Published in final edited form as:

J Comput Chem. 2017 January 15; 38(2): 65–80. doi:10.1002/jcc.24519.

Using polarizable POSSIM force field and Fuzzy-Border continuum solvent model to calculate pK_a shifts of protein residues

Ity Sharma and George A. Kaminski

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

George A. Kaminski: gkaminski@wpi.edu

Abstract

Our Fuzzy-Border (FB) continuum solvent model has been extended and modified to produce hydration parameters for small molecules using Polarizable Simulations Second-order Interaction Model (POSSIM) framework with an average error of 0.136 kcal/mol. It was then used to compute pK_a shifts for carboxylic and basic residues of the turkey ovomucoid third domain (OMTKY3) protein. The average unsigned errors in the acid and base pK_a values were 0.37 and 0.4 pH units, respectively, versus 0.58 and 0.7 pH units as calculated with a previous version of polarizable protein force field and Poisson Boltzmann continuum solvent. This POSSIM/FB result is produced with explicit refitting of the hydration parameters to the pK_a values of the carboxylic and basic residues of the OMTKY3 protein, thus the values of the acidity constants can be viewed as additional fitting target data. In addition to calculating pK_a shifts for the OMTKY3 residues, we have studied aspartic acid residues of Rnase Sa. This was done without any further refitting of the parameters and agreement with the experimental pK_a values is within an average unsigned error of 0.65 pH units. This result included the Asp79 residue that is buried and thus has a high experimental pK_a value of 7.37 units. Thus, the presented model is capable of reproducing pK_a results for residues in an environment that is significantly different from the solvated protein surface used in the fitting. Therefore, the POSSIM force field and the Fuzzy Border continuum solvent parameters have been demonstrated to be sufficiently robust and transferable.

SYNOPSIS TOC

Correspondence to: George A. Kaminski, gkaminski@wpi.edu.

Supporting Information Available

All the potential energy POSSIM and Fuzzy-Border parameters developed in the course of this work and not explicitly introduced in the text.

Our Fuzzy-Border continuum solvent model has been extended and modified to produce hydration parameters for small molecules using Polarizable Simulations Second-order Interaction Model (POSSIM) framework. It was then used to compute pK_a shifts for carboxylic and basic residues of the turkey ovomucoid third domain (OMTKY3) protein. In addition, we have studied aspartic acid residues of Rnase Sa. We conclude that the POSSIM force field and the Fuzzy Border model are sufficiently robust and transferable.

I. Introduction

Structure, stability, solubility, and catalytic functioning of a protein are closely related to pK_a values of its ionizable residues. The locations of acid and basic sites on peptides affect the hydrogen bonding in the biomolecules and therefore impact their biological activity. The protonation states of biomolecules depend on the relative proton affinities of different sites and determining pK_a values is one way of gaining information on the protonation sites of the biomolecules.

Evaluation of pK_a values of protein residues are of particular interest in pharmaceutical applications and designing more robust industrial enzymes that are stable over a wide pH ranges. The pK_a values of ionizable functional groups such as $-\text{COOH}$, $-\text{SH}$, phenol, $-\text{NH}_3^+$, imidazolium and guanidinium are significantly affected by neighboring residues in the protein. The experimental determination of pK_a for the above mentioned ionizable groups is not always straightforward. It is usually obtained by measuring spectroscopic properties of the residue as a function of pH. The acidity constants of amino acids with such ionizable groups as $-\text{SH}$ and phenolic- OH (cysteine and tyrosine) can be measured using UV/vis spectroscopy.¹ For other amino acids, NMR is used to obtain individual pK_a .² However, assignment of NMR chemical shifts in proteins is not a trivial task. Therefore, accurate and rapid prediction of acidity constant shifts of individual protein residues is very important in computational protein research. At the same time, the intrinsically many-body interactions in the protein residue make calculations of these values nontrivial. The protein conformation and the local environment of the residue in the protein affect the pK_a shifts of the residues.

There are numerous computational approaches to calculating pK_a values of proteins and protein residues within different proteins.^{3–27} These can be broadly classified into three major areas. The first one comprises ab initio and DFT quantum mechanics (QM) to treat the residue or the part of protein containing the titrable group and the aqueous solvent as a

dielectric continuum.^{12, 16–18} This method is limited by the size of the protein that can be simulated quantum mechanically, the level of the theory used for quantum calculations and the parameterization of continuum solvation model to be employed with the quantum model used. The above problems can be overcome using combined quantum mechanical/molecular mechanical (QM/MM) methods but this solution leads to the need to parameterize the solvation model, the correct choice of QM and MM regions and careful modeling of QM/MM coupling.

Second, pK_a values and shifts can be evaluated by employing a large database and fitting a linear free energy relationship.²² This is useful only for cases in which the values are close to the fitting data set, and the issue of choosing the optimal way to fit the linear free energy arises.

Finally, empirical force fields using explicit^{11, 13, 25} or continuum solvent models^{2–9, 17–21, 23, 25, 26} can be used to calculate pK_a shifts in proteins. The solvation energy and the energy of interaction of the residue with its environment can be calculated by a number of techniques. In some cases, the “effective dielectric constant” approach is used for simulating the protein. But these methods do not always work well, and the value of the dielectric constant for the interior of the protein varies significantly, sometimes reaching ca. 20,^{5, 7, 8, 21, 23, 28} which raises the question of its physical meaning and undermines the ability of the method to be adequately applicable to a wide variety of cases.

The standard empirical fixed charge force fields like AMBER and CHARMM can be used for calculating pK_a of protein residues. These force fields treat the polarization of the solute molecules in the mean-field manner and lack the explicit polarization term. Simonson *et al*²⁵ carried out molecular dynamics simulations to compute pK_a shifts of two aspartic acid residues (Asp20 and Asp26) in thioredoxin and Asp14 in ribonuclease A using AMBER and CHARMM force fields and explicit water and continuum solvation model. The average error using the explicit solvent model in these residues was 1.91 and 2.64 pH units respectively with calculated 3.2–4.5 unit maximum deviation. These results are not different from the pK_a of aspartic acid residues (Asp7 and Asp27) in OMTKY3 protein presented in Reference 19 using the fixed-charge force field. Explicit treatment of electrostatic polarization is needed to lower the errors. For example, the acidity constants of acidic and basic residues of the turkey ovomucoid third domain (OMTKY3) were determined within 0.6 and 0.7 pH units of the available experimental data, respectively, when PFF was employed. At the same time, the best errors with the fixed-charges OPLS were 3.3 and 2.2 pH units.^{19,48} Moreover, these results were achieved without any explicit fitting of parameters to the pK_a values. Therefore we believe that the advantages of using the polarizable force field result from the correct physics of the force field formalism.

Alternatively the polarization is accounted for in the force fields by incorporating conformation dependent charges.³⁰ In this case, a part of the protein is simulated using quantum mechanics to derive charges by fitting to the electrostatic potential. While this technique is capable of producing reasonably good results, it essentially employs tabulated polarization charge adjustment. At the same time, explicit treatment of polarization achieves the same goal automatically based on more correct physical basis.

Continuum solvation models are used in calculating acidity constants and in a great number of other applications as they are relatively computationally inexpensive and do not require large number of water molecules to be present explicitly. The most popular and accurate implicit solvation model is the Poisson Boltzmann (PB) one based on the Poisson-Boltzmann equation. The electrostatic potential is usually computed by integration over the solute-solvent interface. Both the electrostatic field and the surface density affect each other resulting in self-consistent equations similar to the electrostatic polarization equation.³¹ A further approximation of essentially the same model is known as the Generalized Born (GB) solvation technique.

The solution of the PB and GB model relies on solving the corresponding equations on a numerical grid. This can result in noise in the solvation energy due to grid rebuilding with the movements in the solute. In addition, the self-consistent convergence in solving the PB equations is sometimes difficult to achieve on a numerical grid with arbitrary parameter values. There is a potential to overcome the above problems in using a fixed equally spaced grid independent of the solute coordinates³² and smoothing, antialiasing and proper choice of grid parameters.^{32, 33}

Our recently developed Fuzzy-Border (FB) continuum solvation model³⁴ uses a fixed 3-D grid. It also presents an approximation to the general PB technique that is more accurate than the GB one, and it is in line with the second-order polarization model in POSSIM (although the previously published results of implementing the FB method included combination with the fixed-charges OPLS-AA force field).⁴⁶ This methodology converts the self-consistent equations in the PB model into an analytical one but is closer in reproducing many-body interactions to the Poisson-Boltzmann than to the Generalized Born one.

In this article, we present results of further developing our previously introduced Fuzzy-Border (FB) continuous solvent model, its parameterization for the polarizable POSSIM force field and calculation of pK_a shifts of residues in OMTYK3. We chose the turkey ovomucoid third domain (OMTKY3) protein since this has been extensively studied experimentally and computationally and its pK_a shifts have been measured and calculated.^{37, 38} It is a 56-residue protein with five acidic (Asp7, Glu10, Glu19, Asp27 and Glu43) and six basic (Lys13, Arg21, Lys29, Lys34, His52 and Lys55) ionizable residues. Our parameter fitting for the FB solvation model included explicit fitting to the pK_a values of these residues, while no refitting for the POSSIM parameters for the protein residues themselves was carried out.

