PROTEINS: Structure, Function, and Genetics 48:388–403 (2002)

Empirical Relationships Between Protein Structure and Carboxyl pK_a Values in Proteins

William R. Forsyth,[1] Jan M. Antosiewicz,[2] and Andrew D. Robertson[1*]
[1]*Department of Biochemistry, University of Iowa, Iowa City, Iowa*
[2]*Department of Biophysics, Warsaw University, Warsaw, Poland*

ABSTRACT **Relationships between protein structure and ionization of carboxyl groups were investigated in 24 proteins of known structure and for which 115 aspartate and 97 glutamate pK_a values are known. Mean pK_a values for aspartates and glutamates are ≤ 3.4 (±1.0) and 4.1 (±0.8), respectively. For aspartates, mean pK_a values are 3.9 (±1.0) and 3.1 (±0.9) in acidic (pI < 5) and basic (pI > 8) proteins, respectively, while mean pK_a values for glutamates are approximately 4.2 for acidic and basic proteins. Burial of carboxyl groups leads to dispersion in pK_a values: pK_a values for solvent-exposed groups show narrow distributions while values for buried groups range from < 2 to 6.7. Calculated electrostatic potentials at the carboxyl groups show modest correlations with experimental pK_a values and these correlations are not improved by including simple surface-area-based terms to account for the effects of desolvation. Mean aspartate pK_a values decrease with increasing numbers of hydrogen bonds but this is not observed at glutamates. Only 10 pK_a values are > 5.5 and most are found in active sites or ligand-binding sites. These carboxyl groups are buried and usually accept no more than one hydrogen bond. Aspartates and glutamates at the N-termini of helices have mean pK_a values of 2.8 (±0.5) and 3.4 (±0.6), respectively, about 0.6 units less than the overall mean values. Proteins 2002;48:388–403.** © 2002 Wiley-Liss, Inc.

Key words: ionization; electrostatic; charge

INTRODUCTION

Ionizable residues in proteins play key roles in binding, catalysis, and protein stability. For example, the kinetics and thermodynamics of ligand binding by proteins can be affected quite substantially by electrostatic interactions involving ionizable residues.[1,2] Most enzymes do chemical work with the assistance of ionizable groups that either act directly as acids, bases, and ligands, or less directly through effects on the electrostatic potential at or near active sites.[3] Ionizable residues also make significant contributions to protein folding and other conformational equilibria. This is perhaps best evidenced by the thermodynamic linkage between ionization and unfolding equilibria in proteins[4,5]: in many cases, the pH dependence of protein stability can be explained entirely on the basis of a small number of ionizable residues with perturbed pK_a values in the native state.[6–10] Recent work suggests that the prediction of such pK_a perturbations is a reasonable basis for identifying active-sites in proteins of known structure.[11,12] Other conformational equilibria in a large and growing number of proteins and protein complexes show biologically significant pH dependencies.[13–19] These pH-dependent changes in conformation are also likely to result from a small number of ionizable residues whose pK_a values differ in the different conformations.

Ionizable residues provide a unique experimental window into the relationship between protein structure and the energetics of noncovalent interactions in proteins (see references[20,21] and references therein). This follows from the fact that multidimensional NMR spectroscopy can be used to determine pK_a values for nearly all of the ionizable residues in small globular proteins. These pK_a values provide quantitative information regarding the interaction between a given ionizable side chain and the rest of the protein.

The structural basis for pK_a values in proteins has been the subject of considerable attention from theoreticians and experimentalists (see reviews[20–24]). Electrostatic interactions involving charged groups are the only noncovalent interactions that can act directly at greater than atomic distances, with experimental evidence suggesting that such distances can exceed 10 Å.[25–32] Electrostatics calculations with proteins of known structure can be used to predict pK_a values. However, the large size of protein molecules, the large number of solvent molecules around proteins, and the fact that most experimental pK_a values are not far from values for model compounds present significant challenges to the precision and accuracy of such predictions.[21,33] Consequently, all electrostatics theory for proteins relies on simplifying assumptions regarding such issues as the distribution of charge within the protein, the polarizability of this distribution, and the extent to which

Grant sponsor: National Institutes of Health; Grant number: GM 46869; Grant sponsor: Polish Committee for Scientific Research; Grant number: BST 718/BF.

William R. Forsyth's present address is Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01655.

*Correspondance to: Andrew D. Robertson, Department of Biochemistry, University of Iowa, Iowa City, IA 52242 USA. E-mail: andy-robertson@uiowa.edu

Received 17 January 2002; Accepted 20 February 2002

protein and solvent are treated as homogeneous vs. heterogeneous media. Much of the recent progress in this area has been made by including more molecular detail, such as protein dynamics and explicit solvent molecules. One of the key measures of progress is the extent of agreement between theory and experimental data. In this regard, pK_a values determined from multidimensional NMR data are one of the few observable in proteins that provide quantitative indicators of electrostatic interactions at multiple sites simultaneously.

The last few years have seen an increase in the number of proteins for which many experimental pK_a values are available and, for a variety of reasons, much of the experimental attention has focused on carboxyl groups. The availability of over 200 carboxyl pK_a values in proteins of known structure presents many opportunities. At a simple level, one may be able to identify empirical relationships between protein structure and pK_a values that would serve as the basis for predicting pK_a values in other proteins. This approach has a promising precedent in studies of protein stability (e.g.,[34–37]). One might also identify trends relating pK_a values and protein function. The large body of data also permits some evaluation of precision and accuracy in the experiments and ways in which these might be improved. This point is closely related to another application of this large database of carboxyl pK_a values, the comparison of experimental results with electrostatic theory.

METHODS

Identification of Structural Features

Potential hydrogen bonds were identified using HBPlus.[38] The geometric criteria for identifying hydrogen bonds were those described by McDonald and Thornton and Baker and Hubbard.[38,39] Solvent-accessible surface (SAS) values were calculated using the algorithm of Lee and Richards[40] as implemented in ACCESS (Scott R. Presnell, Zymogenetics, Seattle, WA). The solvent probe radius was 1.4 Å and a slice width of 0.25 Å was used. In cases where the asymmetric unit consisted of multiple chemically identical polypeptide chains, the SAS of the first chain in the PDB file is reported.

Isoelectric Point

Isoelectric points were calculated using "pI wrapper" at www.embl-heidelberg.de/cgi/pi-wrapper.pl. This calculation used the following pK_a values from studies of model compounds: α-carboxyl, 2.34; α-amino, 9.69; Asp, 3.86; Glu, 4.25; His, 6.00; Cys, 8.33; Tyr, 10.00; Lys, 10.50; and Arg, 12.40.

Electrostatic Potential

Investigation of a possible correlation between experimental pK_a values of acidic residues and the electrostatic potential at their location in the protein was performed as follows. First, the X-ray structure of each protein was supplemented with polar hydrogen atoms at positions generated by the HBUILD facility[41] of CHARMm.[42] Positions of the hydrogens were subsequently optimized by 500 steps of steepest descent energy minimization with CHARMm. During these steps, the topology file and parameter file for CHARMm Version 22 (Polar Hydrogens Only) were used. The resulting files contained deprotonated Glu, Asp, and C-termini, with all other titratable residues protonated, included Cys residues not involved in S-S bonds and all remaining polar hydrogens.

Electrostatic potential was calculated within the framework of the Poisson-Boltzmann model for the solute-solvent system using the finite-difference method,[43] for solving the corresponding Poisson-Boltzmann equation.[44] In the Poisson-Boltzmann model, each charge immersed in a molecule can be considered as a charge in a dielectric cavity surrounded by another dielectric medium, corresponding to the aqueous solvent. This surrounding dielectric medium may also contain mobile electrolyte ions. We used the finite-difference Poisson-Boltzmann (FDPB) algorithm implemented in the UHBD program.[45] The Poisson-Boltzmann calculations were carried out with the atomic charges and radii of the CHARMM22 parameter set,[42] with the modifications described below.

The protein dielectric boundary was taken to be a Richard's probe-accessible surface,[46] computed with a spherical probe of radius of 1.4 Å, and an initial dot-density of 280 per atom.[47] A cubic 60^3 grid, with spacing of 1.5 Å was used. The solvent dielectric constant was set to 80, solute dielectric constant was set to 4, temperature was set to 293 K, ionic strength 150 mM, and ion exclusion radius to 2.0 Å.

For each titratable carboxylic group in a given protein, coordinates of a dummy atom situated in the geometric center of the group (two carbon and two oxygen atoms forming the group) were computed, the charges of the atoms forming the group are set to zero, and the potential was calculated and reported for the position of the dummy atom. Potential calculations were performed under two conditions meant to approximate pH 4 and pH 6 and these involved modifications of His residues and all other carboxylic groups, which had nonzero charges. For conditions mimicking pH 4, the total charge on histidine side chains was set to +1.0e and charges of the two carboxyl oxygen atoms were modified in equal amounts so that the total charge of the carboxylic group was −0.5e. For conditions mimicking pH 6, charges of the two nitrogens on the imidazole ring were modified in equal amounts so that the total charge of the titratable group was +0.5e and the total charge on the carboxyl atoms was set to −1.0e.

RESULTS

Selection of Proteins and Molecular Models

The 24 different proteins included in this study are of known structure and have had multiple carboxyl pK_a values determined by NMR spectroscopy as of June 2001 (Table I).For many of the proteins, multiple models are available from the Protein Data Bank[48]; the choice of models for this study was made primarily on the basis of sequence identity with proteins used in the pK_a determinations. The proteins in this survey include four all α proteins, six β proteins, two α/β proteins, and twelve α + β

TABLE I. List of Proteins, Analyzed PDB Files, and Secondary Structure Analysis[†]

Protein	PDB	NMR/Xtal[a]	Residues	Disulfides	% Helix	% Strand	% Turn	% Other
α-Sarcin	1de3	NMR	150	0	5.3	26.7	44.7	23.3
B1 Domain of Protein G	1pga	Xtal	56	0	26.8	42.9	14.3	16.1
B2 Domain of Protein G	1igd[b]	Xtal	56	0	26.8	46.4	14.3	12.5
Barnase	1a2p	Xtal	108	0	24.1	22.2	25.9	27.8
Basic pancreatic trypsin inhibitor	1bhc	Xtal	56	3	26.8	25.0	10.7	37.5
Bovine pancreatic ribonuclease A	1rnz	Xtal	124	4	22.6	33.1	24.2	20.2
Bull seminal inhibitor IIA	2bus	NMR	57	3	21.1	17.5	14.0	47.4
Calbindin D_{9k}	4icb[c]	Xtal	76	0	63.2	0.0	13.2	23.7
Cardiotoxin A5	1kxi	Xtal	62	4	0.0	37.1	40.3	22.6
CD2d1	1cdc[d]	Xtal	99	0	0.0	47.5	29.3	23.2
Chymotrypsin inhibitor 2	2ci2[e]	Xtal	64	0	21.9	29.7	28.1	20.3
Cryptogein	1beo	Xtal	98	3	62.2	6.1	10.2	21.4
Epidermal growth factor	1egf	NMR	53	3	0.0	20.8	49.1	30.2
Hirudin	1hic	NMR	51	3	0.0	37.3	25.5	37.3
HIV-1 protease/KNI-272 complex	1hpx[f]	Xtal	99	0	7.1	55.6	21.2	16.2
Insulin	1mhi	NMR	51	3	47.1	0.0	35.3	17.6
Lysozyme (Hen)	1lys	Xtal	129	4	41.1	9.3	35.7	14.0
Lysozyme (Turkey)	1lz3	Xtal	129	4	45.0	6.2	34.1	14.7
N-terminal domain of L9	1div[g]	Xtal	56	0	40.4	17.5	24.6	17.5
Ribonuclease H1	2rn2	Xtal	155	0	37.4	29.7	13.5	19.4
Subunit *c* of H^+-transporting F_1F_0ATP synthase	1a91	NMR	79	0	92.4	0.0	3.8	3.8
Thioredoxin (oxidized)	1trs	NMR	105	1	35.2	28.6	29.5	6.7
Thioredoxin (reduced)	1trw	NMR	105	0	40.0	28.6	20.0	11.4
Turkey ovomucoid third domain	2ovo[h]	Xtal	56	3	19.6	17.9	25.0	37.5
Xylanase	1xnb	Xtal	185	0	5.9	63.8	15.7	14.6

