pH-Dependence of Protein Stability: Absolute Electrostatic Free Energy Differences between Conformations†

Michael Schaefer, Michael Sommer, and Martin Karplus*

Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, and Laboratoire de Chimie Biophysique, Institut le Bel, Université Louis Pasteur, 4 rue Blaise Pascal, 67000 Strasbourg, France

Received: September 26, 1996; In Final Form: December 5, 1996⊗

A method for calculating the absolute electrostatic free energy of a titrating system as a function of pH is proposed, and a concise formula for the free energy is presented. Based on the theory of linked functions, the electrostatic free energy is calculated by integration of the titration curve. The approach uses pH = ∞, at which the system is in the unprotonated state, as the reference pH for the integration. The finite-difference Poisson−Boltzmann method is used for the electrostatic free energy, and the titration curve is obtained by the Monte Carlo approach of Beroza *et al.*[1] The method is applied to the native and denatured state of hen egg-white lysozyme in aqueous solution. A dielectric constant of 20 is assigned to the protein interior, in accord with the work of Antosiewicz *et al.*[2] X-ray structures are used for the native state, and an extended β-structure is used for the unfolded reference state; good agreement with experiment is obtained. Comparison of the results for the extended β-structure and for a "null" model of noninteracting sites for the unfolded state shows significant differences. This indicates that there are important interactions between titrating sites in the unfolded state, in agreement with recent experimental estimates by Oliveberg *et al.*[3] A number of structures obtained by *in vacuo* minimization of the native and unfolded protein are compared, and it is shown that the absolute pH-stability curve has a strong conformational dependence, although the relative stability curve does not. Calculations of absolute free energy differences, rather than relative changes as a function of pH, are of general interest; for example, they can be used to study the binding energy of ligands involving small titratable compounds.

1. Introduction

The pK_a's and pH-dependent ionization states of titratable groups play an essential role in the catalytic activity[4] and stability of proteins.[5−8] Approaches for calculating pK_a's, protein titration curves, and pH-dependent protein stability have, therefore, been the subject of theoretical studies for many decades.[9−12] All theoretical approaches have in common the assumption that the observed shift of the pK_a of a titrating amino-acid side chain in a protein, relative to the pK_a of the isolated amino acid, is dominated by electrostatic interactions, namely, the interactions with other titrating sites, with the surrounding protein residues, and with the aqueous environment including salts. This hypothesis is supported by the experimental observations of pK_a-shifts due to point mutations[13,14] and due to changes in the ionic strength of the solution,[14,15] as well as by the fact that the titration curve of an unfolded protein is approximately equal to the sum of the titration curves of the constituent sites.[16,17]

For the calculation of the pK_a's, titration curve, and pH-dependent stability of a protein, one or more structures must be selected to represent the different states. For determining the pH-dependence of the stability, which is of primary interest in this paper, the crystallographically determined native structure and an appropriate model of the unfolded protein are required. Some approaches have avoided the requirement for an unfolded structure by restricting themselves to the calculation of relative stabilities.[2,18] By "relative stability", we mean that they have assumed that the unfolded state of a protein can be treated as a state where all sites of the protein possess their standard pK_a's and titrate independently ("zero interaction model"[18] or "null model"[2]). With this assumption, no unfolded reference structure with its associated Coulomb and electrostatic solvation energy is required. The stability is then calculated relative to an arbitrary pH reference value, usually pH = 0, where the free energy difference $\Delta\Delta G$ is set to zero. The main advantage of the null model is that the titration curve and the free energy can be calculated from simple analytic formulas. Another reason for the use of the null model in previous work is that the geometric dimension of unfolded protein models, e.g., an extended structure, leads to accuracy problems in the calculation of electrostatic energies with the finite-difference Poisson−Boltzmann method (see below).

The stability calculation described in this paper is concerned with the effect of pH on the absolute stability. It can be divided in three steps: (1) for each structure, the relevant electrostatic energies must be calculated ("electrostatic energy calculation"); (2) the pH-dependent average charge states of the titrating sites must be determined ("titration calculation"); and (3) the results of the titration calculations for the different conformers are used to obtain the pH-dependent free energy difference ("stability calculation").

Significant theoretical and algorithmic improvements in the methods for calculating the electrostatic energy[2,12,19−21] and titration curves[1,21−23] have been reported in recent years. They are based on the finite-difference Poisson−Boltzmann (FDPB) method[24,25] and the finite-element method[26,27] for the calculation of electrostatic energies. The FDPB method is used in this work. It allows us to include the effects of the solvent and ionic strength on the self-energy and interaction energy of the charges, and it has been employed successfully in numerous

† This work was partly supported by the National Institutes of Health. M.Sch. is currently supported by a research training grant within the Biotechnology Programme of the European Community.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1997.

S1089-5647(96)02972-0 CCC: $14.00

studies of proteins.[28−31] A comparison between different methods for calculating electrostatic energies in the framework of pK_a calculations will be presented elsewhere.[32]

In an exact treatment of the titration curve, which involves the calculation of the generating (partition) function for all possible charge states of the system, the computing time increases exponentially with the number of titratable sites.[1,22] Using the fastest computers that are available today, the exact treatment is limited to systems with up to 25−30 titrating sites. Various approximation methods for the calculation of titration curves for large systems have, therefore, been proposed, e.g., the reduced site (RS) approximation,[22] a Monte Carlo (MC) titration method,[1] and a cluster method.[23] The major disadvantage of the RS approximation is that it is limited to systems where less than 25−30 sites (i.e., the number of sites that can be treated exactly) have similar pK_a's. The cluster method has been shown to yield titration curves at low computational expense (several seconds on a Silicon Graphics Iris 4D/VGX for systems with up to 123 sites[23]); however, its accuracy in comparison with other titration methods has not been fully tested. Consequently, we use the MC titration method, which has the advantage of providing well-defined error bounds, while the required computation time, which depends quadratically on the number of sites, is still small when compared with the time that is necessary to perform the electrostatic energy calculations with the FDPB method.[1,23]

The calculation of a protein pH-stability curve based on the titration curves of the native and unfolded states is possible because, at a given pH, the derivative of the free energy difference between the two states is proportional to the difference between the average charges of the conformers, $\partial(\Delta G)/\partial(\mathrm{pH}) \propto \Delta\langle Q\rangle$. This simple relation between the free energy and the titration curve of a protein results from the theory of linked functions.[33−35] It follows that the stability curve of a protein can be obtained by integrating the difference between the titration curves of the native and unfolded states,[2,11,18] an approach that is termed the "titration curve integration" method in the remainder of this paper.

The method for calculating the absolute, pH-dependent stability of proteins uses an extended structure of the protein as the unfolded reference state. The requirement of defining an arbitrary reference pH and free energy difference in the titration curve integration method is eliminated by integrating from pH = $+\infty$, where the partition function of both the native and the unfolded structure corresponds to the fully unprotonated state. The integration from pH = $-\infty$ where the protonated state dominates would also be possible and should lead to identical results, provided that the electrostatic free energies of the protonated and unprotonated states are calculated with sufficient accuracy. The integration from pH = ∞ is justified because the calculated free energy is independent of the integration path over which it is determined, given a well-defined reference state. In this respect, the fact that a native protein structure is not stable under very low or high pH conditions is irrelevant.

Contributions, other than the electrostatic term, to the free energy difference between the conformational states of a molecule are significant. They include entropic contributions, van der Waals energies, and the hydrophobic effect.[18,36] They are not considered here because this work is aimed exclusively at demonstrating that it is possible to calculate the pH-dependent electrostatic contribution to the absolute free energy. Further, as already mentioned, the pH-dependence is expected to be dominated by the electrostatic contribution. Although the theory presented in section 2 provides a basis for considerations of pH-dependent protein stability that include all energetic and entropic contributions, such an analysis is beyond the scope of this paper.

Also, a methodology for calculating pK_a's that accounts for conformational changes of a protein in response to changes in the ionization state would be more correct than the calculation of site−site interaction energies for a single conformer, as pointed out by Bashford and Karplus.[12] Approximations to take account of multiple conformers have been introduced.[18,37] However, their accuracy is difficult to assess because no Boltzmann weighting of the different conformers associated with a given structure was made.

For numerical calculations, it is necessary to choose a dielectric constant for the protein interior. Although various choices have been made,[2,12,23] we follow Antosiewicz *et al.*,[2] who have found that an electrostatic model where the protein interior is assigned a dielectric constant of $\epsilon_i = 20$ yields good results; they suggest that such a large dielectric constant accounts implicitly for the conformational flexibility of a protein; that is, in conjunction with the use of a single conformer to represent the native state of a protein, this approach has been shown to yield good agreement between calculated and experimental pK_a's for a number of proteins.[2] It is also possible that its success is based on the fact that most of the titrating groups in the proteins that were studied are solvated. The importance of conformational equilibria to the calculation of pK_a's and the accuracy of approaches that account for conformational change implicitly and explictly will be addressed in a separate study.[38]

It should be pointed out that the theory presented in this paper is valid at thermodynamic equilibrium. Use of continuum electrostatics to take account of the solvent effect in molecular dynamics calculations, for example, would require consideration of the time scale of the protonation and deprotonation reaction, relative to the motion of interest.[39]

In section 2, we develop the theory for calculating pK_a's, titration curves, and stability curves based on the assumption that all sites have only two protonation states, protonated and unprotonated ("two-state model"). The pH and protonation state-dependent free energy of a titrating system is expressed in a form that is more concise than formulas available from the literature; some details are given in Appendix A. It allows straightforward extension of the theory to include conformational averages and more complex titrating sites. In Appendix B, formulas for the general case of multiple (>2) protonation states of the sites are given as a basis for future work.[40] In section 3, we describe the methods used for preparing the native and unfolded input structures, for the calculation of electrostatic energies with the FDPB program UHBD,[41] and for the titration calculation with a Monte Carlo titration program provided by Beroza *et al.*[1]

In section 4, we apply the method for calculating absolute pH-dependent free energy differences to the protein lysozyme. Lysozyme is used because experimental data are available and it has been the object of many calculations of electrostatic properties. It is shown that the agreement between the calculated and experimental pK_a's of lysozyme is best when a dielectric constant of $\epsilon_i = 20$ is used for the protein interior, both for the crystallographic structure and for minimized structures of the protein. We present the calculated relative stability curve of lysozyme, using the null model and several explicit models of the unfolded state. For the explicit unfolded models, absolute stability curves are also calculated and compared with the relative stability curves. A strong dependence of the absolute stability, as compared to the relative

stability, on the degree of minimization of the native and unfolded structures is demonstrated.

We discuss extensions of the present approach in section 5. Possible solutions for the problem of conformational equilibria with respect to pK_a's and stability curves are considered. The accuracy and self-consistency of the methodology related to the use of the FDPB method for calculating electrostatic energies are discussed, and the applicability of the present approach to calculating the absolute, pH-dependent free energy of binding between molecules with titrating groups is outlined. In Appendix C, the problem of inconsistencies in the present treatment of nonbonded Coulomb interactions, which is common to all published work on protein pK_a's employing the FDPB method,[2,12,21] is discussed in detail.

2. Theory

2.1. Definitions. We define the standard pK_a of a titrating side chain in a protein as the experimental pK_a of the amino acid ("X") as part of a small peptide (the standard peptide) with electrically neutral atom groups except for the site itself, e.g., the peptide Ala-X-Ala with N- and C-terminal blocking groups.[4,42,43] Correspondingly, the standard pK_a's of the N-terminal ammonium group and of the C-terminal carboxyl group are the experimental pK_a's of these groups in an otherwise electrically neutral peptide. The standard pK_a is denoted by pK_a^{stnd} in the following. The observed pK_a of a titrating atom group in a protein, termed the "effective pK_a", may differ from the standard pK_a by several pK-units. For example, the largest measured pK-shift for titrating sites in the proteins lysozyme,[12,44,45] ribonuclease A,[46] and myoglobin[47] is about 2.5 pK-units. A third pK_a that is commonly used in theoretical studies is the "intrinsic pK_a" of a site, pK_a^{intr}, which is the hypothetical pK_a of the site assuming that all other titrating sites in the system are fixed in their electrically neutral state. In a system with only one titrating group, the effective pK_a of the group is equal to the pK_a^{intr}.

For calculating the pH-dependent properties of a system, a choice of the titrating sites that are included in the analysis has to be made. In this choice, it is possible to exclude atom groups whose standard pK_a's are far from the pH-range of interest, e.g., the Ser hydroxyl group with $pK_a^{stnd} = 13.6$ in considerations of protein stability around pH = 7.

Each titrating site is composed of a set of atoms whose partial charges depend on the protonation state of the site, e.g., all side-chain atoms in the case of aspartate. The atoms of the system that are not part of any site are termed the "background atoms" or "background charges".[12] By definition, the background atoms are assumed to have pH-independent partial charges, even though the theory that is developed in this work makes use of infinite pH-limits; for example, if peptide N−H groups are excluded from the set of titrating sites, it is assumed that they are not ionized in the limit pH $\rightarrow \infty$. This does not cause any difficulties and it is understood that the pH-dependent properties of a system are defined relative to a chosen set of titrating sites.

2.2. Electrostatic Free Energy of a Protonation State. Given a protein with N titratable sites, we describe the protonation state by a vector $\bar{s}$ with N components s_1, ..., s_N. To derive an equation for the pH-dependent energy of protonation state $\bar{s}$, we consider the case where all sites have only the two ionization states "unprotonated" and "protonated", termed the "two-state model" in the following. In Appendix B, the formulas in this section are generalized to be applicable to the case of multiple states. If the components s_i are defined according to

$$s_i = \begin{cases} 0, & \text{if site } i \text{ is unprotonated} \\ 1, & \text{if site } i \text{ is protonated} \end{cases} \quad (1)$$

the pH-dependent free energy difference between the protein in protonation state $\bar{s}$ and the fully unprotonated state is[12]

$$\Delta G(\bar{s},\mathrm{pH}) = (\ln 10) k_B T \sum_{i=1}^{N} s_i (\mathrm{pH} - pK_{a,i}^{intr}) + \sum_{i<j}^{N} [q_i(s_i)\, q_j(s_j) - q_i(0)\, q_j(0)]\, W_{ij} \quad (2)$$

where $pK_{a,i}^{intr}$ is the intrinsic pK_a of site i, $q_i(s_i)$ is the net charge of site i in protonation state s_i, and $q_i(0)$ is the net charge of site i when it is unprotonated. The magnitude W_{ij} of the interaction between the sites i and j is addressed below. Equation 2 differs from the formula given in ref 12 by the subtraction of the term $\sum_{i<j} q_i(0)\, q_j(0)\, W_{ij}$, such that the energy of the unprotonated state ($s_i = 0$ for all i) is equal to 0. This is required since we defined $\Delta G(\bar{s},\mathrm{pH})$ as the energy relative to the unprotonated state. In the literature, the protonation state where all sites are electrically neutral is often used as the reference state for which $\Delta G(\bar{s},\mathrm{pH}) = 0$;[2,21,23] it can be shown that the corresponding equations for the energy as a function of the protonation state are equivalent to eq 2.

The energy W_{ij} is given by[21]

$$W_{ij} = E_{ij}(1,1) - E_{ij}(1,0) - E_{ij}(0,1) + E_{ij}(0,0) \quad (3)$$

where $E_{ij}(s_i,s_j)$ is the electrostatic interaction energy between sites i and j in the protonation states s_i and s_j, respectively. If the effect of the partial charges on sites i and j are included in E_{ij}, all four energy terms in eq 3 can be nonzero, as pointed out by Yang *et al.*[21] The rather complex expression for W_{ij} in eq 3 is required because the $pK_{a,i}^{intr}$ are defined as the pK_a's when all other sites are uncharged; that is, both sums in eq 2 involve interaction terms E_{ij} (see also eq 5). In Appendix A, it is shown that the definition of the $pK_{a,i}^{intr}$ and the energy W_{ij} ensures that for a given charge state $\bar{s}$ of the system the electrostatic interaction terms between all atom groups (sites and background atoms) are counted once and only once. The reason for the four terms and their signs in eq 3 is that the product $q_i(s_i)\, q_j(s_j)\, W_{ij}$ in eq 2 contributes only if both sites i and j are charged. Thus, the expression for W_{ij} includes the interaction energy $E_{ij}(s_i,s_j)$ with the correct (positive) sign for all combinations of anionic and cationic sites i and j ($q_i q_j = \pm 1$ if both sites are charged).

The electrostatic interaction energy between two sites i and j with atoms $i' \in \{i\}$ and $j' \in \{j\}$, respectively, can be determined by first calculating the potential $\phi_{(i,s_i)}(\vec{x})$ due to site i in charge state s_i throughout the protein. Given the potential, the interaction energy with site j in charge state s_j is

$$E_{ij}(s_i,s_j) = \sum_{j'} q_{(j',s_j)}\, \phi_{(i,s_i)}(\vec{x}_{j'}) \quad (4)$$

where $\vec{x}_{j'}$ is the position of atom j' in site j. In eq 4, the charges of the two sites and thus the interaction energy E_{ij} depend on the charge states of the sites, s_i and s_j.

