THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Study of p*K* Values and Effective Dielectric Constants of Ionizable Residues in Pentapeptides and in Staphylococcal Nuclease (SNase) Using a Mean-Field Approach

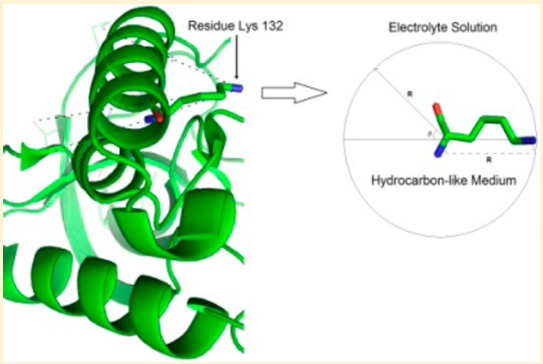
Guilherme Volpe Bossa,[†] Alfred Fahr,[‡] and Tereza Pereira de Souza*[,†,‡]

[†]Instituto de Biociencias, Letras e Ciencias Exatas, Sao Paulo State University, Sao Jose do Rio Preto, 15054-000, Brazil

[‡]Institut für Pharmazie, Friedrich Schiller Universität Jena, Lessingstrasse 8, D-07743 Jena, Germany

Supporting Information

ABSTRACT: The determination of p*K* values of amino acid residues as a function of temperature and ionic concentration is crucial to understanding the dynamics of various biological processes such as adsorption of peptides and their interactions with active sites of enzymes. In this study we developed a mean-field model to calculate the position-dependent dielectric constants of ionizable groups and the mean electrostatic potential on the surface. Such potential, which takes into account the contributions exerted by neighboring groups and ions in solution, is responsible for the fine-tuning of the p*K* value of each residue. The proposed model was applied to the amino acids Asp, Glu, Lys, His, Tyr, and Cys, and since the results were consistent with experimentally obtained values, the model was extended and applied to computation of p*K* values of Gly and Ala pentapeptides and of ionizable residues of the enzyme staphylococcal nuclease (SNase). In this latter case, we used an approach similar to a first-neighbors approximation, and the results turned out to be in good agreement with previously reported data when considering only the interactions of charged groups located at distances of maximally 20 Å. These considerations and the little computational cost involved turn the suggested approach into a promising tool for the modeling of force fields in computational simulations.



INTRODUCTION

The electric charge of amino acids is one of the main factors ruling the dynamics of several physiological processes such as electron transfer in peptides[1] and in their intestinal absorption.[2] However, the complexity of the molecules involved makes it difficult to determine the actual charge that a residue adopts in an electrolyte solution. One of the physicochemical parameters directly related to an amino acid residue's charge is the p*K* values of the residues, which are dependent on the conformation of the molecule, the temperature, and the ionic strength of the solution.[3,4] In order to determine these p*K* values, various approaches based on classical electrostatics have been proposed.[5] As in biological systems, electrostatic interactions are predominant.[5,6] The way in which the dielectric properties of both biomolecules and solvent are dealt with is of crucial importance for the accurate determination of p*K* values of amino acids residues at interior and exterior regions of proteins.[7−10]

Because of the experimental difficulty of establishing the dielectric constant (DC) in inner regions of biomolecules, in some studies phenomenology was combined with the Born model to estimate it.[11] Thus, in a series of studies on glutamate[12] and lysine[13] residues in staphylococcal nuclease (SNase), Isom and co-workers obtained DC values between 8 and 26 for Glu residues, while for Lys residues, values between 9 and 38 were found. Employing a continuum dielectric model,[14] Mellor[15] estimated that in inner regions of *β*-lactoglobulin, DC ranges from 6 to 7. Meanwhile, in other studies calculating this parameter at the protein−solvent interface, reported DC values range from 48 to 120.[11,16] It is important to emphasize here that all of the reported values refer to the effective dielectric constant (DC_{ef}), i.e., a parameter that includes contributions that are not explicitly taken into account by the model used in the calculation of p*K* values, for example.[17,18] The variation of DC_{ef} values in different regions suggests that it is a position-dependent function,[19−21] a consideration dating back to the studies of Debye, Sack, and Lorentz,[22−24] and sigmoidal versions of these functions have been widely used for studying biological systems by computer simulations.[25−27] However, in both Monte Carlo (MC) and molecular dynamics (MD) simulations, the implementation of a position-dependent DC_{ef} may result in an increase of computational cost, stressing the necessity of a model that makes calculation more efficient.[28] Values greater than 20 were used by Antosiewicz[18] to determine the p*K* values of amino acid residues by solving the Poisson−Boltzmann equation (PB)

Received: November 18, 2013
Revised: March 22, 2014
Published: March 24, 2014

   dx.doi.org/10.1021/jp411331p | *J. Phys. Chem. B* 2014, 118, 4053−4061

through the finite differences method which is also used by Gunner[21] for SNase. By employing the generalized Born model and an electrostatic screening factor provided by the Debye–Hückel theory, Brooks[29] uses molecular dynamics to simulate the titration of 10 proteins, including RNase and HEWL. When the results in this study are compared with experimental data, the pK values calculated show an average error between 0.6 or 0.8 units. For studies based on the PB equation, Bashford[30] asserts that this error is close to unity. Nielsen[31] points out that structure-based models used to determine pK values allow establishment of a correlation between protein structure and function. The relation between protein structure and pK values, in particular the exposition of the charged group to the solvent, was also investigated by Forsyth[32] in a study in which they report a series of pK values for Asp and Glu residues in 24 proteins.

It is noteworthy that the acidic carboxylic groups in amino acids (group 1) invariably have a lower pK_a value than the corresponding free alkanoic acid, i.e., ethanoic acid (~2.4 vs ~4.8). Similarly, the basic amino groups in amino acids (group 2) have a lower pK_a value than the corresponding free amine, i.e., methylamine (~9.9 vs ~10.6).

The hypothesis adopted in this study proposes that the difference between the pK value of an ionizable group as part of an amino acid and that of the same group when "free" in solution is caused by an additional electrostatic interaction exerted by the neighboring groups in the amino acid and ions in the solution. Taking into account these contributions, we developed a model that employs pK values of structurally similar molecules for each ionizable group and calculates the mean electrostatic potential in the vicinity of these groups. This model also determines the effective dielectric constant dependent on position, temperature, and ion concentrations. Because this approach yields pK values of amino acids very close to experimentally determined values, the model was extended and applied to Ala and Gly pentapetides and ionizable residues of staphylococcal nuclease. The above-mentioned parameters can be calculated at relatively low computation cost, turning this approach into a promising application for computational simulations and modeling.

