



Published in final edited form as:

J Phys Chem B. 2008 September 4; 112(35): 11122–11136. doi:10.1021/jp709969n.

FAMBE-pH: A Fast and Accurate Method to Compute the Total Solvation Free Energies of Proteins

Yury N. Vorobjev[†], Jorge A. Vila^{‡,§}, and Harold A. Scheraga[‡]

Yury N. Vorobjev: ; Jorge A. Vila: ; Harold A. Scheraga: has5@cornell.edu

[†]Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Science, Novosibirsk, Russia

[‡]Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

[§]Universidad Nacional de San Luis, Instituto de Matemática Aplicada San Luis, CONICET, Ejército de Los Andes 950-5700, San Luis-Argentina

Abstract

A fast and accurate method to compute the total solvation free energies of proteins as a function of pH is presented. The method makes use of a combination of approaches, some of which have already appeared in the literature; (i) the Poisson equation is solved with an optimized fast adaptive multigrid boundary element (FAMBE) method; (ii) the electrostatic free energies of the ionizable sites are calculated for their neutral and charged states by using a detailed model of atomic charges; (iii) a set of optimal atomic radii is used to define a precise dielectric surface interface; (iv) a multilevel adaptive tessellation of this dielectric surface interface is achieved by using multisized boundary elements; and (v) 1:1 salt effects are included. The equilibrium proton binding/release is calculated with the Tanford–Schellman integral if the proteins contain more than ~20–25 ionizable groups; for a smaller number of ionizable groups, the ionization partition function is calculated directly. The FAMBE method is tested as a function of pH (FAMBE-pH) with three proteins, namely, bovine pancreatic trypsin inhibitor (BPTI), hen egg white lysozyme (HEWL), and bovine pancreatic ribonuclease A (RNaseA). The results are (a) the FAMBE-pH method reproduces the observed pK_a 's of the ionizable groups of these proteins within an average absolute value of 0.4 pK units and a maximum error of 1.2 pK units and (b) comparison of the calculated total pH-dependent solvation free energy for BPTI, between the exact calculation of the ionization partition function and the Tanford–Schellman integral method, shows agreement within 1.2 kcal/mol. These results indicate that calculation of total solvation free energies with the FAMBE-pH method can provide an accurate prediction of protein conformational stability at a given fixed pH and, if coupled with molecular mechanics or molecular dynamics methods, can also be used for more realistic studies of protein folding, unfolding, and dynamics, as a function of pH.

Introduction

About 30% of the residues (namely, Asp, Glu, His, Lys, Tyr, and Arg) of proteins are ionizable.¹ The ionization equilibria depend on several solution variables such as the pH, salt concentration, temperature, and so forth, as well as on the conformation of the protein. In particular, ionization equilibria are an important determinant of protein structure and function because they define the charges of the ionizable groups and, consequently, the long-range

Correspondence to: Harold A. Scheraga, has5@cornell.edu.

electrostatic interactions that characterize intra- and intermolecular interactions and protein–solvent interactions. A realistic estimation of the stability of a protein in aqueous solution has to take account of both the solvation free energy, as an interaction with the surrounding solvent, and the ionization free energy of the protein at a given pH, that is, as the free energy of the proton binding/release equilibrium. The observed pK_a 's of the ionizable groups depend on the conformation of the molecule and on the environment of these groups in the macromolecule;² that is, the folding pathway of a polypeptide in aqueous solution is tightly coupled with its ionization equilibria. Traditional approaches for treating polypeptides and proteins in molecular mechanics and molecular dynamics neglect this interdependency between conformation and ionization equilibrium by simply assuming that the charges on the amino acid residues are invariant to conformational changes. For a given pH, the residues are assumed to be charged or neutral at the beginning of a simulation, and this charge distribution is kept constant throughout the entire simulation.

The reason for such a crude approximation is very simple: there is a large computational price to pay when the more realistic approach is used. The 2^ζ possible ionization states of the whole molecule for every conformation have to be considered, with ζ being the number of ionizable groups in the molecule. As a consequence, adoption of a fixed charge distribution during a simulation may introduce an undesired bias to the folding process. Hence, because of its importance in the study of biological processes, the theory of protein titration has been the subject of extensive research for many decades.^{3–12} Recent evidence^{12–14} indicates that a correct description of electrostatic interactions that considers all states of ionization may be crucial for understanding protein stability and, consequently, for discriminating the native state from non-native conformations. To alleviate the effort required for the computation of the 2^ζ possible ionization states in large molecules, different research groups^{15–18} have developed approximate solutions to solve the proton binding/release equilibrium problem. Among them, molecular dynamics (MD) with Monte Carlo (MC) sampling of the protonation states has been adopted.^{15–18} In these applications, the ionization states are selected by using a MC procedure with a Metropolis criterion based on the free-energy difference between protonation states. With the newly selected distribution of charges, a series of MD steps is carried out in an implicit or explicit water solvent. The procedure is repeated iteratively. It is worth noting that MD simulations at fixed pH make use of several other approximations, such as the use of a set of additional continuous coordinates, $0 > \lambda_i > 1$, to describe the ionization degree of site i , with λ_i representing the charge state of site i .^{19–22}

There are two major issues that must be addressed before carrying out simulations in which proton binding/release equilibrium is considered. First, the enormous number of ionization states, that is, $\sim 2^{[N/3]}$ for a protein with N residues (of which $\sim 30\%$ are ionizable), should be treated properly,^{13,14} recognizing that, because its exact treatment is prohibitively expensive, an exact solution to this problem is restricted to proteins with no more than ~ 25 residues.¹⁴ Any approximation used to surmount such a problem forces us to deal with the accuracy of the computed results.²⁰ A solution to this problem, in the treatment of large proteins, relies on the method adopted to calculate the average ionization degree of each of the ionizable sites, for example, by use of a MC random walk in the ionization space and of the Tanford–Schellman integral method, given by Tanford²³ and later by Schellman,²⁴ with an explicit expression by Yang and Honig.²⁵

Second, a fast and accurate method to compute the solvent polarization free energy, for a given fixed protein conformation, is needed. A solution to this problem, representing a further development of the treatment given in refs 10 and 26 will be provided.

A practical implementation of such approaches requires answers to some questions, which are addressed here: (a) what is the optimal algorithm to compute the multisite ionization energy

and (b) what is the optimal set of parameters to produce a fast and accurate method? Solutions of these problems, for a given protein conformation at a fixed pH, are provided here by a new algorithm that generalizes the fast adaptive multigrid boundary element (FAMBE) method.²⁶ This new algorithm enables one to increase the speed of electrostatic calculations markedly for large proteins without loss of accuracy. It will provide us with both the solvation free energies of the ionizable residues in water and an accurate solution to the free energy of ionization equilibrium, with the use of the Tanford–Schellman^{23–25} integral. A test on three proteins, which differ in the number of residues, the three-dimensional topology, and the biological function, illustrates the accuracy of the method.

Methods Section

Theoretical Background

The process of dissolving a protein in water in the presence of hydrogen ions can be modeled as a four-stage thermodynamic process:¹⁰ (1) creation of a solutesized cavity in water; (2) insertion of the zero-charged protein (with all atoms having zero partial charge) into the cavity in water; (3) charging of the protein to the gas-phase partial atomic charges in which all ionizable groups are maintained neutral; and (4) an equilibrium titration of the protein from $\text{pH} = \infty$ (i.e., zero hydrogen ion concentration) to a given pH value. The first three stages of this partition describe the solvation of a neutral protein. The whole thermodynamic cycle defines the total free energy $G(\mathbf{x}, \text{pH})$ of a single protein molecule in water at a given fixed pH in an instantaneous microscopic conformation \mathbf{x}

$$G(\mathbf{x}, \text{pH}) = U_{\text{mol}}^0(\mathbf{x}) + [G_{\text{cav}}(\mathbf{x}) + G_{\text{s,vdw}}(\mathbf{x})] + G_{\text{pol}}^0(\mathbf{x}) + \Delta G_{\text{inz}}(\mathbf{x}, \text{pH}) \quad (1)$$

where $U_{\text{mol}}^0(\mathbf{x})$ is the intramolecular conformational potential energy of the protein computed in a gas-phase approximation, $G_{\text{cav}}(\mathbf{x})$ is the free energy for creation of the molecular cavity in water (stage 1), $G_{\text{s,vdw}}(\mathbf{x})$ is the free energy of van der Waals interactions between the uncharged protein and the water solvent (stage 2),²⁷ $G_{\text{pol}}^0(\mathbf{x})$ is the free energy of polarization of the water solvent by the protein with gas-phase partial charges on all atoms but with the ionizable groups in the neutral state (stage 3), and the last term $\Delta G_{\text{inz}}(\mathbf{x}, \text{pH})$ is the free energy of ionization of the protein at a given pH with respect to the gas-phase atomic charges as a reference state in which all ionizable groups are neutral (stage 4). The sum of all terms but the first one of eq 1 is equal to the total free energy of protein solvation for a given conformation \mathbf{x} at fixed pH.^{10,27}

The only terms of eq 1 that are considered here are $G_{\text{pol}}^0(\mathbf{x})$ and $\Delta G_{\text{inz}}(\mathbf{x}, \text{pH})$. In addition, since eq 1 pertains to a single conformation \mathbf{x} , it is necessary to average over the ensemble of all such conformations.²⁷ However, in interpreting a titration curve, the present practice is to use a single molecular conformation as a representative of the conformational ensemble, although some progress has been made with MC and MD simulations and different electrostatic treatments to treat the whole conformational ensemble.^{15–22}

Multiple-Site Ionization Equilibrium to Compute $\Delta G_{\text{inz}}(\mathbf{x}, \text{pH})$

To define the ionization free energy $\Delta G_{\text{inz}}(\mathbf{x}, \text{pH})$, we assume that a protein has ζ ionizable groups, and the available 2 ionization microstates are described by the vector $\mathbf{z} = (1, 0, 0, 1, \dots, 0)$, where 1 at position i indicates that site i is charged. The underlying theory of multiple proton equilibrium can be found in several previous publications.^{6–8} The free energy for dissociation of hydrogen ions from an amino acid side chain S_i , of the protein P, can be defined relative to the dissociation of hydrogen ions from the isolated amino acid S_i by the following scheme



The value of $\text{p}K_a(S_i^+)$ of the reference system, that is, of the model compound S_i^+ , is taken from experimental data.^{5,6} The $\text{p}K_a^{(1)}$ of a given single ionizable group, indicated by the superscript (1), with all other groups kept neutral, is shifted by $\Delta\text{p}K_a^{(1)}$ due to the protein structure, that is

$$\Delta\text{p}K_i^{(1)} = \text{p}K_a^{(1)}(PS_i^+) - \text{p}K_a(S_i^+) = \frac{\gamma_i}{kT \ln 10} \{ [G(PS_i) - G(PS_i^+)] - [G(S_i) - G(S_i^+)] \} \quad (3)$$

where the value of $\text{p}K_a(S_i^+)$ of the reference system, that is, of the model compound S_i^+ , is taken from experimental data^{5,6} and the parameter γ_i is equal to 1 or -1 if the ionizing group is a base or an acid, respectively. Usually, the structure of a model compound S_i^+ for side chain i is equal to that of the entire residue in the same conformation as it is in the protein PS_i^+ . The free energies in eq 3 consist of two components, namely, the electrostatic intramolecular interactions and the interactions with the solvent. Finally, the free energy of ionization of the single site S_i^+ in protein P at a given pH, that is, with all other ionizable groups in the protein kept neutral, is given by

$$\begin{aligned}
 \Delta G(PS_i^+, \text{pH}) &= [G(PS_i^+) - G(PS_i)] \\
 &= kT(\ln 10)\gamma_i[\text{pH} - \text{p}K_a^{(1)}(PS_i^+)]
 \end{aligned} \quad (4)$$

The multiple-site free energy of ionization of the protein, $\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH})$, from the neutral state $\mathbf{z} = 0$ to the ionization microstate \mathbf{z} , is computed as

$$\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH}) = G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH}) - G^{\text{PS}}(\mathbf{x}) \quad (5)$$

where $G^{\text{PS}}(\mathbf{x})$ is the free energy of the nonionized neutral protein. The energy $\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH})$ is a sum of energies of ionization of single sites and pairwise electrostatic interactions between the ionized sites; hence, eq 5 can be rewritten as

$$\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH}) = \sum_{i=1}^N \Delta G(PS_i^+, \text{pH}) + \frac{1}{2} \sum_{i \neq j}^N \Delta w_{ij}(\mathbf{x}) \quad (6)$$

where $\Delta G(PS_i^+, \text{pH})$ is given by eq 4 and $\Delta w_{ij}(\mathbf{x})$ is the excess electrostatic potential between ionized sites i, j , with respect to the nonionized sites. By inserting eq 4 into eq 6, the final expression for the free energy of ionization is given by

$$\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH}) = kT(\ln 10) \sum_{i=1}^N \gamma_i [\text{pH} - \text{p}K_a^{(1)}(\text{PS}_i^+)] + \frac{1}{2} \sum_{i \neq j}^N \Delta w_{ij}(\mathbf{x}) \quad (7)$$

The ionization free energy $\Delta G_{\text{inz}}^{\text{PS}^+}(\mathbf{x}, \text{pH})$ of the protein in the conformational microstate \mathbf{x} can be calculated exactly from the ionization partition function $Z_{\text{inz}}^{\text{PS}^+}$ as

$$\Delta G_{\text{inz}}^{\text{PS}^+}(\mathbf{x}, \text{pH}) = -kT \ln Z_{\text{inz}}^{\text{PS}^+} \quad (8)$$

where the partition function $Z_{\text{inz}}^{\text{PS}^+}$ is a sum over all 2^N ionization microstates \mathbf{z}

$$Z_{\text{inz}}^{\text{PS}^+} = \sum_{\mathbf{z}} \exp(-\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH})/kT) \quad (9)$$

If the total number of ionizable sites is larger than ~ 25 , then the number of ionization microstates is greater than 10^7 , and hence, it is prohibitively expensive to evaluate the summation in order to calculate $\text{p}K_a$'s.¹⁰ An alternative accurate and fast approximation to solve this problem for proteins with more than ~ 25 ionizable groups follows.

