Incorporating Protein Conformational Flexibility into the Calculation of pH-Dependent Protein Properties

E. G. Alexov and M. R. Gunner
Department of Physics, City College of New York, New York, New York 10031 USA

ABSTRACT A method for combining calculations of residue pK_a's with changes in the position of polar hydrogens has been developed. The Boltzmann distributions of proton positions in hydroxyls and neutral titratable residues are found in the same Monte Carlo sampling procedure that determines the amino acid ionization states at each pH. Electrostatic, Lennard-Jones potentials, and torsion angle energies are considered at each proton position. Many acidic and basic residues are found to have significant electrostatic interactions with either a water- or hydroxyl-containing side chain. Protonation state changes are coupled to reorientation of the neighboring hydroxyl dipoles, resulting in smaller free energy differences between neutral and ionized residues than when the protein is held rigid. Multiconformation pH titration gives better agreement with the experimental pK_a's for triclinic hen egg lysozyme than conventional rigid protein calculations. The hydroxyl motion significantly increases the protein dielectric response, making it sensitive to the composition of the local protein structure. More than one conformer per residue is often found at a given pH, providing information about the distribution of low-energy lysozyme structures.

GLOSSARY

II	ionized form of ionizable group (side chain of Asp, Glu, Tyr, Arg, Lys, His, and the N and C termini of the polypeptide chain (CTR, NTR))
NI	neutral form of ionizable group
MP	multiconformation polar group (buried waters and side chains of Ser, Thr)
FP	fixed polar group (backbone amide and side chains of Asn, Gln, Met, Trp, Cys)
$pK_{int,\,i}$	the apparent pK_a of group i including reaction field energy, interaction with fixed polar groups and backbone and nonelectrostatic energy

Interconvertions of energy units

1 ΔpK unit = 1.36 kcal/mol = 2.30kT.

INTRODUCTION

All proteins contain acidic and basic side chains. These ionizable residues play important roles in determining each protein's structure (Perutz, 1978), ligand binding, functional activity (Matthew, 1985; Sharp and Honig, 1990), and properties of mutants (Raghavan et al., 1989). They are the source of the pH dependence of protein activity and stability (Yang and Honig, 1992, 1993; Yang and Barry, 1994). It is therefore essential that proteins maintain the appropriate side-chain ionization states. pK_a's, the pH at which individual acidic and basic residues are 50% ionized, have been measured for a number of proteins (Bartik et al., 1994; Kuramitsu and Hamaguchi, 1980), and calculations have successfully reproduced these values (Antosiewicz et al., 1994; Bashford and Karplus, 1990; Yang et al., 1993; Oberoi and Allewell, 1993; Gibas and Subramaniam, 1996). These successful calculations can be seen to both test theoretical models and provide insight into how a protein maintains the appropriate charge state of important residues.

The protonation state of a protein changes as a function of pH differently than an assembly of isolated amino acids because the protein modifies the free energy of ionization. There are several established methods for determining the free energy of charges in proteins of known structure (Warwicker and Watson, 1982; Warshel and Russell, 1984; Gilson and Honig, 1987). The continuum electrostatic model assumes that the protein and solvent each have a different dielectric permeability or dielectric constant. The protein is a region of low dielectric response (ϵ_{in}), with a shape determined by the position and radius of the atoms in the protein (Richards, 1977; Gilson et al., 1985). The surrounding water is a uniform medium with a dielectric constant of 80 (ϵ_{out}). The charge distribution due to charged and polar groups in the protein is approximated by partial charges placed at atomic centers. The Poisson-Boltzmann equation provides the electrostatic potential everywhere, given the distribution of charges and dielectric response. Whereas this differential equation has no analytic solution for irregularly shaped objects, the finite-difference Poisson-Boltzmann (FDPB) (Davis and McCammon, 1990; Nicholls and Honig, 1991), boundary-element (Zauhar and Morgan, 1985), and finite-element (You and Harvey, 1993) methods provide numerical solutions. The continuum electrostatic calculations depend on a number of parameters, some of which have been systematically tested. For example, different val-

Received for publication 4 November 1996 and in final form 28 January 1997.

Address reprint requests to Dr. Marilyn R. Gunner, Department of Physics, City College of New York, 138th Street and Convent Ave., New York, NY 10031. Tel.: 212-650-5557; Fax: 212-650-5503; E-mail: gunner@gunner.sci.ccny.cuny.edu.

0006-3495/97/05/2075/19 $2.00

ues of the atomic partial charges, such as those provided by Charmm (Brooks et al., 1983) and Parse (Sitkoff et al., 1994), yield similar results in pK_a calculations (Gilson, 1995). However, the results are sensitive to the position of charges (Oberoi and Allewell, 1993) and to the chosen crystal structure. The inclusion of crystallographic water in a single, rigid conformation does not improve pK_a calculations (Gibas and Subramaniam, 1996). However, raising the protein dielectric constant can shift calculated pK_a's by several pH units (Antosiewicz et al., 1994, 1996).

The assumption that the protein structure is rigid and identical to the x-ray crystal structure is one of the most serious limitations of standard calculations of the changes in free energy when a group changes its charge state (Honig and Nicholls, 1995). First, calculations must address the general problem that there can be differences between a protein structure in the crystal and solution. In addition, crystals are grown at fixed pH in a specific ionization state. This structure may not be appropriate at all pHs, because ionization state changes will induce changes in protein structure to lower the free energy of the new charge distribution (Baker and Mauk, 1992; Turano et al., 1995). The protein dielectric constant incorporates an average response of polarizable electrons and flexible protein dipoles into the continuum electrostatics methodology. Four, the most commonly used value for ϵ_{in} (Harvey, 1989; Bashford and Karplus, 1991; Beroza et al., 1991; Lancaster et al., 1996) accounts for electronic polarizability and only small nuclear motions (Gilson and Honig, 1986). However, Antosiewicz et al. (1994) found the best agreement between calculated and experimental pK_a's with a protein dielectric constant of 20 in several small proteins. The successful match to experiment with calculations using a large dielectric constant suggests that conformation changes occur when residues change ionization state. Unfortunately, using a dielectric constant of 20 averages the effects of the structural changes over the whole protein. The challenge is to incorporate the local response of dipoles to charges in a way that provides atomic-level information about the groups that contribute to the dielectric relaxation processes.

A number of methods have been used to account for changes in protein conformation that accompany changes in ionization state. The simplest use molecular dynamics to generate one protein structure in which every ionizable residue is charged and another in which each residue is neutral (Yang and Honig, 1993; Yang et al., 1993). The first structure provides the free energy of interaction of ionized residues, whereas the second is used for neutral ionizables. You and Bashford (1995) developed a more realistic methodology, generating 36 side-chain conformations sampling χ_1 and χ_2 torsion angles of aspartic and tyrosine side chains, and glutamic acid χ_2 and χ_3 angles. They determined the intrinsic pK_a (the group's pK_a if all other titratable groups are neutral) of these residues given this set of conformations. The pairwise interactions between titratable residues were averaged over their Boltzmann distribution of conformers determined when all other residues were neutral. The pK_a calculation uses the Boltzmann average energies for interactions with other titratable groups, and the residue intrinsic pK_a's were treated as a canonical function of individual conformer pK_a's. Although some information about additional conformations is now included, these conformations are not dynamically accessible during the pH titration with the methods of Yang and Honig (1993), Yang et al. (1993), and You and Bashford (1995).

The goal is therefore to develop a general method for including protein flexibility in the calculation of protein ionization states. This requires a compact and efficient method for generating a set of states that defines each residue's conformation and charge, to obtain each state's energy, and then to determine the Boltzmann distribution of states at a given pH. The difficulty in incorporating realistic protein flexibility into calculations of protein pK_a's is the number of states that must be evaluated. If there are N ionizable residues, where each can be ionized or neutral, the system has 2^N possible states. Combining pH titration and conformational flexibility leads to an enormous increase in this number. For example, if each titratable residue has M different conformations in its ionized and neutral forms, and K nonionizable groups have L conformations, the number of possible states is now $(2M)^N(L)^K$. Because any choice of conformation influences the position of all other groups, all conformations should be given a chance of being realized.

For the purposes of the work presented here, conformational changes are divided into those that change heavy atom positions and those that change hydrogen positions. The first class involves larger displacements and higher energy barriers, and often produces a significant change in the protein's shape. Although these may certainly be important, they will not be considered here. In contrast, proton movements have smaller energy barriers, often can be carried out without the motion of other groups, and involve negligible changes in the protein's shape. We will explore the effects of allowing hydroxyl hydrogens on Ser, Thr, Tyr, and water to take different positions. These protons are not uniquely determined by the heavy atom positions, so they can be influenced by local, nonbonded electrostatic and steric constraints. Neutron diffraction experiments show that these hydrogens often adopt several conformations, demonstrating that there can be more than one accessible energy minimum (Kossiakoff et al., 1990, 1991; McDowell and Kossiakoff, 1995). Most important for pK_a calculations, hydroxyls can be hydrogen bond donors or acceptors, and so they may change their role as the ionization states of nearby residues change (Baker and Hubbard, 1984). Incorporating conformational flexibility for protons also allows neutral states to be treated more realistically. Thus the position of the proton on neutral Asp, Glu, His, or Arg can be determined dynamically by the local electrostatic environment. The method described in this paper fixes none of these proton positions before the pK_a calculation. Rather, interaction energies between a number of predetermined, likely minimum energy residue conformers were calculated. Then the distribution of conformers at a given pH was determined

by Monte Carlo sampling, assuming that the state energies are a linear combination of conformer energies. Triclinic hen egg lysozyme was used to test the method, providing an opportunity to compare our calculations to both the experimental data (Bartik et al., 1994; Kuramitsu and Hamaguchi, 1980) and earlier calculations (Yang and Honig, 1993; Antosiewicz et al., 1994; You and Bashford, 1995; Gibas and Subramaniam, 1996).

MATERIALS AND METHODS

Data file and parameters used

The file 1LZT (Hodsdon et al., 1985) in the Brookhaven Protein Data Bank (Bernstein et al., 1977) provided the atomic coordinates of triclinic hen egg lysozyme grown at pH 4.5. All heavy atoms were maintained in their positions in the crystal structure. All crystal waters were deleted, except for the seven buried in the protein, with less than 1% of their surface accessible to bulk water. (The oxygen coordinates for water 14 and 31 in the 1LZT file differ by only 0.07 Å. Water 14 was deleted.) Buried waters were treated in the same atomic detail as the protein, whereas surface waters were replaced by continuum solvent. All electrostatic energy terms were calculated by the finite-difference Poisson-Boltzmann (FDPB) method with the program DelPhi (Nicholls and Honig, 1991). Three DelPhi focusing runs provided a final scale of 2.13 grids/Å (Gilson and Honig, 1987). A dielectric constant of 4 was used for the protein and 80 for the solvent, the ionic strength was 0.15 M (Hodsdon et al., 1985), the water probe radius was 1.4 Å, and the Stern ion exclusion layer was 2.0 Å. Atomic radii were the same as those used by Gunner et al. (1996). The charges for neutral and ionized acidic and basic residues are provided in Table 1. CHARMM charges were used for all other groups (Brooks et al., 1983).

Nonelectrostatic energy terms are also considered. The parameters for the Lennard-Jones potentials were taken from the x-plor tophar19.pro file (Brünger and Karplus, 1988). Neutral Tyr has two energy minima in the plane of the Tyr ring, with a 2.59 ΔpK units (3.5 kcal/mol) barrier between them. The torsion energy as a function of torsion angle (φ) was modeled as

$$E(\varphi) = -\frac{2.59}{2}\cos(2\varphi). \quad (1a)$$

There are three degenerate low-energy positions for Ser and Thr at torsion angles 60°, 180°, and 300°, with an energy barrier of 0.96 ΔpK units (1.3 kcal/mol) (McDowell and Kossiakoff, 1995):

$$E(\varphi) = -\frac{0.96}{2}\cos 3(\varphi - 60). \quad (1b)$$

In solution the most favorable neutral Asp and Glu hydroxyl position is in the *syn* conformation lying between and in the plane of the two carboxylate oxygens (Gu, 1996). The energy of this hydroxyl proton was calculated as the sum of two terms: 1) a torsion potential with energy minima at 0° and 180° and a barrier of 0.96 ΔpK units; 2) an electrostatic attraction between the hydroxyl proton and the nonbonded carboxylate oxygen, which has a minimum energy of −1.99 ΔpK units (−2.7 kcal/mol) favoring the hydrogen position between oxygens:

$$E(\varphi) = -\frac{1.99}{2}\cos(\varphi) - \frac{0.96}{2}\cos(2\varphi). \quad (1c)$$

This results in a −1.99 ΔpK units preference for the *syn* (torsion angle 0°) over the *anti* conformation (torsion angle 180°). This is comparable to the value found by solid-state NMR measurements (Gu, 1996).

