Computing Ionization States of Proteins with a Detailed Charge Model

JAN ANTOSIEWICZ,* JAMES M. BRIGGS,† ADRIAN H. ELCOCK, MICHAEL K. GILSON,‡ and J. ANDREW McCAMMON

Departments of Chemistry and Biochemistry, and Pharmacology, University of California at San Diego, La Jolla, California 92093-0365

Received 28 September 1995; accepted 21 December 1995

ABSTRACT

A convenient computational approach for the calculation of the pK_as of ionizable groups in a protein is described. The method uses detailed models of the charges in both the neutral and ionized form of each ionizable group. A full derivation of the theoretical framework is presented, as are details of its implementation in the UHBD program. Application to four proteins whose crystal structures are known shows that the detailed charge model improves agreement with experimentally determined pK_as when a low protein dielectric constant is assumed, relative to the results with a simpler single-site ionization model. It is also found that use of the detailed charge model increases the sensitivity of the computed pK_as to the details of proton placement.

Introduction

Ionization equilibria are critical to the structure and function of proteins. As a consequence, the prediction of the ionization states of titratable groups in proteins is an important component of theoretical efforts to account for the properties of proteins. Furthermore, the availability of a substantial body of data on ionization equilibria—chiefly, measured pK_as of ionizable groups—makes it possible to carry out revealing validation studies of theoretical methods for predicting pK_as.

The prediction of ionization constants remains a challenging and still not fully solved problem in molecular biophysics, but considerable progress has been made recently through application of the Poisson–Boltzmann (PB) model for electrostatic interactions.[1–5] One interesting result of this previous work has been the observation that protein pK_as are predicted more accurately when the protein is assumed to have a dielectric constant of 20, than when the protein is assumed to have a dielectric constant of 4.[5] This result has been somewhat

* Current address: Department of Biophysics, University of Warsaw, Warsaw 02-089, Poland.

† Author to whom all correspondence should be addressed.

‡ Current address: Center for Advanced Research in Biotechnology, National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, MD 20850-3479.

CCC 0192-8651 / 96 / 141633-12

surprising because theoretical calculations of the dielectric constant of a protein have typically yielded values closer to 4 than to 20 [6,7] (see also Refs. 5 and 8). Explanations for this apparent discrepancy have been discussed elsewhere.[5]

One of the possible explanations has to do with the model for ionization. Our work to date has modeled ionization as the addition of a single point charge of ± 1 electrons to a single atom in each titrating group. Although this model does yield predictive results when used with a protein dielectric constant of 20,[5] it is clearly rather simplistic: ionization is associated with the gain or loss of a proton, and ionic charge is distributed over more than one atom. Thus, the use of a simplistic single-site ionization model is an obvious candidate as a source of error in the pK_a calculations we have hitherto reported. This article describes the incorporation of a much more detailed model of the charge changes associated with protonation or deprotonation reactions into an efficient algorithm for computing pK_as.

The method is based on the finite-difference PB approach[9] to computing electrostatic interactions, as implemented in the UHBD program.[10,11] This method is termed a "full-charge" titration model, because it takes into account changes in the point charges of all atoms of the titratable amino acids upon protonation or deprotonation. In this respect it is similar to the method described by Yang et al.,[4] and possibly also to that of Bashford and Gerwert.[12] As with procedures developed by others, four PB calculations per ionizable group provide all the necessary energy terms. In addition, the procedure takes advantage of the scripting features of UHBD, and is therefore highly automated.

The titration calculation consists of the following steps:

1. Preparation of a PDB-format[13] file containing the coordinates and appropriate names of all atoms, including all polar and aromatic hydrogens. The procedure for adding hydrogens is described below.
2. Use of the finite difference PB (FDPB) method to calculate the self- and interaction-energies of the ionizable groups in the protein.
3. Use of a Monte Carlo or a "cluster" algorithm to calculate ionization constants of the titratable groups, net average charges, and electrostatic energies as functions of pH. The Monte Carlo method also gives the most stable protonation states; these may be useful in setting up subsequent calculations, such as molecular dynamics simulations.

As noted above, the procedure is computationally efficient. For example, for Ribonuclease A, with 124 residues and 36 titratable groups, an entire calculation requires no more than 3.5 h on an SGI Indigo2 workstation with a 200-MHz MIPS R4400 processor. The bulk of the computer time is spent on the FDPB calculations. The speed of this procedure is important, because it makes it possible to collect results on many systems and with many different sets of parameters in a reasonable amount of time. Thus, improvements to the method can be made based on a broad sampling of systems.

Theoretical Methods

The underlying theory of this work has been presented previously.[14-16] All pH-dependent quantities will be computed using an ionization polynomial.[16,17] This approach requires an expression for the free energy of the protein in each possible ionization state. If the neutral state of the protein is taken as the reference state, its electrostatic energy in a given ionization state[16] is:

$$\Delta G(pH, x_1, \ldots, x_M) = 2.303\,\mathrm{RT} \sum_{i=1}^{M} x_i \gamma_i (\mathrm{pH} - pK_{a_{i,intrinsic}}) + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} x_i x_j \gamma_i \gamma_j \Psi_{ij} \quad (1)$$

where x_i is 1 when the group i is ionized, and 0 when it is neutral; γ_i is $+1$ for bases, and -1 for acids; $pK_{a_{i,intrinsic}}$ is the intrinsic pK_a of group i[14] (see below); Ψ_{ij} is the absolute value of the interaction energy of groups i and j [eq. (19)]; M is the number of ionizable groups in the protein; R is the gas constant; and T is absolute temperature. The intrinsic pK_a of group i is given by:

$$pK_{a_{intrinsic,i}} = pK_{a_{model,i}} - \gamma_i \Delta\Delta G_i / 2.303\,\mathrm{RT} \quad (2)$$

where $\Delta\Delta G_i \equiv \Delta G^{electrostatic}_{protein} - \Delta G^{electrostatic}_{model}$ is the difference between the free energy change of ionization for group i in the otherwise unionized protein and in a model compound of $pK_{a_{model,i}}$. The intrinsic pK_a thus represents the pK_a of the group in the protein with all other titratable amino acids

in the neutral state. Given values for Ψ_{ij} and $\Delta\Delta G_i$ for all groups, eq. (1) provides the basis for the computation of the pH-dependent properties of the protein (see below).