Moreover, we have also computed pK_a value of the significantly buried Asp79 residue of ribonuclease Sa and a terminal Asp1 residue. This former residue has a value of the acidity constant of 7.36 pH units, and the latter one has $pK_a = 3.44$ pH units.⁴⁷ Thus, our choice of the testing residues from the protein included both a buried Asp79 which exists in an environment that is very different from the normal hydration environment of the surface residues and an N-terminal residue. Therefore the latter was, once again, a system that was different from the residues of the OMTKY3 used in the Fuzzy-Border parameter fitting. This choice was designed to obtain information about performance of our methodology in

conditions that differed from those used in the fitting. No refitting of any parameters was done in this case (ribonuclease Sa pKa calculations).

The force field we used for representing the solutes is POSSIM (POLarizable Simulations Second Order Interaction Model). It is a combination of inducible point dipole method and the fast second-order approach that has been successfully used to increase the speed of simulations by ca. an order of magnitude without any loss in accuracy.^{36,40,41} The hydration energies, including those used in obtaining the pK_a values, are assessed with the modified Fuzzy-Border model that is now available to supplement the second-order polarizable POSSIM framework. The introduced modifications of the FB technique included parameterization for used with the polarizable POSSIM force field, changing some general parameter values (such as the spacing between the grid points), and adaptation of the multiple-marching level-set technique⁴² modified for generation of the solvent-accessible grid points list suitable for FB calculations. The physical level of the true solvent energy approximation in our Fuzzy-Border methodology lies between Generalized Born and Poisson-Boltzmann solvation.

The rest of the paper is organized as follows. The methods are described in the second section, results and discussion are given in the third one, and these are followed by conclusions.

II. Methods

A. POSSIM Force Field

A brief outline of POLarizable Simulations Second order Interaction Model (POSSIM) force field used to compute solute energies is as follows. The details of this force field have been discussed previously.^{36, 40, 41} E_{total} or the total energy of a molecule is:

$$E_{total} = E_{electrostatic} + E_{vdW} + E_{stretch} + E_{bend} + E_{torsion} \quad (1)$$

The nonelectrostatic interactions consist of the van der Waals energy, harmonic angle stretch and angle bend and the torsional term. The electrostatic energy is computed using point charges and induced point dipoles, μ and includes the charge-charge, charge-dipole and dipole-dipole interactions. It is taken as the sum of polarization energy and the contribution from interactions of permanent charges.

$$E_{electrostatic} = E_{pol} + E_{additive} \quad (2)$$

The electrostatic polarization energy is computed using the inducible point dipoles, μ :

$$E_{pol} = -\frac{1}{2} \sum_i \mu_i E_i^0 \quad (3)$$

where μ is

$$\mu_i = \alpha_i E_i^{tot} \quad (4)$$

E^0 is the electrostatic field in the absence of the induced dipoles and is computed as product of scalar polarizabilities, α , and the total electric field, E_i^{total} including the dipole-dipole term.

$$E_i^{tot} = E_i^0 + \sum_{j \neq i} T_{ij} \mu_j \quad (5)$$

$$T_{ij} = \frac{1}{R_{ij}^3} \left(\frac{3R_{ij}R_{ij}}{R_{ij}^2} - I \right) \quad (6)$$

T_{ij} is the dipole-dipole interaction tensor and I is the unit tensor. R_{ij} stands for the distance between atomic sites i and j .

$$\mu_i = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j \quad (7)$$

These equations are self-consistent and are usually solved iteratively. The first three iterations lead to “first-order”, “second-order” and “third-order” approximations for the induced dipoles:

$$\mu_i^I = \alpha_i E_i^0 \quad (8)$$

$$\mu_i^{II} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j^I = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j E_j^0 \quad (9)$$

$$\mu_i^{III} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j^{II} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j E_j^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j \sum_{k \neq j} T_{jk} \alpha_k E_k^0 \quad (10)$$

POSSIM employs the second-order iteration given in Equation 9 and has been shown to increase the computational speed by about an order of magnitude without any compensation in the accuracy.^{35,36,41,46} This second order expression also makes the equation analytical,

thus eliminating the danger of polarization catastrophe. The overall electrostatic energy also includes the pairwise-additive Coulomb fixed charges interactions between the atoms i and j :

$$E_{additive} = \sum_{i \neq j} \frac{q_i q_j}{R_{ij}} f_{ij} \quad (11)$$

The factor f_{ij} is equal to zero for 1, 2 and 1, 3 pairs, (atoms which belong to the same valence bond or angle), 0.5 for 1,4 interactions (atoms in the same dihedral angle) and 1.0 for all other pairs. Another distinguishing feature in POSSIM is a short-distance cutoff parameter, R_{cut} (currently set at 0.8 Å) used to reduce unphysical growth of electrostatic interactions at the short distances. If the overall distance between two atoms i and j , R_{ij} , is smaller than sum of these parameters $R_{min}^{ij} = R_{cut}^i + R_{cut}^j$, smoothing function is used:

$$R_{ij}^{eff} = \left[1 - \left(\frac{R_{ij}}{R_{min}^{ij}} \right)^2 + \left(\frac{R_{ij}}{R_{min}^{ij}} \right)^3 \right] R_{min}^{ij} \quad (12)$$

and the actual distance between the atoms is replaced with the effective one defined in Equation 12.

For the non-electrostatic part of POSSIM we use the standard Lennard-Jones formalism for the van der Waal's energy:

$$E_{vdW} = \sum_{i \neq j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right] f_{ij} \quad (13)$$

Following geometric combining rules are applied for Lennard-Jones parameters: $\varepsilon_{ij} = (\varepsilon_i \cdot \varepsilon_j)^{1/2}$ and $\sigma_{ij} = (\sigma_i \cdot \sigma_j)^{1/2}$.

Standard harmonic terms are used to calculate bond stretching and angle bending energy and the torsional term is in the Fourier form:

$$E_{torsion} = \sum_i \frac{V_1^i}{2} [1 + \cos(\phi_i)] + \frac{V_2^i}{2} [1 - \cos(2\phi_i)] + \frac{V_3^i}{2} [1 + \cos(3\phi_i)] \quad (14)$$

The bond stretching and angle bending parameters were adopted from the OPLS-AA force field as implemented in BOSS.⁴³

B. POSSIM Force Field Parameterization

Parameterization started with producing atomic polarizabilities using three-body energies. The procedure was the same as used previously.⁴¹

The target three-body energy is calculated using Equation 15:

$$E_{3body} = E(1+2+3) - E(1+2) - E(1+3) - E(2+3) + E(1) + E(2) + E(3) \quad (15)$$

Jaguar software suite⁴⁵ was used to compute the values of the QM energies using the density functional theory (DFT) with the B3LYP method and cc-pVTZ(-f) basis set. The isotropic atomic polarizabilities, α_i , were fitted with POSSIM calculations to reproduce the quantum mechanical values.

The next step in parameterization is determining values of permanent atomic charges and Lennard-Jones parameters fitted to reproduce quantum mechanical dimerization energies and intermolecular distances. The geometries of these dimers were optimized quantum mechanically by LMP2/cc-pVTZ(-f) optimizations, and the energies were found with a cc-pVTZ(-f)/cc-pVQZ extrapolation routine similarly to the previous POSSIM parameterizations.⁴¹

The torsional parameters in Equation 14 were computed for dihedrals in methylguanidine molecule as well as for dihedrals in ethylamine and propylamine. These parameters were fitted to the constrained geometry optimizations at fixed dihedral angles calculated with LMP2/cc-pVTZ quantum mechanical level of theory.

C. Fuzzy-Border Solvation Model

The solvation energy is found as a sum of electrostatic and non-polar parts:

$$\Delta G_{solvation} = \Delta G_{electrostatic} + \Delta G_{nonpolar} \quad (16)$$

The electrostatic part of the total solvation energy is calculated with an approximation to the Poisson-Boltzmann formalism. It includes integration over the whole solute-solvent interface of surface charge density, σ multiplied by the electrostatic potential ϕ^0 created by the charges of the solute only.

$$\Delta G(el) = \frac{1}{2} \int_S \sigma \phi^0 d^2r \quad (17)$$

The surface charge density σ at the interface is given by the self-consistent equation as the electrostatic field E_I depends on the surface charge density distribution.

$$\sigma = \frac{1}{4\pi} \left(1 - \frac{1}{\epsilon} \right) E_I \cdot n \quad (18)$$

Here n is the unit vector normal to the surface.

