Conformation and Hydrogen Ion Titration of Proteins: A Continuum Electrostatic Model with Conformational Flexibility

Tony J. You and Donald Bashford
Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA

ABSTRACT A new method for including local conformational flexibility in calculations of the hydrogen ion titration of proteins using macroscopic electrostatic models is presented. Intrinsic pK_a values and electrostatic interactions between titrating sites are calculated from an ensemble of conformers in which the positions of titrating side chains are systematically varied. The method is applied to the Asp, Glu, and Tyr residues of hen lysozyme. The effects of different minimization and/or sampling protocols for both single-conformer and multi-conformer calculations are studied. For single-conformer calculations it is found that the results are sensitive to the choice of all-hydrogen versus polar-hydrogen-only atomic models and to the minimization protocol chosen. The best overall agreement of single-conformer calculations with experiment is obtained with an all-hydrogen model and either a two-step minimization process or minimization using a high dielectric constant. Multi-conformational calculations give significantly improved agreement with experiment, slightly smaller shifts between model compound pK_a values and calculated intrinsic pK_a values, and reduced sensitivity of the intrinsic pK_a calculations to the initial details of the structure compared to single-conformer calculations. The extent of these improvements depends on the type of minimization used during the generation of conformers, with more extensive minimization giving greater improvements. The ordering of the titrations of the active-site residues, Glu-35 and Asp-52, is particularly sensitive to the minimization and sampling protocols used. The balance of strong site-site interactions in the active site suggests a need for including site-site conformational correlations.

INTRODUCTION

For many years, the calculation of the pH titration behavior of ionizable groups in proteins has been one of the challenges of computational biochemistry and has provided a testing ground for theories of electrostatic interactions in proteins, the understanding of which is important for the study of protein structure and function (Perutz, 1978). Most of these calculations have been based on the assumption that the difference between the free energy of protonation of a particular group in a protein and the same group in a small model compound can be treated by classical electrostatics (Warshel and Russell, 1984; Sharp and Honig, 1990); and in many cases, a quasi-macroscopic model in which the protein interior has a low dielectric constant and the solvent region has a high dielectric constant has been used. Early versions of this model used spherical geometries for which the resulting Poisson or Poisson-Boltzmann equation could be solved analytically (Tanford and Kirkwood, 1957; Matthew and Gurd, 1986), but in the last decade, the introduction of numerical techniques such as the finite difference method has made it possible to use a dielectric boundary that follows the detailed atomic coordinates of the protein structure (Warwicker and Watson, 1982). This method of macroscopic electrostatics with atomic detail (MEAD) has been applied to the pH titration behavior of ionizable groups in lysozyme (Bashford and Karplus, 1990; Takahashi et al., 1992) and other proteins (McGrath et al., 1992; Bashford and Gerwert, 1992; Yang et al., 1993; Bashford et al., 1993) under the assumption that there is little or no structural flexibility in or around the titrating groups. However, there has been some work touching on conformational change in MEAD-type calculations. Combining conformational flexibility with macroscopic electrostatics is implicit in proposals to incorporate macroscopic electrostatics in molecular dynamics calculations (Sharp, 1991; Gilson et al., 1995), although titration is not included; and the coupling of titration to a two- or three-state model of protein denaturation has been introduced in calculations on myoglobin (Yang and Honig, 1994).

In previous reports of calculations using a rigid two-dielectric model with atomic detail, we have observed a tendency of the calculations to overestimate the magnitude of the difference between the calculated $pK_{1/2}$ values of ionizable groups in proteins and the pK_a of analogous model compounds, and it was suggested that allowing for conformational variation might reduce the magnitude of these shifts (Bashford and Karplus, 1990; Bashford and Gerwert, 1992; Bashford et al., 1993). Recently, Antosiewicz et al. (1994) have found that the rigid two-dielectric model tends to give smaller pK_a shifts and generally more accurate results for a number of proteins when the interior dielectric constant is set to 20 rather than the more commonly used values of 2.0 to 4.0. These authors have also pointed out that a more physically reasonable low-dielectric model that accounts for conformational flexibility may well achieve similar accuracy by mitigating the pK_a shifts through conformational changes.

Received for publication 9 March 1995 and in final form 14 August 1995.

Address reprint requests to Dr. Donald Bashford, Department of Molecular Biology MB-1, Scripps Research Institute, 10666 N. Torrey Pines Road, La Jolla, CA 92037. Tel.: 619-554-9612; Fax: 619-554-3789; E-mail: bashford@scripps.edu.

0006-3495/95/11/1721/00 $2.00

In this paper we present a formalism for calculating differences between the intrinsic pK_a of an ionizable group in a protein and the pK_a of an analogous model compound using a two-dielectric model with atomic detail and a small ensemble of atomic structures that differ from one another in the region of the ionizable group. We also present a study of different minimization protocols to prepare a protein structure either for single-conformer calculations or to serve as a base structure for multi-conformer calculations, and some protocols for selecting the conformations to be included in the ensemble and calculating their relative non-electrostatic energies (which enter in the ensemble weighting). The method is applied to hen lysozyme, and comparisons with conformationally rigid methodologies are presented.

METHODS

The methods used to calculate the electrostatic contributions to titration in proteins in a single-conformer model have been described previously (Bashford and Gerwert, 1992; Bashford et al., 1993; Yang et al., 1993; Antosiewicz et al., 1994). Here, the electrostatic theory and the single-conformer model are presented in a slightly different form that provides a convenient point of departure for the multiple conformation model that follows. Next, the method of selecting conformers to be included and calculating the non-electrostatic terms required by the theory is presented. Finally, the molecular models, software, and parameters are described.

Electrostatic model

Fig. 1 illustrates the physical assumptions of the electrostatic model of hydrogen ion titration used. The model requires the calculation of the electrostatic work of altering charges in model compounds and proteins. We suppose that the electrostatic potential in and around a molecule in solvent is governed by the linearized Poisson-Boltzmann equation,

$$\nabla[\epsilon(r)\nabla\phi(r)] - \kappa^2\epsilon(r)\phi(r) = -4\pi\rho(r), \qquad (1)$$

where ϕ is the electrostatic potential due to the charge distribution, ρ, which is assumed to have the form of point charges on atom centers, and ϵ is the dielectric constant, which takes on the values ϵ_s in the solvent-accessible region and ϵ_m in the molecular interior. The solvent-accessible region is defined as the volume that can be swept out by a solvent-sized probe sphere without overlapping any atomic radii of the molecule; what remains is defined as the molecular interior. (With this definition, the boundary between solvent and interior is the "molecular surface" as defined by Richards, Connolly, and others (Richards, 1977; Connolly, 1983)). A more formal definition of the solvent-accessible volume as well as the algorithm used here for lattice calculations on such volumes has been given by You and Bashford (1995).) Generally, one does not find the analytical solution of Eq. 1 for molecules with complex shapes, so various numerical methods have to be used to solve the equation approximately. In our study, a finite difference approach is employed to find the electrostatic potential, ϕ, in a model system.

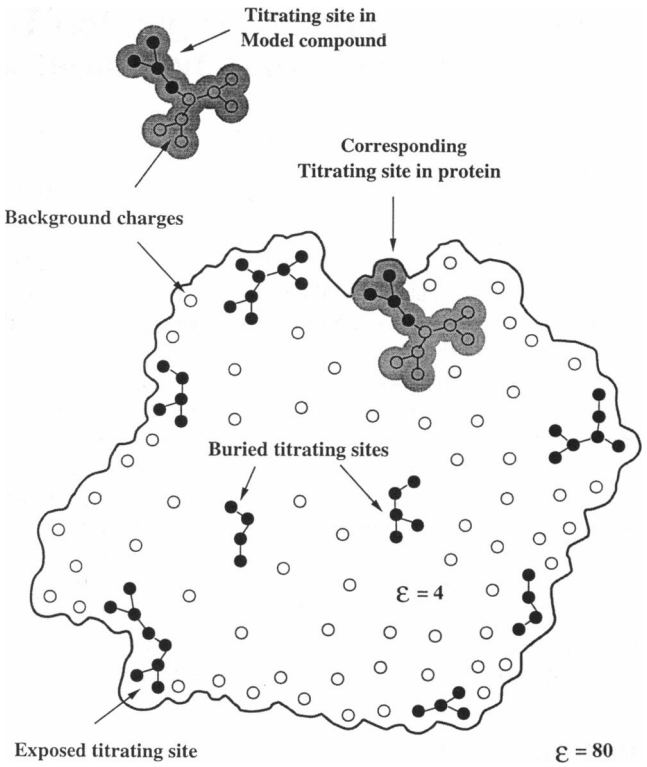

FIGURE 1 The macroscopic idealization used for titration calculations. The protein is treated as a low-dielectric object with embedded titrating and background charges, and the surrounding solvent is represented as a high-dielectric medium. Titration is described as adding a proton charge to the titrating site. The model compound is "carved out" from the protein by taking model compound coordinates from the protein coordinates. The pK_a difference between the site in the protein and the site in the model compound is due to differences in the electrostatic work required to alter the titrating charges in the model compound and protein.

