PROTEINS: Structure, Function, and Bioinformatics 55:689–704 (2004)

The Determinants of Carboxyl pK_a Values in Turkey Ovomucoid Third Domain

Hui Li,[1] Andrew D. Robertson,[2] and Jan H. Jensen[1]*
[1]*Department of Chemistry, The University of Iowa, Iowa City, Iowa*
[2]*Department of Biochemistry, The University of Iowa, Iowa City, Iowa*

ABSTRACT **A computational methodology for protein pK_a predictions, based on ab initio quantum mechanical treatment of part of the protein and linear Poisson–Boltzmann equation treatment of the bulk solvent, is presented. The method is used to predict and interpret the pK_a values of the five carboxyl residues (Asp7, Glu10, Glu19, Asp27, and Glu43) in the serine protease inhibitor turkey ovomucoid third domain. All the predicted pK_a values are within 0.5 pH units of experiment, with a root-mean-square deviation of 0.31 pH units. We show that the decreased pK_a values observed for some of the residues are primarily due to hydrogen bonds to the carboxyl oxygens. Hydrogen bonds involving amide protons are shown to be particularly important, and the effect of hydrogen bonding is shown to be nonadditive. Hydrophobic effects are also shown to be important in raising the pK_a. Interactions with charged residues are shown to have relatively little effect on the carboxyl pK_a values in this protein, in general agreement with experiment. Proteins 2004; 55:689–704.** © **2004 Wiley-Liss, Inc.**

INTRODUCTION

The stability and function of a protein are intimately tied to its acid base chemistry and, hence, to the pK_a values of its ionizable residues. Thus, an understanding of the catalytic function of an enzyme or the design of a more stable protein rests in large part on an understanding of the environmental determinants of pK_a values.

The pK_a values of hundreds of residues in many different proteins have been accurately measured, including several rationally designed mutants.[1–3] X-ray or NMR structures are available for virtually all these proteins, allowing for an analysis of the pK_a values in terms of structure. The following four main determinants of pK_a values are generally invoked[3,4]:

1) Charge–charge interactions (i.e., Coulomb or electrostatic effects), whereby, for example, a Lys residue lowers the pK_a of an Asp residue by preferentially stabilizing the negative (unprotonated) form. The distance dependence of this interaction appears to depend on the protein and residue in question.[5–8] It has been suggested that surface residues are less sensitive to charged residues due to solvent screening of the charge.[5,6]

2) Hydrogen bonding (i.e., charge-dipole interactions), whereby, for example, hydrogen bonding to a Ser residue lowers the pK_a of an Asp residue by preferential stabilization of the negative (unprotonated) form. If the hydrogen-bonding partner is charged (e.g., a Lys residue), the interaction is sometimes classified as a charge–charge interaction. It is not clear whether hydrogen bonds to charged residues tend to induce larger pK_a shifts than do hydrogen bonds to neutral residues.

3) Desolvation effects (i.e., Born or hydrophobic effects), whereby, for example, a hydrophobic environment raises the pK_a of an Asp residue by a net preferential stabilization of the neutral (protonated) form (because solvation preferentially stabilizes the unprotonated form). Desolvation effects can produce very large (4–6 pH units) pK_a shifts and are often important pK_a determinants for active site residues.[9–12]

4) Helix dipole interactions, whereby, for example, the positive (N-terminal) end of an α-helix lowers the pK_a of an Asp group by preferentially stabilizing the negative (unprotonated) form.[1,13–17]

However, the determinants of protein pK_a values are not sufficiently clear to allow for a quantitative pK_a prediction for a given protein residue. For example, a recent survey[1] of the determinants of carboxyl pK_a values concludes that "empirical relationships between protein structure and carboxyl pK_a values reveal interesting and intriguing trends but…these relationships are not very precise." Another survey[2] of His-residue pK_a values reached similar conclusions.