The numerical solution of these equations is done with representing the surface represented by a discrete set of points i . Thus, the above equations change to:

$$\Delta G(el) = \frac{1}{2} \sum_i q_i \phi_i^0 \quad (19)$$

$$q_i = -\frac{1}{4\pi} \left(1 - \frac{1}{\varepsilon}\right) \mathbf{E}_{1,i} \cdot \mathbf{n}_i \quad (20)$$

The electrostatic field at the point i on the surface grid $\mathbf{E}_{1,i}$ is a sum of the fields produced by the solute charges and the solute-solvent interface points.

$$\mathbf{E}_{1,i} = \sum_j \frac{q_j \mathbf{R}_{ij}}{R_{ij}^3} + \sum_{k \neq 1} \frac{q_k \mathbf{R}_{ik}}{R_{ik}^3} \quad (21)$$

$\mathbf{R}_{ij}$ in this equation represents the vector pointing from j to i . The first term in equation 21 correspond to the summation over all the solute charges whereas the second term represents summation for all the interface ones.

The fast first-order polarization approximation used in Fuzzy-Border model constitutes replacing $\mathbf{E}_{1,i}$ with $\mathbf{E}_{1,i}^0$ which is the field created by the solute alone. Although, this does not include the field due to the polarized solute-solvent interface, it still contains many-body interactions as electric field includes interactions from all the solute charges. This approach converts the self-consistent equations 20 and 21 to analytical ones, thus alleviating the convergence problem. The second-order FB model would include the field created by the interface charges.

The solute-solvent interface in FB model is represented by a three-dimensional equally spaced cubic numerical grid. The interface between the solute and solvent is assumed to consist of points on the solvent-accessible solute surface with distances from R_- to R_+ from the solute atoms (where R would be the distance for the exact location of the solute surface accessible by the solvent). Thus, the solvent-accessible surface is “fuzzy”, with a non-zero thickness.

The solvent-accessible part of the grid is chosen using a Fuzzy-Border modification of the level-set method.⁴² Briefly, the technique proceeds in three steps: first, an outward propagation from the interior of the solute creates the surface accessible to centers of solvent molecules. Second, an inward propagation is used to produce the contact surface (the truly solvent-excluded surface). The R referenced above refers to position of this very surface. Finally, another inward propagation is employed to identify interior cavities that can be distinguished from the outside solvent-accessible space. A point of the grid is associated with a particular solute atom if the distance from this point to the atom falls within the $\pm \delta$

range of the solvation radius R_i of the atom. That is, the point belongs to the part of the solute-solvent interface created by the surface of the solute atom. In cases when the surface point is located on the part of the interface linking two or more atomic surfaces, this point is associated with more than one atom, each one contributing with its own weight w_i calculated as described by Equation 22.

The surface points j corresponding to the solute atom i has weights w_j associated with it as the real solvation surface radius is different from points in the R^- to R^+ range that are not exactly at R :

$$w_j = a_{0,i} \left[\left(\frac{R_{ij} - R_i}{\Delta_i} \right)^4 - 2 \left(\frac{R_{ij} - R_i}{\Delta_i} \right)^2 + 1 \right] \quad (22)$$

Here R_{ij} represents the distance between the solvent grid point and the solute atom. R_i and $a_{0,i}$ are adjustable parameters for the atomtype of the atom i . Their values have been fitted in this work to achieve a good level of agreement between the experimental and calculated hydration energies described in the section on results and discussion below. The weights are normalized and the unit vectors between the grid point j and solute atom i , $\mathbf{n}_{ji}$ are calculated with the overall average unit vector being set to $\mathbf{n}_j = \sum_i \mathbf{n}_{ji} w_{ji}$.

The zeroth-order electrostatic field at the surface grid points can now be calculated via summation over all the solute atoms.

$$\mathbf{E}_{1,j}^0 = \sum_i \frac{q_i \mathbf{R}_{ij}}{R_{ij}^3} \quad (23)$$

The fuzzy border first order charge on the surface grid point j can be written as

$$q_j^I = -A_{scale} \frac{1}{4} \pi (\epsilon - 1) w_j \mathbf{E}_{1,j}^0 \cdot \mathbf{n}_j \quad (24)$$

A_{scale} is an adjustable scaling factor. The first-order electrostatic component of the solvation energy in fuzzy-border formalism can now be determined as

$$\Delta G(el)^I = \frac{1}{2} \sum_i q_j^I \phi_j^0 \quad (25)$$

The non-polar part of the total solvation energy is calculated as sum of the two terms.

$$\Delta G(np) = \sum_j w_j A_{np} - \sum_i \sum_j w_j \frac{A_i^{LJ}}{R_{ij}^6} \quad (26)$$

A_i^{LJ} is an adjustable parameter for the atomtype of the atom i . The first term is the contribution of the solvent accessible surface area and is taken as summation over all the grid points. The second term in the above equation approximates the Lennard-Jones attraction between the solute and the solvent atoms and is taken in double summation over all the solute atoms i and the grid points j . Their values have been fitted in this work to achieve a good level of agreement between the experimental and calculated hydration energies described in the section on results and discussion below. If the second-order FB approximation is to be used, the field in Equation 24 is replaced with

$E_{1,j}^I = E_{1,j}^0 + A_{self} \sum_{k \neq j} \frac{q_k R_{kj}}{R_{kj}^3}$, and the resulting charges are used in Equation 25 to calculate the energy.

D. General scheme for pKa calculations

To obtain the values of the pKa shifts, the following two processes in aqueous solution were considered. For deprotonation reaction for the reference system (for example, propanoic acid, C_2H_5COOH for the Asp residues and butanoic acid, C_3H_7COOH , for the Glu residues of the OMTKY3 and Rnase Sa): $Acid-H \rightarrow Acid^- + H^+$. The change of free energy is

$$\Delta G_1 = G(Acid^-) + G(H^+) - G(Acid-H) \quad (27)$$

For the protein residue deprotonation reaction: $A-H \rightarrow A^- + H^+$, the free energy change is:

$$\Delta G_2 = G(A^-) + G(H^+) - G(A-H) \quad (28)$$

The acidity constants $pK_a(Acid)$ and $pK_a(A)$ for these deprotonation reaction are:

$$pK_a(Acid) = \Delta G_1 / (2.303RT) \quad (29)$$

$$pK_a(A) = \Delta G_2 / (2.303RT) \quad (30)$$

The pKa shift is the difference between the acidity constants for the residue and the acid:

$$\Delta pK_a = pK_a(A) - pK_a(\text{Acid}) = [G(A^-) - G(A-H) - G(\text{Acid}^-) + G(\text{Acid}-H)] / (2.303RT)$$

(31)

or

$$pK_a(A) = pK_a(\text{Acid}) + [G(A^-) - G(A-H) - G(\text{Acid}^-) + G(\text{Acid}-H)] / (2.303RT) \quad (32)$$

Equation 31 above is used for calculation of the pK_a values for the residues using the reference values for the acid/base systems.

Values of the energies $G(A^-)$, $G(A-H)$, $G(\text{Acid}^-)$, and $G(\text{Acid}-H)$ were obtained via geometry optimizations with POSSIM/FB using the POSSIM software.

The experimental and calculated pK_a values listed in this study are given at 298.15 K.

The above technique is similar to that used in our previous calculations of pK_a shifts of acid¹⁹ and base⁴⁸ residues of the OMTKY3 protein.

III. Results and Discussion

A. POSSIM Force Field Parameterization

Most of the POSSIM parameters for the pK_a calculations of acid and basic residues of OMTKY3 protein were adopted from previous works.^{36,40,41,49} New parameters were produced whenever they did not exist before. This included methylguanidine molecule needed for arginine pK_a calculations in OMTKY3 basic residues and torsional parameters for C-C-N-H, H-C-C-N dihedrals in ethylamine and C-C-C-N dihedral angle in propylamine for lysine pK_a calculations.

1. Methylguanidine molecule: Three-Body Energies, Dimerization Energies, and Torsional Fitting—The previously established procedure^{36,40,41} was used to fit the electrostatic polarizabilities, permanent electrostatic charges, Lennard-Jones and the torsional parameters. All these calculations were carried out using our POSSIM software suite.

Three body energies were first fitted to the QM results to produce atomic polarizabilities. There were six hydrogen bonding positions of dipolar probes around methylguanidine and fifteen possible three-body energies (pairwise combinations of the probes). Four of these dipolar probes were acceptors and two had positive point charges at the hydrogen bonding distance from the nitrogen atoms as shown on Figure 2.

The only polarizability that was fitted was that for the nitrogen atom with only two covalent bonds. The average absolute error in the three-body energies of methylguanidine was 0.151 kcal/mol, which is consistent with our previously observed results.

The next step in the parameterization involved reproducing quantum mechanical methylguanidinium-water dimerization energies and distances. Values of the Lennard-Jones parameters were adjusted. The previously developed POSSIM water molecule was employed.³⁶ There were two methylguanidine-water dimers, and their geometries were optimized with the LMP2/cc-pVTZ(-f) level of theory using Jaguar program.⁴⁵ The geometries of these dimers simulated with POSSIM are shown in Figure 3.

