Combining Conformational Flexibility and Continuum Electrostatics for Calculating pK_as in Proteins

Roxana E. Georgescu, Emil G. Alexov, and Marilyn R. Gunner
Department of Physics, City College of New York, New York 10031 USA

ABSTRACT Protein stability and function relies on residues being in their appropriate ionization states at physiological pH. In situ residue pK_as also provides a sensitive measure of the local protein environment. Multiconformation continuum electrostatics (MCCE) combines continuum electrostatics and molecular mechanics force fields in Monte Carlo sampling to simultaneously calculate side chain ionization and conformation. The response of protein to charges is incorporated both in the protein dielectric constant (ε_{prot}) of four and by explicit conformational changes. The pK_a of 166 residues in 12 proteins was determined. The root mean square error is 0.83 pH units, and >90% have errors of <1 pH units whereas only 3% have errors >2 pH units. Similar results are found with crystal and solution structures, showing that the method's explicit conformational sampling reduces sensitivity to the initial structure. The outcome also changes little with protein dielectric constant (ε_{prot} 4–20). Multiconformation continuum electrostatics titrations show coupling of conformational flexibility and changes in ionization state. Examples are provided where ionizable side chain position (protein G), Asn orientation (lysozyme), His tautomer distribution (RNase A), and phosphate ion binding (RNase A and H) change with pH. Disallowing these motions changes the calculated pK_a.

INTRODUCTION

Protein stability and function depend on interactions among charged and dipolar groups. All proteins have ionized residues, many bind charged substrates, and changes in charge distribution often occur in the active site with reaction. Ionization state changes determine the pH dependence of protein function (Mellor et al., 1993; Dyson et al., 1997; Alexov et al., 2000) and stability (Hendsch et al., 1996; Spassov et al., 1997; Elcock and McCammon, 1998). Proton transfers between buried residues generate transmembrane proton gradients (Ferguson-Miller and Babcock, 1996; Alexov and Gunner, 1999; Luecke et al., 1999; Sham et al., 1999). Surface-charged residue distribution is critical for protein-protein association (Sheinerman et al., 2000). Computational tools that improve estimates of the energetics of the charge distribution within proteins will thus help us understand how protein function is derived from structure.

The analysis of protein charges and dipoles starts with water as a reference solvent (Warshel and Russell, 1984; Gilson and Honig, 1988; Gunner and Alexov, 2000). Water dipoles, being both polar and polarizable, reorient when a charge is added. Ions and dipoles have favorable interactions with water with its high dielectric constant, which are diminished in protein. In contrast, proteins are highly polar but not very polarizable molecules. Each backbone amide has a dipole greater than water, and more than 50% of all side chains are polar or ionizable. However, proteins have limited, nonuniform flexibility producing a heterogeneous response to charges.

The equilibrium distribution of side chain ionization states and position represents the balance of many interactions and so is hard to calculate. Methods to calculate the free energies of protein charge distributions are often tested by their ability to reproduce experimental pK_as (Bashford and Karplus, 1990; Beroza et al., 1991; Takahashi et al., 1992; Yang and Honig, 1993; Antosiewicz et al., 1994, 1996; Demchuk and Wade, 1996; Alexov and Gunner, 1997; Sham et al., 1997; Havranek and Harbury, 1999; Mehler and Guarnieri, 1999; Nielsen and Vriend, 2001). These benchmark calculations are not ideal because they include many values for surface residues, which have little interaction with protein. Also, many pK_as are not at pHs where the protein is functional or the structure determined, so calculations must also model pH-dependent structural perturbations. However, with more than 100 measured pK_as, calculations can try to match many values with one parameter set and methodology. As reliable computational methods become available, proteins where electrostatic forces play important functional roles can be analyzed (Alexov and Gunner, 1999; Demchuk et al., 2000; Dillet et al., 2000).

The Poisson-Boltzmann equation of continuum electrostatics provides a powerful framework for the analysis of electrostatic interactions in proteins (Warwicker and Watson, 1982; Gilson et al., 1985). These calculations take the distribution of dielectric constants, fixed charges, and mobile ions and return the electrostatic potential. Partial charges are placed at atoms in the protein structure. Water, with a dielectric constant of 80, is the major source of the

Submitted August 3, 2001, and accepted for publication April 19, 2002.

R.E. Georgescu's present address is Department of Cellular Biology, Rockefeller University, 1230 York Ave. P.O. Box 228, New York, NY 10021.

E.G. Alexov's present address is Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th St., New York, NY 10032.

Address reprint requests to M.R. Gunner, Physics Dept., City College of New York, 138th St. and Convent Ave, New York, NY 10031. Tel.: 212-650-5557; Fax: 212-650-6940; E-mail: gunner@sci.ccny.cuny.edu.

0006-3495/02/10/1731/18 $2.00

reaction field (solvation) energy of protein dipoles and charges. Water also screens interatomic interactions in a manner that depends on the distances between protein atoms and the solvent. The Poisson-Boltzmann formalism incorporates the large impact of water easily and reasonably accurately. However, a central problem is how to assign the protein dielectric constant. The dielectric constant of a dried protein is measured to be near 4 (Takashima and Schwan, 1965). Molecular dynamics simulations show the protein inner core is not very polarizable with an effective dielectric constant near 4, whereas the solvent exposed flexible exterior has a value near 30 (Simonson and Brooks, 1996). Calculations using one, low value for protein ($\approx$4) yield a poor match to benchmark pK_as, but there is significant improvement as ε_{prot} is increased to $\approx$20 (Antosiewicz et al., 1994, 1996; Demchuk and Wade, 1996).

The dielectric constant measures the response of a medium to changes in charge. In protein calculations the dielectric constant is a parameter that averages many kinds of reorganization so that they need not be included explicitly (Sham et al., 1997, 1998; Simonson, 1998, 2001; Gunner and Alexov, 2000; Schutz and Warshel, 2001). For example, changes in electronic polarization, backbone or side chain position, residue protonation state, and ion binding can occur as the pH is changed or a reaction proceeds. Electronic polarization may be fairly uniform throughout the protein, and the backbone motions may be small when there is no significant structural change. Thus, their contribution to the electrostatic energy may be estimated from an averaged dielectric constant rather than by explicit atomic calculations. However, much of the protein response is dependent on the local environment.

Several methods that remain based on continuum electrostatics add a nonuniform protein response to charges (for reviews, see Schutz and Warshel, 2001; Simonson, 2001). Different dielectric constants have been assigned to different amino acids (Voges and Karshikoff, 1998) or to different regions of the protein (Demchuk and Wade, 1996; Rocchia et al., 2001). Other methods consider explicit protein motions. Early methods for sampling multiple protein conformations either averaged interactions between different side chain atomic positions (You and Bashford, 1995) or averaged the ionization states calculated with different structures (Ripoll et al., 1996; Zhou and Vijayakumar, 1997). Hybrid methods simultaneously calculate ionization states of acidic and basic residues and the conformation states of surface side chains (Beroza and Case, 1996), hydroxyls and waters (Alexov and Gunner, 1997), and polar and ionizable side chains (Alexov and Gunner, 1999).

Other approaches include: adding explicit protein relaxation into the protein dipole Langevin dipole methodology with a linear response approximation (Sham et al., 1997, 1998); the use of a nonuniform response parameterized on fragment hydrophobic constants, which implicitly draws on the relationship between a group's polarity and its ability to partition into water (Mehler and Guarnieri, 1999); as well as an iterative mobile cluster approach for calculation of multiple-site ligand binding to flexible macromolecules (Spassov and Bashford, 1999).

The multiconformation continuum electrostatic (MCCE) method allows multiple positions of hydroxyl and water protons, side chain rotamers, and ligands in the calculation of the pH dependence of the ionization equilibria of titratable groups (Alexov and Gunner, 1997, 1999). An analysis of the pH titrations of 166 residues in 12 proteins is reported here. The benchmark root mean square errors of these titrations are comparable with those found with other approaches (Antosiewicz et al., 1994, 1996; Mehler and Guarnieri, 1999). Because improvement in pK_a calculations depends on the protein changing side chain position with pH, MCCE highlights motions that may be important for maintaining protein structure and function.

METHODS

The MCCE method

MCCE calculates the equilibrium conformation and ionization states of protein side chains, buried waters, ions, and ligands at a defined pH. Preselected choices for atomic positions and ionization states for side chains and ligands are used. Each specified side chain or ligand position and charge is called a conformer. Look-up tables are calculated of electrostatic and nonelectrostatic conformer self- and pair-wise interactions. Protein microstates have one conformer for each side chain, and ligand. Monte Carlo sampling establishes each conformer's probability in the Boltzmann distribution at a given pH or E_h (Alexov and Gunner, 1997, 1999). The method can compare mutated with wild-type proteins (Alexov et al., 2000) or protein with reactant or product bound (Alexov and Gunner, 1999). The MCCE procedure is divided into three stages: 1) selection of conformers (steps 1–3 below); 2) calculation of energy look up tables (steps 4–5); 3) Monte Carlo sampling and calculation of pK_as (steps 6–7). Early descriptions of the method can be found in Alexov and Gunner (1997, 1999).

Step 1: generate protein data file, striped of surface waters, complete with all protons except those on hydroxyls or waters

All protons in the original protein coordinate file are stripped off. A translation table brings residue and atom names into compliance with the conventions of the MCCE package. Residues with missing atoms are reported. Every independently ionizable group must have a unique residue name and number so backbone N- and C- terminal atoms are given a residue designator that differs from their side chain. The chain terminus conformer interacts with the terminal side chain. The program Proteus (Samir Lipovaca City College of New York) places protons with no degrees of freedom. Protons with no partial charges, such as methylene protons, are added in their torsion minima. Only buried waters are treated in atomic detail. Water accessibility is determined with the program SURFV (Sridharan et al., 1992; Bharadwaj et al., 1995) and waters with $>$10% surface exposure are deleted. Ions are also automatically deleted if they are on the protein surface but retained if they are buried.

Step 2: identify residues with strong electrostatic pair-wise interactions

One DelPhi (Nicholls and Honig, 1991) continuum electrostatics calculation for all polar and ionized acidic and basic residues finds the pair-wise

interactions between residues as well as the interaction between residues and the backbone (see step 4). Residues with pair-wise interactions greater than ±2.1 ΔpK units (2.9 kcal/mol) will be provided with additional conformers from the heavy atom rotamer library (Dunbrack and Karplus, 1994; Dunbrack and Cohen, 1997).

Step 3: generate composite protein structure

Conformers must be generated that could be at low energy over the entire range of the calculation before knowing the position or ionization state of other groups. However, if conformers that make good hydrogen bonds to ionized sites are unavailable, the calculated ionization free energy will be wrong. Inside the protein, extra, bad conformers increase the size of the calculation. However, because they are not selected in Monte Carlo sampling, they do not affect the outcome. On the protein surface, extra conformers change the protein-solvent boundary, introducing errors into the electrostatic energy terms. Therefore, conformer preselection tries to generate as many good choices as possible without including positions that will never be found in low energy microstates.

Side chains identified as making strong pair-wise interactions in step 2 are replicated on the backbone in rotamer libraries positions (Dunbrack and Karplus, 1994; Dunbrack and Cohen, 1997). New side chains with atoms closer than 2.0 Å to fixed atoms (backbone and nonpolar side chains) are deleted. If all rotamers clash with the rigid part of the protein, additional conformers are made by 30° rotations around the final bond.

The working protein data file contains a single copy of the atomic positions for the backbone and side chains with no degrees of freedom. It also includes complete copies of each side chain conformer. The following conformers are generated for the original side chain as well as for each added library rotamer.

Each Arg and Lys gets one ionized and one neutral conformer. Each His gets two ionized and four neutral conformers. Two basic conformers are formed by interchange of atoms, equivalent to rotation around the CB—CG bond (see Fig. 6B). Each of these has two neutral forms in which ND1 or NE2 is unprotonated and one ionized form. Each Asn and Gln has two conformers with the terminal O and N interchanged representing 180° rotation of O—C—N. Each Asp, Glu, and Tyr gets one ionized and two neutral conformers with the proton in the torsion minimum. Each Ser and Thr has three conformers with a proton in each torsion minima. Hydroxyl-containing residues (Ser, Thr, neutral Asp, Glu, and Tyr) have extra conformers that can hydrogen bond with all nearby acceptors. Water gets conformers to make and accept available hydrogen bonds plus a conformer with no interaction with the protein, representing water in bulk solvent.

Step 4: electrostatic self- (reaction field) and pair-wise interactions

As described in Alexov and Gunner (1997), one DelPhi (Nicholls and Honig, 1991) calculation is carried out for each conformer. Parse partial charges and radii (Sitkoff et al., 1994), a water probe radius of 1.4 Å, 0.15 M salt, and a 2.0-Å Stern ion-exclusion radius are used. Focusing yields a resolution better than 2.13 grids/Å (Gilson et al., 1987). Each DelPhi calculation has partial charges on atoms of only one conformer. Net ionized residue charges are ±1, and they are 0 on polar, or neutral, ionizable confomers. The net charge for phosphate is −2 and +2 for calcium and magnesium. The potential (Ψ) is collected at atoms of all other conformers. The pair-wise electrostatic interaction energy between the designated conformer (j), which is the source of the potential and the distal conformer (k) is

$$\Delta G^{el}_{jk} = \sum_{k'=1}^{atoms.in.conf.k} \Psi_{(j\rightarrow k')} q_{k'} \qquad (1)$$

in which $\Psi_{j\rightarrow k'}$ is the potential at atom k' in conformer k from partial charges on conformer j. The partial charge, $q_{k'}$, that would be on k' in the ionization state of conformer k. In the same DelPhi run, the interaction of the conformer j with polar backbone and side chains without conformers is collected in a single interaction energy ($\Delta G_{pol,j}$).