It is useful to develop a Green function formalism for calculations of the work of altering point charges. Because of the linear form of Eq. 1, the potential is a linear function of the charges that can be expressed in the form

$$\phi(r) = \sum_{i=1}^{N_Q} Q_i\Phi(r, r'_i), \qquad (2)$$

where the Q_i are a set of N_Q charges located at points r'_i and Φ is the Green function of Eq. 1, which depends only on ϵ and κ, not on the charges. (The Green function is the solution of the special case of Eq. 1 where ρ consists of a unit point charge at r'.) In the molecular interior the Green function can be shown to have the form

$$\Phi(r, r') = \frac{1}{\epsilon_m|r - r'|} + \Phi^*(r, r'), \qquad (3)$$

where the first term is simply the Coulomb formula for charges in the interior dielectric and the second term arises because of the dielectric boundary. If no charges other than Q_i are present, the work of creating them by a charging process is obtained by integrating the incremental work of adding an increment of charge, $\delta W = \phi \delta q$, from zero charge to the full strength of the charges to obtain

$$W_{self} = \frac{1}{2} \sum_{i,j=1}^{N_Q} \frac{Q_i Q_j}{\epsilon_m |r_i - r_j|} + Q_i Q_j \Phi^*(r_i, r_j), \quad (4)$$

where the factor of 1/2 arises in the integration. The Φ^* term in Eq. 4 is the interaction of the charges with the reaction field that the charges themselves produce by their polarization of the environment. If another set of fixed charges, q_k, is present in the molecular interior, an additional term in the work of creating the charges, Q_i, arises:

$$W_{interaction} = \sum_{j=1}^{N_q} \sum_{i=1}^{N_Q} \frac{Q_i q_j}{\epsilon_m |r_i - r_j|} + Q_i q_j \Phi^*(r_i, r_j). \quad (5)$$

For the present applications, differences in work upon a change of the external dielectric or the position of the dielectric boundary (as in the case of a change from protein to model compound) must be calculated. The difference of W_{self} terms is then

$$\Delta W_{self} = \frac{1}{2} \sum_{i,j=1}^{N_Q} Q_i Q_j \Phi_2^*(r_i, r_j) - Q_i Q_j \Phi_1^*(r_i, r_j), \quad (6)$$

where Φ_1^* and Φ_2^* are the reaction field parts of the Green functions of the Poisson-Boltzmann equation before and after the change in the dielectric environment, respectively, and the coulomb terms have cancelled because the magnitude and disposition of the charges Q_i are the same before and after the change. In the special case where the molecule is a sphere with a point charge at the center, and the change in dielectric environment corresponds to transfer from vacuum to water ($\epsilon_s = 1 \rightarrow 80$), Eq. 6 leads to the well-known Born formula (Born, 1920) for ion solvation.

Many numerical methods used for solving the Poisson-Boltzmann equation, including the finite-difference method used here, do not directly give Green functions; rather, they give the potentials, $\phi(r; Q)$ due to specified sets of generating charges Q_i in a particular dielectric environment. Therefore, in practical calculations, Eqs. 5 and 6 become

$$W_{interaction} = \sum_{k=1}^{N_q} q_k \phi(r_k; Q) \quad (7)$$

and

$$\Delta W_{self} = \frac{1}{2} \sum_{i=1}^{N_Q} Q_i \phi_2(r_i; Q) - Q_i \phi_1(r_i; Q), \quad (8)$$

respectively, where the ϕ are now solutions provided by the numerical methods. In the latter equation, the products of the form $Q\phi(r_i;Q)$ present a potential problem because the finite-difference solutions contain a peak in the potential at each generating charge that is analogous to the Coulomb singularity and strongly dependent on the size and positioning of the finite-difference lattice. But in Eq. 8 these spurious "lattice singularities" cancel provided the interior dielectric constant stays the same and identical lattices are used for the calculation of both ϕ_1 and ϕ_2 (Gilson and Honig, 1988). In Eq. 7 the problem does not arise because the potential is not being sampled at the location of generating charges.

One-conformer formalism for titration in proteins

To isolate the intrinsic titration properties of a particular site in a protein from the complex effects of its interaction with other titrating sites in the protein, the intrinsic pK_a or pK_{intr} is defined as the pK_a that the site would have if all other titrating sites were held in their electrically neutral ionization states (Tanford and Kirkwood, 1957). The pK_{intr} value of an ionizable group of type A in the protein is related to the pK_a of a model compound containing the same type of group by

$$pK_{intr} = pK_{model} + \frac{1}{2.303RT}[\mu^\circ(A_p) - \mu^\circ(A_pH) - \mu^\circ(A_m) + \mu^\circ(A_mH)], \quad (9)$$

where the μ° are standard chemical potentials and p and m denote protein and model compound, respectively. It is assumed that the chemical potentials can be decomposed as

$$\mu^\circ(A) = \mu^\circ_{internal} + \Delta G_{env} \quad (10)$$

(and similarly for AH), where the internal term depends only on the chemical type of the ionizable group and is independent of whether the group is in the protein or the model compound or the conformation of either, these environmental dependencies being contained in the second term, the free energy of the molecule-solvent environment and the interaction of the titrating group with the environment. It is further assumed that for any given conformer, ΔG_{env} is adequately described by the electrostatic work of creating the charges of the titrating group in the context of the dielectric environment and the background of fixed charges.

If only one conformation of the protein and one conformation of the model compound are considered, and if the magnitude and relative disposition of the charges making up a given ionization state of the titrating group are the same in both the protein and the model compound, the $\mu^\circ_{internal}$ terms cancel immediately and the ΔG_{env} terms are conveniently written as two kinds of electrostatic terms,

$$pK_{intr} = pK_{model} - \frac{1}{2.303kT}[\Delta\Delta G_{Born} + \Delta\Delta G_{back}], \quad (11)$$

where the Born-like term, $\Delta\Delta G_{Born}$, arises from the interactions of the group's charges with the reaction field that the charges themselves produce in the environment, and the background term, $\Delta\Delta G_{back}$, arises from the interactions between the charges of the titrating group and any non-titrating charges in the protein or model compound (Bashford and Karplus, 1990). Using Eqs. 4–8 these terms can be written

$$\Delta\Delta G_{Born} = \frac{1}{2}\sum_{i}^{N_Q} Q_i^{(h)}[\phi_p(r_i; Q^{(h)}) - \phi_m(r_i; Q^{(h)})] - \frac{1}{2}\sum_{i}^{N_Q} Q_i^{(d)}[\phi_p(r_i; Q^{(d)}) - \phi_m(r_i; Q^{(d)})] \quad (12)$$

$$\Delta\Delta G_{Back} = \sum_{j}^{N_q} q_j[\phi_p(r_j; Q^{(h)}) - \phi_p(r_j; Q^{(d)})] - \sum_{k}^{N_q} q_k[\phi_m(r_k; Q^{(h)}) - \phi_m(r_k; Q^{(d)})], \quad (13)$$

where the Q_i are the partial charges of the atoms i, making up the titrating group, and $\phi(r; Q)$ is the electrostatic potential produced by them, as determined by the finite-difference solution of the linearized Poisson-Boltzmann equation; the q_j are the fixed background charges of the model compound or protein, including both peptide dipoles and the partial charges of other titrating groups regarded as "fixed" for the purpose of the definition of pK_{intr}. The subscripts p and m represent the protein and the model compound, respectively, and the superscripts h and d denote the protonated and deprotonated states of the titrating group, respectively.