It is also worth giving the expression for the ionization energy relative to the fully deprotonated, rather than neutral, state of the protein:[18]

$$\Delta G(pH, x'_1, \ldots, x'_M) = 2.303\,RT \sum_{i=1}^{M} x'_i(\mathrm{pH} - \mathrm{p}K_{a_{i,intrinsic}}) + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \left(q_i q_j - q_i^0 q_j^0\right)\Psi_{ij} \quad (3)$$

where the parameter x'_i is 1 when the site is protonated, and 0 otherwise; q_i and q_i^0 are the *net* charges of group i in a given ionization state of the protein, and in the deprotonated state, respectively. The parameter x' is implicitly present in the second term on the right side of eq. (3) through charge q_i, because for $x'_i = 0$, $q_i = q_i^0$.

FULL-CHARGE MODEL

As in previous applications of the PB model to the computation of ionization equilibria, the goal is to determine the influence of the protein environment upon the free energy of ionizing each titratable group. This section derives equations showing what FDPB calculations must be carried out to yield the electrostatic energy terms that enter eqs. (1), (2), and (3). The derivation focuses first upon an ionizable model compound; then upon a hypothetically neutral protein containing the same ionizable group and, finally, upon the interactions among ionizable groups in a protein.

To begin, consider the ionization of a titratable group forming part of a model compound in solution. Within the PB model, the electrostatic free energy of the neutral compound is:

$$G_{model}^{(neutral)} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} q_i \phi_{ji}^m \quad (4)$$

were q_i is the charge assigned to atom i, ϕ_{ji}^m is the potential at atom i due to the charge at atom j, and n is the number of charge-bearing atoms in the neutral model compound. The energy of the ionized form of the compound is:

$$G_{model}^{(ionized)} = \frac{1}{2} \sum_{i=1}^{n'} \sum_{j=1}^{n'} q'_i \phi_{ji}^m \quad (5)$$

where the primes indicate that the variables pertain to the ionized compound.

The electrostatic energies of a hypothetically neutral protein containing the same titratable group in its neutral and ionized forms may be written:

$$G_{protein}^{(neutral)} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} q_i \phi_{ji}^p \quad (6)$$

and

$$G_{protein}^{(ionized)} = \frac{1}{2} \sum_{i=1}^{N'} \sum_{j=1}^{N'} q'_i \phi_{ji}^{\prime p} \quad (7)$$

Next, consider the changes in energy associated with ionization. This discussion requires that the atomic charges associated with the titrating group be separated from the remaining "fixed" charges. Thus, for the neutral titratable group, let indices 1 through m_1 correspond to atoms belonging to the titratable group, and let the remaining indices ($m_1 + 1$ through n for the model compound, $m_1 + 1$ through N for the protein) correspond to all other atoms. Adding primes to the indices gives the corresponding ranges for the ionized group.

With this notation, the energy of ionizing the model compound becomes, from eq. (4) and eq. (5):

$$\begin{aligned}
G_{model}^{(ionized)} - G_{model}^{(neutral)} &= \Delta G_{model}^{electrostatic} \\
&= \frac{1}{2}\left(\sum_{i=1}^{m'_1} + \sum_{i=m'_1+1}^{n'}\right)\left(\sum_{j=1}^{m'_1} + \sum_{j=m'_1+1}^{n'}\right) q'_i \phi_{ji}^m \\
&\quad - \frac{1}{2}\left(\sum_{i=1}^{m_1} + \sum_{i=m_1+1}^{n}\right)\left(\sum_{j=1}^{m_1} + \sum_{j=m_1+1}^{n}\right) q_i \phi_{ji}^m \\
&= \frac{1}{2}\left(\sum_{i=1}^{m'_1}\sum_{j=1}^{m'_1} + \sum_{i=1}^{m'_1}\sum_{j=m'_1+1}^{n'} + \sum_{i=m'_1+1}^{n'}\sum_{j=1}^{m'_1}\right) \\
&\quad \times q'_i \phi_{ji}^m \\
&\quad - \frac{1}{2}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} + \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{n} + \sum_{i=m_1+1}^{n}\sum_{j=1}^{m_1}\right) \\
&\quad \times q_i \phi_{ji}^m \\
&= \frac{1}{2}\sum_{i=1}^{m'_1}\sum_{j=1}^{m'_1} q'_j \phi_{ij}^{\prime m} + \sum_{i=1}^{m'_1}\sum_{j=m'_1+1}^{n'} q'_j \phi_{ij}^{\prime m} \\
&\quad - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} q_j \phi_{ij}^m - \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{n} q_j \phi_{ij}^m
\end{aligned}$$

$$= \sum_{j=1}^{n'} q'_j \phi'^{m}_j - \frac{1}{2} \sum_{j=1}^{m'_1} q'_j \phi'^{m}_j - \sum_{j=1}^{n} q_j \phi^{m}_j + \frac{1}{2} \sum_{j=1}^{m_1} q_j \phi^{m}_j \quad (8)$$

here, ϕ_j and ϕ'_j are the potentials at atom j due to all the atoms forming the titratable group, in their neutral and ionized forms, respectively. Thus, for the neutral group:

$$\phi_j \equiv \sum_{i=1}^{m_1} \phi_{ij} \quad (9)$$

Eq. (8) shows that only two FDPB calculations are required to compute the ionization energy of the titratable model compound. One calculation yields the values ϕ^m_j, and the other yields ϕ'^m_j.