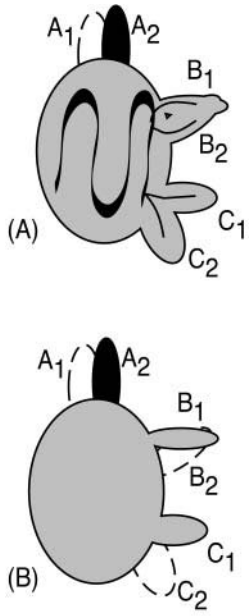

FIGURE 1 Dielectric boundary used for calculation of (*A*) pair-wise interactions and (*B*) reaction field energy. A2 is the only conformer with partial charges in the DelPhi calculation. Speckled region at $\varepsilon_{prot} = 4$. Conformers within dashed lines and the surrounding area at $\varepsilon = 80$.

One approximation is that all conformers are combined in the protein structure adding to the dielectric boundary. Now only N DelPhi calculations are needed. In contrast, N^2 calculations would be needed if the right conformers for each pair-wise interaction were used. However, even this would not provide the correct position for all other conformers in a given microstate. The small proteins studied here are the worst case for errors caused by extra surface conformers. In contrast, buried sites in large proteins will be essentially undisturbed by moderate changes in protein surface.

Several steps reduce the errors in dielectric boundary introduced by extra conformers. For the pair-wise interactions, only the active conformer of the residue with partial charges is included (Fig. 1 *A*). However, other residues must have all conformers present so all pair-wise interactions can be determined in a single calculation. The theoretically symmetrical ΔG^{el}_{ij} and ΔG^{el}_{ji} are calculated with different conformers removed and these pair-wise interactions are averaged. In addition, the reaction field energy (Rocchia et al., 2001) of each conformer is determined in a second series of DelPhi calculations with all residues in their original position except for the conformer with partial charges (Fig. 1 *B*).

Some residues are more sensitive to having their electrostatic interactions improperly calculated because of the low-dielectric material added by the additional conformers (Fig. 1). These can be identified by comparing conformer reaction field energies in protein prepared for calculating pair-wise interactions (Fig. 1 *A*) with those found with boundary used for reaction field energies (Fig. 1 *B*). Approximately 15% of the ionized conformers have differences between these two values of more than 1 ΔpK unit. However, these residues are not found to have larger errors in calculated pK_a.

Step 5: self- (torsion) and pair-wise (Lennard-Jones) nonelectrostatic energies for all conformers

The proton torsion energy is nonzero only for protons placed out of a torsion minimum to make hydrogen bonds. Neutral Tyr has two energy minima in the

plane of the Tyr ring with a 2.59 ΔpK unit (3.5 kcal/mol) barrier between them, so $E(\varphi) = -3.5/2 \cos(2\varphi)$ kcal/mol in which φ is the torsion angle. Ser and Thr have three degenerate low energy positions with $E(\varphi) = -1.3/2 \cos 3(\varphi - 60°)$ kcal/mol. The energy of a neutral Asp and Glu hydroxyl proton is the sum of a torsion potential with minima at 0° and 180° and a 0.96 ΔpK unit barrier and a −2.0 ΔpK unit electrostatic attraction between hydroxyl proton and nonbonded carboxylate oxygen when both are in plane so $E(\varphi) = -((2.7/2)\cos(\varphi) + (1.3/2)\cos(2\varphi))$kcal/mol. This results in a −2.00 ΔpK unit preference for the syn (0°) over anti (180°) conformation, comparable with the value found by solid-state nuclear magnetic resonance (NMR) measurements (Gu et al., 1996). As only library rotamers are used for heavy atom conformers, these are near torsion minima and their torsion energies are set to zero.

The pair-wise Lennard-Jones interactions between all conformers are calculated and added to the pair-wise electrostatic interactions. Lennard-Jones interactions with portions of the protein that are held rigid are added to the pair-wise polar interaction for each conformer. The *A* and *B* constants are found in SM XI.

Step 6: Monte Carlo sampling to determine the Boltzmann averaged conformer occupancy

A given protein microstate is a combination of one conformer for each residue, cofactor, and water. The energy of microstate *n* (ΔG_n) is

$$\Delta G_n = \sum_{i=1}^{M} \delta_n(i)\{\gamma(i)[2.3k_bT(\mathrm{pH} - \mathrm{pK}_{sol,i})] + (\Delta\Delta G_{rxn,i} + \Delta G_{pol,i} + \Delta G_{nonel,i})\} + \sum_{i=1}^{M} \delta_n(i) \sum_{j=i+1}^{M} \delta_n(j)[\Delta G_{ij}^{el} + \Delta G_{ij}^{nonel}] \quad (2)$$

in which k_bT is 0.43 ΔpK units (0.59 kcal/mol), *M* is the number of conformers, $\delta_n(i)$ is 1 for conformers that are present in the state and 0 for all others. $\gamma(i)$ is 1 for bases, −1 for acids, and 0 for neutral conformers (polar groups, waters, neutral acids, and bases). $pK_{sol,i}$ is the solution pK_a of ionizable group *i*, $\Delta\Delta G_{rxn,i}$ is the difference between the reaction field energy for this residue type in solution (SM XII) and that found in the protein (Nicholls and Honig, 1991); $\Delta G_{pol,i}$ and $\Delta G_{nonel,i}$ are the electrostatic and Lennard Jones interactions between conformer *i* and backbone and side chains with no conformational degrees of freedom; $\Delta G_{ij}^{non\text{-}el}$ and ΔG_{ij}^{el} are pair-wise interactions between conformer *i* and *j*. The limits on the summation of the pair-wise interconformer terms ensure that each interaction is counted only once.

$pK_{sol,i}$ is obtained from studies of peptides (Richarz and Wüthrich, 1975; Matthew et al., 1985). Asp and Glu use a carboxylic acid reference pK_a of 4.75. Studies have found that the backbone lowers the pK_a of these residues even in peptides (Oliveberg et al., 1995; Gunner et al., 2000). Using a simple carboxylic acid reference pK_a avoids double counting the impact of the backbone in peptide and protein.

The solution reaction field energies of ionizable groups used with protein dielectric constants 4, 8, and 20 are given in SM XII. These parameters were adjusted so that the average pK_a error for each amino acid type is close to zero.

Several techniques aid Monte Carlo convergence. 1) The number of conformers per residue ranges from 2 to more than 20. Sampling first selects a residue and then chooses one conformer to change. Thus, all residues make similar numbers of changes but in residues with many conformers, each is sampled fewer times. 2) On 50% of the steps, a second residue changes conformation. This is selected randomly from residues that interact by at least 1 ΔpK unit with the first. The new microstate with two changes is then subjected to Metropolis sampling (Beroza et al., 1991). 3) After a predetermined number of steps, conformers with no occupancy are deleted. Those with 100% occupancy are fixed on and their pair-wise and self-energy terms are added to the fixed, polar energy terms. This can reduce the size of the calculation by as much as 70%. 4) The conformer occupancies are saved after a stage of conformer elimination. Monte Carlo sampling is reinitiated with a new seed. If the changes in occupancy after the new cycle are less than a preset limit, the program exits.

Step 7: calculate residue pK_a and n *values from conformer occupancies as a function of pH*

Monte Carlo sampling is performed from pH 0 to 14 in steps of 1 pH unit and the pK_a of each titrating group is obtained from the best nonlinear, least squares fit to

$$\langle \mathrm{Occ}_i \rangle = \frac{10^{\gamma_i n_i(\mathrm{pH}-\mathrm{pK}_i)}}{1 + 10^{\gamma_i n_i(\mathrm{pH}-\mathrm{pK}_i)}} \quad (3)$$

in which $_i$ is the occupancy of the ionized form of residue *i* at a given pH, n_i is the Hill coefficient reflecting the degree of cooperatively between the sites (Cantor, 1980), and $\gamma(i)$ is the charge, −1 for an acidic or 1 for a basic residues.

Timing of the MCCE calculations

The analysis of Bovine pancreatic trypsin inhibitor (BPTI) (58 residues and 173 conformations) takes 3 h, whereas RNase H (155 residues and 608 conformers) takes 8 h on a single processor of an R10,000, 180 MHz SGI Origin 200. Most the time is spent in the DelPhi calculations. Monte Carlo sampling to determine all conformer occupancies at 15 pHs takes 20 min for BPTI and 1.5 h for RNase H.

Single conformer continuum electrostatics (SCCE) calculations

SCCE and MCCE calculations use the same protein structures with the same parameters for atomic charge and radius. In SCCE, waters are deleted and hydroxyl protons are placed in the most occupied MCCE position at pH 7. The most populated form of the neutral acidic residue at low pH is the only available neutral conformer. ε_{prot} is 4.

Atomic coordinates and experimental pK_a values

MCCE was used to calculate pK_as in 12 proteins (Table 1). The 166 measured values are from ^1H -^{13}C heteronuclear resonance experiments unless otherwise specified. The experimental and calculated pK_a for each residue are reported in the supplementary material (SM I-X).

Barnase (SM-I) uses crystal structures 1A2P (Martin et al., 1999), 1B20 (Buckle et al., 1993), and the 20 NMR model structures (1BNR) (Bycroft et al., 1991). Experimental values were measured at low ionic strength (Oliveberg et al., 1995). The pK_a of His-18 was obtained from fluorescence measurements (Loewenthal et al., 1993).

BPTI (SM-II) uses crystal structure (4PTI) (Marquart et al., 1983) and 20 NMR structures (1PIT) (Berndt et al., 1992). pK_a from NMR experiments performed at 41°C were corrected to 25°C using the method described in March et al. (1982).

Intestinal bovine calcium-binding protein (calbindin, CabD) (SM-III) consists of crystal structure in presence of Ca^{2+} (3ICB) (Szebenyi and Moffat, 1986), 24 NMR structures for the holo-protein (1CDN) (Akke et al., 1995), and 33 NMR structures for the apoprotein (1CLB) (Skelton et al., 1995). pK_as of Lys with and without calcium is from Kesvatera et al. (1996). The acidic pK_as in the apoenzyme are from Kesvatera et al. (1999).

Rat T-lymphocyte adhesion glycoprotein (CD2) (SM-IV) uses crystal structure 1HNG (Jones et al., 1992). pK_as of 14 acidic residues determined

TABLE 1 Number of polar and charged residues

Protein	SCOP class	All		D + E			C + N ter			Y			H			K			R		N + Q		S + T		HOH	
		Res	*N*	Res	*N*	Conf	Res	*N*	Conf	Res	*N*	Conf	Res	*N*	Conf	Res	*N*	Conf	Res	Conf	Res	Conf	Res	Conf	Res	Conf
BPTI	Small	58	14	4	4	51	2	2	4	4	4	18	–	–	–	4	4	8	6	14	4	7	4	15	4	56
Barnase	$\alpha + \beta$	110	14	12	12	190	2	1	5	7	–	70	2	1	8	8		36	6	30	10	18	18	70	8	94
CabD-apo	α	75	19	17	8	148	2	1	5	1	–	4	–	–	–	10	10	36	–	–	6	12	8	31	–	–
CabD-holo	α	75	10	17		195	2	–	5	1	–	15	–	–	–	10	10	36	–	–	6	9	8	33	1	14
CD2D1	β	176	14	23	13	153	2	1	5	5	–	26	1		8	13	–	48	11	30	17	33	26	87	–	–
HEWL	$\alpha + \beta$	129	19	9	9	135	2	1	5	3	2	27	1	1	4	6	6	30	11	42	17	33	17	82	6	77
HIV-1	β	99	2	8	2	66	2	–	5	1	–	6	1	–	4	6	–	20	4	10	9	17	9	31	1	5
Omtky3	Small	56	15	5	5	73	2	2	5	3	3	28	1	1	8	4	4	20	1	2	5	10	9	40	1	5
ProtG	$\alpha + \beta$	56	13	10	9	118	2	1	5	3	1	18	–	–	–	6	2	44	–	–	4	8	11	47	–	–
RNaseA	$\alpha + \beta$	124	16	10	10	93	2	2	5	6	–	51	4	4	36	10	–	32	4	18	17	34	25	94	4	21
RNaseHi	α/β	155	25	19	19	265	2	1	5	5	–	42	5	5	36	11	–	44	10	44	15	29	14	57	7	86
RnaseT	$\alpha + \beta$	104	4	12	1	137	2	–	6	9	–	92	3	3	36	2	–	4	1	4	15	22	21	79	7	118
TRX	α/β	108	1	16	1	163	2	–	–	2	–	7	1	–	4	10	–	54	1	2	7	14	9	35	9	64
All			166	145	93	1592	24	12	60	50	10	404	19	15	144	90	36	376	55	196	126	237	179	701	48	526

SCOP class (Murzin et al., 1995); res: Number of residues; N: Number of residues with experimentally determined pK_as; conf: number of MCCE conformers with $\varepsilon_{prot} = 4$ for representative crystal structure (Table 2). HOH: Number of buried waters retained in the calculation.

at low (0.1 mM) and high (1.2 mM) salt concentration were averaged (Chen et al., 2000).

Hen egg-white lysozyme (HEWL) (SM-V) uses triclinic crystal structures (2LZT) (Ramanadham et al., 1981), 1B0D (Vaney et al., 1996), tetragonal crystal structure 1HEL (Wilson et al., 1992), 50 NMR structures (1E8L) (Schwalbe et al., 2001), and the averaged NMR structure (1HWA) (Smith et al., 1993). pK_as of basic residues are from Kuramitsu and Hamaguchi (1980), and values for acidic residues at 100 mM ionic strength are from Bartik et al. (1994).

