLS underestimates the binding energy by 0.73 kcal/mol, whereas the oxygen−oxygen distance is actually overestimated by 0.16 Å. These results were obtained with the existing OPLS-AA parameter values for the CH_3O^-,[16] and we adopted the parameters with no modifications to preserve the original OPLS-AA values wherever possible. It should be noted that the dimerization energy is actually underestimated for this gas-phase dimer, while it is known that fixed-charges force fields usually overestimate gas-phase binding energetics to obtain correct liquid-state properties. Explanation of this fact is given in the next subsection.

Finally, polarizable force field parameters were produced for the imidazole and imidazolium molecules. The values of the parameters were optimized to reproduce structures and energies of the molecule−water dimers shown in Figure 3. Comparison of the results with those achieved by employing the OPLS-AA parameters is given in Table 3.

The OPLS-AA parameters were adopted without change from refs 16 and 20, the imidazolium ones designed for the histidine side-chain description. The energy of the cation−water dimmer is underestimated by about 2.3 kcal/mol, and we will show below that this is the price one has to pay to obtain a reasonable value of the imidazole acidity constant with this fixed-charges force field. At the same time, the PFF allowed the dimerization energies and distances within ca. 0.2 kcal/mol and 0.1 Å from the quantum mechanical results, respectively. Finally, imidazole condensed-state properties are very similar with both the OPLS and PFF, as can be seen from the data in Table 4.

TABLE 4: Liquid-State Properties of the Imidazole Model[a]

	ΔH_{vap}		density	
system	PFF	OPLS[b]	PFF	OPLS[b]
imidazole	15.11	14.94	1.091	1.120

[a] Energies in kcal/mol and densities in g/cm^3. Data for 298.15 K. [b] From ref 20.

TABLE 5: pK_a Values for Substituted Phenols, Computed with the Modified OPLS-AA Force Field

	absolute pK_a		relative pK_a	
system	modified OPLS	experiment[a]	modified OPLS[b]	experiment[a]
phenol	4.50	9.98	0.52	0.00
p-chlorophenol	−1.06	9.38	−5.04	−0.60
p-aminophenol	8.22	10.30	4.24	0.32
average error	5.29		2.96	

[a] Reference 13. [b] Zero shifted to reduce the average error.

TABLE 6: pK_a Values for Substituted Phenols, Computed with the Polarizable Force Field

	absolute pK_a		relative pK_a	
system	PFF	experiment[a]	PFF[b]	experiment[a]
phenol	10.81	9.98	0.07	0.00
p-chlorophenol	9.87	9.38	−0.87	−0.60
p-nitrophenol	8.17	7.15	−2.63	−2.83
average error	0.78		0.18	

[a] Reference 13. [b] Zero shifted to reduce the average error.

B. Acidity Constant Values for Substituted Phenols. In all the cases, the calculations were carried out using eqs 2 and 14, via direct deprotonation of the species. First, absolute pK_a values were computed for substituted phenols using the OPLS-AA force field modified as described above. The results are presented in Table 5. The average error in the relative acidity constants as computed with the modified OPLS-AA was 2.96. This result is not out of the range that could be expected for a fixed-charges force field. Numbers of similar accuracy were obtained with the OPLS-AA in earlier works.[22] The absolute acidity constant error of 5.29 is simply disastrous for any attempt to compute pK_a values with the chemically valuable accuracy of ca. 0.8 units.

Table 6 shows results of substituted phenols acidity constant simulations with exactly the same protocol, but using the polarizable force field. The results seem to be amazingly accurate, with the average error in the relative pK_a values at only 0.18 units, and the absolute pK_a values accurate to 0.78. Moreover, the free energy of solvation difference for phenol and phenoxide molecule was found to be 65.25 kcal/mol. The experimental numbers vary from 65.38 to 68.38 kcal/mol.[13] This result is yet another reason to believe that the protocol and polarizable force field used in this project are adequate in reproducing solvation properties of such electrostatically different species as the protonated and deprotonated forms of phenol.

It should be noted that, while the polarizable force field seems to be adequate in capturing the solvation effects resulting from deprotonation, the fixed-charges modified OPLS-AA underestimates the pK_a values in all the substituted phenols cases. In accordance with eq 2, this means that the anionic species solvation energy is estimated to be too low. To explain this phenomenon, let us consider the following schematic illustrations.