[†] Secondary structure analysis performed using DSSP [120].
[a] NMR and Xtal indicate NMR and crystal derived structure, respectively.
[b] Structure contains an additional 5 amino terminal residues relative to protein used for pK_a determination.
[c] PDB file of calcium-loaded protein. Titrations were performed on apoprotein.
[d] Missing coordinates for residues 1–3.
[e] Structure contains an additional 19 amino terminal residues relative to protein used for pK_a determination.
[f] Homodimer, Interchain interactions included in analysis.
[g] Residues 57–149 in PDB file are not present in protein used for pK_a determination.
[h] PDB file of silver pheasant which contains Met at position 18 instead of Leu.

proteins. The polypeptide chains range in size from 30 to 185 residues and 48% contain disulfide bonds. Three pairs of proteins share significant levels of sequence identity: the B1 and B2 domains of protein G (88% identity); chicken and turkey lysozyme (93% identity); and the bull seminal inhibitor IIA and turkey ovomucoid third domain (56% identity). Two sets of independent or semi-independent pK_a determinations have been made for BPTI, bovine RNase A, and turkey ovomucoid third domain (Table II). In the case of thioredoxin, pK_a determinations have been made for both oxidized and reduced protein (Table II).

Efforts have been made to identify molecular models for proteins with sequences that are the same or very similar to those for the proteins used in pK_a determinations (Table I). This turned out to be somewhat challenging in a few cases because of a large number of molecular models with minor sequence differences and uncertainties related to the fact that sequences are not reported explicitly in papers describing the pK_a determinations.

We are aware of at least two additional studies of carboxyl pK_a values that have been published since June 2001: carboxyl pK_a values have been determined for the eight side chain carboxyl groups in the tenth fibronectin type II domain of human fibronectin (FNfn10)[49] and all 12 carboxyl pK_a values have been determined for mammaliam ubiquitin.[50] While revising this paper, we learned that pH-titration studies of the 18 side chain carboxyl groups in T4 lysozyme have been reported in a book chapter.[51]

Determination of pK_a Values by NMR

For all proteins included in this study, apparent pK_a values for carboxyl groups were determined by monitoring the pH dependence of NMR chemical shifts. This widely used approach is generally based on the following observations and assumptions (see[52] and references therein). First, the chemical shift, δ, of an NMR-active nucleus within a given ionizable residue is sensitive only to the ionization state of that residue. Second, chemical shifts of these nuclei at pH values well below and above the apparent pK_a value are independent of pH and represent the chemical shifts of protonated and unprotonated species, δ_{AH} and δ_A, respectively. Third, an observed chemical shift, δ_{obs}, that is intermediate between values for protonated and unprotonated species is a weighted average of these extrema and, consequently, represents the proportions of protonated, AH, and unprotonated species, A.

$$[AH]/([A] + [AH]) = (\delta_{obs} - \delta_A)/(\delta_{AH} - \delta_A) \quad (1)$$

$$[A]/([A] + [AH]) = (\delta_{AH} - \delta_{obs})/(\delta_{AH} - \delta_A) \quad (2)$$

$$[A]/[AH] = (\delta_{AH} - \delta_{obs})/(\delta_{obs} - \delta_A) \quad (3)$$

pK_a values are determined by fitting the pH dependence of chemical shifts to a Henderson-Hasselbalch equation in which the right hand side of Eqn. 3 has been substituted for the ratio of basic to acid species.

$$pH = pK + \log[(\delta_{AH} - \delta_{obs})/(\delta_{obs} - \delta_A)] \quad (4)$$

Eqn. 4 can be rearranged in a variety of ways to yield equations suitable for fitting the pH dependence of δ_{obs} to obtain apparent pK_a values. The following is the most commonly used fitting equation in studies surveyed in the present work.

$$\delta_{obs} = [\delta_{AH} + \delta_A 10^{(pH-pKa)}]/[1 + 10^{(pH-pKa)}] \quad (5)$$

In most cases, the simple titration behavior described by Eqn. 5 yields a satisfactory fit to the NMR data. However, more complex pH dependencies are occasionally observed. Three different approaches to curve fitting have been used to analyze complex pH dependencies. The first uses a modified Hill equation in which the Hill coefficient, n, describes the breadth of the transition relative to a simple titration, i.e., the case where $n = 1$: broader transitions yield values of $n < 1$ and sharper transitions are characterized by values of $n > 1$.[52]

$$\delta_{obs} = [\delta_{AH} + \delta_A 10^{n(pH-pKa)}]/[1 + 10^{n(pH-pKa)}] \quad (6)$$

The second approach explicitly models the complex pH dependence as two or more independent ionization events involving noninteracting residues. This adds a minimum of two fitting parameters to Eqn. 5, one for the chemical shift of the intermediate ionization state and one for the additional pK_a value. This approach has been used to derive pK_a values for Glu 29, Glu 41, Glu 99, and possibly other residues in CD2d1,[53] for Glu 56, Asp 58, Asp 60, and Asp 61 in thioredoxin[54] and for an unknown number of residues in α-sarcin.[55] The third approach takes into account possible interactions among ionizing residues.[56] This approach has been used in fitting titration data for Asp 10 and Asp 70 in RNase HI[57] for Glu 78 and Glu 172 of xylanase.[58,59]

Most pK_a values in these studies were determined by monitoring the pH dependence of proton chemical shifts for the β and γ methylene groups next to the titrating carboxyl groups of aspartate and glutamate, respectively (Table II). Carboxyl pK_a values in cryptogein and the B1 and B2 domains of protein G were determined using both ^{15}N and ^{1}H chemical shifts. ^{13}C chemical shifts of the titrating carboxyl groups were followed in studies of BPTI, calbindin, the N-terminal domain of CD2, HIV protease, RNase HI, thioredoxin, and xylanase (Table II). In general and for a variety of reasons, ^{13}C NMR data probably yield more accurate pK_a values than either ^{1}H or ^{15}N data.[53,57] One example of the virtues of ^{13}C NMR applies to the small but significant number of cases where no titration is observed over a wide range of pH: solvent isotope effects on the carboxyl chemical shift can be used to determine the protonation state of the carboxyl group.[53,59,60]

In most of the studies, a small number of peptide amide hydrogens show large upfield changes in ^{1}H chemical shift, on the order of 1 ppm, with decreasing pH.[7,9,10,53,55,61–70] These backbone amide protons are often located in nontitratable residues and probably reflect the involvement of these protons in hydrogen bonds with titrating carboxyl groups.[71,72] In fact, the pH dependence of the amide proton chemical shifts is often consistent with pK_a values for carboxyl groups that have been identified as hydrogen bond acceptors in the protein structures. While the precise molecular basis for these pH-dependent chemical shifts is not yet known, the magnitude of the chemical shift changes probably contains useful information regarding changes in hydrogen bonding upon titration of the carboxyl group.[73–76]

A minor but significant issue in any NMR study is the referencing of chemical shifts.[75] The IUPAC and IUBMB have adopted a set of standards for chemical shift referencing in aqueous solutions.[77] For ^{1}H and ^{13}C, referencing should be relative to perdeuterated 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) while the reference for ^{15}N is liquid ammonia. ^{15}N chemical shift can also be referenced indirectly using DSS and knowledge of the absolute frequencies for ^{1}H and ^{15}N on the NMR spectrometer.[75,77] Until recently, however, NMR pK_a determinations were often made using perdeuterated 3-trimethylsilyl-propionic acid (TSP) as the chemical shift standard and the chemical shift for TSP has a small amplitude pH dependence.[78]

Precision and Accuracy of pK_a Values

The precision of pK_a values determined by NMR is partially reflected in the reported fitting errors. In most cases, these uncertainties are about 0.1 pH unit. A more complete assessment of precision includes consideration of the magnitude of the change in δ_{obs}, number of data points describing low and high pH plateaus and the transition, resolution in the NMR spectrum, and the precision of pH measurements. In our estimation, the magnitude of the change in δ_{obs} and the number of data points usually play the biggest role in determining the overall precision. ^{13}C chemical shifts tend to show larger changes with pH than do ^{1}H chemical shifts. The number of data points used to estimate pK_a values ranges from a low of about 5 or 6 in a few cases to well over 10 for most studies (Table II).

Factors affecting the accuracy of pK_a determinations by NMR include the accuracy of resonance assignments and pH determinations, the range of experimental pH values relative to the pK_a value, pH-dependent perturbations of chemical shifts not arising from titration of the nearest carboxyl group, variations in ionic strength during a titration study, and the reversibility of the titration. Ideally, the overall accuracy would be best evaluated by comparing independent determinations under the same solution conditions made by different investigators. Multiple determinations have been made for BPTI, RNase A, and OMTKY3 (Table II). However, the pairs of independent studies for both BPTI and RNase A were done under