Turkey ovomucoid inhibitor domain 3 (OMTKY3) (SM-VI) consists of OMTKY3 complexed with human leukocyte elastase (1PPF) (Bode et al., 1986), *Streptomyces griseus* proteinase B (3SGB) (Read et al., 1983), or α-chymotrypsin (1CHO) (Fujinaga et al., 1987). The proteases are removed. There are 50 NMR structures (1OMU) (Hoogstraten et al., 1995). Acidic residue pK_as are measured at low (10 mM) and high ionic strength (1 M) (Schaller and Roberston, 1995). Basic residue pK_as are measured at 15 mM and 500 mM ionic strengths (Forsyth et al., 1998).

B1-immunoglobulin G-binding domain of protein G (ProtG) (SM-VII) uses crystal structure (1PGA) (Gallagher et al., 1994), 60 NMR structures (1GB1) (Gronenborn et al., 1991), and minimized NMR structure (2GB1) (Gronenborn et al., 1991). pK_as at 100 mM ionic strength are from Khare et al. (1997).

Ribonuclease A (RNase A) (SM-VIII) uses the crystal structure with sulfate (3RN3) (Howlin et al., 1989), ion free protein (7RSA) (Wlodawer et al., 1988), and 32 NMR structures (2AAS) (Santoro et al., 1993). pK_as at 200 mM NaCl (Rico et al., 1991; #4211). pK_as of His-12, -48, -105, and -119 in the absence of phosphate are an average of values from Ruterjans and Witzel (1969), Matthews and Westmoreland (1973), Walters and Allerhand (1980), and Quirk and Raines (1999); His-112 and His-119 pK_as determined at 20 mM and 100 mM phosphate (Meadows et al., 1969; Cohen et al., 1973). Sulfate is treated as phosphate.

Ribonuclease Hi (RNase H) (SM-IX) uses crystal structures 2RN2 (Katayanagi et al., 1992) and 1RNH (Yang et al., 1990) and Mg^{2+} containing form (1RDD) (Katayanagi et al., 1993) and 8 NMR derived structures (1RCH) (Yamazaki et al., 1997). pK_as are from Oda et al. (1993, 1994).

Ribonuclease T1 (RNase T) (SM-X) uses the crystal structure with vanadate (3RNT) (Kostrewa et al., 1989). Vanadate was treated as phosphate. Measured pK_as are from Inagaki et al. (1981) and McNutt et al. (1990).

Human Immunodeficiency Virus Type-1 Protease (HIV-1) uses the crystal structure (1HIV) (Thanki et al., 1992) with dihydroethylene-containing inhibitor, which is removed. pK_as are of the aspartyl dyad from inhibitor free enzyme kinetic studies (Hyland et al., 1991; Ido et al., 1991).

For residues in Barnase, OMTKY3, BPTI, Calbindin, and RNaseHi, there is more than one pK_a reported. The value used is either an average or the pK_a of the site closest to the atom losing the proton. Overall, the uncertainties are small ($\leq$ 0.3 pK units). The *n* values (Eq. 3) were estimated from the published titration curves.

Even in these well-characterized proteins, only 166 of the 374 ionizable residues (Asp, Glu, His, Tyr, Lys, and Arg) have measured pK_as (Table 1). There are no Arg pK_as reported. Residues titrating in the same pH region influence each other. Thus, errors in the titration of Arg may influence Lys or Tyr. Likewise, errors in titration of missing Asp or Glu can impact the titration of the reported residues.

The following have incomplete titration curves that miss the plateau at low or high pH: Lys-34, Asp-7, and Asp-27 in OMTKY3, Lys-25 in CabD, Tyr-3, Tyr-33, Tyr-45, and Lys-13 in ProtG, Asp-93 and Glu-73 in Barnase, Tyr-23 in BPTI, Asp-48 and Asp-66 in HEWL, and His-114 in RNase H. Generally titration curves justify the reported values. There are eight residues where only a limiting value is available. This is taken as the true pK_a, which may overstate the error of the calculations (see supplementary material).

RESULTS AND DISCUSSION

Comparison of MCCE calculated and measured pK_as

The first pK_a calculations using detailed protein structures and energies derived from continuum electrostatics held the protein rigid, allowed only changes in residue ionization states, and used a low protein dielectric constant ($\varepsilon_{prot} = 4$) (Bashford and Karplus, 1990; Lim et al., 1991; Yang et al., 1993; Honig and Nicholls, 1995). SCCE does not match experimentally determined pK_as very well (Antosiewicz et al., 1994; Demchuk and Wade, 1996). However, SCCE provides good benchmarks for comparing different methods. SCCE root mean square (RMS) errors at ε_{prot} of 4 (Table 2) are similar to those found previously. MCCE and SCCE calculations with $\varepsilon_{prot} = 4$ are compared in Fig. 2 and Table 2. Experimental pK_as for each residue and values calculated with different Protein Data Bank files and ε_{prot}

TABLE 2 Errors in calculated pK_as

				*SCCE $\varepsilon_{prot} = 4$			*MCCE $\varepsilon_{prot} = 4$					*MCCE		†MCCE $\varepsilon_{prot} = 4$	
								‡Errors in range				$\varepsilon_{prot} = 8$	$\varepsilon_{prot} = 20$		
Protein	*N*	PDB	R (Å)	RMS error	Max error	‡Errors > 1.5	RMS error	0 to 0.75	0.75 to 1.5	>1.5	Max error	RMS error	RMS error	Added structures	RMS error
BPTI	14	4PTI	1.50	0.80	−1.7	1	0.64	10	4	0	1.3	0.47	0.41	1PIT§	0.45
Barnase	14	1A2P	1.50	2.11	−3.7	7	1.04	7	6	1	−1.7	0.91	0.89	1B20, 1BNR§	0.81
CabD-holo	10	3ICB	2.30	0.87	1.5	1	0.66	8	1	1	1.5	0.38	0.59	1CDN§	0.45
CabD-apo	19	1CLB§		N.D.	N.D.	N.D.	0.71	14	5	0	1.4	N.D.	N.D.	–	N.D.
CD2d1	14	1HNG	2.80	1.62	4.1	4	0.77	12	0	2	1.7	0.76	0.73	–	N.D.
HEWL	19	1B0D	1.84	1.74	−4.1	7	0.63	15	4	0	−1.1	0.74	0.90	2LZT, 1HEL, 1E8§	0.58
HIV-1	2	1HIV	2.00	N.D.	N.D.	N.D.	0.62	1	1	0	−0.8	N.D.	N.D.	–	N.D.
Omtky3	15	1PPF	1.80	2.06	−4.4	7	1.20	7	4	4	−2.9	N.D.	N.D.	1CHO, 3SGB, 1OMU§	1.05
ProtG	14	1PGA	2.07	2.13	−5.0	6	0.67	10	4	0	1.3	0.63	0.75	1GB1§	0.70
RNaseA	16	3RN3	1.45	2.69	−6.8	9	0.99	9	6	1	2.5	0.66	0.44	7RSA, 2AAS§	0.83
RNaseH	24	2RN2	1.48	2.75	8.7	13	1.04	12	8	4	−2.6	0.87	0.77	1RDD, 1RNH, 1RCH§	0.73
RnaseT	4	3RNT	1.80	4.18	5.1	3	0.75	3	1	0	1.1	0.54	0.63	–	N.D.
TRX	1	2TRX	1.68	N.D.	N.D.	N.D.	0.10	1	0	0	0.3	N.D.	N.D.	–	N.D.
Eight proteins	126						0.89								0.73
All proteins	166			2.29	−8.7	58	0.82	109	44	13	~3	0.68	0.70		

N, Number of experimental pK_as; PDB, representative structure file; R, reported resolution; SCCE, single conformer continuum electrostatics; MCCE, multiconformer continuum electrostatics; N.D., not determined.

All electrostatic calculations use DelPhi (Nicholls and Honig, 1991) with a solution dielectric constant of 80; Parse (Sitkoff et al., 1994) charges and radii and a salt concentration of 0.15 M. The protein dielectric constant (ε_{prot}) is 4, 8, or 20.

RMS error, root mean square error (in pH units) is RMS.error = $\mathrm{sqrt}((1/N)\,\Sigma_{i=1}^{N}\,(pK_{exp} - pK_{calc})^2)$ in a protein with *N* measured pK_as; Max error, largest error in pK_a for this protein. Eight proteins, Statistics for the proteins where added structures were considered.

*, Calculations using only the representative structure.

†, Calculations that consider added structures.

‡, Number of residue pK_as with errors in the given range.

§, NMR structures; all other coordinates obtained by x-ray crystallography.

values are in the supplementary material (SM-I-X). For the 166 residues, the average absolute error decreases from 0.46 (SCCE) to 0.11 pH units (MCCE). The RMS error is reduced from 2.29 to 0.81 pH units. Less than 10% of the calculated pK_as differ from experiment by more than 1.5 pH units. With SCCE 40% of the pK_as are in error by this amount and 7% have errors greater than 4 pH units (Table 2).

Eighty-one residues are on the surface with no special interactions. However, 35 experimental pK_as are shifted by >1.0 ΔpK unit from isolated amino acids (Table 3). Thirty-six residues are buried and have lost >3 ΔpK units (4.08 kcal/mol) reaction field (solvation) energy. Fifty-three have ion pair interactions of >3 ΔpK units. Thus, 85 residues are in one or more perturbed categories with many in more than one class. Surface, noninteracting residues are easiest to model. Their SCCE RMS error is only 1.57, and the MCCE error is 0.67 pH units. Of the unperturbed residues His have the largest RMS error because of His-124 in RNAse H (see

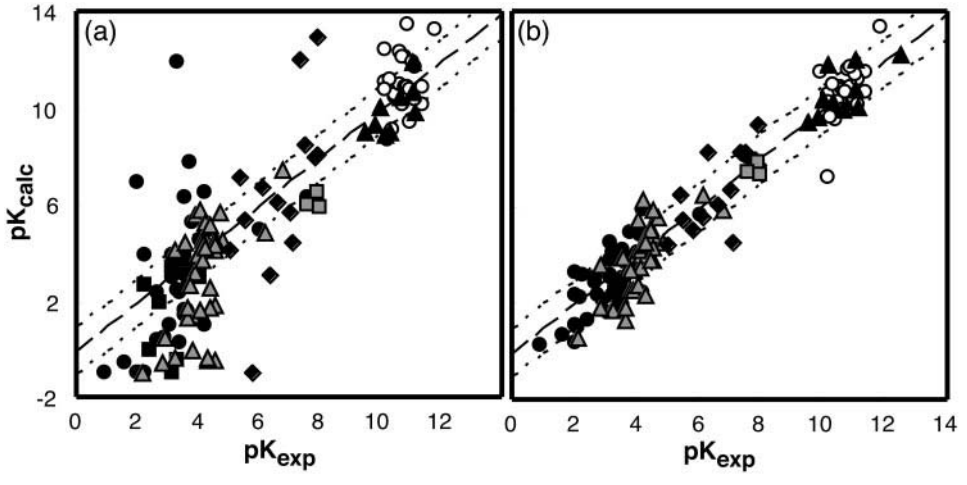

FIGURE 2 Calculated versus measured pK_as for (*a*) single conformer (SCCE) and (*b*) multiconformer (MCCE) continuum electrostatic calculations. $\varepsilon_{prot} = 4$. Asp (●), Glu (▲), Ctr (■), His (♦), Ntr (■), Tyr (▲), and Lys (○).

TABLE 3 Experimental and calculated pK_as for different residue types

		Experimental pK_as				Calculated pK_as										
						ε_{prot} = 4.0		ε_{prot} = 4.0 RMS error (*N*)					ε_{prot} = 8.0		ε_{prot} = 20.0	
Residue	*N*	Null pK_a	Av. pK_a	SD	Max shift	Av. pK_a	Av. err	All	Unperturbed	Shifted	Buried	SE	Av. pK_a	RMS error	Av. pK_a	RMS error
ASP	48	4.00	3.32	1.07	−3.98	3.24	0.08	0.85	0.51 (20)	0.78 (15)	0.90 (16)	1.09 (17)	3.51	0.83	3.57	1.00
C-ter	9	3.80	3.07	0.51	−0.93	3.04	0.02	0.78	0.90 (5)	0.86 (3)	0.65 (1)	0.90 (2)	3.03	0.74	3.03	0.83
GLU	45	4.40	4.10	0.76	−2.63	3.82	0.25	0.92	0.66 (22)	1.07 (8)	1.20 (10)	1.13 (19)	4.03	0.55	4.17	0.88
HIS	15	6.30	6.76	0.96	1.76	6.60	0.16	1.10	1.41 (5)	0.85 (7)	1.74 (2)	1.02 (7)	6.58	0.76	6.62	0.58
LYS	36	10.40	10.70	0.45	−1.11	10.67	0.03	0.84	0.54 (23)	1.56 (1)	0.62 (2)	1.71 (5)	11.28	0.89	11.23	0.83
N-ter	3	7.50	7.82	0.16	0.22	7.47	0.31	0.42	0.42 (3)	(0)	(0)	(0)	7.63	0.26	7.64	0.12
TYR	10	10.20	10.58	0.83	−1.91	10.78	−0.09	0.77	0.54 (4)	0.26 (1)	0.95 (5)	1.18 (3)	10.13	0.67	10.61	0.56
Total				0.74						0.90 (35)	1.00 (36)	1.11 (53)				
MCCE	166							0.83	0.67 (81)*		1.09 (85)†			0.70		0.74
SCCE	144							2.23	1.57*		2.59†					

N, number of residues with experimentally determined pK_as; experimental pK_as, null pK_a, amino side chain pK_as in peptides in solution (Richarz and Wüthrich, 1975; Matthew et al., 1985); Av. pK_a, average of experimental pK_as for residues considered here; SD, standard deviation of measured pK_as; Max Shift, maximum difference between experimental and null pK_as; For the calculated values, Av pK_a, the average calculated pK_a; Av Err, average difference between calculated and experimental pK_a; RMS error, root mean square error (in pH units). Value in parenthesis gives number of residues in each of the following classes. All, all residues; Shifted, residues whose experimental pK_a differs from the null value by more than ±1 pH unit; Buried, residues with a loss of reaction field energy of more than 3 ΔpK units (4.08 kcal/mol); SE (strong electrostatic), Residues where occupied conformers have favorable pair-wise electrostatic interactions of more than 3 ΔpK units; Unperturbed, all other residues.