MODEL AND METHODOLOGY

As a model of the intracellular environment, we initially assumed that each amino acid studied is in aqueous solution and each of them was treated as an "ensemble amino acid"; i.e., they represent the average values of quantities related to molecules like them. Because most amino acids consist predominantly of carbon, the analyzed system is represented by a medium of low dielectric constant immersed in an aqueous electrolyte solution with a high dielectric constant.[33−35] On the basis of this consideration and the hypothesis based on electrostatic interactions, the effective dielectric constants were determined.

The following notation for index and subscripts was applied: group 1 is the ethanoic acid attached to C_α and the pK ascribed to this acid is 4.76; group 2 is the basic group, methylamine, attached to C_α, with an ascribed pK value of 10.64; group 3 is the ionizable group of side chain that, for each amino acid, is treated as the structurally corresponding compound, like pentylamine for the lysine side chain. The pK values that were used to model each side chain are presented in Table 1.

(I) Effective Dielectric Constant, DC_{ef}. To describe a system constituted by three or more ionizable groups, like an amino acid with an ionizable side chain, we first considered that each acid–base pair is fully shrouded by a spherical surface whose interior (medium 1) is characterized by the dielectric constant ε_1. In view of the molecular composition of this medium, it was, for the case of simplicity, treated as a hydrocarbon and, for calculation purposes, ε_1 was taken as equal to 2.0.[38,39]

Table 1. pK Values Assigned, by Structural Similarity Criteria, to Amino Acid Side Chain Residues[a]

residue	pK used[36,37]	residue	pK used[36,37]
Asp	4.87 (propanoic acid)	Arg	12.65 (butylguanidine)
Glu	4.83 (butyric acid)	His	6.95 (methylimidazole)
Cys	10.50 (ethylmercaptan)	Lys	10.63 (pentylamine)
Tyr	10.20 (*p*-ethanephenol)		

[a]In parentheses are the the corresponding "free" molecules.

The outer region of the sphere, denoted by medium 2, is an electrolyte solution. To determine the dielectric constant of this medium as a function of temperature[40] (T in °C) and ion concentration (n_{i0} in moles per liter) we used an interpolation, eq 1. This equation is the sum of the expressions presented by Hamelin et al.[40] and Robinson et. al.[41]

$$\varepsilon_2 = 87.7404 - (0.40014)T + (0.940425 \times 10^{-3})T^2 - (0.140165 \times 10^{-5})T^3 - (0.929212 \times 10^{-10})T^4 + 2\delta n_{i0} \quad (1)$$

For NaCl, δ is 5.5 L/mol.

As the sphere contains a pair of acid–basic groups, inside it there are two charges denoted q_1 and q_2 whose positions are shown in Figure 1. The electrostatic potential in medium 1

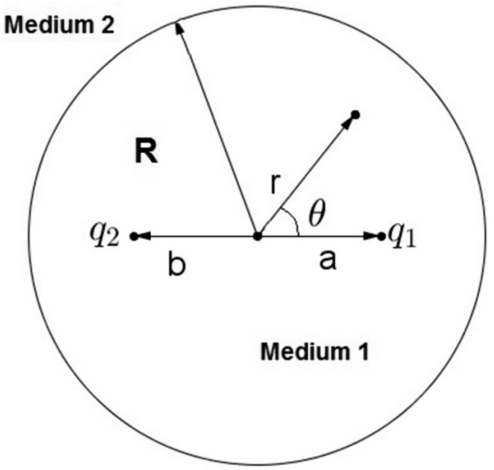


Figure 1. Model for the effective dielectric constant of a region where there are two charges, q_1 and q_2 at a distance of a and b, respectively, from the center of a sphere of radius R.

(Ψ_1) satisfies the Poisson equation, while in medium 2, the potential, denoted by Ψ_{2DH}, obeys the linear form of Poisson–Boltzmann equation,[33,42] which in a more concise form can be written as

$$\nabla^2\Psi_{2DH}(r) = \frac{\Psi_{2DH}(r)}{\lambda^2} \quad (2)$$

The parameter λ is the Debye length, a quantity related to the electrostatic screening exerted by ions. Using azimuthal

dx.doi.org/10.1021/jp411331p | J. Phys. Chem. B 2014, 118, 4053−4061

symmetry and the boundary conditions that both electrostatic potential and normal component of electric displacement must be continuous at the interface between medium 1 and medium 2, we obtain the complete solution of Ψ_1 and Ψ_{2DH}. The mathematical details of this procedure are presented in the Supporting Information. With these potentials fully solved, the effective position-dependent dielectric constant, DC_{ef}, calculated at a position $\vec{r}$, is given by

$$DC_{ef}(\vec{r}) = \begin{cases} \dfrac{1}{\Psi_{1DH}}\left[\dfrac{q_1}{4\pi\varepsilon_0|\vec{r}-\vec{a}|} + \dfrac{q_2}{4\pi\varepsilon_0|\vec{r}-\vec{b}|}\right] \\ \quad \text{inside the sphere} \\ \dfrac{1}{\Psi_{2DH}}\left[\dfrac{q_1}{4\pi\varepsilon_0|\vec{r}-\vec{a}|} + \dfrac{q_2}{4\pi\varepsilon_0|\vec{r}-\vec{b}|}\right] \\ \quad \text{outside the sphere} \end{cases} \tag{3}$$

The radius R of the sphere that delimits the medium and entirely covers an acid−base pair group is

$$R = \frac{r_i + r_j + R_{i\to j}}{2} \tag{4}$$

where r_i and r_j are the radii of the respective ionizable groups covered by the sphere (i and j can denote group 1, 2, or 3) and $R_{i\to j}$ is the distance between these groups. Again, all spatial parameters were determined from Protein Data Bank (PDB) files. In order to calculate the DC_{ef} for each acid−base pair group in an amino acid in its zwitterionic form, one of the charges was taken as zero and, at the very site of this charge, the potential exerted by the remaining charge is calculated using the upper part of eq 3. The DC_{ef} thus obtained is used to calculate the mean electrostatic potential at the surface of each group.