To calculate the hydrogen ion titration behavior of the protein as a whole, the site-site interactions are also required. The interaction between sites μ and ν, $W_{\mu\nu}$, is defined as the additional energy required to protonate site μ if site ν is also protonated. By a simple extension of Eq. 7 it is

$$W_{\mu\nu} = \sum_i [Q_{\mu,i}^{(h)} - Q_{\mu,i}^{(d)}][\phi_p(r_i; Q_\nu^{(h)}) - \phi_p(r_i; Q_\nu^{(d)}], \quad (14)$$

where $Q_\nu^{(h)}$ is the set of atomic partial charges $Q_{\nu,i}^{(h)}$ for the protonated (h) state of site ν, and so on.

Once the pK_{intr} for each of the N titrating sites in the protein and the interaction energy for each pair of sites μ and ν have been obtained, the titration curve of each site can be derived from a Boltzmann average over all possible protonation states at each pH. For a given pH environment, the protonation fraction of site i is

$$\theta_i = \frac{\sum_x^N x_i \exp\left[\sum_\mu^N x_\mu 2.303(pK_{intr,\mu} - pH) - \frac{1}{RT}\sum_{\mu,\nu}^N W_{\mu\nu}(x_\mu + q_\mu^\circ)(x_\nu + q_\nu^\circ)\right]}{\sum_x^N \exp\left[\sum_\mu^N x_\mu 2.303(pK_{intr,\mu} - pH) - \frac{1}{RT}\sum_{\mu,\nu}^N W_{\mu\nu}(x_\mu + q_\mu^\circ)(x_\nu + q_\nu^\circ)\right]}, \quad (15)$$

where q_ν° is the formal charge of the deprotonated state of site ν and x is an N-element protonation state vector whose elements x_i take on values of 1 or 0 according to whether site i is protonated or deprotonated. The summation over x includes all 2^N possible protonation states of the protein. The $pK_{1/2}$ of a titrating site is defined as the pH value at which its protonation fraction is 0.5. In practical calculations for more than about 15 sites the 2^N term summation becomes burdensome, and a reduced-site approximation (Bashford and Karplus, 1991) is used.

Including local conformational variations

Consider the problem of calculating the pK_{intr} of a site if both the protein and the model compound are allowed to take on N_c conformations that differ in the region of the site in question. The chemical potentials in Eq. 9 would then be replaced by logarithms of canonical partition functions such as $\ln(\sum_n e^{-\mu_n^\circ/RT})$, where n labels the conformers. However, when such terms are combined in the multi-conformational analogue to Eq. 9 for the intrinsic pK, the conformation-independent part of the chemical potential (see Eq. 10) will factor out of the sum, and the logarithms of identical factors for the same ionization state in the protein and the model compound will cancel. The resulting expression for pK_{intr} is

$$pK_{intr} = pK_{model} - \frac{1}{2.303}\left[\ln\sum_{n=1}^{N_c} e^{-\Delta G_{env}(A_{p,n})/RT} - \ln\sum_{n=1}^{N_c} e^{-\Delta G_{env}(A_{p,n}H)/RT} - \ln\sum_{n=1}^{N_c} e^{-\Delta G_{env}(A_{m,n})/RT} + \ln\sum_{n=1}^{N_c} e^{-\Delta G_{env}(A_{m,n}H)/RT}\right], \quad (16)$$

where n labels the conformers. As in the one-conformer formulation, it is assumed that the ionization state-dependent interactions between the titrating site and its environment that go into ΔG_{env} can be treated by the Poisson-Boltzmann equation. But it is also necessary to include energy terms that depend on conformation but are independent of ionization state—terms that would cancel in a one-conformer case, but affect the weighting of different conformers in the multi-conformer case. For the deprotonated site in protein,

$$\Delta G_{env}(A_{p,n}) = E_{p,n} + \Delta G_{Born}(A_{p,n}) + \Delta G_{Back}(A_{p,n}), \quad (17)$$

where n denotes a particular conformation, E is an energy that is independent of ionization state and may include non-electrostatic terms, and the Born and background terms are

$$\Delta G_{\mathrm{Born}}(A_{\mathrm{p,n}}) = \frac{1}{2}\sum_{i}^{N_{\mathrm{Q}}} Q_{\mathrm{i}}^{(\mathrm{d})}[\phi_{\mathrm{p,n}}(r_{\mathrm{i}}; Q^{(\mathrm{d})}) - \phi_{\mathrm{unif,n}}(r_{\mathrm{i}}; Q^{(\mathrm{d})})] \tag{18}$$

$$\Delta G_{\mathrm{back}}(A_{\mathrm{p,n}}) = \sum_{j}^{N_{\mathrm{q}}} q_{\mathrm{j}}\phi_{\mathrm{p,n}}(r_{\mathrm{i}}; Q^{(\mathrm{d})}). \tag{19}$$

In Eqs. 18 and 19 the notation for charges is as before, $\phi_{\mathrm{p,n}}$ is the solution of the Poisson-Boltzmann equation for protein conformer n, and ϕ_{unif} is the solution for the case where the dielectric constant everywhere is set to the same value as the dielectric constant of the molecular interior, ϵ_{m}. This extra term is needed to cancel the "lattice singularity" in $\phi_{\mathrm{p,n}}$ (see Eqs. 4 and 8). Such special terms are not needed in the one-conformer calculation because the model compound calculation provides the necessary cancellation. Formulae similar to Eqs. 18 and 19 apply to the protonated state and to the model compound.

The ionization state-independent energy term, E, is mostly non-electrostatic, and for this part of E we use a standard empirical potential energy function composed of two-, three-, and four-atom terms (see below). Energy terms involving atoms whose position does not change from one conformer to the next can be ignored. However, the Poisson-Boltzmann electrostatic model being used here is many-body rather than two-body in form, so conformational dependency can arise even among charges that do not move. For example, consider a side-chain dihedral rotation that causes a backbone peptide group to become more exposed to solvent. This will alter the self energy of the peptide group even though its conformation does not change. Such effects give rise to an electrostatic contribution to E,

$$\Delta G_{\mathrm{selfback}}(A_{\mathrm{p,i}}) = \frac{1}{2}\sum_{j}^{N_{\mathrm{q}}} q_{\mathrm{j}}[\phi_{\mathrm{p,n}}(r_{\mathrm{j}}; q) - \phi_{\mathrm{unif}}(r_{\mathrm{j}}; q)], \tag{20}$$

which is analogous to the ΔG_{Born} term, except that it involves the background charges q rather than the ionizable group charges Q. This term influences pK_{intr} indirectly by altering the relative population of conformers in the canonical ensemble and adds significantly to the computational expense of the calculations, so we have investigated the question of whether it can be omitted in practical calculations (see below).

Because the individual terms in Eq. 16 resemble calculations of absolute free energies, it might be thought that the sampling of conformations would need to be nearly exhaustive for the method to be useful. We argue that the sampling need only be representative. Consider the special case in which all of the $\Delta G_{\mathrm{env}}(A_{\mathrm{p,n}})$ are identical to one another, all of the $\Delta G_{\mathrm{env}}(A_{\mathrm{p,n}}H)$ are identical to one another, and so on. Then Eq. 16 reduces to

$$\begin{aligned} pK_{\mathrm{intr}} = pK_{\mathrm{model}} \\ -\frac{1}{2.303}[&\ln(N_{\mathrm{c,p}}e^{-\Delta G_{\mathrm{env}}(A_{\mathrm{p}})/RT}) - \ln(N_{\mathrm{c,p}}e^{-\Delta G_{\mathrm{env}}(A_{\mathrm{p}}H)/RT}) \\ &- \ln(N_{\mathrm{c,m}}e^{-\Delta G_{\mathrm{env}}(A_{\mathrm{m}})/RT}) + \ln(N_{\mathrm{c,m}}e^{-\Delta G_{\mathrm{env}}(A_{\mathrm{m}}H)/RT})], \end{aligned} \tag{21}$$

where $N_{\mathrm{c,p}}$ and $N_{\mathrm{c,m}}$ are the number of conformers sampled for the protein and model compound, respectively. Because the number of conformers sampled is the same for both the protonated and the deprotonated states, the ln N terms cancel and an expression identical to the single-conformer formula, Eq. 11, is obtained. By a similar argument, it can be shown that if a truly exhaustive ensemble can be grouped into subsets of equal or nearly equal ΔG_{env} values, and only a small number of members of each subset is included in Eq. 16, the correct result will still be obtained provided that the number of members chosen from each subset is proportional to the true population of the subset (that is, the sampling is representative) and provided that the sampling ratio is the same for both the protonated and deprotonated states. Note that the sampling ratio or the number of sample conformers need not be the same for protein and model compound, but only for the protonated and deprotonated states of either. The requirement that the conformer free energies fall into a limited number of subsets implies that only a limited range of conformational variation can be accounted for by this method.