TABLE 1 Partial charges for ionized and neutral acids and bases

Residue type	Atom type	Ionized	Neutral
Asp	CB	−0.160	0.000
	CG	0.360	0.550
	O D1	−0.600	−0.475
	O D2	−0.600	−0.475
	H H	0.000	0.400
Glu	CG	−0.160	0.000
	CD	0.360	0.550
	O E1	−0.600	−0.475
	O E2	−0.600	−0.475
	H H	0.000	0.400
Tyr	CZ	0.000	0.250
	O H	−1.000	−0.650
	H H	0.000	0.400
Ctr	CD	0.140	0.140
	O T1	−0.570	−0.070
	O T2	−0.570	−0.070
Ntr	C	0.250	0.250
	N T	−0.300	−0.300
	H T1	0.350	0.020
	H T2	0.350	0.020
	H T3	0.350	0.010
Arg	CD	0.100	0.100
	N E	−0.400	−0.400
	H E	0.300	0.300
	CZ	0.500	0.500
	N H1	−0.450	−0.450
	N H2	−0.450	−0.450
	H H11	0.350	0.100
	H H12	0.350	0.100
	H H21	0.350	0.100
	H H22	0.350	0.100
His	CG	0.330	0.100
	N D1	−0.540	−0.400
	C E1	0.360	0.300
	N E2	−0.540	−0.400
	C D2	0.190	0.100
	HD1	0.460	0.300
	H E1	0.140	0.000
	H E2	0.460	0.000
	H D2	0.140	0.000
Lys	CE	0.250	0.250
	NZ	−0.300	−0.300
	H Z1	0.350	0.020
	H Z2	0.350	0.020
	H Z3	0.350	0.010

Only the neutral His-ND1 form is given, but the partial charge distribution of His-NE2 can be found by simple exchange of partial charges of atoms "D" and "E".

Multiple conformation problem

Generating multiple conformations

Most polar hydrogens whose presence or position is not fixed by the protein heavy atom positions are assigned alternative positions and/or occupancies. This yields several conformers for these residues. Here the term "conformer" denotes forms of a residue that differ by either hydrogen position or ionization state. Every titratable group within the protein has at least two possible conformers: ionized (II) and neutral (NI) ionizable. Ionizable residues gain or lose their proton on different atoms or in

different positions, generating additional conformers. Thus Glu gains a proton to OE1 or OE2, Asp to OD1 or OD2; His loses a proton from ND1 or NE2, Arg from NE, NH1, or NH2. In addition, the proton on neutral Asp, Glu, and Tyr as well as Ser and Thr can rotate around the oxygen to which it is bound (Kossiakoff et al., 1990). Water hydrogens can be oriented in any direction from the fixed oxygen. Ser, Thr, and waters that form different conformers with different hydroxyl proton positions are classified as mobile polar (MP) groups. All protein dipoles and charges that have only a single conformation are classified as fixed polar (FP) groups.

The first step in the multiconformation analysis is to generate possible positions for MP hydroxyl and NI carboxylate protons so that one will be favorable in every possible ionization state and proton position for the other resides. Protons are generated by three methods, implemented in parallel. In each case multiple minima generate additional conformers. 1) Protons are placed to minimize the distance to any possible hydrogen bond acceptor. Acceptors are all oxygen-containing FP groups, and the oxygen and nitrogen of every acid, base, and hydroxyl group, stripped of their mobile protons. 2) Protons are placed at minima of the electrostatic potential in a protein where every Asp, Glu, Tyr, His, Lys, and Arg is ionized. 3) Protons are placed at electrostatic potential minima when all residues are neutral. The energy landscape for the last two methods is obtained from the electrostatic interactions with all residues, with at least one heavy atom within 4.5 Å of the oxygen to which the mobile proton is bound. Electrostatic interactions are determined with Coulomb's law, using the atomic charges shown in Table 1 and CHARMM charges for other residues. Torsion energies are ignored at this stage, but the Tyr hydroxyl protons were kept near the plane of the ring. Proton positions from the three runs are compared and duplicate positions are discarded. Although other schemes can be used to generate proton positions, this method yielded Ser, Thr, neutral Tyr, and waters with 1 to 3 conformers and neutral Asp and Glu with 3 to 6 conformers (see Table 2). A protein data file is built up that includes all alternative proton positions.

Determining the free energy of each state

A simple "protein" with an Asp, Ser, His and a water shows how states and their free energies are defined (Fig. 1). The neutral Asp has three proton positions, neutral His two positions, and Ser and water two positions. The system contains 11 conformers, including two II conformers: Asp^- and His^+; five NI conformers: Asp^0 with H1, H2, or H3 and His-ND1 and His-NE2; and four MP conformers: Ser with H1 or H2 and water HOH1, HOH2. The FP (fixed polar) groups are any dipoles and charges with a single conformation. A residue's backbone and side chain are treated as separate groups. Thus the backbone atoms of a MP or ionizable residue are considered an FP group. Forty-eight states can be constructed from all combinations of the 4 Asp and 3 His, 2 Ser and 2 water conformers, with each residue occupying only one conformation in a given state. In general, the number of states $= R = \prod_{i=1}^{K}(n_i)$, where K is the number of multiconformation residues and n_i is the number of conformations of the ith group.

The free energy for the ith conformer is a sum of four terms: 1) the nonelectrostatic energy $\Delta G_{\text{nonel,i}}$, which contains the conformer-dependent torsion and the Lennard-Jones energies between the proton and neighboring heavy atoms; 2) the reaction field energy $\Delta G_{\text{rxn,i}}$, also called the Born, solvation, or self energy; 3) the polar energy $\Delta G_{\text{pol,i}}$, which accounts for pairwise charge-charge interactions with FP sites; and 4) the pairwise interactions with all other conformers in the state (ΔG^{ij}). The free energy of a given state is assumed to be the sum of the free energies of the individual conformers. The first three terms are assumed to be independent of the conformer that is occupied for any other residue.

One DelPhi calculation for each conformer provides all three electrostatic energy terms (Beroza et al., 1991; Gilson, 1993; Gilson and Honig, 1988; Lancaster et al., 1996; Nicholls and Honig, 1991). Each DelPhi run has partial charges on only one conformer. The net charge is +1 or −1 for II conformers and 0 for polar NI and MP conformers. The m DelPhi runs for the $m = \Sigma_{i=1}^{K} n_i$ conformers yield 1) a vector of length m with each conformer's reaction field energy obtained by the method described by Nicholls and Honig (1991); 2) a vector of length m containing each conformer's pairwise interactions with all FP sites (this is obtained by summing the product of the potential at each FP atom from the conformer's charges and the partial atomic charge on that FP atom; Gilson and Honig, 1988); 3) an $m \times m$ matrix for pairwise interactions between conformers. This is obtained by summing the product of the potential and the atomic partial charge over the atoms in each II, NI, or MP conformer. The pairwise interaction between conformers belonging to the same residue is set to zero (Table 4).

The free energy of the nth state in a protein having K multiconformation residues with a total of m conformers at a given pH is

$$\Delta G^{n} = \sum_{i=1}^{m} \delta_n(i)\{\gamma(i)[2.3k_bT(\text{pH} - \text{pK}_{\text{sol,i}})] + (\Delta G_{\text{rxn,i}} + \Delta G_{\text{pol,i}} + \Delta G_{\text{nonel,i}} + \Delta G_{\text{entropy,i}})\} + \sum_{i=1}^{m} \delta_n(i) \sum_{j=i+1}^{m} \delta_n(j)\Delta G^{ij}. \quad (2)$$

$\delta(i)$ is 1 for conformers that make up the state and 0 for all others. Each state has one conformer for each residue. $\gamma(i)$ is 1 for bases, −1 for acids, and 0 for NI and MP groups. $\text{pK}_{\text{sol,i}}$ is the pK_a of the ith group in solution (Table 3). The $\text{pH-pK}_{\text{sol,i}}$ term provides the free energy of ionization of the ith amino acid isolated from the protein at this pH. Without the protein, the electrostatic free energy of all neutral conformers is defined as zero, so $\text{pK}_{\text{sol,i}}$ is the free energy difference between ionized and neutral forms in solution. $\Delta G_{\text{rxn,i}}$ is the difference between the ith conformer's reaction field energy and the reference reaction field energy for this type of residue in solution (Table 3). NI and MP solution reaction field energies were calculated by averaging over the value for different conformations of these groups isolated from the protein. $\Delta G_{\text{pol,i}}$ is the ith conformer's interaction with all FP charges and dipoles. In standard, rigid protein calculations $\Delta G_{\text{pol,i}}$ also includes interactions with neutral Tyr, Ser, Thr, and water (Bashford and Karplus, 1990; Yang et al., 1993). However, the multiconformation pK_a calculations treat these interactions in ΔG^{ij}. $\Delta G_{\text{nonel,i}}$ is the energy difference between the ith conformer's energy and the lowest nonelectrostatic energy for this group in solution. Lennard-Jones and torsion angle energies are included as described above. $\Delta G_{\text{entropy,i}}$ is an entropy correction described below. The last, double sum in Eq. 2 describes the pairwise electrostatic interactions between the appropriate conformers of the residues. The limits on the summation ensure that each pairwise interaction is counted only once.

The probability of realizing a given conformation at a particular pH depends on the energy of all the states where this conformer is realized. The conformer population in a Boltzmann distribution of states at a given pH is obtained by Monte Carlo sampling, using the state energies defined in Eq. 2 (Beroza et al., 1991; Lancaster et al., 1996). The only modifications of the standard method for pK_a calculations used by Lancaster et al. (1996) is to allow only one conformer for each residue in any state and to use only one flip for each step (Beroza et al., 1991). At each pH the conformer occupancy for each residue is the average of 300,000 Monte Carlo sample steps after 1000 preequilibration steps. The provides a precision of ±2% for the fractional conformer occupancies. The precision was calculated numerically on the basis of 10 different runs at pH 5. Thus the problem of obtaining the ionizable group charge and the probability of realizing each conformer is solved simultaneously in a self-consistent manner.

The entropy correction required for multiconformation calculations

In conventional pK_a calculations both ionized and neutral forms of a residue have only one conformation and therefore the same entropy. In the work presented here there is only one ionized conformer, whereas there is

TABLE 2 All conformers used in the multiconformation calculations (energy in ΔpK units)

type#	res. num.	ΔG rxn	ΔG pol	hydroxyt torsion angle	ΔG non-el	occu-pancy pH 5 %	type#	res. num.	ΔG rxn	ΔG pol	hydroxyt torsion angle	ΔG non-el	occu-pancy pH 5 %
ARG+	5	1.22	-0.30			100.0	ASP-	66	7.25	-3.86			100.0
ARG^0	5	0.69	-0.33				ASP,1a	66	1.21	0.10	36.43	0.65	
ARG+	14	1.26	-0.06			100.0	ASP,2a	66	1.19	0.37	66.50	1.03	
ARG^0	14	0.52	-0.02				ASP,3a	66	1.23	-0.05	-3.54	1.90	
ARG+	21	1.26	-0.59			100.0	ASP,1b	66	1.11	-1.29	76.24	3.72	
ARG^0	21	0.45	-0.18				ASP,2b	66	1.23	-0.84	-3.79	-0.22	
ARG+	45	2.39	0.17			100.0	ASP-	87	1.15	-1.75			100.0
ARG^0	45	0.90	0.16				ASP,1a	87	0.31	-0.01	27.97	1.04	
ARG+	61	2.73	-0.22			100.0	ASP,2a	87	0.0	0.43	-29.50	2.51	
ARG^0	61	0.87	0.06				ASP,3a	87	0.41	-0.01	1.16	2.25	
ARG+	68	1.01	-0.04			100.0	ASP,4a*	87	0.0	2.07	-171.94	4.58	
ARG^0	68	0.75	0.02				ASP,1b	87	0.56	-0.57	-2.22	-0.12	
ARG+	73	0.98	-1.18			100.0	ASP-	101	2.37	3.80			2.7
ARG^0	73	0.93	-0.41				ASP,1a	101	0.18	-2.10	114.34	4.96	37.8
ARG+	112	2.52	-0.30			100.0	ASP,2a	101	0.86	-0.30	-0.40	1.92	21.5
ARG^0	112	0.84	0.30				ASP,1b*	101	0.45	-0.83	-100.36	3.23	17.8
ARG+	114	1.03	-0.08			100.0	ASP,b2	101	0.86	-0.39	4.31	1.78	20.2
ARG^0	114	0.54	0.01				ASP-	119	1.39	-1.45			100.0
ARG+	125	1.66	-1.41			100.0	ASP,1a*	119	0.62	-0.19	-16.42	1.75	
ARG^0	125	0.86	-0.18				ASP,2a	119	0.0	-0.20	173.67	1.80	
ARG+	128	0.56	0.06			100.0	ASP,3a	119	0.71	-0.15	43.05	-0.24	
ARG^0	128	0.60	0.02				ASP,1b	119	0.00	0.23	130.83	1.38	
ASP-	18	2.02	-1.80			100.0	ASP,2b	119	0.60	0.31	0.89	1.92	
ASP,1a*	18	0.38	0.87	103.06	2.17		ASP,3b	119	0.33	1.05	-159.11	-0.26	
ASP,2a	18	0.37	0.06	3.13	-0.25		GLU-	7	1.35	-1.73			100.0
ASP,1b	18	0.18	-0.71	178.50	1.87		GLU,1a*	7	0.35	1.45	-151.2	4.86	
ASP,2b	18	0.50	-0.74	-1.53	-0.17		GLU,2a	7	0.18	0.33	101.06	5.74	
ASP-	48	3.52	-2.03			100.0	GLU,3a	7	0.47	-0.36	-1.06	1.85	
ASP,1a*	48	0.70	1.54	177.76	1.63		GLU,1b	7	0.0	-1.33	177.48	1.94	
ASP,2a	48	0.50	0.69	107.58	2.69		GLU,2b	7	0.7	-1.21	-2.58	-0.25	
ASP,3a	48	1.04	0.32	-12.40	3.64		GLU-	35	3.15	-1.49			73.7
ASP,1b	48	0.27	-0.89	130.76	1.48		GLU,1a*	35	0.67	-0.88	75.18	1.51	7.8
ASP,2b	48	0.81	-0.66	70.76	1.04		GLU,2a	35	0.61	-0.63	-174.68	1.23	1.3
ASP,3b	48	1.12	-0.56	10.73	1.96		GLU,a3	35	0.91	-0.82	-4.68	1.91	0.1
ASP-	52	3.24	-2.05			28.0	GLU,1b	35	0.49	-0.36	-122.38	0.01	5.3
ASP,1a*	52	0.37	-0.0	121.09	1.94	0.0	GLU,b2	35	0.83	-0.25	-2.39	-0.36	11.9
ASP,1b	52	0.32	-1.24	-71.23	0.28	39.2							
ASP,2b	52	0.82	-1.64	58.78	0.37	32.5							