Figure 4 shows dipole moments of a water molecule in a dimer with phenol, with both OPLS and PFF. Let us assume

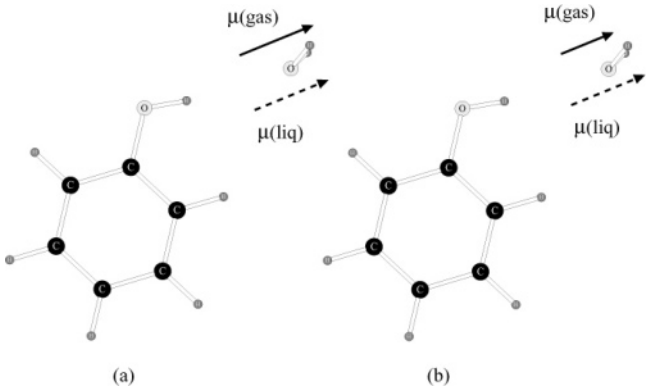

Figure 4. Schematic depiction of a water molecule dipole moment with a fixed-charges (a) and polarizable (b) force field for phenol. Liquid-state and gas-phase dipole moments are shown as dashed and solid arrows, respectively.

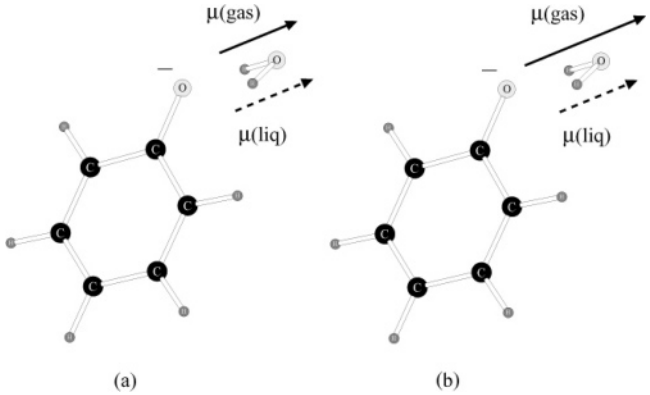

Figure 5. Schematic depiction of a water molecule dipole moment with a fixed-charges (a) and polarizable (b) force field for phenoxide. Liquid-state and gas-phase dipole moments are shown as dashed and solid arrows, respectively.

that the condensed-phase dipole moments, shown as dashed arrows, are adequate in reproducing liquid-state properties (and thus have about the same value for the both force fields). The fixed-charges dipole moment stays the same in the gas phase, as such force fields are not capable of any electrostatic readjustments. On the other hand, the PFF dipole moment in the gas-phase dimer will be smaller. This reflects the well-known fact that molecules are polarized more in aqueous solutions as they are in the gas phase. For example, the dipole moment of a water molecule in bulk water is about 1 D greater that its dipole moment in the gas phase.[23] Therefore, to perform well in the liquid state, the fixed-charges force fields have to overestimate the magnitudes of gas-phase dimerization energies.[24] Alternatively, a well-performing gas-phase fixed-charges model would underestimate the favorable energetics of solution. This is true, no matter if we consider a model for water or for the phenol molecule in Figure 4.

The situation is just opposite in the case of a charged ion shown in Figure 5. In this case, the real response of the water molecule should be an increase of the dipole moment in the gas phase, as the interaction with the charged species should polarize the water molecule more than interactions in the bulk water. A polarizable force field can capture this effect correctly. However, a nonpolarizable one would still attribute the same electrostatic properties to the water molecule (Figure 5a). Therefore, the gas-phase energy for such a dimer would be underestimated in magnitude, not overestimated, as happens with the neutral molecule in Figure 4a. Thus, although the fixed-charges force fields have lower accuracy in assessing gas-phase

TABLE 7: pK_a Values for Imidazole

method	pK_a
OPLS-AA	6.42
PFF	6.96
experiment[a]	7.0

[a] Reference 12.

dimerization energies for both neutral and charged molecules because of the inability to reproduce electrostatic changes adequately, the actual mechanism seems to be exactly opposite in the two cases. Furthermore, because we have fitted the ion parameters for both the modified OPLS and PFF phenoxide anions to gas-phase data, the fixed-charges molecule has a stronger interaction ability than it should (to compensate for the nonpolarization of water), and this apparently leads to the overestimation of the favorable solvation energy and thus to the lowering of the modified OPLS pK_a values. At the same time, it should be noted that the polarizable model functioned well enough in the aqueous solution, even though the actual parametrization only took place for the gas-phase ion−water complex. The author believes that this happens simply by the virtue of the ability of the polarizable model to reproduce elctrostatic effects, no matter what is the particular physical reason for their emergence.