The presence of more than one titrating group in the protein, and the fact that these groups interact, raise the question of conformational correlation. In deriving the expressions for pK_{intr}, we have isolated a specific titrating site from others which, in a multi-site molecule, amounts to an implicit assumption that the conformational changes that influence the pK_{intr} of one site do not influence the pK_{intr} of another. For this assumption to be valid, the conformational changes should be limited to the local region of a titrating site—our current study is focused on side-chain torsion angle changes. Even in the context of an assumption of localized pK_{intr} effects, one could, in principle, make the site-site interaction terms, $W_{\mu\nu}$, depend on the conformation of sites μ and ν; but this would lead to a combinatorial explosion in the multi-site titration expression, Eq. 15. However, for interactions between titrating groups in the protein that are not too strongly coupled, it should be possible to use the average interaction energy of one site with the average field produced by the other to calculate the elements of the site-site interaction matrix,

$$W_{\mu\nu} = \sum_{i=1}^{N_{\mathrm{Q}_\mu}} [Q_{\mu,\mathrm{i}}^{(\mathrm{h})} - Q_{\mu,\mathrm{i}}^{(\mathrm{d})}][\langle\phi_{\mathrm{p}}(r_{\mathrm{i}}; Q_{\nu}^{(\mathrm{h})})\rangle - \langle\phi_{\mathrm{p}}(r_{\mathrm{i}}; Q_{\nu}^{(\mathrm{d})})\rangle], \tag{22}$$

where

$$\langle\phi\rangle = \frac{\sum_n \phi_{\mathrm{n}} e^{-\Delta G_{\mathrm{env,n}}/RT}}{\sum_k e^{-\Delta G_{\mathrm{env,n}}/RT}}, \tag{23}$$

where n labels conformers of site ν and site μ is held fixed in its "base" conformation. A similar expression for site $W_{\nu\mu}$ is obtained when site μ is varied and ν is fixed, and the two are averaged to form a symmetrized W matrix. For titratable groups that are close together, especially those that form salt bridges and titrate at similar pH values, this approach may not be adequate. Nevertheless, we have used it for all site-site interactions in the present study.

Conformational sampling

For the reasons discussed above, the conformational changes considered must be mostly local to particular sites. We have focused on the effects of side-chain conformational changes, specifically, the systematic sampling of the χ_1 and χ_2 torsion angles of aspartic acid and tyrosine residue side chains, and the χ_2 and χ_3 torsions of glutamic acid residue side chains.

Three protocols have been developed for generating 36 conformers differing primarily in the torsion angles of a selected side chain. They differ mainly in the minimization scheme used to release steric and other energetic stresses that may arise from altering side-chain torsion angles. In the first protocol, named STIFF, no energy minimization is conducted. The other two protocols, FMINL and FMING, use minimization of all protein atoms within an 8-Å cutoff distance from the selected side chain, and minimization of all atoms of the protein, respectively. For all three protocols, the sampling procedure starts with 20° steps through the two torsion angles, generating 324 conformers. For the FMINL and FMING protocols, conjugate gradient minimization with a dielectric constant of 4.0 and an energy tolerance of 0.1 kcal/mol is carried out for each conformer. The non-electrostatic part of the empirical potential energy function is then evaluated for each conformer and only those with energies below a cutoff value are retained. The cutoff value is 4–5 kcal/mol above the lowest energy found, depending on the flexibility of the site. The remaining conformers typically fall into a handful of allowed regions of the side-chain torsion map. If the number of conformers remaining exceeds 36, the highest energy members of each allowed region are removed until 36 conformers remain. If the number of conformers is less than 36, a "focusing" scheme is used in which the allowed regions of the torsion map are searched with 5–10° steps, depending on the number of conformers to be generated. These conformers are again subjected to minimization and screening according to non-electrostatic energy terms. These sampling procedures are carried out for the protein only. The model-compound conformers are taken directly from the protein conformers by deleting all protein atoms except those belonging to the model compound. This is consistent with the method of model compound generation used in single-conformer studies.

Molecular models, parameters, and software

The heavy atom coordinates for hen egg lysozyme were initially taken from the data of Hodsdon et al. (1990), which are deposited in the Brookhaven databank (Bernstein et al., 1977) as structure 1LZT. Hydrogen atoms were added to the coordinate set using the HBUILD (Brünger and Karplus, 1988) facility of the CHARMM22 molecular mechanics program (Brooks et al., 1983), and either the CHARMM22 all-hydrogen protein parameter set or the CHARMM22 polar-hydrogen parameter set in which aliphatic carbons and the hydrogen atoms bonded directly to them are treated as single extended atoms. All minimizations and empirical potential energy function evaluations, including those for the $E_{p,n}$ term in Eq. 17, were done using CHARMM22.

Charges and atomic radii for the Poisson-Boltzmann electrostatic calculations were taken from the CHARMM22 parameter sets. The radii used were the $R_{min}/2$ values of the van der Waals term. The protein interior dielectric constant, ϵ_m, is taken as 4.0 following Tanford and Roxby (1972), and the dielectric constant of water is taken as 80. The solvent ionic strength is taken as 0.1 M, and a 2.0-Å ion exclusion radius is used. As in previous calculations (Bashford and Karplus, 1990; Bashford and Gerwert, 1992), the model compounds used were the *N*-formyl-*N*-methylamide derivatives of the amino acids being considered. The pK_{model} values used are the same as those used by Bashford and Karplus (1990). For each conformer of each titrating site, model compound coordinates corresponding to that conformer and site were derived from the coordinates of the protein by deleting all atoms except those of the residue in question, the C and O atoms of the preceeding residue, and the N, H, and C_α atoms of the following residue. The Poisson-Boltzmann calculations were done using the programs multimead and multiflex from the MEAD (Macroscopic Electrostatic with Atomic Detail) program suite, which has been developed in our laboratory. Lattice calculations are done first on a coarse lattice with 1.0 Å spacing and an 80-Å extent, which allows lysozyme to be enclosed with at least 10 Å to spare on all sides, and then on a 0.25 Å grid with a 20-Å extent centered on the titrating site. Previous numerical tests (Bashford and Karplus, 1990) suggest that with this protocol, numerical errors are limited to 0.2 or 0.3 pK units and are usually significantly less. The MEAD programs and data files needed to reproduce some of the results presented here are available by anonymous FTP from ftp.scripps.edu:/pub/electrostatics.

RESULTS AND DISCUSSION

Single-conformer calculations

Structural models of lysozyme have been generated using both all-hydrogen or polar-hydrogen modeling schemes, and minimizations have been carried out using several different dielectric constants in the empirical potential function or by deleting the electrostatic term. Single-conformer ti-

tration calculations were carried out on the resulting structures. For the polar-hydrogen models, titrating groups were treated as a single charge as in previous calculations on lysozyme (Bashford and Karplus, 1990), but for the all-hydrogen models, the CHARMM partial charges were used for the titrating groups. As in the previous calculations, the arginine residues are regarded as non-titrating "background" charges. Table 1 shows the calculated pK_{intr} and p$K_{1/2}$ values for several of these structures; and Table 2 shows statistics comparing experimental and calculated p$K_{1/2}$ values for all the structures generated. A strong dependence on the minimization procedure has been found—for some titrating sites, the maximum deviation of the p$K_{1/2}$ values calculated using different structures can be as large as 3 pK units (Lys-116, for example). Examination of the molecular structures using computer graphics showed that the sites with larger deviations tended either to be partially buried away from solvent or to have strong interactions with other charged sites, whereas those with smaller variations tended to be well exposed to solvent. This is consistent with previous observations that burial away from solvent and site-site interactions can have strong effects on calculations of this type (Bashford and Karplus, 1990; Bashford et al., 1993). Generally, the calculations based on all-hydrogen representations result in a better agreement with the experiment than those including only polar hydrogens. Structures minimized with a dielectric constant of 4.0 or higher usually lead to a better fit between predicted and experimental p$K_{1/2}$ values for all-hydrogen models. Root mean square deviations of the calculated p$K_{1/2}$ values from experiment are very similar for the all-hydrogen models minimized without electrostatics (ALLME0), minimized with a high dielectric constant (ALLME10), or minimized by a two-step procedure: first, 200 steps of deepest decent minimization without an electrostatic term in the potential energy function, then conjugate-gradient minimization with a dielectric constant of 4.0 in the electrostatic term until the convergence criterion, energy step $<$ 0.1 kcal/mol, is satisfied (ALLME0E4). However, the ALLME0E4 procedure gives the largest number of calculated p$K_{1/2}$ values within 1.0 pK unit of experiment (Table 2), so we regard it as the best overall. Minimization protocols that use electrostatic terms with dielectric constants of 4 or less (e.g., POLME2, POLME4, ALLME2, and ALLME4) produce the largest deviations of calculated p$K_{1/2}$ values from experiment.