By definition, the $pK_{a,i}^{intr}$ of site i in a protein differs from the pK_a^{stnd} of the site in the corresponding standard peptide (see section 2.1) by a pK-shift that is due to the electrostatic interaction with the protein environment (protein backbone, nontitrating side chains, titrating side chains in their uncharged

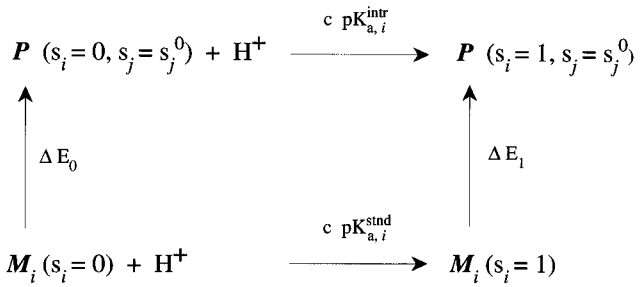

Figure 1. Electrostatic free energy for protonating site i as part of the protein ($\boldsymbol{P}$) with all other sites $j \neq i$ in their uncharged state s_j^0 and as part of the model compound M_i. The quantities ΔE_1 or ΔE_0 are the electrostatic free energy differences between the system with site i protonated or unprotonated and the protonated or unprotonated model compound M_i. From the cycle, it follows that $c\ (\mathrm{p}K_{\mathrm{a},i}^{\mathrm{intr}} - \mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}) = \Delta E_1 - \Delta E_0$, where $c = -(\ln 10)k_{\mathrm{B}}T$.

state) and the change in the interaction with the aqueous environment, i.e., the desolvation effect upon transfer of the site from the standard peptide to the protein. The isolated residue of site i is used as a model of the standard peptide ("model compound"), for which the pK_{a} is assumed to be equal to $\mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}$. This permits the use of empirical data for the $\mathrm{p}K_{\mathrm{a}}^{\mathrm{stnd}}$ of titrating sites in standard peptides or amino acid side-chain analogs. We follow such an approach in this study and write M_i for the model compound for site i.

As in the protein, we use the term "background atoms" for the atoms in M_i that are not part of the titrating site and thus have fixed charges. The background atoms in the protein and in the model compounds are formally assigned the site index 0 and the charge state $s_0 = 0$ in the following. For the interaction between the site i in charge state s_i and the background atoms in the protein (model compound), we can thus write $E_{i0}(s_i,0)$ ($E_{i0}^{M_i}(s_i,0)$) in a notation that is consistent with that for the electrostatic interaction between the titrating sites.

According to the thermodynamic cycle given in Figure 1, the pK-shift $\mathrm{p}K_{\mathrm{a},i}^{\mathrm{intr}} - \mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}$ of site i is given by

$$-(\ln 10)k_{\mathrm{B}}T(\mathrm{p}K_{\mathrm{a},i}^{\mathrm{intr}} - \mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}) = \Delta E_1 - \Delta E_0 = [E_{\mathrm{el}}(s_i{=}1, s_{j\neq i}{=}s_j^0) - E_{\mathrm{el}}^{M_i}(s_i{=}1)] - [E_{\mathrm{el}}(s_i{=}0, s_{j\neq i}{=}s_j^0) - E_{\mathrm{el}}^{M_i}(s_i{=}0)] \quad (5)$$

where we have written s_j^0 for the uncharged state of site j and where $E_{\mathrm{el}}(\bar{s})$ and $E_{\mathrm{el}}^{M_i}(s_i)$ are the electrostatic free energies of the protein and the model compound for site i, respectively. On the basis of the subdivision of the protein or model compound into groups of atoms corresponding to the titrating sites or site and the background atoms, these energies are given by

$$E_{\mathrm{el}}(\bar{s}) = (1/2)\sum_{i,j=0}^{N} E_{ij}(s_i,s_j) \quad (6)$$

$$E_{\mathrm{el}}^{M_i}(s_i) = (1/2)\sum_{k,l\in\{0,i\}} E_{kl}^{M_i}(s_k,s_l) \quad (7)$$

The sum in eq 7 is over all pairs of the indices 0 and i. This leads to a total of four terms, because the model compound for site i is composed of only two atom groups, i.e., the site itself, and the background atoms that constitute the remainder of the residue. In eqs 6 and 7, the terms $(1/2)E_{ii}(s_i,s_i)$ and $(1/2)E_{ii}^{M_i}(s_i,s_i)$ in the summations are the self-energies (described as Coulomb plus "Born"[12] or Coulomb plus "charge-solvent"[48] energies) of site i in charge state s_i in the protein and in the model compound, respectively.[12] The factor (1/2) and the double notation of the charge state s_i in the self-energy are required for consistency with the definition of the energy of interaction between sites, eq 4.

The right-hand side of eq 5 is the difference between the free energies of protonating site i in the protein (with all other sites neutralized) and in the model compound. If this energy difference is positive, it requires more energy to protonate the site in the protein than in the model compound, and it follows from eq 5 that $\mathrm{p}K_{\mathrm{a},i}^{\mathrm{intr}} < \mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}$.

Solving eq 5 for the $\mathrm{p}K_{\mathrm{a},i}^{\mathrm{intr}}$ and insertion in eq 2 lead to an expression for the pH-dependent energy of the protein in protonation state $\bar{s}$, $\Delta G(\bar{s},\mathrm{pH})$, which only depends on the $\mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}$ of the titrating sites and the electrostatic self-energies and interaction energies $E_{ij}(s_i,s_j)$ in the protein and the model compounds. In Appendix A, it is shown that as a result of the cancellation of identical energy terms from the $\mathrm{p}K_{\mathrm{a},i}^{\mathrm{intr}}$ and from the second sum in eq 2 and further simplification, the energy $\Delta G(\bar{s},\mathrm{pH})$ relative to the unprotonated state can be expressed in the concise form

$$\Delta G(\bar{s},\mathrm{pH}) = (\ln 10)k_{\mathrm{B}}T\sum_{i=1}^{N} s_i(\mathrm{pH} - \mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}) + E_{\mathrm{el}}(\bar{s}) - E_{\mathrm{el}}(\bar{0}) - \sum_{i=1}^{N}(E_{\mathrm{el}}^{M_i}(s_i) - E_{\mathrm{el}}^{M_i}(0)) \quad (8)$$

The energy $E_{\mathrm{el}}(\bar{0})$ or $E_{\mathrm{el}}^{M_i}(0)$ is the electrostatic free energy of the protein or M_i in the unprotonated state. It can be calculated by assigning the partial charges of the unprotonated state to all sites (protein) or site i (model compound), assignment of the fixed partial charges to the background atoms, and subsequent calculation of the Coulomb and electrostatic solvation free energy of the molecule (see section 3.3). In eq 8, the energy of the protonation state $\bar{s}$ is expressed as a simple function of the electrostatic free energy of the protein and the model compounds in the protonation states $\bar{s}$ and s_i, respectively, relative to the energies of their unprotonated states.

By adding the electrostatic free energy of the unprotonated system, $E_{\mathrm{el}}(\bar{0})$, on both sides of eq 8, one obtains an expression for the absolute pH-dependent free energy of the protonation state $\bar{s}$,

$$G(\bar{s},\mathrm{pH}) = (\ln 10)k_{\mathrm{B}}T\sum_{i=1}^{N} s_i(\mathrm{pH} - \mathrm{p}K_{\mathrm{a},i}^{\mathrm{stnd}}) + E_{\mathrm{el}}(\bar{s}) - \sum_{i=1}^{N}(E_{\mathrm{el}}^{M_i}(s_i) - E_{\mathrm{el}}^{M_i}(0)) \quad (9)$$

The introduction of the absolute electrostatic free energy $G(\bar{s},\mathrm{pH})$ has the advantage that it makes it possible to use eq 9 to calculate and compare the free energies of different conformers of a system (see section 2.5).

2.3. Titration: Generating Function Treatment. An exact treatment of the statistical mechanics problem of calculating pH-dependent properties of a given protein conformation, e.g., the average net charge or the probability of finding site i in protonation state s_i, can be based on the generating function,[33−35] a macroscopic analog of the partition function,

$$\Xi(\mathrm{pH}) = \sum_{\bar{s}} \exp(-\beta\, G(\bar{s},\mathrm{pH})); \quad \beta = 1/(k_{\mathrm{B}}T) \quad (10)$$

where the pH-dependent electrostatic free energy of the protonation state $G(\bar{s},\mathrm{pH})$ of the protein is given in eq 9. The sum

in eq 10 is over all protonation states of the system. Furthermore, it is assumed that nonelectrostatic contributions to the free energy of the system are independent of the protonation state (that is, they contribute a constant term) and can thus be omitted in the present analysis.

Given the generating function, the pH-dependent free energy is

$$G(\mathrm{pH}) = -k_B T \ln \Xi(\mathrm{pH}) \tag{11}$$

while the pH-dependent average protonation $\langle n \rangle$ and the probability $\langle s_i \rangle$ of finding site i in the protonated state are given by the ensemble averages[1,12,21]

$$\langle n(\mathrm{pH}) \rangle = \frac{1}{\Xi(\mathrm{pH})} \sum_{\bar{s}} n(\bar{s}) \exp(-\beta G(\bar{s},\mathrm{pH})) = -\frac{1}{\ln 10} \frac{\partial \ln \Xi}{\partial \mathrm{pH}} \tag{12}$$

$$\langle s_i(\mathrm{pH}) \rangle = \frac{1}{\Xi(\mathrm{pH})} \sum_{\bar{s}} s_i \exp(-\beta G(\bar{s},\mathrm{pH})) = \frac{1}{\ln 10} \frac{\partial \ln \Xi}{\partial \mathrm{p}K_{a,i}^{\mathrm{stnd}}} \tag{13}$$

In eq 12, we have written $n(\bar{s}) = \sum_{i=1}^{N} s_i$ for the number of protons bound to the system in state $\bar{s}$.

In eqs 10−13, $G(\bar{s},\mathrm{pH})$ can be replaced by the free energy $\Delta G(\bar{s},\mathrm{pH}) = G(\bar{s},\mathrm{pH}) - E_{el}(\bar{0})$ relative to the unprotonated state, eq 8; this leaves the ensemble averages $\langle n \rangle$ and $\langle s_i \rangle$ unchanged, while the free energy in eq 11 becomes the free energy relative to that of the unprotonated state,

$$\Delta G(\mathrm{pH}) = -k_B T \ln(\sum_{\bar{s}} \exp(-\beta \Delta G(\bar{s},\mathrm{pH}))) = G(\mathrm{pH}) - E_{el}(\bar{0}) \tag{14}$$

In the limit of infinite pH, the only protonation state that is energetically accessible at finite temperature is the unprotonated state (see eqs 8 and 9). It follows that the free energy asymptotically approaches the electrostatic free energy of the unprotonated state, while the average protonation approaches zero; that is,

$$G(\mathrm{pH})|_{\mathrm{pH}\to\infty} = E_{el}(\bar{0}) \tag{15}$$

$$\langle n(\mathrm{pH}) \rangle|_{\mathrm{pH}\to\infty} = n(\bar{0}) = 0 \tag{16}$$

The titration curve $\langle Q(\mathrm{pH}) \rangle$ of the system differs, by definition,[10] only by a constant from the average protonation curve, i.e., the net charge $Q(\bar{0})$ of the unprotonated state, $\langle Q \rangle = \langle n \rangle + Q(\bar{0})$. From eq 16, it follows that the titration curve approaches $Q(\bar{0})$ in the limit pH → ∞. In the following, the term "titration curve" will be used for both the average protonation $\langle n \rangle$ and the average charge $\langle Q \rangle$ as functions of pH; it is implied that the latter is obtained from the former by adding $Q(0)$.

From eqs 14 and 15, it follows that the relative free energy $\Delta G(\mathrm{pH})$ approaches zero in the limit pH → ∞; that is, the infinite pH-limit of the relative free energy is independent of the conformation of the system and the electrostatic free energy of the unprotonated state. In contrast, the absolute free energy depends on the electrostatic free energy (sum of Coulomb and solvation energy) of the unprotonated system.

2.4. Titration Curve Integration. As pointed out above, we assume that the free energy of a given protein conformation depends on pH only through eq 11. The derivative of the free energy with respect to pH is then proportional to the average number of protons bound to the titrating system;[33,35] that is, from eqs 9 to 12, we have

$$\frac{\partial G(\mathrm{pH})}{\partial \mathrm{pH}} = (\ln 10) k_B T \frac{1}{\Xi(\mathrm{pH})} \sum_{\bar{s}} n(\bar{s}) \exp(-\beta G(\bar{s},\mathrm{pH})) = (\ln 10) k_B T \langle n(\mathrm{pH}) \rangle \tag{17}$$

This relation and eq 15 for the infinite pH-limit of the free energy make possible the calculation of the absolute, pH-dependent free energy by integration of the titration curve of the system,

$$G(\mathrm{pH}) = G(\infty) - \int_{\mathrm{pH}}^{\infty} \frac{\partial G(\mathrm{pH}')}{\partial \mathrm{pH}'} \mathrm{d}\mathrm{pH}' = E_{el}(\bar{0}) - (\ln 10) k_B T \int_{\mathrm{pH}}^{\infty} \langle n(\mathrm{pH}') \rangle \mathrm{d}\mathrm{pH}' \tag{18}$$

The pH-dependent electrostatic free energy of the protonation state $\bar{s}$, eq 9, can be written as the sum of a pH-independent contribution $G(\bar{s})$ and a contribution that is proportional to the product of pH and the number of protons bound,

$$G(\bar{s},\mathrm{pH}) = G(\bar{s}) + (\ln 10) k_B T n(\bar{s}) \mathrm{pH} \tag{19}$$

It follows that the energy of all protonation states except the unprotonated state increases linearly with pH and that in the limit pH → ∞ the average protonation of the system as given in eq 12 approaches zero exponentially. Consequently, the integral in eq 18 converges and yields a finite free energy for finite pH; the integral diverges only in the limit pH → −∞.

If the number of titrating sites of a system is sufficiently small (about $N < 30$), it is practical to derive the free energy from the generating function (eq 11). With the use of eq 18 and of an accurate method for calculating titration (average protonation) curves, e.g., the MC titration program, it is possible to calculate the absolute free energy of a system with several hundred titrating sites despite the intractability of the statistical mechanics approach.

2.5. pH-Dependent Stability. Given two conformers A and B of a titrating system (e.g., the bound and unbound states of a protein complex or the native and denatured states of a protein), the pH-dependent free energy difference is

$$\Delta G_{AB}(\mathrm{pH}) = G_A(\mathrm{pH}) - G_B(\mathrm{pH}) \tag{20}$$

To calculate the electrostatic free energy difference, we use eq 18 and obtain

$$\Delta G_{AB}(\mathrm{pH}) = E_{el}^{A}(\bar{0}) - E_{el}^{B}(\bar{0}) - (\ln 10) k_B T \int_{\mathrm{pH}}^{\infty} (\langle n(\mathrm{pH}') \rangle_A - \langle n(\mathrm{pH}') \rangle_B) \mathrm{d}\mathrm{pH}' \tag{21}$$

where $E_{el}^{A}(\bar{0})$ and $E_{el}^{B}(\bar{0})$ are the electrostatic free energies of the unprotonated system in conformation A and B and where $\langle n \rangle_A$ and $\langle n \rangle_B$ are the pH-dependent average protonation values of A and B, respectively. The use of the difference between the titration curves of A and B, $\langle Q \rangle_A - \langle Q \rangle_B$, as the integrand in eq 21 yields the same free energy difference as $\langle n \rangle_A - \langle n \rangle_B$, since A and B have the same net charge in the unprotonated state, $\langle Q \rangle_X = \langle n \rangle_X + Q(\bar{0})$. Equation 21 has been given in differential form by Tanford and Roxby,[11] who stated that the derivative of the folding free energy (expressed as the logarithm of the equilibrium constant K_D for the denaturation) with respect to pH is equal to the difference in proton binding to the native

and unfolded protein, i.e., $\partial \ln K_D/\partial \ln a_{H+} = \Delta\langle n\rangle$ (a_{H+} is the proton activity).

According to eq 21, the absolute free energy difference is given by the difference between the electrostatic energies of the unprotonated states minus $(\ln 10)k_BT$ times the area between the titration curves of the two conformers in the pH-range (pH, ∞). Consequently, a stabilization of conformation A relative to conformation B with respect to the limiting value ($E_{el}^A - E_{el}^B$) results for a pH-range where $\langle n\rangle_A$ is larger than $\langle n\rangle_B$.

If the electrostatic free energy $E_{el}(0)$ of the unprotonated state is not available for A or B or for both A and B (see section 2.7), it is still possible to calculate the *relative* free energy difference $\Delta\Delta G_{AB} = \Delta G_A - \Delta G_B$ between the two states on the basis of their titration curves. Using eq 14 for the relative free energy ΔG of conformer A and B and eqs 20 and 21 for the absolute stability ΔG_{AB}, the relative stability is given by

$$\begin{aligned}\Delta\Delta G_{AB}(\mathrm{pH}) &= \Delta G_A(\mathrm{pH}) - \Delta G_B(\mathrm{pH}) \\ &= \Delta G_{AB} - (E_{el}^A(\bar{0}) - E_{el}^B(\bar{0})) \\ &= -(\ln 10)k_BT\int_{\mathrm{pH}}^{\infty}(\langle n(\mathrm{pH}')\rangle_A - \\ &\qquad \langle n(\mathrm{pH}')\rangle_B)\,\mathrm{dpH}' \quad (22)\end{aligned}$$

Since the relative free energies ΔG_A and ΔG_B of both conformers approach zero in the limit of infinite pH (see eqs 14 and 15), the relative free energy difference $\Delta\Delta G_{AB}$ also approaches zero.

Unlike expressions in the literature,[2,18] eq 22 has a reference pH with a uniquely defined free energy difference, namely, $\Delta\Delta G_{AB}(\infty) = 0$. In previous work, an arbitrary pH and free energy difference, usually pH = 0 and $\Delta\Delta G(0) = 0$, were chosen as the reference for integration of the titration curves of the native and unfolded system. This allows for the calculation of relative changes of the free energy as a function of pH, e.g., the stability of a complex relative to the pH-optimum of the binding energy, but it does not allow for an extension of the theory to yield the absolute free energy difference between two conformers, because the free energy difference at an arbitrary reference pH is undefined. In contrast, the present treatment yields both the relative stability curve of a system, eq 22, and the absolute free energy difference, eq 21, which is obtained from the former by adding the electrostatic free energy difference between the conformers in their unprotonated states.