#Residue charge (+1, −1, or 0). The letter with the neutral forms of Asp, Glu denotes a hydrogen bound to (a) OD1/OE1 and (b) to OD2/OE2. ΔpK_{rxn} is the difference between the conformer's reaction field energy and the reference energy in Table 3. Nonelectrostatic torsion energy of hydroxyls was determined with Eq. 1. ΔpK_{none1} is the conformer nonelectrostatic energy minus the lowest nonelectrostatic energy in solution for this type of residue (see reference energy in Table 3). For residues with more than one conformer, the position used in the fixed structure calculations is shown with an asterisk. Ionized conformers are highlighted.

usually more than one neutral form. Without correction, the larger number of neutral conformers stabilizes the neutral form by increasing its entropy. $\Delta G_{entropy,i}$ in Eq. 2 is an entropy correction of $\ln(n_i) * 0.59$ (kcal/mol), where n_i is the number of NI or II conformers. No entropy correction is needed for MP groups because they exist only in one, neutral state.

Formally the entropy correction should be obtained from the logarithm of the canonical partition function (McQuarrie, 1973):

$$\Delta G_{entropy,i} = \ln \sum_{\text{number of conformers}} \exp(-(G_i - G_0)/kT), \quad (3)$$

Table 2—*continued*

type#	res. num.	ΔG rxn	ΔG pol	hydroxyt torsion angle	ΔG non-el	occu-pancy pH 5 %	type#	res. num.	ΔG rxn	ΔG pol	hydroxyt torsion angle	ΔG non-el	occu-pancy pH 5 %
LYS+	1	0.56	0.07			100.0	THR-H1	51	0.78	-2.54	14.30	0.79	42.1
LYS0	1	0.59	0.01				THR-H2*	51	0.80	0.55	74.22	0.00	57.9
LYS+	13	0.31	0.08			100.0	THR-H1	69	0.85	-0.25	-44.57	0.0	2.4
LYS0	13	0.57	0.06				THR-H2*	69	0.85	0.72	75.33	0.53	97.6
LYS+	33	1.10	-1.61			100.0	THR-H1*	89	0.56	-0.05	56.09	0.19	62.8
LYS0	33	0.62	0.21				THR-H2	89	0.61	1.13	96.04	0.00	37.2
LYS+	96	1.93	-2.50			100.0	THR-H1	118	0.46	-0.25	-43.01	0.0	95.4
LYS0	96	0.67	0.04				THR-H2*	118	0.63	0.89	46.83	0.35	4.6
LYS+	97	0.33	-0.61			100.0	TYR-	20	3.22	0.72			
LYS0	97	0.58	-0.05				TYR-H1*	20	0.12	-3.67	29.04	1.24	55.2
LYS+	116	0.01	-0.90			100.0	TYR-H2	20	0.08	-2.53	-20.94	0.92	44.8
LYS0	116	0.55	0.0				TYR-	23	3.07	-1.40			
SER-H1*	24	0.62	-1.52	-146.67	0.0	64.3	TYR-H1	23	0.22	-1.51	-13.23	0.05	59.2
SER-H2	24	0.66	-1.05	143.30	0.98	35.7	TYR-H2	23	0.42	-1.29	1.45	0.42	40.8
SER-H1*	36	0.85	-0.56	-110.34	1.77	86.0	TYR-	53	6.37	-1.49			
SER-H2	36	0.84	0.74	-20.43	0.0	14.0	TYR-H1*	53	0.60	0.29	-158.92	2.70	99.7
SER-H1	50	0.74	-0.47	-76.73	2.18	0.6	TYR-H2	53	0.64	0.62	41.14	0.96	0.3
SER-H2*	50	0.76	0.06	-27.70	0.0	91.4	NTR+	1	2.74	1.44			100.0
SER-H3	50	0.72	-0.53	-116.93	0.31	8.0	NTR0	1	1.20	1.25			
SER-H1*	60	0.86	-0.48	145.05	0.0	65.2	HIS+	15	5.44	-1.09			30.0
SER-H2	60	0.86	-1.09	105.17	1.43	34.8	HIS-ND1*	15	0.99	-0.47			51.7
SER-H1	72	0.64	-2.03	47.99	0.0	94.6	HIS-NE1	15	0.99	0.27			18.3
SER-H2*	72	0.68	-0.12	-91.97	0.46	5.4	HOH-1*	1	0.84	-1.87		0.34	81.1
SER-H1*	81	0.15	-0.24	60.34	0.13	65.2	HOH-2	1	0.80	-1.31		1.04	17.9
SER-H2	81	0.24	-0.06	76.92	0.0	34.8	HOH-3	1	0.83	0.76		0.0	1.1
SER-H1*	85	0.26	-0.06	-143.68	1.15	18.2	HOH-1*	17	0.81	-2.47		0.43	19.8
SER-H2	85	0.23	0.24	176.28	0.00	81.8	HOH-2	17	0.86	-2.54		0.0	80.2
SER-H1*	86	0.40	-0.50	157.23	0.0	57.7	HOH-1*	31	0.86	-1.10		0.25	25.0
SER-H2	86	0.36	0.59	-92.88	1.06	42.3	HOH-2	31	0.86	-1.26		0.0	75.0
SER-H1*	91	0.86	0.78	-11.84	0.0	73.8	HOH-1*	34	0.61	-2.04		0.0	73.8
SER-H2	91	0.85	0.35	-71.52	0.65	26.2	HOH-2	34	0.65	-2.53		1.56	26.2
SER-H1	100	0.38	0.49	-22.94	0.0	99.5	HOH-1*	36	0.85	-1.78		0.0	48.9
SER-H2*	100	0.38	-0.40	88.98	2.43	0.5	HOH-2	36	0.80	-1.02		0.11	51.1
THR-H1*	40	0.81	-0.28	71.44	0.0	39.5	HOH-1*	62	0.86	-2.54		0.52	61.0
THR-H2	40	0.63	0.09	86.20	0.02	60.5	HOH-2	62	0.82	-2.08		0.0	39.0
THR-H1*	43	0.30	-0.16	170.11	0.0	49.3	HOH-1*	169	0.81	-1.20		0.38	40.6
THR-H2	43	0.08	0.07	176.4	0.26	50.7	HOH-2	169	0.85	-1.33		0.0	59.4
THR-H1*	47	0.01	-1.14	-42.05	0.0	55.4	CTR-	129	1.37	-0.50			98.5
THR-H2	47	0.10	-0.15	174.23	0.04	44.6	CRT0	129	1.21	0.11			1.5

where the lowest conformer energy in the protein is G_0 and the sum runs over the neutral and ionized conformers for this residue. Thus, qualitatively it is not the number of conformers, but the number with significant occupancy in a Boltzmann distribution of states that properly defines the entropy. The exact calculation would require an iterative procedure to obtain self-consistent values for the Boltzmann distribution of states and the residue's entropy correction to each state's energy. The entropy correction in Eq. 2 assumes that all of a residue's conformers have the same energies and the same population. The error introduced by this approximation is small. The maximum error in $\Delta G_{\text{nonel},i}$ is 0.7 ΔpK units, which would be found if a residue with six conformers, the maximum for any residue, had only one that was energetically accessible. The error is generally smaller because most residues have fewer conformers and more than one is often occupied. The error in the entropy contribution to the energy difference between neutral and ionized form of a given residue can be calculated, knowing the conformer occupancy. The entropy free energy associated with M_1 conformers is

$$\Delta G_{\text{entropy}} = \ln\left(1 + \sum_{k=2}^{M_1} \frac{\rho_k}{\rho_1}\right), \tag{4}$$

TABLE 3 Reference energy values for residues in isolation from the protein in ΔpK units

Type	pK_a(sol)[a]	G_{rxn}(ion)[b]	G_{pol}(neu)[c]	G_{nonel}(ref)[d]
Arg	12.5	−12.5	−1.2	0
Asp	3.9	−12.8	−1.3	−1.47
Ctr	3.6	−13.7	−0.4	0
Glu	4.3	−12.8	−1.3	−1.47
His	6.5	−10.2	−1.3	0
Lys	10.8	−14.6	−0.8	0
Ntr	8.0	−13.6	−0.9	0
Tyr	10.9	−13.5	−0.9	−1.29

(a) Solution pK_a values; reaction field energies for (b) ionized and (c) neutral forms. (d) Nonelectrostatic energy for the minimum energy conformer in solution.

where ρ_i is the occupancy of the most populated conformer and ρ_κ is the occupancy of the other conformers of that residue. Given the calculated conformer occupancies, the error due to our assumption of equal conformer occupancy ranges from 0.19 ΔpK units for Asp101 (see Table 2, occupancy at pH 5) to 0.57 ΔpK units for Asp87 (see Fig. 2, occupancy at pH 1).

pK_as in a rigid protein structure

The standard, fixed structure and the new, mobile proton pK_a calculations were compared for lysozyme. Hydroxyl hydrogens for the fixed structure were added in two steps. First, MP hydrogens were placed at an energy minimum, assuming that all titratable sites are ionized. Then acid NI protons were added without changing the MP positions. The chosen hydrogens are noted in Table 2. This procedure provides proton positions that are close to the structure having lowest energy only in a limited pH range. Equation 2 describes the state energy for a protein without conformational flexibility (Bashford and Karplus, 1990; Yang et al., 1993) if the summation considers only ionized groups, the entropy and nonelectrostatic energies are omitted, and interactions with MP groups are included in $\Delta G_{pol,i}$. Because all titratable groups have fixed conformation, $\Delta G_{entropy,i}$ is zero. The difference in nonelectrostatic energy due to the titrating proton in II and NI groups is traditionally ignored. In addition, because all polar groups have only one conformation, their reaction field, polar, and nonelectrostatic energies add a constant to the free energy of all states and so can be omitted. $\Delta G_{rxn,i}$ or $\Delta G_{pol,i}$ is now the difference between the reaction field or polar energy in the ionized and neutral residues. In contrast, the multiconformation method uses the residue in solution as the reference state for these two terms. Monte Carlo sampling again provides the Boltzmann distribution of states at each pH.

Calculation of the contribution of proton flexibility to the effective dielectric constant (ϵ_{eff}) of the protein

The interactions between charges in a protein are modulated by electronic polarizability and rearrangements of protein atoms, buried waters, the surrounding media, including water and salt. The effective dielectric constant is defined as the ratio between the pairwise interaction energy of two sites in a homogeneous medium with dielectric constant $\epsilon = 1$ and this energy in the system in question. The first term is obtained with Coulomb's law, where $G_{ij}(\text{Coulomb}) = q_i q_j / r_{ij}$. The latter energy is obtained with DelPhi, given $G_{ij}(\text{DelPhi}) = \Psi_{ij} q_j$ (where Ψ_{ij} is the potential from the charge at site i at the uncharged position j and q_j is the charge to be placed at j). Thus,

$$\varepsilon_{\text{eff}} = \frac{q_i}{r_{ij}\Psi_{ij}} \equiv \frac{G_{ij}(\text{Coulomb})}{G_{ij}(\text{DelPhi})} \tag{5}$$

(Bashford and Karplus, 1990). In an inhomogeneous medium ϵ_{eff} differs for any two charges, because even for sites the same distance apart Ψ_{ij} depends on their positions relative to the molecular surface. Conventional FDPB calculations can determine the effect of the region occupied by water with its high dielectric constant and the additional screening from salt on Ψ_{ij}. The motion of the hydroxyl protons in response to site titration also adds to ϵ_{eff}. The contribution of hydroxyls to ϵ_{eff} between two groups i and j was evaluated by calculating the change of pairwise energy at site j when hydroxyls change their orientation upon ionization of site i. The free energy of interaction between sites i and j is now

$$G_{ij} = \{\Psi_{ij} + (\Psi''_p - \Psi'_p)\}q_j, \tag{6}$$

where Ψ'_p is the potential at j from the hydroxyl protons in their position in the initial charge state of i and Ψ''_p is the potential again at j, given their rearrangement in response to the titration of i. The ratio of the Coulomb interaction and the energy in Eq. 5 is ϵ_{eff}.