TABLE II. Asp and Glu pK_a Values With Experimental Details†

Protein	Residue	pK_a	Reference[a]	Protein	Residue	pK_a	Reference[a]
α-Sarcin	Asp 9	3.9	[55]	Hirudin	Asp 5	4.3	[62]
	Asp 41	< 3	2 mM protein,	(desulfatohirudin:	Asp 33	4.2	6 mM protein,
	Asp 57	4.3	0.2 M NaCl, 35°C,	natural protein	Asp 53	3.8	22°C, 10% D_2O,
	Asp 59	4.1	10% D_2O, pH 3.0–8.5,	contains sulfate at	Asp 55	4.1	pH 1.77–6.75,
	Asp 75	3.9	0.5 unit intervals,	Y63)	Glu 8	4.3	0.4–1 unit intervals,
	Asp 77	< 3	H-1, reversibility NR		Glu 17	3.8	H-1, reversibility NR
	Asp 85	3.8			Glu 35	4.3	
	Asp 91	< 3			Glu 43	4.2	
	Asp 102	< 3			Glu 57	4.6	
	Asp 105	< 3			Glu 58	4.7	
	Asp 109	3.7			Glu 61	4.5	
	Glu 19	4.6			Glu 62	4.5	
	Glu 31	4.6					
	Glu 96	5.1		HIV-1 protease/KNI-272	Asp 25	> 6.2	[121]
	Glu 115	4.9		complex (HXB2	Asp 29	3.2	0.9 mM dimer,
	Glu 140	4.3		isolate, C67A/C95A)	Asp 30	3.9	50 mM Na-acetate, 45°C,
	Glu 144	4.3			Asp 60	3.0	97% D_2O, pH 2.5–6.2,
					Asp 125	< 2.5	0.6–1 pH unit intervals,
B1 domain of protein G (IgG	Asp 22	2.9	[64]		Asp 129	3.7	C-13, reversibility NR
binding domain)	Asp 36	3.8	3.4 mM protein,		Asp 130	3.8	
	Asp 40	4.0	0.1 M Na-acetate, 25°C,		Asp 160	3.0	
	Asp 46	3.6	10% D_2O, pH 1.5–7.0,				
	Asp 47	3.4	0.25 unit intervals,	Insulin (human, Asp	Asp (B9)	2.6	[66]
	Glu 15	4.4	H-1, some N-15,	(B9) mutation)	Glu (A4)	2.6	1.7 mM protein,
	Glu 19	3.7	reversibility checked		Glu (A17)	> 3.7	22°C, 10% D_2O,
	Glu 27	4.5			Glu (B13)	2.2	pH 1.73–3.93,
	Glu 42	4.4			Glu (B21)	3.7	0.2–1 pH unit intervals,
	Glu 56	4.0					H-1, reversibility NR
B2 domain of protein G (IgG	Asp 22	2.9	[64]	Lysozyme (Hen; HEWL)	Asp 18	2.7	[122]
binding domain)	Asp 36	3.9	3.4 mM protein,		Asp 48	< 2.5	2 mM protein,
	Asp 40	4.4	0.1 M Na-acetate, 25°C		Asp 52	3.7	0.1 M NaCl, 35°C,
	Asp 46	3.6	10% D_2O, pH 1.5–7.0,		Asp 66	< 2.0	10% D_2O, pH 1–7,
	Asp 47	3.4	0.25 unit intervals		Asp 87	2.1	0.5 pH unit intervals
	Glu 15	4.3	H-1, some N-15,		Asp 101	4.1	H-1, reversibility checked
	Glu 24	4.2	reversibility checked		Asp 119	3.2	
	Glu 27	4.6			Glu 7	2.9	
	Glu 56	4.2			Glu 35	6.2	
Barnase	Asp 8	2.9	[7]	Lysozyme (Turkey)	Asp 18	2.7	[122]
	Asp 12	3.8	2–4 mM protein, 30°C		Asp 48	< 2.5	2 mM protein,
	Asp 22	3.3	10% D_2O, pH 2.2–5.9,		Asp 52	3.8	0.1 M NaCl, 35°C
	Asp 44	3.4	0.2–1 unit intervals		Asp 66	< 2.0	10% D_2O, pH 1–7,
	Asp 54	≤ 2.2	H-1, reversibility NR		Asp 87	2.1	0.5 pH unit intervals
	Asp 75	3.1	E73, D101 titrations		Asp 119	3.4	H-1, reversibility checked
	Asp 86	4.2	coincide with unfolding		Glu 7	2.70	
	Asp 93	< 2			Glu 35	6.1	
	Asp 101*	≤ 2					
	Glu 29	3.8		N-terminal domain of	Asp 8	3.0	[10]
	Glu 60	3.0		L9 (residues 1–56)	Asp 23	3.1	2 mM protein,
	Glu 73*	≤ 2.1			Glu 17	3.6	10 mM Na-phosphate,
					Glu 38	4.0	0.1 M NaCl, 25°C,
Basic pancreatic trypsin	Asp 3	3.4	[123]		Glu 48	4.2	10% D_2O, pH 1.8–7,
inhibitor (BPTI)	Asp 50	3.1	25–50 mM protein 35°C		Glu 54	4.2	0.5 pH unit intervals
	Glu 7	3.7	100% D_2O, pH 1–10.5,				H-1, reversibility checked
	Glu 49	3.6	0.75 unit intervals.				
	Asp 3	3.6	[124]				
	Asp 50	3.2	22 mM protein,				
	Glu 7	3.9	0.1 M NaCl, 25°C,*				
	Glu 49	4.0	H_2O, pH 1.78–10.41,				
			0.3 unit intervals,				
			C-13, reversibility NR				
Bovine pancreatic	Asp 14	< 2.0	[67]	Ribonuclease H1	Asp 10*	6.1	[57]
ribonuclease A (RNase A)	Asp 38	3.5	8 mM protein,		Asp 70*	2.6	0.7–1.5 mM protein,
	Asp 53	3.9	0.2 M NaCl, 35°C		Asp 94	3.2	0.1 M NaCl, 27°C,
	Asp 83	3.5	pH 1–9,		Asp 102	< 2	99.9% D_2, pH 2–7.8,
	Asp 121	3.1	unknown unit intervals,		Asp 108	3.2	0.3 unit intervals,
	Glu 2	2.8	H-1, reversibility NR		Asp 134	4.1	C-13, reversibility NR
	Glu 9	4.0			Asp 148	< 2	
	Glu 49	4.7			Glu 6	4.5	
	Glu 86	4.1			Glu 32	3.6	
	Glu 111	3.5			Glu 48	4.4	
					Glu 57	3.2	
	Asp 14	1.8	[68]		Glu 61	3.9	
	Asp 38	2.1	3–5 mM protein, 30°C,		Glu 64	4.4	
	Asp 53	3.7	10% D_2O, pH 1.2–7.9,		Glu 119	4.1	
	Asp 83	3.3	0.3–0.7 unit intervals,		Glu 129	3.6	

TABLE II. (Continued)

Protein	Residue	pK_a	Reference[a]	Protein	Residue	pK_a	Reference[a]
Bovine pancreatic ribonuclease A (RNase A)	Asp 121	3.0	H-1, reversibility NR	Ribonuclease H1	Glu 131	4.3	
	Glu 2	2.6			Glu 135	4.3	
	Glu 9	NR			Glu 147	4.2	
	Glu 49	4.3			Glu 154	4.4	
	Glu 86	4					
	Glu 111	NR		Subunit *c* of H^+-transporting F_1F_0ATP synthase	Asp 7	5.6	[125]
					Asp 44	5.6	1.5–2 mM protein,
Bull seminal inhibitor IIA (BUSI IIA)	Asp 6	4.0	[61]		Asp 61	7.0	0.05 M NaCl, 25°C
	Asp 12	3.6	7 mM protein,		Glu 2	5.5	$CDCl_3/CD_3OD/D_2O$
	Glu 9	4.3	0.1 M NaCl, 18°C,		Glu 37	5.5	(4/4/1),
	Glu 20	4.1	10% D_2O, pH 3.2–5.4,				pH 3.3–8.0,
			0.2–0.6 unit intervals,				0.1–0.3 unit intervals,
			H-1, reversibility NR				H-1, reversibility checked
Calbindin D_{9k} (P43G variant)	Asp 47	3.0	[79]				
	Glu 4	3.8	0.5 mM protein, 37°C	Thioredoxin (human mutant; C62A/C69A/ C73A; oxidized)	Asp 16	4.2	[54]
	Glu 5	3.4	10% D_2O, pH 2.4–7.6,		Asp 20	3.8	1 mM protein, 25°C,
	Glu 11	4.7	0.3 unit intervals,		Asp 26	8.1	0.1 M Na-phosphate,
	Glu 17*	3.6	C-13, partial reversibility		Asp 58*	4.0	100% DO_2?, pH 1.9–9.0,
	Glu 26	4.1	check		Asp 60*	3.3	0.3–0.8 unit intervals,
	Glu 48	4.6			Asp 61*	4.3	C-13, reversibility NR
	Glu 64	3.8			Asp 64	3.2	
					Glu 6	4.9	
Cardiotoxin A5	Asp 42	3.2	[65]		Glu 13	4.4	
	Asp 59	< 2.3	4 mM protein, 25°C,		Glu 47	4.3	
	Glu 17	4.0	10% D_2O, pH 2–7,		Glu 56*	3.3	
			0.1–0.2 unit intervals,		Glu 68	5.1	
			H-1, reversibility NR		Glu 70	4.8	
					Glu 88	3.6	
CD2d1 (N-terminal domain of rat CD2)	Asp 2	3.5	[53]		Glu 95	4.1	
	Asp 25	3.5	1.2 mM protein, 25°C,		Glu 98	3.9	
	Asp 26	3.6	20 mM K-phosphate,		Glu 103	4.5	
	Asp 28	3.6	0.5 mM EDTA,				
	Asp 62	4.1	0.1 mM PMSF,	Thioredoxin (human mutant; C62A/C69A/ C73A; reduced)	Asp 16	4.0	[54]
	Asp 71	3.2	10% D_2O, pH 2.0–10.0,		Asp 20	3.8	1 mM protein, 25°C,
	Asp 72	4.1	0.2–0.3 unit intervals,		Asp 26	9.9	0.1 M Na-phosphate,
	Asp 94	3.9	C-13, reversibility NR		Asp 58*	3.6	100% D_2O?, pH 1.9–9.0,
	Glu 29	4.4			Asp 60*	3.3	0.3–0.8 unit intervals,
	Glu 33	4.2			Asp 61*	4.3	C-13, reversibility NR
	Glu 41	6.7			Asp 64	3.2	
	Glu 56	3.9			Glu 6	4.8	
	Glu 99	4.2			Glu 13	4.4	
					Glu 47	4.1	
					Glu 56*	3.3	
					Glu 68	4.9	
					Glu 70	4.6	
					Glu 88	3.7	
					Glu 95	4.1	
					Glu 98	3.9	
					Glu 103	4.4	
Chymotrypsin inhibitor 2 (Barley; C12)	Asp 23	2.4	[9]	Turkey ovomucoid third domain (residues 1–56; OMTK Y3)	Asp 7	2.5	[69]
	Asp 45	3.6	4–6 mM protein,		Asp 27	2.2	5 mM protein,
	Asp 52	2.5	no salt, 27°C,		Glu 10	4.1	10 mM KCl, 25°C
	Asp 55	5.0	10% D_2O, pH 2.05–5.91,		Glu 19	3.2	10% D_2O, pH 1.8–7.0,
	Glu 4	2.9	0.2–1 unit intervals,		Glu 43	4.8	0.2–0.5 unit intervals,
	Glu 7	2.9	H-1, reversibility NR				H-1, reversibility checked
	Glu 14	3.5					
	Glu 15	2.8					
	Glu 26	3.6		Turkey ovomucoid third domain recombinant; residues 5–56; rOM3)	Asp 7	< 2.6	[70]
	Glu 41	3.1			Asp 27	< 2.3	1 mM protein,
					Glu 10	4.1	10 mM KCl 25°C,
Cryptogein	Asp 21	2.5	[126]		Glu 19	3.2	10% D_2O, pH 0.9–7.5,
	Asp 30	2.5	1 mM protein, 40°C		Glu 43	4.8	0.2–0.4 unit intervals,
	Asp 72	2.6	I = 0.045, variety of buffers,				H-1, reversibility checked
			10% D_2O, pH 1.5–11.2,				
			0.5 unit intervals,	Xylanase	Asp 4	3.0	[59]
			H-1 N-15, reversibility NR		Asp 11	2.5	0.5–0.75 mM protein,
					Asp 83	< 2.0	3 mM NaN_3,
Epidermal growth factor (mouse; EGF)	Asp 11	3.9	[127]		Asp 101	< 2.0	25 mM NaPi, 25°C,
	Asp 27	4.0	3.3 mM protein, 28°C,		Asp 106	2.7	10% D_2O, pH 2.06–8.32,
	Asp 40	3.6	99.95% D_2O, pH 1.59–8.79,		Asp 119	3.2	0.2–0.5 unit intervals,
	Asp 46	3.8	0.4–1.5 unit intervals,		Asp 121	3.6	C-13, reversibility NR
	Glu 24	4.1	H-1, reversibility NR		Glu 78	4.6	
	Glu 51	4			Glu 172	6.7	

* Asterisks denote pK_a values derived from transitions that do not conform to a simple Henderson-Hasselbalch relationship.

