Published in final edited form as:

J Comput Chem. 2012 November 15; 33(30): 2388–2399. doi:10.1002/jcc.23074.

Calculating pKa values for substituted phenols and hydration energies for other compounds with the first-order Fuzzy-Border continuum solvation model

Ity Sharma and George A. Kaminski*

Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, MA 01609

Abstract

We have computed pKa values for eleven substituted phenol compounds using the continuum Fuzzy-Border (FB) solvation model. Hydration energies for 40 other compounds, including alkanes, alkenes, alkynes, ketones, amines, alcohols, ethers, aromatics, amides, heterocycles, thiols, sulfides and acids have been calculated. The overall average unsigned error in the calculated acidity constant values was equal to 0.41 pH units and the average error in the solvation energies was 0.076 kcal/mol. We have also reproduced pKa values of propanoic and butanoic acids within ca. 0.1 pH units from the experimental values by fitting the solvation parameters for carboxylate ion carbon and oxygen atoms. The FB model combines two distinguishing features. First, it limits the amount of noise which is common in numerical treatment of continuum solvation models by using fixed-position grid points. Second, it employs either second- or first-order approximation for the solvent polarization, depending on a particular implementation. These approximations are similar to those used for solute and explicit solvent fast polarization treatment which we developed previously. This article describes results of employing the first-order technique. This approximation places the presented methodology between the Generalized Born and Poisson-Boltzmann continuum solvation models with respect to their accuracy of reproducing the many-body effects in modeling a continuum solvent.

Keywords

pKa calculations; hydration energies; continuum solvation model; Poisson-Boltzmann; Generalized Born

I. Introduction

Since many organic, physical and almost all biophysical and biochemical processes take place in water, an adequate representation of the solvent is crucial in computer simulations of such systems. Some properties simply cannot be addressed without having an adequate water model for fundamental reasons. Calculating pKa values is one example of such simulations. Accurate assessment of acidity constants is both an important and a challenging task. Success of many applications depend on robustness of these calculations, including those in pharmaceutical industry and related to protein-ligand binding in general, as protein structure and function depend on the states of ionizable residues. At the same time, the problem of calculating pKa values correctly is not trivial as it involves finding a careful balance between two components of almost equally large magnitude, the bond breaking

Corresponding author: gkaminski@wpi.edu.

(gas-phase) deprotonation energy and the energy of the ionic hydration. Therefore, utilizing a good model for the surrounding water is crucial in pKa calculations.

There are two main ways to simulate aqueous environment around a solute. One of them consists of representing water molecules explicitly. While this method has proved to be successful in a number of studies, it suffers from one significant drawback. Even a small molecular system has to be surrounded by hundreds of water molecules, and this number grows rapidly as the system protein size increases. There are also other issues which may preclude use of an explicit solvation model for particular tasks, for example, the need to approximate an essentially infinitely large bulk solvent with a finite number of solvent molecules.

This is why continuum solvation models have proven to be of great help in assessing solvation free energy. Our lab is just one of those utilizing the benefits of such models, and we have been able to do so in a number of projects, including those which involve calculating acidity constant values for both proteins and smaller molecular systems.¹⁻³

The two most common continuum solvent models in the surface formulation are the Poisson-Boltzmann and the Generalized Born models. In the both models, the whole space is divided into a part occupied by the solute and the rest of the space which is taken up by the solvent. The interface between these two parts of the space is assumed to have a certain charge density. The electrostatic component of the solute-solvent interactions is assumed to be the most important one. The Poisson-Boltzmann continuum solvation technique is based on the Poisson-Boltzmann equation which relates the electrostatic potential to the integral containing this surface density over the interface. In turn, this field affects the surface density itself, as any two different points of the surface polarize each other, and the resulting formalism leads to a self-consistent equation not entirely dissimilar from the regular electrostatic polarization equation.⁴ While the above method permits the exact answer to the problem of finding the electrostatic potential of the solvent as interacting with the solute, and thus finding the exact value of the solvation energy, a more approximate Generalized Born formalism can also be used.⁴

The Poisson-Boltzmann approach relies on solving the corresponding equations on a numerical grid, and some versions of Generalized Born model use a grid to compute effective Born radii. This can and does lead to two kinds of problems. First, even a small movement of a solute can lead to generation of a new grid, in which a point is added, omitted or significantly shifted, depending on the specific solute geometry and the rules applied in grid building. This leads to a noise in the solvation energy which can be as high as several kcal/mol, which presents obvious problems in energy minimizations and thermodynamic sampling, especially molecular dynamics, because of the non-smoothness of the solvation energy as a function of the solute coordinates. The second problem arises from the fact that convergence of the self-consistent Poisson-Boltzmann problem is not always achieved automatically on a numerical grid. Methods have been suggested to address the above issues. For example, the numeric Poisson-Boltzmann solution can be sought on a fixed equally spaced grid which is built independently of the solute coordinates.⁵ Smoothing, antialiasing and careful choosing of the grid parameters can also be employed to ensure convergence.^{5,6}

We employed the Fuzzy-Border (FB) continuum solvation methodology which combines the following two features. First, we follow the abovementioned solution and make use of a fixed-position equally spaced 3-D grid with position of the nodes not dependent on the solute coordinates. Second, we propose an approximation to the full-scale Poisson-Boltzmann procedure which is similar to our previously developed fast second-order

polarization technique for solutes and explicit solutions and pure liquids.^{7,8a} It truncates the self-consistent procedure of solving the complete Poisson-Boltzmann equation at either the first or the second iteration. While this method increases the computational speed as compared to the full-scale Poisson-Boltzmann formalism, its main advantage in the case of applying it to a continuum solvent is in reducing the self-consistent problem to an analytical one. The resulting model, while being an approximation as compared to the true Poisson-Boltzmann technique, is still functionally closer to reproducing the true many-body nature of the solvation energy than the Generalized Born method.

Let us explicitly mention that the FB method is distinct enough from the other continuum solvation techniques available from other research groups. Unlike the PCM models,⁹ it is geared toward use with empirical force fields and not quantum mechanics. It contains no conductor-like screening of COSMO¹⁰ and is not based on the Generalized Born methodology of the SMx models.¹¹ Moreover, while there have been successful attempts in noise reduction in Poisson-Boltzmann calculations by using a fixed grid or smoothing functions,¹² as well as in Generalized Born models,¹³ it has not been combined with the iteration truncation implemented in the FB model to avoid possible convergence problems. This truncation is consistent in spirit with our fast polarization technique for explicit all-atom simulations of solutes and solvents.^{7,8a} And the differences from the standard Poisson-Boltzmann technique have been outline above.

We have applied the first-order version of the FB method, implemented by augmenting our POSSIM software suite,^{8a} to calculating pKa values of substituted phenols and to finding hydration energies of 40 other molecules simulated with the OPLS-AA force field.^{14,15} We have achieved an accuracy of 0.37 pH units and ca. 0.076 kcal/mol in the acidity constants and the hydration energies, respectively. It should be noted that this accuracy was achieved with direct fitting to the pKa and hydration energies data. Therefore, we do not claim that the OPLS/FB methodology exceeds in accuracy the best available quantum mechanical techniques, but only that it is physically correct enough to permit such a fitting and to assure a reasonable level of parameter transferability. We hope that future applications of the model in protein and other biologically relevant simulations will confirm this conclusion.

The rest of the article is organized as follows. The methodology is described in the next section. Results are presented in Section III, followed by the conclusions in Section IV.