RESULTS

Generation of conformers

Lysozyme has 32 titratable residues (11 Arg, six Lys, one His, one Ntr, seven Asp, three Glu, three Tyr, and one Ctr) and 24 mobile polar (MP) groups (10 Ser, seven Thr, and seven buried waters). The goal is to let these groups explore a preselected set of conformers to determine their Boltzmann distribution as a function of pH. As defined in this work, conformers for a given residue differ either by the position of a polar proton or by the residue ionization state. Available conformers include the charged form of each ionizable residue; one neutral form for each Lys, Arg, Ctr and Ntr; and two neutral forms for His. In addition, hydroxyl positions that have been identified as being likely to be at low energy in some ionization or conformation state of the protein are assigned to each Ser, Thr, buried water, and neutral Asp, Glu, and Tyr. One hundred fifty-two conformers were defined (Table 2). Three to six different conformers were generated for each neutral Asp and Glu. For some acids, such as Asp48, each carboxylate oxygen is involved in the same number of conformers. However, this symmetry is easily broken in the complex environment of the protein. For example, Asp87 has four distinct positions for protonation of the buried OD1 and only one position for OD2, which is on the protein's surface. Sampled Tyr conformers have protons near the ring plane, whereas the conformer torsion angles for Ser and Thr are more uniformly distributed. Each buried water samples two to three positions, as suggested by the local electrostatic field and the presence of hydrogen bond donors and acceptors.

The reaction field ($\Delta G_{rxn,i}$), polar ($\Delta G_{pol,i}$), and nonelectrostatic ($\Delta G_{nonel,i}$) energy differences for each conformer are assumed to be independent of the state of any other residue and are listed in Table 2. $\Delta G_{rxn,i}$ is the difference between the reaction field energies of residue i in the protein and in solution. Because the solution reaction field energy (Table 3) is much larger for charged than for neutral groups, $\Delta G_{rxn,i}$ is almost always larger for ionized than for neutral conformers. This loss in reaction field stabilization in the protein that is greater for the ionized than for the neutral

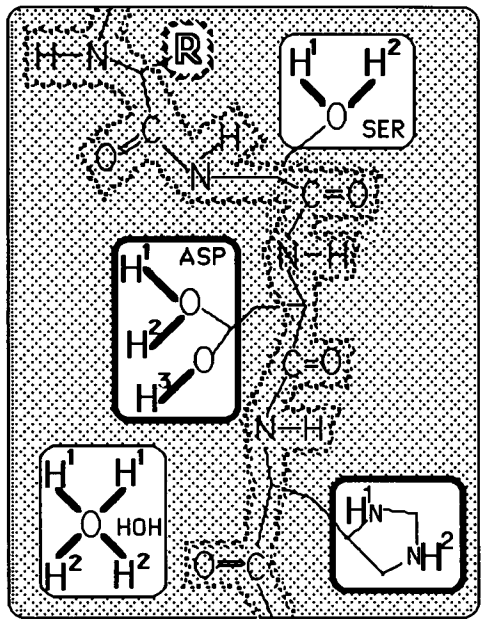

FIGURE 1 A model protein made up of an acid (Asp) with three different positions for the neutral form's proton; a base (His) with two neutral forms; the mobile polar (MP) groups (Ser) with two positions for its hydroxyl and water (HOH) with two orientations. The backbone for each residue and the polar side chain (R) are fixed polar (FP) groups. One state of the protein represents the choice of one proton position for the Ser and HOH. If the Asp is neutral, one proton is selected, and one proton is removed if the His is neutral.

form of a residue shifts the pK_a favoring the neutral form. However, for surface residues, $\Delta G_{rxn,i}$ of both ionized and neutral conformers is small and, depending on the side-chain conformation, can be larger for the neutral form. This is found most often for Lys. The maximum difference between the reaction field energies for the neutral conformers of a given residue is 0.7 ΔpK units. Thus the small changes in hydrogen position have little impact on the interaction with solvent. This supports the assumption that changing one residue's conformation does not effect $\Delta G_{rxn,i}$ for neighboring residues.

$\Delta G_{pol,i}$ sums the conformer pairwise interactions with the amide backbone dipoles and FP side chains such as Asn and Gln. $\Delta G_{pol,i}$ is generally larger for ionized than for neutral conformers of a residue and usually has the opposite sign. For example, $\Delta G_{pol,i}$ destabilizes the ionized Asp101 by 3.8 ΔpK units while stabilizing the neutral form by −2.1 ΔpK units. Polar interactions with neutral conformers of the same residue can vary significantly. For example, $\Delta G_{pol,i}$ of different neutral conformers of Glu7, Asp87, and Asp48 differs by more than 3 ΔpK units.

$\Delta G_{nonel,i}$, which includes the torsion angle, Lennard-Jones energies, as well as the attraction of the nonbonded carboxylate oxygen for the proton in Asp and Glu, modifies the conformer energies of Ser, Thr, and Tyr by less than 2.4 ΔpK units. The difference in nonelectrostatic energy between different neutral Asp and Glu conformers can be as large as 6.0 ΔpK units (Table 2).

TABLE 4 The 11×11 pairwise interaction energy matrix for the model protein shown in Fig. 1

conformer	1	2	3	4	5	6	7	8	9	10	11
	Asp^-	Asp^{H1}	Asp^{H2}	Asp^{H3}	His+	His-ND1	His-NE1	HOH^1	HOH^2	Ser^{H1}	Ser^{H2}
Asp^-	0	0	0	0	$G^{1,5}$	$G^{1,6}$	$G^{1,7}$	$G^{1,8}$	$G^{1,9}$	$G^{1,10}$	$G^{1,11}$
Asp^{H1}	0	0	0	0	$G^{2,5}$	$G^{2,6}$	$G^{2,7}$	$G^{2,8}$	$G^{2,9}$	$G^{2,10}$	$G^{2,11}$
Asp^{H2}	0	0	0	0	$G^{3,5}$	$G^{3,6}$	$G^{3,7}$	$G^{3,8}$	$G^{3,9}$	$G^{3,10}$	$G^{3,1}$
Asp^{H3}	0	0	0	0	$G^{4,5}$	$G^{4,6}$	$G^{4,7}$	$G^{4,8}$	$G^{4,9}$	$G^{4,10}$	$G^{4,11}$
His+	$G^{5,1}$	$G^{5,2}$	$G^{5,3}$	$G^{5,4}$	0	0	0	$G^{5,8}$	$G^{5,9}$	$G^{5,10}$	$G^{5,11}$
His-ND1	$G^{6,1}$	$G^{6,2}$	$G^{6,3}$	$G^{6,4}$	0	0	0	$G^{6,8}$	$G^{6,9}$	$G^{6,10}$	$G^{6,11}$
His-NE1	$G^{7,1}$	$G^{7,2}$	$G^{7,3}$	$G^{7,4}$	0	0	0	$G^{7,8}$	$G^{7,9}$	$G^{7,10}$	$G^{7,11}$
HOH^{H1}	$G^{8,1}$	$G^{8,2}$	$G^{8,3}$	$G^{8,4}$	$G^{8,5}$	$G^{8,6}$		0	0	$G^{8,10}$	$G^{8,11}$
HOH^{H2}	$G^{9,1}$	$G^{9,2}$	$G^{9,3}$	$G^{9,4}$	$G^{9,5}$	$G^{9,6}$	$G^{9,7}$	0	0	$G^{9,10}$	$G^{9,11}$
Ser^{H1}	$G^{10,1}$	$G^{10,2}$	$G^{10,3}$	$G^{10,4}$	$G^{10,5}$	$G^{10,6}$	$G^{10,7}$	$G^{10,8}$	$G^{10,9}$	0	0
Ser^{H2}	$G^{11,1}$	$G^{11,2}$	$G^{11,3}$	$G^{11,4}$	$G^{11,5}$	$G^{11,6}$	$G^{11,7}$	$G^{11,8}$	$G^{11,9}$	0	0
$G_{rxn,i}$	$G_{rxn,1}$	$G_{rxn,2}$	$G_{rxn,3}$	$G_{rxn,4}$	$G_{rxn,5}$	$G_{rxn,6}$	$G_{rxn,7}$	$G_{rxn,8}$	$G_{rxn,9}$	$G_{rxn,10}$	$G_{rxn,11}$
$G_{pol,i}$	$G_{pol,1}$	$G_{pol,2}$	$G_{pol,3}$	$G_{pol,4}$	$G_{pol,5}$	$G_{pol,6}$	$G_{pol,7}$	$G_{pol,8}$	$G_{pol,9}$	$G_{pol,10}$	$G_{pol,11}$
$G_{nonel,i}$	$G_{non,1}$	$G_{non,2}$	$G_{non,3}$	$G_{non,4}$	$G_{non,5}$	$G_{non,6}$	$G_{non,7}$	$G_{non,8}$	$G_{non,9}$	$G_{non,10}$	$G_{non,11}$

Note that pairwise energies between conformers belonging to the same residue are set to zero. The three additional 1×10 vectors for the difference in reaction field energy ($\Delta G_{rxn,i} = \Delta G_{rxn}$ (in protein) − ΔG_{rxn}(sol)), polar $\Delta G_{pol,i}$, and nonelectrostatic interactions ($\Delta G_{nonel,i} = G_{nonel,i}$(protein) − $G_{nonel,i}$(solution)) are required to define a conformer's energy (see Eq. 2). Each DelPhi calculation provides all electrostatic terms in one column of this matrix.

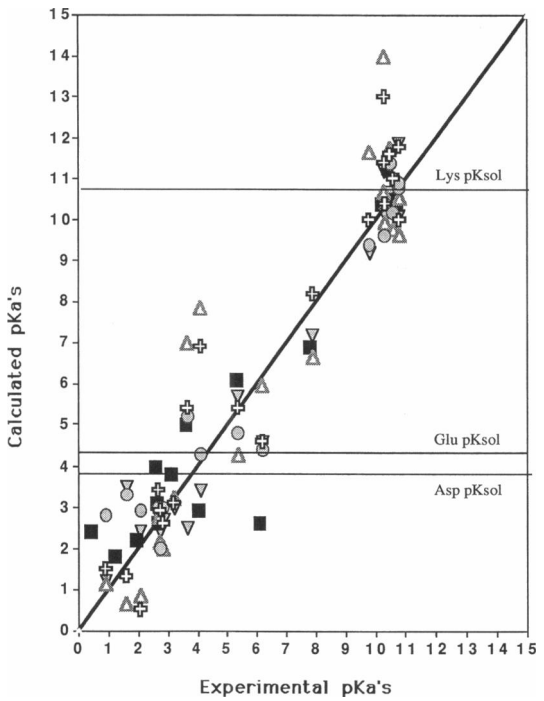

FIGURE 2 Calculated pK_a's plotted against the experimental values. The data for bases are from Karamitsu and Hamaguchi (1980), and those for acids and His are from Bartik et al. (1994). The straight line is the ideal fit. Horizontal lines show solution pK_a values for Asp, Glu, and Lys. See Fig. 4 for the identity of residues with each pK_a. Calculated numbers are from the multiconformation (✣), pK_a^b (Yang and Honig, 1993) (■), SOL20 (Antosiewicz et al., 1995) (▨), POLM0 (You and Bashford, 1995) (△), and CLUS (Gibas and Subramaniam, 1996) (▽) models. Tyr53 is not shown, because either it was not calculated (Yang and Honig, 1993) or its calculated pK_a is outside the graph (multiconformation analysis and You and Bashford, 1995). Tyrosine pK_a's were not reported by Yang and Honig (1993), and Antosiewicz et al. (1995) did not calculate the pK_a's of Asp119 and NTR. Arginines are not shown because there are no available experimental data.

TABLE 5 The number of pairwise interactions greater than 0.43 ΔpK units (1*kT*) for the ionized form of each titratable residue with other groups in the protein

	Number of interactions > 0.43 ΔpK units			
Residue	Bases	Acids	MP groups	FP groups
Arg 5			1	
Arg 14				
Arg 21				
Arg 45	1	2	1	
Arg 61		2		1
Arg 68	1	2		
Arg 73				1
Arg 112				1
Arg 114				
Arg 125				1
Arg 128				
Asp 18	1			1
Asp 48	1	1	1	1
Asp 52	1	4		4
Asp 66	2	2	4	5
Asp 87	1		1	1
Asp 101	1			3
Asp 119				1
Ctr 129	1			
Glu 7				1
Glu 35		1		2
His 15	1	1	1	2
Lys 1	1			
Lys 13		2		
Lys 33				1
Lys 96	1	1		1
Lys 97		1		
Lys 116				
Ntr 1	1		3	4
Tyr 20	1		1	1
Tyr 23				3
Tyr 53	2	2	1	1

The interactions are divided into those with ionized bases, ionized acids, and (neutral) mobile or fixed polar groups.