Several computational methodologies aimed at protein pK_a predictions have been developed in the last two decades.[18–29] In the most popular approach, the protein is treated by a molecular mechanics force field, embedded in a uniform dielectric continuum with dielectric constants of 80 for the solvent and 4–20 for the protein interior. A pK_a shift is calculated from the difference in electrostatic energy of a residue in its charged and neutral form, and this shift is added to a model pK_a value. The protein structure is generally kept fixed.

Although these methods can predict most pK_a values to within 1 pH unit, significantly larger errors are not

Grant sponsor: Research Innovation Award from the Research Corporation; Grant sponsor: National Science Foundation; Grant number: MCB-0209941; Grant sponsor: National Institutes of Health; Grant number: GM-46869.

*Correspondence to: Jan H. Jensen, Department of Chemistry, The University of Iowa, Iowa City, IA 52242. E-mail: Jan-Jensen@uiowa.edu

Received 17 July 2003; Accepted 23 October 2003

Published online 5 March 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20032

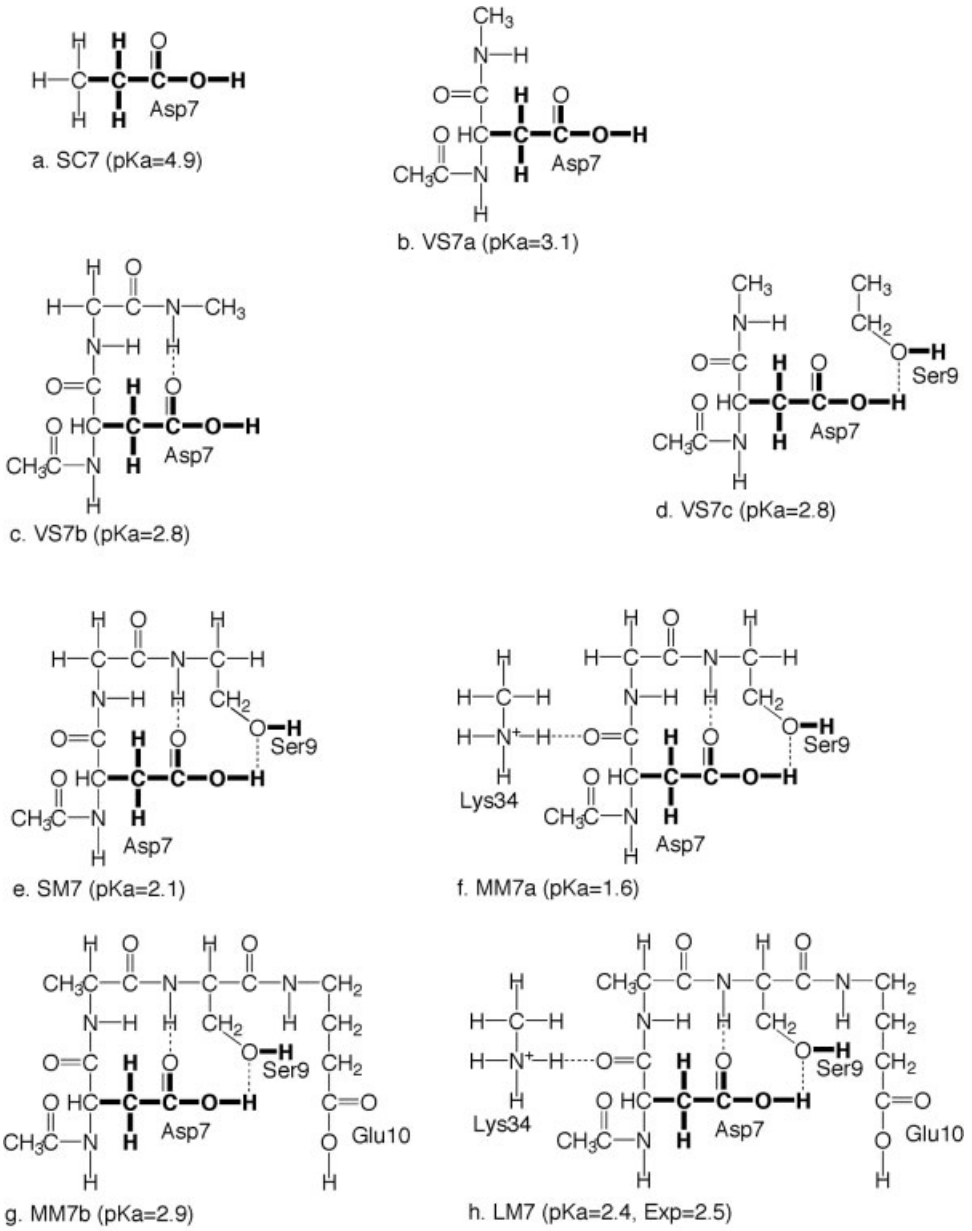


Fig. 1. Model compounds for Asp7 of OMTKY3 and their computed pK_a values. The positions of the atoms in bold were energy minimized. Acid form I is shown.