II. Methodology

A. The General Protocol Used in the pKa calculations

We consider the following closed thermodynamic cycle in order to calculate the acidity constants:

Thus,

$$\Delta G(\text{gas}) + \Delta G_{\text{sol}}(\text{A}^-) + \Delta G_{\text{sol}}(\text{H}^+) - \Delta G(\text{aq}) - \Delta G_{\text{sol}}(\text{AH}) = 0 \quad (1a)$$

$$\Delta G(\text{aq}) = \Delta G(\text{gas}) + \Delta G_{\text{sol}}(\text{H}^+) + \Delta G_{\text{sol}}(\text{A}^-) - \Delta G_{\text{sol}}(\text{AH}) \quad (1b)$$

and

$$\Delta G(\text{aq}) = \Delta G(\text{gas}) + \Delta G_{\text{sol}}(\text{H}^+) + \Delta G(\text{aq}, \text{AH} \rightarrow \text{A}^-) - \Delta G(\text{gas}, \text{AH} \rightarrow \text{A}^-) \quad (2)$$

Finally,

$$\text{pKa} = \Delta G(\text{aq}) / (2.303RT) \quad (3)$$

Since we concentrated on the accuracy of the solvation model in this work, the values of $\Delta G(\text{gas})$ were taken from literature. Thus, we were calculating the $\Delta G(\text{aq}, \text{AH} \rightarrow \text{A}^-)$ and $\Delta G(\text{gas}, \text{AH} \rightarrow \text{A}^-)$ terms in Equation 2. They were found as differences of the corresponding energies, just as we did in the previous studies described in References 1–3 and 16. Moreover, since the proton free energy of solvation $\Delta G_{\text{sol}}(\text{H}^+)$ has the same value for any molecular system, a carefully determined literature result¹⁷ of -269.0 kcal/mol was used. This choice was the same as in our previous work on calculating absolute pKa values for substituted phenols with explicit solvation model and a polarizable force field.¹⁶

Values of the energies $G(\text{A}^-)$, $G(\text{A-H})$, $G(\text{Acid}^-)$, and $G(\text{Acid-H})$ were obtained via geometry optimizations with the Fuzzy-Border (FB) model. All the geometry optimizations were carried out with our POSSIM software suite^{8a,18} which was modified to permit the calculations with the continuum solvent.

The pKa values were obtained at 298.15 K.

B. The Electrostatic Component of the Solvation Energy

The overall solvation energy was calculated as a sum of the electrostatic and non-polar terms:

$$\Delta G(\text{solv}) = \Delta G(\text{el}) + \Delta G(\text{np}), \quad (4)$$

where the $\Delta G(\text{el})$ and $\Delta G(\text{np})$ terms stand for the electrostatic and the non-polar parts of the solvation energy, respectively. The electrostatic part of the energy was calculated by using an approximation to the Poisson-Boltzmann formalism. The Poisson-Boltzmann method as such is described, for example, in Reference 4. Briefly, let us consider the solute-solvent surface, as shown on Figure 1.

Here $\mathbf{E}_1$ and $\mathbf{E}_2$ are the values of the electrostatic field in the solute and solvent regions, respectively, and $\mathbf{n}$ is the unit vector normal to the interface. The continuity of the normal component of the electric displacement across the surface requires that:

$$\varepsilon_1 \mathbf{E}_1 \cdot \mathbf{n} = \varepsilon_2 \mathbf{E}_2 \cdot \mathbf{n}, \quad (5)$$

where ε_1 and ε_2 stand for the dielectric constants inside and outside of the solute, respectively. Setting $\varepsilon_1 = 1$, we can rewrite Equation 5 as:

$$\mathbf{E}_1 \cdot \mathbf{n} = \varepsilon \mathbf{E}_2 \cdot \mathbf{n}, \quad (6)$$

According to the Gauss law,

$$4\pi\sigma = (\mathbf{E}_2 - \mathbf{E}_1) \cdot \mathbf{n}, \quad (7)$$

with σ being the surface charge density at the interface. Combining equations 6 and 7,

$$\sigma = -\frac{1}{4\pi}(1 - 1/\epsilon)\mathbf{E}_1 \cdot \mathbf{n} \quad (8)$$

Since the electrostatic field $\mathbf{E}_1$ itself depends on the surface charge density distribution, Equation 8 describes a self-consistent problem, just like the general electrostatic polarization case. The electrostatic part of the solvation energy can be then found as:

$$\Delta G(el) = \frac{1}{2} \int_S \sigma \varphi^0 d^2r \quad (9)$$

Here φ^0 represents the electrostatic potential created by the charges of the solute only (not by the polarized solvent) and the integration is carried out over the whole solute-solvent interface surface.

When the equation is solved numerically, the surface is represented by a discrete set of points i . In this case, Equations 8 and 9 are approximated by Equations 10 and 11:

$$q_i = -\frac{1}{4\pi}(1 - 1/\epsilon)\mathbf{E}_{1,i} \cdot \mathbf{n}_i \quad (10)$$

$$\Delta G(el) = \frac{1}{2} \sum_i q_i \varphi_i^0 \quad (11)$$

The electrostatic field $\mathbf{E}_{1,i}$ is calculated as:

$$\mathbf{E}_{1,i} = \sum_j \frac{q_j \mathbf{R}_{ij}}{R_{ij}^3} + \sum_{k \neq i} \frac{q_k \mathbf{R}_{ik}}{R_{ik}^3} \quad (12)$$

The first sum is taken over the solute charges (this expression can be easily extended to include higher-order multipoles), while the second one goes over the other solute-solvent interface points. $\mathbf{R}_{ij}$ stands for the vector from point j to point i . Specific details about our choice of the numeric grid points will be discussed in the next subsection.

At the point, we will introduce our fast polarization approximation, which constitutes an integral part of the Fuzzy-Border (FB) model. Equations 10 and 12 together form a self-consistent problem which can be solved iteratively. The first step consists of replacing $\mathbf{E}_{1,i}$ with $\mathbf{E}_{1,i}^0$, the field created by the solute only and not by the polarized solute-solvent interface. If we stop at this stage, we obtain our first-order approximation. It still contains many-body interactions, since the electrostatic field is a vector quantity and it contains contributions from all the solute charges. But the problem is now analytical, not self-consistent, and the convergence is no longer an issue. If we now recompute the electrostatic field $\mathbf{E}_{2,i}$ taking into the account the field created by the interface charges using Equation 12, obtain the new charges with the Equation 10, but do not perform the next iteration, we

obtain the second-order model, which is still safe from the electrostatic charges convergence problems since there are only two iterations and magnitudes of the charges cannot increase beyond the value they achieve at the first of the second iteration. The complete formulation of the Fuzzy-Border model is given in the next subsection.

C. Choosing the Numerical Grid to Represent the Solute-Solvent Interface

We use a fixed cubic three-dimensional equally-spaced grid to minimize the noise resulting from grid rebuilding after moving a solute atom or a group of atoms. The interface between solute and solvent is assumed to consist of points with distances from $R - \Delta$ to $R + \Delta$ from the solute atom, as shown on Figure 2.

The values of parameters R and Δ can be different for different solute atomtypes. Only those solvent points that were located in the solvent-accessible region were used. Therefore, the overall solvation surface is constructed with the knowledge of positions and solvation parameters of all the solute atoms. A point is defined as solvent-accessible if there is no solute atom i which would be closer to the grid point than $R_i - \Delta_i$. Since the “real” radius of the solvation surface is R_i and not $R_i - \Delta_i$ or $R_i + \Delta_i$, the surface points j corresponding to the solute atom i , has weights w_j associated with them calculated as:

$$w_{j=i} = a_{0,i} \left[\left(\frac{R_{ij} - R_i}{\Delta_i} \right)^4 - 2 \left(\frac{R_{ij} - R_i}{\Delta_i} \right)^2 + 1 \right] \quad (13)$$

Here R_{ij} is the distance between the solvent grid point and the solute atom, $a_{0,i}$ is a parameter which depends on the solute atomtype, and the whole weight is maximum at the nominal solvation radius R_i and decreases to zero at distances $R_i - \Delta_i$ and $R_i + \Delta_i$. Since one grid point can be within the solvation surface range of more than one solute charge, several weights w_{ji} may be needed for the same grid point j and different solute atoms i .