The importance of adding conformational flexibility depends on the number and strength of the interactions involving the new degrees of freedom (see Table 5). Three of the nine ionized Asp and Glu, no Lys, and two of the 11 Arg interact with a MP conformer by more than 0.43 ΔpK units (1*kT*). None of the three acidic residues in the active site of lysozyme, Asp52, Asp101, and Glu35, interact this strongly with any MP group. Although the groups with more direct connections do show the largest differences when rigid and multiconformation calculations are compared, all groups will be shown to be affected by the presence of additional, available conformations.

Calculation of pK_a values

The pK_a's of the Asp, Glu, His, Lys, Tyr, and chain termini have been measured in lysozyme (Bartik et al., 1994; Kuramitsu and Hamaguchi, 1980). The Boltzmann distributions of ionization states were calculated as a function of pH, using all conformer interaction energies (Eq. 2). The pH where a residue has a 50% probability of being ionized is designated the pK_a. All residue titration curves can be well described by the Henderson-Hasselbach equation, so there is no ambiguity about pK_a values (Beroza et al., 1991).

Fig. 2 compares the pK_a's calculated by the multiconformation method and by previously reported methods (Yang and Honig, 1993; Antosiewicz et al., 1994; You and Bashford, 1995; Gibas and Subramaniam, 1996) with the experimental data. Table 6 provides these values, together with three different sets of experimental pK_a's (Kuramitsu and Hamaguchi, 1980; Ramanadham et al., 1981; Bartik et al., 1994) and the r.m.s. deviation of each set of calculated pK_a's from the experimental data. Lysozyme is a small molecule without strong pairwise interactions, and values calculated by various methods are close to the experimental pK_a's for many residues. However, some pK_a's do differ significantly from the data, particularly those of Asp66 (Antosiewicz et al., 1994) and Asp101 (You and Bashford, 1995). The multiconformation calculations show significant improvement over all rigid protein methods, with the exception of that of Antosiewicz et al. (1994), which uses a

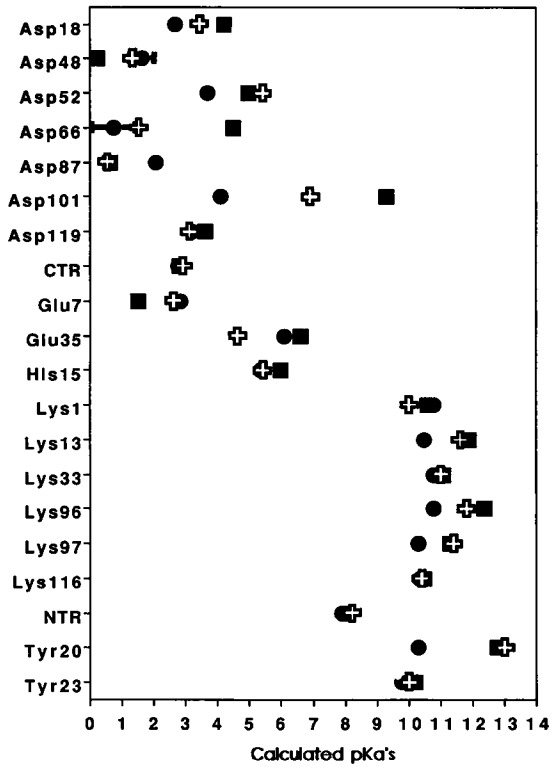

FIGURE 3 pK_a's calculated by the multiconformation (✣) and fixed structure methods (■), and the experimental data (●) (Karamitsu and Hamaguchi, 1980; Bartik et al., 1994) (Tyr53 is not included). Error bars reported for the measurements of Bartik et al. (1994), which are used for acidic pH (Glu, Asp, and His) pK_a's. Data from Kuramistu and Hamaguchi (1980) were used for all other residues. Arg is not shown.

protein dielectric constant of 20. Those calculations can be viewed as incorporating significant, average protein motion with the high dielectric constant. The r.m.s. of calculated pK_a's of Gibas and Subramaniam (1996) is only 0.8, but they used a different crystal structure (2lzt) from the Protein Data Bank, and therefore their results cannot be compared directly.

The multiconfomation method is compared (Fig. 3) with a calculation that fixes one conformer (* in Table 2) for each ionization state. The heavy atom positions, charges, and atomic radii are the same in both calculations. This allows a direct comparison, without the uncertainties in input parameters that are found when values are obtained by different research groups. The multiconformation pK_a's are closer to the data than those of rigid protein values. The calculated values can differ by as much as 3 pH units (Fig. 3). The largest differences are found for the acids, which have a variety of possible neutral conformations, but small changes are found for all residues.

The residues Glu35, Asp52, Asp101, and Tyr53 show significant differences between calculated and experimental pK_a's. Each site has a different source for the discrepancy. Asp52 and Glu35 are in the active site of lysozyme. They are part of the Asp66 cluster (see below). However, this pair of residues can be treated as a separate unit, because at pHs where they titrate, no nearby residues change ionization state. Thus the rest of the protein simply adds a constant shift to their pK_a values. The calculated pK_a for Asp52 is 1 pH unit higher and that for Glu35 1 pH unit lower than the experimental results. The energy of interaction between these residues is 0.8 ΔpK units. Depending on which residue is ionized first (with increasing pH), the other will have its pK_a shifted upward by 0.8 pH units by the charge on the other acid. In the calculations, Glu35 is ionized at lower pHs because the $pK_{int,i}$ of Asp52 is 0.25 pH units lower than the $pK_{int,i}$ of Glu35, but Asp48 interacts 0.6 ΔpK units more strongly with Asp52. If the $pK_{int,i}$ is changed so that Asp52 is ionized at lower pH, the pK_a of Glu35 is increased by 0.8, so both values are now very close to the experimental data (see also You and Bashford, 1995). This example shows how even weak interactions with distant sites can cause much larger changes in the behavior of strongly interacting sites. Thus the calculation of strongly coupled residues can be in error by the magnitude of a site's interactions. Unfortunately, the end is controlled by the accuracy of calculation of the weakest interactions. In these cases only the total charge on the coupled sites can be obtained reliably (Lancaster et al., 1996). If Asp52 and Glu35 were to be considered a unit, the combined ionization of the two sites would closely match the experimental values and would be relatively independent of the interactions with weakly coupled sites (Yang et al., 1993).

The pK_a of Asp101 calculated with the rigid protein method is 5.2 pH units higher than the experimental value and 2.8 pH units too high in the multiconformation calculations. Although Asp101 does not interact strongly with MP groups, the multiconformation method yields a pK_a closer to the experimental data because of the inclusion of $\Delta G_{nonel,i}$. An unfavorable Lennard-Jones repulsion destabilizes the hydroxyl position with the strongest hydrogen bond to Ser100, populating other neutral conformers. Destabilizing the neutral Asp101 lowers the pK_a. Despite the better fit in the multiconformation method, there is clear evidence for larger scale motions on ionization of this residue. In triclinic lysozyme, Asp101 OD1 is 2.29 Å from the carbonyl oxygen of the backbone of Ser100. Therefore there will be significant repulsion between the carboxylate and carbonyl when the Asp is ionized. The pH of crystallization, 4.5, is close to the experimental pK_a of 4.02–4.16 for this Asp. No clear electron density was found for the residue (Hodsdon et al., 1985), suggesting that it may be partially ionized in the crystal and that ionized and neutral forms have different side-chain conformations. Bashford and Karplus (1990) showed that $\Delta G_{pol,\ i}$ for Asp101 is 2 ΔpK units higher in the triclinic (crystallized at pH 6.7–7.3) (Matthews et al., 1973) than in the tetragonal structure. Our calculations comparing triclinic (1LZT) (Hodsdon et al., 1985) and tetragonal (2LZM) (Weaver and Matthews, 1987) lysozyme are in agreement with this observation.

The pK_a of Tyr53 was calculated to be higher than 16, whereas the experimental value is 12.1. This Tyr is hydrogen bonded to the ionized Asp66. Imoto et al. (1972) have shown large-scale conformational changes in the protein on ionization of this Tyr, which Bashford and Karplus (1990)

TABLE 6 Experimental pK_a values

Residue	Experimental values			Calculated values					
	KH	RSJ	BRD	Y&H	AMG	Y&B	G&S	Fixed	Multi
NTR	7.9 ± 0.1	7.9	na	6.9	na	6.64	7.2	8.1	8.2
Glu 7	2.6	2.6	2.60–3.10	3.1	3.0	1.99	2.7	1.5	2.6
Glu 35	6.1	6.1	6.1–6.3	2.6	3.3	5.96	4.6	6.6	4.6
Asp 18	2.0 ± 0.3	2.9	2.58–2.74	4	5.2	2.75	2.8	4.2	3.4
Asp 48	3.4	4.3	1.2–2.0	1.8	2.8	0.61	3.5	0.2	1.3
Asp 52	3.4	3.55	3.60–3.76	5	2.9	6.98	2.5	5.0	5.4
Asp 66	1.6	2.0	0.4–1.4	2.4	4.3	1.12	1.2	4.5	1.5
Asp 87	2.1 ± 0.3	3.65	1.92–2.22	2.2	na	0.84	2.4	0.6	0.5
Asp 101	4.5	4.25	4.02–4.16	2.9	4.3	7.84	3.4	9.3	6.9
Asp 119	2.5 ± 0.3	2.5	3.11–3.29	3.8	na	3.19	3.0	3.6	3.1
Tyr 20	10.3	10.3	na	na	9.6	13.98	10.2	12.8	13
Tyr 23	9.8	9.8	na	na	9.4	11.65	9.2	10.2	10
Tyr 53	12.1	12.1	na	na	11.0	20.66	11.1	>16	>16
His 15	5.8	5.8	5.29–5.43	6.1	4.8	4.27	5.7	6.0	5.4
Lys 1	10.8 ± 0.1	10.6	na	10.4	10.8	9.65	10.1	10.6	10
Lys 13	10.5 ± 0.1	10.3	na	11.3	11.4	11.75	10.7	11.9	11.6
Lys 33	10.6 ± 0.1	10.4	na	10.5	10.2	9.78	11.0	11.1	11
Lys 96	10.8 ± 0.1	10.7	na	10.4	10.9	10.54	11.9	12.4	11.8
Lys 97	10.3 ± 0.1	10.1	na	10.4	10.4	10.7	11.2	11.3	11.4
Lys 116	10.4 ± 0.1	10.2	na	10.4	10.4	9.93	10.1	10.5	10.4
CTR	3.1	2.9	2.63–2.87	2.6	2.0	2.15	2.4	2.8	2.9
r.m.s.	0.70	1.04		1.11	0.91	1.36	0.80	1.70	1.08

KH, Kuramitsu and Hanamaguchi (1980); RST, Ramanadham et al. (1981); BKD, Batrik et al. (1994). Calculated pK_a's: Y&H, pK_a^b (Yang and Honig, 1993); AMG, SOL20 (Antosiewicz et al., 1994); Y&B, POLM0 (You and Bashford, 1995); G&S, CLUS (Gibas and Subramaniam, 1996) models; Fixed, rigid structure in this work; Multi, multiconformation treatment in this work. Y&H use protein heavy atom positions modified by molecular dynamics (pKa's of tyrosines are not reported). AMG use a protein dielectric constant of 20 (pKa's of Asp119 and NTR are not reported). Y&B obtain the intrinsic pK_a and the pairwise interactions with other residues from an energy-weighted distribution of conformer positions. G&S used a different crystal structure (2LZT) for the calculation, and therefore their results cannot be compared directly. The last row contains the standard deviation for these calculations without tyrosine pK_a's. The bottom position in columns of KH and RST shows the r.m.s. deviation of their pK_a's with respect to the available pK_a's of BKD.

proposed results from repulsion between the Tyr^- and Asp^-. Thus the crystal structure of lysozyme is likely to be quite different from the true high pH structure.

Effect of MP groups on the pK_a's: titration of clusters of residues

Groups of residues can be found that have large electrostatic interactions between them. These will be referred to as "clusters" (Lancaster et al., 1996). The conformer occupancies within the clusters are strongly interdependent. Additional residues change their conformer occupancy in response to changes in the clusters but have small interactions with groups in the cluster. These residues will be called "isolated" (Lancaster et al., 1996). Asp52 and Glu35, described above, are an example of the smallest possible cluster. Their pK_a's are mutually interdependent, whereas the order of ionization is determined by small interactions with isolated sites. Two typical clusters with strong interactions between ionizable and MP groups will illustrate how conformer populations are coupled to the titration of the protein.

Asp87 is in a small cluster with His15 and Thr89. The cluster has two II, seven NI, and two MP conformers. The electrostatic interaction energies of ionized and neutral Asp conformers with other cluster members are given in Table 7. This may not seem like a site where multiconformation analysis will provide new information, because there is little difference in the pK_a's calculated by multiconformation and rigid protein methods (Fig. 3). However, following the occupancy of each conformer within the cluster as a function of pH shows the fixed structure analysis provides an incomplete picture (Fig. 4). All three residues have at least two conformers energetically accessible over a wide range of pH. Thus three neutral Asp conformers are found below pH 1, whereas both His conformers are occupied above pH 4. Thr89, the MP group in the cluster, reorients its dipole in response to the ionization of Asp87 and His15. Thr89-H2 is favored below pH 1; H1, 40° away, is more probable between pH 1 and 4.5; and above pH 4.5, H2 again has the highest occupancy.