2.6. Unfolded State. Whereas the X-ray structure is the appropriate model for the native protein in solution, there are insufficient experimental data to determine the structures of unfolded proteins or folding intermediates, in most cases. The available data suggest that the unfolded state of a protein is characterized by an ensemble of rather different conformations.[49−51] Although an ensemble of configurations should be used for the native state, as well as the unfolded state, an average over a sufficient number of conformers is likely to be more important for the latter than for the former.

In the first study of the absolute electrostatic free energy difference between the native and unfolded state, we have simplified the calculation by using an extended, linear structure as a model of the unfolded state. The choice of the linear model corresponds to the limiting case for the unfolded protein. It is thus an important test case for an examination of the interactions between titrating sites in the unfolded state, which are assumed to be zero in the null model (see section 2.7) that has been used previously in protein stability calculations.[2,18] Also, in contrast to random coil structures, which can be generated by various computational techniques for denaturation,[52,53] the extended structure is uniquely defined by a set of ϕ/ψ angles (assuming

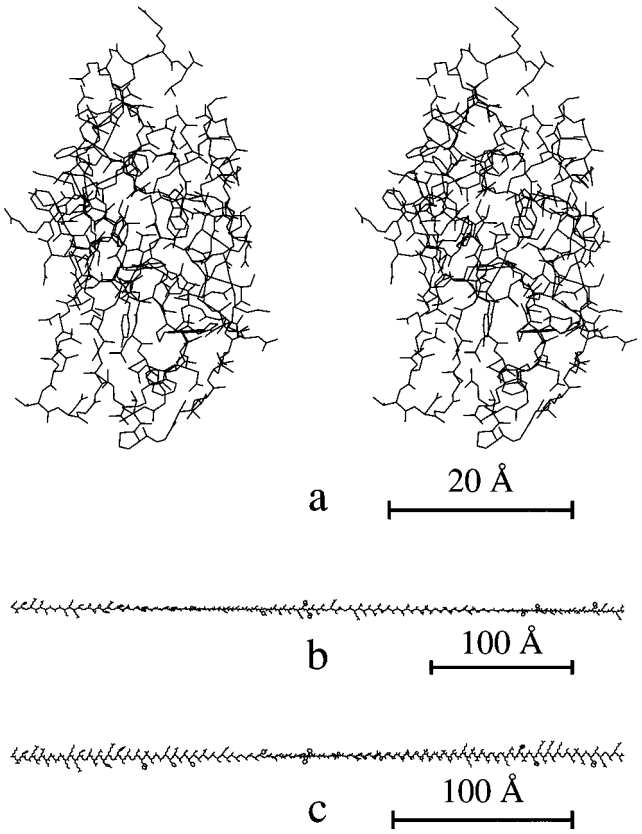

Figure 2. Wire graph of lysozyme (heavy atoms): (a) stereo view of the native state (protein data bank entry 21zt); (b) extended structure "Beta" after 100 steps of steepest descent minimization (all ϕ/ψ angles set to −140°/+135° prior to minimization; N-terminus left); (c) extended structure "Ex72" after 100 steps of steepest descent minimization (all ϕ/ψ angles initially set to −72°/+72°; N-terminus left).

that all side-chain angles are set to 180°) and can easily be generated and reproduced. High accessibility to the solvent is ensured for all side chains so that the maximum effect of denaturation on the stability is likely to be included. Finally, there is a technical reason for the choice of an extended model. In the calculation of electrostatic energies with the FDPB method, a linear structure can be included in an elongated box with a grid constant (resolution) that is comparable to that used for native protein structures. Despite the length of a linear chain with several hundred residues, this is achieved by orienting the chain along the *z*-axis of Cartesian space and using a grid with different dimensions N_x, N_y, N_z along the Cartesian axes, where $N_z \gg N_x, N_y$. This simplifies the calculation of the electrostatic energy of the unfolded state, relative to more random but still highly extended protein conformations.

As a simple test of the dependence of the results on the choice of ϕ/ψ values for an extended structure, we use two different models of the unfolded state (see Figure 2): first, the extended ideal β-structure "Beta" with (ϕ, ψ) = (−140°, +135°) and, second, the "Ex72" model with (ϕ, ψ) = (−72°, +72°). For both models, the choice of the (ϕ, ψ) angles leads to a backbone conformation that is close to a local minimum of the vacuum energy;[54,55] while the total energy of the Ex72 model is higher than that of the ideal Beta model, the Coulomb energy (dielectric constant of 1, no cutoff) of Ex72 is lower than that of Beta (see section 3.1). In the Ex72 model, the choice of ϕ/ψ values leads to a chain of open, seven-membered rings along the backbone, each consisting of the atom sequence O_{i-1}, C_{i-1}, N_i, $C_{\alpha i}$, C_i, N_{i+1}, H_{i+1}, where i is a residue number. For lysozyme, the example used here, Figure 2 illustrates that the longest dimen-

sion of the Ex72 model, 343 Å, is smaller than that of the Beta model, 437 Å.

2.7. Null Model. In the literature, a "zero interaction"[18] or "null"[2] model for the unfolded state of a protein has been proposed where all sites are assumed to have their standard $pK_{a,i}^{stnd}$ and to titrate independently. The null model has also been used as a test of pK_a calculational methods in the folded state;[2] that is, the results of a given calculational methods are compared with what would have been obtained with the null model. The null model assumes that the self-energy difference between the protonated state of site *i* and its unprotonated state is the same for the site in the protein as in the model compound. Combined with the assumption of zero interaction between sites, including no interaction with the background atoms, the relative free energy of protonation states for the unfolded protein $\Delta G_{null}(\bar{s},pH)$ is

$$\Delta G_{null}(\bar{s},pH) = (\ln 10)k_B T\sum_{i=1}^{N} s_i(pH - pK_{a,i}^{stnd}) \quad (23)$$

To derive formulas for the relative folding free energy and protonation curve from the null model, it is useful to introduce the pH-dependent "site energy" $\Delta G_i(s_i,pH)$,

$$\Delta G_i(s_i,pH) = (\ln 10)k_B T s_i\,(pH - pK_{a,i}^{stnd}) \quad (24)$$

such that for the null model the relative energy of the protonation state $\bar{s}$ is the sum of the site energies,

$$\Delta G_{null}(\bar{s},pH) = \sum_{i=1}^{N}\Delta G_i(s_i,pH) \quad (25)$$

For the null model it is possible to calculate the generating function $\Xi_{null}(pH)$ and the relative free energy $\Delta G_{null}(pH)$ for systems with an arbitrary number of titrating sites;[2,18] i.e.,

$$\Xi_{null}(pH) = \sum_{\bar{s}}\{\prod_{i=1}^{N}\exp(-\beta\Delta G_i(s_i,pH))\} = \prod_{i=1}^{N}\{1 + \exp(-\beta\Delta G_i(1,pH))\} \quad (26)$$

$$\Delta G_{null}(pH) = -k_B T\sum_{i=1}^{N}\ln\{1 + \exp(-\beta\Delta G_i(1,pH))\} \quad (27)$$

where we used the definition of the charge states in the two-state model, eq 1, and the fact that the site energy is equal to zero in the unprotonated state, $\Delta G_i(0,pH) = 0$. It is evident that a corresponding factorization of the generating function cannot be done for the general case of interacting sites because the free energy of a protonation state, eq 9, is not a sum of *N* independent terms.

In the null model, the sites titrate independently with their standard pK_a's, such that the pH-dependent average protonation of the system becomes

$$\langle n(pH)\rangle_{null} = \sum_{i=1}^{N}\langle s_i(pH)\rangle_{null} = \sum_{i=1}^{N}\left\{\frac{s_i}{\Xi_i(pH)}\exp(-\beta\Delta G_i(1,pH))\right\} \quad (28)$$

$$\Xi_i(pH) = 1 + \exp(-\beta\Delta G_i(1,pH))$$

In eq 28, we have written $\langle s_i\rangle_{null}$ for the probability that site *i* is protonated.

As pointed out in section 2.3, the relative energy of the protonation state, $\Delta G(\bar{s},pH)$, can be used to calculate the titration curve $\langle n\rangle$ and the relative free energy $\Delta G(pH)$ of a system. It is thus possible to calculate the relative pH-dependent stability of proteins using the null model for the unfolded state. However, the null model provides no information on the electrostatic free energy $E_{el}(\bar{s})$ of the system in protonation state $\bar{s}$, in particular the free energy $E_{el}(\bar{0})$ of the unprotonated state. It is, therefore, not possible to evaluate the absolute free energy of protein folding $\Delta G(pH)$ on the basis of the null model (see eq 21).

A comparison between the titration curves calculated for extended model structures of proteins according to eq 12 and the titration curve according to the null model, eq 28, shows that the assumption of no interactions between titrating sites in the null model is contradicted by the Beta or Ex72 models (see section 4.2). Furthermore, by calculating the pH-dependence of the stability of the extended protein models relative to the null model, it is possible to determine what is essentially a lower bound for the energetic contribution of site−site interactions to the stability of the unfolded state of proteins.

3. Methods

In this section, we describe the methods used in setting up the system and doing the calculations. The description focuses on lysozyme, which is the protein that is studied. However, most of the methodology is valid for other proteins.

3.1. Preparation of Structures. The structures of hen egg-white lysozyme were taken from the protein data bank.[56] Both the triclinic crystal structure (entry 21zt[57]) and the tetragonal crystal structure (entry 1hel[58]) were used for the native state to obtain a measure of the effect of different configurations in the native state; it has been shown[12] that the two structures give significantly different pK_a values for some sites. Hydrogen positions were calculated using the HBUILD command[59] within CHARMM,[60] and van der Waals radii ($R_{min}/2$ of the Lennard-Jones potential) were taken from the all-hydrogen parameter set param22 of CHARMM.[61,62]

To further analyze the effect of local structural changes on the pK_a's, titration curve, and stability curve, we generated minimized structures by 100 steps of steepest descent (sd) minimization using CHARMM. Further structures were generated by 100 and 200 additional minimization steps with the conjugate gradient (cg) algorithm. We thus generated the sequence of structures 21zt (unminimized), 21zt-m1 (100 steps sd), 21zt-m2 (100 steps sd followed by 100 steps cg), and 21zt-m3 (100 steps sd followed by 200 steps cg) for triclinic lysozyme and a corresponding series for tetragonal lysozyme (analogous nomenclature). None of these structures are at a local minimum of the vacuum energy, even though the total energy of the system varies slowly toward the end of the minimization, as compared to the initial 100 minimization steps (see Table 1); further minimization using the conjugate gradient method would lead to a continued change in the structure and the energy (data not shown). The objectives of the minimization are to remove any bad contacts in the crystal structures and examine the effects of small structural changes on the electrostatic properties of the protein. Table 1 shows that the two crystal structures (unminimized and minimized) have very similar values of the total and Coulomb energies. The rms differences between the 21zt/1hel and the minimized 21zt/1hel structures, respectively, are given in Table 2. The rms structural difference between the two crystal structures is 1.4 Å. This difference is almost unaffected by the minimization; that is, the two crystal structures converge toward different minima upon *in vacuo* minimization (see Table 2).

TABLE 1: Energy upon Minimization[a]

name[b]	step[c]	E_{tot}	E_{Coul}	E_{vdW}
21zt	0	−2263	−3224	−388
	100	−3154	−3507	−615
	200	−3543	−3998	−562
	300	−3611	−4057	−573
1hel	0	−2091	−3263	−98
	100	−3105	−3531	−570
	200	−3478	−3959	−553
	300	−3613	−4051	−565
Beta	0	−342	−2263	624
	100	−1687	−2463	−79
	200	−2079	−3051	−31
	300	−2241	−3131	−47
Ex72	0	1×10^6	−2962	1×10^6
	100	−1360	−2864	461
	200	−1803	−3261	284
	300	−2116	−3446	138

[a] All energies in kcal/mol. [b] For the meaning of the name, see text. [c] Minimization step (see text).

TABLE 2: Structural Change upon Minimization

structures that are compared[a]		rms deviation (Å) all atoms	rms deviation (Å) backbone
21zt	21zt-m1	0.11	0.08
21zt	21zt-m2	0.61	0.38
21zt	21zt-m3	0.70	0.44
21zt-m1	21zt-m2	0.56	0.35
21zt-m1	21zt-m3	0.66	0.41
21zt-m2	21zt-m3	0.21	0.13
1hel	1hel-m1	0.17	0.10
1hel	1hel-m2	0.73	0.44
1hel	1hel-m3	0.92	0.56
1hel-m1	1hel-m2	0.68	0.41
1hel-m1	1hel-m3	0.89	0.53
1hel-m2	1hel-m3	0.46	0.30
21zt	1hel	1.36	0.62
21zt-m3	1hel-m3	1.44	0.64
Beta	Beta-m1	0.11	0.07
Beta	Beta-m2	0.84	0.51
Beta	Beta-m3	0.89	0.50
Beta-m1	Beta-m2	0.80	0.50
Beta-m1	Beta-m3	0.85	0.49
Beta-m2	Beta-m3	0.28	0.16
Ex72	Ex72-m1	0.19	0.14
Ex72	Ex72-m2	0.92	0.59
Ex72	Ex72-m3	1.54	1.20
Ex72-m1	Ex72-m2	0.84	0.55
Ex72-m1	Ex72-m3	1.50	1.18
Ex72-m2	Ex72-m3	1.03	0.80

[a] For the meaning of the name, see text.

The extended structures for the unfolded state of lysozyme were generated by assigning the Beta and Ex72 values to the ϕ/ψ dihedral angles (see section 2.6), 180° to all other dihedral angles, and ideal bond length and angles to the atoms of the backbone, and subsequent use of the internal coordinate IC BUILD command within CHARMM. Since the generated structures can have strain from bonded energy terms, especially within prolines, we applied the same minimization scheme to Beta and Ex72 as to the crystal structures (see above), generating the structures Beta-m1, ..., Beta-m3 and Ex72-m1, ..., Ex72-m3, respectively. The rms structural change upon minimization of the extended protein models Beta and Ex72 are also given in Table 2. The data show that Ex72 undergoes a larger structural change upon minimization than Beta; this is explained by the high strain in the unminimized Ex72 structure due to bad contacts (van der Waals repulsion). These bad contacts are removed within the first 100 steps of minimization (see Table 1). Whereas the total energy of Beta is lower than that of Ex72 model at all minimization steps investigated in this study, the Ex72 model has a significantly lower Coulomb energy than the extended Beta structure.

3.2. Titrating Sites. The set of sites that is used in this study, together with the atoms that belong to each site and their partial charges as a function of the protonation state, are given in Table 3. The data are compatible with the all-hydrogen parameter set param22 of CHARMM.[61,62] For simplicity, only two protonation states are considered for each site. This simplification requires the introduction of average charge states for the atoms in cases where there are several states compatible with a given number of protons bound at the site; for example, even though there are two forms of the Asp carboxyl group with one proton bound, we consider a single protonated form without explicit representation of the hydrogen, with a net charge of zero for the site, and with symmetric assignment of the partial charges to the two oxygens.

Serine, threonine, and cysteine are also listed in Table 3 even though they are excluded in the applications to the protein hen egg-white lysozyme that are reported in section 4. Serine and threonine are excluded because their standard $pK_a = 13.6$ is far from pH ≈ 7, which corresponds to the maximum stability of lysozyme.[64] The extension of the theory to include metal ions and cysteines that take part in disulfide bridges is possible and requires the consideration of redox potentials in analogy to the present treatment of the proton chemical potential. To avoid the complication of the theory that would be necessary in applications to hen egg-white lysozyme (all cysteines in the native enzyme are involved in disulfide bridges), we exclude cysteines from the list of sites in section 4.