(II) Mean Electric Potential and Structural Similarity. According to the adopted hypothesis, the ionization constant of each ionizable group is modified by an additional electrostatic interaction resulting from the presence of neighboring groups and ions in solution. The modified ionization constant and ionization degree, K_i^* and α_i^*, respectively, are written as

$$K_i^* = K_i \exp\left(\frac{-z_i e \Psi_i}{k_B T}\right) \tag{5}$$

$$\alpha_i^* = \frac{1}{1 + \frac{[H^+]}{K_i^*}} \tag{6}$$

where k_B is the Boltzmann constant. K_i denotes the intrinsic ionization constant of a group i, which is a constant when this i group is not part of an amino acid molecule. The intrinsic pK (i.e., $-\log_{10} K_i$) assigned to each group corresponds to that of a structurally similar molecule. In all amino acids analyzed, group 1 (ethanoic acid) has a pK equaling 4.76 while the pK of group 2 (methylamine) equals 10.65. The pK values attributed to each individual side chain, based on structural similarity criteria, are shown in Table 1. The symbol pK_i^* is assigned to the i ionizable group constituent of each amino acid and can be inferred from eq 5. All ionizable groups were treated as spheres that, when charged, have a central point charge. The radii of such groups as well as the distance between the centers of the spheres were determined from PDB files.

In eq 5, Ψ_i corresponds to the mean electrostatic potential at the surface of group i. This mean potential is calculated taking into account the electrostatic interaction exerted by the group itself, ions in solution, and neighboring groups located at a distance of no more than 20 Å. This cutoff distance is adopted as equivalent to a "first neighbors" approximation, and it will be explained and analyzed in the Results and Discussion. In order to describe Ψ_i, the linear form of Poisson−Boltzmann equation was used,[43] which in radial coordinates can be written as

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2 \frac{d}{dr}\Psi(r)\right) = \frac{\Psi(r)}{\lambda^2} \tag{7}$$

This approach leads to a large reduction of computational time in comparison with the finite difference method and furthermore avoids problems related to divergence. By use of a change of coordinates ($\xi = r/\lambda$) and the dimensionless electrostatic potential (the reduced potential $\varphi_i = e\Psi/(k_B T)$, the solution of Ψ_i is given in terms of Bessel functions and spherical harmonics.[44] Taking into account only the monopole terms, which in fact dominate the summation, the superposition principle provides the following expression for the mean reduced electrostatic potential at the surface of a group i:

$$\varphi_i(r) = \frac{a_i \exp(-r_i/\lambda)}{(r_i/\lambda)} + \frac{\sinh(r_i/\lambda)}{(r_i/\lambda)} \sum_{j=2}^{3} \frac{a_j \exp(-R_{i-j}/\lambda)}{(R_{i-j}/\lambda)} \tag{8}$$

where r_i is the i group radius and R_{i-j} is the distance between this group and the neighboring group j. The term inside the summation corresponds to "crossed interactions" from each neighboring group j on the group i. In eq 8 the sum is given until 3 because we are dealing with a system with three ionizable groups. For a molecule with N ionizable groups, we would have a sum up to N, totaling $N - 1$ crossed interactions terms. The fact that we only considered monopole terms and ignored angular contributions does not limit the applicability of the model because when the superposition principle is applied, the point where the potential is calculated always establishes a plane with the charge of the group and with the neighboring group under consideration.

The integration constants a_i and a_j are determined by boundary conditions based on phenomenology, i.e., those that relate the electric field at the surface of each group with its ionization degree and, consequently, with its pK^* value. However, both α_i^* and pK_i^* are dependent on Ψ_i which results in a set of transcendental coupled algebraic equations that are solved through a self-consistent method, the Kolmogorov contracting-operations method.[45] This method allows a fast determination of the integration constants as a function of pH, ionic concentration, and temperature (the last two factors are implicitly accounted by the use of Debye's length). Once these constants are established, one proceeds with the calculus of the ionization degree for each ionizable group. Finally, by use of the Henderson−Hasselbalch equation, the pK_i^* value that each group assumes in the conditions in which the integrations constants are calculated can be determined. It is noteworthy that this whole procedure, from the determination of DC_{ef} until the pK_i^* value involves only low computational cost.

RESULTS AND DISCUSSION

As a first step, the model was applied to amino acids with an ionizable group in the side chain: aspartic acid (Asp), glutamic

dx.doi.org/10.1021/jp411331p | J. Phys. Chem. B 2014, 118, 4053−4061

acid (Glu), cysteine (Cys), arginine (Arg), tyrosine (Tyr), lysine (Lys), and histidine (His). From the very structure of the proposed model, it is clear that the DC_{ef} depends on the position where the electrostatic potential is calculated. For a system of two opposite charges 2 Å equidistant from the center of a sphere with a radius R = 5 Å, the DC_{ef} values are shown in Figure 2. The DC_{ef} values display a large variation close to the

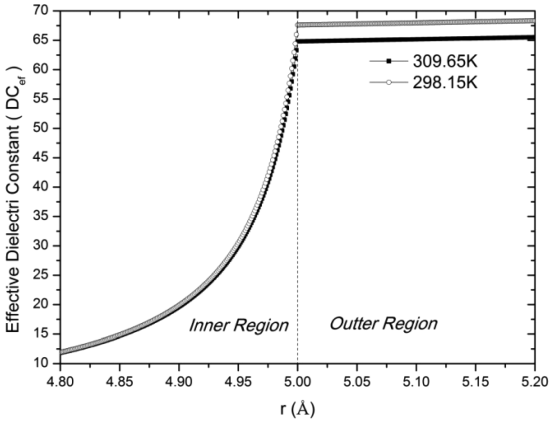


Figure 2. Variation of effective dielectric constant with position (r) in which it is calculated for a system of two opposite charges inside a sphere of radius 5 Å in 150 mM NaCl.

interface between media 1 and 2. One of the main reasons for this is the proximity of a medium (electrolyte solution) with a high dielectric constant, about 40 times higher than that of the inner region.[44] An increase in temperature leads to slightly lower values of DC_{ef}, since this causes a reduction of the dielectric constant of the electrolyte solution, ε_2, given by eq 1. Such a decrease is directly related to solvent destructuralization caused by thermal agitation. Because in medium 2 there are ions in solution, we analyzed the effects on DC_{ef} upon variation of the ionic concentration. The results presented in Figure 3 indicate that in inner regions not close to the interface, there is a small contribution by the ions of medium 2, in agreement with results presented by Mukerjee.[46] This can be explained by the change in the density of polarization charges caused by the ions in solution, an effect that, in the inner region, will become stronger when approaching the interface.