The quantum mechanical distances are reproduced well within ca. 0.03 Å and the dimerization energies are fitted within the error range of 0.09 kcal/mol, as shown in Table 1.

Most of the torsional parameters for methylguanidine were adopted from the methylguanidinium ion.⁴¹ Only the coefficients for key rotations of C-N-C-N(H) and H-N(H)-C-N torsions were obtained by fitting the POSSIM energies to LMP2/cc-pVTZ(-f) quantum mechanical values shown in Table 2. In both the cases, the deviations of the resulting torsional energies were less than 0.1kcal/mol.

2. Ethylamine and Propylamine: Torsional Fitting—The torsion coefficients $V_1 - V_3$ for amines were produced for C-C-N-H and H-C-C-N dihedrals in ethylamine and C-C-C-N angle in propylamine following the same procedure as was utilized before.³⁶ The gas phase LMP2/cc-pVTZ (-f) QM energies of the rotamers were obtained with these dihedral angles restrained to their positions. These energies were then used to fit the torsional coefficients for POSSIM. The overall RMSD error in the torsional fitting of both the amines was less than 0.1kcal/mol as can be seen from data in Table 3.

B. Fuzzy-Border Parameterization

The numerical grid spacing was chosen to be 0.40 Å in each dimension. All the simulations using FB model were carried out in water as the solvent. The radius of water molecule was taken as 1.4 Å and the value of dielectric constant, ϵ , of water was set at 80.4.³⁴ The scaling factor A_{scale} was set at 0.07069. The value of r or the “fuzziness” in the solvation radius was the same for all the atom types at 0.25 Å. The nonpolar component of the total solvation energy, $G(np)$ has a factor, A_{np} , that has the same value of 3.200 for all the solute atoms and all the grid-points at the solute-solvent interface.

All the simulations presented in this article are performed with revised first-order Fuzzy-Border model. Thus, the charges on the solute-solvent interface are created by the electrostatic fields of the solute only and not by the polarized interface interaction with itself. However, the induced solute dipoles do contribute to the field that polarizes the boundary, and the polarized boundary does contribute to the field that is then taken into the account in adjusting the inducible solute dipoles. These calculations are done with modified POSSIM software suite including revised FB continuum solvation model. POSSIM force field^{36,40,41,46,49} was used for simulating the solutes in the pK_a calculations of both the acid and the base residues of OMTKY3 protein. Values of the hydration parameters of the small

molecules were produced by fitting to the experimental or quantum mechanical solvation energies of these systems. These solvation energies are calculated as:

$$\Delta G_{(solv,calculated)} = G_{(solvated)} - G_{(gas)} \quad (33)$$

$G_{(solvated)}$ and $G_{(gas)}$ are the computed energies of the solute molecules in the aqueous and the gas phase respectively. Geometry optimizations were carried out, and the lowest energy conformer was chosen whenever applicable.

The hydration energies of the small molecules parameterized with POSSIM force are listed in Table 4. The overall average error in the hydration energy of these molecules is 0.136 kcal/mol. Parameters were assumed to be transferable and were used in calculations of hydration energies of larger molecules as needed. For instance, the aliphatic hydration parameters were employed in simulations of methyl or ethyl groups of the carboxylic acids. All the hydration energies for small molecules are given in Table 4.

C. pK_a values of Carboxylic Residues and Hydration Energies of related molecules

Coordinates of atoms in the five carboxylic residues (Asp7, Glu10, Glu19, Asp27 and Glu43) of the serine protease inhibitor turkey ovomucoid third domain protein (OMTKY3) were obtained from Protein Data Bank (PDB ID 1PPF). It has been demonstrated both experimentally and computationally that the local environment of a residue plays the most important role in defining the acidity constant shift of these residues.^{18,31} Therefore, only a part of the protein was explicitly simulated to compute value of the pK_a shifts (the same atoms as in References to 19 and 48 were picked. The above references contain figures that explicitly show the simulated part of the protein). Generally, the hydrogen atoms were added with the IMPACT software suite. The carboxylic acid group protons that were added manually. The experimental pK_a values for the acid reference systems propanoic acid (Asp) and butanoic acid (Glu) were adopted at 4.87 and 4.83 units respectively.⁵²

The initial conformations of the five systems (the parts of the protein containing Asp7, Glu10, Glu19, Asp27 and Glu43 residues) were obtained from Reference 19. The deprotonated and the protonated acid forms for the aspartate and glutamate reference system are shown in Figure 4. Protonation of each of the acid oxygen atoms was carried out and the one with the lower total energy was accepted for the final pK_a calculations. The terminal carboxylic groups and the hydrogens in the molecular systems were allowed to move during the geometry optimizations.

Let us briefly present the process of assembling sets of hydration parameters for the systems studied in this subsection from those from small molecules. The hydration parameters of aliphatic carbon and hydrogen atoms were produced by fitting to the experimental solvation energies of methane, ethane, propane, butane and isobutane. It can be seen from the results in Table 4 that the calculated solvation energies of alkanes follow the experimental trend with the overall average error of only 0.06 kcal/mol.

Methanol and ethanol were simulated to produce the hydration parameters of the oxygen and the polar hydrogen atoms. The average error in the solvation energies of these molecules is 0.14 kcal/mol.

Fuzzy-border hydration parameters for C (COOH) and O, O (=C) in carboxylic acids were fitted to the solvation energies of acetic and propanoic acid. The overall average error in the solvation energies of the acids was 0.15 kcal/mol.

Fuzzy Border model for the carbon and oxygen of the carboxylate ion, (-COO^-) was parameterized by fitting to the experimental pK_a values of the two abovementioned acids. Therefore, we did carry out explicit refitting of the parameters in order to reproduce the pK_a values of these protein residues. This applies only to the continuum solvent hydration parameters, as the solute (protein) force field parameters were not adjusted.

The energies of the protonated and deprotonated forms of the reference systems (propanoic acid and butanoic acid) and the pK_a values are shown in Table 5. The values of the gas phase deprotonation energies $G(\text{gas})$ for propanoic and butanoic acids were adopted at 347.20 kcal/mol⁵¹ and 347.26 kcal/mol,⁵² respectively.

The overall average error in the absolute pK_a of the acids was 0.56 pH units.

Hydration parameters for NMA and acetamide were also required in order to calculate the pK_a values of protein residues. The aliphatic carbon and hydrogen atoms had the same solvation parameters as in the alkane molecules discussed above. The hydration parameters of alcohol oxygen and polar hydrogen were transferred without changes as well. The only atoms for which the solvation parameters were fitted in this case was the amide nitrogen atoms. The average error in the solvation energies of the two amides is 0.02 kcal/mol as shown in Table 4.

The computed protonated and deprotonated energies of the carboxylic residues of turkey ovomucoid third domain are shown in Table 6.

The effect of two hydrogen bonds with Ser9 on results for the Asp7 residue can be seen in Table 6, with the overall deprotonation energy change (and thus the pK_a) being the lowest in the series. The pK_a of Asp7 is 2.33 pH unit compared to experimental value of 2.67 pH unit. The Glu10 residue has no hydrogen bonds with neighboring residues. The pK_a calculated for this residue is 3.89 pH unit with an unsigned error of only 0.22 pH unit from the experimental value. Glu19 residue shows a hydrogen bond from the hydrogen bonded to nitrogen of the backbone and also from the Thr17 residue in the crystal structure. Our model of Glu19 includes the Thr17 residue as well as the backbone part of Leu18.¹⁹ The computed pK_a value for this residue is 3.36 pH unit compared to the experimental one of 3.21. The model for the Asp27 residue and its environment used in this study has three intramolecular hydrogen bonds as suggested by the PBD crystal structure. The predicted pK_a of Asp27 is 1.66 compared to the experimental value of 2.28. The strong electrostatic interactions between the carboxyl group and the amide hydrogens are taken into account in the polarizable POSSIM and can be seen reproducing pK_a of this residue with error of 0.62 pH

units. The residue Glu43 has no hydrogen bonds for its carboxylic group. The pK_a value of this residue is 4.29 compared to the experimental result of 4.81 pH units.

The above results can also be compared with the pK_a values for the carboxylic residues obtained with the OPLS-SGB, OPLS-PBF and OPLS-PBF.¹⁹ These results are summarized in Table 7.

The average error in the pK_a values of carboxylic acid residues is 0.37 pH units obtained with POSSIM/FB technique. This error is 0.21 pH units lower than with the previous version of the polarizable force field for proteins. This demonstrates the robustness of the Fuzzy-Border continuum solvation model in such calculations. The results also confirm the importance of the explicit treatment of electrostatic polarization in reproducing or predicting protein pK_a shifts. At the same time, it should be emphasized once again that the above results were produced with explicit refitting of the values of the hydration parameters.