3.3. Electrostatic Energy. For the calculation of the electrostatic interaction energies between sites, eq 4, we employ the program UHBD,[41] which solves the Poisson−Boltzmann equation numerically on a three-dimensional grid. The finite-difference Poisson−Boltzmann (FDPB) method is based on the continuum model where the protein interior and the solvent are described as polarizable continua with dielectric constants ϵ_i and ϵ_s, respectively.[24,65] In principle, more than two different dielectric constants could be introduced to account, for example, for differences in the mobility of side chains between the protein interior ("protein core") and the surface[66] or for differences in the polarizability between lipid head groups and tails in calculations that involve lipid membranes.[67]

To determine the interaction energies $E_{ij}(s_i,s_j)$ between site i in charge state s_i and all other sites $j \neq i$ (in all possible charge states s_j), it is sufficient to calculate the potential $\phi_{(i,s_i)}(\vec{x})$ of the charges of site i in charge state s_i and then use eq 4 for all (j, s_j); this includes the interaction between site i and the background atoms in the protein, $E_{i0}(s_i,0)$. To calculate the interaction energy with the background atoms in the model compound, $E_{i0}^{M_i}$, the potetnial of site i in charge state s_i must be determined for the site as part of the model compound M_i. The self-energies of the sites in the protein and model compounds, $(1/2)E_{ii}$ and $(1/2)E_{ii}^{M_i}$, can be derived from the same potentials using the analytical scheme in UHBD for subtracting the grid-Coulomb and grid-self contributions to the electrostatic energy.[48] This scheme yields the solvation energy of site i, $\Delta E_{ii}^{solv} = E_{ii} - E_{ii}^{Coul}$, where E_{ii}^{Coul} is the sum of all (bonded and nonbonded) Coulomb interactions between the atoms of the site in a homogeneous dielectric ϵ_i. However, in the present methodology it is not necessary to include the nonbonded Coulomb energy contribution to the self-energy of site i in the protein and in the model compound, since these contributions are identical and thus cancel each other in eq 9 for the energy of a protonation sate, $G(\bar{s},pH)$. For the same reason, it is not necessary to exclude the bonded Coulomb interaction between

TABLE 3: Titrating Sites in the Two-State Model[a]

site	pK_a^{stnd}	atom	charge[b] $s=0$	charge[b] $s=1$
Arg	12.48	CD	0.20	0.20
		HD1	0.09	0.09
		HD2	0.09	0.09
		NE	−0.70	−0.70
		HE	0.44	0.44
		CZ	0.44	0.64
		NH1	−0.80	−0.80
		HH11	0.26	0.46
		HH12	0.26	0.46
		NH2	−0.80	−0.80
		HH21	0.26	0.46
		HH22	0.26	0.46
Asp	4.00	CB	−0.28	−0.21
		HB1	0.09	0.09
		HB2	0.09	0.09
		CG	0.62	0.75
		OD1	−0.76	−0.36
		OD2	−0.76	−0.36
C-Ter	3.80	C	0.34	0.76
		OT1	−0.67	−0.38
		OT2	−0.67	−0.38
Cys[d]	10.46	CB	−0.25	−0.11
		HB1	0.05	0.09
		HB2	0.05	0.09
		SG	−0.85	−0.23
		HG1	0.00	0.16
Glu	4.40	CG	−0.28	−0.21
		HG1	0.09	0.09
		HG2	0.09	0.09
		CD	0.62	0.75
		OE1	−0.76	−0.36
		OE2	−0.76	−0.36
His[c]	6.42	CB	−0.08	−0.05
		HB1	0.09	0.09
		HB2	0.09	0.09
		CG	0.08	0.19
		ND1	−0.53	−0.51
(His continued)		HD1	0.16	0.44
		CD2	0.09	0.19
		HD2	0.09	0.13
		CE1	0.25	0.32
		HE1	0.13	0.18
		NE2	−0.53	−0.51
		HE2	0.16	0.44
Lys	10.79	CE	0.18	0.21
		HE1	0.05	0.05
		HE2	0.05	0.05
		NZ	−0.96	−0.30
		HZ1	0.34	0.33
		HZ2	0.34	0.33
		HZ3	0.00	0.33
N-Ter	7.50	N	−0.96	−0.30
		HT1	0.34	0.33
		HT2	0.34	0.33
		HT3	0.00	0.33
		CA	0.19	0.21
		HA	0.09	0.10
Ser[d]	13.60	CB	−0.14	0.05
		HB1	0.05	0.09
		HB2	0.05	0.09
		OG	−0.96	−0.66
		HG1	0.00	0.43
Thr[d]	13.60	CB	−0.05	0.14
		HB	0.09	0.09
		OG1	−0.96	−0.66
		HG1	0.00	0.43
		CG2	−0.35	−0.27
		HG21	0.09	0.09
		HG22	0.09	0.09
		HG23	0.09	0.09
Tyr	10.13	CZ	−0.04	0.11
		OH	−0.96	−0.54
		HH	0.00	0.43

[a] Atoms and charges compatible with the all-hydrogen parameter set of CHARMM; standard pK_a's taken from refs 4, 42, 43, and 63. [b] Partial charges in the unprotonated ($s = 0$) and protonated state ($s = 1$). [c] Protonated His: average of Hsd and Hse in the CHARMM parameter set; "macroscopic" standard pK_a of His as defined in ref 43. [d] Site excluded from calculations reported in section 4.

the atoms of site i and the background atoms when calculating the energies E_{i0} and $E_{i0}^{M_i}$ according to eq 4. This is important because the calculation of the electrostatic potential of site i with the FDPB method does not allow for the exclusion of specific charge-pair interactions.

Finally, for the calculation of the absolute energy as a function of a protonation state, eq 9, it is necessary to calculate the total electrostatic free energy of the unprotonated state, $E_{el}(\bar{0})$, which is the sum of the Coulomb energy and the solvation energy. We use the CHARMM program[60] without cutoff and a dielectric constant of ϵ_i (dielectric constant of the protein interior) for calculating the Coulomb energy of the unprotonated system, and UHBD for the solvation energy, i.e., the energy of transfer from the dielectric ϵ_i to the solvent. The reason for using CHARMM to calculate the Coulomb energy and not UHBD is the fact that in the latter the interactions between bonded atom pairs cannot be excluded, even though they are not contributing to the conformational energy of the system.

For a protein with N titrating sites, where N_i is the number of states of site i, it follows that the calculation of the electrostatic energies E_{ij} requires $\sum_i N_i$ calculations of the electrostatic potential for sites in the protein, the same number of calculations for the sites in the model compounds, and one calculation for the solvation energy of the unprotonated state, i.e., a total of $1 + 2\sum_i N_i$ finite-difference calculations. In the case where all sites are assigned only two states ($N_i = 2$ for all i), the number of FDPB calculations is $1 + 4N$. If the protonation of a site is approximated by the +1 change of the partial charge of a single atom, the number of necessary FDPB calculations is reduced to $1 + 2N$ (see ref 2).

In the present approach for calculating the electrostatic free energy $G(\bar{s},\mathrm{pH})$ of the system, there exists an inconsistency that results from the fact that for each site an individual model compound is used with its associated electrostatic energy $E_{el}^{M_i}(s_i)$, such that intraresidue electrostatic energy contributions are canceled in eq 9 even though they are part of the conformational energy of the system. This problem is addressed in Appendix C; its effect on the energetics of a system with only 20−30 sites is expected to be small.

For all calculations of the electrostatic potential reported in this study, we make use of the focusing technique[68] with a protocol that is similar to focusing schemes used previously.[2,21] This allows a considerable reduction of the computing time relative to the case where a single, large grid with a small grid

constant is used to determine the potential. Four grids (initial grid and three focusing grids) are used for the sites in the protein, and three grids for the sites in the model compounds. In both cases, the boundary values of the initial grid potential are calculated using the Debye−Hückel equation with the dielectric constant and ionic strength of the solution, and the grid constant (distance between grid points) and border space (minimum distance between any atom in the protein/model compound and the surface of the grid) are set to 2.5 and 20.0 Å, respectively. In the above considerations on the number of FDPB calculations that are required to determine the energies E_{ij} in the protein and the model compounds, each sequence of four (protein) or three (model compound) focusing potential calculations is counted as one FDPB calculation, even though it involves the repeated application of the FDPB algorithm. This is justified by the fact that, in general, the final potential calculation of a focusing sequence accounts for 80−90% of the CPU time that is required by the entire sequence.

For the focusing grids of site *i* in the protein, we use the sequence (i) grid constant $a = 1.25$ Å and border space $b = 5.0$ Å around the entire protein, (ii) $a = 0.6$ Å and border space $b = 15.0$ Å around the residue of site *i*, and (iii) $a = 0.3$ Å and border space $b = 4.0$ Å around the residue of site *i*. The two focusing grids of site *i* in the model compound (isolated residue of the site) are identical with the last two focusing grids of the site in the protein (ii and iii, above; same grid origin, orientation, grid constant, and dimension). This improves the accuracy because of a cancellation of errors between the calculated potentials in the protein and in the model compound (see eq 5).

After each focusing calculation, the potential at the atom positions that are included in the focusing grid is updated. The potential of site *i* is thus determined at a resolution of 0.3 Å within a distance of 4.0 Å from the site, at a resolution of 0.6 Å in the distance range 4.0−15.0 Å and at a resolution of 1.25 Å at larger distances. The described focusing scheme ensures that the calculated self-energies and interaction energies, and consequently the calculated pK_a's and titration curves, are nearly independent of the grid setup in the FDPB method. This can be verified by variation of the grid constants used in the calculations.[2]

For the calculation of the solvation energy of the unprotonated system, a separate FDPB calculation is performed with an initial potential grid and one focusing grid. As before, the initial grid is defined by a grid constant of 2.5 Å and a border space of 20.0 Å around the protein. In contrast to that for the titrating sites, however, the final (focusing) grid used for the solvation energy calculation of the unprotonated state must include the entire protein. Because of limitations in the available computer memory, we were restricted to FDPB grids with a maximum number of about 2×10^6 grid points. For consistency between different configurations of the same protein, we use the same final grid constant of 0.7 Å in all calculations of the solvation energy of the unprotonated state.

We set the ionic strength of the aqueous solution in all FDPB calculations to 0.145 M, with the ion-exclusion (Stern) layer equal to 2.0 Å. The dielectric constant of water is set to $\epsilon_s = 80$, whereas the dielectric constant that is assigned to the protein (model compound) interior is varied in the range $\epsilon_i = 1$ to $\epsilon_i = 30$ in the test calculations of the lysozyme pK_a's. In the calculations of the relative and absolute free energy curves, the dielectric constant of the protein is set to the value that leads to the best agreement between calculated and experimental pK_a's in lysozyme; this value is $\epsilon_i = 20$ (see section 4.1).

In all FDPB calculations, the solute interior is determined using the solvent accessibility method,[68] with a probe sphere radius of 1.4 Å and 800 dots per atom for the generation of the Sharke and Rupley[69] dot surface. For the assignment of the dielectric constant to grid links that are passing through the solute surface, the boundary smoothing option (interpolation between protein and solvent dielectric constants) was employed;[70] this improves the grid independence of the calculated energies.

For an average-sized protein (e.g., lysozyme with 1960 atoms in the all-hydrogen representation), the CPU time that is required for the FDPB calculations is on average 90 CPU-seconds per site and charge state of the protein, and 20 CPU-seconds per site and charge state of a model compound, using an HP-735/9000 for the calculations. The calculation time for the solvation energy of the unprotonated state of lysozyme is 110 CPU-seconds. The total time that is required for the calculation of the electrostatic energies for lysozyme with 32 sites (two-state model) is 2 CPU-hours. Thus, the use of the FDPB method in the framework of pK_a and stability calculations is not very computationally intensive.

3.4. Titration: Monte Carlo Method. Because of the exponential dependence of the number of protonation states on the number of titrating sites, the calculation of the generating function, of the free energy, and of the thermodynamic averages according to eqs 10−13 is impractical for systems with more than 25−30 sites, given the speed of current computers. In this study, we use a Monte Carlo (MC) titration program that was kindly provided to us by Beroza *et al*.[1] The Metropolis *et al*.[71] "importance sampling" method is well suited for evaluating statistical ensemble averages of a titrating system, e.g., the average protonation or titration curve.[72] A method for calculating the free energy from the titration curve is given in section 2.4.

The MC titration program of Beroza *et al*. is restricted to two protonation states per site as defined in eq 1, and the calculations that are reported in section 4 are limited to the two-state model. A modified version of the MC titration program that is applicable to the general case of multiple protonation states is being developed and will be described elsewhere.[40] The program of Beroza *et al*. requires the interaction energy matrix W_{ij} and the $pK_{a,i}^{intr}$ of all sites as input; they can be calculated according to eqs 3 and 5. The principal output of the program is the fractional degree of protonation of the sites in the pH-range of interest, which is also specified in the program input. The problem of strongly coupled, neighboring sites is solved by the introduction of "pair steps", where two sites exchange a proton and thus change their protonation states simultaneously. Each pair of sites for which the interaction energy W_{ij}, eq 3, is larger than a lower bound given in the program input is included in the list of coupled sites.

Since the details of the MC program have not been published, we present an outline here that complements the original paper of Beroza *et al*.[1] In the calculation, a "MC step" is defined as a sequence of N_{dof} "Metropolis steps", where N_{dof} is the number of degrees of freedom (single and coupled site transitions) that are selected for the sampling. A Metropolis step consists of the random selection of one degree of freedom, calculation of the change in the free energy of the system upon change of the degree of freedom, and application of the Metropolis criterion[71] to accept or reject the new charge state of the system in which the selected degree of freedom is altered. The change in the free energy upon change in the charge state is calculated using eq 2 for the energy of a protonation state.

At a given pH, the MC program generates an ensemble E of charge states as follows: (i) All titrating sites are assigned a charge state at random. (ii) To equilibrate the system before data collection, N_{eq} MC steps are performed, where N_{eq} is set equal to the number of "full MC steps" that are used for data collection (see below). For the equilibration, only single site transitions are used as the degrees of freedom, $N_{dof} = N_{site}$. (iii) The final charge state of the equilibration calculation is used as the starting point for N_{full} "full" MC steps for data collection. After each MC step, the current charge state of the system is added to the ensemble E. The term "full" MC step is used to make clear that all titrating sites are included. Both single site and pair transitions of the charge state are used as the degrees of freedom such that $N_{dof} = N_{site} + N_{pair}$, where N_{pair} is the number of coupled sites.

Optionally, a second ensemble E_{red} of N_{red} charge states is calculated by selecting a reduced set of titrating sites and allowing only for single site and pair steps in the reduced set. The last charge state in the full MC ensemble E is used as the starting point for the reduced MC calculation. For the selection of the reduced set of sites, the full MC ensemble E is analyzed and the sites are selected for which the average degree of protonation see eq 29) differs by more than a small constant δ from the unprotonated and protonated states; that is, site i is selected if $\delta < p_i < (1 - \delta)$, where we have written p_i for the average protonation of site i.

Given the MC ensemble E, the probability of having site i protonated is

$$\langle s_i(\mathrm{pH})\rangle = \frac{1}{N(E)} \sum_{\bar{s}\in E} s_i \tag{29}$$

where $N(E)$ is the number of states in the ensemble E and where the sum is over all charge states in the ensemble. The average in eq 29 is pH-dependent because the free energy and, therefore, the ensemble E changes with pH. Correspondingly, the average pH-dependent protonation of the system is given by

$$\langle n(\mathrm{pH})\rangle = \frac{1}{N(E)} \sum_{\bar{s}\in E} n(\bar{s}) \tag{30}$$

where $n(\bar{s})$ is the number of protons bound to the system in charge state $\bar{s}$. The last two equations for the MC ensemble are equivalent to eqs 12 and 13 for the exact treatment of the generating function.

To estimate the error for the calculated average protonation of site i or the entire system, the autocorrelation function is derived from the sequence of charge states in the MC ensemble; for example, for site i, the autocorrelation function is[1]

$$C_i(\tau) = \frac{1}{N(E) - \tau} \sum_{t=0}^{N(E)-\tau-1} s_i(t+\tau)\, s_i(t) - \langle s_i\rangle^2 \tag{31}$$

where $s_i(t)$ is the charge state of site i at MC step t and $\langle s_i\rangle$ is the average value of s_i. The pH-dependence of the variables has been omitted for brevity. In the MC program of Beroza *et al*., the step number τ where the autocorrelation function has decreased to 1/10 of its initial value is arbitrarily defined as the autocorrelation time; i.e., $\tau_i = \min\{\tau | C_i(\tau) < C_i(0)/10\}$. Using the variance, $C_i(0)$, of the calculated average protonation state and the autocorrelation time τ_i, the standard deviation σ_i of $\langle s_i\rangle$ is given by

$$\sigma_i^2 = \frac{C_i(0)}{N(E)/\tau_i} \tag{32}$$

where $N(E)/\tau_i$ is an estimate of the number of independent measurements of s_i in the MC ensemble.

If both a full MC ensemble using all sites and an ensemble for a reduced set of sites are calculated, the averages of the two ensembles are combined according to their statistical weight. If the two averages for site i are $\langle s_i\rangle_1$ and $\langle s_i\rangle_2$ and the standard deviations are σ_{i1} and σ_{i2}, respectively, the combined average is

$$\langle s_i\rangle_{1\&2} = \sigma_i^2 \left(\frac{1}{\sigma_{i1}^2}\langle s_i\rangle_1 + \frac{1}{\sigma_{i2}^2}\langle s_i\rangle_2\right);\quad \frac{1}{\sigma_i^2} = \frac{1}{\sigma_{i1}^2} + \frac{1}{\sigma_{i2}^2} \tag{33}$$

where σ_i is the standard deviation of the combined result.

In the MC titration calculations reported in this work, we used $N_{full} = 2000$ full MC steps followed by $N_{red} = 4000$ reduced MC steps. This implies that $N_{eq} = 2000$ MC steps were performed to equilibrate the system before calculating the full MC ensemble. The temperature was set to $T = 300$ K, and the constant δ for the selection of the reduced set of titrating sites was set to $\delta = 10^{-4}$. To determine the set of coupled pairs of sites, the equivalent of 0.5 pK-units was used as the lower bound to the interaction energy, $W_{ij} > 0.5(\ln 10)k_BT$.

For a given number of MC steps, the calculation time that is required by the titration program depends quadratically on the number of sites. To perform 2000 full MC steps and 4000 reduced MC steps at 150 different pH values (i.e., 150 × [2000 + 4000] MC steps) for lysozyme with 32 sites, the titration program takes 13 CPU-minutes on an HP-735/9000 workstation. Thus, the CPU time required by the MC titration program is only a fraction of the time required by the electrostatic energy calculations with the FDPB method (see end of section 3.3).

4. Results

There are many experimental studies on the pK_a's,[44,45,73] titration curve,[11,17] and pH-dependent stability of lysozyme,[64] which makes this protein an excellent test case. In fact, lysozyme was one of the first proteins for which a complete titration curve was measured.[17] Hen egg-white lysozyme has 11 arginines, 7 aspartic acids, 2 glutamic acids, 1 histidine, 6 lysines, 3 tyrosines, and an N- and C-terminus, i.e., a total of 32 titrating sites, with the list of sites given in Table 3 (see section 2.1 for the reasons to exclude serine, threonine, and cysteine).

4.1. Protein Dielectric Constant. To verify which internal dielectric constant ϵ_1 is to be assigned to the protein interior in the FDPB calculations (see section 3.3), we performed a series of pK_a-calculations with ϵ_i in the range 1−30. For each value of the dielectric constant ϵ_i, we calculated the electrostatic free energies (self-energies and interaction energies of all sites and the background charges, protein and model compounds) and then used the MC titration program to determine the pH-dependent protonation of all sites. Figure 3 shows an example of the calculated titration curves for each of the sites in lysozyme with $\epsilon_i = 20$.