*RMS errors for unperturbed residues.

†RMS errors for perturbed residues. Most perturbed residues are counted in more than one category (Shifted, Buried, SE).

below). Perturbed residues have larger errors with SCCE (RMS error 2.58 pH units). MCCE provides significant improvement (RMS error 1.09), but these residues pose the greatest challenge for calculation (Table 3). The largest errors come from residues in strong interactions (RMS error 1.17) as do some of the greatest improvements. One difficulty is that many ion pairs involve Arg for which there are no measured pK_as.

pK_as calculated with different structures of the same protein

Calculations should provide values that are the same for all nominally identical structures of the same protein. The average standard deviation with SCCE of the 32 residues in lysozyme structures (2LZT and 1B0D) is 1.4 pH units and only 0.5 pH units with MCCE (SM-V). Thus, MCCE conformational flexibility greatly diminishes the dependence on the initial structure. In addition, the RMS error for the 19 residues with known pK_as is cut in half.

The pK_as derived from crystal and solution structures were compared for 126 residues in eight proteins (Table 4). MCCE uses a rigid backbone and nonpolar side chains. The sampling of conformational space is increased in an ensemble of NMR structures, where each has somewhat different backbone and side chain positions. Thus, the average RMS variation of residue pK_as is 1.5 pH units when different NMR coordinate files are compared. However, the RMS error of the averaged pK_as is 0.83 pH units, smaller than for the crystal structures (0.94). Similar improvement was found in single conformer studies with ε_{prot} of 20 (Antosiewicz et al., 1996). Calculations with single, deposited average NMR structures have slightly larger errors (RMS error 1.02 pH units). However, the improvement with analysis of ensembles of NMR structures comes at the expense of more calculations. Here, 271 NMR structures were analyzed for the comparison of eight proteins.

Comparison of experimental and calculated *n* values

The Hill coefficient, *n*, in Eq. 3 measures how much residue titrations influence each other. Neighboring residues titrating in the same pH range have smaller *n* values. Experimental *n* values were obtained for 94 residues in CabD, Barnase, OMTKY3, CD2d1, and RNaseH, either from the published values or from the shape of published titration curves. Even in these small proteins, half of the residues have *n* values smaller than 0.85, whereas only two have values significantly larger than 1. The MCCE derived values (Eq. 3) are well correlated with those determined experimentally (Table 5).

The correspondence between calculated titration curves and MCCE site occupancies become worse as *n* decreases. This is because Eq. 3 provides only an approximation for the behavior of individual residues, missing the complex pH dependence of clusters of interacting residues (Onufriev et al., 2001).

TABLE 4 MCCE calculations with NMR and x-ray derived structures

	X-ray			NMR						
Protein	PDB	RMS error	Max error	PDB	*n*	RMS error	Max error	RMSV	RMS error*	Max error*
BPTI	4PTI	0.64	1.30	1PIT	20	0.62	1.78	1.40	0.85	1.54
Barnase	1A2P	1.02	1.50	1BNR	20	0.94	1.66	1.63		
CabD-holo	3ICB	0.67	1.60	1CDN	24	0.47	0.99	1.17	0.67	1.21
HEWL	2LZT	0.86	1.75	1E81/1HWA	50	0.74	1.75	1.77	1.22	2.88
Omtky3	1PPF	1.20	3.00	1OMU	50	1.10	2.54	0.87		
ProtG	1PGA	0.67	1.17	1GB1/2GB1	60	0.65	1.44	0.60	0.87	2.17
RNaseA	3RN3	0.99	2.46	2AAS	32	1.28	2.77	2.21		
RNaseH	2RN2	1.04	2.60	1RCH	8	1.15	2.44	1.60	1.33	2.83
All		0.93	3.00			0.84	2.77	1.54	1.02	

$\varepsilon_{prot} = 4.0$. For NMR structures. *n*, the number of structures analyzed; RMS error and Max Error. consider the average of the calculated residue pK_a given all *n* structures. The root-mean-square-variation (RMSV) of pK_as for a residue (*j*) in a group of *n* NMR derived structures is: $RMSV = \mathrm{sqrt}((1/n)\,\Sigma_{j=1}^{n}\,(pK_{av} - pK_j)^2)$ in which pK_{av} is the average ensemble pK_a. RMS* and MaxErr* represent calculations using the single averaged structure deposited with the NMR structures.

Calculation of water and ion binding

Solvent exposed waters are removed, whereas buried waters are treated in the same detail as side chains with conformations allowing rotation about the oxygen. Each water has an extra conformer in bulk solvent with no interactions with the protein. The reaction field energy of free, solvated water controls the occupancy of interior sites. The water/hexane partition coefficient suggests a −3.56 ΔpK units (−4.84 kcal/mol) reaction field energy for an isolated water in bulk water (Wolfenden and Radzicka, 1994). For water to be bound, the remaining reaction field energy plus added interactions with the protein must be more favorable than this. There are 38 buried, crystallographic waters (Table 1). With a solution reaction field energy of −3.56 ΔpK units, the average water site occupancy is only 33% (pH 7). The fraction increases to 77% as the penalty for removing the water from bulk was decreased to 0.87 ΔpK units. Similar results were found in earlier MCCE calculations (Alexov and Gunner, 1999). In the group of proteins studied here only Asp-66 in HEWL has a pK_a that depends on bound water (see below). The indifference to water in this data set can be seen in the similarity of the pK_as derived from crystal structures and those using NMR structures where no waters are included. However, in other systems, bound water can have a large impact on the stability of buried charges (Gibas and Subramaniam, 1996; Garcia-Moreno et al., 1997; Alexov and Gunner, 1999).

TABLE 5 Distribution of Hill coefficients (*n*) values

		Experimental *n* values		
Parameters		$n < 0.85$	$0.85 \le n \le 1.1$	$n > 1.1$
Calculated	$n < 0.85$	32	4	0
n	$0.85 \le n \le 1.1$	15	41	0
Values	$n > 1.1$	0	0	2

Experimental *n* values for 94 residues obtained either from published values or estimated from the shape of the published titration curves. MCCE values calculated with the representative structure (Table 2), $\varepsilon_{prot} = 4$.

Analysis of salt-bridges and corrections for strong electrostatic interactions

It is often difficult to obtain the correct pK_as for residues in ion pairs (Sindelar et al., 1998). Continuum electrostatics calculations often over-stabilize the ionization of charges, lowering acid pK_as and raising those of bases. MCCE with no correction yields pK_as for the 98 acids −0.59 pH units too low and for the 43 bases 0.54 pH units too high. An empirical function (SOFT) (Eqs. 4a and 4b, Fig. 3) was introduced to weaken strong pair-wise electrostatic interactions with other conformers (ΔG_{ij}^{el}) and with backbone and nonconformer containing side chains ($\Delta G_{pol,\,i}$).

$$\Delta G_{ij}^{el}(\mathrm{SOFT}) = \frac{\Delta G_{ij}^{el}}{kT \times \left[1 + \dfrac{e^{(\mathrm{Abs}(\Delta G_{ij}^{el}/\mathrm{kT}) - 5.5)}}{1 + e^{(\mathrm{Abs}(\Delta G_{ij}^{el}/\mathrm{kT}) - 5.5)}}\right]} \tag{4a}$$

$$\Delta G_{pol,i}(\mathrm{SOFT}) = \frac{\Delta G_{pol,i}}{kT \times \left[1 + 1.5 \times \dfrac{e^{(\mathrm{Abs}(\Delta G_{pol,i}/\mathrm{kT}) - 5.5)}}{1 + e^{(\mathrm{Abs}(\Delta G_{pol,i}/\mathrm{kT}) - 5.5)}}\right]} \tag{4b}$$

With full strength interactions the global RMS error is 1.68 pH units (Tables 2 and 6), whereas it is 0.86 with the SOFT function. The average errors for the acids and bases are now −0.16 and 0.07 pH units, respectively. This function has little effect on the pK_a of most groups, but for 20 residues the error is decreased by >1.5 ΔpK units (Fig. 4).

The SOFT function can be viewed as increasing the effective dielectric constant but only for relatively close neighbors (Fig. 3). The largest change halves the interaction energy, equivalent to an ε_{prot} of 8. There are several reasons

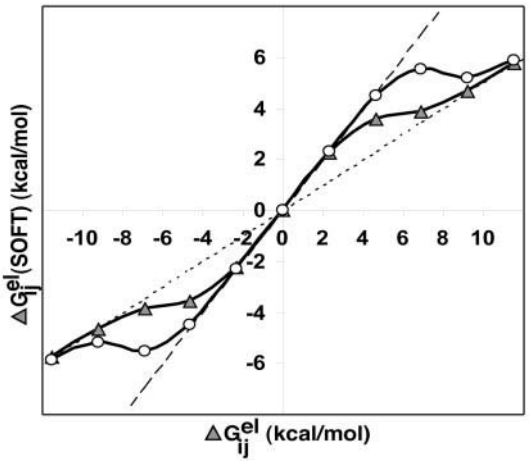

FIGURE 3 SOFT function applied to reduce strong pair-wise (▲) conformer-conformer (Eq. 4a); (○) conformer-polar group (Eq. 4b) electrostatic interactions. Dashed line, initial ΔG_{ij}^{el}; dotted line, energy halved.

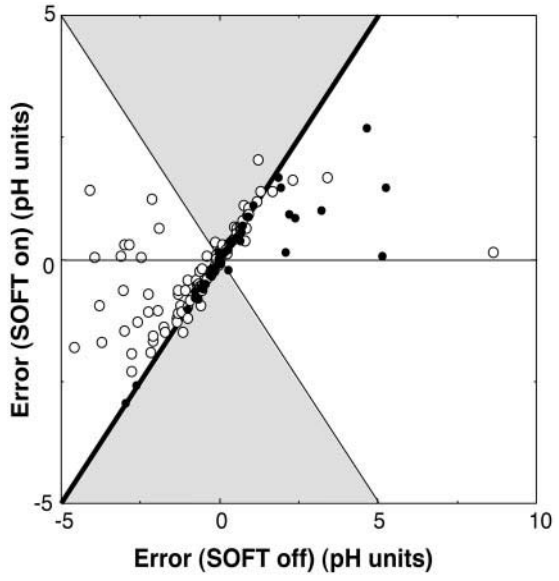

FIGURE 4 Comparison of the error in the pK_as calculated with and without the SOFT function (Eq. 4). (○) acids; (●) bases. Along heavy line pK_as are unchanged. Residues in the shaded region have larger errors with the SOFT function. Those in the white region have their errors reduced.

that this may be necessary. Interaction energies are very sensitive to small changes in position when groups are close together. If structures report the closest allowable separation for salt-bridges as the time-averaged position then the interactions will be too large. Small excursions, weakening electrostatic interactions are not taken into account. In addition, MCCE uses Lennard-Jones parameters (SM-XI) that yield good values for hydrogen bonds. These may underestimate the cost of bringing ion pairs together, overestimating the electrostatic interactions. Lastly, the available conformers may not be optimal for stabilizing the charge-dipole state of the ion pair at low and high pH. Thus, the SOFT function represents a pragmatic solution to the problem of obtaining good values for ion pairs. However, future implementations of MCCE will aim to use enhanced local conformer flexibility to render this function unnecessary.

TABLE 6 Accuracy of pK_a calculations with and without the SOFT function

			RMS error		Max error	
Protein	PDB	*N*	Soft-off	Soft-on	Soft-off	Soft-on
BPTI	4PTI	14	0.69	0.64	1.29	1.30
Barnase	1A2P	14	2.14	1.04	3.11	1.50
CabD-apo	3ICB	19	1.25	0.71	2.25	1.39
CabD-holo	1CLB	10	0.75	0.66	1.92	1.56
CD2d1	1HNG	14	1.21	0.77	2.84	1.67
HEWL	1B0D	19	0.86	0.63	1.96	1.11
HIV-1	1HIV	2	0.58	0.62	0.73	0.81
OMTKY3	1PPF	15	1.49	1.20	3.48	2.97
ProtG	1PGA	14	1.32	0.67	2.36	1.17
RNaseA	3RN3	16	2.20	0.99	3.80	2.46
RNaseH	2RN2	24	1.94	1.04	3.71	2.60
RNaseT	3RNT	4	5.80	0.75	8.30	1.11
All		165	2.16	0.83	4.8	3

ε_{prot} = 4.0. *N* is the number of residues with measured pK_as for each protein. SOFT function is Eq. 4.

MCCE calculations with different protein dielectric constants

Residue pK_as were calculated at ε_{prot} of 4, 8, and 20 (Tables 2 and 3). With SCCE, there is marked improvement as ε_{prot} is increased (Antosiewicz et al., 1996; Demchuk and Wade, 1996). However, MCCE calculations are only moderately sensitive to ε_{prot}. The RMS error at ε_{prot} = 8 (0.69) is slightly smaller than with ε_{prot} = 4 (0.84) or ε_{prot} = 20 (0.70). Here, site titration is stabilized by either explicit conformation change at small ε_{prot} or by the averaged dielectric response as ε_{prot} increases. Fewer conformational changes are seen as ε_{prot} is increased.