Fast Approximate Method to Calculate $\text{p}K_a$'s

The ionization free energy, $\Delta G_{\text{inz}}^{\text{PS}^+}(\mathbf{x}, \text{pH})$, can be calculated by the thermodynamic integration method as a titration process from zero hydrogen ion concentration to a given value of pH by means of the Tanford–Schellman integral.^{23–25} From eqs 7–9, the following relations can be obtained:

$$\begin{aligned} \frac{\partial \Delta G_{\text{inz}}^{\text{PS}^+}(\mathbf{x}, \text{pH})}{\partial \text{pH}} &= \left\langle \frac{\partial \Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH})}{\partial \text{pH}} \right\rangle \\ &= kT(\ln 10) \sum_{i=1}^N \gamma_i \langle \gamma_i(\mathbf{x}, \text{pH}) \rangle \end{aligned} \quad (10)$$

where $\langle \gamma_i(\mathbf{x}, \text{pH}) \rangle$ is the average ionization degree of site i , and

$$\langle \gamma_i(\mathbf{x}, \text{pH}) \rangle = \frac{1}{Z_{\text{inz}}^{\text{PS}^+}} \sum_{\mathbf{z}} \gamma_i \exp(-\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH})/kT) \quad (11)$$

From eqs 10 and 11, the Tanford–Schellman integral^{23–25} can be deduced:

$$\Delta \Delta G_{\text{inz}}(\mathbf{x}, \text{pH}) - \Delta \Delta G_{\text{inz}}(\mathbf{x}, \infty) = kT(\ln 10) \sum_{i=1}^N \gamma_i \int_{\infty}^{\text{pH}} [\langle \gamma_i(\mathbf{x}, \text{pH}) \rangle - \langle \gamma_i^{\text{isol}}(\text{pH}) \rangle] \text{d} \text{pH} \quad (12)$$

with the functions $\langle z_i(\mathbf{x}, \text{pH}) \rangle$ and $\langle z_i^{\text{isol}}(\text{pH}) \rangle$ representing the average ionization degree of site S_i^+ in the protein (PS_i^+) in conformational microstate \mathbf{x} and in the isolated model compound S_i^+ , respectively, and

$$\Delta\Delta G_{\text{inz}}(\mathbf{x}, \text{pH}) = \Delta G_{\text{inz}}^{\text{PS}^+}(\mathbf{x}, \text{pH}) - \Delta G_{\text{inz}}^{\text{S}^+}(\mathbf{x}, \text{pH}) \quad (13)$$

In eq 13, the free energy term $\Delta G_{\text{inz}}^{\text{S}^+}(\mathbf{x}, \text{pH})$ for the sum of all isolated ionizable residues is given by

$$\Delta G_{\text{inz}}^{\text{S}^+}(\mathbf{x}, \text{pH}) = -kT \sum_{i=1}^N \ln\{1 + \exp[-\gamma_i \ln 10(\text{pH} - \text{p}K_a(S_i^+))]\} \quad (14)$$

For site i in protein conformation \mathbf{x} at a given pH, the average ionization degrees $\langle z_i(\mathbf{x}, \text{pH}) \rangle$ can be calculated by a Monte Carlo random walk in the space of the ionization microstates. For $\text{pH} = \infty$, there is only one populated ionization microstate, namely $\mathbf{z}^a = (a^-, b^0)$, that is, when all acidic residues, \mathbf{z}^a , are negatively charged, a^- , while all of the basic residues are neutral, b^0 . Therefore, a solution for $\Delta\Delta G_{\text{inz}}(\mathbf{x}, \infty)$ can be obtained with eq 6 by computing

$$\Delta\Delta G_{\text{inz}}(\mathbf{x}, \infty) = \Delta G_{\text{inz}}^{\text{PS}^+}(\mathbf{x}, \mathbf{z}^a) - \Delta G_{\text{inz}}^{\text{S}^+}(\mathbf{x}, \mathbf{z}^a) \quad (15)$$

For the complementary situation (a^0, b^+), when all acidic residues are neutral, a^0 , while all the basic residues are positively charged, b^+ , an alternative expression for the free energy of ionization can be obtained by integration in eq 12 over the pH interval $(-\infty, \text{pH})$ as

$$\Delta\Delta G_{\text{inz}}(\mathbf{x}, \text{pH}) - \Delta\Delta G_{\text{inz}}(\mathbf{x}, -\infty) = kT(\ln 10) \sum_{i=1}^N \gamma_i \int_{-\infty}^{\text{pH}} [\langle z_i(\mathbf{x}, \text{pH}) \rangle - \langle z_i^{\text{isol}}, \text{pH} \rangle] d\text{pH} \quad (16)$$

while the corresponding $\Delta\Delta G_{\text{inz}}(\mathbf{x}, -\infty)$ value is given by

$$\Delta\Delta G_{\text{inz}}(\mathbf{x}, -\infty) = \Delta G_{\text{inz}}^{\text{PS}^+}(\mathbf{x}, \mathbf{z}^b) - \Delta G_{\text{inz}}^{\text{S}^+}(\mathbf{x}, \mathbf{z}^b) \quad (17)$$

where the ionization state is $\mathbf{z}^b = (a^0, b^+)$.

The ionization free energy can be calculated in two different ways, namely, by integrating from $+\infty$ or from $-\infty$, that is, by using eqs 12 and 15 and 16 and 17, respectively. This offers the opportunity to test the accuracy as well as the internal consistency of the procedure because both ways of integration should give the same results.

Finally, the value of the $\text{p}K_a(PS_i^+)$ of ionizable site S_i^+ in protein P with multiple-site ionization is defined from a titration curve of eq 11 as the value of the pH, that is, $\text{pH}_{1/2}$, at which

$$\langle z_i(\mathbf{x}, \text{pH}_{1/2}) \rangle = 1/2 \quad (18)$$

In order to compute $\langle z_i(\mathbf{x}, \text{pH}_{1/2}) \rangle$, it is necessary to compute $\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH})$ of eq 7.

Computation of $\langle z_i(\mathbf{x}, \text{pH}_{1/2}) \rangle$

Continuum Dielectric Model—For a protein embedded in a polar (water) solvent, it is common practice^{8–10,26,28} to use a continuum dielectric model (Figure 1) to calculate the electrostatic potential $\varphi(\mathbf{r})$ and the solvent reaction field by solving the Poisson equation

$$\nabla[D(\mathbf{r})\nabla\varphi(\mathbf{r})] = -4\pi \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (19)$$

where the charges are q_i at position \mathbf{r}_i in the protein conformation \mathbf{x} , with \mathbf{r} being the Cartesian coordinates at given positions, ∇ is the gradient, and the dielectric function $D(\mathbf{r})$ is modeled as a sharp dielectric interface, that is, as the surface $S(\mathbf{t})$ of the protein molecular cavity created, in stage 1, by excluding water. The internal volume of this excluded solvent cavity has a low dielectric constant D_1 , while the solvent has the bulk water dielectric constant D_0 . The position of the dielectric boundary separating the solute protein from the solvent is chosen empirically in the dielectric model and is defined by the set of atomic radii R_{D_i} which are used for calculating the smooth dielectric interface surface.²⁹ The method used to compute the dielectric interface must be defined with precision because it is a crucial component for an accurate prediction in macromolecular applications.^{27,29–31} For this reason, we use the smooth invariant molecular surface (SIMS) method²⁹ in this work for calculating the dielectric interface surface, that is, the surface of the cavity consisting of the solvent-excluded volume of the protein molecule and an optimized set of Born atomic radii for an accurate modeling of the solvent polarization energy of the protein in water. The SIMS method constitutes an improved algorithm over that of the Richards–Connolly method for calculating molecular surfaces (MS); it smooths all singularities of the Richards–Connolly MS by rolling a probe sphere inside of the MS in order to detect and remove discontinuities in the vector normal to the MS at each point, aimed at producing a uniform distribution of equal-sized surface elements.²⁹

Fast Adaptive Multigrid Boundary Element Method for Solution of the Poisson

Equation: The Poisson equation (eq 19) can be converted into an integral equation²⁶ for the induced charge density $\zeta(\mathbf{t})$ on the dielectric surface $S(\mathbf{t})$, where \mathbf{t} is a surface point

$$\sigma(\mathbf{t}) = f \int_s \frac{\sigma(\mathbf{s})(\mathbf{t} - \mathbf{s})\mathbf{n}(\mathbf{t})\text{d}s}{|\mathbf{t} - \mathbf{s}|^3} + \frac{f}{D_1} \sum_i \mathbf{n}_i \cdot \mathbf{E}_i(\mathbf{t}) \quad (20)$$

where $f = (1/2\pi)(D_1 - D_0)/(D_1 + D_0)$, $\mathbf{E}_i(\mathbf{t})$ is the electrostatic field $[\sum_k q_{i,k}(\mathbf{t} - \mathbf{r}_{i,k})/|\mathbf{t} - \mathbf{r}_{i,k}|^3]$, where k runs over all charges belonging to ionizable groups i generated by the set of charges belonging to the charged group i ($i = 1, \dots, N_g$, where N_g is the total number of charged groups)²⁶ of the solute, and $\mathbf{n}(\mathbf{t})$ is the vector normal to the molecular surface at point \mathbf{t} . The induced charge density $\zeta(\mathbf{t})$ approximates the average solvent-induced charge density. Since the term $\mathbf{E}_i(\mathbf{t})$ is linear in the charges q_i , it is possible to split $\zeta(\mathbf{t})$ given by eq 20 into a sum of terms, each one of which represents the polarization charge density, $\zeta_i(\mathbf{t})$, generated by a single group of charges of ionizable group i .²⁶ Hence, eq 20 for $\zeta(\mathbf{t})$ can be decomposed into a set of

independent minor integral equations, one for each of the polarization charge densities, $\zeta_i(\mathbf{t})$ ^{26,27}

$$\sigma_i(\mathbf{t}) = f \int_s \frac{\sigma_i(\mathbf{s})(\mathbf{t} - \mathbf{s})n(\mathbf{s})d\mathbf{s}}{|\mathbf{t} - \mathbf{s}|^3} + \frac{f}{D_i} n(\mathbf{t}) \cdot \mathbf{E}_i(\mathbf{t}) \quad i=1, 2, \dots, N_g \quad (21)$$

The reason for such a representation is that the integral in eq 21 for each component $\zeta_i(\mathbf{t})$ can be converted into a discrete linear equation of low dimensionality of a matrix \mathbf{M}_i . By this representation over only the set i of adaptive multisized boundary elements,²⁶ we obtain

$$\sigma_i(\mathbf{t}) = \mathbf{M}_i(\mathbf{t}, \mathbf{s})\sigma_i(\mathbf{s}) + \mathbf{E}_i(\mathbf{t}) \quad (22)$$

as an analogue of eq 9 of ref²⁶. For each charge group, i , the size of the boundary element increases steadily with the distance from the source of the molecular electrostatic field of that charged group, as shown in Figure 1. Hence, for any charge group i , the number of multisized boundary elements, that is, the dimensions of the vectors σ_i and \mathbf{E}_i and the matrix \mathbf{M}_i in eq 22, is significantly lower (see Result and Discussion section) than the total number of surface elements that would be encountered if the surface was tessellated by the finest uniform boundary elements of surface area s in eq 21. The number of multilevel boundary elements N_{MBE} which tessellate a MS with area A_S scales as

$$N_{\text{MBE}} \approx n_{\text{loc}} \log(A_S/A_{\text{loc}}) \quad (23)$$

where n_{loc} and A_{loc} are the average number of boundary elements and the size of the local area with the finest tessellation, respectively, if within R_1 .

The present adaptive tessellation algorithm is a generalization of the tessellation method by the boundary elements of three size levels, that is, small, large, and surface patches described in ref²⁶ (a surface patch represents the whole surface of one surface atom). The current tessellation method considers a generalized set of mutually inserted surface boundary elements (BE), that is, surface grids, of different average sizes, with average BE dimensions of d_i , $i = 1, 2, 3, \dots, n$, where i is the level of the BE and n is the total number of levels. The BEs of level i are a set of surface elements with areas s_α^i and coordinate of center \mathbf{t}_α^i , so that the whole molecular surface can be completely covered by the BEs of each level. Each BE has an average value of polarization charge density σ_α^i which is assumed to be constant over this BE. The BEs of the first, finest level $i = 1$ are calculated as a set of surface elements by the SIMS²⁹ method with high dot density and small average dimension d_1 ; the BEs of the second level are also calculated by the SIMS²⁹ method with low dot density, with average dimension d_2 . Each BE of the third level represents the whole surface of one surface atom. Each BE of levels 4, 5, and so forth represent a united surface of a group of nearest atoms. Each BE of level i is a collection of integer numbers of BEs of the previous level $i - 1$. The surface area, average polarization charge density, and position of the BE of level i are defined as the average values over the respective set of BEs of level $i - 1$ and are inserted into the BE of level i . That is why the total number of variables in eq 22 is small.

The FAMBE method constructs the adaptive tessellation, which depends on a specific center on the protein surface, for each σ_i (as given by eq 22). The FAMBE method uses a set of distances R_i for level i (Figure 1), using the boundary element set of level i to tessellate the

molecular surface (MS) in the vicinity of point \mathbf{t} if the distance r_{it} from the charged group q_i to the point \mathbf{t} of the MS is in the range of $R_{i-1} > r_{it} > R_i$ (the region between two arcs in Figure 1). The dimensions d_i of the multisized BEs and R_i determine both the accuracy of the numerical solution and the CPU time. The important primary parameters are the size of the finest BEs, d_1 , and the surface area covered by d_1 and defined by R_1 . All other parameters d_i and R_i are defined from d_1 and R_1 recursively by the relations d_i/d_{i-1} and R_i/R_{i-1} described below. An analysis of the dependence of the accuracy and CPU time on the tessellation parameters d_1 and R_1 for the three-level version of the FAMBE method has been carried out in ^{ref 26}. The optimization of the primary parameters of the method for determining d_1 and R_1 was carried out by Vorobjev and Scheraga²⁶ by comparing the numerical solution of eq 22 with an analytical one for a sphere with a single charge off center. The optimal values obtained for d_1 and R_1 were ~ 0.5 and ~ 5.0 Å, respectively. These optimal values of the parameters d_1 and R_1 are independent of the total number of multisized levels n . Each boundary element (BE) of level i consists of an integer number of boundary elements of the previous level $i - 1$, in the range of 3–6, with an average value of ~ 4 ; therefore, the average ratio for sizes d_i/d_{i-1} is ~ 2 , based on the square root of the ratio of the areas of BE i and $i - 1$. The relation $R_i/R_{i-1} \sim 2 - 1.5$ follows from the ratio $d_i/d_{i-1} \sim 2$ and from the required numerical accuracy of the solution of eq 22.

In summary, the use of the SIMS²⁹ method to compute the molecular surface area and the whole tessellation procedure described above are improvements of the corresponding methods used in previous work.²⁶

Each linear equation for σ_i , as given by eq 22, can be solved iteratively by the preconditioned biconjugated gradient method²⁶ after a few iterations, namely, 4–6. Therefore, the total numerical complexity of the FAMBE method scales linearly with the size of the protein, that is, the number of such equations is linear in protein size, which grows linearly with the number of ionizable groups.