The weights are normalized in the following way:

$$w_j = \frac{\sum_i w_{ji}^2}{\sum_i w_{ji}} \quad (14)$$

The unit normal vector $\mathbf{n}_{ji}$ for each solute atom i and the grid point j is assumed to be in the direction from the solute atom to the grid point. The overall unit normal vector for the grid point is then calculated as:

$$\mathbf{n}_j = \sum_i \mathbf{n}_{ji} w_{ji} \quad (15)$$

We can now write down the overall FB continuum solvation formalism. Once the solvation surface grid points j are defined as described above, the zeroth-order electrostatic field at those points is found as:

$$\mathbf{E}_{2,j}^0 = \sum_i \frac{q_i \mathbf{R}_{ij}}{R_{ij}^3} \quad (16)$$

The summation goes over all the solute points. The first-order FB charge on the grid point j is then found as shown in Equation 17:

$$q_j^I = -A_{scale} \frac{1}{4} \pi (\epsilon - 1) w_j \mathbf{E}_{1,j}^0 \cdot \mathbf{n}_j \quad (17)$$

A_{scale} is a scaling factor and an adjustable parameter of the theory. The first-order electrostatic part of the solvation energy can then be calculated as:

$$\Delta G(el)^I = \frac{1}{2} \sum_j q_j^I \varphi_j^0 \quad (18)$$

If the second-order approximation is to be produced, the first-order electrostatic field is found as described in Equation 19:

$$\mathbf{E}_{1,j}^I = \mathbf{E}_{1,j}^0 + A_{self} \sum_{k \neq j} \frac{q_k \mathbf{R}_{kj}}{R_{kj}^3} \quad (19)$$

with the additional summation done over the solute-solvent interface points k , and A_{self} being another adjustable parameter (with its value being the same for all the grid points). Equation 17 is then modified to include the first-order and not the zeroth-order field:

$$q_j^I = -A_{scale} \frac{1}{4} \pi (\epsilon - 1) w_j \mathbf{E}_{1,j}^I \cdot \mathbf{n}_j, \quad (20)$$

and the resulting second-order energy is found in accordance with Equation 20:

$$\Delta G(el)^{II} = \frac{1}{2} \sum_j q_j^{II} \varphi_j^0 \quad (21)$$

It should also be noted that effect upon the energy of shifting the position of the fixed cubit grid by one half of the spacing between the grid points is assessed every time the hydration energy is calculated, and the energy is averaged with respect to this transformation. Finally, the electrostatic part of the energy, regardless of whether it is calculated with the first- or second-order model, is multiplied by 332.0657418 in order to obtain the final result in kcal/mol.

We believe that this formulation of the electrostatic part of the Fuzzy-Border (FB) method places it between the Poisson-Boltzmann (PB) and Generalized Born (GB) models in the sense of being close to the physical correctness of PB one. The difference between the FB and PB techniques is clear – the former truncates some advanced iterations in the self-consistent procedure which is usually involved in solving the equations of the latter one. As far as the GB continuum solvation model is concerned, its surface formulation includes two terms:⁴

$$U = \sum_k U_{se}(q_k, \mathbf{r}_k) + \sum_{i \neq j} U_{pr}(q_i, q_j, \mathbf{r}_i, \mathbf{r}_j), \quad (22)$$

Where the second one represents screened pairwise electrostatic interactions:

$$U_{pr} = -\frac{1}{2}(1-1/\epsilon) \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2} e^{-D}}, \quad \alpha_{ij} = \sqrt{\alpha_i \alpha_j}, \quad D = \frac{r_{ij}^2}{(2\alpha_{ij})^2} \quad (23)$$

and the first term stands for the single-charge solvation energy:

$$U_{se} = -\frac{q_k^2}{8\pi} (1-1/\epsilon) \int_S \frac{(\mathbf{R}-\mathbf{r}_k) \cdot \mathbf{n}(\mathbf{R})}{|\mathbf{R}-\mathbf{r}_k|^4} d^2\mathbf{R} \quad (24)$$

It should be noted that this first term contains no reference to electrostatic interactions between the charge q_k and other charges of the solute (except in an indirect form, by shaping the solute-solvent interface S). And the value of the U_{se} term is what ultimately defines the Born parameter α in Equation 23. Therefore, the Generalized Born formalism contains no directly defined many-body electrostatic interactions, while the Fuzzy-Border one includes this part of the physical picture explicitly as the electrostatic field is calculated with all the solute charges, even in the first-order FB model. Thus, we place the FB technique between its PB and GB counterparts.

At the same time, it should be emphasized very strongly that such a relation among the physical bases of these three methods does not at all mean that the actual accuracy of computing hydration energies will be the best with the Poisson-Boltzmann formalism, somewhat worse with the Fuzzy-Border method and even worse with the Generalized Born one (as is witnessed by the data presented in the next Section of this paper). There are many factors which are affecting the accuracy. They include the size of the fitting set, the number of parameters used in the fitting, the fitting technique, etc. Generally speaking all physically reasonable models should be able to produce reasonably accurate results for solvation energies of small molecular systems. And our aim is, first and foremost, to demonstrate that the Fuzzy-Border technique is in fact one of such reasonable and robust approaches. Moreover, the FB model is formulated without the true self-consistency of the true Poisson-Boltzmann method. Thus, in principle, all the expressions for the surface charges can be written analytically. This should make it possible to directly derive analytical gradients for this model in the future.

D. The Non-Polar Part of the Solvation Energy

The non-polar part of the solvation energy was calculated as a sum of two terms, one with a positive and one with a negative contribution:

$$\Delta G(np) = \sum_j w_j A_{np} - \sum_i \sum_j w_j \frac{A_i^{LJ}}{R_{ij}^6}, \quad (25)$$

The first term contains a sum taken over all the grid points. This is essentially the overall solvent-accessible surface area (SASA) contribution which is commonly employed in continuum solvation models (see for example Reference 19) adapted to the Fuzzy-Border formalism. The second term is calculated with a double summation going over all the grid points j and all the solute atoms i . It approximates the attraction part of the Lennard-Jones energy for interactions between the solute and solvent atoms.

Once the electrostatic and non-polar terms of the solvation energy are found, the overall solvation energy can be calculated according to Equation 4.

III. Results and Discussion

Having tried different combinations of the values of the solvation parameters, we finally decided to use the grid spacing of 0.25Å in each dimension. All calculations were carried out for water as the solvent. The radius of the solvent molecule was set at 1.4Å, and the value of the dielectric constant was $\epsilon = 80.4$.²⁰ All the calculations presented in this article were done with the first-order Fuzzy-Border model. The scaling factor $A_{scale} = 0.07069$. The non-polar factor A_{np} in the part of the non-polar solvation energy $\Delta G(np)$ which corresponded to the positive contribution associated with the solvent-accessible surface area, had the same value of 3.747 for all the solute atoms and all the grid points at the solute-solvent interface. Finally, we eventually picked a value of the parameter Δ (the “fuzziness” of the border) of 0.25Å for all the atomtypes considered

OPLS-AA force field^{14,15} was used for all the solute parameters employed in this study. The calculations were carried out with the POSSIM software^{8a} modified to include the Fuzzy-Border (FB) continuum solvent model. The solvation energies were found as:

$$\Delta G(solv, calculated) = E(solvated) - E(gas), \quad (26)$$

where $E(solvated)$ and $E(gas)$ are the computed energies of the system in the aqueous solution and in gas-phase, respectively. Both these energies were obtained by energy minimizations, therefore, the resulting calculated hydration energy could differ from the nominal solvation energy obtained for the hydrated solute molecule in accordance with the Equation 3. All the geometry optimizations were completely unconstrained, and when applicable, the lowest-energy conformations were used.

The calculated pKa values are listed in Table 1, the related hydration energies are given in Table 2, and the remaining energies of hydration are presented in Table 3. The values of the hydration parameters are shown in Table 4.

In addition to calculating the hydration energies with the POSSIM/FB formalism, we have also used Poisson-Boltzmann (PBF) and Generalized Born (SGB) continuum solvent models as implemented in IMPACT software suite^{8b,c} for comparison. The results of these additional calculations are presented in Table 3.

A. Hydration Energies of Benzene, Phenol and Phenoxide and the Phenol pKa Value

These calculations required that we would first develop parameters for the unsubstituted phenol. We took the natural course in developing aromatic parameters by starting with benzene. The final FB hydration radii for the benzene carbon and hydrogen atomtypes were 2.050Å and 1.320Å, respectively. The values of the parameter a_0 for the aromatic carbon and hydrogen atoms were set at 0.02748 and 0.01034, and we tried to keep these unchanged for all the further development of the carbon and hydrogen FB parameters to avoid overparameterization. The best values of the A^{LJ} parameters for the aromatic carbon and hydrogen atoms were found to be 161.5 and 10.00 kcal/mol·Å⁶. The error in the calculated hydration energy of benzene (−0.866kcal/mol) as compared with the experimental data (−0.87kcal/mol)¹⁹ was less than 0.01kcal/mol.