The cluster with Asp66 as the central residue contains eight other residues, including four Asp, three Arg, one Glu, one Tyr, two Ser, one Thr, and one water, resulting in nine II, 29 NI, and 10 MP conformers (see Table 8, where strong interactions with Asp66 are shown). The pH dependence of the three residues with strongest interactions and the isolated residue HOH31, which responds to cluster titration, is shown in Fig. 6. Below pH 3, three neutral Asp conformers are present in significant numbers. In addition, Tyr53, Thr69, and HOH31 are all distributed in two conformers. When the Asp is ionized the picture changes. The isolated

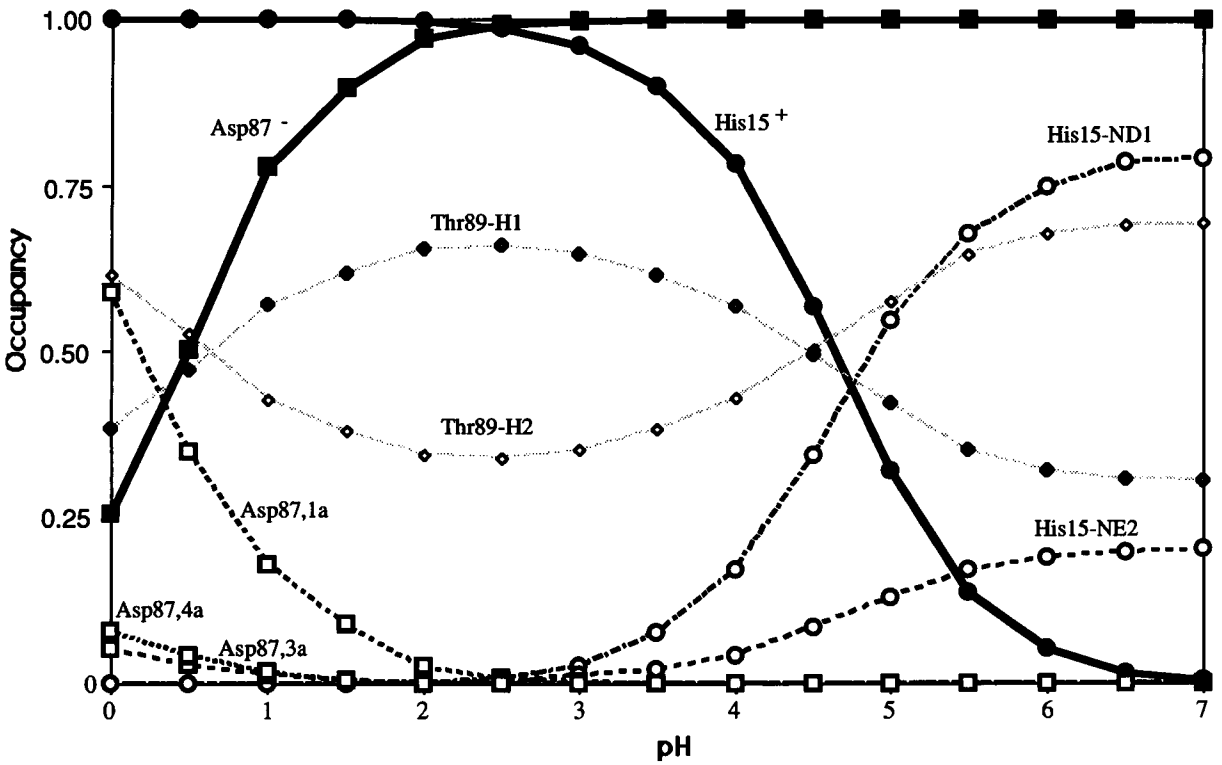

FIGURE 4 Conformer fractional occupancies for the three residues in the Asp87 cluster as a function of pH. See Table 7 for intracluster interaction energies.

water still has more than one conformer populated. However, Thr69 and Tyr53, which are closely coupled to the Asp, are now found in a single conformation. Thus ionization of the Asp lowers the cluster entropy.

The reorientation of the microdipoles of MP groups can have a large impact on the free energy of ionization. For example, the pK_a of Asp66 is much closer to the experimental value in the multiconformation than the rigid protein calculations, largely because of a more accurate treatment of the interaction between this Asp and Thr69. This pair of residues demonstrates how the reorientation of MP dipoles stabilizes titratable sites in each ionization state and makes the free energy difference between neutral and ionized groups smaller than in a rigid structure. The low-energy conformer for the neutral Asp is determined by the Thr69 conformer (Fig. 7). Two states differing by a 120° rotation of the Thr hydroxyl and which Asp oxygen is protonated, have a difference in free energy of 0.32 ΔpK units. The state free energy is even more strongly dependent on the Thr69 hydroxyl position when Asp66 is ionized. One Thr conformer stabilizes Asp^- by −1.61 ΔpK units, whereas the other destabilizes it by 1.83 ΔpK units. This pair of residues highlights the problem with a rigid structure analysis, because the lowest energy for the ionized Asp is found when Thr69 takes the conformation that is at higher energy with the neutral Asp. Depending which Thr conformer is chosen, the calculated pK_a of Asp66 in this two-residue system differs by 3.43 pH units. However, if Thr69 is allowed to change conformation when the Asp changes ionization state, it stabilizes both Asp ionization states. The energy of Asp ionization coupled to Thr dipole rotation now favors the ionized relative to neutral Asp by only −0.61 ΔpK units.

Conformer occupancy

Although calculating pK_a's may be a good test of theory, it is the rare Asp, Glu, Lys, or Arg that is not ionized at pHs where proteins function. His is the only residue that has significant uncertainty in the ionization state. However, the multiconformation calculations provide added information at the pHs where lysozyme will function. The conformer occupancies at pH 5, the pH of maximum activity, are listed in Table 2. No Tyr is ionized, His15 is partially charged, and most of the other acidic and all basic sites are fully ionized. However, it is seen that most of the conformers of the MP groups, neutral His, and Tyr and partially neutral Asp are occupied. The fractional occupancy of many conformers highlights the lack of a unique low-energy structure for the protein. In each low-energy state all conformer occupancies are interdependent.

Estimation of the local dielectric constant (dielectric permeability of protein)

The effective dielectric constant for the pairwise interactions between Asp66, the key residue in the biggest cluster,

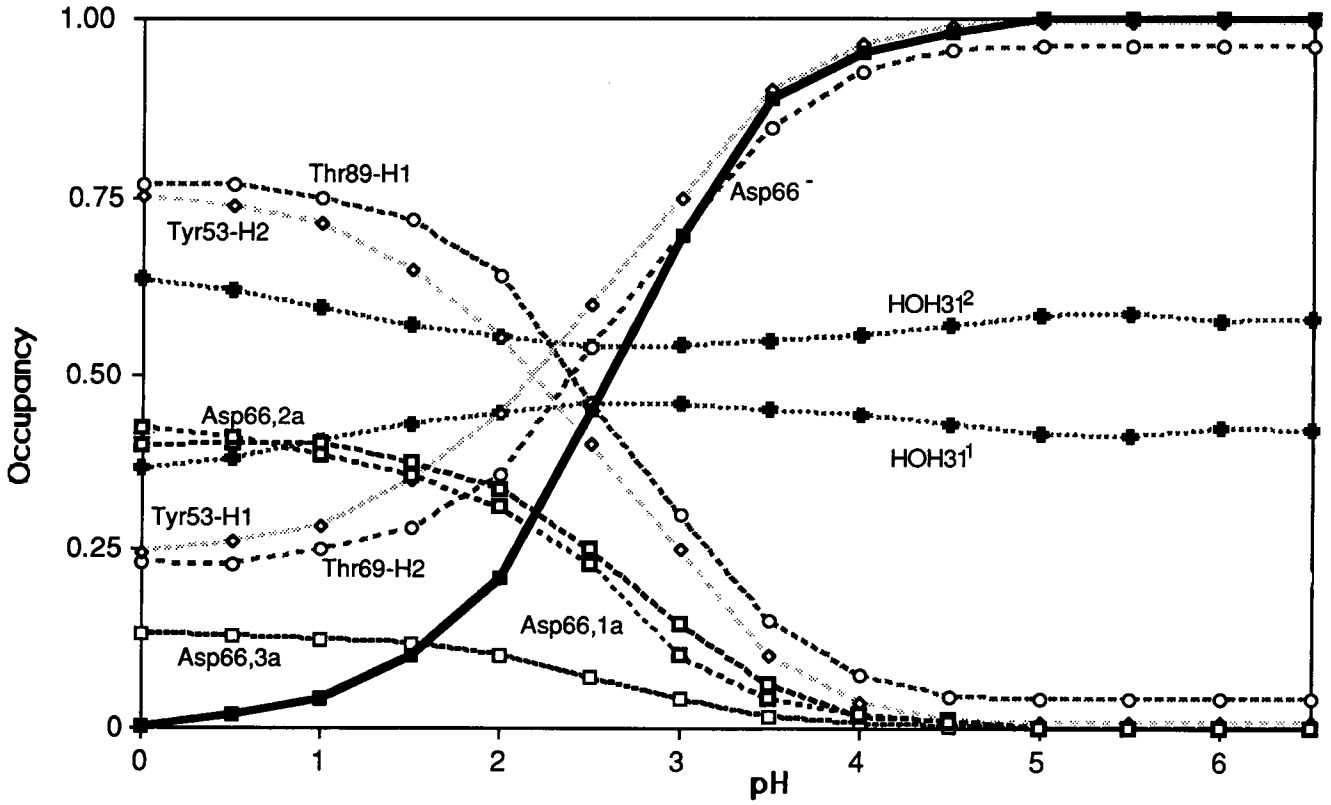

FIGURE 5 Conformer fractional occupancies for Asp66, two of the MP groups in its cluster (Tyr53 and Thr69), and one isolated MP group HOH31, which is weakly coupled to the cluster. See Table 8 for intracluster interaction energies.

and Arg45, Arg68, and Asp52 has been determined (Table 9). The choice of internal dielectric constant, the surrounding water, salt, and MP reorientation all influence ϵ_{eff}. To find the contribution of each factor, pairwise energies were calculated under different conditions. First the interaction energies were obtained with a uniform dielectric constant of 4. The relative values of G_{ij} show that Arg68 is closest to the Asp, whereas Asp52 is furthest away. The addition of salt to the protein exterior screens the interactions between the charges, raising ϵ_{eff} by about twofold. Including the surrounding water with its dielectric constant of 80 further reduces the interaction energy. The screening by salt and then by water depends on the molecular shape and the distance to the surface. Thus it can be seen in the continuum electrostatics treatment, even though the interior dielectric constant is 4, the effective dielectric constant for pairwise interactions is often much larger. Increasing the internal dielectric constant by 5 to $\epsilon = 20$ only increases the ϵ_{eff} by twofold. The modest impact of the value for ϵ_{in} will be found for sites where the screening by the surrounding media has a large impact on ϵ_{eff}. The pairwise interactions modified by hydroxyl reorientation are obtained with Eq. 5. The energy of interaction is reduced by 20–40%. The impacts on the free energy of interactions are different for the three sites. The distance and orientation of the changing dipole as well as its position relative to the surface all determine its contribution to ϵ_{eff}. More distant MP groups have been found to have a negligible impact.

DISCUSSION

What is needed to make accurate pK_a calculations?

Residue pK_a's measure the free energy of site ionization in the protein. Previous calculations have generally been limited by the assumption that site ionization is determined only by electrostatic interactions in a rigid protein. Methods relying on this model have been able to reproduce many but not all experimental findings. Several attempts have been made to improve the method to provide more accurate calculated values (see reviews by Gilson, 1995; Honig and Nicholls, 1995). The calculations presented here do generally yield a closer match between the experimental and calculated pK_a's for three reasons:

1. The inclusion of conformational flexibility provides low-energy configurations in each ionization state of the protein, reducing the free energy difference between neutral and ionized sites (e.g., see Fig. 6). For example, reorientation of four MP groups shifts the pK_a of Asp66 by 3 pH units. It is noteworthy that allowing only polar proton flexibility significantly changes the outcome of the calculations. The improvement in the match between experiment and calculation suggests that polar protons on hydroxyls and water do play an important role in the response to cnarge changes at a significant number of sites in this protein (see Table 5). There are, of course, residues such as Asp101 and

TABLE 7 Pairwise interaction energies between Asp87 conformers and the residues within the Asp87 cluster

	Asp87$^-$	Asp87, 1a	Asp87, 1b	Asp87, 2a	Asp87, 3a	Asp87, 4a
His15+	**−0.95**	0.04	−0.09	−0.15	0.09	0.04
His15-ND1	0.12	−0.03	−0.00	0.03	−0.02	0.01
His15-NE1	−0.12	0.04	−0.01	−0.01	0.03	−0.03
Thr89-H1	**−0.56**	**1.66**	0.08	0.32	**0.65**	**−0.51**
Thr89-H2	**−0.84**	**−0.59**	−0.15	**−0.66**	−0.25	0.27

Interactions ≥ 0.43 ΔpK units are shown in bold.