Calculation of the Ionization Equilibrium—The key advantage of the FAMBE method over other solutions of the Poisson equation^{7–9} is that this method calculates the full set of minor charge densities $\sigma_i(\mathbf{t})$ simultaneously, that is, the system of N_g independent equations, given by eq 21 for $\sigma_i(\mathbf{t})$, can be solved simultaneously in N_g processors. These quantities are needed for calculating individual polarization free energies of the ionizable groups as well as for computing the electrostatic potential of the mean force Δw_{ij} between ionized groups i and j , in eq 7.^{10,26}

The partial charges on all atoms of an ionizable group differ depending on whether the group is charged or neutral, that is, q_i^+ and q_i^0 , respectively, as in the model compounds S_1^+ and S_1 . The set of charges of the whole structure, PS_1^+ , can be represented by q_i^+ plus the set of all of the other charges Q_i^P of the protein; likewise the total charge of PS_1 is q_i^0 plus Q_i^P . Each set of atomic charges induces a corresponding set of solvent polarization charges, namely, σ_i^+ , σ_i^P , and σ_i^0 , on the boundary elements of the MS, of the molecule (PS_1^+ or PS , respectively). The set of charges in microstate \mathbf{z} can be represented by the set of $q_{\mathbf{z}}^+$ charges containing the atomic charges of all ionized groups in the ionization microstate \mathbf{z} , while the rest of all of the atomic charges of the protein, except the atoms that belong to the ionized groups \mathbf{z} , are represented by $Q_{\mathbf{z}}^P$. For the neutral state of the molecule, the set of charges in microstate \mathbf{z} is represented by the set of $q_{\mathbf{z}}^0$ charges, while the rest of all of the atomic charges of the protein, except the atoms that belong to the groups \mathbf{z} , are represented by the same $Q_{\mathbf{z}}^P$. The set of atomic charges $q_{\mathbf{z}}^+$, $Q_{\mathbf{z}}^P$, $q_{\mathbf{z}}^0$ induces a corresponding set of solvent polarization charges, namely, $\sigma_{\mathbf{z}}^+$, $\sigma_{\mathbf{z}}^P$, $\sigma_{\mathbf{z}}^0$, on MS ($PS_{\mathbf{z}}^+$ or PS , respectively). Using this definition of charges, the total electrostatic free energy of the protein can be represented as a sum of charge–charge interactions between any pair of sets of atomic charges plus the interactions between the charges

and the surface-induced charges. After using the reciprocal relation for the product between potential and charge,³² the product of a set of atomic charges with a set of induced charges over the corresponding MS can be described by the following equation

$$[q_i^+ \cdot \sigma_i^P]_{\text{PS}_i^+} = [\sigma_i^+ \cdot Q_i^P]_{\text{PS}_i^+} \quad (24)$$

where

$$[q_i^+ \cdot \sigma_i^P]_{\text{PS}_i^+} = \frac{1}{2} \sum_{\alpha} q_{\alpha} \int_{\text{PS}_i^+} \frac{\sigma_i^P(s) ds}{|s - r_{\alpha}|} \quad (25)$$

is the product between the set q_i^+ and the set of induced charges σ_i^P on the MS of the structure PS_i^+ , and the subindex α runs over all atoms of the ionizable group i . With these equations, the excess free energy $\Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH})$ of eqs 5 and 6 can be written as

$$\begin{aligned} \Delta G^{\text{PS}^+}(\mathbf{x}, \mathbf{z}, \text{pH}) = & (Q_z^P \cdot q_z^+) + 2[Q_z^P \cdot \sigma_z^+]_{\text{PS}_z^+} - \\ & (Q_z^P \cdot q_z^0) - 2[Q_z^P \cdot \sigma_z^0]_{\text{PS}_z^+} + [Q_z^P \cdot \sigma_z^P]_{\text{PS}_z^+} - [Q_z^P \cdot \sigma_z^P]_{\text{PS}^+} + \\ & (q_z^+ \cdot q_z^+) + [q_z^+ \cdot \sigma_z^+]_{\text{PS}_z^+} - (q_z^0 \cdot q_z^0) - [q_z^0 \cdot \sigma_z^0]_{\text{PS}^+} \end{aligned} \quad (26)$$

where an example of the charge–charge product is

$$(Q_z^P \cdot q_z^+) = \frac{1}{2} \sum_{\text{P}_{\beta}} q_{\text{P}_{\beta}} \sum_{\eta} q_{\eta} \frac{1}{|r_{\text{P}_{\beta}} - r_{\eta}|} \quad (27)$$

where the index η runs over all ionizable atomic charges in the ionization microstate \mathbf{z} and P_{β} runs over all of the remaining atomic charges of the protein. A similar expression pertains to the other terms of eq 26. It should be noted that the terms in eq can be rearranged into the first sum of eq 6 and the sum of the pair interactions between ionizable sites, that is, the last term of eq 6. With the aid of eq 3, the value of $\text{p}K_a^{(1)}(\text{PS}_i^+)$ for the ionizable group in eq 7 can be reduced exactly to the following expression

$$\begin{aligned} \text{p}K_a^{(1)}(\text{PS}_i^+) &= \text{p}K_a(\text{S}_i^+) + \Delta \text{p}K_i^{(1)} \\ \Delta \text{p}K_i^{(1)} &= - \frac{\gamma_i}{kT(\ln 10)} \{ \Delta g_{\text{P}^+} - \Delta g_{\text{P}^0} + \Delta g_{\text{PP}} \} \end{aligned} \quad (28)$$

where $\Delta \text{p}K_i^{(1)}$ is the shift in the $\text{p}K_a$ of the ionizable group S_i^+ due to the protein environment while all other ionizable groups are neutral, and

$$\Delta g_{\text{P}^+} = [q_i^+ \cdot \sigma_i^+]_{\text{PS}_i^+} - [q_i^+ \cdot \sigma_i^+]_{\text{S}_i^+} + (Q_i^P \cdot q_i^+) + 2[Q_i^P \cdot \sigma_i^+]_{\text{PS}_i^+} \quad (29)$$

$$\Delta g_{p0} = [q_i^0 \cdot \sigma_i^0]_{ps} - [q_i^0 \cdot \sigma_i^0]_{s_i} + (Q_i^P \cdot q_i^0) + 2[Q_i^P \cdot \sigma_i^0]_{ps} \quad (30)$$

$$\Delta g_{pp} = [Q_i^P \cdot \sigma_i^P]_{ps_i^+} - [Q_i^P \cdot \sigma_i^P]_{ps} \approx 0 \quad (31)$$

The term Δg_{p+} is the desolvation penalty for ionization of residue S_i^+ in the protein environment, relative to the model compound, and is given by the first two terms of eq 29. On the other hand, the last two terms of eq 29 describe the effect of the protein field on the ionizable residue S_i^+ . Similarly, the terms of eq 30 describe the desolvation penalty and the protein field effect for the neutral residue S_i . The two terms in eq 31 represent the difference in polarization free energy because of the change in the MS due to protonation of the i th group and have been found to be negligible (see Results and Discussion section).

Rearranging expression 26 as a sum of pair terms gives the excess potential of mean force $\Delta w_{ij}(\mathbf{x})$, shown in eq 6, as

$$\Delta w_{ij} = w_{++} - w_{+0} \quad (32)$$

Where

$$w_{++}(\mathbf{x}) = (q_i^+ \cdot q_j^+) + 2[q_i^+ \cdot \sigma_j^+]_{ps^+} = w_{++}^{CL} + w_{++}^{RF} \quad (33)$$

It should be noted that the potential w_{++} consists of a sum of two terms, w_{++}^{CL} representing the direct charge–charge Coulomb interaction, plus w_{++}^{RF} representing the solvent reaction field effect, as given by the interaction with the polarization charge density on the protein surface. The term w_{+0} represents a correction to the interaction between ionized group i and group j due to atomic charges in the neutral state⁸ and is given by

$$w_{+0} = (q_i^+ \cdot q_j^0) + 2[q_i^+ \cdot \sigma_j^0]_{ps} + (q_i^0 \cdot q_j^+) + 2[q_j^+ \cdot \sigma_i^0]_{ps} - (q_i^0 \cdot \sigma_j^0) - 2[q_i^0 \cdot \sigma_j^0]_{ps} \quad (34)$$

The 1:1 Salt Effect in the FAMBE Method: The FAMBE method solves the Poisson equation by considering the solvent polarization under zero salt conditions. A rigorous generalization of the boundary element method to an electrolyte solutions is considered in a number of papers.^{33,34} The main results of such a generalization can be summarized as follows: (i) the conversion of the linear Poisson–Boltzmann equation into the boundary element method gives rise to two coupled integral equations and becomes considerably more complicated than that given by eq 21; (ii) at physiological concentrations of a 1:1 salt, that is, ~ 0.1 – 0.2 M, the free-energy contribution to the total free energy due to the mobile salt ions is small, about 1–2% of the value of the polarization energy of the water solvent in the linear Poisson–Boltzmann method;^{33,34} and (iii) the major salt effect on the pair of electrostatic interactions is the Debye–Hückel screening of the electrostatic interactions between charged sites.^{35,36}

The 1:1 salt effect is included, indirectly, in the FAMBE-pH method, as was done for the salt-dependent generalized Born method.²² The correction consists of two terms; the first one is a correction of the free energy $\Delta G_i^{\text{salt}}(\mathbf{x})$ due to interactions of the ionizable group i with the mobile salt ions and is given by

$$\Delta G_i^{\text{salt}}(\mathbf{x}) = - \frac{q_i^2 \kappa_D}{2(1 + \kappa_D R_{Q_i}) D_0} \quad (35)$$

where the Debye screening constant (inverse length) is given by $\kappa_D = 8\pi I / D_0 kT$, with I being the ionic strength, k the Boltzmann constant, T the temperature in K, and D_0 the bulk dielectric constant; R_{Q_i} is the effective radius of the charged group i due to the protein structure and includes the radius of the Stern layer, as explained below in eq 37. The radius R_{Q_i} is defined in the framework of the generalized Born (GB) method^{37–39} by calculation of the solvent polarization free energy of the ionized group as

$$R_{B_i} = \frac{q_i^2}{2G_{\text{pol},i}(\mathbf{x})} \left(\frac{1}{D_0} - \frac{1}{D_1} \right) \quad (36)$$

where $G_{\text{pol},i}(\mathbf{x})$ is the solvent polarization free energy of the ionizable group i , for the protein in conformation \mathbf{x} , in the FAMBE method. The effective Born radius, R_{B_i} , should be increased by the radius of the hydrated salt ion R_{ion} as

$$R_{Q_i} = R_{B_i} + R_{\text{ion}} \quad (37)$$

where the value of R_{ion} represents the Stern radius ($\sim 2 \text{ \AA}$).²⁸ The contribution of the salt ions to the pK shift is given by

$$\Delta \text{p}K_i^{\text{salt}} = - \frac{\gamma_i}{kT(\ln 10)} \left(\Delta G_{\text{PS}_i^+}^{\text{salt}} - \Delta G_{\text{S}_i^+}^{\text{salt}} \right) \quad (38)$$

This term should be added to $\Delta \text{p}K_i^{(1)}$ of eq 28. This contribution is usually low (~ 0.01 pK unit) because of the small magnitude of the free energy and the mutual cancelation of two terms in eq 38, that is, one of them corresponding to the energy for the isolated site S_i^+ and the other one as the energy corresponding to this site in the protein, PS_i^+ . Another contribution from the 1:1 salt correction is related to the electrostatic potential of mean force, Δw_{ij} , between two ionized sites i and j , which is known³⁵ to be modified by the Debye–Hückel screening factor as

$$\Delta w_{ij}^{\text{salt}}(r) = \Delta w_{ij}(r) \frac{\exp(-\kappa_D(r-b))}{(1 + \kappa_D b)} \quad (39)$$

where $b = R_{Q_i} + R_{Q_j}$ is an effective distance of minimal approach. This correction is valid for the linear regime of screening, when the Poisson–Boltzmann equation can be linearized under

the assumption that the magnitude of the electrostatic potential at the ion-accessible surface of the ionized protein in water is about kT .^{35,36} At physiological pH, most proteins are in the vicinity of their isoelectric pH with a total charge of only a few units, and hence, they are expected to have a moderate electrostatic surface potential. At low or large values of pH, compared to the isoelectric pH, proteins can have large positive or negative total charge and possess a large electrostatic surface potential, that is, significantly greater than kT . Under these extreme pH conditions, an accurate treatment of the protein electrostatic energy is challenged since the Poisson–Boltzmann equation for mobile ions is no longer valid because it does not take into account the ion–ion correlations. Some discussion of this problem is given in ^{ref 40}, but analysis of these limits deserves more research and is beyond the goal of the current work.

Effective Dielectric Constant for the Protein: To compare computed pH-dependent effects of a protein with experimental data, the thermal fluctuations and the response of the protein–solvent system to the protein ionization must be taken into account, that is, the protein dynamics cannot be ignored. Thus, the dynamic response of the polar groups to the ionization of one of the groups in the sequence can be treated, equivalently, as the polarization response of a polar solvent to the charging of a solute atom.⁴¹ In this respect, and in the context of the dielectric continuum model, the protein polarization response can be modeled by assigning an appropriate dielectric constant. For example, according to explicit simulation of the dielectric properties of a protein–water system, proper consideration of the polarization response can be obtained by assigning a dielectric constant of $D_1 \sim 15$ and $D_0 \sim 50$ – 100 for the protein and water surrounding the protein, respectively.⁴¹ For this reason, in the FAMBE method, the pH-dependent properties are computed for a given fixed conformation \mathbf{x} , although the protein dynamic polarization response is taken into account implicitly by assigning a dielectric constant of $D_1 = 16$ to the volume occupied by the protein and $D_0 = 80$ for the solvent-excluded volume outside of the protein.

Results and Discussion

Optimal Set of Dielectric Interface Atomic Radii R_{B_i}

The optimal set of atomic radii R_{B_i} for calculating the dielectric surface was determined by fitting the polarization free energy of the FAMBE method, as given by eq 35 of ^{ref 26}, to the “experimental” polarization free energies obtained for a set of terminally blocked amino acid residues.⁴² The “experimental” polarization free energies were simulated by a slow charging process^{30,43,44} with the molecular dynamics program SigmaX,⁴⁵ in explicit SPC water,⁴⁶ with periodic boundary conditions, and with the particle-mesh Ewald (PME)⁴⁷ approximation to treat long-range electrostatic interactions. On the basis of the hypothesis that the optimal radius R_{B_i} should be independent of both the charge distribution and the molecular conformation, the calculation of the solvent polarization free energy shown in Table 1 was carried out for each listed amino acid residue X in the sequence Ac-X-NMe in an extended conformation with the charges taken from the SigmaX⁴⁵ force-field. Fifteen out of the 30 groups in Table 1 were taken as the training (fitting) set for the calculation of the atomic radii R_{B_i} to be used by the FAMBE method. During the fitting procedure, the dielectric constant of the solute cavity, defined by the FAMBE method, was set equal to $D_1 = 1.0$, to be consistent with the dielectric constant for a fixed protein conformation in a molecular dynamics simulation of the charging process in explicit water solvent, while the value of the solvent dielectric constant was set to $D_0 = 80.0$. The calculated atomic radii R_{B_i} are listed in Table 2, together with the radii obtained by the PARSE⁴⁸ method shown for comparison. The overall error of the FAMBE method (compared to the slow-charging method) in reproducing the polarization free energies of terminally blocked amino acids (given in Table 1) is equal to 1.6% for the training set (indicated by asterisks in Table 1) and about 2% for the set of 13 groups in Table 1 not included in the fitting procedure.

Internal Self-Consistency Test of the FAMBE Method

The asymptotical form for the total electrostatic potential $\varphi(\mathbf{r})$ at a large distance from a protein is equal to the sum of the classical multipole terms due to the protein charges q_i at the points \mathbf{r}_i scaled by a factor $1/D_0$. This condition imposes limits on the multipole moments of induced charge density distribution $\sigma(\mathbf{t})$ over the dielectric interface surface. Thus, a test (called a Q test) for computing the total charge requires that $Q_\sigma \equiv Q_q$, defined in eq 40

$$Q_\sigma \equiv \int_s \sigma(\mathbf{t}) d\mathbf{t}$$

$$Q_q \equiv \left(\frac{1}{D_0} - \frac{1}{D_1} \right) \sum_i q_i \quad (40)$$

In the FAMBE method, $\sigma(\mathbf{t})$ satisfies the Q test exactly²⁶ by using a sum rule normalization and smoothing of the matrix elements of \mathbf{M}_i of eq 22 to increase the accuracy and numerical stability of the solution. Moreover, the surface-induced dipole moment \mathbf{M}_σ should be proportional to the dipole moment of the protein for the set of atomic charges. Then, the M test requires that $\mathbf{M}_\zeta \equiv \mathbf{M}_q$, defined in eq 41

$$\mathbf{M}_\sigma \equiv \int_s \sigma(\mathbf{t}) (\mathbf{t} - \mathbf{r}_0) d\mathbf{t}$$

$$\mathbf{M}_q \equiv \left(\frac{1}{D_0} - \frac{1}{D_1} \right) \sum_i q_i (\mathbf{r}_i - \mathbf{r}_0) \quad (41)$$

By contrast with eq 40, eq 41 depends on the numerical accuracy of the method. Table 3 shows the results for the self-consistency test for the calculations of the total charge Q and the induced dipole moments \mathbf{M} for a set of five proteins. It can be seen from Table 3 that the FAMBE method exactly satisfies the Q test given by eq 40. Moreover, the average error for the M test, given by eq 41, namely, the average ratio $|(\mathbf{M}_q - \mathbf{M}_\sigma)|/|\mathbf{M}_q|$ (given in the last column of Table 3), does not exceed 2% for the optimal parameters given in Table 4, which provides a compromise between speed, scalability, and accuracy of the solution.