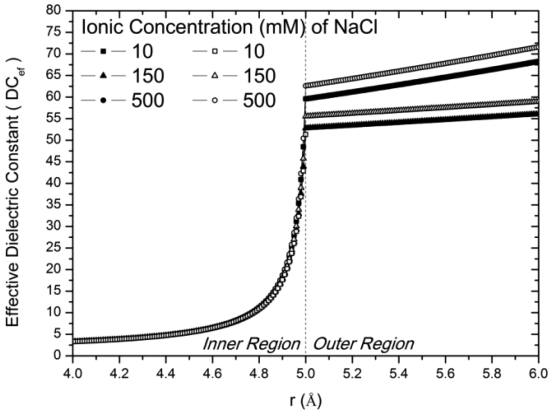


Figure 3. Effective dielectric constant versus position for three distinct values of ionic concentration for a system of two opposite charges inside a sphere of radius 5 Å (in the graphic, delineated by the dashed line): solid points, 309.95 K; open points, 298.15 K.

The effective dielectric constant of the region covering each pair of acid−basic groups in amino acids is computed by using the radius of each ionizable group and the distance between them. The results obtained are shown in Table 2. The high

Table 2. Amino Acids and Respective Effective Dielectric Constants (DC_{ef}) at 298.15 K and 150 mM NaCl[a]

amino acid	region	DC_{ef}
Arg	acid−basic	15.66
	acid side chain	79.95
	basic side chain	79.95
His	acid−basic	15.66
	acid side chain	26.11
	basic side chain	79.95
Lys	acid−basic	15.66
	acid side chain	79.95
	basic side chain	79.95
Cys	acid−basic	25.25
	basic side chain	14.77
	acid side chain	79.95
Tyr	acid−basic	15.66
	basic side chain	79.95
	acid side chain	79.95
Asp	acid−basic	16.70
	basic side chain	25.25
	acid side chain	79.95
Glu	acid−basic	25.25
	basic side chain	29.01
	acid side chain	79.95

[a]Acid is the ethanoic acid attached to C_α, group 1. Basic is the methylamine attached to C_α, group 2. Side chain denotes the side chain ionizable group.

values of DC_{ef} in regions "acid side chain" and "basic side chain" of Arg, Lys, and Tyr are due to the geometry of these molecules, which are characterized by large and linear side chains. As a result of an abundance of water molecules in the region between these groups, the interaction between their electric charges is attenuated only by a dielectric constant practically equal to ε_2 (i.e., 79.95 for 150 mM of NaCl and 298.15 K). For groups with the same charge (as acid−acid side chain and basic−basic side chain regions) we measured distances greater than in the regions of oppositely charged groups. Upon application of these conformational and compositional arguments, according to our model, in regions covering groups of the same kind the calculated DC_{ef} in all amino acids equals ε_2. In all amino acids analyzed the distance between groups 1 and 2 (acid basic) is in the range of 3−4 Å. This variation, even not so large, implies DC_{ef} values between 15 and 25, approximately.

Acid−basic regions of His, Arg, Lys, and Tyr have identical DC_{ef} values because the calculated distance between these groups is the same in each of these molecules (3.75 Å). This argument also explains the equality seen at the same region of Cys and Glu. By the proposed model, Lys has a DC_{ef} equal to 16 in the region between ethanoic acid and methylamine. For the same amino acid, Isom et al.[13] calculated values between 8 and 26. For the surrounding regions of glutamic acid, we obtained values from 25 to 29 while Isom et al. in another paper[12] report values in the range 9−38, depending on the residue position in SNase. Studying Asp and Glu residues of this same protein, Castañeda[47] employed a DC_{ef} equal to 20.

dx.doi.org/10.1021/jp411331p | J. Phys. Chem. B 2014, 118, 4053−4061

Considering all reported values, it is clear that the DC_{ef} values calculated and presented here are consistent with those reported by other approaches. Furthermore, they are validated by our first-principle model describing a system of two media with different dielectric constants and taking into account, even if only implicitly, contributions of ionic concentration and temperature. As we obtained a position-dependent DC_{ef} values, any other system can possibly be analyzed by a redimensionalization of spatial parameters as the radius of the sphere that fully covers a desired group pair. Because of this feature and the low computational cost involved, our method to determine the DC_{ef} for amino acids as well as residues can be used in force fields or in approaches that solve the Poisson−Boltzmann equation via a finite difference method. We performed the calculus of integration constants and the mean electrostatic potential at the surface of each *i* group. For this purpose, we used the DC_{ef} of the corresponding region and performed all calculations as a function of the pH of the medium, implying that our calculations simulate a titration process. Using these potential values, we determined the corresponding pK_i* values and, in order to validate our theoretical model, we compared them with those measured by Henchóz[48] as indicated in Table 3. In this connection, the following notations were used: pK_1* is the p*K** of the acidic group attached to C_α, group 1; pK_2* is the p*K** of the basic group attached to C_α, group 2; pK_3* is the p*K** of the ionizable group of the side chain, group 3. The results thus obtained agree with those obtained experimentally at the same conditions of temperature and ionic concentration (298.15 K and 150 mM of NaCl). This agreement supports our assumption that the electrostatic interaction between neighboring groups and ions in solution is the main factor[6] dictating the difference between *K** and *K* (i.e., the ionization constant when the group is a constituent of an amino acid and when this same group is free in solution, respectively). The modeling assigned to the groups 1 and 2, in all amino acids treated as ethanoic acid and methylamine, respectively, confers efficiency and versatility to the model because for each individual amino acid only the parameters related to the side chain are needed. A well-built model also based in structural similarity was used by Makowski[28] in the development of a force field to describe the interaction between side chains shaped as a nonpolar ellipsoids.

Table 3. Calculated p*K Values for Each Amino Acid at 298.15 K and 150 mM NaCl[a]**

amino acid	p*K**	calcd	exptl[44]
Asp	pK_1*	2.14	2.02
	pK_2*	9.98	10.03
	pK_3*	3.70	3.66
Glu	pK_1*	2.21	2.29
	pK_2*	9.89	9.87
	pK_3*	4.17	4.17
Cys	pK_1*	2.23	2.05
	pK_2*	10.10	10.19
	pK_3*	8.48	8.44
Tyr	pK_1*	2.26	2.33
	pK_2*	9.32	9.15
	pK_3*	10.05	10.57
Arg	pK_1*	2.12	1.95
	pK_2*	8.98	9.00
	pK_3*	12.49	>12.00
His	pK_1*	1.60	1.54
	pK_2*	9.20	9.18
	pK_3*	6.44	6.07
Lys	pK_1*	2.28	2.21
	pK_2*	9.19	9.19
	pK_3*	10.45	10.75

[a] pK_1* denotes the p*K** of the acid group attached to C_α. pK_2* denotes the p*K** of the basic group attached to this carbon, and pK_3* denotes that of the ionizable group in the side chain. The experimental (exptl) column corresponds to results obtained from Henchoz.[44]

With the increase of ionic concentration, all the calculated p*K** values rise and tend to reach the intrinsic p*K* value, which is the value for a group in its free form (i.e., not as part of an amino acid). Such behavior is due to electrostatic screening, since the larger the amount of ions in solution, the further is the interaction between a given ionizable group, and its neighboring groups will be attenuated.