With larger ε_{prot}, buried charges have a smaller $\Delta\Delta G_{rxn}$ and smaller pair-wise interactions. Because most pair-wise interactions are favorable, changes in these two terms tend to offset each other. However, explicit conformational flexibility can stabilize a charge more than ε_{prot} of 20. For example, the MCCE pK_a of Asp-7 in OMTKY3 increases from 2.8, the correct value, to 5.0 as ε_{prot} increases from 4 to 20. In other cases the ionized form is too stable with ε_{prot} of 20 indicating the high dielectric constant either overestimates the local protein flexibility or weakens unfavorable pair-wise interactions.

Conformational flexibility in the MCCE calculations

In the proteins used here, 55% of the residues have multiple conformers either because they are polar (23%) or ionizable (32%). Overall 62% of the available conformers have some

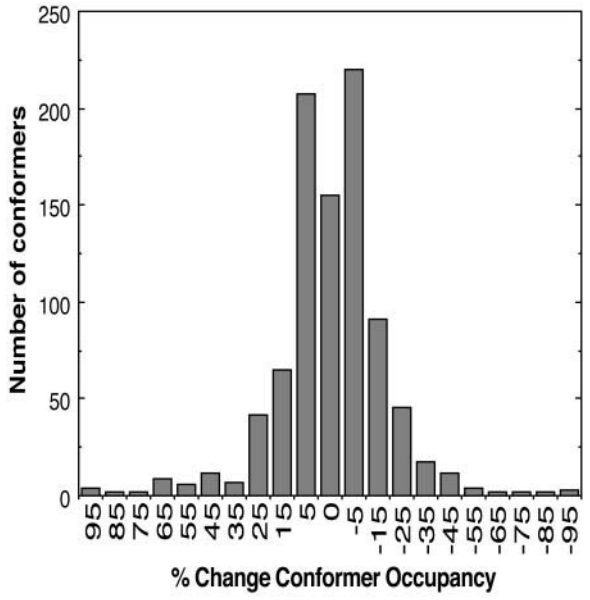

FIGURE 5 Number of polar (Ser, Thr, Asn, Gln) conformers that have changed occupancy by the given percentage as the pH is changed from 0 to 14.

TABLE 7 Outcome with different methods for pK_a calculations: the results for lysozyme

Method	ε_{prot}	RMS error	Glu-35	Asp-52	Asp-66
Exp pK_a			6.2 ± 0.1	3.7 ± 0.1	0.9 ± 0.5
MCCE	4	0.86	**6.2**	3.0	2.6
MCCE*	**4**	0.98	5.3	**7.1**	1.2
SCCE	4	2.05	5.3	**7.8**	−1.0
M-G	4	0.72	**5.9**	3.5	2.6
S-C-W	4	1.36*	**5.9**	4.9	3.3
Y-B	4	1.30	4.2	**6.8**	2.0
D-W	15/75	0.81	**4.5**	3.0	2.9
V-S-K	20	0.70	**4.3**	3.3	1.8
A-MC-G	20	0.80	3.6	3.4	1.4
W	20	0.98	4.7	NR	NR

Calculations of lyzozyme active site cluster, Glu-35, Asp-52, and Asp-66 pK_as using coordinate file 2LZT. Residue with the highest pK_a is highlighted. SCCE, MCCE data presented here; MCCE*, Asn-46 is fixed in conformer found in 2LZT; RMS error, all calculations compared to the 19 experimental pK_as in (Bartik et al., 1994) with the exception of S-C-W (Sham et al., 1997) where only seven residue pK_as were reported. M-G (Mehler and Guarnieri, 1999); Y-B (You and Bashford, 1995); D-W (Demchuk, and Wade, 1996); V-S-K (Van Vlijmen et al., 1998); A-MC-G (Antosiewicz et al., 1996); W (Warwicker, 1999); NR, values not reported.

occupancy and 48% are occupied by more than 10% during the titration. Focusing on changes in polar residues (Asn, Gln, Ser, Thr, and waters), 17% do not change conformation between pH 0 and 14, 65% show small changes of 1% to 20%, whereas only 4% change occupancy by >50% (Fig. 5). Most motions are hydroxyl reorientation or rotation of Asn or Gln, which would not be seen by x-ray crystallography. However, these rotations rearrange hydrogen bond networks and so can have significant impact on calculated pK_as (Hooft et al., 1996; Nielsen and Vriend, 2001). In addition, several residues take heavy atom conformations different from that found in the protein data file at some point in the pH titration, and these have significant impact on calculated pK_as (see below).

MCCE analysis of individual proteins

The benchmark pK_a calculations show the MCCE method yields as good fit to data as other current methods (Antosiewicz et al., 1996; Mehler and Guarnieri, 1999) (see below). The unique advantage of this method is that it follows explicit movement of side chains and waters during a titration and so highlights important conformational changes (Alexov and Gunner, 1999).

Lysozyme

Lysozyme is a glycosyl transfer enzyme (Vocadlo et al., 2001). The cleaved bond is held between Glu-35 and Asp-52. There are 31 ionizable residues and all have measured pK_as with the exception of the 11 Arg, 1 Tyr, and the N terminus. Glu-7-Lys-1 and Asp-48-Arg-61 are ion pairs. Glu-35, Asp-52, and Asp-66 are in a cluster of residues with like charge. With the exception of Lys-1, these interacting residues are all buried, with $\Delta\Delta G_{rxn}$ 3 ΔpK units. Many pK_a calculations have been carried out on this protein (see below). The MCCE RMS error for 1BOD is 0.63 pH units. Three residues have errors of ≈1.1 pH units, whereas all other errors are smaller (SM-V).

Polar side chain motions

The active site, Glu-35 has a pK_a of 6.2, Asp-52 of 3.7, and Asp-66 of 0.9. The interaction between 35 and 52 is only ≈1 ΔpK unit. The high Glu pK_a has been hard to calculate. With SCCE the order of ionization of 35 and 52 are interchanged (Table 7). However, MCCE calculations with several x-ray and NMR structures reproduce the pK_as. One determinant is Asn-46. At low pH, conformers with OD1 and ND2 near Asp-52 are both occupied. At high pH, the Asn always has ND2 closer to the ionized Asp stabilizing it by ≈2 ΔpK units. Monte Carlo sampling with the Asn fixed in the conformer found in the crystal structure (OD1 closer to Asp-52) interchange the order of titration of Glu-35 and Asp-52 (Table 7).

Small motions yield significant changes

When Asn-46 is fixed the pK_a of Asp-66 decreases by 1.4 pH units. The difference in interaction with ionized Asp-52 and Glu-66 of ≈0.5 ΔpK unit is too small to explain the change. However, reorientation of Thr-51, Ser-60, Thr-69, two waters, and Arg-68^+ all stabilize Asp66^- more when Asn-46 is fixed. The Asp-66 pK_a also decreases as the water

TABLE 8 pK_as of interacting residues in RNaseA and CD2: the impact of ion binding and mutation

	Experimental pK_as		SCCE	MCCE			
CD2	Wild type	Mutant	Wild type	Wild type	WT*	E29Q	E41Q
Glu-29	4.5	4.5 (E41Q)	1.8	4.8	3.9	–	3.2
Glu-41	6.5	5.0 (E29Q)	7.5	5.8	6.2	4.1	–

RNAse A	Experimental pK_as		SCCE	MCCE		
	No PO_4	$+PO_4$	No PO_4	No PO_4	Dynamic	Fixed PO_4
His-12	5.8	7.0	−1.0	4.8	6.8	7.6
His-119	6.1	7.1	6.7	5.4	6.2	6.4

Experimental pK_as for wild-type (WT) and mutant CD2 (Chen et al., 2000). The structure 1HNG used for all calculations. WT* has Glu-29 fixed in the heavy atom position in 1HNG. Calculations of the E29Q and E41Q mutants disallowed all ionized conformers of the appropriate Glu forcing the residue to remain neutral. Previous calculations have shown similar pair-wise interactions of a neutral Glu and a Gln with surrounding residues (Alexov et al., 2000).

Experimental pK_as for RNase A (3RN3) in the absence of phosphate (Ruterjans and Witzel, 1969; Matthews and Westmoreland, 1973; Walters and Allerhand, 1980; Quirk and Raines, 1999) and with phosphate (Meadows et al., 1969; Cohen et al., 1973). The SO_4 in 3RN3 is treated as a dianionic PO_4. All simulations use the same energy look-up tables, which has one conformer with the ion as found in the data file and another with no interactions with the protein. No PO_4, Only the noninteracting conformer allowed; dynamic, both conformers allowed; fixed PO_4, only the bound conformer allowed.

solution reaction field energy is diminished, increasing the occupancy of stabilizing, buried waters.

CD2d1

Rat CD2d1, in the immunoglobulin superfamily, is the N-terminal domain of the T-cell antigen CD2 (Driscoll et al., 1991). It is important for cell surface recognition (McAlister et al., 1996) and transmembrane signaling during T-cell activation (Sunder-Plassmann and Reinherz, 1998). The CD2 ligand-binding surface has a high proportion of charged and polar residues including Glu-28, Glu-29, and Glu-41, Arg-31, and Thr-37 so ionic interactions are thought to be important for binding and specificity. The pK_as of 14 of the 55 ionizable residues have been determined (Chen et al., 2000) (SM -IV).

Rotamer changes on the protein surface

Asp-28, Glu-29, and Glu-41 are in a cluster on the binding surface (Fig. 6 *A*, Table 8). Glu-41, with a pK_a of 6.5, helps bind target proteins (van der Merwe et al., 1995; Chen et al., 2000). At low pH, Glu-41^0, Glu-29^- occupy rotamers close to Arg-31^+ and Lys-43^+ further from Asp-28^-. As Glu-41 is ionized, Glu-29^- reorients. Fixing Glu-29 in its original rotamer lowers its pK_a by 0.7 pH units, whereas that of Glu-41 is raised by 0.4 pH units. This cluster is modeled less well at $\varepsilon_{prot} = 20$.

Calculation of pK_as in mutated proteins

Glu-29 and Glu-41 have been mutated to Gln and pK_as measured for nearby residues (Chen et al., 2000). MCCE calculates pK_as in mutated proteins by supplying conformers for the replacement residue. These are added to the wild-type structure, so interactions for both wild-type and mutant side chain are in the same energy look-up table. Monte Carlo sampling then allows only appropriate residue conformers (Alexov et al., 2000). Previous calculations have shown that interactions with neutral Asp are comparable with that of Asn (Alexov et al., 2000). The mutation of Glu-29 and Glu-41 in CD2 were therefore modeled with the wild-type energy look-up table where the mutated acid was simply forced to remain in a neutral conformer. If Glu-29 is mutated to Gln, the pK_a of Glu-41 is lowered by 1.7 pH units in agreement with the experiment shift of 1.5 pH units (Table 8). However, mutation of Glu-41 to Gln lowers the calculated pK_a of 29 by 1.6 pH units in contrast to the measured negligible change.

Prot G

Streptococcal protein G, an immunoglobin binding protein with small (55–75 amino acids) repeating domains (B1, B2, and B3), binds immunoglobin-G constant regions (Tashiro and Montelione, 1995). The pK_as of 13 of the 21 ionizable residues have been determined (Khare et al., 1997) and residue ionization studied previously by an SCCE based approach with ε_{prot} of 20 (Khare et al., 1997). Glu-56 and Lys-28 are buried ($\Delta\Delta G_{rxn} > 3$). Ion pairs include Glu-27 with both Lys-28 (3.0 Å) and Lys-31 (3.0 Å), Glu-15 with Lys-4 (3.4 Å), Asp-47 with Lys-50 (3.0 Å), and Glu-19 with the N terminus (3.4 Å). Each pair has >3 ΔpK units pair-wise interaction with each other.