Table 4 shows the set of optimal parameters of the FAMBE adaptive tessellation method for four proteins containing 58–448 residues. The dimensions N_{MBE} of the matrices \mathbf{M}_i of eq 22 for the adaptive tessellation are low, in the range of 346–1502, while the average size N_{MBE} of the matrix \mathbf{M}_i is about 744–869 for proteins of dimensions of 33–70 Å, as shown in Table 4. Thus, the adaptive tessellation by the multisized boundary elements considerably reduces the dimensions N_{MBE} of the matrices \mathbf{M}_i of eq 22, compared to the dimensions for the uniform and three-level tessellation method of ref²⁶, and greatly increases the speed for solving the primary boundary element eq 20 within an error of ~1–2%.

Calculations of pK_a Shifts of Proteins

Three proteins, BPTI (PDB code 4PTI), hen egg white lysozyme HEWL (PDB code 2LZT), and bovine pancreatic ribonuclease RNaseA (PDB code 3RN3), for which experimental data are available, have been used as a test set for the calculation of pK_a 's in several published works.^{9,12,49–52} Our results for this set of proteins, using the new FAMBE-pH method, follows.

The BPTI Protein—The structure of BPTI was used here to compute the pK_a 's of the ionizable groups in the multiple-site titration equilibrium to determine the optimal value of the solvent-excluded cavity dielectric constant D_1 of the protein. This optimal value was estimated as the value for which the average absolute deviation and the maximum absolute error in the pK_a 's each reach minima. A detailed description of the calculation of the $\Delta pK_i^{(1)}$ shift, as a

function of D_I using the FAMBE-pH method for each of the 20 ionizable residues of this protein is shown in Table 5. From Table 5, it can be seen that the major factors affecting the $\Delta pK_i^{(1)}$ shift are the free energy of desolvation $\Delta g_{p^+}^{ds}$ of the ionized site PS_i^+ (the first two terms of eq 29), the reaction-field free energy $\Delta g_{p^+}^{RF}$ (the third and fourth term of eq 29), the desolvation free energy of the respective neutral site PS_i , $\Delta g_{p_0}^{ds}$ (the two first terms in eq 30), and the reaction-field free energy $\Delta g_{p_0}^{RF}$ (the third and fourth term of eq 30). The term Δg_{pp} (given by eq 31) represents a negligible contribution, and it is omitted in Table 5. It should be noted that the desolvation penalties $\Delta g_{p^+}^{ds}$ have large magnitudes for buried residues such as GLU7, TYR23, and TYR35. As a consequence, the $\Delta pK_i^{(1)}$ shifts for these buried residues are large (see Table 5); the desolvation free energies of the buried groups decrease significantly with increasing dielectric constant D_I , as shown in Table 5 for $D_I = 2$ and 4. The term $\Delta g_{p_0}^{RF}$ describes the interaction between the nonionized neutral group and the protein and indicates that such interactions, for example, involving ARG1, GLU7, and so forth, are favorable.

Results for the calculations of the potential of mean force, Δw_{ij} , eq 32, are shown for BPTI in Table 6, in which it can be seen that the value of the PMF, w_{++} , is a result of a mutual cancelation of two terms: the direct Coulombic interaction, terms Δw_{++}^{CL} , and the solvent reaction field term Δw_{++}^{RF} , eq 33. The PMF term w_{+0} describes the correction to the electrostatic interaction between groups i and j and has a small value for the majority of interacting sites. The large value of the PMF term w_{+0} for some pairs reflects strong interaction between those pairs, for example, hydrogen bonds for the pair ASP50-ARG53, in Table 6. The values of the PMF Δw_{ij} are less sensitive than the values of $\Delta pK_i^{(1)}$ to the value of the protein dielectric constant D_I , as can be seen from Tables 5 and 6.

For a pair of ionizable groups, the ratio of the PMF Δw_{ij} to the value of the direct Coulombic electrostatic energy, $u^{CL} = q_i q_j / D_0 r_{ij}$, is shown in Figure 2. Here, the distance r_{ij} is taken as the distance between the centers of charges of groups i and j . As seen from Figure 2, for pairs of residues at distances of $r_{ij} > 20$ Å, there is a significant deviation (up to 100%) with respect to the ratio 1.0. On the other hand, at large distances, $r_{ij} < 25$ Å, the PMF can be approximated by the Coulomb energies for the interaction of the ionized groups in water.

It should be noted that, in the range of $10 > r_{ij} > 20$ Å, the PMF values for some charged surface groups i, j are over-screened. In other words, they have an effective dielectric constant larger than the water solvent dielectric constant D_0 . A similar effect was found for the PMF between charged LYS⁺-LYS⁺ pairs at positions (1,3), (1,6), and (1,7) of α -helical polylysine, which interact through the low dielectric protein cavity, calculated on the basis of a finite difference solution of the Poisson equation.^{35,36} It should be noted that the generalized Born (GB) approximation³⁷⁻³⁹ cannot reproduce such an overscreening effect between two charges, as shown in ref³⁸ because the GB approximation is a smooth interpolation of the effective dielectric constant between two limits, that is, D_i and D_0 .

Table 7 shows the results of calculations of the ionization constants pK_a for the ionizable residues of BPTI for different values of the dielectric constant D_I of the protein with and without 1:1 salt effects. From Table 7, it can be seen that the calculated pK_a^{cal} 's are sensitive to the value of the protein dielectric constant D_I . The 1:1 salt corrections, from eqs 35, 38, and 39 [column (pK_a^{cal})^e in Table 7] show improvement of the agreement between the calculated pK_a^{cal} values and the experimental, pK_a^{exp} , as indicated by both the average absolute deviation, $\Delta_{av} = 0.3$, and the maximum absolute error in the pK_a 's, $\Delta_{max} = 0.8$, with respect to the same values calculated without salt [see Table 7, (pK_a^{cal})^d columns]. For BPTI, better agreement is obtained between pK_a^{cal} and pK_a^{exp} for $D_I = 16$. This value is close to the values of the protein

dielectric constant $D_p \sim 15$ found by direct simulations of the dielectric response of the protein molecule in water solution.⁴¹ The necessity to use a high value for the protein dielectric constant, namely, ~ 20 , to obtain results for pK_a which are comparable to experimental data has been pointed out in a number of publications.^{9,12,49–51} The results of the calculations presented here for the pK_a 's, in Table 7, are in a good agreement with experimental data, as indicated by the average absolute deviation, $\Delta_{av} = 0.3$ and the maximum absolute error in pK_a 's, $\Delta_{max} = 0.8$ pK units. In particular, this accuracy is better than the one reported by Antonisiewicz et al.⁵⁰ for the same system, namely, BPTI, as shown in Table 7, footnote f. Their results were obtained by using the standard two-dielectric continuum model and the finite difference method for solution of the linear Poisson–Boltzmann equation for electrostatic calculations in a water solution of a 1:1 salt and a detailed model for atomic charges of ionized and neutral groups. The method of Demchuck and Wade⁵¹ shows an accuracy similar to that of the present work. However, the Demchuck and Wade method assigns a different dielectric constant for different ionizable groups of the protein in order to improve the accuracy. On the other hand, the recent empirical method of Li et al.⁵² shows large errors in the predictions of the pK_a 's for BPTI.

The optimal value of 16 for D_I , obtained for BPTI, has been adopted for the remaining tests on HEWL and RNase A.

Proteins HEWL and RNase A

HEWL—The calculated ionization constants pK_a of lysozyme (HEWL) and ribonuclease (RNase A) are shown in Tables 8 and 9, respectively. It can be seen in Table 8 that the average absolute deviation from the experimental pK_a 's, Δ_{av} , is equal to 0.5 pK units for HEWL and that the maximum deviation, Δ_{max} , is equal to 1.2 pK units. The achieved accuracy is better than those of other methods for the same protein^{12,50–52} (see (Δ_{av}) and (Δ_{max}) in Table 8). Six residues, namely, GLU7, ASP18, GLU35, TYR53, ASP66, and ASP119, have the largest shift for pK_a^{exp} with respect to the corresponding pK_a^0 , namely, more than 1.5 pK units. For these six residues (indicated with the superscript j in Table 8), the average absolute deviation (Δ_{av}) of the calculated values, pK_a^{cal} , from the experimental data, pK_a^{exp} , is 0.7 pK units.

RNase A—The results for RNase A (PDB code 3RN3), shown in Table 9, were obtained by considering both the SO_4^{2-} ion and two alternative, observed side-chain positions of HIS119, namely, positions A and B, shown in Figure 3. Positions A and B of HIS119, as computed from their PDB coordinates, have different torsion angles χ^1 , namely, 158.5 and -72.6° , respectively. In position A, HIS119 is involved in a hydrogen bond with ASP121, and the distance between atoms NE2 of HIS119 and OD1 of ASP121 is equal to 2.59 Å. On the other hand, in position B, HIS119 is exposed to the solvent, and the hydrogen bond between NE2 of HIS119 and OD1 of ASP121 is broken. In both positions, HIS119 interacts electrostatically with the SO_4^{2-} ion. In the structures reported in the PDB, the SO_4^{2-} ion was bound to the protein surface at 57% occupancy in both positions A and B. Both conformations, namely, with HIS119 in positions A or B, were used to compute the pK_a shifts. Without considering HIS119 in position B, the computed pK_a^{cal} values are significantly less accurate, as shown in column $(pK_a^{cal})^e$ ($\Delta_{av} = 0.7$, $\Delta_{max} = 2.4$). In fact, the best average deviation computed with HIS119 in position B gave a Δ_{av} equal to 0.4 and $\Delta_{max} = 1.0$ pK unit, as shown in column $(pK_a^{cal})^f$ of Table 9. The results presented in columns $(pK_a^{cal})^e$, $(pK_a^{cal})^f$, and $(pK_a^{cal})^g$ in Table 9 show the importance of taking into account the counterion SO_4^{2-} and the conformational mobility of the protein structure, that is, through multiple positions of HIS119, for the calculation of the pK's. Taking HIS119 in the B position, the hydrogen bond between the surface residues HIS119 and ASP121 is disrupted. This should not be unexpected since, in water, this hydrogen bond is probably unstable primarily because of exchange with neighboring water molecules. Moreover, a small conformational move of the side chain of

ARG10 also disturbs the hydrogen bond between GLU2 and ARG10. Breaking the hydrogen bond between GLU2 and ARG10 induces changes of the pK_a 's of the partners, as shown by comparing columns $(pK_a^{\text{cal}})^e$ and $(pK_a^{\text{cal}})^f$ of Table 9 for GLU2 and ARG10 and also for HIS119 and ASP121; for example, the pK_a of GLU2 is shifted down, and the pK_a of ARG10 is shifted up by formation of the hydrogen bond. In general, such shifts upon formation of a hydrogen bond were predicted in *ref*². Such up/down shifts are as large as ~ 2 pK units. The SO_4^{2-} ion decreases the pK_a 's of the neighboring residue HIS12, at the distance of ~ 4.5 Å, by about 1.2 pK units and affects other more distant residues (such as GLU86, HIS119, and ASP121) in the range of 0.2–0.4 pK units. It should be noted that recent methods,^{20–22} which combine molecular dynamics with ionization equilibrium calculations, do not lead to better accuracy for predicted pK_a 's for RNase A; the values reported²² for Δ_{av} and Δ_{max} are equal to 1.0 and 2.1 pK units, respectively.

Calculation of the pH-Dependent Ionization Free Energy

The pH-dependent ionization free energy $\Delta G_{\text{inz}}^{\text{PS}+}(x, \text{pH})$ was calculated from a direct summation over all ionization states, as given by the partition function in eq 9, only for protein BPTI, while the Tanford–Schellman integration method, eqs 12 and 16, was used for BPTI, HEWL, and RNase A. The partition function method cannot be used for HEWL and RNase A because these proteins contain more than 25 ionizable groups, as shown in Tables 7–9. The calculated ionization free energies are shown in Figure 4. A comparison of the calculations of

$\Delta G_{\text{inz}}^{\text{PS}+}(x, \text{pH})$ by different methods enabled us to test the self-consistency and accuracy of the Tanford–Schellman method. From Figure 4, for example, it can be seen that, for BPTI, there is good agreement between the exact computation of the free energy, that is, calculated from the partition function, and the results of the Tanford–Schellman integral method by using both intervals of integration, namely, $(-\infty, \text{pH})$ and $(+\infty, \text{pH})$. The pH-dependent free energies of ionization of BPTI, calculated by the back and forth integrals $(+\infty, \text{pH})$ and $(-\infty, \text{pH})$, respectively, differ within only 0.2 kcal/mol and are equal to the exact free energy, calculated from the partition function, within a difference of only 1.2 kcal/mol. The differences between the back and forth ionization free energy for lysozyme and RNase are within 0.4 kcal/mol. From this analysis, we conclude that the implemented Monte Carlo method of calculation of the free energy of ionization by the Tanford–Schellman integral gives reasonable accuracy for pH-dependent effects in proteins. The calculations of the average ionization degrees

$\langle z_i(\mathbf{x}, \text{pH}) \rangle$ of residues (by eq 11) were carried out with a Monte Carlo (MC) random walk in the ionization microstate space, that is, by partitioning the pH range of integration given by eqs 12 and 16, for example, between -25 and 25 pH units, in small increments, that is, $\Delta \text{pH} = 0.25$. Since the ionization microstates at extremely large and low pH values are known, the MC titration method coupled with small pH increments constitutes an accurate approximation since it enables us to explore the ionization microstates more efficiently.