C. Acidity Constant Values for Imidazole. The imidazole pK_a's were computed in exactly the same way as the ones for the substituted phenols. The results are presented in Table 7. Both the fixed-charges OPLS and the polarizable force field produce results, which are close to the experimentally measured 7.0 pK_a units, even though the PFF 6.96 is in a better agreement with the experiment than the OPLS-AA 6.42. However, as noted above, the accuracy of the OPLS in the acidity constant value is a result of underestimating the gas-phase imidazolium−water dimerization energy by ca. 2.3 kcal/mol. Therefore, this fixed-charges force field is not capable of yielding an adequate response to all the variety of the electrostatic environments, unlike the polarizable force field, which is reproducing the physical behavior of the system correctly. This funding is in a perfect agreement with the explanation given above for the underestimation of the phenol pK_a values by the fixed-charges force field. If the gas-phase imidazolium−water dimerization energy were correct, the magnitude of the solvation energy of the imidazolium cation would likely be overestimated. This is exactly the same direction of error that leads to the underestimation of the phenol acidity constants described above. Therefore, the performance of the PFF is once again superior. This is a very promising result in the view of the importance of the imidazole-imidazolium system as a model for the histidine residues of proteins. To predict the pK_a shifts in protein, it is necessary to make sure that the computational model in hand is capable of reproducing the responses to the changing electrostatic environment, and this is what the PFF seems to permit better than the fixed-charges force field.

D. Acidity Constant of Methanol and Inportance of the Ewald Summation Procedure. We have also computed the absolute acidity constant for methanol. In this case, all the OPLS-AA parameters were available, and thus the only ones to produce were those for the polarizable force field CH_3O^- anion. Therefore, the comparison here is really between the OPLS-AA and the PFF. Moreover, additional calculations were preformed with the BOSS program[16] to assess the pK_a value with the OPLS-AA force field but without the Ewald summation to account for the images of the system resulting from application of the periodic boundary conditions. Solute−solvent interactions were not computed explicitly for solute−solvent distances over 10 Å. To take into the account the ionic nature of the phenoxide, the simple Born model was used, with the following term added to the total ion−solvent free energy:

$$\Delta G_{Born} = -\frac{164.0}{R_{cutoff}} = -\frac{164.0}{10} = -16.4 \text{ kcal/mol} \quad (15)$$

TABLE 8: Computed Methanol Absolute pK_a Values

method	pK_a
OPLS-AA with Ewald summation	21.8
OPLS-AA with cutoff	16.2
PFF	13.9
experiment[a]	15.5

[a] Reference 12.

TABLE 9: Absolute pK_a Values

system	fixed charges	PFF	DFT[a]	experimental
phenol	4.5	10.81	9.8	9.98[b]
p-chlorophenol	−1.06	9.87	9.6	9.38[b]
methanol	21.8	13.9	16.4	15.5[a]
imidazol	6.42	6.96	6.8	7.0[a]
average error	5.70	0.74	0.38	

[a] Reference 12. [b] Reference 13.

All the results of the methanol pK_a calculations are given in Table 8. Two facts can be observed from these data. First of all, although the error obtained with the PFF force field is equal to 1.6 and thus somewhat larger than the average one for substituted phenols, it is still significantly smaller than the OPLS/Ewald error of 6.3 units. Moreover, the overestimation of the pK_a value by the OPLS is consistent with the fact that the OPLS water−CH_3O^- dimerization energy is almost 1 kcal/mol too high, as compared to the PFF and quantum mechanical results in Table 2. Second, it is clear that the OPLS-AA number obtained with the cutoff procedure is very different from the OPLS result calculated with the Ewald summation procedure. It has been reported that the Ewald procedure is, in fact, necessary for accurate reproduction of properties of solvated molecular systems (and especially ionic species).[25] Therefore, we can conclude that employing the Ewald summation is crucial in accurate prediction of acidity constants with empirical force fields.

E. Comparison with Available Density Functional Theory (DFT) Results. Given in Table 9 is a summary comparison of the computed pK_a values with their experimental counterparts and with density functional results from ref 12. The following observations can be made. First of all, the DFT results are the closest ones to the experimental data with the average deviation of only 0.38 pK_a units. Second, the PFF results for the systems studied are also rather accurate with the average error of 0.74 units. The former and the latter are both within the ca. 0.8 units target accuracy. Though the DFT results are of a better level of accuracy, the PFF is an empirical technique, requiring no direct quantum mechanical calculations in the process of the simulations. Therefore, it can be employed more easily in biologically relevant modeling of proteins, protein−ligand complexes, and other large systems. Finally, the fixed-charges modified OPLS-AA yields results deviating by more than 5 pK_a units from the experiment. The only case of a close agreement is the imidazole one, and the agreement here is achieved at the price of underestimating the imidazolium−water gas-phase dimerization energy by about 2.3 kcal/mol. This is a great illustration for the fact that fixed-charges force fields can perform well in the gas phase or in condensed state, but not in both at the same

time, because fixed charges cannot readjust themselves depending on the environment. Therefore, the PFF can be expected to be superior in obtaining pK_a shifts in proteins, because this is precisely the case when the electrostatics of a residue should respond adequately to differences in the surroundings.