Accurate pK_as of ion pairs can rely on side chain motions

Protein G has five residues in salt bridges. SCCE at $\varepsilon_{prot} = 4$ for the five residues with known pK_as have an RMS error of 3.1 pH units, whereas it is 0.75 with MCCE. In the crystal structure (1PGA) Glu-27 is close to Lys-28 and Lys-31 so the SCCE pK_a of Glu-27 is too low and that of Lys-28 too high. In the average NMR structure (1GB1) the Glu is further than 8 Å from either Lys. MCCE provides good pK_as for both Lys-28 and Glu-27 starting with the 1PGA structure because a new Glu conformer is selected that is closer to that found by NMR. Changes in conformation reducing ion-pair stabilization had been suggested in previous SCCE based calculations at $\varepsilon_{prot} = 20$, which could not obtain the correct pK_a for this Glu (Khare et al., 1997). The

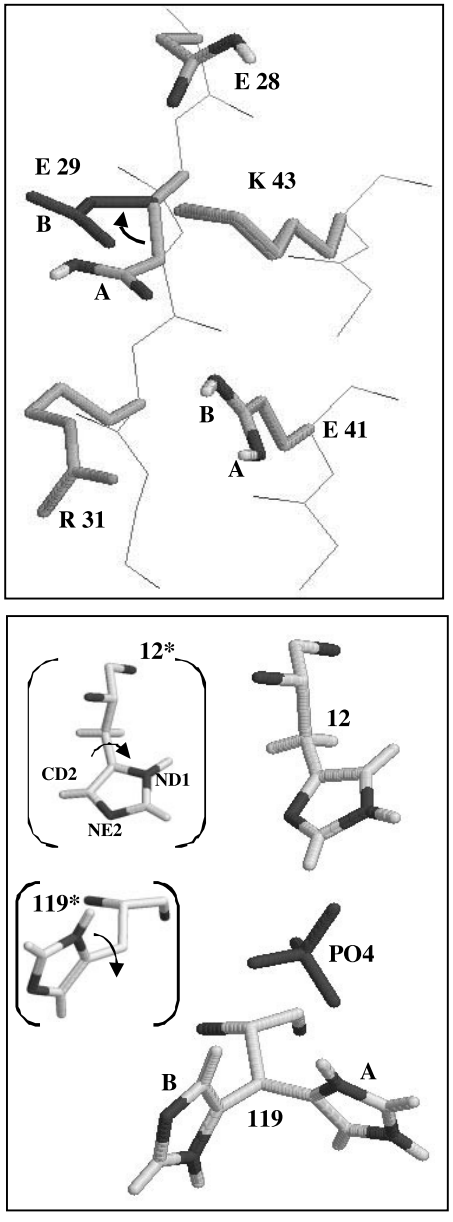

FIGURE 6 (*Top*) Position of occupied conformers of interacting residues on the binding surface of rat CD2d1 (1HNG). The order of ionization is Glu-28, Glu-29, and then Glu-41 as the pH is raised. Glu-29 moves from neutral conformer A to ionized position B. The B conformer has more favorable interactions with Lys-43 and Arg-31. At low pH, the neutral Glu-41 is found with the proton at position A (78%) and B (18%). (*Bottom*) Conformer positions in the active site of RNaseA (3RN3). His-119, His-12, and the doubly ionized PO_4 shown. The imidazole nitrogens (ND1 and NE2) in darker gray. His-12* and His-119*, found in the deposited structure, differ from His-12 and His-119 by rotation around the CB—CG bond. His-119 positions A and B are included in the deposited structure file. At low pH ionized His-12* predominates (70% occupied) and His-119B* and His-119B are both found. Between pH 3 and 5, a small amount of ionized His-119A is occupied. Both His-119 proton tautomers (ND1 or NE2 protonated) are occupied. PO_4 dissociates from the protein as His-12 looses its proton.

accurate pK_a for Glu-15 also relies on several rotamers being occupied during its titration.

OMTKY3

The avian ovomucoids are glycoproteins composed of three tandem, homologous structural domains, each acting as a serine protease inhibitor (Laskowski et al., 1987). Structures of the third domain, complexed to different proteases, have been determined. The ionization states of the 15 the 16 ionizable residues in the inhibitor have been measured in the absence of protein target (Schaller and Roberston, 1995; Forsyth et al., 1998). Previous calculations of pK_as were made using an SCCE based approach with $\varepsilon_{prot} = 20$ (Forsyth et al., 1998). NMR studies show the binding region in solution is located between Lys-13-Pro-22 and Asn-33-Ala-40 (Song and Markley, 2001). Lys-13 and Lys-34, Glu-19, and Tyr-20 are found at this interface (SM-VI).

Problems calculating pK_as of residues in clusters

Lys-13, Lys-34, and Tyr-11 form a triad titrating near the same pH, a motif that is often causes problems. In 1PPF the pK_a of Lys-13 and Tyr-11 are both 1.6 pH units too high, whereas Lys-34 is 2.6 pH units too low. The ionized Lys-34 is destabilized by loss of reaction field energy and unfavorable interactions with backbone. A favorable interaction of -1.2 ΔpK units between Lys-13^0 and Tyr-11^- helps stabilize the anionic Tyr. In 3SGB the two Lys pK_as are close to experiment, whereas Tyr-11 is >4 pH units too high. The $\Delta\Delta G_{rxn}$ for Lys-34 and its interaction with backbone are reduced by 3 pH units shifting its pK_a up. This Lys takes different conformations when it is neutral and ionized. However, the occupied conformer of Lys-13^0 has little interaction with the Tyr-11^-, which now remains neutral even at pH 14.

RNaseA

Bovine pancreatic RNaseA is a ribonuclease with a well-defined binding cleft containing His-12 and His-119 and Lys-7, Lys-41, and Lys-66. Of the 36 ionizable residues 16 have known pK_as (Rico et al., 1991; Quirk and Raines, 1999) and residue ionization has been studied previously by an SCCE based approach with $\varepsilon_{prot} = 20$ (Antosiewicz et al., 1996).

Selection of His tautomers

His-12 and His-119 are functionally important residues. There are two positions for His-119 in 3RN3. Previous studies have shown the calculated pK_a depends on the choice of rotomer and neutral His tautomer, which must be preassigned in SCCE based methods (Antosiewicz et al., 1996). With MCCE all rotamers and tautomers are available

during Monte Carlo sampling (Fig. 6 *B*). With this freedom the calculated pK_as are 0.7 to 1.6 pH units lower than the experimental value (phosphate free) and relatively independent of the structure analyzed (SM VIII).

Ion binding

His-12 and His-119 bind phosphate increasing the pK_a of each by 1 to 1.2 pH units (Table 8) (Cohen et al., 1973; Matthews and Westmoreland, 1973; Walters and Allerhand, 1980). In MCCE ion binding is modeled by providing the phosphate with two conformers. One interacts with the protein, whereas the second isolated conformer has −45.6 ΔpK units (−62 kcal/mol) of reaction field energy. When bound the phosphate looses 9.5 ΔpK units of reaction field energy (3RN3), which must be replaced by favorable interactions, primarily with ionized His-12 and His-119, for the protein bound conformer to be occupied. Although it is possible to dynamically calculate the phosphate ionization state within MCCE, this was not done here. Rather, phosphate is always doubly ionized. Monte Carlo sampling was carried out under three conditions using the energy look-up tables for 3RN3. In one, the phosphate conformer bound to the protein is forced to be occupied, in another the unbound conformer is fixed on, lastly both conformers are allowed and so the ion remains in equilibrium with the protein. Comparing phosphate bound and free, the pK_a of His-12 increases by 2.8 and His-119 by 1.0 pH unit (Table 8). The shift is intermediate if phosphate remains in equilibrium with the protein. Here the phosphate dissociates, as His-12 looses its proton. The calculations thus reproduce the order of ionization of 12 and 119 as well as the magnitude of the pH shifts with and without phosphate. The experimental pK shift of 1 to 1.2 pH with added phosphate is closer to that found in the dynamic calculations suggesting that the ion is lost when the two His are neutral.

Unstable calculated pK_as for residues in a buried ion pair

His-48 has a measured pK_a of 6.3 (Table 3). With MCCE the pK_a is 2.5 pH units too high in the crystal structure 3RN3, whereas in NMR structures (2AAS) the average is 1.7 pH units too low with an RMS variance of 4.4 pH units. The variability of His-48 is coupled to changes in Asp-14 with which it has an interaction of ≈ -5 ΔpK units when both are ionized. NMR structures, which gave very different pK_as for these groups, were compared. When the Asp $\Delta\Delta G_{rxn}$ is moderate and its interaction with the backbone favorable it titrates at low pH and the His remains ionized to very high pH (10 or greater). However, there are structures where the Asp $\Delta\Delta G_{rxn}$ is >9 ΔpK units so it now stays neutral until pH 10. Here the favorable His^+:Asp^- pairwise interactions are never important because the His titrates well before the Asp. Thus, small shifts in Asp $\Delta\Delta G_{rxn}$ significantly changes the pK_a of both Asp and His.

RNase T1

Aspergillus oryzae RNase T1 and Barnase from *Bacillus amyloliqefaciens* are guanine-specific microbial RNases with an active site similar to RNase A (Mauguen et al., 1982).

Stabilization of ionized His

RNase T has three histidines (27, 40, and 92) all with elevated pK_as, measured between 7.3 to 7.9. The MCCE pK_as lie between 7.1 and 9.0, reproducing the high values. His-40 and His-92 are near Glu-58 in a cluster that binds phosphate. In MCCE the phosphate leaves the protein near pH 4, slightly lower than where Glu-58 becomes ionized. $Glu\text{-}58^-$ then stabilizes the ionized His as suggested previously (McNutt et al., 1990). His-92, with the higher calculated pK_a, has a stronger interaction with Glu 58^- than does His-40. The high pK_a of His-27, which is distant from 40 and 90, relies on a favorable interaction with $Glu\text{-}82^-$.

An isolated, buried Asp stabilized by polar side chains

Asp-76 has a profound effect on the conformational stability of RNase T (Giletto and Pace, 1999). Protein unfolding studies suggests that it is fully ionized at physiological pH with a pK_a below 1. The calculated pK_a is 3.4. The residue is deeply buried with a $\Delta\Delta G_{rxn}$ 6 ΔpK units, Dipolar interactions with Asn-9, Tyr-11, and Thr-91 alone provide −6.5 ΔpK units stabilization. The Asn is in the conformer found in the protein data file at high pH but is rotated by 180° when the Asp is neutral. The hydroxyl in the Tyr and Thr take a distribution of positions that change with pH. In addition, favorable interactions with Arg-77 and His-92 add another −2 ΔpK units when they are ionized.

RNaseH

RNaseH is a small endonuclease (124 residues), which cleaves the RNA strand of a DNA-RNA hybrid. Four acidic residues (Asp-10, Asp-70, Asp-134, and Glu-48) and a Mg^{+2} form the active site. There are 36 ionizable residues. All acidic residues, as well as the His and Ntr, have measured pK_as (SM IX). Previous analysis of acidic residue pK_as used an SCCE-based approach with ε_{prot} of 4 and 10 (Oda et al., 1994).

Effects of conformational changes not considered by MCCE

His-124, near the active site, has a pK_a of 7 (Oda et al., 1994). However, its calculated pK_a is 4.5 in 2RN2, 1RDD, and in the NMR structures 1RCH. These structures were obtained above pH 8 where the His is neutral and the calculations show this His form is too stable (Katayanagi et

al., 1992). All have His-124 pointed out toward the surface with a $\Delta\Delta G_{rxn}$ of less than 0.25 ΔpK units. However, all have ≈ -1.5 ΔpK units interaction with the backbone destabilizing the positively charged His. The nearby Lys-122 also shifts the pK_a down slightly (<0.5 ΔpK unit). Given only structures with an outward conformation for His-124 the MCCE yields a firm, incorrect prediction that the pK_a is lower than 5. However, 1RNH has the His buried within the protein close to the active site acidic cluster (Oda et al., 1993). Although this structure is also made above pH 8, the calculated pK_a is 8.7, so the ionized His is now too stable. Here it has lost more reaction field energy but has favorable interactions with four acids. Thus, it appears that the correct pK_a may be obtained if the His can move between the buried (ionized) and exposed (neutral) forms, which would require changes in backbone conformation. The His is in a loop with a sequence GHAG. The Gly may aid mobility. However, backbone changes cannot currently be considered in MCCE conformer sampling.

Errors in the calculated pK_a for Asp-10

There are two acids, Asp-70 and Asp-13 near Asp-10. 1RDD has a Mg^{+2} bound to the cluster, whereas the other structures do not. Experimental and calculated titration curves (Oda et al., 1994) show coupling of Asp-10 and Asp-70 with most of 70 titrating near pH 2.5 and 10 near pH 6.5. With Mg^{+2}, the measured Asp-10 pK_a decreases by 2 pH units, whereas the other two acidic residues show moderate shifts (0.8 pH units) (Oda et al., 1994). Although Asp-70 and Asp-134 are reasonably calculated in presence (1RDD) and absence (1RNH) of Mg^{+2} the Asp-10 calculated pK_a is at least 4 pH units too high. Several factors contribute. Asp-10 is the most deeply buried of the three acids and the most destabilized by the others. In addition, the neutral Asp-10 is actually stabilized by interactions with ionized Asp-134 and Asp-70. A variety of Monte Carlo calculations were carried out in an attempt to lower the Asp-10 pK_a. The Mg^{+2} position was varied. pK_as for Asp-70 and Asp-134 remained close to experimental values, but Asp-10 remained neutral. There are several exposed crystallographic waters in the active site that are stripped off during protein preparation. These were added back, changing the calculated Asp-10 pK_a by less than 0.5 pH units. Thus, MCCE calculations of this residue are so aberrant that this is the one residue whose calculated pK_a is not included in the RMS error for the method. Given its position it is hard to justify a pK_a as low as 6 in any available crystal or solution structure. It should be noted that in MCCE calculations at $\varepsilon_{prot} = 20$ the pK_a is far closer to that measured. Thus, it appears that there are conformational changes on ionization that are not captured in the current MCCE methodology or in any of the available structure.

Error assessments

Benchmark calculations are carried out to validate a method so it can be used to estimate unknown pK_as. Thus, it is important to recognize when a calculated pK_a is more likely to be wrong. Of the 166 residues considered here, in the standard structure chosen to represent each protein, 26% have errors >1 ΔpK unit and only 3% 2 ΔpK units. The four residues with errors >2 ΔpK units are Asp-10 and His-124 in RNase H, His-48 in RNase A, and Lys-34 in OMTKY3, all described above.

Several parameters were compared to determine if residues with a poor fit to the experimental data could be identified. The average value of n (0.89) in Eq. 3 is the same for the residues with errors greater and smaller than 1-pH unit. The fit to theoretical titration curve (Eq. 3) was also similar for both classes. However, when multiple protein structures were available, the variability is significantly greater for residues with greater error. Thus, if the standard deviation of the calculated pK_a is <0.8-pH units most residues (88%) have errors <1-pH unit, whereas, if it is >1.2-pH units most (72%) have calculated pK_as with errors >1-pH unit. Three of the four residues with errors >2-pH units have standard deviations of >2. The variability in calculated pK_a sometimes highlights clear changes in structure (such as for His-124 in RNase H). However, often as in the case of His-48 and Asp-14, the structures are not very different. Rather modest differences in energy terms determine which residue in a cluster is ionized first.