The relative stability of the pH-dependent ionization free energy, $\Delta \Delta G_{\text{inz}}(\mathbf{x}, \text{pH})$, shown in Figure 5, was computed by using eq 13, which describes the ionization free energy of the folded protein conformation $\Delta G_{\text{inz}}^{\text{PS}+}(x, \text{pH})$ with respect to the unfolded reference state $\Delta G_{\text{inz}}^{\text{S}+}(x, \text{pH})$, in which all ionizable groups are well exposed to solvent and do not have electrostatic interactions between them. It can be seen from Figure 5 that the calculated ionization free energy stabilizes the respective native structures [$\Delta \Delta G_{\text{inz}}(\mathbf{x}, \text{pH}) > 0$] in the pH range of ~ 4 to ~ 10 for BPTI, ~ 5 to ~ 7.5 for RNase A, and ~ 7 to ~ 9.5 for HEWL. For the particular case of HEWL, the calculated stabilization $\Delta \Delta G_{\text{inz}}(\mathbf{x}, \text{pH})$ in this pH range is in good agreement, that is, within ~ 2 kcal/mol, with the experimental free energy of stabilization in the pH range of ~ 4 to ~ 6 . Although the calculated pH of maximum stability ($\text{pH}_{\text{max}}^{\text{calc}} = 8.5$) is shifted with respect to the experimental one ($\text{pH}_{\text{max}}^{\text{exp}} = 4.5$),⁵³ the pH range for which the structure is stable is in

qualitative agreement, for example, $\text{pH}_{\text{stab_interval}}^{\text{calc}} = 6.0 - 10.0$ and $\text{pH}_{\text{stab_interval}}^{\text{exp}} = 4.0 - 6.0$, between the calculated and experimental free energy of stabilization (see Figure 5). It is worth noting that the experimental pH-dependent relative stability $\Delta\Delta G_{\text{inz}}^{\text{exp}}(x, \text{pH})$ for HEWL was calculated with the Tanford–Schellman integral method from the experimental average ionization charge⁵³ $Q_{\text{inz}}^{\text{exp}}(\text{pH})$ at 0.1 M salt concentration and 298 K, which is shown in Figure 6. In this figure, the calculated total charge as a function of the pH, $Q_{\text{inz}}^{\text{calc}}(\text{pH})$, is also shown. The results indicate that, in the pH range of 4–10, in which there is an ensemble of different ionization states, the calculated charge $Q_{\text{inz}}^{\text{calc}}(\text{pH})$ (shown with small black-filled squares) is lower than the experimental charge $Q_{\text{inz}}^{\text{exp}}(\text{pH})$ (shown with large black-filled diamonds), which are both in the charge range of +6 to +10 units. Conceivably, a possible explanation of the systematic low charge for $Q_{\text{inz}}^{\text{calc}}(\text{pH})$ in this pH range is due to a nonlinear effect or specific counterion effects, which are not included in the current method.

Conclusion

The FAMBE-pH method, described for the calculation of the pH-dependent properties of proteins, is based on the dielectric continuum model of a protein in a water solvent and on the efficient boundary element method FAMBE to solve the Poisson equation taking into account 1:1 salt effects. The FAMBE-pH method shows internal self-consistency and is faster and more accurate to compute pH-dependent properties in proteins than existing methods.^{6–9,12,22,25,49–54} Moreover, it can also be implemented together with any existing force field to study a variety of applications such as protein stability, protein folding, and protein–ligand binding. However, it should be noted that, despite the good accuracy for the prediction of pK_a 's for ionizable residues, that is, with an average error of ~ 0.5 pH unit for lysozyme (as shown in Table 8), such accuracy is not good enough for a precise prediction of the pH value of maximum stability in a protein such as lysozyme (as shown in Figure 5). In addition, the importance of considering the counterion effect was shown in Table 9 for RNase A. Here, it was shown that counterion effects become a significant contribution if the pH is far from the isoelectric point, and such effects certainly cannot be ignored for an accurate treatment of pH-dependent protein stability. We have also been able to show (Table 9) that, taking the conformational mobility of the protein side chains into account, for example, by considering small changes of $\sim 2\text{--}3$ Å, could lead to significant changes in the hydrogen bond network. Such changes in the hydrogen bond network might be very important for calculating pH-dependent phenomena, as was shown for RNase A.

The solvent polarization free energy of the protein 1FCS (fasciculin II, consisting of 153 residues) was computed by Lu and McCammon⁵⁵ in ~ 20 s on a 2 GHz Pentium IV computer by using an improved boundary element method for solving the Poisson–Boltzmann equation at 0.0 M salt concentration. We subsequently computed the solvent polarization free energy of this protein on a 2 GHz Pentium IV computer by using the FAMBE method in a comparable time. Although we cannot compare the total free energy from both laboratories because we both used different sets of charges and atomic radii, our FAMBE-pH method presented here is not limited to computing only the solvent polarization free energy but can also compute the total free energy of solvation. In addition, the FAMBE-pH method provides a very efficient and accurate computational treatment of the multiple-site titration problem for proteins containing a large number of ionizable groups, for example, 150 for a 448 residue protein 1EA1 (Table 4). As noted in the subsection “Calculation of pK_a Shifts of Proteins”, our FAMBE-pH method shows internal self-consistency in the computation of the free energy of ionization and accurate prediction of the pK_a shifts in three proteins, namely, BPTI, HEWL, and RNase A. Moreover, very good agreement was found for the computation of the free energy of ionization

by using both the exact calculation of the partition function and the Tanford–Schellman integral method.

All of the accumulated results indicate that the FAMBE-pH method presented here could lead to a more realistic treatment of pH-dependent phenomena, when used together with molecular dynamics or Monte Carlo conformational sampling.

Acknowledgments

This work is supported by the Russian Academy of Sciences and the Russian Fund for Basic Research, Grant 05-04-48322 (Y.V.) and by grants from the National Science Foundation (MCB05-41633) and the National Institutes of Health (GM-14312 and GM-24893). Support was also received from the National Research Council of Argentina (CONICET), FONCyT-ANPCyT (PAE 22642/22672), and from the Universidad Nacional de San Luis [UNSL] (P-328501), Argentina. The authors are grateful to J. Hermans for providing the Sigma program.

References and Notes

1. Rost B, Sander C. *J Mol Biol* 1993;232:584. [PubMed: 8345525]
2. Laskowski M, Scheraga HA. *J Am Chem Soc* 1954;76:6305.
3. Hill TL. *J Am Chem Soc* 1956;78:5527.
4. Tanford C, Kirkwood JG. *J Am Chem Soc* 1957;79:5333.
5. Warshel A. *Biochemistry* 1981;20:3167. [PubMed: 7248277]
6. Bashford D, Karplus M. *Biochemistry* 1990;29:10219. [PubMed: 2271649]
7. Beroza P, Fredkin DR, Okamura MY, Feher G. *Proc Natl Acad Sci U S A* 1991;88:5804. [PubMed: 2062860]
8. Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B. *Proteins: Struct, Funct, Genet* 1993;15:252. [PubMed: 7681210]
9. Antosiewicz J, McCammon JA, Gilson MK. *J Mol Biol* 1994;238:415. [PubMed: 8176733]
10. Ripoll DR, Vorobjev YN, Liwo A, Vila JA, Scheraga HA. *J Mol Biol* 1996;264:770. [PubMed: 8980685]
11. Zhou HX, Vijayakumar M. *J Mol Biol* 1997;267:1002. [PubMed: 9135126]
12. Alexov EG, Gunner MR. *Biophys J* 1997;72:2075. [PubMed: 9129810]
13. Ripoll DR, Vila JA, Scheraga HA. *J Mol Biol* 2004;339:915. [PubMed: 15165859]
14. Vila JA, Ripoll DR, Arnautova YA, Vorobjev YN, Scheraga HA. *Proteins: Struct, Funct, Bioinf* 2005;61:56.
15. Baptista AM, Teixeira VH, Soares CM. *J Chem Phys* 2002;117:41840.
16. Bürgi R, Kollman PA, VanGunsteren VF. *Proteins: Struct, Funct, Bioinf* 2002;47:469.
17. Dlugosz M, Antosiewicz JM. *Chem Phys* 2004;302:161.
18. Mongan J, Case DA, McCammon JA. *J Comput Chem* 2004;25:2038. [PubMed: 15481090]
19. Leezfn MS, Salsbury FR Jr, Brooks CL III. *Proteins: Struct, Funct, Bioinf* 2004;56:738.
20. Khandogin J, Brooks CL III. *Biophys J* 2005;89:141. [PubMed: 15863480]
21. Khandogin J, Chen J, Brooks CL III. *Proc Natl Acad Sci U S A* 2006;103:18546. [PubMed: 17116871]
22. Khandogin J, Brooks CL III. *Biochemistry* 2006;45:9363. [PubMed: 16878971]
23. Tanford C. *Adv Protein Chem* 1970;24:1. [PubMed: 4912353]
24. Schellman JA. *Biopolymers* 1975;14:999.
25. Yang SA, Honig B. *J Mol Biol* 1993;231:459. [PubMed: 8510157]
26. Vorobjev YN, Scheraga HA. *J Comput Chem* 1997;18:569.
27. Vorobjev YN, Almagro JC, Hermans J. *Proteins: Struct, Funct, Bioinf* 1998;32:399.
28. Sharp KA, Honig B. *Annu Rev Biophys Chem* 1990;19:301.
29. Vorobjev YN, Hermans J. *Biophys J* 1997;73:722. [PubMed: 9251789]
30. Vorobjev YN, Hermans J. *Biophys Chem* 1999;78:195. [PubMed: 10343388]
31. Vorobjev YN, Hermans J. *Protein Sci* 2001;10:2498. [PubMed: 11714917]

32. Landau, LD.; Lifshitz, EM. *Electrodynamics of Continuous Media Landau and Lifshitz Course of Theoretical Physics Volume 8 Translated from the Russian*. Pergamon Press; Oxford, U.K.: 1984. p. 34-36.p. 39
33. Yoon BJ, Lenhoff AM. *J Comput Chem* 1990;11:1080.
34. Juffer AH, Botta EFF, Bert AM, van Keulen BAM, van der Ploeg A, Berendsen HJC. *J Comp Phys* 1991;97:144.
35. Vorobjev YN, Scheraga HA, Hitz B, Honig B. *J Phys Chem* 1994;98:10940.
36. Vorobjev YN, Scheraga HA, Honig B. *J Phys Chem* 1995;99:7180.
37. Qiu D, Shenkin PS, Hollinger FP, Still WC. *J Phys Chem A* 1997;101:3005.
38. Sigalov G, Scheffel P, Onufriev A. *J Chem Phys* 2005;122:094511. [PubMed: 15836154]
39. Onufriev A, Case DA, Bashford D. *J Comput Chem* 2002;23:1297. [PubMed: 12214312]
40. Scheraga HA, Katchalsky A, Alterman Z. *J Am Chem Soc* 1969;91:7242.
41. Loffler G, Schreiber H, Steinhauser O. *J Mol Biol* 1997;270:520. [PubMed: 9237916]
42. Nina M, Beglov D, Roux B. *J Phys Chem B* 1997;101:5239.
43. Hummer G, Pratt LR, Garcia AE. *J Phys Chem* 1995;99:14188.
44. Aqvist J, Hansson T. *J Phys Chem* 1996;100:9512.
45. Hermans, J. University of North Carolina; 1999. Sigma Documentation. <http://hekto.med.unc.edu/HERMANS/software/SIGMA/index.html>
46. Berendsen, HJC.; Postma, JPM.; van Gunsteren, WF.; Hermans, J. *Jerusalem Symposia on Quantum Chemistry and Biochemistry*. Pullmann, B., editor. D Reidel; Dordrecht, The Netherlands: 1981.
47. Essmann U, Perera L, Berkowitz L, Darden T, Lee H, Pedersen LG. *J Chem Phys* 1995;103:8577.
48. Sitkoff D, Sharp KA, Honig B. *J Phys Chem* 1994;98:1978.
49. Antosiewicz J, McCammon JA, Gilson MK. *Biochemistry* 1996;35:7819. [PubMed: 8672483]
50. Antosiewicz J, Briggs JM, Elcock AH, Gilson MK. *J Comput Chem* 1996;17:1633.
51. Demchuk E, Wade RC. *J Phys Chem* 1996;100:17373.
52. Li H, Robertson AD, Jensen JH. *Proteins: Struct, Funct, Bioinform* 2005;61:704.
53. Tanford C, Roxby R. *Biochemistry* 1972;11:2192. [PubMed: 5027621]
54. Schafer M, Sommer M, Karplus M. *J Phys Chem B* 1997;101:1663.
55. Lu B, McCammon A. *J Chem Theory Comput* 2007;3:1134.

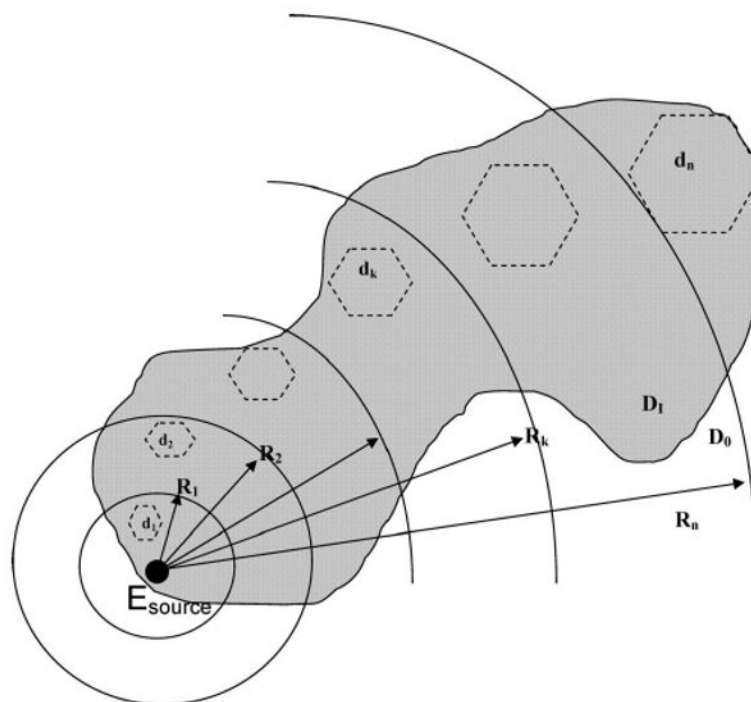


Figure 1. Tessellation of the dielectric interface surface by multisized boundary elements d_i , with $i = 1, 2, \dots, n$, used in the FAMBE-pH method. The large black-filled dot indicates the source E_{source} . The distances, R_i , from the charged group are shown. Values of d_i and R_i are reported in Table 4. The shaded region represents the dielectric interface surface of a protein surrounded by water; D_1 and D_0 are the dielectric constants inside and outside of the surface, respectively.

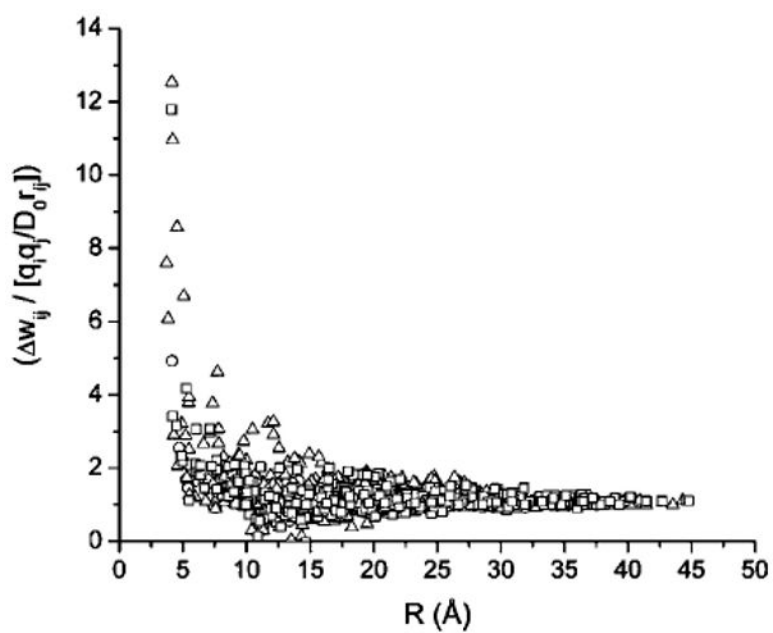


Figure 2. The ratio of the PMF Δw_{ij} to the Coulombic energy ($u^{\text{CL}} = q_i q_j / D_0 r_{ij}$) as a function of the distance between pairs of ionizable residues (i, j) is plotted for BPTI (open circles), HEWL (open squares), and RNase A (open triangles).