We then produced FB hydration parameters for the −OH group in phenol. The best performance was found with the phenol oxygen radius of 1.700Å. The parameter a_0 was equal to 0.02748 (the same as for the aromatic carbons), and the Lennard-Jones factor was 143.3 kcal/mol·Å⁶. For the hydrogen, the values were 1.300Å, 0.01034 and 1.327. Moreover, this set of hydrogen parameters was found to be suitable for the other polar hydrogens as well, and the only difference in the aliphatic hydrogen was in a slight change

in the radius. The error in the hydration energy of phenol was only 0.01 kcal/mol, with the calculated value of -6.612 kcal/mol and the reference of -6.62 kcal/mol.¹⁹

The next step in our pKa calculations required hydration parameters for the deprotonated phenol, C₆H₅O⁻, to be produced. We used the OPLS-AA atomtype 420 for the oxygen atom, with the Lennard-Jones parameters of $\sigma = 3.15$ Å and $\epsilon = 0.25$ kcal/mol (the O⁻ in CH₃ O⁻). The hydration parameters for the aromatic ring remained unchanged, as we were only fitting those for the oxygen. We chose the following strategy for this fitting. The target hydration energy for the ion was chosen in such a way as to lead to the pKa value of the unsubstituted phenol to be equal to the experimental value of 9.98 pH units.¹⁷ This required the C₆H₅O⁻ system to have a hydration energy of -73.02 kcal/mol with the reference $\Delta G(\text{gas})$ of 349.0 kcal/mol.²¹ The -73.02 kcal/mol solvation energy for the ion is consistent with the experimental range of -72 – -75 kcal/mol.¹⁷ We have obtained the target hydration energy with the O⁻ radius of 2.690 Å, $a_0 = 0.01734$ and the Lennard-Jones factor of 142.9 kcal/mol·Å⁶. This led to the phenol pKa which was exactly the same as the experimental value of 9.98 pH units.

B. pKa Values of the Substituted Phenols and Hydration Energies of Related Molecules

The next series of substituted phenol systems we wanted to consider were the *ortho*- *meta*- and *para*-methyl phenol. However, in order to demonstrate transferability of our hydration parameters and to avoid overparameterization, we decided to produce parameters for aliphatic carbon and hydrogen atoms first and then to use them in the methyl phenols as well in all the other compounds which contain alkyl groups without any further modifications.

The series of species considered contained methane, ethane, propane, butane and iso-butane. The solvation parameters for all the aliphatic carbons were kept the same, and the same condition was observed for the aliphatic hydrogens. Moreover, we used the same aliphatic carbon and hydrogen hydration parameters in all cases when such a group was present, regardless of the chemical functionality of the remaining molecule (for example, in the -CH₃ groups in methyl phenol, NMA and methanol, in the -C₂H₅ group of ethanol and diethyl ether, the methyl group of the toluene molecule, etc.).

As has been mentioned earlier, the value of Δ was set to be the same for all the atomtypes we considered and equal to 0.25 Å. The radius of the carbon atom was $R = 1.9$ Å. In the OPLS formalism, the Lennard-Jones radius of an atom is $2^{1/6}$ times one half of the parameter ϵ which has a value of 3.5 Å for the OPLS aliphatic carbon. Therefore, the Lennard-Jones radius for these atoms is 1.964 Å. The solvation radius is fairly close to this number. While we did not specifically have a target of the Lennard-Jones radii as the solvation radii, it is worth noting that they were generally not too different. The value of a_0 for the aliphatic carbons in the FB model is equal to 0.02748 (the same as for the aromatic carbons and phenol oxygen), and the Lennard-Jones factor of 166.3 kcal/mol·Å⁶. The corresponding hydrogen parameter values were 1.357 Å, 0.01034 and 1.327 kcal/mol·Å⁶.

It can be seen from the data in Table 3, that the hydration energies for methane, ethane, propane, butane and *iso*-butane follow the general experimental trend. Moreover, we correctly reproduce the hydration energy of the *iso*-butane molecule to be higher than that of *n*-butane, which is not always a property represented correctly by continuum solvation models.¹⁹ The overall average error for the hydration energies of the aliphatic hydrocarbons was ca. 0.1 kcal/mol.

The pKa values for the methyl phenols were in a very good agreement with the experimental data. The calculated values for the *o*-, *m*- and *p*-methyl phenols were found to be 10.02, 9.11 and 10.50 pH units, respectively, with their experimental counterparts being 10.29, 10.08

and 10.14 units. Therefore, the general trend of the *m*-methyl phenol being more acidic was followed (although our *p*-methyl phenol is somewhat more basic than it should be), and the overall error for these three acidity constants was 0.53 pH units, well within the range for which meaningful comparison with experiment. Moreover, it should be explicitly noted that no specific fitting for methyl phenols was carried out, with the hydration parameters taken directly from the unsubstituted phenol calculations and the alkane hydration energy fitting. This attests to the robustness of our methodology and portability of the produced parameter values, which is especially important for the aliphatic groups which are present in a variety of compounds.

The next series of the compounds for which we calculated pKa values were the *o*-, *m*-, and *p*-chlorophenols. In this case, the hydration parameters for the chlorine atom were fitted specifically to reproduce the experimental acidity constants of these substituted phenols. The parameter fitting has led to the following final values of the Cl- hydration parameters: $R = 1.905 \text{ \AA}$, $a_0 = 0.00808$ and the $A^{LJ} = 250.0 \text{ kcal/mol} \cdot \text{\AA}^6$. The final calculated pKa values were 8.57, 8.51 and 10.07 pH units for the *o*-, *m*-, and *p*-chlorophenol, respectively. The experimental counterparts are 8.56, 9.02 and 9.38 units.¹⁷ The average error is just under 0.40 pH units, and the calculated value of the acidity constant generally follows the experimental trend of growing from the *ortho*- to *para*-compound.

In the case of *m*- and *p*-cyanophenols we also did the substituent ($-\text{C}\equiv\text{N}$) hydration parameter fitting with the explicit goal of reproducing the experimental pKa values of these substituted phenols. The final values of the parameters for the carbon and nitrogen atoms were $R = 1.300 \text{ \AA}$, $a_0 = 0.00600$ and the $A^{LJ} = 100.0 \text{ kcal/mol} \cdot \text{\AA}^6$ and $R = 1.825 \text{ \AA}$, $a_0 = 0.02100$ and the $A^{LJ} = 200.0 \text{ kcal/mol} \cdot \text{\AA}^6$, respectively. The calculated values of the acidity constant for the *m*- and *p*-cyanophenols were 8.48 and 7.42 pH units (as can be seen from the data in Table 1). These values have the same relative order as the experimental numbers of 8.61 and 7.95 pH units,¹⁷ and the average error is 0.33 units.

The last set of the substituted phenols which we considered was the *m*- and *p*-nitrophenols. The experimentally measured pKa values for these compounds are 8.40 and 7.15 pH units.¹⁷ We have fitted the N and O hydration parameters for the nitrogroup and obtained the values of the acidity constants of 8.14 and 7.87 units. Therefore, the *para*- isomer is more acidic, in agreement with the experiment, and the average error is 0.49 pH units. The values of the solvation parameters for the nitrogen and oxygen atoms in the nitrogroup are $R = 1.550 \text{ \AA}$, $a_0 = 0.00821$ and the $A^{LJ} = 110.0 \text{ kcal/mol} \cdot \text{\AA}^6$ and $R = 1.370 \text{ \AA}$, $a_0 = 0.00900$ and the $A^{LJ} = 1.327 \text{ kcal/mol} \cdot \text{\AA}^6$, respectively.