D. pK_a values of Basic Residues and Hydration Energies of related molecules

The pK_a of six basic residues of OMTKY3 (Lys13, Arg21, Lys29, His52, and Lys55) were computed using POSSIM force field and FB continuum solvation model. The reference model systems for these residues are n-pentylamine (experimental $pK_a = 10.6^{47}$) for lysine, 4-ethylimidazole (experimental $pK_a = 7.55^{47}$) for histidine and n-butylguanidine (experimental pK_a is 12.48, from methylguanidine⁴⁸) for arginine.

The initial conformations for the six basic residues are obtained from the experimental OMTKY3 geometry (PDB ID 1omt⁵⁵). These residues were capped with an acetyl group on the N-terminus and an N-methylamine group on the C-terminus except for Lys29 which will be discussed in the following section. The geometry optimizations of both the protonated and deprotonated forms were employed to determine their relative acidity constants.

Lys13, Lys29, Lys34 and Lys55 Residues—FB hydration parameters for amines were produced by fitting parameters for the nitrogen atom to the amines solvation energies. The primary amines chosen were methylamine, ethylamine, n-propylamine and n-butylamine. The solvation parameters for the polar amine hydrogen atoms were same as before. The overall average error in the hydration energies was 0.42 kcal/mol and the trend of decreasing hydration energy with increasing carbon chain length was successfully reproduced as shown in Table 4. The resulting hydration parameters for nitrogen atom are $R = 1.703 \text{ \AA}$, $a_0 = 0.2010$ and $A^{LJ} = 237.7 \text{ kcal/mol \AA}^6$.

The absolute acidity constant of lysine reference system (n-pentylamine) was computed by simulating its protonated and deprotonated forms, Figure 6, in gas and in aqueous solution. The final values of the energies and acidity constants for the reference systems are given in Table 5.

We followed the procedure established previously^{19,47} and used only a part of the protein in the immediate vicinity of the ionizable residues to calculate the pK_a shifts. The terminal functional group of the basic residue and the side chains were unconstrained for all the calculations. For Lys13, Lys34 and Lys55, there is no hydrogen bond with the other residues,

Figure 7. The errors in the calculated pK_a for Lys13, Lys34 and Lys55 are 0.49, 0.22 and 1.07 pH units, respectively.

Because of the close proximity between the Lys29 and Asp27 residues in the crystal structure of OMTKY3, the model for Lys29 includes the sequence of Ac-Asp-Asn-Lys-NMe as shown on Figure 8.

Although the side chain in Lys29 has potential hydrogen bonding in the NMR structure, the free movement allowed in the side chain in neutral Lys29 structure permitted more accurate calculation of its pK_a value. The distance between the O and N atoms in NMR structure and the optimized neutral Lys29 structures are 4.252 Å and 5.120 Å respectively. The overall error in the pK_a of Lys29 is 1.48 pH units. The energies of the protonated, deprotonated lysine residues and the pK_a values are summarized in Table 8.

Arg21 Residue—The reference system for Arg21 is n-butylguanidine ($pK_a = 12.48^{47}$). The POSSIM potential energy parameters produced for methylguanidine were used to compute the pK_a shift of Arg21.

First the FB hydration parameters n-propylguanidine were produced. The standard hydration parameters of alkanes and polar hydrogens produced in this work were retained. Only the hydration parameters for the central C atom and the N (-NH, -NH₂) bonded to the central C atom were fitted to reproduce the solvation energy of n-propylguanidine. The average unsigned error in the solvation energy of n-propylguanidine was 0.04 kcal/mol given in Table 4.

These solvation parameters were used to compute the pK_a of n-butylguanidine (Figure 9).

The computed pK_a of n-butylguanidine was 13.12 pH units, compared to the experimental result of 12.48 pH units.⁴⁸

The same parameters were also used to compute pK_a of Arg21 (Figure 10).

The average quantum value of the n-methylguanidine⁵⁷ gas-phase deprotonation energy was used in calculating pK_a of Arg21.

The error in the pK_a of Arg 21 was 0.41 pH units.

His52 Residue—4-ethylimidazole was used as the reference model for calculating pK_a shift in His52. The solvation parameters of the heterocyclic molecule are required to calculate its pK_a . First, the solvation parameters of benzene and phenol were produced. The C and H hydration radius in benzene are 2.050 Å and 1.320 Å respectively. The a_0 for both the atoms is 0.04000 and the Lennard-Jones factor is 128.7 kcal/mol Å⁶ and 12.00 kcal/mol Å⁶ for C and H, respectively. An unsigned error of 0.002 kcal/mol was obtained for the solvation energy of benzene as presented in Table 4. In case of phenol, the same parameters for aromatic carbon and hydrogen as well as polar hydrogen were employed. Only the parameters for the oxygen atom was refitted. The solvation parameters of oxygen in phenol

are $R = 1.750 \text{ \AA}$, $a_0 = 0.04000$ and Lennard-Jones factor as $75.40 \text{ kcal/mol \AA}^6$. The error in the solvation energy was less than 0.001 kcal/mol .

The hydration energy of toluene was calculated using the hydration parameters of benzene and aliphatic groups without any further fitting. The error in the toluene solvation energy was 0.05 kcal/mol .

The transferability of hydration parameters of aromatic carbon and hydrogen for 4-methylimidazole was found to be good. Only nitrogen atoms were refitted to the hydration energy. The calculated solvation energy of 4-methylimidazole is -10.20 kcal/mol compared to the experimental value of -10.25 kcal/mol . The solvation parameters of the nitrogen atoms are radii of $1.700 \text{ \AA}$ for NA and $1.750 \text{ \AA}$ for NB, $a_0 = 0.04000$ and A^{LJ} of $228.0 \text{ kcal/mol \AA}^6$ and $200.1 \text{ kcal/mol \AA}^6$ for NA and NB, respectively.

The same set of parameters were employed to calculate the energy of deprotonated 4-ethylimidazole molecule. Since there are two hydrogens covalently bonded to two different nitrogen atoms (N_δ and N_ϵ) in the imidazole ring, two deprotonated structures shown in Figure 11 were simulated for both the His52 residue and the reference molecule 4-ethylimidazole.

Only the hydration parameters for the nitrogen atoms in the 4-ethylimidazolium were fitted to reduce the pK_a error, with the final value of the latter being 0.79 pH units (Table 5). The gas-phase free energy of deprotonation was set at 216.6 kcal/mol .⁵⁸

The same parameters were used to compute the acidity constant of His52 (Table 8).

The deprotonated His52 structure with hydrogen covalently bonded to N_ϵ has lower energy value and was used to determine acidity constant of this residue. The calculated pK_a of His52 residue is 7.51 pH units compared to experimental value of 7.50 pH units .

The pK_a results of the basic residues of OMTKY3 produced with POSSIM force field and FB continuum solvation model can be compared with the results of previous pK_a calculations⁴⁸ using SGB/OPLS, PBF/OPLS and PBF/PFF in Table 9.

The overall average pK_a error in the basic residues using POSSIM/FB is 0.4 pH units which is by 0.30 pH units more accurate than the previous pK_a calculations using PFF force field and PBF solvation model⁴⁸. Although these results are produced by fitting to the experimental hydration energy and pK_a targets of the reference systems, it clearly demonstrates the Fuzzy-Border solvation model is robust in predicting the protein acidity constants.

While the overall accuracy of the pK_a values calculated for the OMTKY3 protein is good, it can be noted that these values tend to be slightly lower than their experimental counterparts. Therefore, the deprotonated forms of the residues are a little more favored than they should be. While the magnitude of this bias is small and does not cause any concerns, we can probably conclude that the reason is that the Fuzzy-Border polar hydrogen hydration radius value is at the higher side of the acceptable range. Once again, the POSSIM/FB results consistently agree with the experimental literature values, no significant outliers have been

observed in this work, and thus we believe that the overall quality of fitting of the parameters is adequate.

E. Comparison of the Fuzzy-Border Solvation Radii Values with the OPLS-AA Lennard-Jones Radii

In order to confirm that the fitted values of our FB solvation radii are reasonable and do not contain obvious artifacts of optimization, we have included Table 10 that lists both these FB radii and the Lennard-Jones radii of the standard OPLS-AA²⁸ atomic types. The comparison clearly demonstrates that the values of the radii in the FB and OPLS-AA sets are generally close to each other, and the chemically justifiable trends in their values are reproduced by the Fuzzy-Border model. The only exception is with the polar hydrogen atoms, that have zero Lennard-Jones parameters in the OPLS-AA formalism.

F. Calculating the pKa value of the Ribonuclease Sa Residues Asp1 and Asp79

These calculations were carried out without any additional refitting of the solute or solvation parameters. We chose the above two residues for the following reason. In the case of Asp79, the ionizable residue in question was significantly buried, therefore we used it to validate the performance of the presented technique in a situation when the application environment is different from that in which the parameter fitting was carried out. It was not sufficient to include only the residues with hydrogen bonds in the simulation model. We have included residues with atoms within ca. 5 Å of Asp79. The initial protein structure was that from the Protein Data Bank entry 1RGG. The structure of the final model (the deprotonated form of the system) used in this work for the Asp79 of Rnase Sa is shown on Figure 12a.