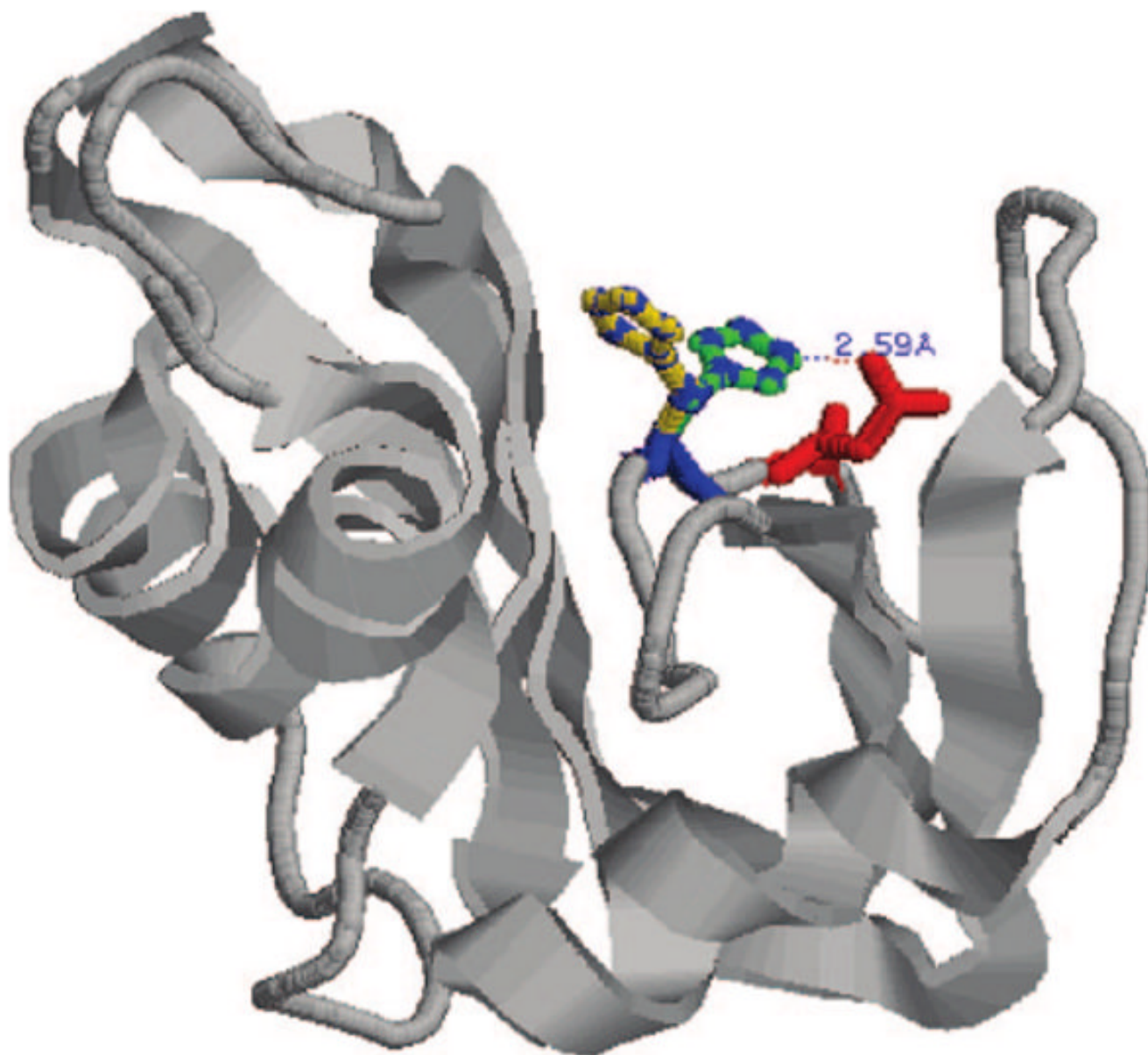


Figure 3. Structure of RNase A (PDB code 3RN3), showing two positions of the side chain of HIS119. In position A (blue), NE2 is hydrogen-bonded to OD1 of ASP121 (red) within 2.59 Å. In the alternative position B, this hydrogen bond does not exist.

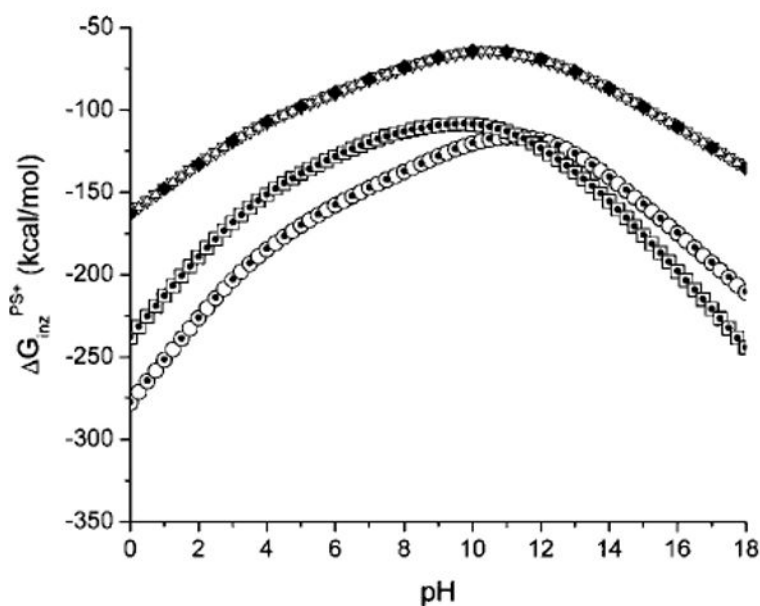


Figure 4.

The free energy of ionization ($\Delta G_{\text{inz}}^{\text{PS}^+}$) as a function of pH is shown for the proteins BPTI, HEWL, and RNase A. The upper curve shows the results obtained for BPTI (the black-filled diamonds represent the values computed by using the exact partition function, while the open up and down triangles are the result of calculations using the Tanford–Schellman integral approach after back and forth integration intervals, respectively). The results obtained for HEWL are shown in the middle curve (the squares, and dots within them, represent the results of the calculations using the Tanford–Schellman integral approach over back and forth integration intervals, respectively). The results obtained for RNase A are shown in the lower curve (the circles, and the dots within them, represent the results of the calculations using the Tanford–Schellman integral approach over back and forth integration intervals, respectively).

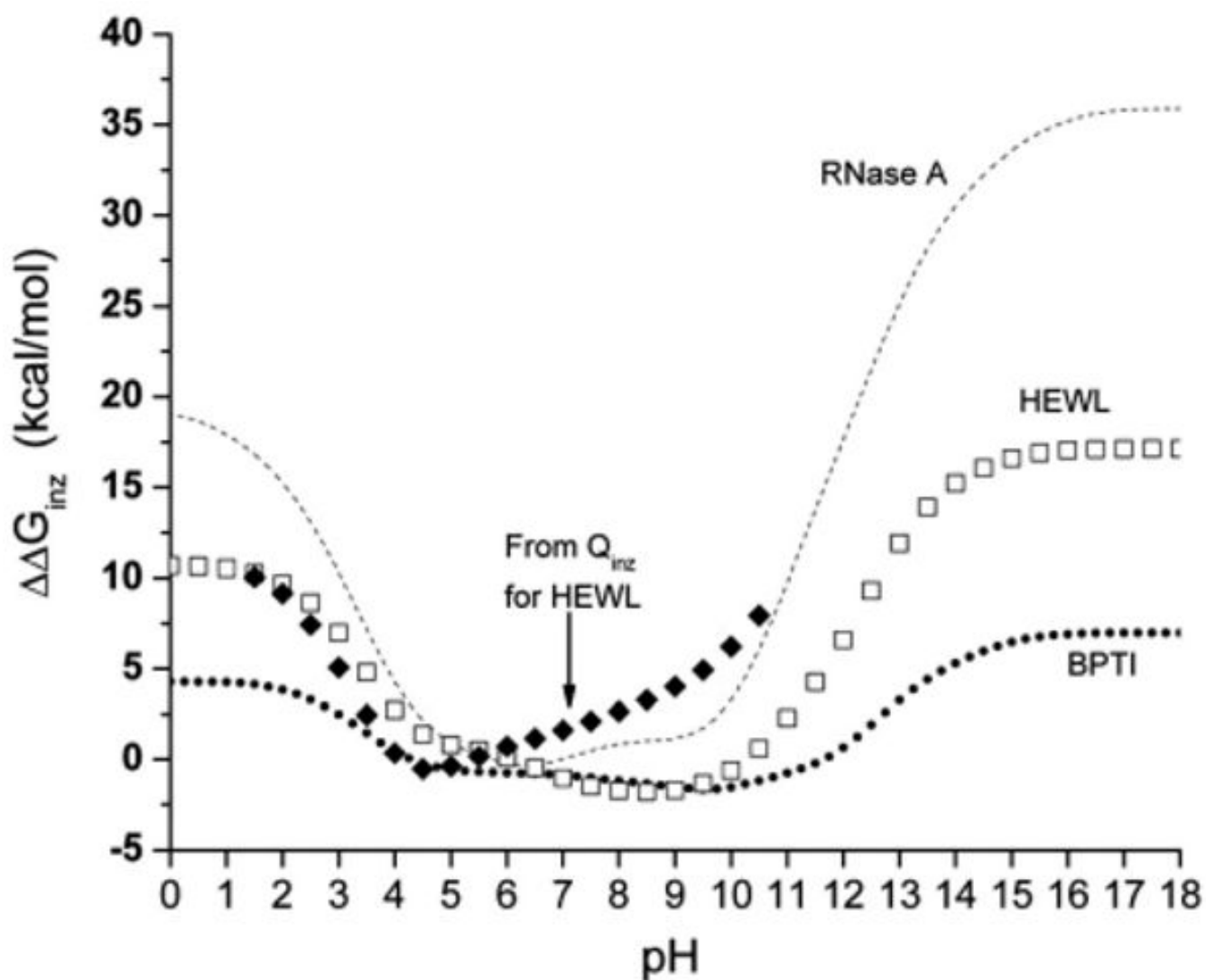


Figure 5. The computed stabilization free energy of ionization ($\Delta\Delta G_{inz}$, eq 13) of the native structure with respect to the non-native (unfolded one), as a function of pH, is shown for three proteins, namely, BPTI, HEWL, and RNase A. For the non-native representation of the protein structure, the average ionization was approximated by ΔG_{inz}^{S+} of eq 14. The dotted line is the result for BPTI, the dashed line is for RNase A, and the open squares are for HEWL. The black-filled diamonds represent the computed $\Delta\Delta G_{inz}$ from the observed titration curve, $Q_{inz}(\text{pH})$ (see Figure 6), only for HEWL.

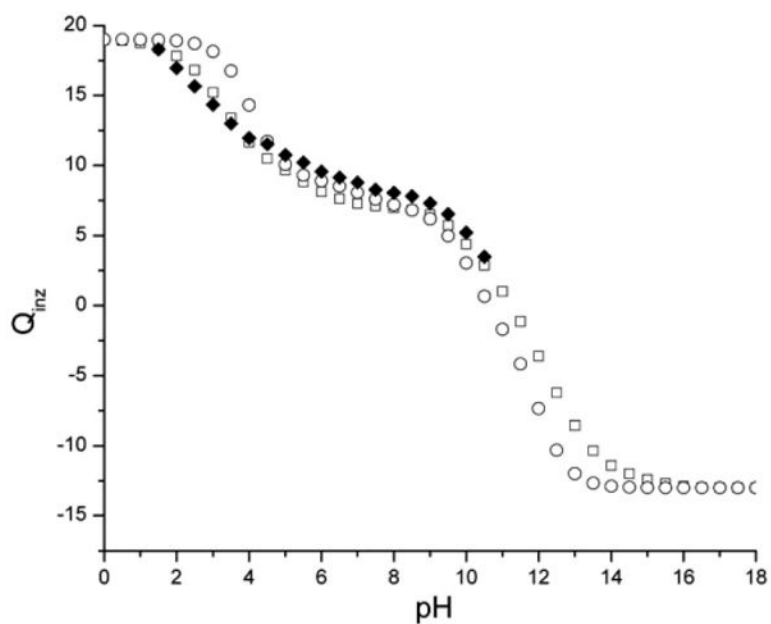


Figure 6.

The pH-dependent total ionization charge [$Q_{inz}^{cal}(pH)$] for HEWL is shown as small black-filled squares for the curve computed by the FAMBE-pH method, the observed [$Q_{inz}^{exp}(pH)$]⁵³ as large black-filled diamonds, and the curve for free groups in solution as open circles.

TABLE 1
Solvent Polarization Free Energy (kcal/mol)^a

group ^b	X	slow-charging method	FAMBE method
AcCO	ALA	-2.98	-2.93
pep ^c CO	ALA	-2.96	-2.95
AcCO-pep ^c CO*	ALA	-5.40	-5.66
pep ^c NH	ALA	-1.73	-1.74
NMeNH	ALA	-1.89	-1.94
H ₂ O ^d		-8.50	-8.46
AcALANMe*	ALA	-10.7	-10.30
pep ^c NHCO*	ALA	-3.65	-3.74
pep ^c NHCO*	GLY	-3.83	-3.90
pep ^c NHCO*	VAL	-3.50	-3.58
pep ^c NHCO*	LEU	-3.74	-3.73
side ^e COH	SER	-8.20	-8.22
side ^e COH*	THR	-7.31	-7.15
pep ^c CONH-side ^e CHOH*	THR	-11.30	-11.58
pep ^c CONH-side ^e CHOH*	SER	-11.60	-12.04
side ^e ASP	ASP	-88.4	-88.3
side ^e GLU ⁻ *	GLU	-96.1	-96.9
side ^e LYS ⁺	LYS	-90.5	-90.6
side ^e ARG ⁺	ARG	-62.1	-63.4
side ^e HIS ⁰	HIS	-10.8	-10.8
side ^e NH ₂	GLN	-6.35	-6.22
side ^e GLN*	GLN	-8.95	-8.59
AcGLNNMe*	GLN	-16.1	-16.1
pep ^c CONH-side ^b	GLN	-10.8	-10.8
AcASNNMe*	ASN	-21.1	-20.6
side ^e ASN	ASN	-8.90	-9.03
side ^e (NH ₂)*	ASN	-5.2	-5.23
pep ^c -side ^e ASN*	ASN	-13.7	-13.45
side ^e PHE	PHE	-0.24	-0.32
side ^e TYR	TYR	-5.75	-5.63

^a Computed for the charged groups in the terminally blocked amino acid X in Ac-X-NMe by slow-charging simulations³¹ and also by using the dielectric model (FAMBE) method as described in the Results and Discussion section in the subsection entitled "Optimal Set of Dielectric Interface Atomic Radii R_{B_i} ". Asterisks denote the groups used for fitting; "pep" indicates backbone groups, and "side" indicates side-chain groups.

^b Atomic group undergoing the slow-charging process. The rest of the atoms of the terminally blocked amino acids have zero partial charges. This set covers all polar and ionizable groups in proteins.