The model used to describe amino acids can be applied to other molecules such as pentapeptides. To that end, we considered the C-terminus as an ethanoic acid and N-terminus as methylamine and determined the p*K** values of Gly and Ala pentapeptides. The pentapeptides analyzed have the following structures:

H·ALA−Ala−X−Ala−Ala·OH

H·Gly−Gly−X−Gly−Gly·OH

They are chosen to enable a comparison between our results and those reported in other studies.[47,48] In these peptides, at the X position a residue with ionizable side chain is placed, and that can be Asp, Cys, Arg, His, Cys, Tyr, or Lys. Thus, this approach has general applicability, allowing determination of not only p*K** but also the DC_{ef} of Val, Leu, or "mixed" pentapeptides. The intrinsic p*K* assigned to the X-position residue follows the same criteria of structural similarity used in amino acids. The calculated effective dielectric constant for the C-terminal side chain and for the N-terminal side chain is very close to that of the solvent, given by eq 1, and is approximately equal to 80. This result is explained by the conformation that such peptides adopt in aqueous solution because the interaction between the groups in the two above-mentioned regions is attenuated, practically, only by water molecules and ions in solution. Using the mean electrostatic potential, the p*K** calculated at the surface of the X residue is listed in Table 3 which allows a comparison with those provided by Thurlkill[49] and Gurd (Table 4).[50]

Figure 4 is a correlation plot for both amino acids and pentapeptides with, on the *x*-axis, the p*K** values reported in literature[48−50] and, on the *y*-axis, the p*K** values determined by our model. The Pearson correlation factor calculated from the data presented in Tables 2 and 3 is equal to 0.9983. This value being very close to unity indicates that our theoretical model efficiently produces values very near those obtained experimentally in a wide range of temperatures and ionic concentrations. Most of our results are well within the range presented by Creighton[49] which covers the p*K** values typically found on the surface of proteins. As in this case the charged residues are exposed deep into the solvent, the use of a DC_{ef} value equal to that of the solvent is entirely justified. Another major consideration is that when the p*K** of the side chain ionizable group is compared with its intrinsic p*K*, it appears that the difference between them is smaller in peptides than in amino acids. This occurs because in peptides, the neighboring

dx.doi.org/10.1021/jp411331p | J. Phys. Chem. B 2014, 118, 4053−4061

Table 4. p*K Values Calculated at the Conditions Applied in the Cited Studies: 100 mM KCl and 299.15 K for Gly Pentaptides and 100 mM KCl and 295.15 K for Ala Pentapeptides[a]**

	p*K**			
residue	calcd	Gurd et al.[50]	Thurlkill et al.[49]	Creighton et al.[51]
		Ala Pentapeptides		
Asp	3.97		3.67	3.9−4.0
Glu	4.,2		4.25	4.3−4.5
His	6.71		6.54	6.0−7.0
Cys	9.03		8.55	9.0−9.5
Tyr	10.25		9.84	10.0−10.30
Lys	10.35		10.40	10.4−11.1
		Gly Pentapeptides		
Asp	3.81	3.9		3.9−4.0
Glu	3.99	4.2		4.3−4.5
His	6.70	6.8		6.0−7.0
Tyr	10.25	10.0 (33 °C)		10.0−10.30
Lys	10.33	10.5		10.4−11.1

[a]Creighton[51] shows a range of values for residues commonly found at the surface of proteins.

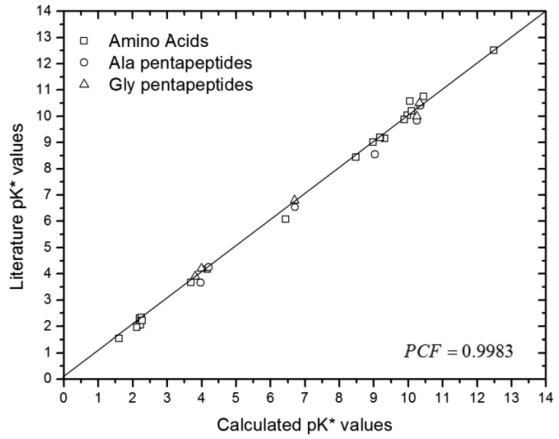


Figure 4. Correlation between calculated and experimental p*K** values from amino acids and pentapeptides. Literature values are provided by Henchoz[48] for amino acids, Thurlkill[49] for Ala pentapeptides, and Gurd[50] for Gly pentapeptides.

ionizable groups (C and N terminus groups) are at larger distance from the X-position residue than in amino acids, and thus, the side chain group in pentapeptides will be less sensitive to interactions with the two terminal groups. These results indicate that to study ionizable residues in proteins, we must adopt a consideration such as a first-neighbors approximation, i.e., that the p*K** of group *i* predominantly senses the influence of relatively close residues. To establish the minimal (cutoff) interaction distance and the number of groups to be taken into account for the calculation of the p*K** of each *i* group, we have performed calculations considering the two and four neighboring groups nearest the residue under consideration. Intrinsic p*K* values were assigned again according to the structural similarity criteria. With this in mind, the model was applied to determine the p*K** of Lys and Glu residues in SNase variants (3BDC and 1TQO PDB files, respectively), and the calculated values are presented in Table 5. We also calculated the effective dielectric constant and compared both results with those reported by Isom[13] and Castañeda.[47] The spatial parameters of group dimensions and distance between them were determined by PyMol[52] and corresponding PDB files.

In order to determine the DC_{ef} of the region between acid and basic pair groups, we used eq 3. The results are shown in Table 5. It appears that when the ionizable residues are further than 20 Å apart, the difference between our results and those reported by Isom and Castañeda tends to become substantial. This is because for these distances, which correspond to approximately twice the Debye length, effects that are ignored by our model become appreciable and thus affect the calculus of the electrostatic potential at each group surface and, consequently, its p*K**. In Table 5 we highlighted the values of those parameters that are in better agreement (within 0.1 unity) with the literature data.[13,47] The results for glutamic acid residues appear to be in better agreement than those for lysine. A possible explanation is that the calculated DC_{ef} for Glu has the same value as was used by Castañeda, whereas for the Lys A132K residue we obtained a DC_{ef} equal to 75, while Isom et al. used a slightly lower value (78.5). For the residue Glu 135, in 1M, the p*K** values obtained both for three and five groups are practically "equidistant" from the value reported by the reference.