[a] References for and experimental conditions under which the pK_a values were determined. H-1, C-13, and N-15 indicate nucleus monitored. Unit intervals indicate pH increments used over the titration. Reversibility and reversibility NR (reversibility not reported) designate whether a post-titration experiment was performed to assess reversibility and reproducibility of the chemical shifts.

different solution conditions. In BPTI, differences in the four side chain carboxyl pK_a values range from 0.1 to 0.4 U (pH units). Two determinations are available for 8 out of the 10 side chain carboxyl groups in RNase A. Six of these differ by no more than 0.2 pH units. The larger differences of 0.4 pH units at Glu 49 and 1.4 pH units at Asp 38 are attributed to the particular sensitivity of these residues to differences in solution conditions.[68] In the case of OMTKY3, two determinations have been made under the same solution conditions on slightly different forms of the protein and the 5 side chain carboxyl pK_a values differ by no more than 0.2 pH units. Similar levels of agreement are also seen in the highly homologous B1 and B2 domains of protein G and in the two lysozymes (Table II).

At least two studies have identified significant protein concentration dependencies for pK_a values.[53,79] Given that protein concentrations in Table ii range from 0.1 to at least 25 mM, this is a significant issue. In the study of lysine residues in calbindin, the pK_a differences were interpreted as evidence of the nonnegligible effects of protein molecules as mobile ions in solution.[79] In the case of the N-terminal domain of CD2, differences in carboxyl pK_a values were attributed to possible self-association at higher protein concentrations. Another possible interpretation is that ionic strength is varying in different ways at different protein concentrations.[80] In going from neutral to acidic pH, the addition of titrant leads to increases in ionic strength. For a protein with 15 titratable groups, titration to acid pH of all of these groups in a 1-mM protein solution will require the addition of at least 15 mM titrant. This effect will be proportional to the protein concentration. Other possible complications include specific binding of titrant or its counterion and the presence of other salts that might not be removed in the final stages of protein purification. These possibilities are usually not addressed in the NMR studies.

About half of the studies report some assessment of reversibility in the pH titrations (Table II). Covalent modifications of proteins are possible during the many hours or days needed to acquire the titration data and some of these involve introduction or modification of a charge.[81–84] If such modifications are occurring during titration, then pK_a values could change during the experiment. Consequently, some uncertainty in accuracy should be assigned to pK_a values derived from studies in which reversibility or chemical integrity are not investigated.

Distribution of Carboxyl pK_a Values

The distributions of pK_a values for over 200 aspartate and glutamate carboxyl groups are presented as histograms in Figure 1. The mean values for aspartate and glutamate in proteins are $\leq$ 3.4 ($\pm$1.0) and 4.1 ($\pm$0.8), respectively; the former is an upper limit because some of the aspartate pK_a values are $<$ 2. These are similar to values identified previously in a smaller data set[33] and less than typical model compound values. The reference or model compound pK_a values for side chain carboxyl groups in oligopeptides are most often quoted as 3.9 to 4.0 for aspartate and 4.3 to 4.5 for glutamate.[85] These ranges are

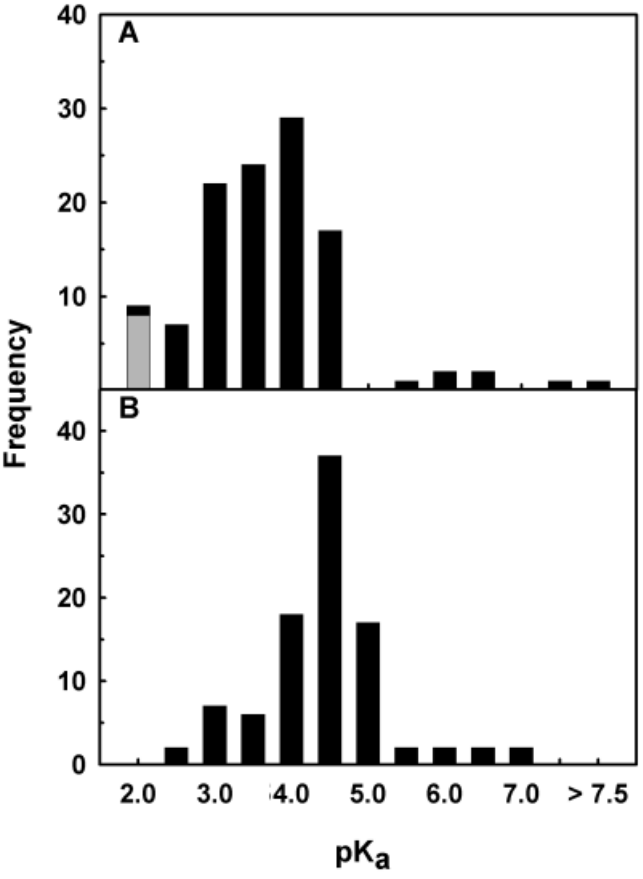

Fig. 1. Distributions of (**A**) aspartate and (**B**) glutamate carboxyl pK_a values in proteins. Each column entry represents the count for pK_a values, in 0.5-pH unit increments, that are less than or equal to the indicated value. The gray bar in A represents pK_a values $<$ 2.

derived from a variety of studies carried out under a variety of solution conditions. A more recent study examined pK_a values at 25°C for aspartate and glutamate in oligopeptides derived from the N-terminal domain of L9.[10] In 0.1 M NaCl, the two aspartates had pK_a values of 3.8 and 4.1 while pK_a values for the four glutamates ranged from 4.1 to 4.6. Both asparate pK_a values were 3.8 in 0.75 M NaCl and pK_a values for glutamates were 4.1 to 4.4. Given these results and the range of salt concentrations represented in Table II, we estimate that appropriate ranges of model compound values are 3.8 to 4.1 for aspartate and 4.1 to 4.6 for glutamate. These ranges thus reflect both local sequence effects and different solution conditions for the different model compound studies.

The mean pK_a value for glutamate in proteins thus falls at the low end of model compound values and the mean pK_a for aspartate is 0.4 pH units less than the smallest model compound value; note that 8 of the aspartate pK_a values are $<$ 2. Only 15 pK_a values are greater than 5 and, in a particularly striking relationship, all but one of the pK_a values that are $>$ 5.5 falls in enzyme active sites. The exception is Glu 41 of the N-terminal domain of CD2. This residue is primarily responsible for the pH dependence of

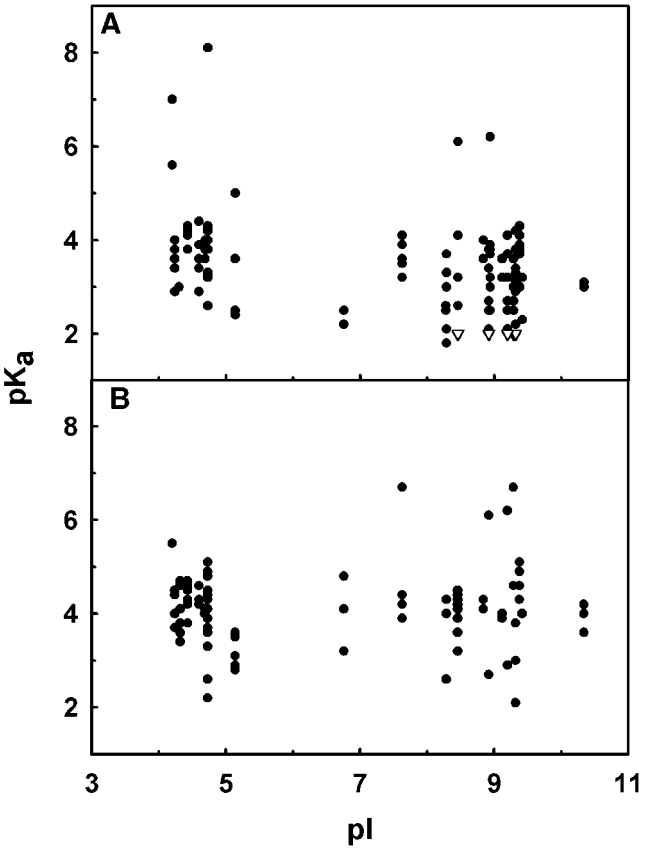

Fig. 2. pK_a values for (**A**) aspartate and (**B**) glutamate vs. calculated pI values in proteins. The open triangles represent the eight pK_a values that are < 2; some of these overlap with one another.

self-association but the biological significance of this phenomenon is unclear.[53]

The distributions for aspartate and glutamate have distinctly different shapes although both are skewed to low pK_a values. Glutamate pK_a values show a more normal distribution while aspartate pK_a values show a broader distribution that is skewed more significantly to low values. In general, negative charges on both glutamate and aspartate side chains are stabilized by protein structure, and this effect is most pronounced for asparate residues. The possible structural basis for these distributions is addressed in the following sections.

pK_a and pI Values

One simple hypothesis to explain perturbed carboxyl pK_a values is that carboxyl groups respond to a protein's net charge: pK_a values in positively charged proteins should be less than pK_a values in negatively charged proteins.[80,86,87] If indeed the net charge on a protein influences carboxyl pK_a values, then one would expect that carboxyl pK_a values will be greater in proteins with low pI values than in proteins with high pI values. This expectation is realized for aspartate pK_a values in Figure 2, with mean pK_a values of 3.9 ($\pm$1.0) and $\leq$ 3.1 ($\pm$0.9) for proteins with pI values < 5 and > 8, respectively; all of the aspartate pK_a values < 2 are found in basic proteins, so 3.1 is an upper limit on the mean pK_a value in these proteins. However, no relationship is observed between pI and glutamate pK_a values: mean pK_a values for glutamates are 4.2 in both acidic and basic proteins.

Considerations for Structural Analysis

While investigating the possible relationships between protein structure and carboxyl pK_a values, we encountered difficulties in interpreting structures determined by NMR. Most of these difficulties were attributable to uncertainties in positioning of side chains for aspartate and glutamate. These uncertainties may reflect the real dynamics of these side chains, a lack of NMR observables regarding the position of these side chains, or both. Atoms in carboxyl groups are usually not observed directly in NMR experiments and this introduces significant uncertainties into positioning these groups in protein structures. For the subsequent structural analysis, we have thus chosen to focus on the 16 proteins for which X-ray structures are available. For these proteins, pK_a values are known for 79 aspartate and 61 glutamate residues.