Residues in clusters, where several interacting groups titrate near the same pH, are the most difficult to calculate. There are two clusters in barnase, Asp-8, Asp-12, and the Ctr, and in the active site Asp-54, Asp-75, Asp-86, and Glu-73. In OMTKY3 there is Tyr-11 and Lys-13 and Lys-34. In RNase H Asp-10, Asp-70, and Asp-134 are in the active site. Asp-52, Asp-66, and Glu-35 are in the active site of lysozyme. Of the 17 residues with errors of more than 1.4 pH units, eight are in these four clusters. Here small changes in energies change the order of ionization leading to large pK_a shifts. Finer sampling of MCCE conformers just for these clusters may lead to more reliable results. For example, the rearrangement of Asn-46 is the key to obtaining the correct order of ionization in the lysozyme cluster (Table 7).

As expected, including calculations with more structures improves the reliability of the calculations by increasing the conformational space sampled. One structure is used as representative for each protein. Multiple structures were analyzed for eight proteins. The RMS error for the representative structures is 0.89, whereas it is 0.73 when pK_as obtained from all proteins are averaged.

Comparison between the results with different methods for pK_a calculation

The MCCE calculations provide a global RMS error of ~0.8 ΔpK units. This is comparable to calculations with

data sets of ≈90 residues using SCCE based methods at $\varepsilon_{prot} = 20$ (Antosiewicz et al., 1994, 1996; Khare et al., 1997). Another approach using a nonhomogeneous environment with a microenvironment-modulated screening function (sigmoidally screened Coulomb potentials) gives an excellent match to measured data (RMS error ~ 0.5 for ≈100 residues (Mehler and Guarnieri, 1999)).

The RMS error for a group of residues provides a rough view of the goodness of fit of the calculations. However, there are differences in the residues considered for each protein and in the experimental values chosen for comparison. Different methods also obtain good values for different residues. A second approach is to consider the outcome for a small group of residues. The cluster in lysozyme containing Glu-35, Asp-52, and Asp-66 has been considered in detail in earlier work, all of which compare the structure in 2LZT. All $\varepsilon_{prot} = 4$, single conformation calculations (SM-V), or calculations with limited flexibility (You and Bashford, 1995) find that Asp-52 has a pK_a higher then Glu-35. With $\varepsilon_{prot} = 20$, single conformer calculations (Antosiewicz et al., 1996; Warwicker, 1999) or averages of multiple structures generated by molecular dynamics (Van Vlijmen et al., 1998) find the pK_as of these two groups to be similar, close to the pK_as of these acids in solution. Methods that raise the interior dielectric constant are strongly biased to return a pK_a close to the solution value. Methods where the interior dielectric is defined for individual residues (Demchuk and Wade, 1996) also return low pK_as for both Glu-35 and Asp-52 because significant portions of the protein have an ε_{prot} of 15 or higher. The average experimental residue pK_as are slightly stabilized relative to what is found in small peptides with a modest standard deviation (Table 2). Thus, high ε_{prot} calculations model the bulk of the data well but have difficulties calculating the rare but interesting shifted pK_a.

Linear response approximation calculations by Warshel and collaborators (Sham et al., 1997), which use the semimicroscopic version of the protein dipole Langevein dipole method to incorporate the effects of relaxation, obtain the correct order of titration for the residues. This methodology allows backbone and side chain positions to respond to changes in charge. The calculation of Mehler and Guarnieri reproduce the pK_as of the cluster very well. The screening function introduced in this method depends heavily on resides surrounding a residue. If they are more polar the electrostatic interactions are more heavily screened. The empirical function appears to be able to find good pK_a values for groups throughout a protein (Mehler and Guarnieri, 1999). As described above, MCCE obtains good values for the pK_as. The rotation of Asn-46 occurs with the titration of Asp-52. This conformational change represents a specific mechanism for dielectric relaxation of this particular Asp, diminishing the energy difference between ionized and neutral forms of the acid.

CONCLUSIONS

The ionization state of proteins as a function of pH, metal ion, and mutation was calculated for several soluble proteins. In each, MCCE calculations show the importance of electrostatic and dynamic factors in evaluating the free energy of a residue as it changes ionization state. The loss of reaction field energy, pair-wise interactions with the backbone, and polar and ionized side chains captured in classical, single conformer continuum electrostatic studies remain important. However, the conformational flexibility in MCCE improves the match to benchmark pK_as without blunting interaction energies as is done in methods that use high average dielectric constants. Rather MCCE allows the protein explicit mechanisms of relaxation as the charge state changes. The structure can change with pH to follow changes in hydrogen bonding patterns and ion binding site occupancy.

MCCE allows only limited changes in conformation. Thus, the motions seen here as coupled to residue ionization represent possible conformational changes. In particular, the importance of backbone motions can now be seen only by comparison of different starting structures. Future calculations will aim to sample more degrees of freedom to allow ever more realistic view of the interaction of protein with charged groups.

We would like to acknowledge Samir Lipovaca's contributions to PROTEUS, Demetri Karpman's to conformer placing, and Nataliya Ozhegova's to *n* calculations, and the financial support of National Science Foundation MCB 9629047.

REFERENCES

Akke, M., S. Forsen, and W. J. Chazin. 1995. Solution structure of (Cd^2)1-calbindin D9k reveals details of the stepwise structural changes along the Apo 224 (Ca^{2+})II1 224 (Ca^2)I, II2 binding pathway. *J. Mol. Biol.* 252:102–121.

Alexov, E., and M. Gunner. 1999. Calculated protein and proton motions coupled to electron transfer: the electron transfer from Q_A^- to Q_B in bacterial photosynthetic RCs. *Biochemistry.* 38:8253–8270.

Alexov, E., J. Miksovska, L. Baciou, M. Schiffer, D. Hanson, P. Sebban, and M. R. Gunner. 2000. Modeling the effects of mutations on the free energy of the first electron transfer from Q_A^- to Q_B in photosynthetic reaction centers. *Biochemistry.* 39:5940–5952.

Alexov, E. G., and M. R. Gunner. 1997. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* 72:2075–2093.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties in 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.* 35:7819–7833.

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.

Berndt, K. D., P. Guntert, L. P. Orbons, and K. Wuthrich. 1992. Determination of a high-quality nuclear magnetic resonance solution structure of

the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J. Mol. Biol.* 227:757–775.

Beroza, P., and D. Case. 1996. Including side chain flexibility in continuum electrostatic calculations of protein titration. *J. Phys. Chem.* 100: 20156–20163.

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. U. S. A.* 88:5804–5808.

Bharadwaj, R., A. Windemuth, S. Sridharan, B. Honig, and A. Nicholls. 1995. The fast multipole boundary element method for molecular electrostatics: an optimal approach for large systems. *J. Comp. Chem.* 16:898–913.

Bode, W., A. Z. Wei, R. Huber, E. Meyer, J. Travis, and S. Neumann. 1986. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J.* 5:2453–2458.

Brown, L. R., A. De Marco, G. Wagner, and K. Wuthrich. 1976. A study of the lysyl residues in the basic pancreatic trypsin inhibitor using ^{1}H nuclear magnetic resonance at 360 Mhz. *Eur. J. Biochem.* 62:103–107.

Buckle, A. M., K. Henrick, and A. R. Fersht. 1993. Crystal structural analysis of mutations in the hydrophobic cores of barnase. *J. Mol. Biol.* 234:847–860.

Bycroft, M., S. Ludvigsen, A. R. Fersht, and F. M. Poulsen. 1991. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry.* 30:8697–8701.

Cantor, C. R. 1980. Biophysical Chemistry. Freeman, New York, NY.

Chen, H. A., M. Pfuhl, M. S. McAlister, and P. C. Driscoll. 2000. Determination of pK_a values of carboxyl groups in the N-terminal domain of rat CD2.: anomalous pK_a of a glutamate on the ligand-binding surface. *Biochemistry.* 39:6814–6824.

Cohen, J. S., J. H. Griffin, and A. N. Schechter. 1973. Nuclear magnetic resonance titration curves of histidine ring protons: IV. The effects of phosphate and sulfate on ribonuclease. *J. Biol. Chem.* 248:4305–4310.

Demchuk, E., U. K. Genick, T. T. Woo, E. D. Getzoff, and D. Bashford. 2000. Protonation states and pH titration in the photocycle of photoactive yellow protein. *Biochemistry.* 39:1100–1113.

Demchuk, E., and R. C. Wade. 1996. Improving the continuum dielectric approach to calculating pK_a's of ionizable groups in protein. *J. Phys. Chem.* 100:17373–17387.

Dillet, V., R. L. Van Etten, and D. Bashford. 2000. Stabilization of charges and protonation states in the active site of the protein tyrosine phosphatases: a computational study. *J. Phys. Chem. B.* 104: 11321.1–11333.

Driscoll, P. C., J. G. Cyster, I. D. Campbell, and A. F. Williams. 1991. Structure of domain 1 of rat T lymphocyte CD2 antigen. *Nature.* 353: 762–765.

Dunbrack, R. L., and F. E. Cohen. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* 6:1661–1681.

Dunbrack, R. L., and M. Karplus. 1994. Conformational analysis of the backbone-dependent rotamer preferences of protein chains. *Nat. Struct. Biol.* 1:334–340.

Dyson, H. J., M. Jeng, L. L. Tennant, I. Slaby, M. Lindell, D. Cui, S. Kuprin, and A. Holmgren. 1997. Effects of buried charged groups on cysteine thiol ionization and reactivity in *Escherichia coli* thioredoxin: structural and functional characterization of mutants of Asp 26 and Lys 57. *Biochemistry.* 36:2622–2636.

Elcock, A. H., and J. A. McCammon. 1998. Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* 280:731–748.

Ferguson-Miller, S., and G. T. Babcock. 1996. Heme/copper terminal oxidases. *Chem. Rev.* 96:2889–2907.

Forsyth, W. R., M. K. Gilson, J. Antosiewicz, O. R. Jaren, and A. D. Robertson. 1998. Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. *Biochemistry.* 37:8643–8652.

Fujinaga, M., A. R. Sielecki, R. J. Read, W. Ardelt, M. Laskowski, and M. N. James. 1987. Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J. Mol. Biol.* 195:397–418.

Gallagher, T., P. Alexander, P. Bryan, and G. L. Gilliland. 1994. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry.* 33: 4721–4729.

Garcia-Moreno, B., J. J. Dwyer, A. G. Gittis, E. E. Lattman, D. S. Spencer, and W. E. Stites. 1997. Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* 64: 211–224.

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

Giletto, A., and C. N. Pace. 1999. Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T1. *Biochemistry.* 38:13379–13384.

Gilson, M. K., and B. Honig. 1988. Energetics of charge-charge interactions in proteins. *Proteins.* 3:32–52.

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.

Gilson, M. K., K. A. Sharp, and B. H. Honig. 1987. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comp. Chem.* 9:327–335.

Gronenborn, A. M., D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, and G. M. Clore. 1991. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science.* 253:657–661.

Gu, Z., C. F. Riendenour, C. E. Bronnimann, T. Iwashita, and A. McDermott. 1996. Hydrogen and distance studies of amino acids and peptides using solid state 2D ^{1}H-^{13}C heteronuclear correlation spectra. *J. Am. Chem. Soc.* 118:822–829.

Gunner, M. R., and E. Alexov. 2000. A pragmatic approach to structure based calculation of coupled proton and electron transfer in proteins. *Biochim. Biophys. Acta.* 1458:63–87.

Gunner, M. R., M. Saleh, E. Cross, A. ud-Doula, and M. Wise. 2000. Backbone dipoles generate positive potentials in all proteins: origins and implications of the effect. *Biophys. J.* 78:1126–1144.

Havranek, J. J., and P. B. Harbury. 1999. Tanford-Kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. U. S. A.* 96:11145–11150.

Hendsch, Z. S., T. Jonsson, R. T. Sauer, and B. Tidor. 1996. Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. *Biochemistry.* 35:7621–7625.

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

Hooft, R. W. W., C. Sander, and G. Vriend. 1996. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins.* 26:363–376.

Hoogstraten, C. G., S. Choe, W. M. Westler, and J. L. Markley. 1995. Comparison of the accuracy of protein solution structures derived from conventional and network-edited NOESY data. *Protein Sci.* 4:2289–2299.

Howlin, B., D. S. Moss, and G. W. Harris. 1989. Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallogr. A.* 45:851–861.

Hyland, L. J., T. A. Tomaszek, Jr., and T. D. Meek. 1991. Human immunodeficiency virus-1 protease 2: use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry.* 30:8454–8463.

Ido, E., H. P. Han, F. J. Kezdy, and J. Tang. 1991. Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. *J. Biol. Chem.* 266:24359–24366.

Inagaki, F., Y. Kawano, I. Shimada, K. Takahashi, and T. Miyazawa. 1981. Nuclear magnetic resonance study on the microenvironments of histidine residues of ribonuclease T1 and carboxymethylated ribonuclease T1. *J. Biochem. (Tokyo).* 89:1185–1195.

Jones, E. Y., S. J. Davis, A. F. Williams, K. Harlos, and D. I. Stuart. 1992. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature.* 360:232–239.

Katayanagi, K., M. Miyagawa, M. Matsushima, M. Ishikawa, S. Kanaya, H. Nakamura, M. Ikehara, T. Matsuzaki, and K. Morikawa. 1992.

Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* 223:1029–1052.

Katayanagi, K., M. Okumura, and K. Morikawa. 1993. Crystal structure of *Escherichia coli* RNase Hi in complex with Mg^{+2} at 2.8 Å resolution: proof of a single Mg^{+2} binding site. *Proteins.* 17:337–346.

Kesvatera, T., B. Jonsson, E. Thulin, and S. Linse. 1996. Measurement and modeling of sequence-specific pK_a values of lysine residues in calbindin D9k. *J. Mol. Biol.* 259:828–839.