By definition, the effective pK_a of site i in the protein is the point at which the average protonation is equal to 1/2. We derived the pK_a's of all sites in lysozyme from the output of the MC titration program and compared them with experimental data.[44,45,74] In Figure 4, the average error between the calculated and experimental pK_a's in lysozyme is shown as a function of the internal dielectric constant used in the continuum electrostatic calculations; curves for the crystal structures 21zt and 1hel as well as for the minimized structures (see section 3.1) are shown.

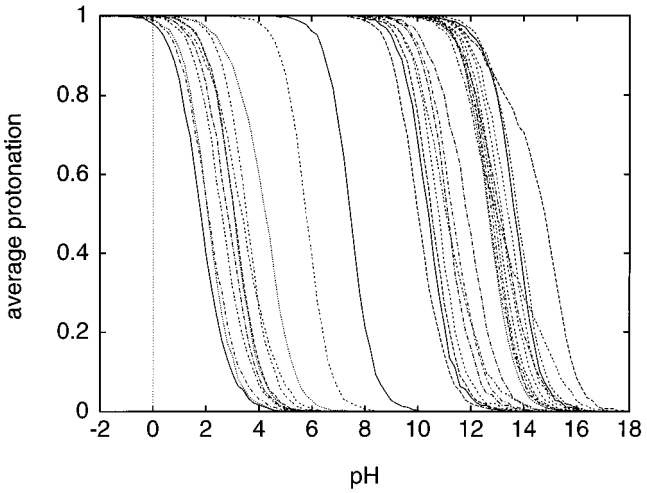

Figure 3. pH-dependent protonation of the 32 titratable sites in lysozyme (21yz) as calculated with the MC titration program. Protein dielectric constant $\epsilon_i = 20$. The two protonation curves that depart markedly from curves of independently titrating sites (pH-range 12−16) are for Tyr 53 and Arg 68 with effective pK_a's of 13.0 and 14.8, respectively.

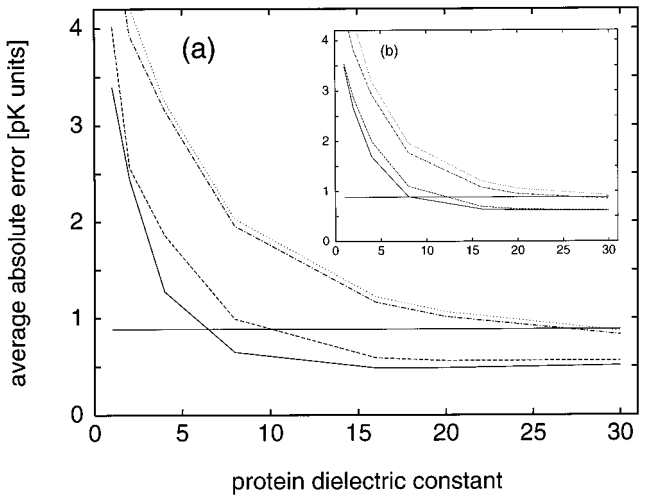

Figure 4. (a) Average error of calculated pK_a's relative to experimental data for triclinic hen egg-white lysozyme as a function of the protein dielectric constant used in the FDPB calculations. From bottom to top: crystal structure 21zt and minimized structures 21zt-m1, 21zt-m2, and 21zt-m3. (b) Same as part a for tetragonal hen egg-white lysozyme. From bottom to top: crystal structure 1hel and minimized structures 1hel-m1, 1hel-m2, and 1hel-m3.

For all structures, the average error decreases continuously when the dielectric constant of the protein interior is increased from 1 to 20. An increase of ϵ_i beyond 20 does not lead to a significant change of the error. For both 21zt and 1hel, and independent of the choice of the internal dielectric constant, the calculated pK_a's of the crystal structure are in better agreement with experiment than the pK_a's of the minimized structures. The smallest average error for the 21zt and 1hel crystal structures are 0.49 pK_a-units at $\epsilon_i = 20$ (rms deviation 0.67) and 0.61 pK_a-units at $\epsilon_i = 30$ (rms deviation 0.83), respectively. This is significantly less than the average error of 0.89 pK_a-units (rms deviation 1.23) between the standard pK_a's of the titrating sites and the experimentally observed pK_a's in the protein, i.e., the error of the null model (see section 2.7).

Our results confirm and extend the findings of Antosiewicz *et al.*,[2] who explored the agreement between calculated and experimental pK_a's of three proteins for values of the protein dielectric constant of $\epsilon_i = 2$, 4, and 20. The best result was obtained by them at the maximal value. In the following, $\epsilon_i = 20$ for the protein dielectric constant is used for the electrostatic free energy calculations. Furthermore, the stability calculations are based on the crystal structure as the model of the native protein (see also Figure 7 and related text).

4.2. Titration Curve and pK_a's. Because of the large number (32) of titrating groups, the free energy and pH-stability of lysozyme was calculated using the titration curve integration method, eq 21. To this end, we determined the titration curves for the crystal structures 21zt and 1hel and for the unfolded protein models null, Beta, and Ex72. Figures 5 and 6 show, respectively, calculated and experimental titration curves for the folded and unfolded protein. From Figure 5a, there are only minor differences between the titration curves of the two crystal structures; the maximum difference is $|\Delta\langle Q\rangle|_{max} = 0.72$ at pH = 2.2. The maximum differences between the titration curves of the unfolded lysozyme models are also small, 0.32 at pH = 3.8 between null and Beta-m1 and 0.80 at pH = 4.0 between null and Ex72-m1 (see Figure 5b). This is to be compared with the maximum difference between the calculated titration curves of the crystal structure 21zt and the null model; the maximum is 5.02 at pH = 3.4 (see Figure 5a). The maximum difference between the experimental titration curves of the native and unfolded lysozyme is 3.45 at pH = 3.0 (see Figure 6). Thus, the results from the two crystal structures on the one hand, and between the three models of the unfolded lysozyme on the other hand, are much more similar to each other with respect to the calculated titration curves than the titration curves of the native and unfolded protein. It is interesting to note that in Figures 5a and 6, the largest difference occurs in the same pH-region (between pH 2.5 and 4). This region involves the carboxyl groups, which have the largest pK_a-shifts in the native relative to the denatured state. The denatured state is reasonably approximated by the null model (Figure 5b), although the extended chain models give better results (see below).

The differences between the calculated titration curves can also be expressed in terms of pK_a-shifts. The average absolute difference between the pK_a's of the null model (i.e., the standard pK_a's of all sites) and the pK_a's of the lysozyme structures are Beta-m1 $\langle|\Delta_{pK}|\rangle = 0.18$; Ex72-m1, 0.27; 21zt, 0.82; and 1hel 0.97 pK-units. It follows that the average pK_a-shift calculated for the extended model structures Beta-m1 and Ex72-m1 of unfolded lysozyme is about one-fourth to one-sixth of the average pK_a-shift of the native protein structure. If the analysis is restricted to the carboxyl groups (there are 10 in lysozyme, out of a total of 32 sites), the average absolute pK_a-shift is Beta-m1 0.20; Ex72-m1, 0.31; 21zt, 1.20; and 1hel 1.37 pK-units, while the average (signed) shift is Beta-m1 $\langle\Delta_{pK}\rangle = -0.18$; Ex72-m1, −0.29; 21zt, −1.20; and 1hel −1.27 pK-units. This confirms that the carboxyl groups dominate in the structural effects on pK_a-values. The calculated average pK_a-shift for the carboxyl groups in the unfolded lysozyme structure, −0.18 (Beta-m1) and −0.29 (Ex72-m1), is in approximate agreement with recent studies by Oliveberg *et al.*,[3] who estimated that, on the average, the pK_a's of carboxyl groups in denatured barnase are 0.4 pK-units lower than the standard pK_a's of the sites.

In Figure 7a,b, the calculated titration curves for the native and unfolded lysozyme structure are shown in comparison with the experimental titration curves taken from ref 11 (see Figure 6). The average difference between the calculated and the experimental titration curves in the experimental pH-range from 1.5 to 10.5 is 0.82 (in elementary charge units) for the crystal structure 21zt. For the minimized structures, the average error is 21zt-m1, 0.92; 21zt-m2, 1.13; and 21zt-m3, 1.24. For the 1hel crystal structure, the average error is also smaller than for the minimized structures (titration curves not shown). It follows

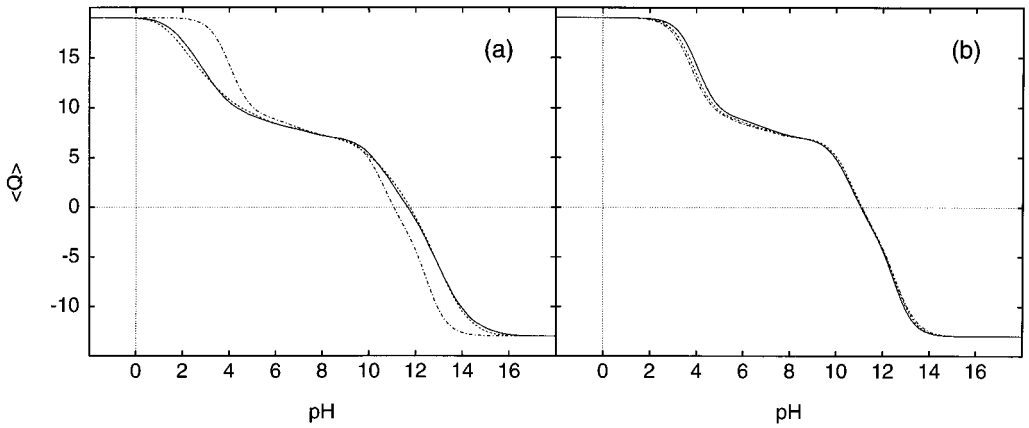

Figure 5. Titration curve of lysozyme: (a) crystal structure 21zt (—) and 1hel(- - -); for comparison, the titration curve according to the null model is shown (- • -); (b) models of the unfolded protein, null (—), Beta-m1 (- - -), and Ex72-m1 (- • -).

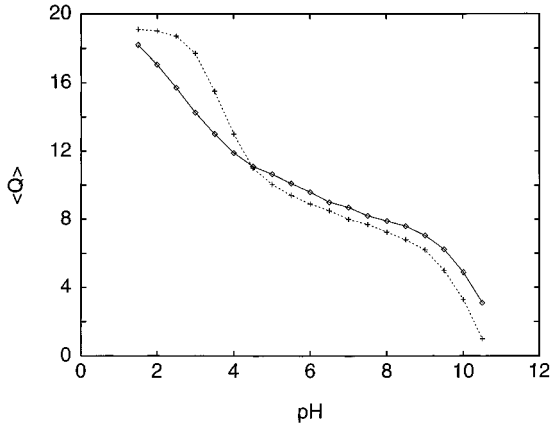

Figure 6. Experimental titration curves of native (◇; solution 0.1 M KCL) and denatured lysozyme (+; solution 6 M guanidine hydrochloride); data taken from ref 11.

that the best agreement between the calculated and experimental titration curves is obtained from the crystal structures of lysozyme, which is in agreement with our findings for the calculated pK_a's of the titrating sites.

For the titration curve of the unfolded lysozyme, we also find good agreement between theory and experiment (see Figure 7b). The average error between the calculated and experimental titration curve is Beta, 0.56; Beta-m1, 0.52; Beta-m2, 0.69; and Beta-m3, 0.79 for the Beta model of the unfolded protein and its minimized structures. The error for the Ex72 models are larger, Ex72, 0.73; Ex72-m1, 0.69; Ex72-m2, 0.70; and Ex72-m3, 0.73 (titration curves not shown). For both extended models of unfolded lysozyme, the agreement with the experimental titration curve is significantly better than for the null model, for which we determined an average error of 1.13. This implies that the extended structures Beta and Ex72 are useful models of the unfolded protein in the context of titration and pH-stability calculations. Furthermore, the agreement between the calculated and the experimental titration curves is best for the unfolded models Beta-m1 and Ex72-m1, i.e., after the first 100 steps of minimization. This may be a consequence of the high conformational energy (i.e., low probability) of the unminimized Beta and Ex72 structures, which are generated by setting all side-chain dihedral angles to 180° and bond length, bond angles, and dihedrals of the protein backbone (including prolines) to ideal values (see section 2.6).

4.3. Relative Stability. From eq 22, the titration curves of two conformers of a system are sufficient for calculating the relative free energy difference as a function of pH. Since we found the best agreement between the calculated and the experimental pK_a's and titration curve of lysozyme when using the crystal structure 21zt and 1hel, we report the results obtained with them. For the same reason, we use Beta-m1 and Ex72-m1 as the models of the unfolded protein.

In Figure 8a, the relative stability $\Delta\Delta G$, eq 22, of triclinic lysozyme is shown relative to the null, Beta-m1, and Ex72-m1 models of the unfolded protein. The experimental stability curve,[64] which has been determined in the pH-range 1.5−7, is also given; the aggregation of lysozyme has so far prevented accurate measurements at high pH.[75] For comparison, we also calculated a "semiexperimental" stability curve using the titration curve integration method with the experimental titration curves

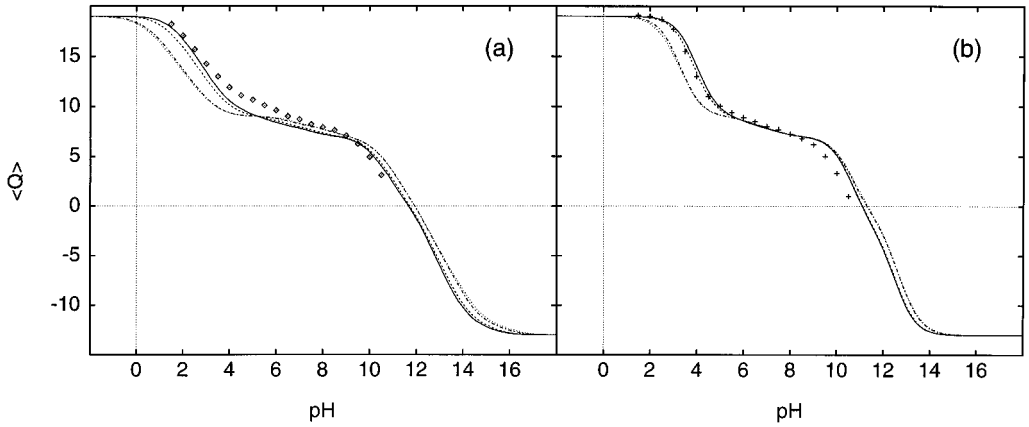

Figure 7. Titration curve of lysozyme: (a) triclinic crystal structure 21zt (—), minimized structures 21zt-m1 (- - -), 21zt-m2 (- • -), and 21zt-m3 (•••); (b) extended model of unfolded lysozyme, Beta (—), and minimized structures Beta-m1 (- - -), Beta-m2 (- • -), and Beta-m3 (•••). Experimental data (◇ and +) same as in Figure 6.

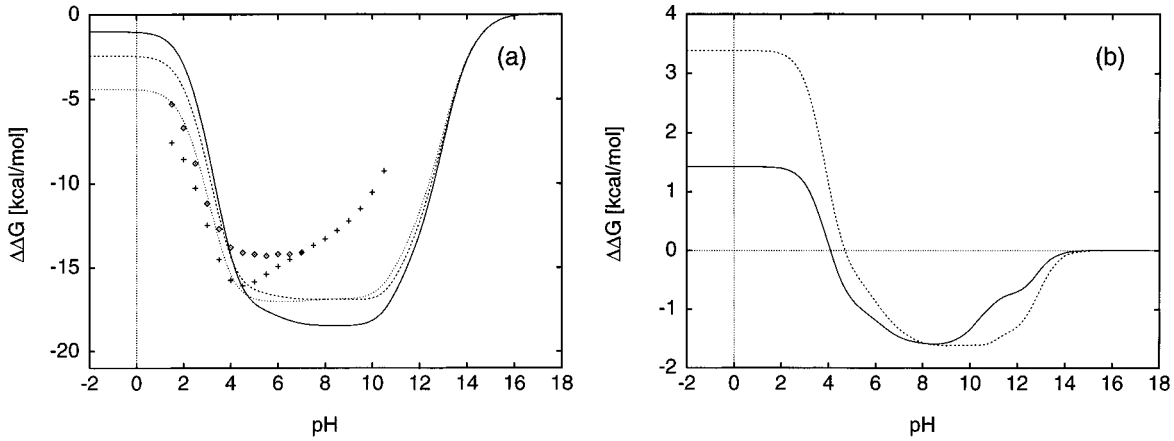

Figure 8. Relative pH-dependent stability: (a) native lysozyme (21zt) using the null (—), the Beta-m1 (- - -), and the Ex72-m1 (···) models of the unfolded protein as the reference state; experimental data (◇) for the net stability (total folding free energy, including nonelectrostatic contributions) of native lysozyme in 0.1 M NaCl solution taken from ref 64; data points (+) calculated with the titration curve integration method, eq 21, using the experimental titration curves[11] of the native and the unfolded lysozyme for the integration and setting $\Delta\Delta G$ at pH = 7 equal to the experimental value from ref 64; (b) unfolded models Beta-m1 (—) and Ex72-m1 (- - -) of lysozyme, using the null model as the reference.

of the folded and unfolded lysozyme.[11] We used pH = 7 with its associated experimental stability from ref 64 as the reference pH and reference stability, since the experimental titration curves do not extend to a pH where the unprotonated state dominates (see eq 22 and related text). The difference between the experimental stability curve and that estimated from the experimental titration curves suggests that there are significant inaccuracies in calculating the stability in the latter. Potential sources for this inaccuracy are the assumption that the nonelectrostatic contributions to protein stability are pH-independent and differences in the experimental conditions in the titration curve measurements of the native and unfolded states (see below).