For two titrating groups in the active site, Glu-35 and Asp-52, the experimental data indicate widely separated pK_a values: one increased (Glu-35) and one decreased (Asp-52), relative to model compound values. The calculations presented in Table 1 also have widely separated p$K_{1/2}$ values for these two groups, but vary as to their ordering, with most calculations having orderings opposite to exper-

TABLE 1 pK_a values calculated from different energy minimized structures

		POLM0	ALLME0	ALLME10	ALLME0E4
Site	Exp. pK_a	p$K_{1/2}$(pK_{intr})	p$K_{1/2}$(pK_{intr})	p$K_{1/2}$(pK_{intr})	p$K_{1/2}$(pK_{intr})
NTLYS-1	7.8–8.0*	6.64 (6.76)	7.56 (7.63)	7.36 (7.46)	7.86 (8.04)
HIS-15	5.29–5.43‡	4.27 (4.05)	3.46 (3.72)	5.15 (5.37)	6.16 (6.03)
GLU-7	2.60–3.10‡	1.99 (2.59)	2.24 (2.76)	2.41 (2.89)	2.62 (3.07)
GLU-35	6.1–6.3‡	5.96 (5.80)	6.22 (5.25)	5.06 (4.75)	4.21 (4.27)
ASP-18	2.58–2.74‡	2.75 (3.34)	3.24 (3.88)	2.91 (3.64)	2.68 (3.43)
ASP-48	1.2–2.0‡	0.61 (0.61)	3.31 (3.00)	4.31 (3.76)	3.95 (3.57)
ASP-52	3.60–3.76‡	6.98 (5.23)	5.13 (3.67)	6.03 (3.92)	6.79 (4.44)
ASP-66	0.4–1.4‡	1.12 (1.04)	1.83 (1.96)	2.09 (2.24)	1.95 (2.12)
ASP-87	1.92–2.22‡	0.84 (2.45)	1.92 (3.54)	1.97 (3.50)	2.77 (4.38)
ASP-101	4.02–4.16‡	7.84 (8.26)	4.47 (5.42)	4.72 (5.42)	4.89 (5.61)
ASP-119	3.11–3.29‡	3.19 (3.22)	3.36 (3.41)	3.26 (3.30)	3.39 (3.41)
TYR-20	10.3*	13.98 (13.05)	12.06 (10.83)	12.54 (11.33)	12.87 (11.75)
TYR-23	9.8*	11.65 (11.02)	11.12 (10.38)	10.66 (10.09)	10.34 (9.88)
TYR-53	12.1*	20.66 (11.43)	18.81 (10.33)	19.00 (10.42)	18.60 (10.08)
LYS-1	10.7–10.9*	9.65 (8.92)	10.45 (9.75)	10.59 (9.91)	11.13 (10.45)
LYS-13	10.4–10.6*	11.75 (9.55)	11.53 (9.39)	11.41 (9.38)	11.20 (9.36)
LYS-33	10.5–10.7*	9.78 (9.32)	9.70 (9.24)	10.28 (9.82)	10.90 (10.45)
LYS-96	10.7–10.9*	10.54 (10.15)	8.66 (8.29)	9.05 (8.76)	9.73 (9.41)
LYS-97	10.2–10.4*	10.70 (9.99)	10.59 (9.45)	10.80 (9.92)	11.01 (10.14)
LYS-116	10.3–10.5*	9.93 (9.69)	7.53 (7.23)	8.57 (8.29)	9.40 (9.12)
CTLEU-129	2.63–2.87‡	2.15 (3.61)	2.34 (3.71)	2.41 (3.67)	2.50 (3.62)
ASP & GLU rms§		1.69 (2.32)	0.69 (1.12)	1.17 (0.97)	1.40 (1.00)
TYR rms§		5.49 (2.40)	4.08 (0.94)	4.22 (1.14)	4.05 (1.28)
All site rms§		2.41 (1.91)	1.84 (1.36)	1.86 (1.03)	1.83 (0.93)

Minimization protocol names as given in Table 2.

*Experimental data from Kuramitsu and Hamaguchi (1980).

‡Experimental data from Bartik et al. (1994).

§Number outside parentheses is $\langle(pK_{1/2} - pK_{exp})^2\rangle$, where p$K_{exp}$ is the nearest end of the experimental range; number inside parentheses is $\langle(pK_{intr} - pK_{model})^2\rangle$.

TABLE 2 Statistical summary of single conformer calculations on structures generated by different hydrogen atom models and minimization procedures

	Range of $pK_{1/2}$–pK^*_{exp}				RMSD‡		
	<0.5	0.5–1.0	1.0–2.0	>2.0	ASP & GLU	TYR	All sites
POLM0	8	3	6	4	1.69 (2.32)	5.49 (2.40)	2.41 (1.91)
POLME0	5	7	4	5	0.74 (2.27)	4.11 (2.82)	2.25 (2.39)
POLME2	5	4	1	11	3.90 (4.86)	2.93 (3.32)	3.08 (3.89)
POLME4	3	5	4	9	2.35 (3.50)	3.00 (2.45)	2.16 (2.81)
POLME10	4	4	8	5	1.57 (3.11)	3.65 (2.55)	2.05 (2.65)
POLME0E4	4	2	7	8	2.33 (3.77)	3.24 (2.54)	2.26 (2.95)
ALLME0	11	2	5	3	0.69 (1.12)	4.08 (0.94)	1.84 (1.36)
ALLME2	7	3	5	6	1.77 (0.83)	3.73 (1.56)	1.95 (0.97)
ALLME4	7	8	2	4	1.49 (0.99)	4.27 (1.39)	1.93 (0.94)
ALLME10	10	4	3	4	1.17 (0.97)	4.22 (1.14)	1.86 (1.03)
ALLME0E4	7	9	2	3	1.40 (1.00)	4.05 (1.28)	1.83 (0.93)

Notation for hydrogen atom models and minimization protocols: POL, polar hydrogens only model; ALL, all-hydrogen model; M0, no minimization; ME0, minimized without electrostatics; MEϵ, minimized with dielectric constant ϵ; ME0E4, minimized in two steps, first without electrostatics then with ϵ = 4 (see text).

*Deviation from experimental pK_a values measured from the nearest end of experimental range given in Table 1.

‡Root mean square of deviation from experimental pK_a (rms deviation between calculated pK_{intr} and model compound).

iment. A closer examination of the electrostatic terms involved shows that this change of ordering is related to a complex balance of energetic terms involving an electrostatically driven competition between the two sites. For example, in the ALLME0 calculation (all-hydrogen minimization without electrostatics) the glutamic acid model compound pK_a of 4.40 is modified by a Born term of 4.79 pK units and a background term of −3.94 pK units resulting in a pK_{intr} value of 5.25 for Glu-35; whereas for Asp-52 in the same calculation, the pK_{model} value of 4.0 is modified by a Born term of 3.09 and a background term of −3.42 so that pK_{intr} = 3.67. It would appear at first from Table 1 that the site-site interactions simply raise both sites' pK_a values by 1 to 1.5 units to produce the $pK_{1/2}$ values of 6.22 and 5.13 for Glu-35 and Asp-52, respectively. But the same pK_{intr} and $W_{\mu\nu}$ values used in Eq. 15 can also be used to calculate the free energy of moving a proton from the protonated Glu-35 to the deprotonated Asp-52 while holding all other carboxylates deprotonated and all lysine, arginine, and tyrosine residues protonated—conditions likely to prevail in the neutral range. For ALLME0 it is found that this transfer energy is +0.36 pK units, which is considerably less than the 1.09 unit difference between the $pK_{1/2}$ values of these sites or the 1.48 pK_{intr} difference. The apparent discrepancy with the $pK_{1/2}$ difference is caused by a site-site interaction term between the two sites that is equivalent to 1.04 pK units. This term makes it more difficult to deprotonate one of the sites if the other is already deprotonated (negatively charged) so that the $pK_{1/2}$ values are farther apart than the transfer energy. The difference between the transfer energy and the pK_{intr} is caused primarily by interactions with two nearby sites: Asp-66 and Asp-48 have interactions with Asp-52 that tend to raise its pK_a by 1.41 units if 48 and 66 are both deprotonated (which they are in the relevant pH range), whereas other titrating sites have relatively little influence on Glu-35. This difference in third-site influence is almost enough to overcome the difference of 1.48 in the pK_{intr} values of the two sites. The small transfer energy relative to the larger terms that contribute to it implies that small changes in the sites' relative pK_{intr} values or changes in the interaction of third sites with the pair could reverse the ordering of the titration of the sites.