Kesvatera, T., B. Jonsson, E. Thulin, and S. Linse. 1999. Ionization behavior of acidic residues in calbindin D(9k). *Proteins.* 37:106–115.

Khare, D., P. Alexander, J. Antosiewicz, P. Byran, M. Gilson, and J. Orban. 1997. pK_a measurements from nuclear magnetic resonance for the B1 and B2 immunoglobulin G-binding domains of protein G: comparison with calculated values for nuclear magnetic resonance and X-ray structures. *Biochemistry.* 36:3580–3589.

Kostrewa, D., H. W. Choe, U. Heinemann, and W. Saenger. 1989. Crystal structure of guanosine-free ribonuclease T1, complexed with vanadate (V), suggests conformational change upon substrate binding. *Biochemistry.* 28:7592–7600.

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

Laskowski, M., Jr., I. Kato, W. Ardelt, J. Cook, A. Denton, M. W. Empie, W. J. Kohr, S. J. Park, K. Parks, B. L. Schatzley, O. L. Schoenberger, M. Tashiro, G. Vicho, H. E. Whatley, A. Wieczorek, and M. Wieczorek. 1987. Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. *Biochemistry.* 26:202–221.

Lim, C., L. Bashford, and M. Karplus. 1991. Absolute pK_a calculations with continuum dielectric methods. *J. Phys. Chem.* 95:5610–5620.

Loewenthal, R., J. Sancho, T. Reinikainen, and A. R. Ferst. 1993. Long-range surface charge-charge interactions in proteins: comparison of experimental results with calculations from a theoretical method. *J. Mol. Biol.* 232:574–583.

Luecke, H., B. Schobert, H. T. Richter, J. P. Cartailler, and J. K. Lanyi. 1999. Structural changes in bacteriorhodopsin during ion transport at 2 Ångstrom resolution. *Science.* 286:255–261.

March, K. L., D. G. Maskalick, R. D. England, S. H. Friend, and F. R. Gurd. 1982. Analysis of electrostatic interactions and their relationship to conformation and stability of bovine pancreatic trypsin inhibitor. *Biochemistry.* 21:5241–5251.

Marquart, M., J. Walter, J. Deisenhofer, W. Bode, and R. Huber. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. Sect. B.* 39:480–490.

Martin, C., V. Richard, M. Salem, R. Hartley, and Y. Mauguen. 1999. Refinement and structural analysis of barnase at 1.5 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* 55:386–398.

Matthew, J. B., F. R. N. Gurd, B. Garcia-Moreno, M. A. Flanagan, K. L. March, and S. J. Shire. 1985. pH-dependent processes in proteins. *CRC Crit. Rev. Biochem.* 18:91–197.

Matthews, C. R., and D. G. Westmoreland. 1973. A Fourier transform NMR study of the thermal denaturation of ribonuclease A at low pH. *Ann. N.Y. Acad. Sci.* 222:240–254.

Mauguen, Y., R. W. Hartley, E. J. Dodson, G. G. Dodson, G. Bricogne, C. Chothia, and A. Jack. 1982. Molecular structure of a new family of ribonucleases. *Nature.* 297:162–164.

McAlister, M. S., H. R. Mott, P. A. van der Merwe, I. D. Campbell, S. J. Davis, and P. C. Driscoll. 1996. NMR analysis of interacting soluble forms of the cell-cell recognition molecules CD2 and CD48. *Biochemistry.* 35:5982–5991.

McNutt, M., L. S. Mullins, F. M. Raushel, and C. N. Pace. 1990. Contribution of histidine residues to the conformational stability of ribonuclease T1 and mutant Glu-58→Ala. *Biochemistry.* 29:7572–7576.

Meadows, D. H., G. C. Roberts, and O. Jardetzky. 1969. Nuclear magnetic resonance studies of the structure and binding sites of enzymes: 8. Inhibitor binding to ribonuclease. *J. Mol. Biol.* 45:491–511.

Mehler, E. L., and F. Guarnieri. 1999. A self-consistent, microenvironment modulated screened Coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys. J.* 77:3–22.

Mellor, G. W., M. Patel, E. W. Thomas, and K. Brocklehurst. 1993. Clarification of the pH-dependent kinetic behavior of papain by using reactivity probes and analysis of alkylation and catalysed acylation reactions in terms of multihydronic state models: implications for electrostatics calculations and interpretation of the consequences for site-specific mutations such as Asp-158-Asn and Asp-158-Glu. *Biochem. J.* 294:201–210.

Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.

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.

Nielsen, J. E., and G. Vriend. 2001. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK_a calculations. *Proteins.* 43:403–412.

Oda, Y., T. Yamazaki, K. Nagayama, S. Kanaya, Y. Kuroda, and H. Nakamura. 1994. Individual ionization constants of all the carboxyl groups in ribonuclease HI from *Escherichia coli* determined by NMR. *Biochemistry.* 33:5275–5284.

Oda, Y., M. Yoshida, and S. Kanaya. 1993. Role of histidine 124 in the catalytic function of ribonuclease HI from *Escherichia coli. J. Biol. Chem.* 268:88–92.

Oliveberg, M., V. L. Arcus, and A. R. Fersht. 1995. pK_a values of carboxyl groups in the native and denatured states of barnase: the pK_a values of the denatured state are on average 0.4 units lower than those of model compounds. *Biochemistry.* 34:9424–9433.

Onufriev, A., D. A. Case, and G. M. Ullmann. 2001. A novel view of pH titration in biomolecules. *Biochemistry.* 40:3413–3419.

Quirk, D. J., and R. T. Raines. 1999. His. . . Asp catalytic dyad of ribonuclease A: histidine pK_a values in the wild-type, D121N, and D121A enzymes. *Biophys. J.* 76:1571–1579.

Ramanadham, M., L. C. Sieker, and L. H. Jensen. 1981. Structure of triclinic lysozyme and its Cu^{2+} complex at 2 Å resolution. *Acta Crystallogr. A.* 37:33.

Read, R. J., M. Fujinaga, A. R. Sielecki, and M. N. James. 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry.* 22:4420–4433.

Richarz, R., and K. Wüthrich. 1975. Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers.* 17: 2133–2141.

Rico, M., J. Santoro, C. Gonzalez, M. Bruix, and J. L. Neira. 1991. Solution Structure of Bovine Pancreatic Ribonuclease A and Ribonuclease-Pyrimidine Nucleotide Complexes as Determined by ^1H NMR. C. M. Cuchillo, R. de Llorens, M. V. Nogues, X. Pares, editors. Sant Feliu de Guixols, Girona, Spain. 9–14.

Ripoll, D. R., Y. N. Vorobjev, A. Liwo, J. A. Vila, and H. A. Scheraga. 1996. Coupling between folding and ionization equilibria: effects of pH on the conformational preferences of polypeptides. *J. Mol. Biol.* 264: 770–783.

Rocchia, W., E. Alexov, and B. Honig. 2001. Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B.* 105:6507–6514.

Ruterjans, H., and H. Witzel. 1969. NMR-studies on the structure of the active site of pancreatic ribonuclease A. *Eur. J. Biochem.* 9:118–127.

Santoro, J., C. Gonzalez, M. Bruix, J. L. Neira, J. L. Nieto, J. Herranz, and M. Rico. 1993. High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* 229:722–734.

Schaller, W., and A. D. Roberston. 1995. pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. *Biochemistry.* 34:4714–4723.

Schutz, C. N., and A. Warshel. 2001. What are the "dielectric constants" of proteins and how to validate electrostatic models? *Proteins.* 44: 400–417.

Schwalbe, H., S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, and L. J. Smith. 2001. A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci.* 10:677–688.

Sham, Y. Y., Z. T. Chu, and A. Warshel. 1997. Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem.* 101:4458–4472.

Sham, Y. Y., I. Muegge, and A. Warshel. 1998. The effect of protein relaxation on charge-charge interactions and dielectric constant of proteins. *Biophys. J.* 74:1744–1753.

Sham, Y. Y., I. Muegge, and A. Warshel. 1999. Simulating proton translocations in proteins: probing proton transfer pathways in the *Rhodobacter sphaeroides* reaction center. *Proteins.* 36:484–500.

Sheinerman, F. B., R. Norel, and B. Honig. 2000. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10:153–159.

Simonson, T. 1998. Dielectric constant of cytochrome c from simulations in a water droplet including all electrostatic interactions. *J. Am. Chem. Soc.* 120:4873–4878.

Simonson, T. 2001. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* 11:243–252.

Simonson, T., and C. L. Brooks. 1996. Charge screening and the dielectric constant of proteins: insights from molecular dynamics. *J. Am. Chem. Soc.* 118:8452–8458.

Sindelar, C., Z. Hendsch, and B. Tidor. 1998. Effects of salt bridges on protein structure and design. *Protein Sci.* 7:1898–1914.

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.

Skelton, N. J., J. Kordel, and W. J. Chazin. 1995. Determination of the solution structure of Apo calbindin D9k by NMR spectroscopy. *J. Mol. Biol.* 249:441–462.

Smith, L. J., M. J. Sutcliffe, C. Redfield, and C. M. Dobson. 1993. Structure of hen lysozyme in solution. *J. Mol. Biol.* 229:930–944.

Song, J., and J. L. Markley. 2001. NMR chemical shift mapping of the binding site of a protein proteinase inhibitor: changes in the ^{1}H, ^{13}C and ^{15}N NMR chemical shifts of turkey ovomucoid third domain upon binding to bovine chymotrypsin A(alpha). *J. Mol. Recogn.* 14:166–171.

Spassov, V. Z., and D. Bashford. 1999. Multiple site ligand binding to flexible macromolecules separation of global and local conformational change and an iterative mobile clustering approach. *J. Comp. Chem.* 20:1091–1111.

Spassov, V. Z., R. Ladenstein, and A. D. Karshikoff. 1997. Optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins. *Protein Sci.* 6:1190–1196.

Sridharan, S., A. Nicholls, and B. Honig. 1992. A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.* 61:A174.

Sunder-Plassmann, R., and E. L. Reinherz. 1998. A p56lck-independent pathway of CD2 signaling involves Jun kinase. *J. Biol. Chem.* 273: 24249–24257.

Szebenyi, D. M., and K. Moffat. 1986. The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine: molecular details, ion binding, and implications for the structure of other calcium-binding proteins. *J. Biol. Chem.* 261:8761–8777.

Takahashi, T., H. Nakamura, and A. Wada. 1992. Electrostatic forces in two lysozymes: calculations and measurements of histidine pK_a values. *Biopolymers.* 32:897–909.

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

Tashiro, M., and G. T. Montelione. 1995. Structures of bacterial immunoglobulin-binding domains and their complexes with immunoglobulins. *Curr. Opin. Struct. Biol.* 5:471–481.

Thanki, N., J. K. Rao, S. I. Foundling, W. J. Howe, J. B. Moon, J. O. Hui, A. G. Tomasselli, R. L. Heinrikson, S. Thaisrivongs, and A. Wlodawer. 1992. Crystal structure of a complex of HIV–1 protease with a dihydroxyethylene-containing inhibitor: comparisons with molecular modeling. *Protein Sci.* 1:1061–1072.

van der Merwe, P. A., P. N. McNamee, E. A. Davies, A. N. Barclay, and S. J. Davis. 1995. Topology of the CD2-CD48 cell-adhesion molecule complex: implications for antigen recognition by T cells. *Curr. Biol.* 5:74–84.

Van Vlijmen, H. W. T., M. Schaefer, and M. Karplus. 1998. Improving the accuracy of protein pK_a calculations: conformational averaging versus the average structure. *Proteins Struct. Funct. Genet.* 33:145–158.

Vaney, M. C., S. Maignan, M. Ries-Kautt, and A. Ducruix. 1996. High-resolution structure (1.33 Ång) of HEWL Lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpacHab-01 mission. *Acta Crystallogr. Sect. D.* 52:505–517.

Vocadlo, D. J., G. L. Davies, R. Laine, and S. G. Withers. 2001. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature*. 23:835–838.

Voges, D., and A. Karshikoff. 1998. A model of a local dielectric constant in proteins. *J. Chem. Phys.* 108:2219–2227.

Walters, D. E., and A. Allerhand. 1980. Tautomeric states of the histidine residues of bovine pancreatic ribonuclease A: application of carbon 13 nuclear magnetic resonance spectroscopy. *J. Biol. Chem.* 255: 6200–6204.

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. 1999. Simplified methods for pK_a and acid pH-dependent stability estimation in proteins: removing dielectric and counterion boundaries. *Protein Sci.* 8:418–425.

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

Wilson, K. S., B. A. Malcolm, and B. W. Matthews. 1992. Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *J. Biol. Chem.* 267:10842–10849.

Wlodawer, A., L. A. Svensson, L. Sjolin, and G. L. Gilliland. 1988. Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochemistry.* 27:2705–2717.

Wolfenden, R., and A. Radzicka. 1994. On the probability of finding a water molecule in a nonpolar cavity. *Science.* 265:936–937.

Yamazaki, T., T. Hasebe, J. Shouguchi, H. Amano, S. Kajiwara, and K. Shishido. 1997. Structure and function in *Escherichia coli* of plasmids containing pyrimidine/purine-biased stretch originated from the 5′-flanking region of the basidiomycete ras gene. *J. Biochem. (Tokyo).* 122:696–702.

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

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, W., W. A. Hendrickson, R. J. Crouch, and Y. Satow. 1990. Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science.* 249:1398–1405.

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.

Zhou, H., and M. Vijayakumar. 1997. Modeling of protein conformational fluctuations in pK_a predictions. *J. Mol. Biol.* 267:1002–1011.