Lys A109K p*K** value shows a large discrepancy with that reported in ref 13 because we have taken into account groups at a distance of >20 Å, both for three and five groups (i.e., two and four neighbors, respectively). When considering only the interaction of a neighbor closer than 20 Å, a p*K**of 9.9 is obtained for this residue that, indeed, is considerably closer to 9.20. With respect to the computational cost, for all the analyzed residues in SNase the time required to determine its p*K** is always less than 1 s. Using the data presented in Table 5, we show a correlation plot, presented in Figure 5. We have obtained a Pearson correlation factor (PCF) of 0.9976 when taking into account all groups inside a cutoff distance of 20 Å of the *i*-residue analyzed. In this study, the maximum number of residues included was 5 because of the intrinsic sequence and geometry of the SNase regions analyzed. These results once more demonstrate the large degree of agreement between our methodology and those used in other studies and prove the effectiveness of the adopted considerations and approximations. The difference between amino acids−pentapeptides and residues of SNase PCF values can be explained by the consideration that in this latter case we have to take into account additional groups to determine the mean electric potential while in addition the effects of larger distances are more pronounced.

■ CONCLUSIONS

Our method to calculate the effective dielectric constant, based on first principles, renders values that are consistent with those presented by Isom et al.[12,13] and Castañeda et al.[47] for residues in staphylococcal nuclease. However, in our model the effects of ionic concentration and temperature are implicitly taken into account by the use of the Debye−Hückel theory, the effective dielectric constant remaining sensitive to solvent restructuration. The values found for DC_{ef} display substantial variability, in particular in the region close to the interface between media with different dielectric constants,[9] and for regions buried inside the protein, these values are only weakly affected by the ions in solution, in agreement with data reported by Mukerjee.[46] The method developed to compute the position-dependent DC_{ef} can be implemented in force fields that

dx.doi.org/10.1021/jp411331p | J. Phys. Chem. B 2014, 118, 4053−4061

Table 5. Calculated Values of Effective Dielectric Constant (DC_{ef}) and pK* Values for Residues in SNase and Comparison with Those Presented by Isom[13] (Lys Residues) and Castañeda[47] (Glu Residues)[a]

		3[b] groups		5[b] groups				max R[e] (Å)	
i residue	DC_{ef}	pK*	CT[c]	pK*	CT[c]	pK* [13,47]	Δ[d] 3^b–5^b	3^b	5^b
Lys A132K	75	10.65	91.62	10.50	108.03	10.40[13]	0.25–0.10	13.14	15.28
Glu 101	20	4.08	70.69	3.84	129.02	3.81[47]	0.27–0.03	13.75	18.04
Glu 101 (1 M)	20	4.67	70.82	4.49	140.08	4.41[47]	0.26–0.08	13.75	18.04
Lys A109K	24	10.82	137.09	10.62	145.38	9.20[13]	1.62–1.42	23.0	32.15
Glu 135	20	3.84	76.74	3.81	115.17	3.76[47]	0.08–0.05	12.4	15.85
Glu 135 (1.0 M)	20	3.96	93.44	4.19	185.63	4.08[47]	0.12–0.11	12.4	15.85
Glu 10	21	2.84	73.63	3.20	109.45	2.82[47]	0.02–0.38	10.8	27.01
Glu 10 (1 M)	21	3.70	78.22	4.02	110.74	3.45[47]	0.25–0.57	10.8	27.01
Glu 129	79	4.55	73.00	3.85	113.28	3.75[47]	0.80–0.10	13.85	14.01
Glu 129 (1 M)	79	5.12	265.47	4.61	683.06	4.32[47]	0.80–0.29	13.85	14.01

[a]When not specified, the calculation was done in 0.1 M NaCl. CT is the computational time, in milliseconds, required by the calculations. [b]Groups considered in the calculation of mean electrostatic potential. [c]In milliseconds, in an Intel Core i3-2310M, 2.10 GHz, 4.0 GB RAM, 64 bit operational system. [d]Δ is the difference (in module) between the calculated pK* values and those reported by the reference for each residue when considering 3 or 5 groups inside a cutoff distance presented in the max R column. [e]Distance between the i residue (ionizable side chain group) and the farthest ionizable group taken into account to calculate the pK_i*.

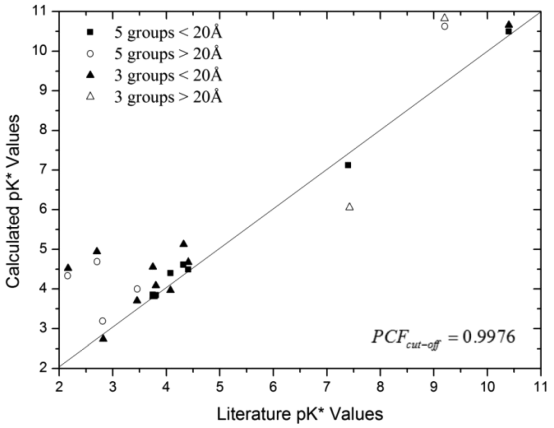


Figure 5. Correlation between calculated and experimental pK* values from SNase residues and comparison of cutoff distances. References are Isom et al.[13] for Lys residues and Castañeda et al.[47] for Glu residues.

describe the interactions between charged groups or interfaces of different media.

By the calculus of the mean electrostatic potential at the surface of ionizable groups, the developed model allows us to determine the pK* that each group assumes in specified conditions of temperature and ionic concentration. As the results obtained for amino acids are in good agreement with reported experimental data,[48] the model was also applied to larger molecules, such as Ala and Gly-pentapeptides and again the outcome was in accordance with published data.[49,50] The Pearson correlation factor (PCF) obtained for both amino acids and pentapeptides was 0.9983.