Tyr53, where ionization appears to initiate motions of groups other than hydroxyls. The limitations of the method presented here are seen in the poor fit of the data to experiment for these sites. A more general application of the multiconformation methodology will be required to accurately calculate ionization state changes for these groups.

2. Nonelectrostatic Lennard-Jones and torsion potentials are included in the state energies. The most important is the Lennard-Jones repulsion, which is a necessary correction when close contact between proton and hydrogen bond acceptor yields a very favorable electrostatic interaction. This term is responsible for the better fit for Asp101, changing the pK_a by 2.4 pH units. Most conventional protocols for electrostatic calculations fix the atomic position of hydroxyls. These positions are usually obtained by energy minimization, which accounts for the nonelectrostatic and electrostatic energies. But conventional pK_a calculations omit the nonelectrostatic terms in determining the state energy. This is appropriate only if this term is close to zero or is the same for both ionization states of each residue. If the difference between ionized and neutral residues is the same for all residues of a given type, a single correction of solution pK_a could be added. However, Table 2 shows that in lysozyme the nonelectrostatic energies are different from zero and they differ for each residue. These terms become essential in the multiconformation method, because the differences in nonelectrostatic and electrostatic energies of different conformers are often of similar magnitude.

3. Neutral acids are treated more accurately by including an explicit proton on one carboxylate. However, this term on its own does not greatly increase the accuracy of the calculations, as can be seen in the rigid protein calculations (Fig. 3 and Table 6). Part of the problem may be that a single, localized proton is not a good model for the neutral

TABLE 8 Pairwise interaction energies between Asp66 conformers and the residues within the Asp66 cluster

	Asp66	Asp66,1a	Asp66,1b	Asp66,2a	Asp66,2b	Asp66,3a
Arg45+	**-0.87**	0.17	0.13	0.17	0.04	-0.13
Arg45^0	-0.26	-0.04	0.00	-0.09	0.00	-0.09
Trh51-H1	0.17	0.09	-0.09	0.09	0.00	0.04
Thr51-H2	**-0.61**	-0.09	-0.04	-0.13	0.09	-0.09
Asp52$^-$	**0.52**	0.03	0.06	0.03	-0.02	0.04
Asp52,1a	-0.07	-0.00	-0.01	-0.00	0.00	-0.01
Asp52,1b	0.12	0.00	0.02	0.00	-0.00	0.01
Asp52,2b	0.04	-0.00	0.01	-0.00	-0.00	0.00
Tyr53	**9.47**	**2.74**	**0.34**	**2.86**	**1.09**	**2.74**
Tyr53-H1	**-0.87**	**-0.47**	-0.04	**-0.47**	0.04	**0.43**
Tyr53-H2	**2.56**	**0.90**	0.26	**0.95**	**0.47**	**0.91**
Ser60-H1	**-1.22**	0.19	-0.33	0.09	0.23	0.07
Ser60-H2	**-0.61**	0.37	-0.20	0.36	-0.04	0.16
Arg68+	**-1.01**	-0.16	0.17	-0.22	0.03	-0.06
Arg68^0	-0.39	-0.06	0.08	-0.09	0.01	-0.01
Thr69-H1	**1.83**	**-1.00**	**0.69**	**-0.74**	0.09	**-0.43**
Thr69-H2	**-1.61**	**1.95**	**-0.68**	**1.38**	**-0.61**	**0.76**
HOH1	**-1.03**	0.23	**-0.45**	**0.52**	-0.28	-0.06

Interactions ≥0.43 ΔpK units are shown in bold. The energies of the conformers with occupancy shown in Fig. 7 are shaded.

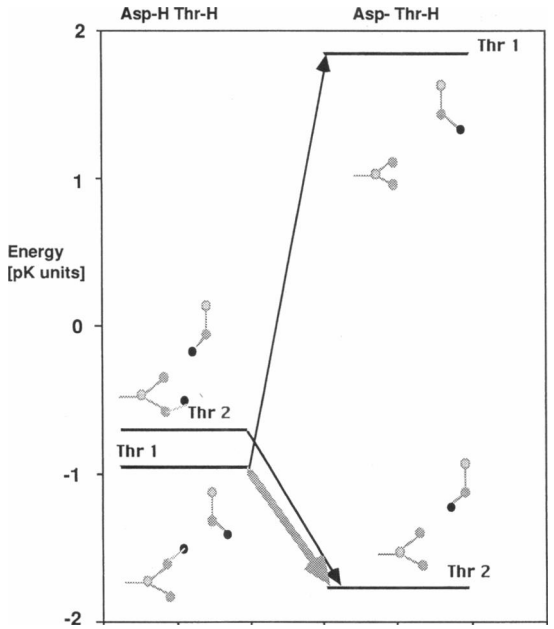

FIGURE 6 The pairwise interaction energy (ΔpK units) between Asp66 and Thr69. The interaction energies with all other residues are set to zero. The conformation of Thr that is at the lowest energy with the neutral Asp and with the ionized Asp is shown. In the sketch the hydroxyl positions are illustrated with filled black circles on the lighter drawing of the heavy atom side-chain positions. The two lines in the left part of the figure show the lowest interaction energy between neutral Asp and Thr (conformations Thr1 and Thr 2). The lines on the right side of the figure give the interaction energy between the ionized Asp and Thr in the same two conformations (Thr 1 and Thr 2). Black arrows show the change in free energy on ionization of Asp with the hydroxyl position of Thr fixed. The gray arrow shows the energy change when the Tyr is allowed to rotate to find its optimum position in each ionization state of the Asp.

conformation. The multiconformation method demonstrates that more than one conformer is often energetically accessible. Distributing the loss of charge over the whole carboxylate may mimic the distribution of proton positions found at many sites better then choosing one neutral conformation.

In this paper, experimental values for the pK_a's were taken from two sources. Bartik et al. (1994) determined the pK_a's of the acids from the transition of individually assigned NMR resonances. Kuramitsu and Hamaguchi (1980) provided values for the bases from measurements of the pH titration of the entire protein. The latter technique is less sensitive. However, a comparison of the pK_a's for the acids obtained by both techniques shows that the potentiometric measurements are within 0.5 pH units of the NMR measurements for 8 of the 10 Asp and Glu.

Generally, the outcome of pK_a calculations depends on many parameters: charge distribution (Oberoi and Allewell, 1993), atomic radii (Sitkoff et al., 1994), dielectric constant for protein (Antosiewicz et al., 1994, 1996) and solution, accuracy of the method for obtaining protein surface and for solving the Poisson-Boltzmann equation, impact of the water molecules (Gibas and Subramaniam, 1996), the crystal structure used, method for generating hydrogen positions, and finally, the treatment of conformation changes. Some of these represent technical problems, whereas others address issues relevant to protein function. Hydroxyl flexibility is shown here to make a significant contribution to the calculated pK_a's. Unfortunately, with so many parameters modified independently by many authors, it is very difficult to determine which parameters are most important.

The choice of possible motions

In the work presented here the preselected conformers are limited to a maximum of six per Glu and Asp; three for Thr, Ser, Tyr, HOH; and two for His. In addition, no heavy atoms are flexible. Although this is a highly restricted subset of possible conformers, it does provide substantial modification of the pK_a's in lysozyme, improving calculated values (see Fig. 3 and Table 6). Rotation of a hydroxyl moves a polar proton by 1.54 Å, requiring little motion of the rest of the protein. In contrast, similar displacements of side chain or backbone dipolar atoms require the coupled motion of several atoms. In the close-packed protein interior, displacements of 1–2 Å may require the movement of several groups. The protein surface is the one region where large side-chain motions can occur easily. However, because surface groups are well solvated, they have little impact on the ionization or conformer state of other residues. Thus focusing on proton flexibility provides a significant improvement in the calculations with a minimum increase in conformers.

Restricting protein motions to proton motions has the added advantage of making the calculations significantly simpler. The analysis assumes that the energy of every state is a linear function of individual conformer energies, which include the self (reaction field and nonelectrostatic) and pairwise energies and no higher terms. This assumption may not always hold if larger motions are included. For example, if side chain A buries site B, the reaction field, nonelectrostatic, and pairwise interaction energies for the n_b conformers of B must now be calculated for each of A's n_a conformers, requiring $n_b^{n_a}$ calculations. In contrast, in the method presented here, only one DelPhi calculation is required per conformer. The observation that the reaction field energy of different conformers of the same residue differs by less than 0.7 ΔpK units (see Table 2) suggests that the change in proton positions has a negligible influence on the protein-solvent boundary.

What is the "dielectric constant" of a protein?

A discussion of the dielectric constant of a protein must consider two distinct but interrelated values: ϵ_{in}, the dielectric response of the protein itself; and ϵ_{eff}, which describes how the interactions between charges are screened by the combined protein and solvent response. The experimental values of ϵ_{in}, estimated with dried protein powders, are between 2 and 4 (Takashima and Schwan, 1965; see Har-

TABLE 9 Pairwise energy of interaction between Asp66 and three other residues within the Asp cluster calculated using different electrostatic parameters

Residue interacting with $Asp66^-$	$\varepsilon_{in} = 4$, $\varepsilon_{out} = 4$, $I = 0$	$\varepsilon_{in} = 4$, $\varepsilon_{out} = 4$, $I = 0.15$	$\varepsilon_{in} = 4$, $\varepsilon_{out} = 80$, $I = 0.15$	$\varepsilon_{in} = 4$, $\varepsilon_{out} = 80$, $I = 0.15$, flexible hydrogens	$\varepsilon_{in} = 20$, $\varepsilon_{out} = 80$, $I = 0.15$
Arg^{45+}	−8.6 (4.0)	−5.4 (6.4)	−0.9 (38.4)	−0.7 (53.4)	−0.4 (77.7)
Asp^{52-}	5.9 (4.0)	3.1 (7.7)	0.5 (45.4)	0.3 (72.5)	0.23 (100.7)
Arg^{68+}	−9.4 (4.0)	−6.1 (6.1)	−1.0 (36.0)	−0.74 (51.6)	−0.49 (76.8)

The effective dielectric constant is given in parentheses.
*The contribution of Tyr53, Thr69, Ser60, and HOH1 to the interaction energy. The effective dielectric constant is calculated using Eq. 5.

vey, 1989, for a review). In contrast, the high dielectric constant of 20–30 measured for proteins in solution (Pethig, 1979; Bone and Pethig, 1982) really represents ϵ_{eff}. A similar difference between "wet" and "dry" values for the dielectric constant has been found in molecular dynamics simulations (Simonson et al., 1991; Smith et al., 1993). As can be seen in Table 9, a protein with a low, uniform dielectric constant of 4 surrounded by water with a dielectric constant of 80 often results in pairwise interaction energies that would be the same if the interaction had occurred in a homogeneous medium with an ϵ_{eff} greater than 20. The combination of ϵ_{in}, ϵ_{out}, and the shape of the dielectric boundary determines ϵ_{eff}.

The value of ϵ_{in} has been an important question in electrostatics calculations. $\epsilon_{in} = 2$ is a well-accepted minimum value, because this represents the screening by the polarizable electrons in any condensed medium (Bottcher, 1993). Higher values allow static calculations to account for an average atomic dipole rearrangement that will also screen charges. The value that is used thus expresses the importance of these unknown motions. Gunner and Honig (1992) compared calculated and experimental values for the electrochemical titration of the four hemes in the cytochrome subunit of *Rhodospeudomonas viridis* photosynthetic reaction centers with a ϵ_{in} of 2, 4, and 6. The best values were obtained with an ϵ_{in} of 4. The reaction center is a system in which the protein greatly modified the behavior of the hemes. In contrast, side chain pK_a's are generally close to their solution values (Antosiewicz et al., 1994). Antosiewicz et al. (1994), compared the pK_a's of several proteins calculated with an ϵ_{in} of 2, 4, and 20. The best fit was found with 20. Of course, a high ϵ_{in} makes it easier to reproduce intraprotein pK_a's that are unchanged from solution, because it yields smaller losses of reaction field energy and weaker interactions between sites in the protein.

Antosiewicz et al. (1994) and others (You and Bashford, 1995; Honig and Nicholls, 1995) are in agreement that 20 is an unrealistically high value of ϵ_{in} for protein as a whole. The high apparent value suggests that specific conformational changes near ionizable residues can generate a highly effective, local dielectric response. This can be approximated by a uniform, high internal dielectric constant. However, the simple approach provides no microscopic information about the source of the motion. In contrast, the multiconformation method retains the distinction between regions of the protein where ionization produces large changes in the protein (like Asp66) and sites that generate little dielectric response (like Asp18). In addition, the dipoles and charges that move can be identified. Thus it is seen that hydroxyls reorient in such a way that they stabilize different charge distributions (Fig. 6) and decrease the pairwise interaction energies (Table 9). These motions can be called a dielectric response, and they contribute to the increased ϵ_{eff} in the calculations reported here.