Such a reversal occurs in the ALLME10 calculation (all-hydrogen minimization with ϵ = 10). Although the pK_{intr} of Glu-35 remains significantly higher than that of Asp-52, the difference in pK_{intr} values for the two sites narrows from 1.48 in the ALLME0 calculation to 0.88 in the ALLME10 calculation, primarily because of a slightly reduced Born term for Glu-35, which is somewhat more exposed to solvent in the ALLME10 structure. The influence of Asp-66 and Asp-48 remain about the same as in ALLME0. The energy of transferring a proton from Glu-35 to Asp-52 is −0.34 pK units in ALLME10, a change of sign compared to ALLME0, but is about the same magnitude. This reverses the order of titration of these groups. As before, interaction between Glu-35 and Asp-52 leads to a $pK_{1/2}$ difference that is larger than the proton transfer energy. This example shows how large, observed pK_a differences may be the result of a subtle balance of energetic terms that is easily reversed by conformational changes or protonation changes of a third site. Competitions and sensitivities of this kind have also been observed in calculations on bacteriorhodopsin (Bashford and Gerwert, 1992) and myoglobin (Bashford et al., 1993).

Calculations with conformational flexibility

Several of the model structures generated by the different hydrogen building and minimization schemes described above were used as base structures for the side-chain sampling protocols described in Methods. Fig. 2 compares the

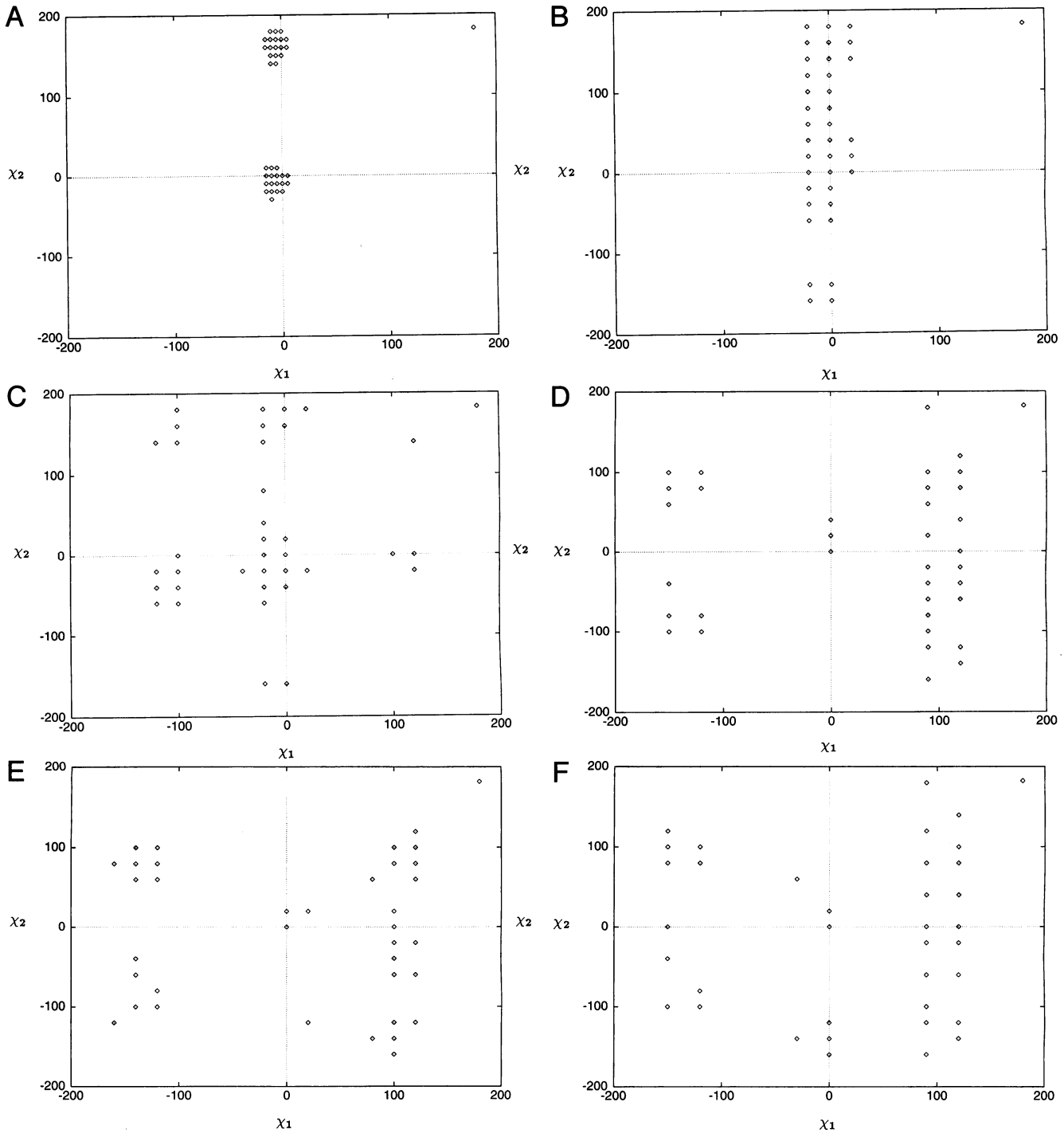

FIGURE 2 Sampling of Asp-52 *(a, b, c)* and Glu-7 *(d, e, f)*-) side-chain torsion angle space by different protocols: *(a, d)* STIFF; *(b, e)* FMINL; *(c, f)* FMING. The base structure torsions are taken as the origin.

results of the STIFF, FMINL, and FMING sampling protocols for two representative sites and shows that for a less exposed site such as Asp-52, the STIFF protocol results in a sample set concentrated near the base structure conformation, whereas the FMINL and FMING protocols sample a much wider region of torsion angle space, but for a well-exposed site such as Glu-7, the same regions of torsion angle space are sampled by all three protocols.

Trial calculations on five sites using the FMINL protocol were done to determine whether the inclusion of the "self-back" term of Eq. 20 led to significant differences in the results. The results (not shown) demonstrated that the dom-

inant conformers in the partition function remained the same and the calculated pK_{intr} values remained the same to within 0.1 pK units. All other calculations reported here were therefore carried out without the "selfback" terms.

Table 3 gives the $pK_{1/2}$ and pK_{intr} values for all Asp, Glu, and Tyr residue side chains calculated by the multiple-conformation method using ALLME0E4 as the base structure and molecular conformers sampled by three different protocols. Also listed are the experimental data and values calculated from the base structure. The rms deviations of the calculated $pK_{1/2}$ values from experiment are reduced by including conformational flexibility. The results become progressively better as one goes from the single-conformer calculation to the STIFF, FMINL, and FMING multi-conformer calculations, with the largest incremental reduction in both the overall deviation and the carboxylic acid groups' deviations being seen upon going from local to global minimization in the conformer generating protocol (FMINL to FMING). It is particularly interesting that the order of titration of the active-site residues Glu-35 and Asp-52 is predicted correctly in the FMING calculations and the $pK_{1/2}$ values are within 0.5 units of the experimental range. The rms deviations between the pK_{model} and pK_{intr} values also tend to decline as more flexibility and more global minimization is used, but the trend is not as strong or consistent as the corresponding trend in $pK_{1/2}$ deviations.