The cancellation of the fourth expansion terms in eq. (8) deserves special attention. It is assumed that those energy terms associated only with the nontitrating atoms of the compound (atoms $m_1 + 1$ to n, which are identical to atoms $m'_1 + 1$ to n') are unaffected by ionization. That is:

$$\sum_{i=m'_1+1}^{n'} \sum_{j=m'_1+1}^{n'} q'_i \phi^{m}_{ji} - \sum_{i=m_1+1}^{n} \sum_{j=m_1+1}^{n} q_i \phi^{m}_{ji} = 0 \quad (10)$$

Formally, this holds within the PB model only when the dielectric boundary of the molecule is unchanged by ionization. Even when a fixed conformation is assumed, as is typical in current implementations of the PB model, the assumption is not strictly valid, because addition or removal of a proton produces small changes in the dielectric boundary. Thus, the use of this cancellation constitutes an approximation, because the dielectric boundary does change when a group is ionized in the FDPB calculations.

This approximation has been checked by comparison with exact calculations. It is found that the model used here does not lead to inaccuracies greater than 0.84 kJ/mol in $\Delta G^{electrostatic}_{model}$. Note also that the same approximation is used or each group in the protein environment. Therefore, in many cases, the error associated with the approximation will be canceled. The advantage of using this modest approximation is that it leads to a conveniently simple formulation of the problem, and a reduction in the number of FDPB runs that must be performed to carry out the calculations.

The derivation above applies to the model compound. A completely analogous derivation for the titratable group in the hypothetically neutral protein is now presented. Subtraction of the right-hand side of eq. (6) from that of eq. (7) yields:

$$\begin{aligned}
\Delta G^{electrostatic}_{protein} &= \frac{1}{2}\left(\sum_{i=1}^{m'_1} + \sum_{i=m'_1+1}^{N'}\right)\left(\sum_{j=1}^{m'_1} + \sum_{j=m'_1+1}^{N'}\right) q'_i \phi'^{p}_{ji} \\
&\quad - \frac{1}{2}\left(\sum_{i=1}^{m_1} + \sum_{i=m_1+1}^{N}\right)\left(\sum_{j=1}^{m_1} + \sum_{j=m_1+1}^{N}\right) q_i \phi^{p}_{ji} \\
&= \frac{1}{2}\left(\sum_{i=1}^{m'_1}\sum_{j=1}^{m'_1} + \sum_{i=1}^{m'_1}\sum_{j=m'_1+1}^{N'} + \sum_{i=m'_1+1}^{N'}\sum_{j=1}^{m'_1}\right) \times q'_i \phi'^{p}_{ji} \\
&\quad - \frac{1}{2}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} + \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{N} + \sum_{i=m_1+1}^{N}\sum_{j=1}^{m_1}\right) \times q_i \phi^{p}_{ji} \\
&= \frac{1}{2}\sum_{i=1}^{m'_1}\sum_{j=1}^{m'_1} q'_j \phi'^{p}_{ij} + \sum_{i=1}^{m'_1}\sum_{j=m'_1+1}^{N'} q'_j \phi'^{p}_{ij} \\
&\quad - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} q_j \phi^{p}_{ij} - \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{N} q_j \phi^{p}_{ij} \\
&= \sum_{j=1}^{N'} q'_j \phi'^{p}_j - \frac{1}{2}\sum_{j=1}^{m'_1} q'_j \phi'^{p}_j \\
&\quad - \sum_{j=1}^{N} q_j \phi^{p}_j + \frac{1}{2}\sum_{j=1}^{m_1} q_j \phi^{p}_j
\end{aligned} \quad (11)$$

where the notation and the steps parallel those of eq. (8).

Eq. (11) shows that two FDPB calculations are required to compute the ionization energy of the titratable group in the neutral protein. One calculation yields the values ϕ^p_j, and the other yields ϕ'^p_j. These potentials are generated by the ionizable group in its neutral and ionized forms, respectively. Thus, the FDPB calculations require that charges be placed upon the ionizable group only; the interactions of the group with all other charge-bearing atoms, j, are obtained by multiplying the potential at each such atom by the potential due to the group. In this respect, the method is in strict analogy to the single-point titration site model.[5]

The remainder of this section demonstrates that the same set of FDPB calculations also yields the interaction energies between titratable groups in their ionized states relative to the corresponding interaction energy in the unionized states.

For simplicity, consider first a protein with only two titratable groups. For the neutral state of the protein, let atoms 1 through m_1 be part of the first titratable group, atoms $m_1 + 1$ through m_2 be atoms of the second titratable group. The remaining atoms, $m_2 + 1$ to N, belong to atoms whose charges do not depend upon pH. Corresponding primed indices will be used for the state with both groups ionized. Thus, the change in electrostatic energy when both groups go from neutral to ionized reads:

$$\begin{aligned}\Delta G_{two\ sites} &= \frac{1}{2}\sum_{i=1}^{N''}\sum_{j=1}^{N''} q_i'\phi_{ji}'^{p} - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i\phi_{ji}^{p} \\ &= \frac{1}{2}\left(\sum_{i=1}^{m_1'} + \sum_{i=m_1'+1}^{m_2'} + \sum_{i=m_2'+1}^{N''}\right) \\ &\quad\times\left(\sum_{j=1}^{m_1'} + \sum_{j=m_1'+1}^{m_2'} + \sum_{j=m_2'+1}^{N''}\right) q_i'\phi_{ji}'^{p} \\ &\quad -\frac{1}{2}\left(\sum_{i=1}^{m_1} + \sum_{i=m_1+1}^{m_2} + \sum_{i=m_2+1}^{N}\right) \\ &\quad\times\left(\sum_{j=1}^{m_1} + \sum_{j=m_1+1}^{m_2} + \sum_{j=m_2+1}^{N}\right) q_i\phi_{ji}^{p}\end{aligned} \tag{12}$$