The distribution of low-energy structures?

Proteins are not rigid structures. Thermal energy allows for a wide array of low-energy motions (Case, 1994) and makes possible the occupancy of many conformational substates with similar state energies (McDowell and Kossiakoff, 1995). These internal protein motions have been studied theoretically (Gibrat and Go, 1990) and are demonstrated by a variety of experimental techniques (Krimm et al., 1986; Callendar and Deng, 1994). Multiple low-energy conformations are seen in highly resolved protein structures (Holmes and Stenkamp, 1991). Changes such as site ionization will modify the state energy and therefore the distribution of different conformers. The challenge is to follow protein motions and ionization changes computationally.

Multiconformation Monte Carlo calculations show that the movable polar (MP) groups in hen egg lysozyme are not found in a single conformation, even in the pH range where a majority of titratable groups are fully ionized (pH 5–9) (see Table 2, occupancy at pH 5). Thus hydroxyl positions are not unique, even at fixed pH, in agreement with experimental neutron density profiles (McDowell and Kossiakoff, 1995). The fact that hydrogen can occupy several different positions at fixed pH means that there are a variety of states with similar energy. The Monte Carlo calculations show that the energy of the system has a broad and quite flat minimum energy well. There are many low-energy states with energies that differ by less than 0.1 kcal/mol at the bottom of the energy well. Residues where several conformers have nonzero occupancy can be found for isolated sites as well as for residues in clusters. In isolated sites (such as Asp101) the conformers have similar energies because there are few interactions. In other cases, such as the Asp66 cluster (see Fig. 5), the choice of conformers is interdepen-

dent. Considering an ensemble of lysozyme molecules (at pH < 2), 25% will have Tyr53 in conformation H1 and Thr69 in conformation H2, whereas 75% of the molecules will have Tyr53-H2 and Thr69-H1. Almost no protein will be found with Tyr53-H1 and Thr69-H1. The behavior of this cluster also shows how site ionization can dramatically change the conformer distribution. Thus, when Asp66 is neutral, the other residues in the cluster are found in many conformations. However, when the Asp is ionized, the coupled residues are much more localized. This represents a reduction in the entropy in this part of the protein, suggesting that the protein becomes more rigid. Thus the effect of introducing a new charge increases the number of strong interactions between groups and restricts the effective degrees of freedom by reducing the number of states at similar energy.

New directions and possible improvements

Monte Carlo analysis can give only information about equilibrium populations, telling us nothing about the kinetics of conformational change. However, the analysis presented here can provide some insight into reasonable pathways for proton transfer. First, the presence of multiple occupied conformers under particular conditions identifies sites that exist in a broad energy minimum. In contrast, situations occur when the proton is localized in a steep energy minimum (see Fig. 5 at pH > 3). Proton rearrangement should be easier in sites with more accessible states. Study of the protons that move collectively can also provide clues to pathways of proton transfer, especially if these groups form a hydrogen-bonded network. The changes in conformation when a charge is placed on a titratable residue or cofactor can suggest pathways for proton equilibration within the protein and with the solvent.

The present work allows the calculation of the dielectric response in different regions of the protein to specific changes in ionization states. The calculated group reorientation can explain the experimental observation that the effect of mutation of an ionizable group is often smaller than predicted by conventional electrostatic calculations (McGrath et al., 1992; Antosiewicz et al., 1995). Multiconformation Monte Carlo simulations show that the removal of the electrostatic field of a charged group will be partially compensated by the changing field produced by the motion of neighboring hydroxyls. Thus the effective pairwise interactions with Asp66 presented in Table 9 provide one set of predictions for testing this method.

The multiconformation calculations require an initial choice of possible conformations for each movable site. Introducing more conformers per residue would allow a smoother response of the protein to changes in ionization states. However, the approximation of the entropy correction used in Eq. 2 will introduce larger errors as the number of conformers increases. As described in Materials and Methods, the correct calculation of the entropy term requires knowledge of the distribution of conformer occupancies. Thus an iterative procedure would be needed to determine the Boltzmann distribution of states and conformer occupancies.

Several more degrees of freedom can be added with no modification of the method used here. Examples include specifying the position of deprotonation of neutral Arg, which was ignored here because there is no available experimental data for these groups; determining the position of the N and O in the side chains of Asn and Gln; and dynamically altering the occupancy of ion and water binding sites. A multiconformation method can also be extended to large-scale motions. However, the state energies will no longer be a linear combination of self energy of conformers and their pairwise interaction energies, but second-order terms would have to be included. The challenge will be to maintain a reasonable number of dependent conformations that are needed to define the state energies. This may be achieved by assuming that residue conformations effect only their neighbors.

Note added in proof. Recent calculations of Scheraga et al. (1996. *J. Mol. Biol.* 264:770–778.) have explored the coupling between molecular titration and conformation in a 17-residue polypeptide.

The authors thank Dr. Mike Gilson for making manuscripts available before publication.

This work was supported by the International Human Frontier Science Program (RG-329/95) and the National Institutes of Health (G.M. 48726).

REFERENCES

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

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pK_a's in proteins. *Biochemistry.* 35:7819–7831.

Antosiewicz, J., J. A. McCammon, S. T. Wlodek, and M. K. Gilson. 1995. Simulation of charge-mutant acetylinesterases. *Biochemistry.* 34: 4211–4219.

Baker, E. N., and R. E. Hubbard. 1984. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* 44:97–179.

Baker, P. D., and A. G. Mauk. 1992. pH-Linked conformational regulation of a metalloprotein oxidation-reduction equilibrium: electrochemical analysis of the alkaline form of cytochrome *c*. *J. Am. Chem. Soc.* 114:3619–3624.

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

Bashford, D., and M. Karplus. 1990. The 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, M. D. Brice, J. R. Rodgers, O. Kennard, T. F. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Beroza, P., D. R. Fredkin, M. Y. Okamura, and G. Feher. 1991. Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. USA.* 88:5804–5808.

Bone, S., and R. Pethig. 1982. Dielectric studies of the binding of water to lysozyme. *J. Mol. Biol.* 157:571–575.

Bottcher, C. J. F. 1993. Theory of Electric Polarization. Elsevier, New York.

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. Comput. 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.

Callendar, R., and H. Deng. 1994. Nonresonance Raman difference spectroscopy: a general probe of protein structure, ligand binding, enzymatic catalysis, and the structures of other biomacromolecules. *Annu. Rev. Biophys. Biomol. Struct.* 23:215–214.

Case, D. 1994. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* 4:285–290.

Davis, M. E., and J. A. McCammon. 1990. Solving the finite difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. *J. Comp. Chem.* 10:386.

Gibas, C. J., and S. Subramaniam. 1996. Explicit solvent models in protein pK_a calculations. *Biophys. J.* 71:138–147.

Gibrat, J. F., and N. Go. 1990. Normal mode analysis. *Proteins.* 8:258.

Gilson, M. K. 1993. Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins.* 15:266–282.

Gilson, M. 1995. Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* 5:216–223.

Gilson, M. K., and B. H. Honig. 1986. The dielectric constant of a folded protein. *Biopolymers.* 25:2097–2119.

Gilson, M. K., and B. H. Honig. 1987. Calculation of electrostatic potentials in an enzyme active site. *Nature.* 330:84–86.

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

Gilson, M. K., A. Rashin, R. Fine, and B. Honig. 1985. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol.* 183:503–516.

Gu, Z. 1996. Hydrogen bonding in biological systems: what solid state NMR can tell. Ph.D. thesis. Columbia University, New York.

Gunner, M. R., and B. Honig. 1992. Electrostatic control of midpoint potentials in the cytochrome subunit of the *Rhodopseudomonas viridis* reaction center. *Proc. Natl. Acad. Sci. USA.* 88:9151–9155.

Gunner, M. R., A. Nicholls, and B. Honig. 1996. Electrostatic potentials in *Rhopseudomonas viridis* reaction center: implications for the driving force and directionality of electron transfer. *J. Phys. Chem.* 100: 4277–4291.

Harvey, S. 1989. Treatment of electrostatic effects in macromolecular modeling. *Proteins.* 5:78–92.

Hodsdon, J., G. Brown, and L. Jensen. 1985. Refinement of Triclinic Lysozyme. Brookhaven Protein Bank, file 1LZT.

Holmes, M. A., and R. E. Stenkamp. 1991. The structures of Met and azidomet hemerythrin at 1.66 angstroms resolution. *J. Mol. Biol.* 220: 723.

Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

Honig, B., K. Sharp, and A.-S. Yang. 1993. Macroscopic models of aqueous solutions: biological and chemical applications. *J. Phys. Chem.* 97:1101–1109.

Hush, N. S., and J. R. Reimers. 1995. Vibrational stark spectroscopy. I. Basic theory an application to the CO stretch. *J. Phys. Chem.* 99: 15798–15805.

Kossiakoff, A. A., J. Shpungin, and M. D. Sintchak. 1990. Hydroxyl hydrogen conformations in trypsin determined by the neutron diffraction solvent difference map method: relative importance of steric and electrostatic factors in defining hydrogen-bonding geometries. *Proc. Natl. Acad. Sci. USA.* 87:4468–4472.

Kossiakoff, A. A., M. Ultsch, S. White, and C. Eigenbrot. 1991. Neutron structure of subtilisin BPN′: effects of chemical environment on hydrogen-bonding geometries and the pattern of hydrogen-deuterium exchange in secondary structure elements. *Biochemistry.* 30:1211–1221.

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

Lancaster, C. R. D., H. Michel, B. Honig, and M. R. Gunner. 1996. The coupling of light-induced electron transfer and proton uptake: electrostatic calculations on the photosynthetic reaction center of *Rhodopseudomonas viridis. Biophys. J.* 70:2469–2492.

Matthew, J. B. 1985. Electrostatic effects in proteins. *Annu. Rev. Biophys. Chem.* 14:387–417.

Matthews, B. W., G. W. Danlquist, and A. Y. Maynard. 1973. Crystallographics data for lysozyme from bacteriophage T4. *J. Mol. Biol.* 78: 575–576.

McDowell, R., and A. Kossiakoff. 1995. A comparison of neutron diffraction and molecular dynamics structures: hydroxyl group and water molecule orientation in trypsin. *J. Mol. Biol.* 250:553–570.

McGrath, M. E., J. R. Vasquez, C. S. Craik, A. S. Yang, B. Honig, and R. J. Fletterich, 1992. Perturbing the planar environment of Asp102 in trypsin: consequence of replacing conserved Ser214. *Biochemistry.* 31: 3059–3064.

McQuarrie, D. M. 1973. Statistical Mechanics. Harper and Row, New York.

Nicholls, A., and B. Honig. 1991. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* 12:435–445.

Oberoi, H., and N. Allewell. 1993. Multigrid solution of the nonlinear Poisson-Boltzmann equation and calculation of titration curves. *Biophys. J.* 65:48–55.

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

Pethig, R. 1979. Dielectric and Electronic Properties of Biological Materials. John Wiley, New York.

Raghavan, V., D. G. Lambright, and Boxer, S. G. 1989. Electrostatic interactions in wild-type and mutant recombinant human myoglobins. *Biochemistry.* 28:3771–3781.

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

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

Simonson, T., D. Perahia, and A. T. Brunger. 1991. Microscopic theory of the dielectric properties of proteins. *Biophys. J.* 59:670–690.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Smith, P. E., R. M. Brunne, A. E. Mark, and W. F. van Gunsteren. 1993. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. *J. Phys. Chem.* 97:2009–2014.

Takashima, S., and H. P. Schwan. 1965. Dielectric dispersion of crystalline powders of amino acids, peptides, and proteins. *J. Phys. Chem.* 69: 4176–4182.

Turano, P., J. Ferrer, M. Cheesman, A. Thomson, L. Banci, I. Bertini, and A. Mauk. 1995. pH, electrolyte, and substrate-linked variation in active site structure of the Trp51 Ala variant of cytochrome c peroxidase. *Biochemistry.* 34:13895–13905.

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

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

Weaver, H. L., and B. W. Matthews. 1987. Structure of bacterrophage T4 lysozyme refined at 1.7 angstroms resolution. *J. Mol. Biol.* 193:189.

Yang, A.-S., and H. Barry. 1994. Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.* 273:602–614.

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

Yang, A.-S., and B. Honig. 1992. Electrostatic effects on protein stability. *Curr. Opin. Struct. Biol.* 2:40–45.

Yang, A., and B. Honig. 1993. On the pH dependence of protein stability. *J. Mol. Biol.* 231:459–474.

Yang, A., and B. Honig. 1995. Free energy determinants of secondary structure formation. I. α-Helices. *J. Mol. Biol.* 252:351–365.

You, T. J., and D. Bashford. 1995. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. *Biophys. J.* 69:1721–1733.

You, T., and S. Harvey. 1993. Finite element approach to the electrostatics of macromolecules with arbitrary geometries. *J. Comp. Chem.* 14: 484–501.

Zauhar, R., and R. J. Morgan. 1985. A new method for computing the macromolecular electric potential. *J. Mol. Biol.* 185:815–820.