^c Peptide main chain, for example, pepCO is the CO group, and pepNHCO is NH and CO of the peptide main chain, and so forth.

^d Internal water molecule in proteins.

^eGroup of atoms from the side-chain X.

TABLE 2

Optimal Set of Atomic Radii (R_{B_i})^a

atom type	(R_{B_i}) ^b (Å)	(R_{B_i}) ^c (Å)
CH, CH ₂ , CH ₃ aliphatic (united-atom)	2.21	2.0
CH, C aromatic	1.72	1.70
N	1.50	1.50
N (NH ₃ ⁺)	1.70	2.00
O=	1.53	1.40
O ⁻ (COO ⁻)	1.30	1.58
-O- water (SPC)	1.73	
-O- TYR	1.65	
-O- SER, THR	1.45	
H-polar atom ^d	1.10	1.0
H-O SER, THR	0.85	
H (NH ₃ ⁺)	0.87	

^aComputed as described in subsection “Optimal Set of Dielectric Interface Atomic Radii R_{B_i} ”

^bComputed in this work.

^cPARSE set of radii.⁴⁸

^dAny H atom bonded to an electronegative donor other than SER and THR, for example, N-H₂, N-H, O-H, and so forth.

TABLE 3
Total Charge and Induced Dipole Moment on the Dielectric Surface Interface^a

protein	Q_σ^b (e.u.)	Q_σ^c (e.u.)	M_G^b (Å e.u.)	M_G^c (Å e.u.)	$ \Delta M / M $
4PTT ^d	-0.4875	-0.4875	-13.432	1.290	-12.887
	-0.9750	-0.9750	-9.350	-29.321	-7.706
	0.0000	0.0000	-0.357	-1.641	0.814
2LZT ^e	-0.3000	-0.3000	-4.517	-4.296	-0.103
	0.2375	0.2375	5.225	1.746	5.158
	-0.2375	-0.2375	0.373	-2.845	0.321
3RN3 ^f	-0.4750	-0.4750	10.053	1.017	-2.603
	-0.2375	-0.2375	-2.048	-0.385	-6.166
	0.2375	0.2375	7.489	5.614	7.502
1UBQ ^g	0.000	0.000	-15.317	-1.219	-15.545
1EA1 ^h	17.7750	17.7750	-82.532	1144.115	-76.276

^a Computed by the FAMBE method for different charged states (chosen arbitrarily) of proteins, indicated by each horizontal line for the proteins listed in column 1 by their Protein Data Bank code.

^b Q_σ and M_G are the induced total charges and dipole moment components computed with eqs 40 and 41, respectively. The three columns under M_G represent components of the dipole moment M_G .

^c Q_σ and M_G are the total molecular charges and the molecular dipole moment components, computed with eqs 40 and 41, respectively, while $|\Delta M|/|M|$ represents the relative error of the induced dipole moment with respect to the value (M_G) given by eq 41. The three columns under M_G represent components of the dipole moment M_G .

^d Bovine pancreatic trypsin inhibitor (BPTI).

^e Hen egg white lysozyme (HEWL).

^f Bovine pancreatic ribonuclease A (RNase A).

^g Ubiquitin.

^h Cytochrome P450.

TABLE 4
Multigrad Surface Boundary Elements on the Dielectric Surface Interface^d

protein	protein		N_{res}	n_{lev}^d	N_{BEM}^e	d_f^f	R_f^g	N_{MBE}^h			N_{ins}^i	CPU _{time} ^j (s)
	size ^b (Å)	MS ^c (Å ²)						min	aver	max		
4PTI	33.7	3119.7	58	1	12720	0.500	5.000	407	863	1128	20	242
				2	3835	1.000	9.000					
				3	269	2.000	12.000					
				4	108	4.968	16.968					
				5	32	9.935	26.903					
2LTZ	46.3	5672.4	129	1	23477	0.500	4.500	401	861	1085	32	508
				2	7173	1.000	8.500					
				3	539	2.000	12.000					
				4	379	3.272	15.272					
				5	124	6.543	21.815					
				6	33	13.087	34.902					
3RN3	45.5	5711.5	124	1	23457	0.500	4.500	346	744	1076	37	417
				2	7086	1.000	8.500					
				3	518	2.000	12.000					
				4	399	3.219	15.219					
				5	123	6.437	21.656					
				6	36	12.874	34.530					
1EAI	69.9	17943.3	448	1	75377	0.500	4.000	705	869	1502	150	3581
				2	22587	1.000	8.000					
				3	1758	2.000	15.000					
				4	715	4.744	19.744					
				5	171	9.488	29.232					
				6	43	18.976	48.208					

^aGenerated by the FAMBE method. Proteins in column 1 are named by their Protein Data Bank code.

^bLargest dimension of the globular protein in its native conformation (Å).

^cTotal area (Å²) of the dielectric interface surface, computed by using the SIMS smoothed molecular surface method.²⁹

- ^d Number of levels of the multisized tessellation (see Figure 1).
- ^e Number of boundary elements of level n which completely cover the tessellated dielectric interface surface.
- ^f Average size (\AA) of the boundary element of level i .
- ^g Distance (\AA) from the central charged group for region i which is tessellated by boundary elements of size d_i (\AA).
- ^h Number of multisized boundary elements in the tessellated dielectric surface.
- ⁱ Number of ionizable residues;
- ^j Time in seconds to calculate the ionization free energy of a single conformation for the full pH range in a single processor of a 2 GHz Pentium IV computer.

TABLE 5
The Terms of the $pK^{(1)}$ Shift of the Titratable Groups of BPTI

residue	$D_I^a = 2$			$D_I^a = 4$			$D_I^a = 8$			$D_I^a = 16$				
	$(\Delta g_{P^+}^{ds})^b$	$(\Delta g_{P^+}^{RF})^c$	$(\Delta g_{P^0}^{ds})^d$	$(\Delta g_{P^0}^{RF})^e$	$\Delta pK^{(1)f}$	$\Delta g_{P^+}^{ds}$	$\Delta g_{P^0}^{ds}$	$\Delta pK^{(1)f}$	$\Delta g_{P^+}^{ds}$	$\Delta g_{P^0}^{ds}$	$\Delta pK^{(1)f}$	$\Delta g_{P^+}^{ds}$	$\Delta g_{P^0}^{ds}$	$\Delta pK^{(1)f}$
NEND1	1.82	-0.39	0.27	0.09	-0.78	0.83	0.19	-0.46	0.19	-0.46	-0.29	0.19	-0.46	-0.19
ARG1	5.22	-5.21	2.06	-2.46	-0.27	-0.08	-0.17	-0.07	-0.08	-0.07	-0.03	-0.03	-0.07	-0.02
ASP3	0.11	-0.45	0.10	0.14	-0.43	-0.15	0.13	-0.21	-0.15	-0.21	-0.10	0.13	-0.21	-0.05
GLU7	17.13	-5.67	1.28	-1.96	8.80	5.54	-0.30	4.21	5.54	-0.30	1.97	-0.30	4.21	0.82
TYR10	4.55	-0.26	1.11	-0.24	2.48	2.23	0.42	1.31	2.23	0.42	0.71	0.42	1.31	0.38
LYS15	0.57	-0.32	0.03	0.03	-0.14	0.06	0.02	-0.03	0.06	0.02	0.02	0.02	-0.03	0.04
ARG17	0.16	-0.05	0.02	0.04	-0.05	0.04	0.04	0.0	0.04	0.04	0.02	0.04	0.0	0.04
ARG20	5.44	-3.30	2.34	-2.07	-1.35	1.08	0.11	-0.70	1.08	0.11	-0.38	0.11	-0.70	-0.20
TYR21	3.51	-0.82	0.20	0.78	1.75	1.54	0.37	0.85	1.54	0.37	0.39	0.37	0.85	0.14
TYR23	16.84	2.22	1.83	-0.36	12.75	9.74	0.72	6.55	9.74	0.72	3.41	0.72	6.55	1.78
LYS26	0.83	-0.20	0.09	0.02	-0.36	0.23	0.04	-0.14	0.23	0.04	-0.13	0.04	-0.14	-0.04
TYR35	16.36	-8.82	1.37	-2.60	6.36	3.67	-0.60	3.10	3.67	-0.60	1.50	-0.60	3.10	0.64
ARG39	0.30	0.24	0.23	-0.01	-0.23	0.31	0.12	-0.15	0.31	0.12	-0.11	0.12	-0.15	-0.09
LYS41	0.52	0.12	0.39	-0.04	-0.22	0.48	0.21	-0.19	0.48	0.21	-0.22	0.21	-0.19	-0.19
ARG42	0.89	-0.10	0.27	0.03	-0.36	0.50	0.15	-0.25	0.50	0.15	-0.20	0.15	-0.25	-0.16
LYS46	0.10	-0.02	0.07	0.12	0.09	0.01	0.10	0.06	0.01	0.10	0.01	0.10	0.06	0.00
GLU49	2.79	-1.00	0.28	-0.36	1.28	0.69	-0.07	0.55	0.69	-0.07	0.18	-0.07	0.55	-0.01
ASP50	4.29	-2.72	0.43	-1.50	1.88	0.61	-0.52	0.82	0.61	-0.52	0.29	-0.52	0.82	0.17
ARG53	2.37	-1.81	1.48	-1.17	-0.19	0.35	0.14	-0.15	0.35	0.14	-0.13	0.14	-0.15	-0.11
CEND58	0.64	0.31	0.31	0.20	0.31	0.62	0.23	0.26	0.62	0.23	0.23	0.23	0.26	0.21

^a Dielectric constant value for the (solvent-excluded) protein cavity volume used by the FAMBE method. Values of each of the free-energy components are given only for $D_I = 2$ and only the desolvation free-energy components for $D_I = 4$. All of the free-energy contributions are omitted for $D_I = 8$ and 16 to make the table clear. The term Δg_{pp} (given by eq 31) is not shown in this table because it represents a negligible contribution to the computation of the $pK^{(1)}$ shifts.

^b Desolvation free energy of the ionized group due to the protein structure and computed as the sum of the first two terms in eq 29.

^c Electrostatic interaction of a given ionized group in the protein with the electrostatic reaction field (due to the solvent) and computed as the sum of the third and fourth terms in eq 29.

- ^d Desolvation free energy of the neutral group due to the protein structure and computed as the sum of the first two terms in eq 30.
- ^e Electrostatic interaction of the neutral group in the protein with the electrostatic field (due to the solvent) and computed as the sum of the third and fourth terms in eq 30.
- ^f The $\Delta pK_1^{\text{prot}}$ is calculated by using eq 28.

Potential of Mean Force (in kcal/mol) Between Selected Pairs of Ionized Residues in BPTI

TABLE 6

residue pair	$(r_{ij})^a$ (Å)	$D_I^b = 2$		$D_I^b = 4$		$D_I^b = 8$		$D_I^b = 16$		
		$(w_{++})^{CL,c}$	$(w_{++})^{RF,d}$	$(w_{++})^e$	$(w_{+0})^f$	$(\Delta w_{ij})^g$	Δw_{ij}	Δw_{ij}	$(\alpha^{CL})^h$	
NEND1-ARG1	5.24	29.98	-28.19	1.79	0.01	1.78	1.34	1.10	0.97	0.79
NEND1-ASP3	8.53	-19.65	19.07	-0.57	0.00	-0.57	-0.56	-0.55	-0.54	-0.49
NEND1-TYR23	8.37	20.06	19.25	-0.81	0.01	-0.82	-0.66	-0.59	-0.55	-0.50
NEND1-CEND58	7.43	-21.55	21.04	-0.51	0.00	0.51	-0.53	-0.55	-0.55	-0.56
ARG1-TYR23	4.14	-43.11	34.59	-8.52	-0.06	-8.46	-4.93	-3.13	-2.15	-1.0
ARG1-CEND58	7.73	-21.40	20.54	-0.86	-0.05	-0.81	-0.72	-0.66	-0.62	-0.54
GLU7-ASP 3	9.78	16.27	-15.74	0.53	0.01	0.52	0.48	0.45	0.43	0.42
GLU7-TYR 10	9.91	15.79	-15.19	0.60	-0.01	0.59	0.52	0.47	0.44	0.42
GLU7-LYS41	5.89	-27.08	25.64	-1.44	-0.01	-1.43	-1.14	-0.98	-0.88	-0.70
GLU7-ARG42	8.76	-19.69	18.93	-0.76	-0.02	-0.74	-0.66	-0.60	-0.56	-0.47
TYR10-TYR35	8.55	18.65	-18.00	0.65	-0.06	0.71	0.65	0.51	0.49	0.49
TYR10-ARG39	9.47	-18.83	18.45	-0.36	0.01	-0.37	-0.44	-0.47	-0.48	-0.44
TYR10-LYS41	5.46	-28.17	26.68	-1.49	-0.00	-1.49	-1.16	-0.98	-0.88	-0.76
ARG20-TYR35	5.45	-27.42	25.13	-2.28	-0.08	-2.20	-1.59	-1.25	-1.04	-0.76
ARG20-LYS46	5.40	-29.24	27.78	1.46	0.11	1.35	1.12	0.99	0.90	0.77
TYR21-GLU49	7.00	24.59	-23.39	1.20	-0.01	1.19	0.99	0.88	0.80	0.59
TYR35-ARG39	9.63	-18.60	17.85	-0.75	-0.00	-0.75	-0.63	0.60	-0.53	-0.43
LYS46-ASP50	8.65	-19.68	18.99	-0.69	0.00	-0.69	-0.65	-0.62	-0.59	-0.48
GLU49-ASP50	7.64	21.44	-20.84	0.60	-0.00	0.60	0.58	0.57	0.56	0.54
GLU49-ARG53	7.54	20.80	20.29	-0.50	-0.01	-0.49	-0.50	-0.51	-0.52	-0.55
ASP50-ARG53	4.70	-41.18	36.57	-4.61	-1.31	-3.29	-2.24	-1.70	-1.40	-0.88

^aDistance in Angstroms between ionizable groups.

^bValue of the intramolecular dielectric constant.

^cEnergy of the intramolecular Coulombic interaction between two ionized groups, given by the first term of eq 33.

^dSolvent reaction field energy of the interaction between two ionized groups, given by the second term of eq 33.

^ePotential of mean force due to the interaction between the ionized groups, computed by eq 33.

^fCorrection to the potential of mean force between pairs of ionized groups, computed by eq 34.

^gThe total potential of mean force, computed by eq 32.

^hThe energy of the Coulombic interaction, $u_{ij}^{\text{CL}} = q_i q_j / D_0 r_{ij}$, between ionized groups using a solvent dielectric constant of $D_0 = 80$.