The observed pK_a values reflect a balance between desolvation, which increases carboxyl pK_a values relative to model compounds, and interactions with charges and dipoles within the proteins. Most carboxyl pK_a values in proteins are within 1 pH unit of model compound values, suggesting that these interactions tend to offset the effect of desolvation. Three simple properties of proteins that are related to desolvation, charge-charge, and charge-dipole interactions, respectively, are solvent-accessible surface area (SAS), calculated electrostatic potential (EP), and intramolecular hydrogen bonds. Possible correlations between these properties and the experimental carboxyl pK_a values are explored below.

pK_a Values and Solvent Accessibility

Solvation plays a key role in ionization equilibria. In general, desolvation associated with placing an ionizable group in a protein will destabilize the charged species relative to the neutral species. In the absence of interactions with other charges and dipoles, desolvation of carboxyl groups through burial in protein structure should lead to carboxyl pK_a values that are elevated by many units.[21] Previous studies and the data in Figure 3 show that the relationship between solvation, as reflected here in solvent-accessibility, and pK_a values in proteins is more complicated (e.g.,[21, 88]). In fact, mean pK_a values tend to show modest declines with increasing burial in protein structure. This is probably due to favorable interactions between the charged carboxyl groups and buried polar groups in the protein.[89–91]

SAS of carboxyl groups is strongly correlated with the dispersion in pK_a values (Fig. 3). About 30% of the glutamate carboxyl oxygens have a combined SAS > 40 Å^2 and the pK_a values for these groups are distributed narrowly about the mean pK_a value of 4.1 ($\pm$0.3). With the exception of Asp 38 in RNase A, a somewhat similar pattern is observed at aspartates that have SAS values $\geq$ 45 Å^2, where the mean pK_a value is 3.5 ($\pm$0.3), not

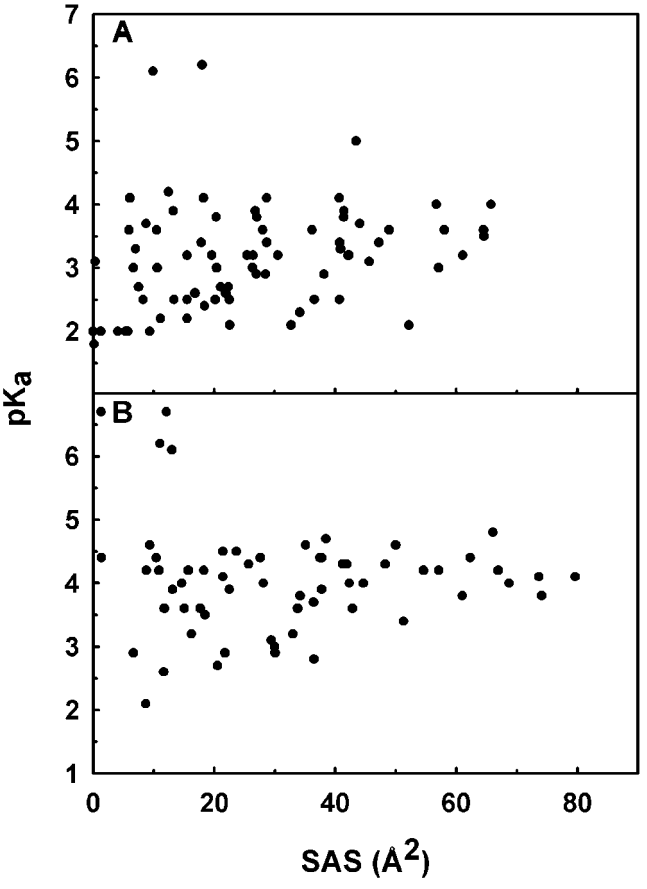

Fig. 3. pK_a values for (**A**) aspartate and (**B**) glutamate in proteins vs. solvent-accessible surface area (SAS) of the two carboxyl oxygen atoms. SAS values for the side chain carboxyl oxygens of aspartate and glutamate in Ala-Asp-Ala and Ala-Glu-Ala tripeptides with extended conformations are 65.5 and 76.9 $Å^2$, respectively.

including Asp 38. However, only 14% of the aspartate carboxyl groups belong to this group. If a threshold of 40 $Å^2$ is used instead, then 26% of the aspartate carboxyl groups are included but the resulting standard deviation, 0.5, about the mean value of 3.5 is nearly twice as large. In the case of Asp 38 in RNase A, the carboxyl group is well exposed in the 1.9 $Å^2$ crystal structure used for this survey but partially buried in the 0.87 $Å^2$ structure of a single amino acid variant (PDB file 1dy5). Increasing burial of carboxyl groups leads to increasing dispersion in pK_a values: pK_a values for the 48 carboxyl groups with SAS $<$ 20 $Å^2$ range from less than 2 to 6.7. Interestingly, all pK_a values $>$ 5.5 are for carboxyl groups with accessible surface areas that are $\leq$ 20 Å.

pK_a Values and Electrostatic Potential

In keeping with the goal of trying to identify simple relationships between protein structure and carboxyl pK_a values, we explored the relationship between calculated electrostatic potentials and experimental carboxyl pK_a values. In principle, these pK_a values might be correlated with pK_a values calculated using electrostatic models such as UHBD. However, calculation of pK_a values based on protein structure generally involves a multi-step thermodynamic cycle, many judgments regarding selection of parameters, and multiple rounds of iteration.[88,90,92,93] In contrast, calculation of electrostatic potentials alone is relatively straightforward, accessible to a large number of investigators and, consequently, one of the most popular applications with new molecular models for proteins. Ideally, the electrostatic potential contributed by protein charges (EP; in kcal/mol e^-) should have the following effect on the observed pK_a for a carboxyl group, pK_{obs}:

$$pK_{obs} = pK_{mod} - EP/2.3RT \quad (7)$$

where pK_{mod} is the pK_a for model compounds, R is the gas constant (1.987×10^{-3} kcal $\cdot$ $mol^{-1} \cdot K^{-1}$) and T is the absolute temperature. Perfect agreement of experimental pK_a values with Eqn. 7 would yield intercepts, pK_{mod}, at pK_a values near those for model compounds and slopes of $-(1/2.3RT)$ or about -0.73 $kcal^{-1}$ mol at 298 K. Such agreement was not anticipated because this treatment undoubtedly ignores important contributions to pK_a perturbations from desolvation, polarization, and the reaction field resulting from introduction of charge near a dielectric boundary.[94,95] However, given the major role played by the EP in predicted pK_a values, one might expect to detect a significant correlation between the calculated potentials and experimental pK_a values. Moreover, these correlations may show significant improvements if the energetic effects of desolvation can be captured in suitable additional terms for solvent-accessible surface areas. Finally, and as mentioned above, calculated potentials of this sort are a very popular application of electrostatics theory. Consequently, the correlations reported here serve as a valuable illustration of the extent to which such potentials alone accurately reflect the energetics of charges at protein surfaces.

For calculation of potentials, pH 4 and pH 6 have been roughly approximated by the use of simple combinations of charge states on carboxyl and imidazole groups, as described in Methods. The results at both pH values are similar and only those at pH 6 are reported in Figure 4. Plots of experimental pK_a values vs. the calculated electrostatic potential, in kcal/mol e-, show considerable scatter, about $\pm$1 pH unit, and the correlation coefficients are very small, but the expected trend of decreasing pK_a values with increasing potential is observed for aspartate and glutamate (Fig. 4). Interestingly, the intercepts from linear fits fall at about 3.9 for aspartate and 4.3 for glutamate, in good agreement with expectations based on model compounds. However, the slopes range from -0.06 to -0.14 and are thus about 10 to 20% of the values expected if the potential was the only factor perturbing pK_a values. The mean calculated potentials for aspartates and glutamates are 4.8 kcal/mol e- and 2.8 kcal/mol e-, respectively, which is qualitatively consistent with the observation that aspartate pK_a values are generally less than glutamate pK_a values.

The scatter plots in Figure 4 suggest that other factors play a significant role in modulating carboxyl pK_a values in proteins. Among the other factors, desolvation should

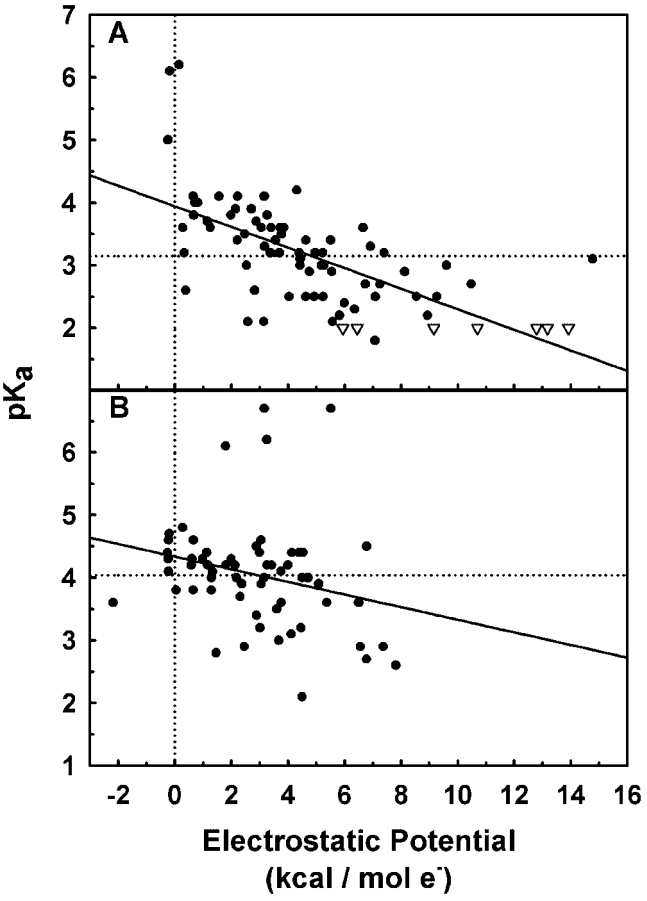

Fig. 4. pK_a values for (**A**) aspartate and (**B**) glutamate as a function of the calculated electrostatic potential at pH 6. The open triangles represent the eight pK_a values that are < 2; some of these overlap with one another. Electrostatic potentials were calculated as described in Methods. The lines are the result of linear regression. The fitted intercepts and slopes are 3.9 and 0.14, respectively, in A and 4.3 and 0.09, respectively, in B. The correlation coefficients are 0.51 and 0.24 for the linear regressions in A and B, respectively.

make a major contribution. We hypothesized that the effect of desolvation was, in some way, proportional to SAS and that inclusion of a term based on this proportionality would improve the correlation between pK_a values and electrostatic potentials. A precedent for such an approach can be found in modified versions of the Tanford-Kirkwood model for protein electrostatics,[20,96] where interaction energies between charged groups are scaled according to the degree of solvent-accessibility of these groups. However, we were unable to improve correlations between pK_a values and calculated potentials by including a term for SAS, regardless of whether SAS was included in linear, exponential, or reciprocal form.

pK_a Values and Intramolecular Hydrogen Bonds

Hydrogen bonds can play a major role in modulating pK_a values in proteins.[89,90] Suitable hydrogen bond donors within proteins can stabilize negatively charged carboxyl groups while the presence of a well-placed hydrogen bond acceptor could conceivably stabilize the neutral form. The

TABLE III. Analysis of pK_a Values and All Hydrogen Bonds†

	Number of Hydrogen Bonds				
	0	1	2	3	4
Aspartates					
Quantity	24	18	23	11	2
Average pK_a	3.5 (0.7)	3.6 (1.0)	2.9 (0.6)	< 2.3 (0.6)	< 2.0
Glutamates					
Quantity	24	18	18	1	0
Average pK_a	4.1 (0.6)	4.0 (1.1)	3.9 (1.0)	3.6	—

†Hydrogen bonds in the crystal structures were identified using HBPlus.[38]

results in Table III suggest that hydrogen bonding and pK_a values for aspartate are related, at least when two or more hydrogen bonds are present: mean pK_a values for aspartate carboxyl groups accepting 0, 1, 2, and 3 hydrogen bonds are 3.5, 3.6, 2.9, and < 2.4, respectively. However, this conclusion is somewhat tentative given that most of these values overlap at one standard deviation (Table III). In contrast, the number of hydrogen bonds does not appear to be related to pK_a values for glutamates (Table III).