uncommon.[30,31] Unfortunately, these larger errors are often encountered for residues with unusual pK_a values, which are of particular interest. The errors may result from errors in the method used to calculate energies and/or errors in the protein structures used to calculate the energies. For example, it has been argued that the use of a uniform protein dielectric is fundamentally incorrect,[27,31] whereas a recent study by Laurents et al.[32] concludes that current continuum approaches tend to "underestimate the contributions of both desolvation and charge-dipole interactions to the pKs of buried ionizable groups." Alternatively, others have noted that predicted pK_a values can be affected by modest changes in a crystal structure (perhaps induced by crystal packing forces).[30,33,34] It has been argued that including the dynamical motion of the protein will improve the accuracy of continuum-based pK_a predictions, although such approaches have had mixed success.[34–39]

The five carboxyl pK_a values of the 56-residue protease inhibitor Turkey ovomucoid third domain (OMTKY3) provide good systems for the study of pK_a determinants. The pK_a values of all the carboxyl residues of the wild-type[40] and several mutants[7,41] have been determined by NMR. The carboxyl pK_a values span a relatively wide range of 2.2–4.8 pH units, and even the lowest pK_a values can be determined relatively accurately, given the unusual stability of this protein at low pH.[42] Several computational predictions[27,30,39,43] of the carboxyl pK_a values in OMTKY3