Overall, the following residues were included into the model: 18, 19, 22, 23, 69–74, 76–81, 90–95. Additionally, the Glu78 side chain that points away from the ionizable Asp79 site and thus is rather far from it was replaced with a methyl. Positions of all the atoms in the system were kept fixed except for the Asp79 side chain and the –CONH₂ group of the Gln94 residue. Otherwise, the protocol was the same as for the Asp residues of the OMTKY3 protein.

The Asp1 residue was chosen as a system that had a solvent-exposed carboxylic residue and, at the same time, had the terminal –NH₃⁺ group right next to it, thus providing additional testing for the model, as this was the first time we calculated a pK_a value for a terminal residue.

The resulting optimized energies of the protonated and the deprotonated Asp79 systems were 3554.12 and 3491.33 kcal/mol, respectively. This gives a difference of –62.79 kcal/mol. Combined with the reference data in Table 5, this gives the final pK_a value of 6.56 pH units, which is about 0.8pH units different from the experimental results of 7.37.⁴⁷ This is a good agreement given that no specific refitting was done and the electrostatic conditions are very different from those in which the fitting was carried out further illustrate transferability and physical robustness of the polarizable POSSIM/FB method.

The optimized energies of the protonated and the deprotonated Asp1 models were –187.25 and –121.34 kcal/mol. The difference is –65.91 kcal/mol, and the calculated Asp1 pK_a value

is 3.92pH units, while the experimental result is 3.44 units.⁴⁷ The error is 0.48 units, which is consistent with the level observed for our OMTKY3 calculations. Once again, the results for the ribonuclease Sa were obtained without any additional refitting of the parameters.

G. Change of Electrostatic Potential as a Result of Residue Ionization

Protonation or deprotonation of an ionizable residue causes a significant change in the local electrostatic environment for the protein. In order to illustrate this, we have created electrostatic potential surfaces for two residues of the OMTKY3, Asp27 and Arg21. POSSIM does not calculate the full electrostatic potential map as it uses point charges and dipoles. Therefore, we have employed the following procedure. We used the final POSSIM/FB structures as the input geometries. Then we ran single point DFT/B3LYP 6-31G** calculations in aqueous solution to produce the maps with Spartan software.⁶⁰ The resulting electrostatic potential maps are shown in Figures 13 and 14.

It can be seen that ionization of the residues causes significant changes in the potential (both in magnitude and in some cases in sign) in the neighboring parts of the protein. Therefore, we can conclude that it is natural that explicit treatment of electrostatic polarizability of the whole systems is important in reproducing protein pKa shifts. This is the same conclusion that can be made by considering the dramatic improvement in the accuracy shown by the data in Tables 7 and 9.

H. CPU Time Required to Run Fuzzy-Border Calculations

Each specific simulation run required a different amount of CPU time depending on the size of the system, the closeness of the initial guess to the final optimized geometry and other factors. However, just to illustrate a typical run, let us consider a Fuzzy-Border geometry optimization for the Rnase Sa Asp79 residue, with the simulated system shown in Figure 12a. This simulation took 35 geometry optimization iterations (in addition to the program initialization, input, and output) and took 498 seconds using a Linux machine with a 1.56Ghz Intel CPU.

IV. Conclusions

We have modified and extended our previously developed Fuzzy-Border continuum solvation model to reproduce hydration energies of a number of small molecules and pK_a values of carboxylic and the basic residues of proteins. The distinguishing features of FB model for calculating solvation energies include a three-dimensional fixed position grid independent of solute coordinates and an approximation to the Poisson-Boltzmann (PB) formalism for assessing solvent polarization. This approximation is consistent with the second-order polarization formalism of POSSIM force field.

The OMTKY3 protein was employed in validating calculations of protein pKa shifts with the combined POSSIM-FB model. The Fuzzy-Border hydration parameters were obtained for small molecules and then transferred into the protein residues with further refitting when necessary, and new POSSIM parameters for solutes were produced when required. All the simulations were carried out with a new version of POSSIM software including implementation of the FB model.

The pK_a of five carboxylic (Asp7, Glu10, Glu19, Asp27 and Glu43) and six basic (Lys13, Arg21, Lys29, Lys34, His52 and Lys55) residues of OMTKY3 were reproduced with an overall average error of 0.37 and 0.4 pH units, respectively. This was an improvement from the previously reported values obtained with the PFF polarizable force field for proteins^{19, 48} and this result validates both the Fuzzy-Border solvation model and the complete POSSIM protein force field⁴⁶ employed for simulation of the solutes. At the same time, it should be noted that the pK_a calculations for the OMTKY3 protein were utilized as a part of the hydration parameters fitting process. Therefore, the Fuzzy-Border parameters were adjusted to achieve the above result.

We then simulated the ribonuclease Sa protein and calculated pK_a shifts for residues Asp1 and Asp79. The average unsigned error was 0.645 pH units. In this case, no refitting of parameters took place, and so these results are entirely a validation of the robustness of the parameters produced during the preceding stages of the reported work.

The overall average unsigned error in the hydration energy of small molecules was ca. 0.14 kcal/mol.

The presented results demonstrate robustness of Fuzzy-Border continuum solvation model in calculating pK_a values of both the acidic and the basic residues of the protein. It also offers another piece of evidence that explicit treatment of electrostatic polarization is crucial to permit accurate calculations of acidity constants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was partially supported by Grant Number R01GM074624 from the National Institutes of Health. 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. The authors express gratitude to Schrödinger, LLC for the Jaguar software.

References

1. Jensen JH, Li H, Robertson AD, Molina PA. J Phys Chem A. 2005; 109:6634–6643. [PubMed: 16834015]
2. Wuthrich, K. NMR of proteins and nucleic acids. Wiley; 1986. p. 1-12.
3. Antosiewicz J, McCammon JA, Gilson MK. J Mol Biol. 1994; 238:415–436. [PubMed: 8176733]
4. Antosiewicz J, McCammon JA, Gilson MK. Biochemistry. 1996; 35:7819–7833. [PubMed: 8672483]
5. Barth P, Alber T, Harbury PB. Proc Natl Acad Sci USA. 2007; 104:4898–4903. [PubMed: 17360348]
6. Dixit SB, Bhasin R, Rajasekaran E, Jayaram B. J Chem Soc, Faraday Trans. 1997; 93:1105–1113.
7. Forsyth WR, Antosiewicz JM, Robertson AD. Proteins. 2002; 48:388–403. [PubMed: 12112705]
8. Forsyth WR, Gilson MK, Antosiewicz J, Jaren OR, Robertson AD. Biochemistry. 1998; 37:8643–8652. [PubMed: 9628726]
9. Forsyth WR, Robertson AD. Biochemistry. 2000; 39:8067–8072. [PubMed: 10891089]
10. Havranek JJ, Harbury PB. Proc Natl Acad Sci USA. 1999; 96:11145–11150. [PubMed: 10500144]