Aspartate pK_a values show correlations with both pI values (Fig. 2) and intramolecular hydrogen bonding (Table III). To investigate the extent to which these phenomena are independent of one another, correlations between aspartate pK_a values and hydrogen bonding were explored in acidic (pI < 5) and basic (pI > 8) proteins with known X-ray structures (Table IV).As expected, glutamate pK_a values continue to show no detectable relationship to pI values and intramolecular hydrogen bonding. For aspartates that are not involved in intramolecular hydrogen bonds, the mean pK_a value is 4.1 in acidic proteins and 3.3 in basic proteins. Mean aspartate pK_a values in acidic and basic proteins tend to decrease with increasing numbers of hydrogen bonds, but most of these values overlap at one standard deviation. The results suggest that relationships of pI values and hydrogen bonding to aspartate pK_a values may not be completely independent, although this conclusion is again somewhat tentative given the large standard deviations and relatively small number of known aspartate pK_a values in acidic proteins (Table IV).

Intermolecular hydrogen bonding between carboxyl groups and water molecules is likely to be an important determinant of carboxyl pK_a values in proteins.[97] Attempts to quantify such hydrogen bonding in the X-ray structures listed in Table I revealed that the extent to which water molecules were included in the structure refinement process was highly variable. Consequently, this analysis was not pursued further.

pK_a Values and Helical Structure

The concept of a helix macrodipole and the related concept of helix capping suggest that the location of ionizable residues in helices may have an impact on pK_a values at these residues.[98–105] The expectations are that aspartate and glutamate residues in helical structure will

TABLE IV. Influence of Isoelectric Point on pK_a Values and Hydrogen Bonds[†]

	Number of Hydrogen Bonds				
	0	1	2	3	4
Aspartates					
Quantity (pI < 5)	3	4	4	0	0
Average pK_a (pI < 5)	4.1 (0.3)	3.5 (0.4)	3.2 (0.3)	—	—
Quantity (pI > 8)	17	11	14	11	2
Average pK_a (pI > 8)	3.3 (0.6)	3.7 (1.3)	2.9 (0.6)	< 2.4 (0.6)	≤ 1.9 (0.1)
Glutamates					
Quantity (pI < 5)	8	5	3	0	0
Average pK_a (pI < 5)	4.3 (0.3)	3.9 (0.5)	4.2 (0.3)	—	—
Quantity (pI > 8)	11	7	12	1	0
Average pK_a (pI > 8)	4.2 (0.7)	4.0 (1.2)	4.0 (1.1)	3.6	—

[†] Hydrogen bonds in the crystal structures were identified using HBPlus.[38]

tend to be found at or near the N-termini of helices and that pK_a values for residues at N-termini will be less than values at C-termini of helices. For the proteins in this study and for which X-ray structures are available, 427 (27%) of the 1,590 residues are in helical conformations (Table I). Of the 21 aspartate residues in helices, 11 are within one turn of the N-termini of the helices and only five are located within one turn of the C-termini of the helices. A similar bias towards the N-termini of helices is observed for the 28 glutamates in helices, with 10 located near N-termini and five located near C-termini. These biases are consistent with those identified previously in protein helices.[101]

Interesting trends are observed in the mean pK_a values for aspartate and glutamate residues in helices: aspartates and glutamates at the N-termini of helices have mean pK_a values of 2.8 (±0.5) and 3.4 (±0.6), respectively, about 0.6 pH units less than the overall mean values. In contrast, mean pK_a values for residues at other helical positions are similar to the overall mean values. These results suggest a significant and specific effect of helical N-termini on carboxyl pK_a values. This effect is consistent with previous studies in proteins and peptides.[102–106] In addition, the average electrostatic potentials at these carboxyl groups, 5.8 (±3.3) and 3.8 (±2.9) kcal/mol e- for aspartate and glutamate, are greater than the overall mean values. This agrees qualitatively with expectations articulated in the early work on helix dipoles.[98,99]

DISCUSSION

Carboxyl pK_a Determinations by NMR

Many investigators have used of various subsets of the pK_a values reported in Table II as a basis for comparison with and refinement of electrostatic theory. One outcome of the present study is an opportunity to evaluate the various pK_a determinations and to determine which sets of pK_a values might be most useful for such comparative studies. In our judgment, the qualities one seeks in such studies are large chemical shift changes (preferably monitored at ^{13}C), experimental precision (i.e., small increments in pH and many data points describing the titration curves), well-described solution conditions with minimal ionic strength changes during titration, some estimate of reproducibility, some evidence for reversibility in the titrations, and a relatively large number of pK_a values. In addition, a high-resolution X-ray structure of the protein used for titrations studies is highly desirable. In this regard, one prototype of an ideal study is the case of histidines in RNase A where interpretation of pK_a values is greatly facilitated by a series of six 1.1 Å crystal structures from pH 5 to pH 9.[107]

The studies of RNase HI and the N-terminal domain of CD2 seem to come closest to satisfying these criteria. Thioredoxin has many of the most desirable features but no X-ray structure is available for the variant used in the NMR studies. Many of the other studies satisfy most of the criteria regarding precision, solution conditions, reproducibility and reversibility and these can be identified on the basis of information in Table II. More recent comprehensive studies of xylanase and calbindin D_{9k} have appeared and are likely to improve the overall reliability of carboxyl pK_a determinations for these proteins.[108,109] In addition, carboxyl pK_a values have recently been reported for RNase T1.[110] In general, analysis of cases where multiple independent pK_a determinations have been performed or where highly similar proteins have been studied suggest that the average experimental error in such determinations is 0.1 to 0.2 pH units.

Carboxylates Are Favored by Protein Structure

Carboxyl pK_a values in globular proteins are generally less than values for model compounds. The mean pK_a value for aspartate in proteins is at least 0.4 pH units less than model compound values. The effect of protein structure on glutamate pK_a values is less pronounced: mean pK_a values are at the extreme low end of model compound values. As discussed below, the structural basis for the difference between aspartate and glutamate pK_a values is not entirely clear but some results point to possible contributions from hydrogen bonding and the electrostatic potential.

The fact that aspartate pK_a values tend to be decreased to a larger extent than glutamate pK_a values suggests that, on average, aspartate residues play a more significant role than glutamates in the acid pH dependence of protein stability. This follows from the thermodynamic

linkage between protein stability and ionization equilibria,[4,5] where only ionizable groups with pK_a values that are different in the native and denatured states make significant contributions to the pH dependence of stability. In other words, because these pK_a differences reflect different proton affinities in the native and denatured states, mass action dictates that protein stability, i.e., the equilibrium constant for unfolding, will change with varying pH. If one assumes that model compound values accurately reflect pK_a values in denatured proteins, then the lower pK_a for aspartates in proteins relative to model compound values should lead to more dramatic changes in protein stability with decreasing pH. Considerable evidence suggests that pK_a values in denatured proteins may also be perturbed relative to model compound values.[7–10,111] Our conclusion regarding the larger role played by aspartate in the acid pH dependence of protein stability is thus predicated on the assumption that pK_a values for aspartates and glutamates in denatured proteins are perturbed to similar extents.

Carboxylic Acids Are Favored by Active Sites

Nearly all pK_a values > 5.5 are located in active sites and the one exception is Glu 41 of rat CD2, which is located on the protein's ligand-binding surface. Approximately half of the proteins in Table II are enzymes and about half of these possess carboxyl groups with pK_a values > 5.5. Based on the thermodynamic linkage argument outlined above, these carboxyl groups will destabilize the proteins at neutral pH relative to acid pH. Conflicting relationships between activity and stability in enzymes have been identified in previous studies.[11,112–114] The requirements for catalysis often include burial of polar and charged groups and these phenomena are likely to destabilize native protein structures. The fact that the highest carboxyl pK_a values are found in active sites leads us to hypothesize that this is a general phenomenon in proteins: carboxyl pK_a values > 5.5 are found only in active sites or ligand-binding sites. This hypothesis is consistent with the premise of a recent computational study in which putative active site residues were identified on the basis of their unfavorable electrostatic properties.[11]

Structural Basis for Carboxyl pK_a Values

One of the clearest results from the present study is that carboxyl groups that are well exposed to solvent have pK_a values that are narrowly distributed about mean values of 3.5 ($\pm$0.3) and 4.1 ($\pm$0.3) for aspartate and glutamate, respectively. This result is complicated a bit by the use of different SAS values for defining solvent exposure at aspartate and glutamate, > 45 Å^2 for aspartate oxygens vs. > 40 Å^2 for glutamate oxygens, but these yield more precise pK_a values for use in structure-based predictions. Interestingly, the mean pK_a values for solvent-exposed carboxyl groups are similar to the mean values for all of the carboxyl groups. However, standard deviations about the overall mean values are about three times greater.

Very few of the carboxyl pK_a values are > 5.5 and all of these groups are located in active sites or binding sites. All of these groups are buried, with < 20 Å^2 of SAS for the oxygen atoms, and they are usually involved in less than two intramolecular hydrogen bonds. However, these structural criteria are also satisfied in the crystal structures by nine aspartates and eight glutamates with pK_a values that are < 5.5. For aspartate, the major feature distinguishing the high and low pK_a values is a near zero potential at the groups with high pK_a values (Fig. 4) vs. potentials of at least 4 kcal/mol e- at the buried groups with smaller pK_a values. The picture is not so clear at glutamates, where calculated potentials are positive at groups with high and low pK_a values. Overall, burial and minimal hydrogen bonding appear to be necessary but not sufficient criteria for predicting high pK_a values for carboxyl groups in globular proteins. Recent studies of Staphylococcal nuclease suggest that buried water molecules may play a significant role in modulating pK_a values of buried ionizable groups.[97] More specifically, buried waters raise pK_a values for buried carboxyl groups and such effects have not been considered in the present study.

The observation that the mean pK_a values for aspartate and glutamate in proteins are less than model compound values by about 0.4 and 0.1 pH units, respectively, suggests that, on average, the negative charges on aspartate residues are stabilized by native protein structure to a greater extent than are negative charges on glutamate residues. This trend might be explained by differences in electrostatic potential and hydrogen bonding at aspartate vs. glutamate carboxyl groups: the mean potential at aspartates, 4.8 kcal/mol e-, is nearly twice the value at glutamates and aspartate carboxyl groups are, on average, involved in 1.4 intramolecular hydrogen bonds vs. 0.9 at glutamate carboxyl groups. Differences in hydrogen bonding at aspartate and glutamate have been observed previously in a number of studies.[38,39,115]

Inspection of Figure 4 and Tables III and IV suggests that the relationships between pK_a values, electrostatic potential, and hydrogen bonding are complex. In fact, each of the structural features explored in this study must be making contributions to the observed pK_a values and, moreover, all of these features are probably correlated with one another to some extent. For example, the calculated electrostatic potential and hydrogen bonding are intimately related because of the charges on the hydrogens in hydrogen bonds. In the calculations, these hydrogens are typically positioned within 1.5 to 2.0 Å of carboxyl oxygens and, consequently, they make a significant contribution to the positive potential. In this regard, computational results suggest that much of the positive potential at side chain carboxyl groups and the overall greater potentials at aspartate vs. glutamate is contributed by the peptide backbone.[116,117] Similarly, hydrogen bonding and pI values might be correlated because basic proteins have more potential hydrogen bond donors than acidic proteins.