The average error in the pKa values of the substituted phenols, as calculated with the Fuzzy-Border model, is only 0.41 pH units. This is comparable with the best quantum error of 0.38 units for these molecules¹⁷ and is better than some lower-level quantum mechanical data from the same reference. Of course, a direct comparison of these two sets of data would not be appropriate. A big advantage of the QM calculations is that they were performed without any specific refitting for the pKa calculations, and many of our hydration parameters were fitted to reproduce the pKa values (though even in these cases the relative order of pKa values of phenols with substituents in different positions was still reasonably good, thus we can assume at least some transferability of the parameters in every single case). The values of the parameters R , a_0 and A^{LJ} were fitted for the following atomtypes: the O^- in phenoxide, Cl atom, carbon and nitrogen of the $-\text{C}\equiv\text{N}$ group, as well as nitrogen and oxygen atoms in $-\text{NO}_2$. But we have clearly demonstrated that our Fuzzy Border model can, with proper fitting of the parameters, reproduce magnitudes and relative orders of the absolute acidity constants rather well.

Moreover, listed in Table 5 are the pKa values for the unsubstituted phenol and *para*-chlorophenol, as calculated with the Fuzzy –Border continuum solvation model in this work and calculated with the explicit aqueous solvation.¹⁶ Both sets of the results were obtained for the OPLS-AA solute model. It can be easily seen that the explicit solvent simulations produce a much greater average error of 6.90 pH units, while the FB model permits to reduce the deviation to the average of only 0.35. Although, once again, this is achieved by the specific pKa-targeted fitting of the solvation parameters, this is an additional reason to believe that the Fuzzy-Border model will be a useful tool in modeling of hydrated molecular systems described by the generally very successful OPLS-AA force field.

The gas-phase deprotonation energies for the phenol and substituted phenol molecules used in this work are given in Table 6.

C. Hydration Energies for the Other Non-Phenol Compounds

Let us now consider the results of producing parameters for other compounds not related to the above pKa calculations for the substituted phenols.

We have simulated four alkenes, ethylene, 1-propene, 1-butene and butadiene. The hydration parameters for the hydrogen atoms were adopted directly from the alkanes and not refitted in any way. The sp² carbon parameters were slightly different from those for the alkane carbons, with $R = 2.015 \text{ \AA}$, $a_0 = 0.02748$ and the $A^{LJ} = 196.5$. The resulting hydration energies are listed in Table 3. It can be seen that the general agreement with the experimental data is good, the average error is only 0.14 kcal/mol, and the only mismatch in the order of these calculated solvation energies is observed for the ethylene, which is less hydrophobic than 1-propene (1.056 kcal/mol vs. 1.512 kcal/mol), while their experimental solvation energies are the same within 0.01 kcal/mol.¹⁹

We have also produced hydration parameters for the following alkynes: acetylene, 1-propyne and 1-butyne. In this case, solvation parameters for both the sp-hybridized carbons and for the hydrogens were refitted. The values for the carbon and hydrogen atoms are, respectively: $R = 2.050 \text{ \AA}$, $a_0 = 0.01700$ and the $A^{LJ} = 250.0$ and $R = 1.500 \text{ \AA}$, $a_0 = 0.02000$ and the $A^{LJ} = 1.327$. As can be seen from the data in Table 3, the average error in the solvation energy for these compounds is 0.16 kcal/mol.

Methanol and ethanol were simulated to represent aliphatic mono-alcohols. As shown in Table 3, the average error is only ca. 0.015 kcal/mol, with the trend of methanol being solvated better by about 0.1 kcal/mol reproduced correctly. The solvation radius of the oxygen atom was 1.735 Å. This correlates well with the OPLS-AA Lennard-Jones radius of 1.751 Å for this atom type.

As was mentioned above, all the polar hydrogen atoms have the same set of the first-order FB hydration parameters for OPLS solutes. Such atoms are assigned a van-der-Waals radius of zero in the OPLS-AA force field. The methanol and ethanol OH hydrogen was given a hydration radius of 1.3 Å in our model.

The next class of compound that we considered was that of ketones. We simulated hydration for acetone, 2-butanone, 2-pentanone and 3-pentanone. As with all the other cases, the hydration parameters for analogous atoms in these systems were not dependent on the specific substance (thus, for example, the oxygen in acetone had exactly the same set of FB parameters as the oxygen atom in 3-pentanone). Moreover, the oxygen parameters were rather similar to those of the alcohols. The hydration radius of the ketone oxygen was slightly larger than that of the aliphatic alcohols (1.75 Å s. 1.735 Å), otherwise the parameters were the same. The overall average error in the ketone hydration energies, as

compared with their experimental counterparts,¹⁹ was just 0.065 kcal/mol. And the trend of the slight energy magnitude reduction from acetone (calculated solvation energy of -3.900 kcal/mol and experimental result of -3.85 kcal/mol) to 3-pentanone (calculated and experimental energies of -3.537 kcal/mol and -3.41 kcal/mol, respectively) was reproduced successfully. Moreover, the success with these compounds has demonstrated a good level of transferability of both our ketone and aliphatic hydration parameter values.

We then considered the dimethyl- and diethyl-ethers. As can be seen from the table, the errors were somewhat greater in this case, though still in the acceptable range, with the both hydration energies slightly overestimated, with an average error of ca. 0.15 kcal/mol. At the same time, the oxygen solvation parameters were taken directly from the alcohol oxygen, and the methyl and ethyl parameters were the same as produced in fitting hydration energies of the saturated hydrocarbons. This is an added proof of the robustness of the method. Moreover, the trend in reduction of the magnitude of the hydration energy with transition from the dimethyl- to diethyl-ether was reproduced correctly.

The next series of compounds in Table 3 are amines – methyl amine, ethyl amine, *n*-propyl amine, *n*-butyl amine, dimethyl amine and diethylamine. The solvation parameters for the polar hydrogen atoms were the same as for the other polar hydrogens, as discussed above. Only the nitrogen solvation parameters were fitted. The resulting values are $R = 1.650 \text{ \AA}$, $a_0 = 0.00900$ and the $A^{LJ} = 262.0$ As shown in Table 3, the average error in the solvation energies is 0.12 kcal/mol, and the trend of the hydration energy magnitude decreasing in this series from methyl amine to diethyl amine is generally reproduced. Ammonia nitrogen parameters were fitted separately, with the final values of $R = 1.357 \text{ \AA}$, $a_0 = 0.00650$ and the $A^{LJ} = 122.9$. Deviation of the calculated hydration energy for ammonia (-4.301 kcal/mol) from the experimental value of -4.310 kcal/mol¹⁹ is about 0.01 kcal/mol.

We employed the benzene parameters described above together with the parameters for alkanes to calculate the hydration energy of toluene, and found it within an error of 0.04 kcal/mol from experiment. This result was obtained with no parameter refitting.

Acetamide and NMA are compounds which are important in their own right and also have a great significance as building blocks of proteins. This is especially true for the NMA, which essentially represents the repeating unit in the protein and peptide backbones. Fitting parameters for amides often represents a challenge, and it is not unusual to have different parameter sets for the acetamide and NMA cases.²² Our solvation parameters were the same for similar atoms in these two amides except for the nitrogens (which also have different atomtype designations in the OPLS-AA). The methyl groups had the standard FB solvation parameters for both the carbon and hydrogen atoms. The hydration parameters for the polar hydrogens were also preserved at the same values as for the previously discussed molecule. We have discovered that the aliphatic version of the carbon hydration parameters and the aliphatic alcohol values for the oxygen work well for the amide C=O group. The only adjusted parameters were those of the amide nitrogen atoms. In the both cases, the radii were greater than those for the amines (1.770 Å and 1.870 Å for the NMA and acetamide nitrogens, respectively, compared to the 1.650 Å for the R-NH₂). As shown in Table 3, the average error in the hydration energy of the amides was 0.05 kcal/mol.

In order to include examples of heterocyclic molecules and to have parameters for the histidine and tryptophan residues in the potential future development of a FB solvation parameter set for peptides and proteins, we carried out fitting for 4-methyl-imidazole and 3-methyl-indole. Once again, a good transferability of the solvation parameters was observed. All the aromatic carbon and hydrogen atomtypes retained the same values as those in benzene and toluene. The same is true with relation to the aliphatic carbons and hydrogens

in these two molecules and all the other saturated hydrocarbon groups. The polar hydrogens had our standard FB polar hydrogen hydration parameter values. Only the nitrogen parameter values were refitted, and the parameters for the N(H) in both these molecules were the same. And the average error in hydration energy for the two compounds was less than 0.01 kcal/mol.