The results on Lys and Glu residues of staphylococcal nuclease were obtained by means of an approximation similar to first neighbors. In this connection, we considered that the pK* of a residue i is given by interactions of three or five groups within a cutoff distance of 20 Å. Even for complex conformations, as in SNase, the computational time required for determine the pK* of each residue was on the order of milliseconds. The resulting agreement with literature data (PCF = 0.9976) confirms the validity of our theoretical model which in addition implies a substantial reduction of the computational time and avoids the necessity, for the studied situations, of calling ions explicitly and to consider all atoms (in fact, we only took into consideration the nitrogen and oxygen atoms). With the pK* values of each residue in both peptides and proteins being known, one could infer the mean electric charge of the entire protein as well as of a specific region, such as the active site where the considered residues are located.[53,54]

We are currently improving our model to analyze other proteins in a general way and to allow the determination of pK* values and the effective dielectric constant in more complex situations.

■ ASSOCIATED CONTENT

Supporting Information

Detailed solution of the boundary problem of two charges inside a sphere that delimits two distinct dielectric media. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: terezapsouza@gmail.com. Phone: 1-617- 496-5953. Current address: Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02142, U.S.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge Dr. Augusto Agostinho Neto (in memoriam) for his crucial contributions to this paper. G.V.B. was supported by The State of Sao Paulo Research Foundation (FAPESP, Process Grant 2011/03150-8), National Council of Technological and Scientific Development—CNPq, and Coordination of Superior Level Staff Improvement—CAPES as a master student. T.P.d.S. was supported by the Alexander von Humboldt Foundation as a postdoctoral fellow. The authors thank Drs. V. B. P. Leite, M. A. Fossey, E. Drigo Filho, and J. R. Ruggiero for suggestions and fruitful discussions during the manuscript preparation. We also thank Dr. Gerrit Scherphof (University of Groningen, The Netherlands) for language editing of the manuscript.

REFERENCES

(1) Gao, J.; Müller, P.; Wang, M.; Eckhardt, S.; Lauz, M.; Fromm, K. M.; Giese, B. Electron Transfer in Peptides: The Influence of Charged Amino Acids. *Angew. Chem., Int. Ed.* **2011**, *50*, 1926−1930.

(2) Nishikawa, M.; Hasegawa, S.; Yamashita, F.; Takakura, Y.; Hashida, M. Electrical Charge on Protein Regulates Its Absorption from the Rat Small Intestine. *Am. J. Physiol.: Gastrointest. Liver Physiol.* **2002**, *282*, G711−G719.

(3) Harms, M. J.; Castañeda, C. A.; Schlessman, J. L.; Sue, G. R.; Isom, D. G.; Cannon, B. R.; García-Moreno E, B. The pK(a) Values of Acidic and Basic Residues Buried at the Same Internal Location in a Protein Are Governed by Different Factors. *J. Mol. Biol.* **2009**, *389*, 34−47.

(4) Ellenbogen, E. J. Dissociation Constants of Peptides. I. A Survey of the Effect of Optical Configuration. *J. Am. Chem. Soc.* **1952**, *74*, 5198−5201.

(5) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144−1149.

(6) Laurents, D. V.; Huyghues-Despointes, B. M. P.; Bruix, M.; Thurlkill, R. L.; Schell, D.; Newsom, S.; Grimsley, G. R.; Shaw, K. L.; Treviño, S.; Rico, M.; et al. Charge−Charge Interactions Are Key Determinants of the pK Values of Ionizable Groups in Ribonuclease Sa (pI=3.5) and a Basic Variant (pI=10.2). *J. Mol. Biol.* **2003**, *325*, 1077−1092.

(7) Schutz, C. N.; Warshel, A. What Are the Dielectric "Constants" of Proteins and How To Validate Electrostatic Models? *Proteins* **2001**, *44*, 400−417.

(8) Israelachvili, J. N. *Intermolecular and Surface Forces*, 3rd ed.; Academic Press: London, 2010.

(9) Pack, G. R.; Garrett, G. A.; Wong, L.; Lamm, G. The Effect of a Variable Dielectric Coefficient and Finite Ion Size on Poisson−Boltzmann Calculations of DNA−Electrolyte Systems. *Biophys. J.* **1993**, *65*, 1363−1370.

(10) Petrowsky, M.; Frech, R. Application of the Compensated Arrhenius Formalism to Self-Diffusion: Implications for Ionic Conductivity and Dielectric Relaxation. *J. Phys. Chem. B* **2010**, *114*, 8600−8605.

(11) García-Moreno, B. E.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. Experimental Measurement of the Effective Dielectric in the Hydrophobic Core of a Protein. *Biophys. Chem.* **1997**, *64*, 211−224.

(12) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; Velu, P. D.; García-Moreno E, B. Charges in the Hydrophobic Interior of Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 16096−16100.

(13) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; García-Moreno E, B. Large Shifts in pKa Values of Lysine Residues Buried Inside a Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5260−5265.

(14) Nakamura, H.; Sakamoto, T.; Wada, A. Roles of Electrostatic Interaction in Proteins. *Protein Eng.* **1988**, *2*, 177−183.

(15) Mellor, B. L.; Cruz Cortés, E.; Busath, D. D.; Mazzeo, B. A. Method for Estimating the Internal Permittivity of Proteins Using Dielectric Spectroscopy. *J. Phys. Chem. B* **2011**, *115*, 2205−2213.

(16) Rees, D. C. Experimental Evaluation of the Effective Dielectric Constant of Proteins. *J. Mol. Biol.* **1980**, *141*, 323−326.

(17) Warshel, A.; Aqvist, J. Electrostatic Energy and Macromolecular Function. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267−298.

(18) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent Properties of Proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.

(19) Karp, D. A.; Gittis, A. G.; Stahley, M. R.; Fitch, C. A.; Stites, W. E.; García-Moreno E, B. High Apparent Dielectric Constant inside a Protein Reflects Structural Reorganization Coupled to the Ionization of an Internal Asp. *Biophys. J.* **2007**, *92*, 2041−2053.

(20) Patargias, G. N.; Harris, S. A.; Harding, J. H. A Demonstration of the Inhomogeneity of the Local Dielectric Response of Proteins by Molecular Dynamics Simulations. *J. Chem. Phys.* **2010**, *132*, 235103.

(21) Gunner, M. R.; Zhu, X.; Klein, M. C. MCCE Analysis of the pKas of Introduced Buried Acids and Bases in Staphylococcal Nuclease. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3306−3319.

(22) Mehler, E. L. The Lorentz−Debye−Sack Theory and Dielectric Screening of Electrostatic Effects in Proteins and Nucleic Acids. In *Molecular Electrostatic Potentials: Concepts and Applications*; Elsevier: Amsterdam, 1996; Vol. 3, pp 371−405.

(23) Debye, P. *Polar Molecules*; Dover: New York, 1929.