If intramolecular hydrogen bonding can indeed explain differences between aspartate and glutamate pK_a values, then pK_a values for both types of residues should show correlations with the number of hydrogen bonds. The fact that this appears to be the case for aspartates but not for

glutamates is a puzzle. One hypothesis to explain this discrepancy is that uncertainties in the positioning of glutamate side chains in molecular models are greater than for aspartate side chains and, consequently, hydrogen bonding at glutamates has been overestimated. The longer side chain for glutamate and the fact that most carboxyl groups are located on the surfaces of proteins could lead to lower electron densities for glutamate carboxyl groups. A modest trend in this direction is detected in the slightly larger average B factor at glutamate carboxyl carbons, 32 ($\pm$23) Å^2, vs. 27 ($\pm$23) Å^2 aspartate carboxyl carbons.

More generally, some anecdotal observations suggest that there may be significant uncertainty in the positioning of some carboxyl groups in crystal structures. Asp 38 in RNase A was cited earlier as an example of a residue whose conformation is significantly different in different crystal structures. Our experience with ovomucoid third domain and, more recently, with ubiquitin suggest that the side chains of solvent-exposed aspartate and glutamate residues can adopt different conformations in different crystal forms. Another source of uncertainty is the possibility of changes in side chain conformation during pH titration (see[118] and references therein]. In other words, in cases where carboxyl side chains are positioned accurately in crystal or NMR structures, these positions may change as pH is varied.

In spite of these uncertainties, at least one other feature of protein structure shows significant correlations with carboxyl pK_a values: mean pK_a values for carboxyl groups at the N-terminal turns of helices are significantly less than mean pK_a values for groups located elsewhere in protein structures. This effect could be due to the greater than average potentials at the N-termini of helices, hydrogen bonding between the carboxyl groups and peptide amide protons at the ends of helices, or some combination of these two phenomena.

Empirical relationships between protein structure and carboxyl pK_a values reveal interesting and intriguing trends but, as evidenced in the standard deviations, these relationships are not very precise. The lack of precision probably reflects, among other things, the combined effects of different experimental conditions, uncertainties in positioning of atoms in the molecular models, differences in X-ray and solution conformations for solvent-exposed side chains, and the relatively crude descriptions of structure used in the present study. In addition, solvent-exposed side chains are able to sample multiple conformations and these fluctuations are going to affect electrostatic potentials and structural parameters such as solvent-accessibility. Nevertheless, the annotated database of carboxyl pK_a values (Tables I and II) will probably be useful in more detailed computational studies of protein electrostatics. This database can be obtained from the authors or from the electronic version of this article. With respect to experimental approaches, more precise information regarding the molecular basis for pK_a values is likely to be obtained from studies that combine mutagenesis with pK_a determinations by NMR (see[32,70,119] and references therein). In addition, both experimental and computational studies rely heavily on knowledge of protein structure and these structures are probably changing with varying pH.[107,108,118] Consequently, more information about such conformational changes would be very useful.

ACKNOWLEDGMENTS

The authors thank Adrian Elcock and Kenneth P. Murphy for helpful discussions, Michael K. Gilson for careful reading of an early version of this manuscript, and the reviewers for helpful comments.

REFERENCES

1. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268:1144–1149.
2. Elcock AH, Sept D, McCammon JA. Computer simulations of protein-protein interactions. J Phys Chem B 2001;105:1504–1518.
3. Fersht A. Structure and mechanism in protein science. 1999, New York: W.H. Freeman & Co.
4. Tanford C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. Adv Prot Chem 1970;24:1–95.
5. Yang A-S, Honig B. Electrostatic effects on protein stability. Curr Opin Struct Biol 1992;2:40–45.
6. Anderson DE, Becktel WJ, Dahlquist FW. pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. Biochemistry 1990;29:2403–2408.
7. Oliveberg M, Arcus VL, Fersht AR. 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 1995;34:9424–9433.
8. Swint-Kruse L, Robertson AD. Hydrogen bonds and the pH dependence of ovomucoid third domain stability. Biochemistry 1995;34:4724–4732.
9. Tan YJ, Oliveberg M, Davis B, Fersht AR. Perturbed pK_a-values in the denatured states of proteins. J Mol Biol 1995;254:980–992.
10. Kuhlman B, Luisi D, Young P, Raleigh DP. pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. Biochemistry 1999;38:4896–4903.
11. Elcock AH. Prediction of functionally important residues based solely on the computed energetics of protein structure. J Mol Biol 2001;312:885–896.
12. Ondrechen M, Clifton J, Ringe D. THEMATICS: a simple computational predictor of enzyme function from structure. Proc Natl Acad Sci USA 2001;98:12473–12478.
13. Wyman J. Linked functions and reciprocal effects in hemoglobin: a second look. Adv Prot Chem 1964;19:223–286.
14. Kielian M, Jungerwirth S. Mechanisms of enveloped virus entry into cells. Mol Biol Med 1990;7:17–31.
15. Dewan J, Mikami B, Hirose M, Sacchettini J. Structural evidence for a pH-sensitive dilysine trigger in the hen ovotransferrin N-lobe: implications for transferrin iron release. Biochemistry 1993;32:11963–11968.
16. Carrasco L. Entry of animal viruses and macromolecules into cells. FEBS Lett 1994;350:151–154.
17. Richter C, Tanaka T, Yada R. Mechanism of activation of the gastric aspartic proteinases: pepsinogen, progastricsin and prochymosin. Biochem J 1998;335:481–490.
18. Khan A, James M. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Sci 1998;7:815–836.
19. Lord J, Smith D, Roberts L. Toxin entry: how bacterial proteins get into mammalian cells. Cell Microbiol 1999;1:85–91.
20. Matthew JB, Gurd FR, Garcia-Moreno B, Flanagan MA, March KL, Shire SJ. pH-dependent processes in proteins. CRC Crit Rev Biochem 1985;18:91–197.
21. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 2001;44:400–417.
22. Warshel A, Russell S. Calculations of electrostatic interactions in

biological systems and in solutions. Q Rev Biophys 1984;17:283–422.
23. Nakamura H. Roles of electrostatic interaction in proteins. Q Rev Biophys 1996;29:1–90.
24. Simonson T. Macromolecular electrostatics: continuum models and their growing pains. Curr Opin Struct Biol 2001;11:243–252.
25. Rees DC. Experimental evaluation of the effective dielectric constant of proteins. J Mol Biol 1980;141:323–326.
26. Russell AJ, Thomas PG, Fersht AR. Electrostatic effects on modification of charged groups in the active site cleft of subtilisin by protein engineering. J Mol Biol 1987;193:803–813.
27. Loewenthal R, Sancho J, Reinikainen T, Fersht AR. Long-range surface charge-charge interactions in proteins. Comparison of experimental results with calculations from a theoretical method. J Mol Biol 1993;232:574–583.
28. Cederholm MT, Stuckey JA, Doscher MS, Lee L. Histidine pK_a shifts accompanying the inactivating Asp121—Asn substitution in a semisynthetic bovine pancreatic ribonuclease. Proc Natl Acad Sci USA 1991;88:8116–8120.
29. Jackson SE, Fersht AR. Contribution of long-range electrostatic interactions to the stabilization of the catalytic transition state of the serine protease subtilisin BPN'. Biochemistry 1993;32:13909–13916.
30. Fisher BM, Schultz LW, Raines RT. Coulombic effects of remote subsites on the active site of ribonuclease A. Biochemistry 1998;37:17386–17401.
31. Lee KK, Fitch CA, Lecomte JTJ, Garcia-Moreno B. Electrostatic effects in highly charged proteins: salt sensitivity of pK_a values of histidines in staphylococcal nuclease. Biochemistry 2002;41:5656–5667.
32. Lee KK, Fitch CA, Garcia-Moreno B. Distance dependence and salt sensitivity of pairwise coulombic interactions in a protein. Prot Sci 2002;11:1004–1016.
33. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pK_as in proteins. Biochemistry 1996;35:7819–7833.
34. Livingstone JR, Spolar RS, Record MT. Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. Biochemistry 1991;30:4237–4244.
35. Privalov PL, Gill SJ. Stability of protein structure and hydrophobic interaction. Adv Prot Chem 1988;39:191–234.
36. Murphy KP, Freire E. Thermodynamics of structural stability and cooperative folding behavior in proteins. Adv Prot Chem 1992;43:313–361.
37. Robertson AD, Murphy KP. Protein Structure and the energetics of protein stability. Chem Rev 1997;97:1251–1267.
38. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
39. Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984;44:97–179.
40. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
41. Brunger AT, Karplus M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. Proteins 1988;4:148–156.
42. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4:187–217.
43. Warwicker J, Watson HC. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. J Mol Biol 1982;157:671–679.
44. Garrett AJM, Poladian L. Refined derivation, exact solutions, and singular limits of the Poisson-Boltzmann equation. Ann Physics 1988;188:386–435.
45. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BE, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comput Phys Comm 1995;91:57–95.
46. Richards FM. Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng 1977;6:151–176.
47. Gilson MK, Sharp KA, Honig BH. Calculating the electrostatic potential of molecules in solution: method and error assessment. J Comp Chem 1988;9:327–335.
48. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
49. Koide A, Jordan MR, Horner SR, Batori V, Koide S. Stabilization of fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface. Biochemistry 2001;40:10326–10333.
50. Sundd M, Iverson N, Ibarra-Molero B, Sanchez-Ruiz JM, Robertson AD. Electrostatic interactions in ubiquitin: stabilization of carboxylates by lysine amino groups. Biochemistry 2002;41:in press.
51. Anderson DE, Lu J, McIntosh L, Dahlquist FW. The folding, stability and dynamics of T4 lysozyme: a perspective using nuclear magnetic resonance. In: Clore M, Gronenborn A, editors. NMR of proteins. London: Macmillan Press Ltd; 1993. p. 258–304.
52. Markley JL. Observation of histidine residues in proteins by means of nuclear magnetic resonance spectroscopy. Acc Chem Res 1975;8:70–80.
53. Chen HA, Pfuhl M, McAlister MSB, Driscoll PC. 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 2000;39:6814–6824.
54. Qin J, Clore GM, Gronenborn AM. Ionization equilibria for side-chain carboxyl groups in oxidized and reduced human thioredoxin and in the complex with its target peptide from the transcription factor NF kappa B. Biochemistry 1996;35:7–13.
55. Perez-Canadillas JM, Campos-Olizas R, Lacadena J, Martinez del Pozo A, Gavilanes JG, Santoro J, Rico M, Bruix M. Characterization of pK_a values and titration shifts in the cytotoxic ribonuclease alpha-sarcin by NMR. Relationship between electrostatic interactions, structure, and catalytic function. Biochemistry 1998;37:15865–15876.
56. Shrager RI, Cohen JS, Heller SR, Sachs DH, Schechter AN. Mathematical models for interacting groups in nuclear magnetic resonance titration curves. Biochemistry 1972;11:541–547.
57. Oda Y, Yamazaki T, Nagayama K, Kanaya S, Kuroda Y, Nakamura H. Individual ionization constants of all the carboxyl groups in ribonuclease HI from *Escherichia coli* determined by NMR. Biochemistry 1994;33:5275–5284.
58. McIntosh LP, Hand G, Johnson PE, Joshi MD, Korner M, Plesniak LA, Ziser L, Wakarchuk WW, Withers SG. The pK_a of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a ^{13}C-NMR study of bacillus circulans xylanase. Biochemistry 1996;35:9958–9966.
59. Joshi MD, Hedberg A, McIntosh LP. Complete measurement of the pK_a values of the carboxyl and imidazole groups in Bacillus circulans xylanase. Prot Sci 1997;6:2667–2670.
60. Ladner H, Led J, Grant D. Deuterium isotope effects on ^{13}C chemical shifts in amino acids and dipeptides. J Magn Reson 1975;20:530–534.
61. Ebina S, Wüthrich K. Amide proton titration shifts in bull seminal inhibitor IIA by two-dimensional correlated ^1H nuclear magnetic resonance (COSY). Manifestation of conformational equilibria involving carboxylate groups. J Mol Biol 1984;179:283–288.
62. Szyperski T, Antuch W, Schick M, Betz A, Stone SR, Wüthrich K. Transient hydrogen bonds identified on the surface of the NMR solution structure of Hirudin. Biochemistry 1994;33:9303–9310.
63. Haruyama H, Qian YQ, Wüthrich K. Static and transient hydrogen-bonding interactions in recombinant desulfatohirudin studied by ^1H nuclear magnetic resonance measurements of amide proton exchange rates and pH-dependent chemical shifts. Biochemistry 1989;28:4312–4317.
64. Khare D, Alexander P, Antosiewicz J, Byran P, Gilson M, Orban J. 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 1997;36:3580–3589.
65. Chiang C-M, Chang S-L, Lin H-J, Wu W-G. The role of acidic amino acid residues in the structural stability of snake cardiotoxins. Biochemistry 1996;35:9177–9186.
66. Sorensen MD, Led JJ. Structural details of Asp(B9) human insulin at low pH from two-dimensional NMR titration studies. Biochemistry 1994;33:13727–13733.
67. Rico M, Santoro J, Gonzalez C, Bruix M, Neira JL. Solution structure of bovine pancreatic ribonuclease A and ribonuclease-pyrimidine nucleotide complexes as determined by H NMR. In:

Structure, Mechanism and Function of Ribonuleases Proceedings of the 2nd International Meeting. 1990. Sant Feliu de Guixols, Girona, Spain: Department de Bioquimica i Biologia Molecular and Institut de Biologia Fonamental Vicent Villar Palasi, Universitat Autonoma de Barcelona, Bellaterra, Spain.
68. Baker WR, Kintanar A. Characterization of the pH titration shifts of ribonuclease A by one- and two-dimensional nuclear magnetic resonance spectroscopy. Arch Biochem Biophys 1996; 327:189–199.
69. Schaller W, Robertson AD. pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. Biochemistry 1995;34:4714–4723.
70. Forsyth WR, Robertson AD. Insensitivity of perturbed carboxyl pK(a) values in the ovomucoid third domain to charge replacement at a neighboring residue. Biochemistry 2000;39:8067–8072.
71. Bundi A, Wüthrich K. Use of amide ^{1}H-NMR titration shifts for studies of polypeptide conformation. Biopolymers 1979;18:299–311.
72. Mayer R, Lancelot G, Spach G. Side chain-backbone hydrogen bonds in peptides containing glutamic acid residues. Biopolymers 1979;18:1293–1295.
73. Wagner G, Pardi A, Wüthrich, K. Hydrogen bond lengths and ^{1}H NMR chemical shifts in proteins. J Am Chem Soc 1983;105:5948–5949.
74. Sternberg U, Brunner E. The influence of short-range geometry on the chemical shift of protons in hydrogen bonds. J Magn Reson A 1994;108:142–150.
75. Wishart DS, Case DA. Use of chemical shifts in macromolecular structure determination. Methods Enzymol 2001;338:3–34.
76. Grzesiek S, Cordier F, Dingley AJ. Scalar couplings across hydrogen bonds. Methods Enzymol 2001;338:111–133.
77. Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K. Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. J Biomol NMR 1998;12:1–23.
78. DeMarco A. pH dependence of internal references. J Magn Reson 1977;26:527–528.
79. Kesvatera T, Jonsson B, Thulin E, Linse S. Measurement and modelling of sequence-specific pK_a values of lysine residues in calbindin D9k. J Mol Biol 1996;259:828–839.
80. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238:415–436.
81. Zale SE, Klibanov AM. Why does ribonuclease irreversibly inactivate at high temperature? Biochemistry 1986;25:5432–5444.
82. Stephenson RC, Clarke S. Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. J Biol Chem 1989;264:6164–6170.
83. Fágáin CO. Understanding and increasing protein stability. Biochim Biophys Acta 1995;1252:1–14.
84. Li S, Schöneich C, Borchardt RT. Chemical instability of protein pharmaceuticals: mechanisms of oxidation and strategies for stabilization. Biotechnol Bioeng 1995;48:490–500.
85. Creighton TE. Proteins: structure and molecular properties, 2nd ed. New York: W. H. Freeman and Company; 1993. 507 p.
86. Edsall JT. Proteins as acids and bases. In: Cohn EJ, Edsall JT, editors. Proteins, amino acids, and peptides as ions and dipolar ions. New York: Reinhold Publishing Co; 1943. p 444–505.
87. Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. Protein Sci 2001;10:1206–1215.
88. Yang AS, Gunner MR, Sampogna R, Sharp L, Honig B. On the calculation of pK_as in proteins. Proteins 1993;15:252–265.
89. Warshel A. Electrostatic basis of structure-function correlation in proteins. Acc Chem Res 1981;14:284–290.
90. Warshel A, Russel ST, Churg AK. Macroscopic models for studies of electrostatic interactions in proteins: limitations and applicability. Proc Natl Acad Sci USA 1984;81:4785–4789.
91. Rashin AA, Honig B. On the environment of ionizable groups in globular proteins. J Mol Biol 1984;173:515–521.
92. Gilson MK. Theory of electrostatic interactions in macromolecules. Curr Opin Struct Biol 1995;5:216–223.
93. Juffer AH. Theoretical calculations of acid-dissociation constants of proteins. Biochem Cell Biol 1998;76:198–209.
94. Kirkwood JG. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. J Chem Phys 1934;2:351–361.
95. Bottcher CJF. Theory of electric polarization. Vol. 1. Amsterdam: Elsevier; 1973
96. Ibarra-Molero B, Loladze VV, Makhatadze GI, Sanchez-Ruiz JM. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. Biochemistry 1999;38:8138–8149.
97. Dwyer JJ, Gittis AG, Karp DA, Lattman EE, Spencer DS, Stites WE, Garcia-Moreno B. High apparent dielectric constants in the interior of a protein reflect water penetration. Biophys J 2000;79:1610–1620.
98. Wada A. The alpha-helix as an electric macro-dipole. Adv Biophys 1976;9:1–63.
99. Hol WGJ. The role of the alpha-helix dipole in protein function and structure. Prog Biophys Mol Biol 1985;45:149–195.
100. Presta LG, Rose GD. Helix signals in proteins. Science 240:1632–1641.
101. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of alpha helices. Science 1988;240:1648–1652.
102. Nicholson H, Anderson DE, Dao-pin S, Matthews BW. Analysis of the interaction between charged side chains and the alpha-helix dipole using designed thermostable mutants of phage T4 lysozyme. Biochemistry 1991;30:9816–9828.
103. Sancho J, Serrano L, Fersht AR. Histidine residues at the N- and C-termini of alpha-helices: perturbed pK_as and protein stability. Biochemistry 1992;31:2253–2258.
104. Lockhart DJ, Kim PS. Electrostatic screening of charge and dipole interactions with the helix backbone. Science 1993;260:198–202.
105. Kortemme T, Creighton TE. Ionization of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pK_a values in proteins of the thioredoxin family. J Mol Biol 1995;253:799–812.
106. Huyghues-Despointes BM, Scholtz JM, Baldwin RL. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. Protein Sci 1993;2:1604–1611.
107. Berisio R, Lamzin VS, Sica F, Wilson KS, Zagari A, Mazzarella L. Protein titration in the crystal state. J Mol Biol 1999;292:845–854.
108. Joshi MD, Sidhu G, Nielsen JE, Brayer GD, Withers SG, McIntosh LP. Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. Biochemistry 2001;40:10115–10139.
109. Kesvatera T, Jönsson B, Thulin E, Linse S. Focusing of the electrostatic potential at EF-hands of calbindin D_{9k}: titration of acidic residues. Proteins 2001;45:129–135.
110. Spitzner N, Lohr F, Pfeiffer S, Koumanov A, Karshikoff A, Ruterjans H. Ionization properties of titratable groups in ribonuclease T1. I. pK_a values in the native state determined by two-dimensional heteronuclear NMR spectroscopy. Eur Biophys J 2001;30:186–197.
111. Elcock AH. Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability. J Mol Biol 1999;294:1051–1062.
112. Meiering EM, Serrano L, Fersht AR. Effect of active site residues in barnase on activity and stability. J Mol Biol 1992;225:585–589.
113. Zhang J, Liu ZP, Jones TA, Gierasch LM, Sambrook JF. Mutating the charged residues in the binding pocket of cellular retinoic acid-binding protein simultaneously reduces its binding affinity to retinoic acid and increases its thermostability. Proteins 1992;13:87–99.
114. Shoichet BK, Baase WA, Kuroki R, Matthews BW. A relationship between protein stability and protein function. Proc Natl Acad Sci USA 1995;92:452–456.
115. Ippolito JA, Alexander RS, Christianson DW. Hydrogen bond stereochemistry in protein structure and function. J Mol Biol 1990;215:457–471.
116. Peters GH, Frimurer TM, Olsen OH. Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases. Biochemistry 1998;37:5383 –5393.
117. Gunner MR, Saleh MA, Cross E, ud-Doula A, Wise M. Backbone

dipoles generate positive potentials in all proteins: origins and implications of the effect. Biophys J 2000;78:1126–1144.
118. Yang F, Phillips GN. Crystal structures of CO-, deoxy- and metmyoglobins at various pH values. J Mol Biol 1996;256:762–774.
119. Tishmack PA, Bashford D, Harms E, Van Etten RL. Use of ^1H NMR spectroscopy and computer simulations to analyze histidine pK_a changes in a protein tyrosine phosphatase: experimental and theoretical determination of electrostatic properties in a small protein. Biochemistry 1997;36:11984–11994.
120. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
121. Wang YX, Freedberg DI, Yamazaki T, Wingfield PT, Stahl SJ, Kaufman JD, Kiso Y, Torchia DA. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. Biochemistry 1996;35:9945–9950.
122. Bartik K, Redfield C, Dobson CM. Measurement of the individual pK_a values of acidic residues of hen and turkey lysozymes by two-dimensional ^1H NMR. Biophys J 1994;66:1180–1184.
123. Richarz R, Wüthrich K. High-field ^{13}C nuclear magnetic resonance studies at 90.5 MHz of the basic pancreatic trypsin inhibitor. Biochemistry 1978;17:2263–2269.
124. March KL, Maskalick DG, England RD, Friend SH, Gurd FR. Analysis of electrostatic interactions and their relationship to conformation and stability of bovine pancreatic trypsin inhibitor. Biochemistry 1982;21:5241–5251.
125. Assadi-Porter FM, Fillingame RH. Proton-translocating carboxyl of subunit c of F_1F_o H(+)-ATP synthase: the unique environment suggested by the pK_a determined by ^1H NMR. Biochemistry 1995;34:16186–16193.
126. Gooley PR, Keniry MA, Dimitrov RA, Marsh DE, Keizer DW, Gayler KR, Grant BR. The NMR solution structure and characterization of pH dependent chemical shifts of the beta-elicitin, cryptogein. J Biomol NMR 1998;12:523–534.
127. Kohda D, Sawada T, Inagaki F. Characterization of pH titration shifts for all the nonlabile proton resonances in a protein by two-dimensional NMR: the case of mouse epidermal growth factor. Biochemistry 1991;30:4896–4900.