Both the thiol and the sulfide molecules, represented by the methane- and ethane-thiol and dimethyl- and diethyl-sulfides, respectively, used our standard hydration parameters for the methyl and ethyl groups, but the sulfur parameters were different for the two groups of compounds. The solvation radii of the both S atomtypes were greater than that of the aliphatic carbon. Values of these parameters are listed in Table 4. The average error in the hydration energy was only about 0.01 kcal/mol.

Finally, we produced FB solvation parameters for two acids, acetic (CH_3COOH) and propanoic ($\text{C}_2\text{H}_5\text{COOH}$). The aliphatic tails had the standard FB parameter values for both carbon and hydrogen atoms. In the $-\text{COOH}$ groups, the hydration parameters for the $-\text{OH}$ part were transferred from the FB aliphatic alcohols without change, and the $=\text{O}$ atom had parameters adopted directly from the acetone oxygen value. The only adjusted parameter set was that of the carbon, with the radius slightly greater than that of the aliphatic carbon and the Lennard-Jones parameter reduced by ca. 10%. As can be seen from the data in table 3, the resulting average error in the hydration energies for the acid molecules was only about 0.02 kcal/mol. We believe that this, once again, is evidence, albeit still anecdotal, of the generally good transferability of the FB parameters.

D. Comparison of Fuzzy-Border Hydration Energies with Poisson-Boltzmann and Generalized Born Results

Let us now compare the Fuzzy-Border (FB) results with those obtained with the Poisson-Boltzmann (PBF) and Generalized Born (SGB) formalisms. All the energies are listed in Table 3. It can be observed that the average error in the FB solvation energies (0.076 kcal/mol) is significantly smaller than those of the PBF and SGB models (0.527 kcal/mol and 0.639 kcal/mol, respectively). The Poisson-Boltzmann continuum solvation gives a somewhat better result than the Generalized Born one, as can be expected from the more physically grounded formalism of the former. It should be emphasized very strongly that the smaller average error afforded by the Fuzzy-Border continuum solvent does not mean that the FB methodology is intrinsically better than the PBF and SGB ones. The latter models implemented in IMPACT have been parameterized for a larger set of molecules, and thus a somewhat greater error is natural to observe. What we do conclude though is that our technique is robust enough, and, even if extending the currently available parameter sets to more classes of compounds does reduce the overall average accuracy, it is still likely to stay within the respectable range observed for the Poisson-Boltzmann and Generalized Born methods.

In order to further compare the FB, PBF and SGB results, we are also listing the values of the electrostatic component of the hydration energy as computed with these techniques in Table 7. We believe that the conclusion that can be drawn by looking at these results is as follows. Generally speaking, the Poisson-Boltzmann electrostatic energy tends to be somewhat more negative than the Fuzzy-Border one, with the average signed difference of -0.662 kcal/mol. At the same time, the Generalized Born electrostatic component is on average about the same as the Fuzzy-Border one (only 0.065 kcal/mol more negative). At the same time, deviations for particular components can be noticeably greater, and even the PBF model does not yield uniformly more negative results. For example, the PBF electrostatic component for the compounds containing sulfur can be about two times smaller than that produced with the Fuzzy-Border model. At the same time, all the solutes were

modeled with the same OPLS-AA force field. Therefore, we conclude that the differences in the electrostatic hydration energies given by the three continuum solvent models are not representing greater or smaller deviations from a physically correct set of results but rather are simply following from differences in fitting techniques employed to produce overall hydration energies, and the quality of these overall energies is adequately good for all the three methods (as can be seen from the results in Table 3).

E. Absolute Acidity Constants for Propanoic and Butanoic Acids

Finally, we have fitted hydration parameters for reproducing absolute pKa values of propanoic and butanoic acids. These compounds were chosen as relevant in calculating pKa shifts of aspartic and glutamic acid residues of proteins (such as those calculated in Reference 1). Relevant results are presented in Table 8. The only two atom types for which solvation parameters were fitted are the -COO^- carbon and oxygen in the deprotonated forms of the acids.

The hydration energies of the protonated propanoic and butanoic acids as calculated with POSSIM/FB were equal to -6.443 kcal/mol and -6.197 kcal/mol, respectively (as can be seen from Table 8). These results deviate by an average of only 0.1 kcal/mol from the experimental numbers. These data were obtained with the same parameters as the hydration energies for CH_3COOH and $\text{C}_2\text{H}_5\text{COOH}$, no refitting was done. The overall values of the acidity constants for these acid were 4.90 and 4.66 pH unites (compared to the reference 4.87 and 4.83 units). The agreement is good, but it was achieved with the direct fitting of the parameters for the deprotonated acids, therefore the ultimate quality of this result will have to be tested by using the same parameters with other systems in future work.

IV. Conclusions

We have applied the first order Fuzzy-Border (FB) solvation model with the OPLS-AA solute to calculate absolute pKa values of several substituted phenols and hydration energies of a number of other compounds. The compounds were chosen to represent several classes which are important not only in themselves but also as building blocks in protein simulations. The FB model was implemented in a modified version of our POSSIM software suite for molecular simulations.

The overall average unsigned error in the calculated acidity constant values was equal to 0.41 pH units and the average error in the solvation energies was ca. 0.08 kcal/mol.

While these results were achieved with fitting of the hydration parameters to the specific pKa and hydration energy targets, the results still prove that the physical and numerical basis of the model is robust enough to permit such a good level of the performance, and the model can be expected to work well in further simulations of organic and biophysical systems. The parameter transferability also seems to be good.

The features of the FB model include utilizing a fixed three-dimensional grid for finding continuum solvation energy and an approximation to the Poisson-Boltzmann formalism which is designed to speed up the calculations and, more importantly, to remove the unfortunate potential problems which accompany the accuracy of the complete self-consistency of the standard Poisson-Boltzmann method and thus to avoid any issues related to convergence of the solvation energy. At the same time, the overall FB technique is still closer to the exact electrostatic model than the Generalized Born approximation.

We plan to continue developing Fuzzy-Border parameters in the direction of creating a complete FB continuum solvation model for proteins.

Acknowledgments

The project described was supported by Grant Number R01GM074624 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. We thank Schrödinger, LLC for the IMPACT software suite.