It should be remembered that the ALLME0E4 structure, which is the basis for all of the results presented in Table 3, was among the best for agreement between single-conformer results and experiment, so it might have been expected that improvements due to conformational flexibility would be more modest for this structure than for others. Therefore, we conducted FMINL protocol calculations for a limited number of sites starting from three alternative basis structures for comparison to the equivalent single-conformer calculations using the same basis structures. The results are presented in Table 4. The five sites chosen for study were among those displaying the largest calculated pK shifts. Including flexibility reduces rms deviation of the calculated $pK_{1/2}$ value versus experiment for all basis structures; and the largest reduction occurs for the basis structure for which the single-conformer results were the worst, ALLME4. Flexibility also reduces the magnitude of the shift between the pK_{intr} and pK_{model} values, although for some basis structures the change in rms of this shift is quite small.

It might be hoped that in addition to improving overall accuracy, including flexibility would reduce the sensitivity of the calculations to the choice of the basis structure. The columns on the right-hand side of Table 4 contain the average and standard deviation across basis structures for the calculated $pK_{1/2}$ and pK_{intr} values. The standard deviation of pK_{intr} values is significantly reduced by including flexibility, but there is no clear trend in the standard deviation of calculated $pK_{1/2}$ values. The reduction of the sensitivity of pK_{intr} calculations to basis structure choice might have been expected because conformational variation enters mainly in the calculation of pK_{intr}. The lack of a clear trend in the sensitivity of $pK_{1/2}$ values may reflect the lack of site-site correlations in the treatment of conformational change and the fact that the "flexible" sites are interacting with "inflexible" ones whose positions are fixed by the choice of basis structure.

Individual terms in the partition functions of Eq. 16 for several sites in the FMINL and FMING ensembles were examined to determine which conformers are significantly populated (partition terms not shown). It was found that only a small number of conformational states contribute to the partition function for the protein terms—that is, the

TABLE 3 $pK_{1/2}$ values calculated for all Asp Glu and Tyr side chains using conformations from different sampling protocols

Site	Exp. pK_a	ALLME0E4* $pK_{1/2}(pK_{intr})$	STIFF‡ $pK_{1/2}(pK_{intr})$	FMINL‡ $pK_{1/2}(pK_{intr})$	FMING‡ $pK_{1/2}(pK_{intr})$
GLU-7	2.60–3.10	2.62 (3.07)	2.62 (3.10)	2.42 (2.89)	3.00 (3.44)
GLU-35	6.1–6.3	4.21 (4.27)	3.56 (3.70)	3.73 (3.84)	5.67 (4.69)
ASP-18	2.58–2.74	2.68 (3.43)	2.74 (3.50)	2.40 (3.18)	2.74 (3.56)
ASP-48	1.2–2.0	3.95 (3.57)	3.46 (3.10)	3.86 (3.41)	3.67 (3.31)
ASP-52	3.60–3.76	6.79 (4.44)	6.26 (3.96)	6.20 (3.89)	4.14 (3.28)
ASP-66	0.4–1.4	1.95 (2.12)	2.00 (2.14)	1.67 (1.85)	2.65 (2.67)
ASP-87	1.92–2.22	2.77 (4.38)	2.79 (3.94)	2.75 (3.91)	2.84 (3.95)
ASP-101	4.02–4.16	4.89 (5.61)	4.81 (5.48)	4.96 (5.52)	3.78 (4.08)
ASP-119	3.11–3.29	3.39 (3.41)	3.32 (3.34)	3.33 (3.34)	3.37 (3.41)
TYR-20	10.3	12.87 (11.75)	12.11 (11.22)	12.39 (11.43)	12.06 (11.26)
TYR-23	9.8	10.34 (9.88)	10.41 (9.88)	10.57 (10.03)	10.78 (10.22)
TYR-53	12.1	18.60 (10.08)	18.18 (9.93)	17.50 (10.06)	17.07 (10.19)
ASP & GLU rms§		1.40 (1.00)	1.33 (1.02)	1.34 (1.11)	0.76 (0.69)
TYR rms§		4.05 (1.28)	3.68 (0.97)	3.37 (1.12)	3.10 (1.08)
All site rms§		2.36 (1.08)	2.17 (1.01)	2.05 (1.11)	1.68 (0.81)

*ALLME0E4 is the base structure used for conformational sampling. See text and Table 2.
‡STIFF, FMINL, and FMING: conformational sampling protocols (see text).
§Root mean square deviations as in Table 1.

TABLE 4 Comparison of single-conformer versus flexible calculations for selected sites and basis structures

Site	Exp. pK_a	ALLME0* $pK_{1/2}(pK_{intr})$	ALLME4 $pK_{1/2}(pK_{intr})$	ALLME10 $pK_{1/2}(pK_{intr})$	ALLME0E4 $pK_{1/2}(pK_{intr})$	Ave. ± SD§ $pK_{1/2}$	Ave. ± SD§ (pK_{intr})
GLU-35	6.1–6.3	6.22 (5.25)	4.20 (4.29)	5.06 (4.75)	4.21 (4.27)	4.92 ± 0.83	(4.64 ± 0.40)
ASP-52	3.60–3.76	5.13 (3.67)	6.98 (4.60)	6.03 (3.92)	6.79 (4.44)	6.23 ± 0.73	(4.16 ± 0.38)
ASP-66	0.4–1.4	1.83 (1.96)	2.12 (2.30)	2.09 (2.24)	1.95 (2.12)	2.00 ± 0.12	(2.16 ± 0.13)
ASP-87	1.92–2.22	1.92 (3.54)	2.74 (4.37)	1.97 (3.50)	2.77 (4.38)	2.35 ± 0.40	(3.95 ± 0.43)
TYR-20	10.3	12.06 (10.83)	12.99 (11.85)	12.54 (11.33)	12.87 (11.75)	12.61 ± 0.36	(11.44 ± 0.40)
All site rms		1.02 (1.16)	2.10 (1.30)	1.53 (1.14)	2.00 (1.30)		
		ALLME0CV‡ $pK_{1/2}(pK_{intr})$	ALLME4CV $pK_{1/2}(pK_{intr})$	ALLME10CV $pK_{1/2}(pK_{intr})$	ALLME0E4CV $pK_{1/2}(pK_{intr})$	$pK_{1/2}$	(pK_{intr})
GLU-35	6.1–6.3	5.20 (4.52)	4.09 (4.16)	4.46 (4.31)	3.73 (3.84)	4.37 ± 0.55	(4.21 ± 0.25)
ASP-52	3.60–3.76	3.89 (3.10)	6.15 (3.91)	5.50 (3.51)	6.20 (3.89)	5.44 ± 0.93	(3.60 ± 0.33)
ASP-66	0.4–1.4	1.97 (2.09)	2.15 (2.32)	2.18 (2.29)	1.67 (1.85)	1.99 ± 0.20	(2.14 ± 0.19)
ASP-87	1.92–2.22	2.58 (3.75)	2.72 (3.91)	1.94 (3.35)	2.75 (3.91)	2.50 ± 0.33	(3.73 ± 0.23)
TYR-20	10.3	11.31 (10.70)	12.05 (11.28)	11.84 (11.05)	12.39 (11.43)	11.90 ± 0.39	(11.11 ± 0.27)
All site rms		0.68 (1.07)	1.65 (1.07)	1.32 (1.07)	1.81 (1.29)		

*Notations such as ALLME0 indicate single conformer calculations on the structures listed in Table 2.
‡When followed by CV the notation indicates conformational variation applied to the basis structure denoted by the part before the CV.
§Average and standard deviations for calculated $pK_{1/2}$ (calculated pK_{intr}).

probability distribution of protein conformational states is narrow. For the model compound the distribution is broader.

CONCLUDING REMARKS

Evaluation of the methods

Previous studies have demonstrated the sensitivity of single-conformer calculations to the particular crystal structure used (Bashford and Karplus, 1990; Bashford et al., 1993) or to conformational samples taken from a rough annealing process (Bashford and Gerwert, 1992), either of which might have been expected to generate significant conformational differences. The present study shows that results can also be sensitive to minimization protocols. In principle, the most consistent minimization scheme for the present calculations would be one that used an electrostatic energy function equivalent to the electrostatic theory used to calculate the electrostatic contributions to pK_a, that is, minimization with Poisson-Boltzmann electrostatics. Because this would greatly increase the cost of the calculations, we have explored various treatments of the electrostatic term in conventional minimization programs. Raising the dielectric constant to mimic some features of solvation in a vacuum calculation is often used in such programs (Brooks et al., 1983; Weiner et al., 1984), and comparisons of minimized structures to well-determined protein structures suggests that higher dielectric constants may yield minimized structures closer to the actual protein structure (Whitlow and Teeter, 1986).