Regardless of the structures used, the relative stability curves always approach zero at high pH. This is a consequence of the definition of $\Delta\Delta G$, which is the difference between the free energies ΔG_A and ΔG_B of the two conformers, where ΔG_A and ΔG_B are the free energies relative to those of the unprotonated states, $\Delta G_X = G_X - E_{el}^X(\bar{0})$. At high pH, only the unprotonated state contributes to the free energies G_A and G_B of the conformers, such that the relative free energies ΔG_A and ΔG_B as well as their difference approach zero (see eqs 14 and 15).

In this work, the fully unprotonated state has been defined as the reference state for the relative free energies ΔG_A and ΔG_B of the conformers. Other choices are possible, and they would lead to a different offset for the relative stability curves in Figure 8. For example, the fully protonated state of the system or the charge state where all sites are electrically neutral could be used. In the former case, this would lead to relative free energy curves that approach zero in the limit pH $\rightarrow -\infty$, whereas in the latter case, no pH for which $\Delta G_{AB}(\text{pH}) = 0$ can be given, which is inconsequential for the interpretation of the relative stability curves. It follows that the absolute value of the *relative* stability curves in Figure 8 contains no information and that only the change in the stability upon variation of the pH (e.g., relative to the pH of maximum stability) is of interest. (The absolute stability is discussed in section 4.4.)

It is coincidental, therefore, that the calculated relative stability in Figure 8 and the experimental stability of lysozyme are in the same energy range. First, nonelectrostatic free energy contributions to protein stability are not accounted for in the present theory, and second, the electrostatic free energy difference between the native and the unfolded protein in the unprotonated reference state is not accounted for in the relative stability results (see eq 22 and the thermodynamic cycle in Figure 9).

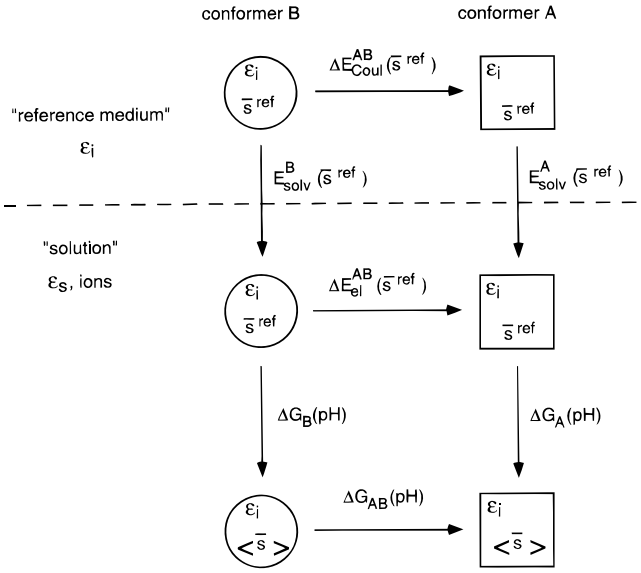

Figure 9. Thermodynamic cycle describing the contributions to the absolute, pH-dependent electrostatic free energy difference $\Delta G_{AP}(\text{pH})$ between two conformations A and B of a titrating system. The unprotonated state is used as the reference state in the upper half of the cycle, $\bar{s}^{ref} = \bar{0}$. For the calculation of Coulomb energies (E_{Coul}) and solvation free energies (E_{solv}), the "reference medium" (on top of the horizontal line) is assigned zero ionic strength and the dielectric constant of the protein interior, ϵ_i. The calculation of relative pH-stability curves ($\Delta\Delta G_{AB}(\text{pH}) = \Delta G_A(\text{pH}) - \Delta G_B(\text{pH})$, eq 22) is concerned with the lower half of the cycle, setting ΔE_{el}^{AB} formally equal to zero. In absolute pH-stability calculations, ΔE_{el}^{AB} is accounted for and is derived from the upper half of the cycle (see text and eqs 20 and 21).

As is expected from the good agreement between the titration curves of Beta-m1 and Ex72-m1 (see Figure 5b), the dependence of $\Delta\Delta G$ on the unfolded model structure used in the calculation is small when compared to the change in the free energy between pH = 7 and the very low and very high pH-limits. In accord with experiment, the theoretical pH-stability curve predicts a plateau for $\Delta\Delta G$ in the pH-range $4 \leq \text{pH} \leq 7$ and an increase of the relative free energy on the order of 10 kcal/mol when the pH is decreased from 4 to 1.5. Pfeil and Privalov[64] report an overall Gibbs free energy difference

between the folded and the denatured state of lysozyme equal to −14.5 kcal/mol under standard conditions (pH = 7, T = 25 °C); at pH = 1.5, the experimental value is −5.3 kcal/mol. Thus, the change in experimental stability between pH = 1.5 and 7 is −9.2 kcal/mol. This is to be compared with the calculated stability difference between pH = 1.5 and 7 of −13.6 kcal/mol (21zt/Beta-m1) and −11.8 kcal/mol (21zt/Ex72-m1). If the null model is used as the unfolded reference state, the predicted change in the stability is −16.5 kcal/mol. The use of the explicit Beta and Ex72 models of the unfolded protein clearly leads to a better agreement with the experimental pH-dependence of the stability of lysozyme.

From Figure 8a, there is good agreement between the change in lysozyme stability from experiment (−8.5 kcal/mol) and from the experimental titration curves of the native and the unfolded protein ("semiexperimental" stability, −8.2 kcal/mol) in the pH-range 1.5−4. On the other hand, the stability change between pH = 1.5 and 7 obtained from the experimental titration curves is −6.5 kcal/mol, as compared to −9.2 kcal/mol from experiment. Since there is good agreement in the pH-range 1.5−4, the disagreement between the experimental and semiexperimental stability in the pH-range 1.5−7 is attributed to an inconsistency in the experimental titration curves, which is most apparent in the pH-range 4−7. Both the experimental and the calculated stability of lysozyme show that there is little change in the stability in the pH-range 4−7. The titration curves of the native and the unfolded lysozyme should thus coincide in this pH-interval. The observed difference between the experimental titration curves in Figure 6 may be a result of differences in the experimental conditions, in particular the ionic strength of the solution used for the measurements.

To analyze the difference between the extended models of the unfolded protein and the null model in detail, we calculated the stability curves of Beta-m1 and Ex72-m1 relative to the null model. From Figure 8, it follows that the relative free energy of Beta-m1 (Ex72-m1) changes by 1.4 (1.3) kcal/mol upon variation of the pH from 7 to 14, and by −2.9 (−4.7) kcal/mol upon pH-change from 2 to 7. This implies that the interactions between titrating sites in the unfolded state, although small, are not negligible. The calculated changes in the stability of 3−5 kcal/mol and about 1.5 kcal/mol upon variation of the pH from 7 to 2 and from 7 to 14, respectively, are likely to be a lower bound to the error that results from the assumption of zero interaction between sites and in the null model, since the extended models are designed for maximum solvent exposure of the titrating sites and, as a consequence, minimal interaction between sites. In a random coil structure or a molten globule state of a protein,[76] residual secondary and tertiary structure would cause the interaction between titrating sites to be larger than for the extended models used in this study.

4.4. Absolute Stability. Use of the extended model of the unfolded protein makes it possible to calculate a meaningful absolute electrostatic free energy difference between the native state and the unfolded state. According to eq 21, this requires the calculation of the electrostatic free energy of the fully unprotonated protein in solution for both conformations (here, the native and the extended structures). The absolute electrostatic free energy difference between the conformers A and B is given by the relative free energy difference plus the difference between the electrostatic free energies of the unprotonated reference states, $\Delta G_{AB}(\mathrm{pH}) = \Delta\Delta G_{AB}(\mathrm{pH}) + \Delta E_{el}^{AB}(\bar{0})$ (see eq 22). It follows that the absolute electrostatic free energy curves in Figure 10 are shifted by $\Delta E_{el}^{AB}(\bar{0}) = E_{el}^{A}(\bar{0}) - E_{el}^{B}(\bar{0})$ relative to the curves in Figure 8.

TABLE 4: Electrostatic Free Energy[a] for the Charge States[b] $\bar{0}$, $\bar{s}^0$, $\bar{s}^{std}$, and $\bar{1}$

structure	$\bar{s}$	$E_{Coul}(\bar{s})$	$E_{solv}(\bar{s})$	$E_{el}(\bar{s})$[c]
21zt	$\bar{0}$	−80.4	−102.5	−182.8
	$\bar{s}^0$	−128.2	−17.1	−145.3
	$\bar{s}^{std}$	−161.2	−77.1	−238.3
	$\bar{1}$	2.4	−186.4	−184.0
1hel	$\bar{0}$	−82.6	−102.8	−185.3
	$\bar{s}^0$	−129.8	−16.8	−146.6
	$\bar{s}^{std}$	−163.1	−77.1	−240.2
	$\bar{1}$	0.8	−186.2	−185.4
Beta-m1	$\bar{0}$	−104.4	−79.4	−183.8
	$\bar{s}^0$	−104.2	−31.2	−135.4
	$\bar{s}^{std}$	−122.9	100.0	−222.9
	$\bar{1}$	−62.7	−119.0	−181.7
Ex72-m1	$\bar{0}$	−120.6	−73.3	−193.9
	$\bar{s}^0$	−125.4	−21.4	−146.7
	$\bar{s}^{std}$	−143.0	−90.0	−233.1
	$\bar{1}$	−75.3	−116.6	−191.9

[a] Energies in kcal/mol; protein $\epsilon_i = 20$, solvent $\epsilon_s = 80$, monovalent ion concentration 0.145 M. [b] $\bar{0}$, unprotonated state; $\bar{s}^0$, all sites neutral; $\bar{s}^{std}$, standard charge state at pH = 7; $\bar{1}$, protonated state. [c] $E_{el} = E_{Coul} + E_{solv}$.

The relationship between the relative and the absolute pH-stability is illustrated by the thermodynamic cycle in Figure 9. The calculation of the relative free energy difference between two conformers A and B involves the lower half of the cycle. In the present theory, we use the unprotonated state $\bar{s}^{ref} = \bar{0}$ as the reference for the relative pH-dependent free energy of each conformer, $\Delta G_A(\mathrm{pH})$ and $\Delta G_B(\mathrm{pH})$; that is, the free energy difference $\Delta E_{el}^{AB}(\bar{0})$ is formally set equal to zero in the relative pH-stability calculations. For the absolute pH-stability, the electrostatic free energy difference between conformers A and B in their unprotonated reference states is accounted for explicitly. From the upper half of the thermodynamic cycle, it is given by the sum of the Coulomb energy difference and the electrostatic solvation free energy difference between A and B, $\Delta E_{el}^{AB}(\bar{0}) = \Delta E_{Coul}^{AB}(\bar{0}) + \Delta E_{solv}^{AB}(\bar{0})$.

In Table 4, the Coulomb energy, the electrostatic solvation free energy, and their sum, the electrostatic free energy in solution, are given in four different charge states for the structures 21zt, 1hel, Beta-m1, and Ex72-m1 used for the stability calculations. The electrostatic free energy in solution is always the lowest for the standard charge state (all sites in the standard state at pH = 7); this is in accord with the fact that lysozyme is most stable in the pH-range from 5 to 9. Interestingly, the Coulomb energy alone is also the lowest for the standard charge state of each structure, but the relative ranking of the four charge states with respect to E_{Coul} and E_{el} is different. Whereas the two crystal structures 21zt and 1hel have very similar values in each of the charge states, the unfolded structure Ex72-m1 has energies in all charge states that are about 10 kcal/mol lower than for Beta-m1. Since the only electrostatic energy term that all four charge states have in common is the interaction of (nontitrating) polar groups, the more negative electrostatic free energy of Ex72-m1 can be attributed to the hydrogen bonds of the backbone that are present in Ex72-m1, but not in Beta-m1.

The electrostatic free energy differences between the four pairs of native and unfolded lysozyme structures are given in Table 5. The same set of charge states as in Table 4 are compared. Since the values of $E_{el}^{A}(\bar{s})$ for 21zt and 1hel are very similar, $\Delta E_{el}^{AB}(\bar{s})$ depends on whether Beta-m1 or Ex72-m1 is used for the unfolded state. As explained above, the folding free energy $\Delta E_{el}^{AB}(\bar{0})$ of the unprotonated state is required to calculate the absolute pH-stability from the relative stability. While the other charge states could, in principle, also

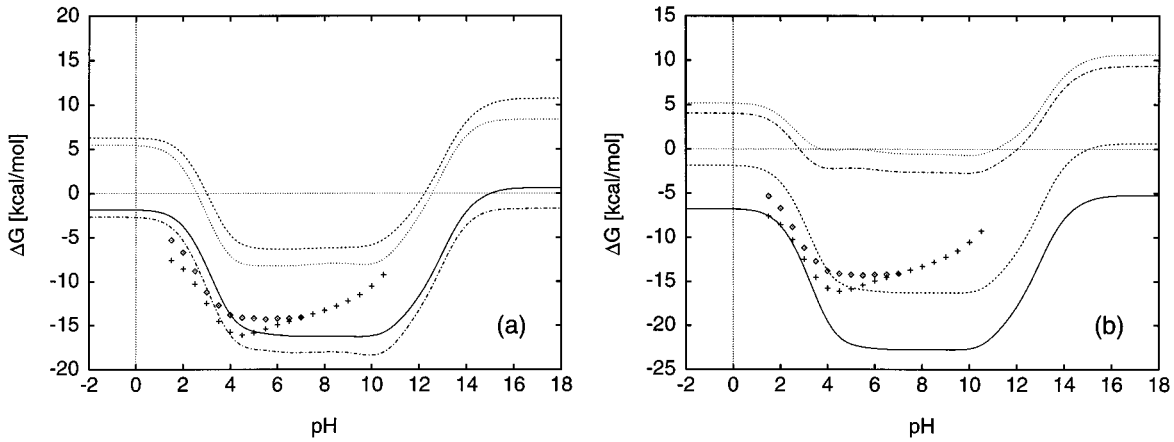

Figure 10. pH-dependent electrostatic contribution to the absolute stability of lysozyme: (a) triclinic crystal structure 21zt relative to the Beta-m1 (—) model, 21zt relative to the Ex72-m1 (- - -) model, tetragonal crystal structure 1hel relative to the Beta-m1 (- · -) model, and 1hel relative to the Ex72-m1 (···) model of the unfolded protein; (b) crystal structure 21zt relative to Beta (—), Beta-m1 (- - -), Beta-m2 (- · -), and Beta-m3 (···). Experimental (◇) and semiexperimental (+) stability data same as in Figure 8a.

TABLE 5: Electrostatic Free Energy Difference[a] for the Charge States[b] 0, $\bar{s}^0$, $\bar{s}^{std}$, and 1

structures			
native	unfolded	$\bar{s}$	$\Delta E_{el}^{AB}(\bar{s})$[c]
21zt	Beta-m1	$\bar{0}$	1.0
		$\bar{s}^0$	−9.9
		$\bar{s}^{std}$	−15.4
		$\bar{1}$	−2.3
21zt	Ex72-m1	$\bar{0}$	11.0
		$\bar{s}^0$	1.4
		$\bar{s}^{std}$	−5.2
		$\bar{1}$	7.9
1hel	Beta-m1	$\bar{0}$	−1.5
		$\bar{s}^0$	−11.2
		$\bar{s}^{std}$	−17.2
		$\bar{1}$	−3.7
1hel	Ex72-m1	$\bar{0}$	8.5
		$\bar{s}^0$	0.1
		$\bar{s}^{std}$	−7.1
		$\bar{1}$	6.5

[a] Energies in kcal/mol; parameters as in Table 4. [b] Charge states as defined in Table 4. [c] $\Delta E_{el}^{AB} = E_{el}^{A} - E_{el}^{B}$.

be used as the reference states in the thermodynamic cycle in Figure 9, their folding free energies are mainly of interest for comparison with the calculated pH-stability curves. The folding free energies in Table 5 are all in a range from about 0 to −20 kcal/mol. Even with the use of a relatively high dielectric constant of $\epsilon_i = 20$ for the protein interior, the solvation free energies in Table 4 are in a range from −10 to −190 kcal/mol. The error in the FDPB calculation of solvation free energies must, therefore, be less than a few percent to obtain meaningful free energy differences between conformers.

In Figure 10a, the absolute stability curves for the four combinations of the native and the unfolded state 21zt/Beta-m1, 21zt/Ex72-m1, 1hel/Beta-m1, and 1hel/Ex72-m1 are shown. They have almost identical shape, which is consistent with the good agreement between the relative stability curves in Figure 8a (21zt only, curves for 1hel not shown). However, the calculated absolute free energy difference at a given pH varies by approximately 15 kcal/mol, depending on the choice of the native and unfolded structure. Use of the tetragonal crystal structure 1hel instead of the triclinic crystal structure 21zt leads to a decrease by about 2 kcal/mol of the calculated free energy of the native state, i.e., to an increase of the calculated stability (negative shift). On the other hand, use of the Ex72-m1 structure instead of Beta-m1 for the unfolded conformation leads to a decrease of the free energy of the unfolded state by about 10 kcal/mol, i.e., to a decrease of the calculated stability (positive shift, upper two curves in Figure 10a).

To estimate the dependence of the calculated pH-stability on the degree of minimization of the unfolded model structure, we calculated stability curves with the 21zt crystal structure for the native protein and with the extended models Beta, Beta-m1, Beta-m2, and Beta-m3 for the unfolded state. The stability curves in Figure 8b show a variation of the absolute free energy difference equal to 22 kcal/mol; at pH = 7, the absolute free energy difference varies from −22.8 to −0.5 kcal/mol.