11. Jorgensen WL, Briggs JM. *J Am Chem Soc.* 1989; 111:4190–4197.
12. Kallies B, Mitzner R. *J Phys Chem B.* 1997; 101:2959–2967.
13. Kaminski GA. *J Phys Chem B.* 2005; 109:5884–5890. [PubMed: 16851640]
14. Kaminski GA. *J Chem Theory Comput.* 2005; 1:248–254. [PubMed: 26641295]
15. Khandogin J, York DM. *Proteins: Struct, Funct, Bioinf.* 2004; 56:724–737.
16. Klicic JJ, Friesner RA, Liu SY, Guida WC. *J Phys Chem A.* 2002; 106:1327–1335.
17. Li H, Hains AW, Everts JE, Robertson AD, Jensen JH. *J Phys Chem B.* 2002; 106:3486–3494.
18. Li H, Robertson AD, Jensen JH. *Proteins: Struct, Funct, Bioinf.* 2004; 55:689–704.
19. MacDermaid CM, Kaminski GA. *J Phys Chem B.* 2007; 111:9036–9044. [PubMed: 17602581]
20. Mehler EL, Guarnieri F. *Biophys J.* 1999; 75:3–22.
21. Nielsen JE, Vriend G. *Proteins.* 2001; 43:403–412. [PubMed: 11340657]
22. Ohno K, Sakurai M. *J Comp Chem.* 2006; 27:906–916. [PubMed: 16550537]
23. Sandberg L, Edholm O. *Proteins.* 1999; 36:474–483. [PubMed: 10450090]
24. Schaller W, Robertson AD. *Biochemistry.* 1995; 34:4714–4723. [PubMed: 7718577]
25. Simonson T, Carlsson J, Case DA. *J Am Chem Soc.* 2004; 126:4167–4180. [PubMed: 15053606]
26. Song JK, Laskowski M, Qasim MA, Markley JL. *Biochemistry.* 2003; 42:2847–2856. [PubMed: 12627950]
27. Zhang MJ, Vogel HJ. *J Biol Chem.* 1993; 268:22420–22428. [PubMed: 8226750]
28. Tanford C, Kirkwood JG. *J Am Chem Soc.* 1957; 79:5333–5339.
29. Harris TK, Turner GJ. *IUBMB Life.* 2002; 53:85–98. [PubMed: 12049200]
30. Ji CG, Mei Y, Zhang JZH. *Biophys J.* 2008; 95:1080–1088. [PubMed: 18645195]
31. Ghosh A, Rapp CS, Friesner RA. *J Phys Chem B.* 1998; 102:10983–10990.
32. Bruccoleri RE, Novotny J, Davis ME, Sharp KA. *J Comp Chem.* 1997; 18:268–276.
33. Grant JA, Pickup BT, Nicholls A. *J Comp Chem.* 2001; 22:608–640.
34. Sharma I, Kaminski GA. *J Comput Chem.* 2012; 33:2388–2399. [PubMed: 22815192]
35. Kaminski GA, Zhou R, Friesner RA. *J Comput Chem.* 2003; 24:267–276. [PubMed: 12548718]
36. Kaminski GA, Ponomarev SY, Liu AB. *J Chem Theory Comput.* 2009; 5:2935–2943. [PubMed: 20209038]
37. Hoogstraten CG, Choe S, Westler WM, Markley JL. *Protein Sci.* 1995; 4:2289–2299. [PubMed: 8563625]
38. Bode W, Wei AZ, Huber R, Meyer E, Travis J, Neumann S. *EMBO J.* 1986; 5:2453–2458. [PubMed: 3640709]
39. Veluraja K, Margulis CJ. *J Biomol Struct Dyn.* 2005; 23:101–111. [PubMed: 15918681]
40. Ponomarev SY, Kaminski GA. *J Chem Theory Comput.* 2011; 7:1415–1427. [PubMed: 21743799]
41. Li X, Ponomarev SY, Sa Q, Sigalovsky DL, Kaminski GA. *J Comput Chem.* 2013; 34:1241–1250. [PubMed: 23420678]
42. Can T, Chen CI, Wang YF. *J Molecular Graphics and Modelling.* 2006; 25:442–454. [PubMed: 16621636]
43. Jorgensen WL, Maxwell DS, Tirado-Rives J. *J Am Chem Soc.* 1996; 118:11225–11236.
44. Berendsen HJC, Grigera JR, Straatsma TP. *J Phys Chem.* 1987; 91:6269–6271.
45. (a) Jaguar v3.5. Schrodinger, Inc; Portland, OR: 1998. (b) Jaguar v4.2. Schrodinger, Inc; Portland, OR: 2000. (c) Jaguar, v7.6. Schrodinger, LLC; New York, NY: 2009.
46. Li X, Ponomarev SY, Sigalovsky DL, Cvitkovic JP, Kaminski GA. *J Chem Theory Comput.* 2014; 10:4896–4910. [PubMed: 25400518]
47. Laurents DV, Huyghues-Despointes BMP, Bruix M, Thurlkill RL, Schell D, Newsom S, Grimsley GR, Shaw KL, Trevino S, Rico M, Briggs JM, Antosiewicz JM, Scholtz JM, Pace CN. *J Mol Biol.* 2003; 325:1077–1092. [PubMed: 12527309]
48. Click TH, Kaminski GA. *J Phys Chem B.* 2009; 113:7844–7850. [PubMed: 19432439]
49. Ponomarev SY, Sa Q, Kaminski GA. *J Chem Theory Comput.* 2012; 8:4691–4706. [PubMed: 23483820]

50. Gallicchio E, Zhang LY, Levy RM. J Comp Chem. 2002; 23:517–529. [PubMed: 11948578]
51. Sitkoff D, Sharp KA, Honig B. J Phys Chem. 1994; 98:1978–1988.
52. Atkins, P.; de Paula, J. Physical Chemistry. 7. W. H. Freeman; New York: 2001. p. 1088–1089.
53. NIST. [accessed on December 15, 2015] NIST Chemistry WebBook. 2005. Available at <http://webbook.nist.gov>
54. Schmidt am Busch M, Knapp E-W. Chem Phys Chem. 2004; 5:1513–1522. [PubMed: 15535550]
55. Hoogstraten CG, Choe S, Westler WM, Markley JL. Protein Sci. 1995; 4:2289–2299. [PubMed: 8563625]
56. NIST. [accessed on January 20, 2016] NIST Chemistry WebBook. 2005. Available at <http://webbook.nist.gov/cgi/cbook.cgi?ID=C110587&Units=SI&Mask=20#Ion-Energetics>
57. Moser A, Range K, York DM. J Phys Chem B. 2010; 114:13911–13921. [PubMed: 20942500]
58. Lias SG, Liebman JF, Levin RD. J Phys Chem Ref Data. 1984; 13:695–808.
59. (a) Jorgensen WL, Maxwell DS, Tirado-Rives J. J Am Chem Soc. 1996; 118:11225–11236. (b) Jorgensen WL, Tirado-Rives J. J Comp Chem. 2005; 26:1689–1700. [PubMed: 16200637]
60. **Spartan**¹⁴Wavefunction, Inc. Irvine, CA. Except for molecular mechanics and semi-empirical models, the calculation methods used in **Spartan** have been documented in: Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio RA Jr, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Lin CY, Van Voorhis T, Chien SH, Sodt A, Steele RP, Rassolov VA, Maslen PE, Korambath PP, Adamson RD, Austin B, Baker J, Byrd EFC, Dachsel H, Doerksen RJ, Dreuw A, Dunietz BD, Dutoi AD, Furlani TR, Gwaltney SR, Heyden A, Hirata S, Hsu C-P, Kedziora G, Khalliulin RZ, Klunzinger P, Lee AM, Lee MS, Liang WZ, Lotan I, Nair N, Peters B, Proynov EI, Pieniazek PA, Rhee YM, Ritchie J, Rosta E, Sherrill CD, Simmonett AC, Subotnik JE, Woodcock HL III, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer HF, Kong J, Krylov AI, Gill PMW, Head-Gordon M. Phys Chem Chem Phys. 2006; 8:3172. [PubMed: 16902710]

Figure 1.
Calculating three-body energies for a molecule using dipolar probes

Figure 2.

Dipolar probes used for calculating three-body energies for methylguanidine. Symbols P and N in the probe represent positive and negative point charges of the probes, respectively.

Figure 3.
Gas-phase dimers of methylguanidine with one water molecule.

Figure 4.

(a) Propanoic acid was used as the reference pK_a model for aspartate residues. (b) Butanoic acid was used as the reference pK_a model for glutamate residues.

Figure 5. Models of the carboxylic acid residues of the OMTKY3 protein and their environment used in this work. Structures with the H-O-C-C dihedral angles in the carboxylates set at 0° and at 180° were optimized, and those with the lower resulting energies are presented here and used in the pKa calculation.

Figure 6. Optimized geometries of the protonated (a) and deprotonated (b) forms of pentylamine used as reference system for lysine residues

Figure 7.
Models of the Lys13, Lys34 and Lys55 residues of OMTKY3 used to calculate pKa values.

Figure 8.
The Lys29 model (deprotonated) after geometry optimization without constraints on side chain.

Figure 9. Protonated (a) and deprotonated (b) forms of n-butylguanidine. These systems were used as reference system for the lysine residues

Figure 10.
Model for the Arg21 residue of OMTKY3.

Figure 11.

4-ethylimidazole, histidine reference system (deprotonated (N_δ) form (a), protonated form (b) and deprotonated (N_ϵ) form (c)) and histidine residue capped with acetyl group on N-terminus and N-methylamine on the C-terminus (deprotonated (d), (f) and protonated (e))

(a)

(b)

Figure 12.

The ribonuclease Sa fragments used to calculate the acidity constant value of the Asp79 (a) and Asp1 (b) residues.

Figure 13.

Electrostatic potential maps for protonated (a) and deprotonated (b) forms of the models used to simulate the Asp27 residue of the OMTKY3 protein. The numerical values in the legend are given in equivalents of kJ/mol/Red color corresponds to negative values of the potential and blue ones signify positive ones.

Figure 14.

Electrostatic potential maps for deprotonated (a) and protonated (b) forms of the models used to simulate the Arg21 residue of the OMTKY3 protein. The numerical values in the legend are given in equivalents of kJ/mol/Red color corresponds to negative values of the potential and blue ones signify positive ones.

Table 1

Dimerization energies (kcal/mol) and distances (Å) of methylguanidine-water dimers.

Dimer	Energy		O...N Distance	
	QM	POSSIM	QM	POSSIM
Methylguanidine-water(a) ^a	-8.32	-8.21	2.91	2.90
Methylguanidine-water(b) ^a	-6.43	-6.35	3.24	3.26
			3.34	3.04

^a Lables (a) and (b) are as in Figure 3.