TABLE 7
 Comparison Between Computed and Observed pK_a 's in BPTI^a

residue	$(pK_a)^b$	$(pK_a)^{exp,c}$			$(pK_a)^{calc,d}$			$(pK_a)^{calc,e}$			$(pK_a)^f$	$(pK_a)^g$	$(pK_a)^h$
		$D_I = 2$	$D_I = 4$	$D_I = 8$	$D_I = 16$	$D_I = 2$	$D_I = 4$	$D_I = 8$	$D_I = 16$	$D_I = 2$			
NEND1	7.5	8.1	5.3	6.5	7.1	7.6	5.8	6.6	7.2	7.3	7.2	7.5	-
ARG1	12.0	-	12.8	16.1	15.8	15.4	12.8	15.6	14.8	14.5	18.1	13.6	-
ASP3 ⁱ	4.0	3.0	1.8	1.8	1.9	2.0	2.4	2.5	3.3	3.2	3.4	3.3	-
GLU7	4.4	3.7	14.9	6.1	4.1	2.9	14.4	6.4	5.3	4.1	5.4	3.7	-
TYR10	9.6	-	12.7	9.5	8.8	8.5	12.5	9.6	9.4	9.2	9.9	9.6	-
LYS15	10.4	10.6	9.4	10.4	10.8	10.8	9.9	10.4	10.7	10.6	10.4	10.5	-
ARG17	12.0	-	12.2	12.6	12.6	12.6	12.0	12.4	12.3	12.0	12.2	12.2	-
ARG20	12.0	-	9.4	12.8	13.4	13.5	9.5	12.6	12.6	12.7	13.1	12.9	-
TYR21	9.6	-	12.6	10.5	9.7	9.4	12.5	10.7	9.9	9.7	10.1	9.5	-
TYR23	9.6	-	21.5	13.0	10.2	9.5	21.2	13.0	10.2	9.7	11.3	10.6	-
LYS26	10.4	10.6	10.5	10.7	11.1	10.9	10.6	10.8	10.8	10.7	10.4	10.6	-
TYR35	9.6	-	17.1	10.9	8.8	8.1	16.8	11.1	9.5	8.9	9.5	8.5	-
ARG39	12.0	-	11.9	12.8	12.9	12.8	11.8	12.5	12.2	12.2	12.2	12.4	-
LYS41	10.4	10.8	8.8	11.3	11.5	11.4	9.0	11.3	11.2	11.1	10.2	10.8	-
ARG42	12.0	-	11.8	12.9	12.9	12.9	11.7	12.6	12.1	12.1	13.0	12.5	-
LYS46	10.4	10.6	8.9	9.8	10.2	10.6	9.3	9.9	10.3	10.4	10.0	10.3	-
GLU49	4.4	3.8	4.4	3.6	3.5	3.2	4.9	4.3	4.3	4.1	3.8	3.7	-
ASP50	4.0	3.4	1.8	1.5	1.8	2.4	2.5	2.5	2.7	2.6	2.3	2.6	-
ARG53	12.0	-	15.1	14.7	14.2	13.9	14.8	14.3	13.3	13.1	12.9	13.0	-
CEND58	3.8	2.9	1.5	1.7	1.8	2.1	2.3	2.4	3.0	2.9	3.9	3.3	-
$(\Delta_{av})^j$			2.5	1.0	0.8	0.6	2.0	0.8	0.5	0.3	0.7	0.3	0.6
$(\Delta_{max})^k$			3.2	2.4	1.6	1.0	3.1	2.7	1.4	0.8	1.7	0.8	1.5

^aIn pK units, computed by eq 18.

^bValue of pK_a^0 of the residues at $T = 300$ K.⁵¹

^cExperimental values of pK_a^{exp} for the residue in 0.2 M NaCl and at $T = 300$ K.⁵¹

- ^d Calculated pK_a for BPTI using $D_0 = 80$ and $D_I = 2, 4, 8,$ and 16 in 0.0 M salt at $T = 300\text{ K}$.
- ^e Calculated pK_a for BPTI using $D_0 = 80$ and $D_I = 2, 4, 8,$ and 16 in 0.15 M of 1.1 salt at $T = 300\text{ K}$.
- ^f Calculated in ref 50 at 0.15 M ionic strength and $T = 293\text{ K}$.
- ^g Calculated in ref 51.
- ^h Calculated in ref 52. Dashes indicate that the pK_a 's for these ionizable groups were not reported in ref 52.
- ⁱ Shift between $(pK_a)_b$ and $(pK_a)_{exp}$ $c \geq 1.0$.
- ^j Average value of the absolute difference between observed and computed pK_a 's.
- ^k Maximum deviation between observed and computed pK_a 's.

TABLE 8
Comparison Between the Computed and Observed pK_a in HEWL^a

residue	$(pK_a)^b$	$(pK_a^{exp})^c$	$(\Delta pK_a)^d$	$(pK_a^{calc})^e$	$(pK_a)^f$	$(pK_a)^g$	$(pK_a)^h$	$(pK_a)^i$
NEND1	7.5	7.9	-0.89	6.8	5.6	8.2	7.3	-
LYSI	10.4	10.7	-0.36	10.5	10.1	10.6	10.5	-
ARG5	12.0	-	0.14	12.4	12.8	-	12.8	-
GLU7 ^j	4.4	2.6	-0.06	3.4	2.8	2.6	2.9	3.7
LYS13	10.4	10.5	0.38	11.5	10.7	11.6	10.8	-
ARG14	12.0	-	-0.10	12.3	12.4	-	12.6	-
HIS15	6.6	5.8	-0.26	5.9	5.1	5.4	5.8	-
ASP18 ^k	4.0	2.0-2.9	0.09	3.6	2.7	3.4	2.6	-
TYR20	9.6	10.3	1.71	9.8	12.4	13.0	11.8	-
ARG21	12.0	-	0.26	13.2	12.7	-	13.2	-
TYR23	9.6	9.8	0.48	9.3	9.7	10.0	9.4	-
LYS33	10.4	10.6	0.04	10.4	10.9	11.1	10.5	-
GLU35 ^l	4.4	6.1	0.58	5.0	5.6	4.6	4.5	5.0
ARG45	12.0	-	-0.65	11.7	12.2	-	12.4	-
ASP48	4.0	3.4-4.3	0.50	3.4	1.5	1.3	2.7	1.4
ASP52	4.0	3.5	-0.39	3.3	3.6	5.4	3.0	-
TYR53 ^m	9.6	12.1	0.41	12.5	20.5	16.0	11.9	-
ARG61	12.0	-	-0.18	13.4	13.7	-	13.1	-
ASP66 ⁿ	4.0	1.9	1.05	3.0	1.9	1.5	2.9	1.3
ARG68	12.0	-	-0.44	13.8	13.0	-	13.3	-
ARG73	12.0	-	0.14	12.3	12.5	-	12.5	-
ASP87	4.0	2.1-3.7	-0.20	2.6	2.4	0.5	2.3	-
LYS96	10.4	10.7	0.35	11.3	13.6	11.8	11.2	-
LYS97	10.4	10.2	0.07	11.4	11.2	11.4	11.1	-
ASP101	4.0	4.3	0.98	3.9	4.4	6.9	3.8	-
ARG112	12.0	-	-0.24	12.2	11.2	-	12.4	-
ARG114	12.0	-	0.04	12.2	13.3	-	12.5	-
LYS116	10.4	10.3	-0.21	10.3	9.3	10.4	10.4	-
ASP119 ^o	4.0	2.5	0.04	2.9	3.7	3.1	2.8	-

residue	$(pK_a^0)^b$	$(pK_a^{exp})^c$	$(\Delta pK^{(1)})^d$	$(pK_a^{calc})^e$	$(pK_a)^f$	$(pK_a)^g$	$(pK_a)^h$	$(pK_a)^i$
ARG125	12.0	-	-0.05	12.8	12.9	-	12.6	-
ARG128	12.0	-	0.13	12.4	12.2	-	12.3	-
CEND129	3.8	3.0	-0.09	2.8	3.1	2.9	2.6	-
$(\Delta_{av})^k$				$0.5 (0.7)^m$	1.0	1.0	0.5	0.7
$(\Delta_{max})^l$				$1.2 (1.1)^n$	2.9	2.7	1.6	1.7

^aIn pK units, computed by eq 18 with $D_I = 16$, 0.15 M salt concentration at $T = 300$ K.

^bValue of pK_a^0 of the residues at 300 K.⁵¹

^cExperimental values of pK_a^{exp} for the residues in 0.15 M NaCl and at $T = 300$ K.¹²

^d $\Delta pK^{(1)}$ is the shift in pK_a^0 due to the transport of the ionizable residue from the solution into the protein and is computed by using eq 28.

^eCalculated pK_a for HEWL by using $D_0 = 80$, $D_I = 16$, and 0.10 M 1:1 salt, at $T = 300$ K.

^fCalculated in ref 50.

^gCalculated in ref 12. Dashes indicate that the pK_a 's for these ionizable groups were not reported in ref 12.

^hCalculated in ref 51.

ⁱCalculated in ref 52. Dashes indicate that the pK_a 's for these ionizable groups were not reported in ref 52.

^jShift between $(pK_a^0)^b$ and $(pK_a^{exp})^c \geq 1.5$.

^kAverage value of the absolute difference between observed and computed pK_a 's.

^lMaximum deviation between observed and computed pK_a 's.

^mAverage value of the absolute difference between observed and computed pK_a 's only for those residues indicated by footnote j in column 1.

ⁿMaximum deviation between observed and computed pK_a 's only for those residues indicated by footnote j in column 1.

TABLE 9
Comparison Between the Computed and Observed pK in RNase A^a

residue	$(pK_a^0)_b$	$(pK_a^{exp})_c$	$(\Delta pK^{(0)})_d$	$(pK_a^{calc})_e$	$(pK_a^{calc})_f$	$(pK_a^{calc})_g$	$(pK_a)_h$	$(pK_a)_i$	$(pK_a)_j$	$(pK_a)_k$	$(pK_a)_l$
NEND1	7.5	7.6	0.05	7.0	7.1	7.0	6.3	7.1	-	-	-
LYS1	10.4	-	0.01	10.7	10.8	10.7	10.5	10.6	-	-	-
GLU2 ^m	4.4	2.8	0.62	1.3	3.2	3.0	2.0	2.6	2.6	0.0	2.4
LYS7	10.4	-	-0.34	9.9	10.7	10.1	9.6	10.7	-	-	-
GLU9	4.4	4.0	0.04	3.9	4.1	3.9	4.7	4.3	-	2.6	2.8
ARG10	12.0	-	-0.24	14.8	12.6	12.3	17.6	13.7	-	-	-
HIS12	6.6	6.2	-0.31	4.8	6.0	4.8	4.2	6.2	5.8	5.1	6.6
ASP14 ^m	4.0	2.0	1.22	2.1	1.8	1.9	0.0	2.3	1.7	3.1	0.8
TYR25	9.6	-	1.65	16.0	16.1	16.1	19.6	15.0	-	-	-
LYS31	10.4	-	-0.28	10.3	10.3	10.3	9.7	10.2	-	-	-
ARG33	12.0	-	0.24	13.5	13.7	13.7	13.4	13.7	-	-	-
LYS37	10.4	-	-0.04	10.9	11.0	11.0	10.9	11.0	-	-	-
ASP38	4.0	3.1	0.10	2.9	3.0	2.9	3.2	2.8	-	1.5	2.2
ARG39	12.0	-	-0.12	12.4	12.7	12.5	12.5	12.9	-	-	-
LYS41	10.4	-	-0.98	8.8	8.9	8.5	9.7	9.3	-	-	-
HIS48	6.6	6.3	-1.33	7.3	7.2	7.4	9.7	4.9	-	6.3	8.3
GLU49	4.4	4.7	0.67	4.8	4.7	4.8	5.4	5.1	-	5.0	3.7
ASP53	4.0	3.9	0.18	3.9	3.9	3.9	3.9	3.5	-	3.8	3.6
LYS61	10.4	-	0.31	11.4	11.1	11.4	10.4	11.2	-	-	-
LYS66	10.4	-	0.05	11.7	11.7	11.6	12.6	11.4	-	-	-
TYR73	9.6	-	1.13	11.6	11.6	11.6	14.0	11.0	-	-	-
TYR76	9.6	-	0.74	9.9	9.7	9.8	10.3	10.1	-	-	-
ASP83	4.0	3.5	1.39	3.0	2.9	3.1	1.7	1.8	-	-0.9	2.2
ARG85	12.0	-	-0.05	13.3	13.2	13.3	13.2	13.2	-	-	-
GLU86	4.4	4.1	0.67	4.1	4.3	4.1	4.7	3.4	-	3.9	3.2
LYS91	10.4	-	-0.02	10.7	10.8	10.8	11.2	11.0	-	-	-
TYR92	9.6	-	0.86	9.8	9.8	9.8	10.8	10.4	-	-	-
TYR97	9.6	-	2.37	13.6	13.9	13.8	17.6	12.7	-	-	-
LYS98	10.4	-	0.06	10.5	10.4	10.4	10.2	10.5	-	-	-

residue	$(pK_a^0)^b$	$(pK_a^{exp})^c$	$(\Delta pK^{(1)})^d$	$(pK_a^{calc})^e$	$(pK_a^{calc})^f$	$(pK_a^{calc})^g$	$(pK_a)^h$	$(pK_a)^i$	$(pK_a)^j$	$(pK_a)^k$	$(pK_a)^l$
LYS104	10.4	–	0.55	11.9	10.7	11.9	10.7	11.3	–	–	–
HIS105	6.6	6.6	0.66	7.8	7.6	7.8	5.6	6.6	–	7.3	8.3
GLU111	4.4	3.5	-0.14	3.7	3.7	3.5	4.2	3.7	–	3.7	3.3
TYR115	9.6	–	1.12	12.3	12.6	12.3	10.8	10.1	–	–	–
HIS119	6.6	6.1	-0.68	7.5	6.8	6.3	6.3	6.2	–	5.7	6.8
ASP121	4.0	3.1	0.76	0.7	2.9	2.6	1.1	2.1	3.7	0.4	1.0
CEN124 ^m	3.8	2.4	0.39	2.4	2.6	2.3	2.3	2.4	0.0	0.0	1.6
$(\Delta_{in})^n$				0.7	0.4	0.4	1.1	0.5	0.7	1.3	1.0
$(\Delta_{max})^o$				2.4	1.0	1.4	3.4	1.7	1.2	4.4	2.1

^aIn pK units, computed by eq 18 with $D_I = 16, 0.15$ M salt concentration at $T = 300$ K.

^bValue of pK_a^0 of the residues at 300 K.⁵¹

^cExperimental values of the pK_a^{exp} for the residues in 0.15 M NaCl and at $T = 300$ K.⁵¹

^d $\Delta pK^{(1)}$ is the shift in pK_a^0 due to the transport of the ionizable residue from the solution into the protein and is computed by using eq 28.

^eCalculated pK_a for RNase A with the SO_4^{2-} ion taken with an occupancy of 0.57 and the HIS119 taken in position A. The calculation was carried out for $D_0 = 80, D_I = 16, 0.15$ M 1:1 salt, and at $T = 300$ K.

^fCalculated pK_a for RNase A with the SO_4^{2-} ion taken with an occupancy of 0.57, the HIS119 taken in position B, and with the ionized group of ARG10 displaced by about 1.5 Å, resulting in rupture of the GLU2⁻...ARG10⁺ hydrogen bond. The calculation was carried out for $D_0 = 80, D_I = 16, 0.15$ M 1:1 salt, and at $T = 300$ K.

^gCalculated pK_a for the same structure as described in footnote f but with the SO_4^{2-} ion removed.

^hCalculated in ref 49.

ⁱCalculated in ref 51.

^jCalculated in ref 52. Dashes indicate that the pK_a 's for these ionizable groups were not reported in ref 52.

^kCalculated in ref 20. Dashes indicate that the pK_a 's for these ionizable groups were not reported in ref 20.

^lCalculated in ref 22. Dashes indicate that the pK_a 's for these ionizable groups were not reported in ref 22.

^mShift between $(pK_a^0)^b$ and $(pK_a^{exp})^c \geq 1.4$.

ⁿ Average value of the absolute difference between observed and computed pK_a 's.

^o Maximum deviation between observed and computed pK_a 's.