(24) Lorentz, H. A. *Theory of Electrons*; Dover Publications: New York, 1952.

(25) Hassan, S. A. Liquid-Structure Forces and Electrostatic Modulation of Biomolecular Interactions in Solution. *J. Phys. Chem. B* **2006**, *111*, 227−241.

(26) Gong, H.; Hocky, G.; Freed, K. F. Influence of Nonlinear Electrostatics on Transfer Energies between Liquid Phases: Charge Burial Is Far Less Expensive than Born Model. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11146−11151.

(27) Wang, L.; Hingerty, B. E.; Srinivasan, A. R.; Olson, W. K.; Broyde, S. Accurate Representation of a B-DNA Double Helical Structure with Implicit Solvent and Counterions. *Biophys. J.* **2002**, *83*, 382−406.

(28) Makowski, M.; Liwo, A.; Sobolewski, E.; Scheraga, H. A. Simple Physics-Based Analytical Formulas for the Potentials of Mean Force of the Interaction of Amino-Acid Side Chains in Water. V. Like-Charged Side Chains. *J. Phys. Chem. B* **2011**, *115*, 6119−6129.

(29) Khandogin, J.; Brooks, C. L. Toward the Accurate First-Principles Prediction of Ionization Equilibria in Proteins. *Biochemistry* **2006**, *45*, 9363−9373.

(30) Bashford, D. Macroscopic Electrostatic Models for Protonation States in Proteins. *Front. Biosci.* **2004**, *9*, 1082−1099.

(31) Nielsen, J. E.; Gunner, M. R.; García-Moreno, B. E. The pKa Cooperative: A Collaborative Effort To Advance Structure-Based Calculations of pKa Values and Electrostatic Effects in Proteins. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3249−3259.

(32) Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D. Empirical Relationships between Protein Structure and Carboxyl pKa Calues in Proteins. *Proteins: Struct., Funct., Bioinf.* **2002**, *48*, 388−403.

(33) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications: New York, 1986.

(34) Levy, A.; Andelman, D.; Orland, H. Dielectric Constant of Ionic Solutions: A Field-Theory Approach. *Phys. Rev. Lett.* **2012**, *108*, 227801.

(35) Grosse, C. Extension of a Classic Theory of the Low Frequency Dielectric Dispersion of Colloidal Suspensions to the High Frequency Domain. *J. Phys. Chem. B* **2010**, *114*, 12520−12527.

(36) Williams, R. pKa Data Compiled by R. Williams. http://research.chem.psu.edu/brpgroup/pKa_compilation.pdf (2004).

(37) *CRC Handbook of Chemistry and Physics*, 92nd ed.; CRC Press: Boca Raton, FL, 2011.

(38) Wallace, J. A.; Shen, J. K. Predicting pKa Values with Continuous Constant pH Molecular Dynamics. In *Biothermodynamics. Part B*; Johnson, M. L., Ackers, G. K., Holt, J. M., Eds.; Academic Press: San Diego, CA, 2009; Vol. *466*, pp 455−475.

(39) Montal, M.; Mueller, P. Formation of Bimolecular Membranes from Lipid Monolayers and a Study of Their Electrical Properties. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 3561−3566.

(40) Hamelin, J.; Mehl, J. B.; Moldover, M. R. The Static Dielectric Constant of Liquid Water between 274 and 418 K Near the Saturated Vapor Pressure. *Int. J. Thermophys.* **1998**, *19*, 1359−1380.

(41) Robinson, R. A.; Stokes, R. H.; Bates, R. G. Electrolyte Solutions: The Measurement and Interpretation of Conductance, Chemical Potential and Diffusion in Solutions of Simple Electrolytes. *J. Electrochem. Soc.* **1960**, *107*, 205C−206C.

(42) Drigo Filho, E.; Agostinho Neto, A. Equilibrium Ionic Distribution as a Boundary Value Problem. *Trends Phys. Chem.* **2004**, *10*, 65−77.

(43) Agostinho Neto, A.; Drigo Filho, E.; Fossey, M. A.; Ruggiero Neto, J. pK Determination. A Mean Field, Poisson−Boltzmann Approach. *J. Phys. Chem. B* **1999**, *103*, 6809−6814.

(44) Arfken, G. B.; Weber, H. J. *Física Matemática: Métodos Matemáticos para Engenharia e Física*; Campus Elsevier: Rio de Janeiro, Brazil, 2007.

(45) Kolmogorov, A. N.; Fomin, S. V. *Elementos da Teoria das Funções e de Análise Funcional*; Mir: Moscou, Russia, 1982.
(46) Mukerjee, P.; Desai, N. R. *Nature* **1969**, *223*, 1056−1057.
(47) Castañeda, C. A.; Fitch, C. A.; Majumdar, A.; Khangulov, V.; Schlessman, J. L.; García-Moreno, B. E. Molecular Determinants of the pKa Values of Asp and Glu Residues in Staphylococcal Nuclease. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 570−588.
(48) Henchoz, Y.; Schappler, J.; Geiser, L.; Prat, J.; Carrupt, P.-A.; Veuthey, J.-L. Rapid Determination of pKa Values of 20 Amino Acids by CZE with UV and Capacitively Coupled Contactless Conductivity Detections. *Anal. Bioanal. Chem.* **2007**, *389*, 1869−1878.
(49) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK Values of the Ionizable Groups of Proteins. *Protein Sci.* **2006**, *15*, 1214−1218.
(50) Keim, P.; Vigna, R. A.; Morrow, J. S.; Marshall, R. C.; Gurd, F. R. N. Carbon 13 Nuclear Magnetic Resonance of Pentapeptides of Glycine Containing Central Residues of Serine, Threonine, Aspartic and Glutamic Acids, Asparagine, and Glutamine. *J. Biol. Chem.* **1973**, *248*, 7811−7818.
(51) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman and Company: New York, 1993.
(52) DeLano, W. *The PyMOL Molecular Graphics System*; DeLano Scientific: South San Francisco, CA, 2002
(53) Velyvis, A.; Kay, L. E. Measurement of Active Site Ionization Equilibria in the 670 kDa Proteasome Core Particle Using Methyl-TROSY NMR. *J. Am. Chem. Soc.* **2013**, *135*, 9259−9262.
(54) Cottrell, J. W.; Scott, L. G.; Fedor, M. J. The pH Dependence of Hairpin Ribozyme Catalysis Reflects Ionization of an Active Site Adenine. *J. Biol. Chem.* **2011**, *286*, 17658−17664.