Table 2

Torsional energies of methylguanidine rotamers (kcal/mol).

Molecule	Dihedral	Angle Value	Energy, QM	Energy, POSSIM
Methylguanidine	C-N-C-N(H)	0°	0.00	0.00
		15°	3.01	3.39
		30°	12.88	12.77
	H-N(H)-C-N	0°	0.00	0.00
		15°	6.41	6.25
		30°	11.55	8.99

Table 3

Relative energies in kcal/mol, for ethylamine and propyl amine rotamers.

molecule	dihedral	Angle values	Energy, QM	Energy, POSSIM
Ethylamine	C-C-N-H	60°	0.03	0.00
		120°	2.34	2.14
		180°	0.00	0.00
		240°	1.94	2.14
	H-C-C-N	0°	3.32	3.32
		60°	0.00	0.00
Propylamine	C-C-C-N	0°	5.08	4.72
		60°	0.01	0.02
		120°	3.61	3.96
		180°	0.00	0.00

Table 4

Calculated and experimental values of hydration free energies for small molecules in kcal/mol

Compound	Hydration Energy, calculated	Hydration Energy, experimental ^a	Error
CH ₄	1.904	1.9 – 2.0 ^b	
C ₂ H ₆	1.736	1.82	–0.084
C ₃ H ₈	1.958	1.96	–0.002
C ₄ H ₁₀	2.205	2.08	0.125
iso-C ₄ H ₁₀	2.405	2.32	–0.085
CH ₃ OH	–5.22	–5.11	–0.110
C ₂ H ₅ OH	–4.84	–5.01	0.170
CH ₃ COOH	–6.854	–6.70	–0.150
C ₂ H ₅ COOH	–6.325	–6.48	0.150
CH ₃ CONH ₂	–9.74	–9.71	–0.030
NMA	–10.09	–10.08	–0.010
C ₆ H ₆	–0.872	–0.870	–0.002
Toluene	–0.936	–0.89	–0.046
C ₆ H ₅ OH	–6.62	–6.62	0.00
methylamine	–4.571	–4.560	–0.01
ethylamine	–4.430	–4.500	0.07
n-propylamine	–3.765	–4.390	0.63
n-butylamine	–3.327	–4.290	0.96
4-methylimidazole	–10.199	–10.250 ^b	0.05
n-propylguanidine	–10.958	–10.920 ^c	–0.04
Average error			0.136

^aReference 50.^bReference 34.^cReference 51

Experimental and calculated pK_a values of reference systems for acid and basic residues of OMTKY3 protein with the POSSIM force field and Fuzzy-Border (FB) continuum solvation model

Table 5

System	Energy, kcal/mol		pK_a	
	Protonated	Deprotonated	FB ^a	Exp ^b
Propanoic acid	gas	-15.82	4.20	4.87
	aq	-94.48		
Butanoic acid	gas	-25.85	5.27	4.83
	aq	-102.41		
n-pentylamine	gas	22.35	11.41	10.63 ^c
	aq	19.87		
n-butylguanidine	gas	14.06	13.12	12.48 ^c
	aq	11.91		
4-ethylimidazole	gas	-6.41 (N _e)	4.31 (N _e)	
	aq	-14.92 (N _e)	-1.72 (N _e)	7.55 ^c

^a Calculations with POSSIM/FB (this work).

^b Reference 19.

^c Reference 48

Table 6

Experimental and calculated pK_a values of carboxylic acid residues of OMTKY3 obtained with POSSIM and Fuzzy-Border (FB) continuum solvation model

Carboxylic Residues OMTKY3	Energy, kcal/mol		pK_a	
	Protonated	Deprotonated	POSSIM/FB ^a	Experimental ^b
Asp7	-120.37	-188.45	2.33	2.67
Glu10	-88.84	-162.57	3.89	4.11
Glu19	-90.93	-165.38	3.36	3.21
Asp27	-119.28	-188.26	1.66	2.28
Glu43	-110.47	-183.64	4.29	4.81

^a Calculations with POSSIM/FB, this work.

^b Reference 19

Comparison of pK_a values for carboxylic acid residues of OMTKY3 calculated in this work with results of previous studies¹⁹**Table 7**

System	OPLS-SGB ^a		OPLS-PBF ^a		PFF-PBF ^a		POSSIM/FB ^b	
	pKa	Error	pKa	Error	pKa	Error	pKa	Error
Asp7	8.76	6.09	6.26	3.59	1.87	0.80	2.33	0.34
Glu10	1.59	2.52	4.95	0.84	4.41	0.30	3.89	0.22
Glu19	3.06	0.15	4.90	1.69	3.61	0.40	3.36	0.15
Asp27	14.16	11.88	11.60	9.32	1.66	0.62	1.66	0.62
Glu43	9.24	4.43	5.79	0.98	4.02	0.79	4.29	0.57
Average error		5.01		3.28		0.58		0.37

^aReference 19.^bCalculations with POSSIM/FB, this work.

Table 8

Experimental and calculated pK_a values of basic residues of OMTKY3 computed with POSSIM and Fuzzy-Border (FB) continuum solvation model

System	Energy, kcal/mol		pK_a , Calculated ^a	pK_a , Experiment ^b	Error
	Protonated	Deprotonated			
Lys13	-93.30	-38.26	9.41	9.9	0.49
Arg21	-155.75	-88.75	12.39	12.80	0.41
Lys29	-321.07	-263.65	11.16	11.1	0.06
Lys34	-71.13	-15.45	9.88	10.1	0.22
His52	-103.40	-64.26 (N _e) -22.69 (N _δ)	7.51	7.50	0.01
Lys55	-86.03	-30.14	10.03	11.1	1.07

^aCalculations with POSSIM-FB, this work.

^bReference 8

Table 9
Comparison of pK_a values for basic residues of OMTKY3 computed with different techniques.

System	OPLS-SGB ^b			OPLS-PBF ^b			PFF-PBF ^b			POSSIM/FB ^c			Exp ^d	NMR ^a
	pK _a	Error	Expt	pK _a	Error	Expt	pK _a	Error	Expt	pK _a	Error	Expt		
Lys13	16.7	6.9	-	9.0	0.8	10.4	0.5	9.41	0.5	9.9	0.5	11.2	-	12.8
Arg21	15.0	-	12.4	-	-	12.3	-	12.39	0.4 ^d	-	-	-	-	-
Lys29	23.4	12.3	17.2	6.0	12.9	1.7	11.16	0.1	11.1	11.1	0.1	11.2	-	11.7
Lys34	16.3	6.2	9.9	0.3	9.9	0.3	9.88	0.2	10.1	10.1	0.2	11.7	-	6.2
His52	11.8	4.3	11.2	3.7	7.6	0.1	7.51	0.0 ^e	7.5	7.5	0.0 ^e	11.3	-	11.3
Lys55	15.7	4.6	9.5	1.6	10.1	1.0	10.03	1.1	11.1	11.1	1.1	11.3	-	11.3
Average	6.1	2.2	0.7	2.2	0.7	0.7	0.4	0.4	0.4	0.4	0.4	0.4	-	0.4

^aReference 8.

^bReference 48.

^cCalculations with POSSIM-FB, this work.

^dArg21 error reported from NMR value.

^eLess than 0.1 pH unit.

Table 10Comparison of Fuzzy-Border radii with the standard OPLS-AA^a Lennard-Jones radii

Atom	OPLS-AA atom types	Fuzzy-Border R, Å	OPLS R ^b , Å
Aliphatic C	135, 136, 137, 138, 148, 157, 223, 224, 242, 244, 235, 274, 292, 906	1.952	1.964
Aliphatic H	140, 156, 911	1.415	1.403
Aromatic C	145, 166	2.050	1.992
Aromatic H	146	1.320	1.358
Polar H	155, 168, 240, 241, 270, 909	1.200	–
N in amine	900, 901	1.703	1.824
O, C _n H _{2n+1} OH, carboxylic acids, NMA	154, 236, 268	1.725	1.751
O, phenol	167	1.750	1.723
N, acetamide	237	1.960	1.824
N, NMA	238	1.700	1.824
C(OOH) in carboxylic acids	267	1.942	2.105
O, O(=C) in carboxylic acids	269	1.74	1.661
C(O), -COO ⁻ , carboxylate ion	271	1.900	2.105
O ⁻ , -COO ⁻ , carboxylate ion	272	1.530	1.661
N(RNH ₃ ⁺)	287	1.700	1.824
H(RNH ₃ ⁺)	290	1.350	–
N, primary amine	900	1.703	1.852
NA, heterocycle	503	1.700	1.824
NB, heterocycle	511	1.750	1.824
NA, protonated heterocycle	512	1.700	1.824
H, protonated heterocycle	513	1.296	–
N(H ₂), methylguanidinium	300	1.703	1.824

^aReference 59.^bCalculated as one half of the Lennard-Jones parameter σ multiplied by $2^{1/6}$.