For the particular case of an all-hydrogen model of lysozyme, minimization without electrostatics, with weak ($\epsilon = 10$) electrostatics or with a two-stage minimization, first using no electrostatic terms and then using electrostatics with a dielectric constant of 4.0, all gave similar agreement with experiment, with the latter method giving the largest number of sites within one pK unit of the experimental range. On the other hand, minimization with a dielectric constant of 2 or 4 gives significantly poorer agreement with experiment. The two-step procedure is in accord with an intuitive sense that a good minimization procedure should relieve steric overlaps without moving the structure too far (the non-electrostatic step) and should place the structure near a minimum of a potential energy function that reflects some degree of solvent screening of the electrostatic interaction (the $\epsilon = 4.0$ or 10 step). However, one cannot conclude from the single example of lysozyme that this procedure is best in general, and the sensitivity of single-conformer calculations to minimization shows that caution is called for in the preparation of structures for calculations and the interpretation of results. Particular care is needed in identifying situations such as the Glu-35–Asp-52 case, where site-site couplings heighten the sensitivity of the results to conformational details.

Including local conformational mobility tends to reduce the magnitude of the pK_{model}–pK_{intr} shift and to partially mitigate the sensitivity of the calculations to the minimization protocol for the case of lysozyme. It seems likely that this will be true for other proteins as well because the rigid, one-conformer model implies conformational stress associated with pK_{intr} shifts and the introduction of conformational variation provides partial relief. However, not all sites display reduced pK_{intr} shifts upon the introduction of conformational variation. Including conformational flexibility by the methods developed here increases computational cost roughly in proportion to the number of conformers included. Therefore, it may be useful to include conforma-

tional variation only for those sites that are of particular interest—enzyme active sites, for example—and to use the single-conformer formalism for other sites. The methods presented here could also be used to study pH-dependent conformational changes.

The conformer selection methods with no minimization or only localized minimization (STIFF and FMINL, respectively) gave relatively little overall improvement in agreement with experiment, whereas the protocol using global minimization FMING gave the best results. Although the multi-conformer formalism here was developed for localized conformational change, it appears that the small global changes allowed by global minimization do no harm as far as agreement with experiment is concerned.

Possible improvements

The method of generating model compounds for particular sites directly from protein conformations is an unattractive feature of both the new methods presented here and the older single-conformer method. Real model compounds do not "belong" to any particular residue position in a protein, so there should not be one "model compound" for Asp-52 and a different one for Asp-119, and so on. In the single-conformer method this questionable treatment of model compounds was made necessary by the assumption of rigidity that the model implied, and it served as a convenient way of canceling lattice artifacts. However, in the multi-conformer formalism there is no inherent need for the model compound to be sampled in the same way or with the same number of conformers as the corresponding site in the protein, and the lattice artifacts should, for the most part, be canceled by the ϕ_{unif} terms in Eq. 18. Therefore, it should be possible to separate the conformational sampling of the model compound from that of the protein and break the connection between model compound calculations and particular residue positions in the protein sequence.

We have found that in the partition-function-like terms that make up Eq. 16, only a few terms in the sums for the protein contribute significantly to the sum. In other words, only a few of the 36 conformers sampled turn out to be important. It would reduce computational cost to include only these, but we do not yet have a way to know which ones are important without doing the full calculation. An easily evaluated empirical energy function that could predict in advance the most important conformers would be useful in this regard. The significant differences between the FMINL and FMING results suggest that the methods used for structure refinement will need careful testing. For cases where side chains move between exposed and buried positions, including a surface area-dependent term for the non-electrostatic interaction of the side chain with solvent might also improve results.

The complex interplay of site-site interactions observed in the Glu-35–Asp-52 pair suggests that the neglect of correlations between conformational changes at different side chains may be a serious limitation in some cases. However, a full treatment in which N sites could each go through M different conformations as well as two different protonation states (or more if there is tautomerism) would require a calculation of the order $(2M)^N$, which is obviously prohibitive if carried out with the numbers of conformers and sites used in the present study. The problem might become tractable if the number of conformers to be sampled for each side chain could be reduced, as suggested above, and if suitable approximations could be made to treat explicitly only conformational correlations between strongly coupled sites and to treat more distant pairs in the averaged way presented here.

This work was supported by the National Institutes of Health (GM45607).

REFERENCES

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Bartik, K., C. Redfield, and C. M. Dobson. 1994. Measurement of the individual pK_a values of acidic residues of hen and turkey lysozymes by two-dimensional ^{1}H NMR. *Biophys. J.* 66:1180–1184.

Bashford, D., D. A. Case, C. Dalvit, L. Tennant, and P. E. Wright. 1993. Electrostatic calculation of side-chain pK_a values in myoglobin and comparison with NMR data for histidines. *Biochemistry*, 32: 8045–8056.

Bashford, D., and K. Gerwert. 1992. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* 224: 473–486.

Bashford, D., and M. Karplus. 1990. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry.* 29: 10219–10225.

Bashford, D., and M. Karplus. 1991. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* 95:9556–9561.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Born, M. 1920. Volumes and heats of hydration of ions. *Z. Phys.* 1:45–48.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

Brünger, A. T., and M. Karplus. 1988. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins.* 4:148–156.

Connolly, M. L. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science.* 221:709–713.

Gilson, M. K., J. A. McCammon, and J. D. Madura. 1995. Molecular dynamics simulation with a continuum electrostatic model of the solvent. *J. Comp. Chem.* 16:1081–1095.

Gilson, M. K., and B. Honig. 1988. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins.* 4:7–18.

Hodsdon, J., G. Brown, L. Sieker and L. Jensen. 1990. Refinement of triclinic lysozyme. I. Fourier and least-squares methods. *Acta Cryst. B.* 46:54–62.

Kuramitsu, S., and K. Hamaguchi. 1980. Analysis of the acid-base titration curve of hen lysozyme. *J. Biochem.* 87:1215–1219.

Matthew, J., and F. Gurd. 1986. Calculation of electrostatic interactions in proteins. *Methods Enzymol.* 130:413–436.

McGrath, M. E., J. R. Vásquez, C. S. Craik, A.-S. Yang, B. Honig, and R. Fletterick. 1992. Perturbing the polar environment of Asp102 in trypsin: consequences of replacing conserved Ser214. *Biochemistry.* 31: 3059–3064.

Perutz, M. 1978. Electrostatic effects in proteins. *Science.* 201:1187–1191.

Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

Sharp, K. 1991. Incorporating solvent and ion screening into molecular dynamics using the finite-difference Poisson-Boltzmann method. *J. Computat. Chem.* 12:454–468.

Sharp, K., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and experiment. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332.

Takahashi, T., H. Nakamura, and A. Wada. 1992. Electrostatic forces in two lysoyzmes: calculations and measurements of histidine pK_a values. *Biopolymers.* 23:897–909.

Tanford, C., and J. G. Kirkwood. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79: 5333–5339.

Tanford, C., and R. Roxby. 1972. Interpretation of protein titration curves. *Biochemistry.* 11:2192–2198.

Warshel, A., and S. Russell. 1984. Calculation of electrostatic interactions in biological systems and in solution. *Q. Rev. Biophys.* 17: 283–422.

Warwicker, J., and H. C. Watson. 1982. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.* 157:671–679.

Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. S. Alagona, J. Profeta, and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

Whitlow, M., and M. M. Teeter. 1986. An empirical examination of potential energy minimization using the well-determined structure of the protein crambin. *J. Am. Chem. Soc.* 108:7163–7172.

Yang, A.-S., M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig. 1993. On the calculations of pK_as in proteins. *Proteins.* 15:252–265.

Yang, A.-S., and B. Honig. 1994. Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.* 237:602–614.

You, T., and D. Bashford. 1995. An analytical algorithm for the rapid determination of the solvent accessibility of points in a three-dimensional lattice around a solute molecule. *J. Comp. Chem.* 16: 743–757.