Eq. (12) has its analog in the single site model[5] and, as in that case, this expression may be separated into three groups of terms that correspond to the energy of each group by itself, and an interaction term. This separation is accomplished as follows. Eq. (12) is expanded to yield:

$$\begin{aligned}&\Delta G_{two\ sites} \\ &= \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} + \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &\quad + \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} + \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} \\ &\quad + \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &\quad + \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} \\ &\quad + \frac{1}{2}\sum_{i=m_2'+1}^{N''}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} \\ &\quad + \frac{1}{2}\sum_{i=m_2'+1}^{N''}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &\quad + \frac{1}{2}\sum_{i=m_2'+1}^{N''}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &\quad - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p} \\ &\quad - \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} - \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} \\ &\quad - \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p} - \frac{1}{2}\sum_{i=m_2+1}^{N}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &\quad - \frac{1}{2}\sum_{i=m_2+1}^{N}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} - \frac{1}{2}\sum_{i=m_2+1}^{N}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p}\end{aligned} \tag{13}$$

The assumption of a fixed dielectric boundary, and use of the reciprocal relation, $q_i\phi_{ji} = q_j\phi_{ij}$, yield that:

$$\begin{aligned}&\Delta G_{two\ sites} \\ &= \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} + \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &\quad + \sum_{i=1}^{m_1'}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} + \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} \\ &\quad + \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &\quad + \sum_{i=m_1'+1}^{m_2'}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &\quad - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} - \sum_{i=1}^{m_1}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p} \\ &\quad - \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &\quad - \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} \\ &\quad - \sum_{i=m_1+1}^{m_2}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p}\end{aligned} \tag{14}$$

This expression contains terms resembling those for a single titratable group in a neutral protein [eq. (11)]. A rearrangement of eq. (14) that recovers one such term for each of the two titratable groups

will leave a remainder that represents the interaction of the two groups. Initial rearrangement yields:

$$\begin{aligned}\Delta G_{two\ sites} &= \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} + \sum_{i=1}^{m_1'}\sum_{j=m_2'+1}^{N''} q_1'\phi_{ji}'^{p} \\ &+ \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_1'+m_2-m_1} q_i'\phi_{ji}'^{p} - \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_1'+m_2-m_1} q_i'\phi_{ji}'^{p} \\ &- \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} - \sum_{i=1}^{m_1}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p} \\ &- \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_1+m_2-m_1} q_i\phi_{ji}^{p} + \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_1+m_2-m_1} q_i\phi_{ji}^{p} \\ &+ \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &+ \sum_{i=m_1'+1}^{m_2'}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} + \sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1} q_i'\phi_{ji}'^{p} \\ &- \sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1} q_i'\phi_{ji}'^{p} - \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} \\ &- \sum_{i=m_1+1}^{m_2}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p} - \sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &+ \sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} + \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &+ \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} \\ &- \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \end{aligned} \tag{15}$$

For the sake of clarity, this expression explicitly includes terms that are added and subtracted as part of the rearrangement. One additional step then yields:

$$\begin{aligned}\Delta G_{two\ sites} &= \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} + \sum_{i=1}^{m_1'}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} \\ &+ \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_1'+m_2-m_1} q_i'\phi_{ji}'^{p} - \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &- \sum_{i=1}^{m_1}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p} - \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_1+m_2-m_1} q_i\phi_{ji}^{p} \\ &+ \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} \\ &+ \sum_{i=m_1'+1}^{m_2'}\sum_{j=m_2'+1}^{N''} q_i'\phi_{ji}'^{p} + \sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1} q_i'\phi_{ji}'^{p} \\ &- \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} \\ &- \sum_{i=m_1+1}^{m_2}\sum_{j=m_2+1}^{N} q_i\phi_{ji}^{p} - \sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &+ \frac{1}{2}\sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} + \frac{1}{2}\sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1'} q_i'\phi_{ji}'^{p} \\ &- \frac{1}{2}\sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} - \frac{1}{2}\sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \\ &- \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_1'+m_2-m_1} q_i'\phi_{ji}'^{p} + \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_1+m_2-m_1} q_i\phi_{ji}^{p} \\ &- \sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1} q_i'\phi_{ji}'^{p} + \sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \end{aligned} \tag{16}$$

Comparison of eq. (16) with eq. (11) shows that the first two lines of eq. (16) equal the change in electrostatic free energy when the first group is ionized with the rest of the neutral protein. The second pair of lines represents the corresponding term for the second group. The third pair of lines, therefore, represents the effective interaction energy between the two groups:

$$\begin{aligned}\Delta G_{12} &= \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} - \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_2} q_i\phi_{ji}^{p} \\ &- \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_1'+m_2-m_1} q_i'\phi_{ji}'^{p} + \sum_{i=1}^{m_1}\sum_{j=m_1+1}^{m_1+m_2-m_1} q_i\phi_{ji}^{p} \\ &- \sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1} q_i'\phi_{ji}'^{p} + \sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_i\phi_{ji}^{p} \end{aligned} \tag{17}$$

Cancellation of the second and fourth terms on the right-hand side of this equation, and use of the reciprocal relation $q_i\phi_{ji} = q_j\phi_{ij}$, yields the following form for the interaction term:

$$\begin{aligned}\Delta G_{12} &= \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_2'} q_i'\phi_{ji}'^{p} - \sum_{i=1}^{m_1'}\sum_{j=m_1'+1}^{m_1'+m_2-m_1} q_i'\phi_{ji}'^{p} \\ &- \sum_{i=m_1'+1}^{m_2'}\sum_{j=1}^{m_1} q_j'\phi_{ij}'^{p} + \sum_{i=m_1+1}^{m_2}\sum_{j=1}^{m_1} q_j\phi_{ij}^{p} \end{aligned} \tag{18}$$

Finally, because the potentials in the second and third terms are due to the second and first groups, respectively, in their neutral forms, it is possible to write:

$$\Delta G_{12} = \sum_{i=1}^{m_1'} q_1' \phi_i'^{p} - \sum_{i=1}^{m_1'} q_i' \phi_i^{p} - \sum_{j=1}^{m_1} q_j \phi_j'^{p} + \sum_{j=1}^{m_1} q_j \phi_j^{p} \equiv \gamma_1 \gamma_2 \Psi_{12} \quad (19)$$

where the ϕ_i are the potentials at group one due to the charges in group two in the ionized (prime) and neutral (unprimed) states. This expression is directly related to that previously given without derivation by Yang and co-workers.[4] Eq. (19) is useful because it shows that the same set of ϕ_j and ϕ_j' values used in eq. (11), and thus the same pair of FDPB calculations, suffice for computation of the effective interaction energy as well. Because superposition holds for the linearized PB equation, which is the form used here, analogous expressions will yield the interactions in a real protein with more than two ionizable groups.