A comparison between the folding free energy $\Delta E_{el}^{AB}(\bar{s}^{std})$ of lysozyme in the standard charge state from Table 5 with the calculated pH-stability $\Delta G_{AB}(pH)$ at pH = 7 from Figure 10 shows that there are only minor differences; for the native/unfolded structure pairs 21zt/Beta-m1, 21zt/Ex72-m1, 1hel/Beta-m1, and 1hel/Ex72-m1, one has $\Delta E_{el}^{AB}(\bar{s}^{std}) = -15.4$, −5.2, −17.2, and −7.1 kcal/mol, while $\Delta G_{AB}(pH=7) = -15.9$, −6.0, −17.9, and −8.0 kcal/mol, i.e., a maximum difference of 0.9 kcal/mol. It follows that the standard charge state represents well the equilibrium distribution of charge states in the pH-range between 6 and 10, in which there is little change in the calculated pH-stability. In part, this is a consequence of the fact that only two gropus (His, N-terminus) titrate in this pH-range and that their titration behavior is similar in the folded and unfolded state.

The electrostatic free energy of folding for the electrically neutral state, $\bar{s} = \bar{s}^0$, differs from the folding free energy of the standard charge state (corresponding to the broad minima in the stability curves in Figure 10) by about 6−7 kcal/mol. The value is nearly independent of the pair of crystallographic and extended structures that are used to represent the native and the unfolded states. Furthermore, the folding free energy of the uncharged state (polar interactions only) exhibits changes as a function of the native/unfolded structures that are very similar to the changes observed for the minimum of the pH-stability in Figure 10 (and for the folding energy of the standard charge state). It follows that it is the variation in the interactions of polar atom groups in the different lysozyme structures that accounts for the variation in the calculated absolute pH-stability of lysozyme and that the interaction of titrating sites is comparatively independent of the choice for the native and unfolded structure. One reason for this difference between the contributions from polar and charged atom groups is that the latter are involving mainly long-range interactions, whereas the former are due to short-range interactions (e.g., hydrogen bonds), which are more sensitive to structural changes.

TABLE 6: Electrostatic Free Energy Difference[a] for the Protonated State

structures		$\Delta E_{el}^{AB}(\bar{s})^b$	$\Delta G_{AB}(pH)^c$	
native	unfolded	$\bar{s} = 1$	pH = −5	error[d]
21zt	Beta-m1	−2.29	−1.44	−0.85
21zt	Ex72-m1	7.85	6.64	1.22
1hel	Beta-m1	−3.67	−2.42	−1.26
1hel	Ex72-m1	6.47	5.66	0.81
21zt	Beta	−6.84	−6.16	−0.69
21zt	Beta-m1	−2.29	−1.44	−0.85
21zt	Beta-m2	4.02	4.64	−0.62
21zt	Beta-m3	4.90	5.84	−0.94

[a] Energies in kcal/mol; parameters as in Table 4. [b] $\Delta E_{el}^{AB} = E_{el}^{A} - E_{el}^{B}$; energies E_{el} from Table 4. [c] ΔG_{AB}(pH=−5) from stability curves in Figure 10, eq 18. [d] Differences between previous two columns.

To test the self-consistency of the present method for calculating absolute pH-stability, the electrostatic energy of folding for the protonated state (see Table 5) is compared with the low pH-limit (pH = −∞) of the absolute pH-stability. The eight pairs of native/unfolded lysozyme structures for which the absolute stability is shown in Figure 10 were used for the comparison. Since there is virtually no variation in the calculated pH-stability of lysozyme below pH = 0, the values of the stability curves at pH = −5 (outside of the range of the plot in Figure 10) are used to represent the limit ΔG_{AB}(pH→ −∞). According to eq 21, the pH-stability curve is calculated by integrating the difference between the titration curves of the native and unfolded structures, beginning with the folding free energy of the unprotonated state in the limit pH → ∞. Since the fully protonated charge state dominates the generating (partition) function in the limit pH → −∞, the pH-stability curve should approach the folding free energy of the protein in the protonated state in the low pH limit.

As explained in section 3.3, the electrostatic self-energies and interaction energies E_{ij} (used in the titration and subsequent pH-stability calculation) for all pairs of titratable sites are calculated by a series of FDPB calculations, while the Coulomb and solvation free energies of the fully unprotonated and protonated states are calculated separately to obtain the folding free energies for the various charge states. From Table 6, the energies $\Delta E_{el}^{AB}(\bar{s} = \bar{1})$ and ΔG_{AB}(pH=−5) differ on average by about 1 kcal/mol. This small difference demonstrates the consistency of the methodology. Potential contributions to the 1 kcal/mol difference arise from the FDPB calculations, the titration calculation with the MC program, the titration curve integration method to obtain the pH-stability, and the exclusion of some nonbonded Coulomb interactions in the self-energies and interaction energies E_{ij} of the titrating sites despite the fact that these terms are part of the conformational energy of the protein (see eq 6, section 3.3, and Appendix C). By analyzing the dependence of the results on the number of MC steps used for the titration calculation and on the pH-intervals used for the titration curve integration, we found that only the FDPB calculations and the inconsistent treatment of nonbonded Coulomb terms in the present approach remain as the sources of the difference that is reported in Table 6. In the present analysis, the error is small when compared with the relevant energies, the stability of lysozyme, and its pH-dependence. However, in systems with large numbers of sites, the error could be more substantial. A more detailed analysis of the accuracy and self-consistency of the methodology is beyond the scope of this work and will be the subject of a future publication.

The experimental data on lysozyme stability include all contributions to the free energy, while the present theory accounts only for the electrostatic energy contribution. Therefore, the comparison with experiment in Figure 10 is only qualitative. The comparison demonstrates that the calculated electrostatic free energy contribution to lysozyme stability is within ±15 kcal/mol from the net stability. However, there is considerable uncertainty in the electrostatic contribution depending on which structures are used to represent the native and the unfolded state. In an improved theoretical treatment of the pH-stability of proteins, it will, therefore, be necessary to include conformational equilibria or to devise a method for selecting "good" conformational representatives for the native and the unfolded state (see below).

5. Discussion

In this study, a method for calculating the absolute pH-dependent electrostatic free energy difference between two conformers of a system has been described and applied to the calculation of the free energy difference between native and unfolded lysozyme. The method represents a significant extension of published studies on the pH-stability of proteins,[2,18] which were restricted to relative free energy differences as a function of pH. For the calculation of absolute free energy differences, it is necessary to evaluate the electrostatic free energy difference ΔG_{AB} between the two conformes A and B in a chosen reference state. In the present work, we have used the fully unprotonated state as a reference, which corresponds to $\Delta G_{AB}(\bar{0})$ or ΔG_{AB} at pH = ∞ (see eq 21).

The importance of the absolute, over the relative, stability is that the former includes all electrostatic contributions to the free energy difference between conformations. Consequently, from the absolute pH-dependent free energy difference ΔG_{AB}, it is possible to predict whether conformer A or conformer B is electrostatically favored at a given pH. By contrast, the relative pH-stability can be used to determine only the *change* of the stability as a function of pH. Since the unprotonated state was chosen as the reference for the relative free energy of each conformer, the relative free energy difference approaches zero in the high pH limit, $\Delta\Delta G_{AB}$(pH→∞) = 0. The thermodynamic cycle in Figure 9 illustrates the relationship between the relative and the absolute pH-stability. Whereas in the absolute stability calculation, the electrostatic free energy difference $\Delta E_{el}^{AB}(\bar{s}^{ref})$ for the reference charge state is included, it is formally set equal to zero for the relative stability. As a consequence, the absolute pH-stability curves in Figure 10 are shifted by $\Delta E_{el}^{AB}(\bar{s}^{ref})$, as compared with the relative stability curves in Figure 8.

The calculation of the electrostatic free energy of lysozyme in the unfolded state is made possible by the use of extended structures as models of the denatured protein. The extended model of lysozyme was oriented parallel to the *z*-axis of the Cartesian space and included in an elongated finite-difference grid, such that the FDPB calculation is tractable using a workstation with 64 MB of computer memory. Despite these improvements, we were restricted in the FDPB calculations of $\Delta G_{AB}(\bar{0})$ to a grid constant of 0.7 Å. The use of a grid constant that is of the same order as the smallest van der Waals radius in the system can lead to an error as large as several kcal/mol in the calculated solvation energy of a protein.[70,77] An added difficulty in the accurate calculation of $\Delta G_{AB}(\bar{0})$ resulted from the fact that a protein in its unprotonated state has a high negative net charge, in general. This introduces a large solvation free energy when compared with that of the system in its standard charge state at pH = 7. However, the consistency of the calculated absolute electrostatic free energy difference between folded and unfolded lysozyme in the low pH-limit, ΔG_{AB}(pH→−∞), and the folding free energy of the protonated state, $\Delta G_{AB}(\bar{1})$, indicates that the actual error is significantly smaller (see Table 6).

From the calculation of the relative pH-stability of lysozyme, where the free energy difference $\Delta\Delta G = 0$ at pH $= \infty$ is used as the reference, we found at pH $= 7$ a stabilization by -17 kcal/mol relative to the high pH-limit (pH > 16) and by -12 to -13.5 kcal/mol relative to the low pH-limit (pH < 1.5). This result, which was obtained using the PDB structure for the native protein and the extended model for the unfolded state, is in reasonable agreement with the experimental observation that the stability of lysozyme varies by -9.2 kcal/mol when the pH is increased from 1.5 to 7. The result obtained with the extended model for the unfolded state is better than the null model of noninteracting sites, which leads to a predicted change in the stability by -16.5 kcal/mol when the pH is varied from 1.5 to 7. There are significant interactions between the titrating residues for the fully extended structure of lysozyme, which represents a limiting case of the denatured protein. In the case of lysozyme, the interactions in the extended structure give rise to a net stabilization of -1.5 to -4.7 kcal/mol at pH $= 7$, relative to the null model.

By comparing the calculated pH-stability of various models for the native and unfolded structure of lysozyme, we have demonstrated that there is a strong dependence of the absolute pH-stability on the choice of structures, particularly that of the denatured state. There is a variation of as much as 25 kcal/mol in $\Delta G_{AB}(\bar{0})$, the value for the fully unprotonated state. However, other than this "offset" of the absolute stability curve at pH $= \infty$, the pH-stability of lysozyme is almost unaffected by the variation in the structures used for the calculations; that is, the absolute stability curves were found to be parallel over the entire pH range. The data in Table 5 show that the variation of the folding free energy $\Delta E_{el}^{AB}(\bar{s}^{ref})$ for the electrically neutral charge state ($\bar{s}^{ref} = \bar{s}^0$) as a function of the native and unfolded structures A and B is the same as the variation of the folding free energy for the standard charge state. Since only the polar atom groups, in particular hydrogen bonds, contribute to the electrostatic free energy of the neutral charge state, it follows that small structural changes have a strong effect on the interactions of polar groups (and, consequently, on the electrostatic free energy of the unprotonated state), whereas the pK_a's and the titration curve of a structure (and, consequently, the relative pH-dependent free energy difference between conformers) are less sensitive to conformational change.

The method for calculating pK_a's and absolute pH-stability curves introduced in this paper can be improved with respect to several aspects: First, a continuum electrostatic model where the protein "core" ($\epsilon_{core} \approx 2-4$) and atom groups at the surface ($\epsilon_{surf} \approx 20-80$) are treated differently can be introduced;[66] that is, the assignment of more appropriate dielectric constants may lead to improved predictions, particularly for the pK_a's of buried sites.[47] Second, standard electrostatic free energies of protonating titrating sites in the model peptides could be used, instead of the current practice of using the energies calculated for individual model compounds with different structures for each site. This would eliminate the error due to the inconsistent treatment of intraresidue Coulomb interactions in the present methodology (see Appendix C). Third, a more detailed treatment of the interaction between titrating sites (including the background atoms) would result from the consideration of multiple protonation states, instead of the average-charge two-state model employed here. Fourth, conformational averages could be introduced for the calculation of the pH-dependent electrostatic free energy of structures, particularly for multiple side-chain conformers.[12,37] This would reduce the strong dependence of the absolute pH-stability on the choice of structures observed in this study. The average over multiple conformations could be made at various levels of the theory, e.g., at the conformational level, at the electrostatic free energy level, at the generating function level, and at the titration curve level. The various improvements in the methodology described here are under investigation.[78]

Even though this paper has been concerned with the pH-stability of proteins, the present theory can be directly applied to the calculation of the absolute, pH-dependent binding free energy of a complex of two molecules with titrating sites, e.g., the binding energy for the inhibitor of a protein, given the structure of the complex. For the calculation of binding energies, the bound/unbound states correspond to the folded/unfolded state (i.e., conformer A/B) in the present work. Since in the unbound state the two molecules of a complex titrate independently, the titration curve for the unbound state is given by the sum of the titration curves of the constituent molecules. Correspondingly, the electrostatic free energy of the unbound state is the sum of the electrostatic free energies (Coulomb energy plus solvation free energy, see Figure 9) of the two molecules. On the basis of the present work, it is thus possible to predict the total electrostatic contribution to the binding free energy of a complex of two molecules at any given pH, including the contribution from the interaction of polar atom groups. To obtain the full binding interaction, other contributions (e.g., hydrophobic and van der Waals terms, configurational entropy) have to be included.

Acknowledgment. We are grateful to J. A. McCammon and co-workers for making available a copy of the UHBD[41] program, and to P. Beroza *et al.*[1] for a MC titration program. We thank H. van Vlijmen for many helpful discussions.

Appendix A. Energy of a Protonation State

To derive eq 8 for the electrostatic free energy of a protonation state, $\Delta G(\bar{s},\mathrm{pH})$, we first rewrite eq 5 for the pK-shift between the $\mathrm{p}K_{a,i}^{intr}$ and the $\mathrm{p}K_{a,i}^{stnd}$ of site i (see Figure 1) in terms of atom group contributions,

$$\begin{aligned}
&-(\ln 10)k_B T(\mathrm{p}K_{a,i}^{intr} - \mathrm{p}K_{a,i}^{stnd}) \\
&= \Delta E_1 - \Delta E_0 \\
&= [E_{el}(s_i{=}1, s_{j\neq i}{=}s_j^0) - E_{el}^{M_i}(s_i{=}1)] - \\
&\qquad [E_{el}(s_i{=}0, s_{j\neq i}{=}s_j^0) - E_{el}^{M_i}(s_i{=}0)] \\
&= {}^1/_2[\Delta E_{ii}(1,1) - \Delta E_{ii}(0,0)] + \Delta E_{i0}(1,0) - \Delta E_{i0}(0,0) + \\
&\qquad \sum_{j\neq i}^{N} [E_{ij}(1,s_j^0) - E_{ij}(0,s_j^0)] \quad \text{(A-1)}
\end{aligned}$$

where s_j^0 is defined as the uncharged state of site j and where

$$\begin{aligned}
\Delta E_{ii}(s_i,s_i) &= E_{ii}(s_i,s_i) - E_{ii}^{M_i}(s_i,s_i) \\
\Delta E_{i0}(s_i,0) &= E_{i0}(s_i,0) - E_{i0}^{M_i}(s_i,0) \qquad \text{(A-2)}
\end{aligned}$$

The sum over $j \neq i$ is over all titrating sites except i and does not include the interaction with the background atoms (site index 0), which is accounted for explictly (ΔE_{i0} terms). In going from eq 5 to eq A-1, the self-energy of the background atoms, $(1/2)E_{00}$, in the protein and in the model compound cancels because it contributes equally to the energy of the protein (model compound) with site i protonated or unprotonated.

From eq 1 and the definition of the net charge of site i in charge state s_i and in the unprotonated state, $q_i(s_i)$ and $q_i(0)$, it follows that $q_i(s_i) = q_i(0) + s_i$. Furthermore, the uncharged

state s_i^0 and the net charge $q_i(0)$ of site i in the unprotonated state are related according to $q_i(0) = -s_i^0$, since in the two-state model

$$q_i(0) = \begin{cases} -1, & \text{if site } i \text{ anionic} \\ 0, & \text{if site } i \text{ cationic} \end{cases}$$

$$s_i^0 = \begin{cases} 1, & \text{if site } i \text{ anionic} \\ 0, & \text{if site } i \text{ cationic} \end{cases} \quad \text{(A-3)}$$

Using eq A-1 for $pK_{a,i}^{intr}$ and eq 3 for W_{ij}, it is thus possible to rewrite eq 2 for the energy of a protonation state in the form

$$\begin{aligned}\Delta G(\bar{s},\mathrm{pH}) &= (\ln 10)k_B T\sum_{i=1}^{N} s_i(\mathrm{pH} - pK_{a,i}^{stnd}) + \\ &\sum_{i=1}^{N} s_i[{}^1/_2(\Delta E_{ii}(1,1) - \Delta E_{ii}(0,0)) + \Delta E_{i0}(1,0) - \\ &\Delta E_{i0}(0,0)] + \sum_{i<j}^{N} \{s_i[E_{ij}(1,s_j^0) - E_{ij}(0,s_j^0)] + s_j[E_{ij}(s_i^0,1) - \\ &E_{ij}(s_i^0,0)]\} + \sum_{i<j}^{N} (s_i s_j - s_i s_j^0 - s_i^0 s_j)\,[E_{ij}(1,1) - \\ &E_{ij}(1,0) - E_{ij}(0,1) + E_{ij}(0,0)] \\ &= (\ln 10)k_B T\sum_{i=1}^{N} s_i(\mathrm{pH} - pK_{a,i}^{stnd}) + \\ &\sum_{i=1}^{N} [{}^1/_2(\Delta E_{ii}(s_i,s_i) - \Delta E_{ii}(0,0)) + \Delta E_{i0}(s_i,0) - \\ &\Delta E_{i0}(0,0)] + \sum_{i<j}^{N} [E_{ij}(s_i,s_j) - E_{ij}(0,0)] \quad \text{(A-4)}\end{aligned}$$

In deriving eq A-4, we made use of equalities of the type

$$\begin{aligned}s_i s_j[E_{ij}(1,1) - E_{ij}(1,0) - E_{ij}(0,1) + E_{ij}(0,0)] &= [E_{ij}(s_i,s_j) - \\ E_{ij}(s_i,0) - E_{ij}(0,s_j) + E_{ij}(0,0)] \quad &\text{(A-5)}\end{aligned}$$

which can be verified by inserting all possible values for s_i, $s_j \in \{0,1\}$.