IV. Conclusions

Results of the work presented in this paper demonstrate that empirical force fields can be used to predict absolute acidity constants with the accuracy of ca. 0.8 units. However, it has also been demonstrated that such empirical force fields have to incorporate an explicit representation of electrostatic polarization to be sufficiently accurate. The modified OPLS-AA force field could only provide absolute acidity constants values within ca. 5 units from their experimental counterparts. Moreover, it has been shown by computing pK_a values of methanol in water that carrying out Ewald is necessary if an explicit water model with periodic boundary conditions is employed.

It should be specifically noted that the above accuracy in assessing absolute pK_a values of substituted phenols was achieved without any explicit parametrization of the deprotonated ionic species in solution. The same is true for the positively charged imidazolium system. The results can probably be improved even further by additional fitting of ionic solvation energies to available experimental data. Moreover, such fitting can be carried out with the polarizable force field without any significant deterioration of the gas-phase dimerization results, whereas the latter is not possible with fixed-charges models, which cannot readjust their electrostatic properties according to changes in the environment.

The proof of the concept provided by this project opens road to parametrizing a more comprehensive polarizable force field for organic molecules. Moreover, the scope of the systems can now be broadened to include such important and generally interesting cases as computing pK_a values of proteins and other biochemically important molecular systems. The next logical step will be to predict pK_a shifts of histidine residues, such as, for example, those studied experimentally in the work reported in ref 1d.

Acknowledgment. This work was supported in part by the Central Michigan University Faculty Research and Creative Endeavors (FRCE), Grant No. 48650.

Supporting Information Available: Tables of all the polarizable force field parameters not previously published and all refitted modified OPLS-AA parameter values. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) (a) Ohno, K.; Kamiya, N.; Asakaw, N.; Inoue, Y.; Sakurai, M. *Chem. Phys. Lett.* **2001**, *341*, 387. (b) Himo, F.; Noodleman, L.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Phys. Chem. A* **2002**, *106*, 8757. (c) Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D. *Proteins* **2002**, *48*, 388. (d) Huyghues-Despointes, B. M. P.; Thurlkill, R. L.; Daily, M. D.; Schell, D.; Briggs, J. M.; Antosiewicz, J. M.; Pace, C. N.; Scholtz, J. M *J. Mol. Biol.* **2003**, *325*, 1093.

(2) See, for example: Hemmateenejad, B.; Safarpour, M. A.; Taghavi, F. *Theochem* **2003**, *638*, 183 and references therein.

(3) For representative publications see: (a) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219. (b) Mehler, E. L.; Guarnieri, F. *Biophys. J.* **1999**, *77*, 3. (c) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400.

(4) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(5) (a) Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195. (b) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(6) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(7) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730 and references therein.

(8) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

(9) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(10) Jaguar v3.5, Schrödinger, Inc. Portland, OR, 1998.

(11) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. *J. Comput. Chem.*, **2002**, *23*, 1515.

(12) Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. *J. Phys. Chem. A* **2002**, *106*, 1327.

(13) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421.

(14) SIM/PFF, Schrödinger, Inc. Portland, OR, 2002.

(15) Smith, W. *CCP5 Newsletter* **1998**, *46*, 18.

(16) Jorgensen, W. L. BOSS Version 4.3; Yale University, New Haven, CT, 2001.

(17) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.

(18) (a) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926. (b) Cieplak, P.; Kollman, P.; Lybrand, T. *J. Chem. Phys.* **1990**, *92*, 6755. (c) Bernardo, D. N.; Ding, Y.; Krogh-Jespersent, K.; Levy, R. *J. Phys. Chem.* **1994**, *98*, 4180. (d) Fox, T.; Kollman, P. A. *J. Phys. Chem. B.* **1998**, *102*, 8070. (e) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910. (f) Sorescu, D. C.; Rice, B. M.; Thompson, D. L. *J. Phys. Chem. A* **2001**, *105*, 9336.

(19) Tsui, H. H. Y.; van Mourik, T. *Chem. Phys. Lett.* **2001**, *350*, 565.

(20) McDonald, N. A.; Jorgensen, W. L. *J. Chem. Phys. B* **1998**, *102*, 8049.

(21) Nagy, P. I.; Durant, G. J.; Smith, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 2912.

(22) See, for example, Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J. Am. Chem. Soc.* **1987**, *109*, 6858.

(23) Dang, L. X.; Chang, T.-M. *J. Chem. Phys.* **1997**, *106*, 8149.

(24) See, for example, Kaminski, G. A.; Zhou, R.; Friesner, R. A. *J. Comput. Chem.* **2003**, *24*, 267.

(25) See, for example: (a) Roberts, J. E.; Schnitker, J. *J. Phys. Chem.* **1995**, *99*, 1322. (b) Kalko, S. G.; Sese, G.; Padro, J. A. *J. Chem. Phys.* **1996**, *104*, 9578. (c) Lisal, M.; Kolafa, J.; Nezbeda, I. *J. Chem. Phys.* **2002**, *117*, 8892.