To conclude, all the required energy terms may be obtained by a series of four FDPB calculations for each ionizable group: two for the group in a model compound, and two for the group in its protein environment. In each case, one of the two calculations is for the group in its neutral form, and the other is for the group in its ionized form. These calculations yield the potential at every charge-bearing site in the molecule. These potentials are used to compute the energy of ionizing each group by itself, and the interaction of each ionizable group with each other group.

IMPLEMENTATION

The model presented above requires a minimum of four UHBD calculations for each titratable group: two for the protein environment and two for the model compound environment. Each calculation is carried out with charges on the titratable group in question appropriate to either the neutral or ionized state, and with all other charges set to zero. Because of the use of the "focusing" method,[5] more than four calculations are actually carried out for each group. However, the focusing method saves computer time by permitting the use of less extensive finite-difference grids.

The model compounds here are hypothetical compounds with an ionizable side chain and neutral carboxy and amino groups, having the same conformation as the corresponding group in the protein. Similar constructs are used for the N- and C-terminal except that the carboxy (for C-terminal) and amino (for N-terminal) groups are treated as ionizable along with their potentially ionizable side chains.

This set of calculations results in an output file containing all of the energies required to solve for the pH-dependent properties of interest. The form and content of the file are as follows:

$$\begin{array}{llll} M & & & \\ pK_{1,model} & \gamma_1 & \Delta\Delta G_1 & 1 \\ \Psi_{1,2} & & & \\ \vdots & & & \\ \Psi_{1,M} & & & \\ pK_{2,model} & \gamma_2 & \Delta\Delta G_2 & 2 \\ \Psi_{2,3} & & & \\ \vdots & & & \\ \Psi_{2,M} & & & \\ \vdots & & & \\ pK_{M,model} & \gamma_M & \Delta\Delta G_M & M \end{array}$$

The first line contains the number of ionizable sites, M. Subsequent lines are organized in blocks whose first line contains information about the group itself, and whose subsequent lines contain information about its interactions with other groups. The first line of each block contains the model compound pK_a of the group ($pK_{1,model}$), the type of the group (γ_1), the electrostatic free energy difference for ionization of the group in the protein with all other sites in neutral state relative to the same change in the model compound ($\Delta\Delta G_1$), and the index of the group. The remaining $M - i$ lines of each block contain the absolute values of the effective interaction potentials (Ψ_{ij}) with the remaining $M - i$ sites ($j = i + 1, \ldots, M$).

The data in this file may be used as input either to the HYBRID program[16] for rapid evaluation of mean charges; pK_as and electrostatic energies for proteins; or to the DOPS program, a modified version of a Monte Carlo procedure described earlier.[18,19] The latter is much slower, but provides results not only for mean charges and pK_as but also for fluctuations of charges. It also provides a list of the most stable protonation states found

during the Monte Carlo search. Another Monte Carlo program described elsewhere[20,21] is worth mention here as it has been used to carry out the same analysis, although the algorithm used differs significantly from that described below.

The DOPS program uses the Metropolis algorithm for sampling protonation states.[22] The equilibrium distribution is generated based on the free energy of states. The initial state is always the protein with all sites protonated. This is done because starting from a state closer to the equilibrium distribution (e.g., initial distribution based on intrinsic pK_as and pH values) may make it more difficult to find a desired number of low energy states using the algorithm employed in the program. Given a starting state, then, the Gibbs free energy is computed by use of eq. (3). A random number in the range {0 − 1} is then picked for each ionizable group in the protein, and the ionization state is switched for each group whose random number is greater than some cutoff value s. Thus, an average of $(1 - s)M$ groups are toggled at each step. The new ionization state is then accepted or rejected depending upon its free energy, again calculated by use of eq. (3), using the usual Metropolis criterion. Because the starting state is always far from equilibrium, the initial 10,000 states are discarded, and the subsequent 100,000 are used in computing thermodynamic averages. The value of s is adjusted so that approximately 50% of new states are accepted.

PROTEIN STRUCTURES AND PARAMETERS

The method described above has been applied to four proteins whose structures have been solved crystallographically: hen egg white lysozyme (triclinic form, Protein Data Bank[13] Accession Code 2LZT[23]); bovine pancreatic trypsin inhibitor (BPTI) (4PTI[24]); turkey ovomucoid third domain (1PPF[25]); and bovine pancreatic Ribonuclease A (3RN3[26]). In the case of Ribonuclease A, the structural file provides two conformations of His-119 residue. Form B is taken for calculations, because it gave better agreement with experimental data for the single-site method.[5]

The solvent dielectric constant, ionic strength, and temperature are chosen to fit the conditions of experimental studies. All other parameters used with the single-site method are the same as described previously.[5] Full-charge calculations were carried out with the atomic charges and radii of the PARSE parameter set, developed by Honig and co-workers.[27] This parameter set is attractive because, when used with a solvation model based on the PB equation, it yields accurate solvation energies for a variety of small molecules. Values of $pK_{a_{model,i}}$ have previously been tabulated.[5] The protein dielectric boundary is taken to be the Richards probe-accessible surface,[28] computed with a probe of radius 1.4 Å, and an initial dot density of 500 per atom.[29]

The present calculations require that coordinates be established for hydrogen atoms. In the present study, polar and aromatic hydrogens are added with the HBUILD command[30] of CHARMm.[21] For the single-site model, all titratable groups are treated as neutral at this stage. For the full-charge model, all are treated as fully protonated. The positions of the hydrogens are further optimized by 500 steps of steepest descent energy minimization. By default, the protonation site for all carboxylic acids is the second oxygen atom in the structure file. No default exists for histidines and so the user is required to specify which nitrogen, ND1 or NE2, is the one that deprotonates in the neutral form. In all calculations reported here, NE2 is taken to be deprotonatable.