References

- MacDermaid CM, Kaminski GA. *J Phys Chem B*. 2007; 111:9036–9044. [PubMed: 17602581]
- Click TH, Kaminski GA. *J Chem Theory Comput*. 2009; 5:2935–2943. [PubMed: 20209038]
- Click TH, Ponomarev SY, Kaminski GA. *J Comp Chem*. 2012; 33:1142–1151. [PubMed: 22370900]
- Ghosh A, Rapp CS, Friesner RA. *J Phys Chem B*. 1998; 102:10983–10990.
- see for example: Bruccoleri RE, Novotny J, Davis ME, Sharp KA. *J Comp Chem*. 1997; 18:268–276.
- Grant JA, Pickup BT, Nicholls A. *J Comp Chem*. 2001; 22:608–640.
- Kaminski GA, Zhou R, Friesner RA. *J Comp Chem*. 2003; 24:267–276. [PubMed: 12548718]
- (a) Kaminski GA, Ponomarev SY, Liu AB. *J Chem Theory Comput*. 2009; 5:2935–2943. [PubMed: 20209038] (b) Impact version 3.0. Schrödinger, LLC; Portland, OR: 2003. (c) Impact version 3.5. Schrödinger, LLC; Portland, OR: 2004.
- Barone V, Cossi M, Tomasi J. *J Chem Phys*. 1997; 107:3210–3221.
- (a) Klamt A, Jonas V, Bürger T, Lohrenz JCW. *J Phys Chem A*. 1998; 102:5074–5085. (b) Klamt A. *WIREs Comput Mol Sci*. 2011; 1:699–709.
- See, for example: Cramer CJ, Truhlar DG, Maroulis G, Simos TE. *SMx Continuum Models for Condensed Phases. Trends and Perspectives in Modern Computational Science; Lecture Series on Computer and Computational Sciences*. 6Brill/VSPLeiden2006; :112–140.
- See, for example: Bruccoleri RE, Novotny J, Davis ME, Sharp KA. *J Comp Chem*. 1997; 18:268–276. Grant JA, Pickup BT, Nicholls A. *J Comp Chem*. 2001; 22:608–640.
- Im W, Lee MS, Brooks CL. *J Comp Chem*. 2003; 24:1691–1702. [PubMed: 12964188]
- Jorgensen WL, Maxwell DS, Tirado-Rives J. *J Am Chem Soc*. 1996; 118:11225–11236.
- Jorgensen WL, Tirado-Rives J. *J Comp Chem*. 2005; 26:1689–1700. [PubMed: 16200637]
- Kaminski GA. *J Phys Chem B*. 2005; 119:5884–5890. [PubMed: 16851640]
- Liptak MD, Gross KC, Seybold PG, Feldgus S, Shields GC. *J Am Chem Soc*. 2002; 124:6421. [PubMed: 12033873]
- Ponomarev SY, Kaminski GA. *J Chem Theory Comput*. 2011; 7:1415–1427. [PubMed: 21743799]
- Gallicchio E, Zhang LY, Levy RM. *J Comp Chem*. 2002; 23:517–529. [PubMed: 11948578]
- Weast, RC., editor. *CRC Handbook of Chemistry and Physics*. 60. CRC Press; Boca Raton, FL: 1979. Dielectric constant for 20°C was used
- Klicic JJ, Friesner RA, Liu SY, Guida WC. *J Phys Chem A*. 2002; 106:1327.
- Maple JR, Cao YX, Damm W, Halgren TA, Kaminski GA, Zhang LY, Friesner RA. *J Chem Theory Comput*. 2005; 1:694–715.
- (a) Kaminski GA, Duffy EM, Matsui T, Jergensen WL. *J Phys Chem*. 1994; 98:13077–13082. (b) Marten B, Kim K, Cortis C, Friesner RA, Murphy RB, Ringnalda MN, Sitkoff D, Honig B. *J Phys Chem*. 1996; 100:11775–11788.
- Lopes PEM, Lamoureux G, Mackerell AD. *J Comp Chem*. 2009; 30:1821–1838. [PubMed: 19090564]
- Fujio M, McIver RT, Taft RW. *J Am Chem Soc*. 1981; 103:4017–4029.
- Bachrach, Steven M. *Computational Organic Chemistry*. accessed on 10 May, 2012 <http://comporgchem.com/blog/?p=27#cystNIST>. NIST Chemistry WebBook. 2005 <http://webbook.nist.gov/>, as referenced in (26a)
- Schmidt Busch M, Knapp E-W. *Chem Phys Chem*. 2004; 5:1513–1522. [PubMed: 15535550]

Figure 1.
Electrostatic field at the solute-solvent interface.

Figure 2.
Schematic depiction of a solvated atom and the solvent grid.

Table 1

Calculated and Experimental Values of Acidity Constants

System	pKa, OPLS/FB	pKa, experiment ^a
phenol	9.98	9.98
o-chlorophenol	8.57	8.56
m-chlorophenol	8.51	9.02
p-chlorophenol	10.07	9.38
m-cyanophenol	8.48	8.61
p-cyanophenol	7.42	7.95
m-nitrophenol	8.14	8.40
p-nitrophenol	7.87	7.15
o-methylphenol	10.02	10.29
m-methylphenol	9.11	10.08
p-methylphenol	10.50	10.14
Average Error	0.41	

^aReference 17

Table 2

Values of Hydration Energy for Compounds Related to Phenol, in kcal/mol

System	Calculated Hydration Energies OPLS/Fuzzy-Border	Hydration Energy, Quantum Mechanics, ^a
Phenol	-6.61	-7.21, -7.92
Phenoxide ion	-73.02	-73.26, -73.49
ortho-chlorophenol	-8.34	-4.20, -4.61
ortho-chlorophenoxide ion	-64.75	-67.32, -67.58
meta-chlorophenol	-8.49	-7.16, -7.73
meta-chlorophenoxide ion	-63.18	-65.83, -66.34
para-chlorophenol	-7.77	-7.50, -8.08
para-chlorophenoxide ion	-61.54	-67.13, -67.80
meta-cyanophenol	-17.08	-9.71, -10.48
meta-cyanophenoxide ion	-65.51	-63.99, -64.55
para-cyanophenol	-18.05	-10.42, -11.21
para-cyanophenoxide ion	-64.44	-61.35, -62.10
meta-nitrophenol	-9.20	-9.64, -10.54
meta-nitrophenoxide ion	-56.71	-63.20, -63.89
para-nitrophenol	-11.37	-10.65, -11.58
Para-nitrophenoxide ion	-56.44	-57.92, -58.89
ortho-methylphenol	-6.41	-6.60, -7.18
ortho-methylphenoxide ion	-73.70	-70.65, -70.91
meta-methylphenol	-6.46	-7.01, -7.70
meta-methylphenoxide ion	-75.79	-72.90, -73.26
para-methylphenol	-6.89	-7.04, -7.72
para-methylphenoxide ion	-75.71	-73.23, -74.06

Reference 17.

Table 3

Calculated and Experimental Values of Hydration Energy for Other Compounds, in kcal/mol

Compound	Hydration Energy, calculated			Hydration Energy, experimental ^a
	Fuzzy-Border	Poisson- Boltzmann	Generalized Born	
CH ₄	1.936	1.587	1.446	1.9 – 2.0 ^b
C ₂ H ₆	1.948	1.658	1.674	1.82
C ₃ H ₈	2.005	1.760	1.813	1.96
C ₄ H ₁₀	2.118	1.870	1.994	2.08
iso-C ₄ H ₁₀	2.154	1.936	1.966	2.32
C ₂ H ₄	1.056	1.332	1.361	1.27
1-propene	1.512	1.350	1.116	1.27
1-butene	1.429	1.463	1.228	1.38
butadiene	0.643	0.870	0.564	0.60
acetylene	–0.224	1.607	–0.076	–0.010
1-propyne	–0.190	0.963	–0.052	–0.310
1-butyne	–0.016	0.732	–0.042	–0.160
CH ₃ OH	–5.100	–5.410	–5.013	–5.11
C ₂ H ₅ OH	–4.989	–5.345	–4.037	–5.01
CH ₃ COCH ₃	–3.900	–3.653	–3.709	–3.85
2-butanone	–3.675	–3.564	–3.521	–3.640
2-pentanone	–3.567	–3.251	–2.899	–3.530
3-pentanone	–3.537	–3.381	–3.636	–3.410
CH ₃ OCH ₃	–2.092	–1.537	–1.470	–1.90
C ₂ H ₅ OC ₂ H ₅	–1.728	–1.552	–1.765	–1.63
methyl amine	–4.566	–5.104	–2.803	–4.560
ethyl amine	–4.375	–4.309	–3.033	–4.500
<i>n</i> -propyl amine	–4.486	–4.331	–2.309	–4.390
<i>n</i> -butyl amine	–4.173	–5.750	–2.269	–4.290
dimethyl amine	–4.485	–4.533	–3.443	–4.290
diethylamine	–3.879	–3.901	–3.026	–4.070
ammonia	–4.301	–4.839	–0.660	–4.310
C ₆ H ₆	–0.866	–1.466	–0.256	–0.870
toluene	–0.933	–1.917	–0.703	–0.89
C ₆ H ₅ OH	–6.612	–6.720	–5.624	–6.62
CH ₃ CONH ₂	–9.797	–8.910	–8.349	–9.71
NMA	–10.069	–7.670	–10.637	–10.08

Compound	Hydration Energy, calculated			Hydration Energy, experimental ^a
	Fuzzy-Border	Poisson- Boltzmann	Generalized Born	
4-methyl-imidazole	-10.243	-11.356	-11.955	-10.25 ^c
3-methyl-indole	-5.879	-9.104	-6.088	-5.88 ^c
CH ₃ SH	-1.246	-1.423	-0.782	-1.24
C ₂ H ₅ SH	-1.299	-1.045	-0.803	-1.30
CH ₃ SCH ₃	-1.528	-1.709	-1.297	-1.540
C ₂ H ₅ SC ₂ H ₅	-1.451	-1.360	-1.133	-1.43
CH ₃ COOH	-6.702	-7.056	-6.051	-6.70
C ₂ H ₅ COOH	-6.448	-6.828	-6.440	-6.48
Average error	0.076	0.527	0.639	

^a Experimental data are from Reference 19, except where noted.