Equation A-4 can be brought into the form of eq 8 by adding and subtracting the self-energy of the background atoms in the protein, $(1/2)E_{00}(0,0)$, adding and subtracting the sum of the self-energies of the background atoms in the model compounds, $(1/2)\sum_{i=1}^{N} E_{00}^{M_i}(0,0)$, and regrouping,

$$\begin{aligned}\Delta G(\bar{s},\mathrm{pH}) &= (\ln 10)k_B T\sum_{i=1}^{N} s_i(\mathrm{pH} - pK_{a,i}^{stnd}) + \\ &{}^1/_2\sum_{i,j=0}^{N} [E_{ij}(s_i,s_j) - E_{ij}(0,0)] - \sum_{i=1}^{N} {}^1/_2 \sum_{k,l\in\{0,i\}} [E_{kl}^{M_i}(s_k,s_l) - \\ &E_{kl}^{M_i}(0,0)] \quad \text{(A-6)} \\ &= (\ln 10)k_B T\sum_{i=1}^{N} s_i(\mathrm{pH} - pK_{a,i}^{stnd}) + E_{el}(\bar{s}) - \\ &E_{el}(\bar{0}) - \sum_{i=1}^{N} (E_{el}^{M_i}(s_i) - E_{el}^{M_i}(0)) \quad \text{(A-7)}\end{aligned}$$

The sum over $k, l \in \{0, i\}$ in eq A-6 denotes the sum over the four pairs of indices $(k, l) = (0, 0)$, $(0, i)$, $(i, 0)$, and (i, i).

Appendix B. Multiple Protonation States

In general, the partial charges of the entire side chain of a titrating residue may vary as a function of the protonation state. A treatment of sites with more than two protonation states as two (or more) independent titration sites leads, therefore, to a considerable complication of the methodology.[47] The effect of including all possible protonation states of the sites in the calculation of pK_a's and titration curves, as compared with the present two-state model, will be the subject of a future study.[40] This is of considerable interest for the prediction of the average and the most probable protonation states of titrating groups at a given pH, e.g., histidines or aspartates, in particular where a specific protonation state is required for catalytic activity.

To extend the theory to the case of multiple protonation states, we do not require that the component s_i is equal to the number of protons bound to site i; that is, the component s_i is merely an index for enumerating all possible protonation states of the site. Without loss of generality, we assume that the charge index of site i is from 0 to $N_i - 1$, where N_i is the number of charge states of the site and where $s_i = 0$ is the unprotonated state. For the number of protons bound to site i in charge state s_i we write $n_i(s_i)$. Furthermore, we introduce the parameter $pK_{a,i}^{stnd}(s_i)$, which is the standard pK_a of the reaction

$$M_i(0) + n_i(s_i)H^+ \rightarrow M_i(s_i) \quad \text{(A-8)}$$

With these definitions, eq 8 can be rewritten in the form

$$\begin{aligned}\Delta G(\bar{s},\mathrm{pH}) = (\ln 10)k_B T\sum_{i=1}^{N}(n_i(s_i)\mathrm{pH} - pK_{a,i}^{stnd}(s_i)) + E_{el}(\bar{s}) - \\ E_{el}(\bar{0}) - \sum_{i=1}^{N}(E_{el}^{M_i}(s_i) - E_{el}^{M_i}(0)) \quad \text{(A-9)}\end{aligned}$$

which is applicable to the general case of multiple protonation states. The general form of the formula for the absolute energy $G(\bar{s},\mathrm{pH})$ of a protonation state, eq 9, is obtained by the same change in the first, pH-dependent term on the right-hand side.

In the following, the equations beyond eq 9 in section 2 which are restricted to the two-state model are given in the form applicable to the case of multiple states; the equation numbers are those of section 2, preceded by the letter M. While eq 12 for the average protonation of the system remains unchanged, with the number of protons bound to the system in state $\bar{s}$ given by $n(\bar{s}) = \sum_{i=1}^{N} n_i(s_i)$, the probability $P(s_i')$ of finding site i in state s_i' is now given by the ensemble average

$$P(s_i')(\mathrm{pH}) = \frac{1}{\Xi(\mathrm{pH})}\sum_{\bar{s}} \delta_{s_i s_i'} \exp(-\beta G(\bar{s},\mathrm{pH})) \quad \text{(M-13)}$$

The Kronecker symbol $\delta_{s_i s_i'}$ is equal to 1 if $s_i = s_i'$ and equal to 0 otherwise.

According to the null model of noninteracting sites (multiple states), the energy of protonation state $\bar{s}$ relative to the unprotonated state is given by

$$\Delta G_{null}(\bar{s},\mathrm{pH}) = (\ln 10)k_B T\sum_{i=1}^{N}(n_i(s_i)\mathrm{pH} - pK_{a,i}^{stnd}(s_i)) \quad \text{(M-23)}$$

To derive formulas for the relative free energy and protonation curve from the null model, we introduce the pH-dependent "site energy" $\Delta G_i(s_i,\mathrm{pH})$,

$$\Delta G_i(s_i,\mathrm{pH}) = (\ln 10)k_B T(n_i(s_i)\mathrm{pH} - pK_{a,i}^{stnd}(s_i)) \quad \text{(M-24)}$$

The generating function $\Xi_{null}(pH)$ and the relative free energy $\Delta G_{null}(pH)$ according to the null model are

$$\Xi_{null}(pH) = \sum_{\bar{s}} \{\prod_{i=1}^{N} \exp(-\beta \Delta G_i(s_i,pH))\} = \prod_{i=1}^{N} \{\sum_{k=0}^{N_i-1} \exp(-\beta \Delta G_i(k,pH))\} \quad \text{(M-26)}$$

$$\Delta G_{null}(pH) = -k_B T \sum_{i=1}^{N} \ln\{\sum_{k=0}^{N_i-1} \exp(-\beta \Delta G_i(k,pH))\} \quad \text{(M-27)}$$

where N_i is the number of charges states of site i. In the null model, the sites titrate independently with their standard pK_a's, such that the pH-dependent average protonation of the system is

$$\langle n(pH)\rangle_{null} = \sum_{i=1}^{N} \left\{ \frac{1}{\Xi_i(pH)} \sum_{k=0}^{N_i-1} n_i(k) \exp(-\beta \Delta G_i(k,pH)) \right\} \quad \text{(M-28)}$$

$$\Xi_i(pH) = \sum_{k=0}^{N_i-1} \exp(-\beta \Delta G_i(k,pH))$$

When applying the Monte Carlo titration method to the case of multiple protonation states, the probability of having site i in charge state s_i' is

$$P(s_i')(pH) = \frac{1}{N(E)} \sum_{\bar{s} \in E} \delta_{s_i s_i'} \quad \text{(M-29)}$$

where $N(E)$ is the number of states in the ensemble E and where the sum is over all charge states in the Monte Carlo ensemble. Equation 30 for the average protonation of the system remains unchanged, with the number of protons bound to the system given by $n(\bar{s}) = \sum_{i=1}^{N} n_i(s_i)$.

Appendix C. Treatment of Coulomb Interactions

For a detailed understanding of the energetic contributions to the pH-dependent energy $G(\bar{s},pH)$ of the system, eq 9, we write the interaction ($i \neq j$) or self ($i = j$) energy of two charge groups as a sum of the Coulomb energy with the dielectric constant $\epsilon = \epsilon_i$ and the solvation energy contribution for the transfer from ϵ_i to the solvent,[79] $E_{ij} = E_{ij}^{Coul} + E_{ij}^{solv}$, where we omitted the dependence of all terms on the charge state of the atom groups i and j for brevity. When using the FDPB method for calculating the energies E_{ij} as described in section 3.3, the following points related to the Coulomb term should be noted.

(i) The FDPB interaction energy E_{i0} between site i and the background charges in the protein/model compounds also includes bonded Coulomb energy terms which, in theory, do not contribute to the conformational energy of the system. These "unwanted" energy terms, however, are canceled in calculating the energy $G(\bar{s},pH)$, eq 9, because they are equally present in the system, E_{i0}, and in the model compound, $E_{i0}^{M_i}$. The current method of calculating the energies E_{i0} with the FDPB method thus relies on the fact that for each site a model compound is used that has the same structure as the residue of the site.

(ii) In general, there are nonbonded Coulomb energy terms between site i and the background charges (energy E_{i0}) and within a given site i (energy E_{ii}). Even though these energy terms contribute to the conformational energy $E_{el}(\bar{s})$, they are canceled in eq 9 due to subtraction of the corresponding model compound energies. The fact that some terms of the electrostatic free energy are not accounted for in calculating the charge state and conformation-dependent energy of the system, $G(\bar{s},pH)$, represents an inconsistency, since long-range electrostatic interactions are accounted for, while intraresidue electrostatic (Coulomb) interactions are excluded. This inconsistency becomes apparent when comparing the low pH-limit of the absolute pH-stability curve of a protein with the electrostatic free energy of folding of the same protein in the fully protonated state (see Table 6 and related text). The inconsistency can only be resolved by the use of a single, standard electrostatic free energy of protonating sites of a given type, instead of the current approach of using an individual model compound for each site. Such standard model compound energies can be determined by calculating an average over the model compounds of a given type, for example, by averaging over model compounds taken from highly resolved structures in the protein data bank.

References and Notes

(1) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Fehler, G. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5804.
(2) Antosiewicz, J.; McCammon, J. A.; Gilson, M. *J. Mol. Biol.* **1994**, *238*, 415.
(3) Oliveberg, M.; Arcus, V. L.; Fersht, A. R. *Biochemistry* **1995**, *34*, 9424.
(4) Fersht, A. *Enzyme Structure and Mechanism*, 2 ed.; W. H. Freeman and Company: New York, 1985.
(5) Becktel, W. J.; Schellman, J. A. *Biopolymers* **1987**, *26*, 1859.
(6) Schellman, J. A. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 115.
(7) Baldwin, R. L.; Eisenberg, D. In *Protein Engineering*; Oxender, D., Ed.; Alan, R. Liss, Inc.: New York, 1987; p 127.
(8) Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 2403.
(9) Linderstrøm-Lang, K. *C. R. Trav. Lab. Carlsberg* **1924**, *15*, 1.
(10) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333.
(11) Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192.
(12) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219.
(13) Sternberg, M. J. E.; Hayes, F. R. F.; Russell, A. J.; Thomas, P. G.; Fersht, A. R. *Nature* **1987**, *330*, 86.
(14) Gilson, M. G.; Honig, B. H. *Nature* **1987**, *330*, 84.
(15) Cannan, R. K.; Palmer, A. H.; Kibrick, A. C. *J. Biol. Chem.* **1942**, *142*, 803.
(16) Nozaki, Y.; Tanford, C. *J. Am. Chem. Soc.* **1967**, *89*, 742.
(17) Roxby, R.; Tanford, C. *Biochemistry* **1971**, *10*, 3348.
(18) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1993**, *231*, 459.
(19) Zauhar, R. J.; Morgan, R. S. *J. Comput. Chem.* **1988**, *9*, 171.
(20) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.
(21) Yang, A.-S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins* **1993**, *15*, 252.
(22) Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9556.
(23) Gilson, M. K. *Proteins* **1993**, *15*, 266.
(24) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.
(25) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(26) Rashin, A. A. *J. Phys. Chem.* **1990**, *94*, 1725.
(27) Zauhar, R. J.; Varnek, A. *J. Comput. Chem.* **1996**, *17*, 864.
(28) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986**, *1*, 47.
(29) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* **1991**, *30*, 9686.
(30) Rodgers, K. K.; Sligar, S. G. *J. Am. Chem. Soc.* **1991**, *113*, 9419.
(31) Mohan, V.; Davis, M. E.; McCammon, J. A.; Pettitt, B. M. *J. Phys. Chem.* **1992**, *96*, 6428.
(32) Schaefer, M.; Karplus, M. Manuscript in preparation.
(33) Wyman, J. *Adv. Protein Chem.* **1964**, *19*, 223.
(34) Szabo, A.; Karplus, M. *J. Mol. Biol.* **1972**, *72*, 163.
(35) Schellman, J. A. *Biopolymers* **1975**, *14*, 999.
(36) Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 231.
(37) You, T. J.; Bashford, D. *Biophys. J.* **1995**, *69*, 1721.
(38) van Vlijmen, H.; Schaefer, M.; Karplus, M. Manuscript in preparation.
(39) Kirkwood, J. G.; Shumaker, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **1952**, *38*, 855.
(40) Schaefer, M.; Karplus, M. Manuscript in preparation.
(41) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187.

(42) Nozaki, Y.; Tanford, C. *Methods Enzymol*. **1967**, *11*, 715.
(43) Tanokura, M. *Biochim. Biophys. Acta* **1983**, *742*, 576.
(44) Kuramitsu, S.; Hamaguchi, K. *J. Biochem*. **1980**, *87*, 1215.
(45) Bartik, K.; Redfield, C.; Dobson, C. M. *Biophys. J*. **1994**, *66*, 1180.
(46) Rico, M.; Santoro, J.; Gonzalez, C.; Bruix, M.; Neira, J. L. In *Structure, Mechanism and Function of Ribonucleases*, Proceedings of the 2nd International Meeting, Santa Feliu de Guíxols, Girona, Spain, 1990; Cuchillo, C. M., de Llorens, R. d., Nogués, M. V.; Parés, X., Eds.; Department de Bioquímica i Biologia Molecular and Institut de Biologia Fondamental, Unviersitat Autònoma de Barcelona: Bellaterra, Spain, 1991; p 9.
(47) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993**, *32*, 8045.
(48) Luty, B. A.; Davis, M. E.; McCammon, A. *J. Comput. Chem*. **1992**, *13*, 768.
(49) Privalov, P. L. *Physical Basis of the Stability of the Folded Conformations of Proteins: 5. Denatured States of Proteins*; W. H. Freeman: New York, 1992; p 91.
(50) Fersht, A. R.; Itzhaki, L. S.; ElMasry, N. F.; Matthews, J. M.; Otzen, D. E. *Proc. Natl. Acad. Sci. U.S.A*. **1994**, *91*, 10426.
(51) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem*. **1996**, *100*, 2661.
(52) Caflisch, A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A*. **1994**, *91*, 1746.
(53) Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. *Proteins* **1995**, *21*, 196.
(54) Schulz, G. E.; Schirmer, R. H. *Principles of Protein Structures*; Springer: Heidelberg, 1978.
(55) Maccallum, P. H.; Poet, R.; Milner-White, E. J. *J. Mol. Biol*. **1995**, *248*, 374.
(56) Bernstein, F. C.; Koetzle, T. F.; Williams, T. F.; Meyer, G. J. B., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Schimanouchi, T.; Tasumi, M. *J. Mol. Biol*. **1977**, *112*, 535.
(57) Ramanadham, M.; Sieker, L. C.; Jensen, L. H. *Acta Crystallogr. C* **1981**, *37*, 33.
(58) Malcolm, B. A.; Wilson, K. P.; Matthews, B. W.; Kirsch, J. F.; Wilson, A. C. *Nature* **1990**, *345*, 86.
(59) Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148.
(60) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem*. **1983**, *4*, 187.
(61) MacKerell, A. D., Jr.; et al. *FASEB J*. **1992**, *6*, A143.
(62) MacKerell, A. D., Jr.; et al. Manuscript in preparation.
(63) Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Principles of Biochemistry*; Worth Publishers: New York, 1993.
(64) Pfeil, W.; Privalov, P. L. *Biophys. Chem*. **1976**, *4*, 41.
(65) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem*. **1990**, *19*, 301.
(66) States, D. J.; Karplus, M. *J. Mol. Biol*. **1987**, *197*, 111.
(67) Karshikoff, A.; Spassov, V.; Cowan, S. W.; Ladenstein, R.; Schirmer, T. *J. Mol. Biol*. **1994**, *240*, 372.
(68) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem*. **1988**, *9*, 327.
(69) Shrake, A.; Rupley, J. A. *J. Mol. Biol*. **1973**, *79*, 351.
(70) Davis, M. E.; McCammon, J. A. *J. Comput. Chem*. **1991**, *12*, 909.
(71) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys*. **1953**, *21*, 1087.
(72) Binder, K. *Introduction: Theory and "Technical" Aspects of Monte Carlo Simulations*, 2nd ed.; Springer: Heidelberg, 1986; p 1.
(73) Merz, K. M., Jr. *J. Am. Chem. Soc*. **1991**, *113*, 3572.
(74) Bartik, K.; Dobson, C. M.; Redfield, C. *Eur. J. Biochem*. **1993**, *215*, 255.
(75) Dobson, C. M. Private communication.
(76) Ptitsyn, O. B. *The Molten Globule State*; W. H. Freeman: New York, 1992; p 243.
(77) Bruccoleri, R. E. *J. Comput. Chem*. **1993**, *14*, 1417.
(78) Schaefer, M.; Karplus, M. Work in progress.
(79) Schaefer, M.; Karplus, M. *J. Phys. Chem*. **1996**, *100*, 1578.