Results and Discussion

pK_A VALUES

Tables I through IV present pK_as for the four proteins listed above. Results are presented for the full-charge model and the single-site model; experimental results are included as well. The final lines of each table provides the root-mean-square deviations of the computed values from the available experimental values.

It was discovered that, for the full-group procedure, a change in the protonation site of aspartic residues from OD2 to OD1, and of glutamic residues from OE2 to OE1, sometimes had a significant effect upon the agreement of the computed pK_as with experiment. Therefore, the possibility of making these changes was added as an option to the procedure. Such changes were tested for lysozyme, BPTI, and ovomucoid, and the results are given as the second column of full-charge data.

As summarized by the root-mean-square deviations listed in the tables, pK_as computed by the full-charge method are generally more accurate in each case than those computed by the single site

TABLE I.
Lysozyme — Computed and Experimental[1,32] Ionization Constants.

Residue	Single[a]	Full-A[b]	Full-B[c]	Exptl.
TERN-1	5.4	5.6	5.6	
LYS-1	9.1	10.1	10.1	10.6
ARG-5	11.6	12.8	12.8	
GLU-7	3.2	2.9	2.8	2.9
LYS-13	10.4	10.7	10.7	10.3
ARG-14	11.7	12.4	12.4	
HISA-15	0.3	5.1	5.1	5.4
ASP-18	2.8	2.7	2.7	2.7
TYR-20	15.1	12.4	12.4	10.3
ARG-21	11.9	12.7	12.7	
TYR-23	12.6	9.7	9.7	9.8
LYS-33	10.9	10.9	10.9	10.4
GLU-35	6.2	3.5	5.6	6.2
ARG-45	10.3	12.2	12.2	
ASP-48	2.3	1.5	1.5	1.6
ASP-52	4.0	4.3	3.6	3.7
TYR-53	23.5	20.5	20.5	12.1
ARG-61	13.5	13.8	13.7	
ASP-66	2.4	1.9	1.9	0.9
ARG-68	14.1	13.0	13.0	
ARG-73	12.0	12.5	12.5	
ASP-87	3.4	2.4	2.4	2.1
LYS-96	10.6	13.6	13.6	10.7
LYS-97	10.8	11.2	11.2	10.1
ASP-101	5.4	4.4	4.4	4.1
ARG-112	11.5	11.2	11.2	
ARG-114	12.3	13.3	13.3	
LYS-116	8.8	9.3	9.3	10.2
ASP-119	3.8	3.7	3.7	3.2
ARG-125	13.0	12.9	12.9	
ARG-128	12.2	12.2	12.2	
TERC-129	2.8	3.1	3.1	2.8
rms	3.2	2.2	2.1	
rms (no res #17)	1.9	1.1	1.0	

Computations are for 293 K, 150 mM ionic strength.
[a] Single-site results.
[b] Full-charge calculations, with default protonation sites for carboxylic acids.
[c] Results when OE1 of GLU-35 is treated as its protonatable atom.

TABLE II.
BPTI-Computed and Experimental[33-35] Ionization Constants.

Residue	Single[a]	Full-A[b]	Full-B[c]	Exptl.
TERN-1	6.2	7.2	7.2	8.1
ARG-1	12.5	18.1	18.1	
ASP-3	3.3	3.3	3.4	3.0
GLU-7	5.0	4.7	5.4	3.7
TYR-10	10.5	9.9	9.9	
LYS-15	10.3	10.4	10.4	10.6
ARG-17	12.1	12.2	12.2	
ARG-20	10.7	13.1	13.1	
TYR-21	10.4	10.1	10.1	
TYR-23	16.5	11.3	11.3	
LYS-26	10.4	10.4	10.4	10.6
TYR-35	15.4	9.4	9.5	
ARG-39	12.1	12.2	12.2	
LYS-41	9.7	10.3	10.2	10.8
ARG-42	12.1	12.9	13.0	
LYS-46	10.0	10.0	10.0	10.6
GLU-49	3.6	3.5	3.8	3.8
ASP-50	2.8	2.5	2.3	3.4
ARG-53	13.5	12.9	12.9	
TERC-58	3.7	3.8	3.8	2.9
rms	0.9	0.7	0.8	

Calculations are for 293 K, 150 mM ionic Strength.
[a] Single-site results.
[b] Full-charge results with default protonation sites for carboxylic acids.
[c] Full-charge results when all ASP and GLU side chains are considered protonatable at OD1 and OE1, respectively.

method, for an assumed protein dielectric constant of 4. The cumulative root-mean-square deviation for the full-charge model is 1.3 pK_a units when the default set of carboxylic acid protonation sites is used ("Full-A" column of each table). Adjustment of the carboxylic protonation sites ("Full-B" columns) leads to a root-mean-square deviation of 1.1 pK_a units. The overall result for the single-site model is 2.5 pK_a units. Changing the protonation site of carboxylic acids has a negligible effect upon the results of the single-site model, however (results not shown). This represents a substantial improvement. This result is potentially practical, and is also inherently interesting. It suggests that accurate pK_as may be computed even when a low protein dielectric constant is used, and thus that the model of a protein as a low-dielectric body may be basically correct.