^b Reference 23.

^c Reference 24.

Table 4

Fuzzy-Border Hydration Parameters

Atom	OPLS-AA Atomtypes ^a	R, Å	Δ, Å	a ₀	A ¹² , kcal/mol·Å ⁶
Aliphatic C and C(=O) in NMA	135, 136, 137, 138, 148, 157, 181, 182, 206, 209, 210, 217, 235, 242, 505, 903, 906	1.900	0.25	0.02748	166.3
sp ² C, alkenes	142, 143, 150	2.015	0.25	0.02748	196.5
Aliphatic H and H on sp ² carbons	140, 144, 156, 185, 911	1.357	0.25	0.01034	1.327
sp C, alkynes	925, 927	2.050	0.25	0.01700	250.0
H, alkynes	926	1.500	0.25	0.0200	1.327
Aromatic C	145, 166, 500, 501, 502, 506, 507, 508, 514	2.050	0.25	0.02748	161.5
Aromatic H	146	1.320	0.25	0.01034	10.00
Polar H	155, 168, 204, 270, 240, 241, 504, 909	1.300	0.25	0.01034	1.327
N in amines	900, 901	1.650	0.25	0.00900	262.0
N in ammonia	127	1.357	0.25	0.00650	122.9
O, (C ₁₀ H _{2n+1})OH, ethers, O(H) in carboxylic acids, NMA	154, 180, 235, 268	1.735	0.25	0.02748	118.7
O, phenol	167	1.700	0.25	0.02748	143.3
O ⁻ in phenoxide	420	2.690	0.25	0.01734	142.9
Cl, chlorophenols	264	1.905	0.25	0.00808	250.0
C (CN)	261	1.300	0.25	0.00600	100.0
N (CN)	262	1.825	0.25	0.02100	200.0
O (NO ₂)	761	1.550	0.25	0.00821	110.0
N (NO ₂)	767	1.370	0.25	0.00900	1.327
O, O(=C) in carboxylic acids and acetone	269, 281	1.750	0.25	0.02748	118.7
C(OOH) in carboxylic acids	267	1.950	0.25	0.02748	150.0
C(O), -COO ⁻ , carboxylate ion	271	1.900	0.25	0.01000	150.0
O ⁻ , -COO ⁻ , carboxylate ion	272	1.700	0.25	0.01800	1.327
C(=O), acetone	280	2.050	0.25	0.02748	140.0
N, primary amines	900	1.700	0.25	0.02000	210.0
N, acetamide	237	1.870	0.25	0.01800	135.0
N, NMA	238	1.770	0.25	0.02000	190.0
NA (-N-H) in heterocycles	503	1.700	0.25	0.02000	167.0

Atom	OPLS-AA Atomtypes ^a	R, Å	Δ, Å	a ₀	A ^{LJ} , kcal/mol·Å ⁶
NB in heterocycles	511	1.850	0.25	0.02000	167.0
S, thiols	200	2.050	0.25	0.07270	183.8
S, sulfides	202	2.150	0.25	0.04600	150.0

^aAtomtypes used in this work. According to in implementation in BOSS, see Reference 15.

Table 5pKa Values for Phenol and *p*-chlorophenol

System	pKa			
	Modified OPLS ^a , explicit solvent	OPLS/FB (this work)	QM ^b	Experiment ^b
Phenol	4.50	9.98	9.88/10.23	9.98
<i>p</i> -chlorophenol	−1.06	10.07	9.84/9.77	9.38
Average Error	6.90	0.35	0.28/0.32	–

^aReference 16.^bReference 17.

Table 6

Gas-Phase Deprotonation Energies for the Phenol Systems and Propanoic and Butanoic Acids, in kcal/mol

Molecule	Deprotonation Energy	Source
Phenol	349.0	Reference 21
<i>o</i> -chlorophenol	337.1	Reference 17
<i>m</i> -chlorophenol	335.3	Reference 17
<i>p</i> -chlorophenol	336.5	Reference 17
<i>m</i> -cyanophenol	329.0	Reference 17
<i>p</i> -cyanophenol	325.5	Reference 17
<i>m</i> -nitrophenol	327.6	Reference 17
<i>p</i> -nitrophenol	324.8	References 17 and 21
<i>o</i> -methylphenol	349.95	Reference 25
<i>m</i> -methylphenol	350.75	Reference 25
<i>p</i> -methylphenol	352.13	References 21 and 25
propanoic acid	347.20	Reference 26
butanoic acid	347.26	Reference 27

Table 7

Calculated Values of Electrostatic Components of Hydration Energy for Other Compounds, in kcal/mol

Compound	Electrostatic Hydration Energy, calculated		
	Fuzzy-Border	Poisson-Boltzmann	Generalized Born
CH ₄	−0.040	−0.043	−0.004
C ₂ H ₆	−0.144	−0.098	−0.003
C ₃ H ₈	−0.228	−0.123	0.017
C ₄ H ₁₀	−0.288	−0.145	0.086
iso-C ₄ H ₁₀	−0.299	−0.120	0.081
C ₂ H ₄	−0.945	−0.642	−0.664
1-propene	−0.511	−0.607	−0.635
1-butene	−0.711	−0.612	−0.553
Butadiene	−1.145	−1.244	−1.199
Acetylene	−1.202	0.000	0.000
1-propyne	−0.853	−0.080	−0.084
1-butyne	−0.821	−0.249	−0.229
CH ₃ OH	−5.575	−6.473	−7.269
C ₂ H ₅ OH	−5.611	−7.193	−5.300
CH ₃ COCH ₃	−4.996	−4.988	−5.224
2-butanone	−4.914	−4.942	−5.273
2-pentanone	−4.837	−4.750	−4.857
3-pentanone	−4.887	−4.776	−5.592
CH ₃ OCH ₃	−2.776	−2.951	−3.106
C ₂ H ₅ OC ₂ H ₅	−2.696	−2.874	−3.186
methyl amine	−2.349	−6.316	−3.309
ethyl amine	−2.439	−5.477	−3.231
<i>n</i> -propyl amine	−2.687	−5.285	−2.737
<i>n</i> -butyl amine	−2.560	−6.297	−2.721
dimethyl amine	−2.868	−5.082	−3.609
Diethylamine	−2.569	−4.536	−3.088
Ammonia	−3.078	−7.224	−2.158
C ₆ H ₆	−2.108	−2.488	−2.508
Toluene	−2.551	−2.970	−2.989
C ₆ H ₅ OH	−5.52	−7.749	−7.521
CH ₃ CONH ₂	−8.949	−10.175	−10.791
NMA	−7.650	−8.135	−9.161

Compound	Electrostatic Hydration Energy, calculated		
	Fuzzy-Border	Poisson-Boltzmann	Generalized Born
4-methyl-imidazole	-7.695	-10.427	-10.989
3-methyl-indole	-4.180	-6.880	-7.529
CH ₃ SH	-4.902	-2.642	-2.303
C ₂ H ₅ SH	-4.667	-2.382	-2.069
CH ₃ SCH ₃	-6.253	-3.005	-2.790
C ₂ H ₅ SC ₂ H ₅	-5.985	-2.889	-2.549
CH ₃ COOH	-5.294	-8.346	-7.470
C ₂ H ₅ COOH	-5.106	-8.155	-7.975
Average signed difference from FB	-	-0.662	-0.065

Table 8

Data and Results from pKa Calculations for Propanoic and Butanoic Acids. Energies are in kcal/mol

System	Hydration Energy, Protonated Form		Hydration Energy, Deprotonated Form		pKa	
	FB ^a	Reference ^b	FB ^a	Reference	FB ^a	Reference ^c
propanoic acid	-6.443	-6.480	-77.964	-79.100 ^b	4.90	4.87
butanoic acid	-6.197	-6.360	-78.112		4.66	4.83

^aThis work.

^bReference 19.

^cReference 1.