On the other hand, the persistence of some large deviations from experimental pK_a values obtained with the full-group procedure indicate that there are important factors still missing. One possibility is that the crystal structure does not correspond to the solution structures for which experimental data are obtained. Examination of this possibility is in progress. It is also of interest that the results of the full-charge model are quite sensitive to the precise location of hydrogens. For example, the shifted pK_a of Glu-35 of lysozyme is accurately reproduced only when Glu-35 OE1 is treated as the

TABLE III.
Ovomucoid Third-Domain-Computed and Experimental[36] Ionization Constants.

Residue	Single	Full-A[a]	Full-B[b]	Exptl.
TERN-1	6.2	5.3	5.3	
ASP-7	2.9	3.2	3.2	2.6
GLU-10	3.5	3.5	3.4	4.1
TYR-11	17.7	9.9	9.9	
LYS-13	10.4	12.6	12.5	
GLU-19	2.7	2.0	2.9	3.2
TYR-20	14.9	10.8	10.8	
ARG-21	12.3	12.4	12.4	
ASP-27	6.8	7.2	3.7	2.3
LYS-29	11.5	11.6	11.6	
TYR-31	20.8	14.0	14.1	
LYS-34	10.5	14.4	14.4	
GLU-43	4.4	4.2	4.5	4.7
HISA-52	4.5	6.2	6.4	
LYS-55	10.1	10.5	10.5	
TERC-56	2.8	2.9	2.9	2.3
rms	1.9	2.1	0.7	

Calculations are for 308 K, 10 mM ionic strength.
[a] Full-charge results with default protonation sites.
[b] Full-charge results with ASP-27 OD1 and GLU-19 OE1 considered protonatable.

TABLE IV.
Ribonuclease — Computed and Experimental[37] Ionization Constants.

Residue	Single	Full	Exptl.
TERN-1	6.7	6.3	7.6
LYS-1	10.4	10.5	
GLU-2	2.3	2.0	2.8
LYS-7	8.9	9.6	
GLU-9	4.7	4.7	4.0
ARG-10	13.4	17.6	
HISA-12	−0.4	4.2	5.8
ASP-14	6.0	0.0	2.0
TYR-25	26.5	19.6	
LYS-31	9.6	9.7	
ARG-33	14.1	13.4	
LYS-37	10.8	10.9	
ASP-38	2.8	3.2	3.1
ARG-39	11.9	12.5	
LYS-41	5.2	9.7	
HISA-48	−8.3	9.7	6.3
GLU-49	7.0	5.4	4.7
ASP-53	5.3	3.9	3.9
LYS-61	10.2	10.4	
LYS-66	11.5	12.6	
TYR-73	13.6	14.0	
TYR-76	11.3	10.3	
ASP-83	2.5	1.7	3.5
ARG-85	12.4	13.2	
GLU-86	4.6	4.7	4.1
LYS-91	10.4	11.2	
TYR-92	11.9	10.8	
TYR-97	25.4	17.6	
LYS-98	10.2	10.2	
LYS-104	10.7	10.7	
HISA-105	4.9	5.6	6.6
GLU-111	4.3	4.2	3.5
TYR-115	18.0	10.8	
HISA-119	5.7	6.3	6.1
ASP-121	2.2	1.1	3.1
TERC-124	2.3	2.3	2.4
rms	4.2	1.4	

[a] Calculations are for 308 K, 200 mM ionic strength.

protonation site (compare "Full-A" and "Full-B" results in Table I). This seems physicochemically reasonable, and suggests that further improvement may be obtained if more attention is paid to this feature of the model. In particular, an objective way of treating the rotational isomerization of the neutral carboxylic acids and of histidine will be quite important. It seems likely that special attention should be paid to ionizable groups that form salt-bridges. A partial treatment of this problem has been presented by Bashford and co-workers.[21]

DISTRIBUTION OF PROTONATION STATES

There are a number of situations in which theoretical predictions of the ionization states of protein residues would be of value. Examples include the setup of molecular dynamics simulations and of protein–ligand binding calculations. The DOPS program can be used to determine highly occupied protonation states, given a set of electrostatic energies computed as outlined above. Table V presents sample results of this type for BPTI, based upon the full-charge model. The five most stable ionization states at pH 7.0 are given, along with their energies. It is of particular interest that two states of similar low energy are identified by this procedure. This is directly related to the prediction that the pK_a of the N-terminal amine is close to 7. A more complex cluster of three states of similar energy lies about 12.6 kJ/mol higher.

Conclusions

This work describes a novel and efficient method for computing the ionization constants of titratable

TABLE V.
BPTI — Charges of Individual Sites (e), Total Charges (e), and Electrostatic Energies (kJ / mol) of Five Lowest Energy Protonation States Found During Monte Carlo Search at pH 7.0[a]

Residue	Charges of Titratable Residues				
TERN-1	1.	0.	1.	0.	1.
ARG-1	1.	1.	1.	1.	1.
ASP-3	−1.	−1.	−1.	−1.	−1.
GLU-7	−1.	−1.	0.	0.	−1.
TYR-10	0.	0.	0.	0.	0.
LYS-15	1.	1.	1.	1.	1.
ARG-17	1.	1.	1.	1.	1.
Arg-20	1.	1.	1.	1.	1.
TYR-21	0.	0.	0.	0.	0.
TYR-23	0.	0.	0.	0.	0.
LYS-26	1.	1.	1.	1.	1.
TYR-35	0.	0.	0.	0.	−1.
ARG-39	1.	1.	1.	1.	1.
LYS-41	1.	1.	1.	1.	1.
ARG-42	1.	1.	1.	1.	1.
LYS-46	1.	1.	1.	1.	1.
GLU-49	−1.	−1.	−1.	−1.	−1.
ASP-50	−1.	−1.	−1.	−1.	−1.
ARG-53	1.	1.	1.	1.	1.
TERC-58	−1.	−1.	−1.	−1.	−1.
Charge	6.	5.	7.	6.	5.
Energy	−367.4	−366.1	−354.8	−354.4	−354.4

[a] See Table II for parameters.

residues in proteins, using a detailed charging model. The method is based upon the widely distributed program UHBD.

The most important result of this work is the significant improvement in the prediction of pK_a's in proteins with a full-charge methodology in comparison to the previous single-site methodology[5] while using low values for the dielectric constant of the protein. Moreover, usage of a more detailed description of the ionizable groups focuses attention upon the problem of the optimal assignment of hydrogen coordinates prior to execution of the titration calculations. This presents new avenues of exploration for the titration methodology. The high degree to which these calculations are automated will facilitate further methodologic development.

Acknowledgments

This work was supported in part by the National Institutes of Health, the National Science Foundation, the Metacenter Program of the National Science Foundation supercomputer centers, and Molecular Simulations, Inc. Certain commercial equipment or materials are identified in this article in order to specify the methods adequately. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment indentified are necessarily the best available for the purpose.

References

1. D. Bashford and M. Karplus, *Biochemistry*, **29**, 10219–10225 (1990).
2. D. Bashford and M. Karplus, *J. Phys. Chem.*, **95**, 9556–9561 (1991).
3. C. Lim, D. Bashford, and M. Karplus, *J. Phys. Chem.*, **95**, 5610–5620 (1991).
4. A.-S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig, *Prot. Struct. Funct. Genet.*, **15**, 252–265 (1993).
5. J. Antosiewicz, J. A. McCammon, and M. K. Gilson, *J. Mol. Biol.*, **238**, 415–436 (1994).
6. M. K. Gilson and B. H. Honig, *Biopolymers*, **25**, 2097–2119 (1986).
7. T. Simonson, D. Perahia, and A. T. Brunger, *Biophys. J.*, **59**, 670–690 (1991).
8. P. E. Smith, R. M. Brunne, A. E. Mark, and W. F. van Gunsteren, *J. Phys. Chem.*, **97**, 2009–2014 (1993).
9. J. Warwicker and H. C. Watson, *J. Mol. Biol.*, **157**, 671–679 (1982).
10. M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, *Comp. Phys. Commun.*, **62**, 187–197 (1991).
11. J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon, *Comput. Phys. Commun.*, **91**, 57 (1995).
12. D. Bashford and K. Gerwert, *J. Mol. Biol.*, **224**, 473–486 (1992).
13. F. C. Bernstein, T. F. Koettzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. J. Tasumi, *J. Mol. Biol.*, **123**, 557–594 (1977).
14. C. Tanford and J. G. Kirkwood, *J. Am. Chem. Soc.*, **79**, 5333–5339 (1957).
15. D. Bashford and M. Karplus, *Biochemistry*, **9**, 327–335 (1990).
16. M. K. Gilson, *Proteins*, **15**, 266–282 (1993).
17. J. A. Schellman, *Biopolymers*, **14**, 999–1018 (1975).
18. J. Antosiewicz and D. Porschke, *Biophys. J.*, **68**, 655–664 (1995).
19. J. Antosiewicz and D. Porschke, *Biochemistry*, **28**, 10077–10078 (1989).
20. P. Beroza, D. R. Fredkin, M. Y. Okamura, and G. Feher, *Proc. Natl. Acad. Sci. USA*, **88**, 5804–5808 (1991).
21. D. Bashford, D. A. Case, C. Dalvit, L. Tennant, and P. E. Wright, *Biochemistry*, **32**, 8045–8056 (1993).

22. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087–1092 (1953).
23. M. Ramanadham, L. C. Sieker, and L. H. Jensen, *Acta Crystallogr. (Sect. A)*, **37C**, 33 (1981).
24. M. Marquart, J. Walter, J. Desenhofer, W. Bode, and R. Huber, *Acta Crystallogr. (Sect. B)*, **39**, 480 (1983).
25. W. Bode, A.-Z. Wei, R. Huber, E. Meyer, J. Travis, and S. Neumann, *EMBO J.*, **5**, 2453–2458 (1986).
26. B. Howlin, D. S. Moss, and G. W. Harris, *Acta Crystallogr. (Sect. A)*, **45**, 851 (1989).
27. D. Sitkoff, K. A. Sharp, and B. Honig, *J. Phys. Chem.*, **98**, 1978–1988 (1994).
28. F. M. Richards, *Ann. Rev. Biophys. Bioeng.*, **6**, 151–176 (1977).
29. M. K. Gilson, K. A. Sharp, and B. H. Honig, *J. Comput. Chem.*, **9**, 327–335 (1988).
30. A. T. Brunger and M. Karplus, *Prot. Struct. Funct. Genet.*, **4**, 148–156 (1988).
31. B. R. Brooks, R. E. Bruccoleri, B. D. O. D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187–217 (1982).
32. K. Bartik, C. Redfield, and C. M. Dobson, *Biophys. J.*, **66**, 1180–1184 (1994).
33. L. R. Brown, A. D. Marco, G. Wagner, and K. Wüthrich, *Eur. J. Biochem.*, **62**, 103–107 (1976).
34. L. R. Brown, A. D. Marco, R. Richarz, G. Wagner, and K. Wüthrich, *Eur. J. Biochem.*, **88**, 87–95 (1978).
35. R. Richarz and K. Wüthrich, *Biochemistry*, **17**, 2263–2269 (1978).
36. W. Schaller and A. D. Robertson, *Biochemistry*, **34**, 4714–4723 (1995).
37. M. Rico, J. Santoro, C. Gonzalez, M. Bruix, and J. L. Neira, Solution structure of bovine pancreatic ribonuclease A and ribonuclease–pyrimidine nucleotide complexes as determined by ^{1}H NMR, in C. M. Cuchillo, R. de Llorens, M. V. Nogués, and X. Parés, Eds., *Structure, Mechanism and Function of Ribonucleases. Proceedings of the 2nd International Meeting, Held in Sant Feliu de Guíxols, Girona, Spain, 1990*, pp. 9–14, Bellaterra, Spain, Departament de Bioquímica i Biologia Molecular and Institut de Biologia Fonamental Vicent Villar Palasí, Universitat Autònoma de Barcelona, 1991.