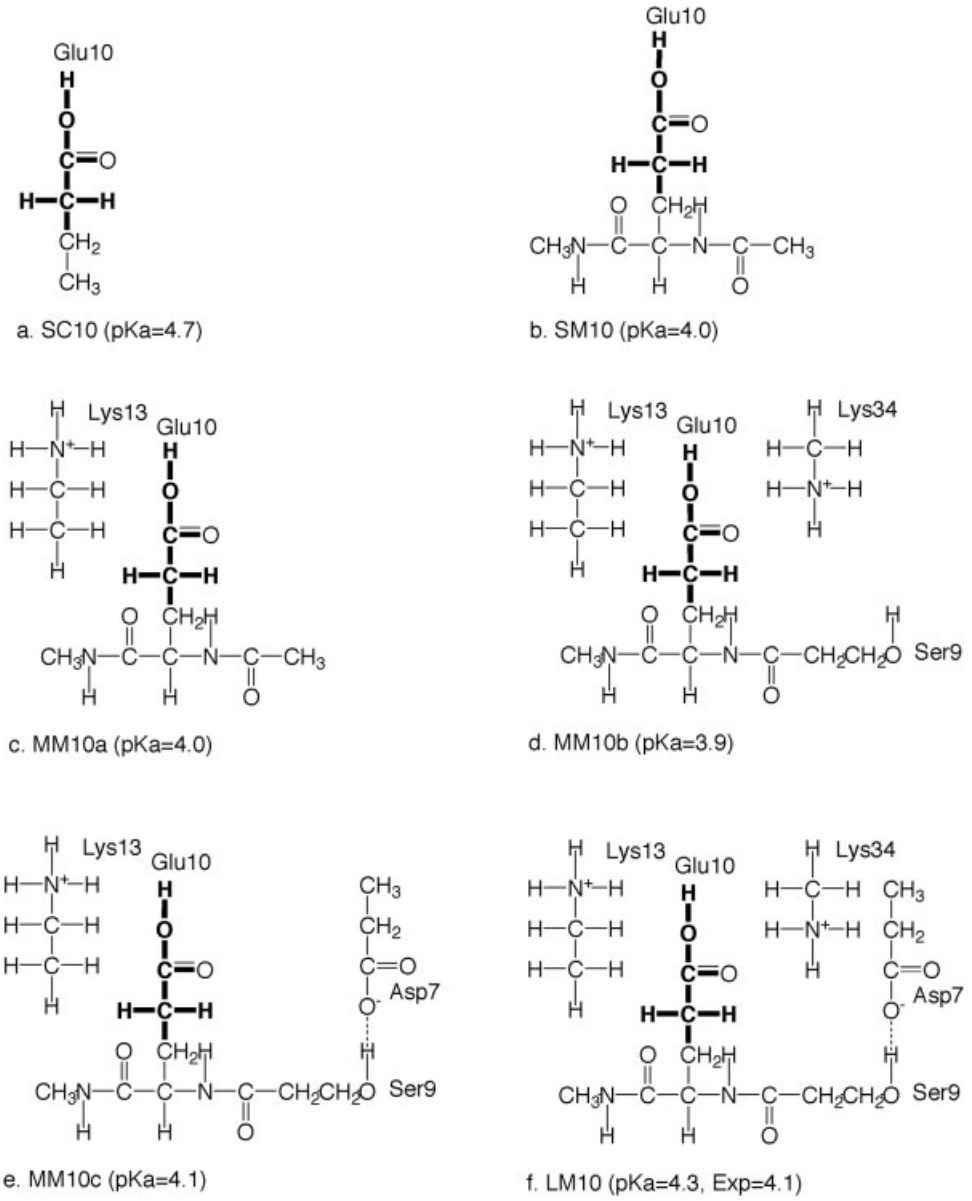


Fig. 2. Model compounds for Glu10 of OMTKY3 and their computed pK_a values. The positions of the atoms in bold were energy minimized. Acid form I is shown.

have been made with root-mean-square deviations (RMSDs) in the 0.5–0.9 range (see below). However, with the exception of the simplest method (a modified Tanford–Kirkwood method[39]), all fail to predict the most acidic pK_a (Asp27) to within 1 pH unit.

In this study, we use an ab initio quantum mechanics (QM) representation of the ionizable residues and their immediate environment combined with a continuum description of bulk solvation to study the determinants of the carboxyl pK_a values in OMTKY3. This general computational approach has been used successfully by us[44] and others[45–60] to predict the pK_a of small organic molecules. This approach can be extended to the prediction of a pK_a of an ionizable residue in a protein. However, only part of the protein, including the ionizable residue, can be treated with QM, because of the computational expense of the QM treatment. The rest of the protein can be treated by molecular mechanics (MM) using a hybrid QM/MM methodology.[44] Alternatively, the rest of the protein can be neglected, as is done in this study. This simpler approach provides a clearer understanding of the physical forces that determine the pK_a values. We note that this general approach has been used successfully by Zheng et al.[61] and Ullmann et al.[62,63] to predict protein pK_a values of residues near metal atoms.

The article is organized as follows. First, we briefly outline the computational methodology. Second, we discuss the determinants of each carboxyl pK_a in some detail. Third, we compare and contrast these determinants to draw more general conclusions about the structural deter-

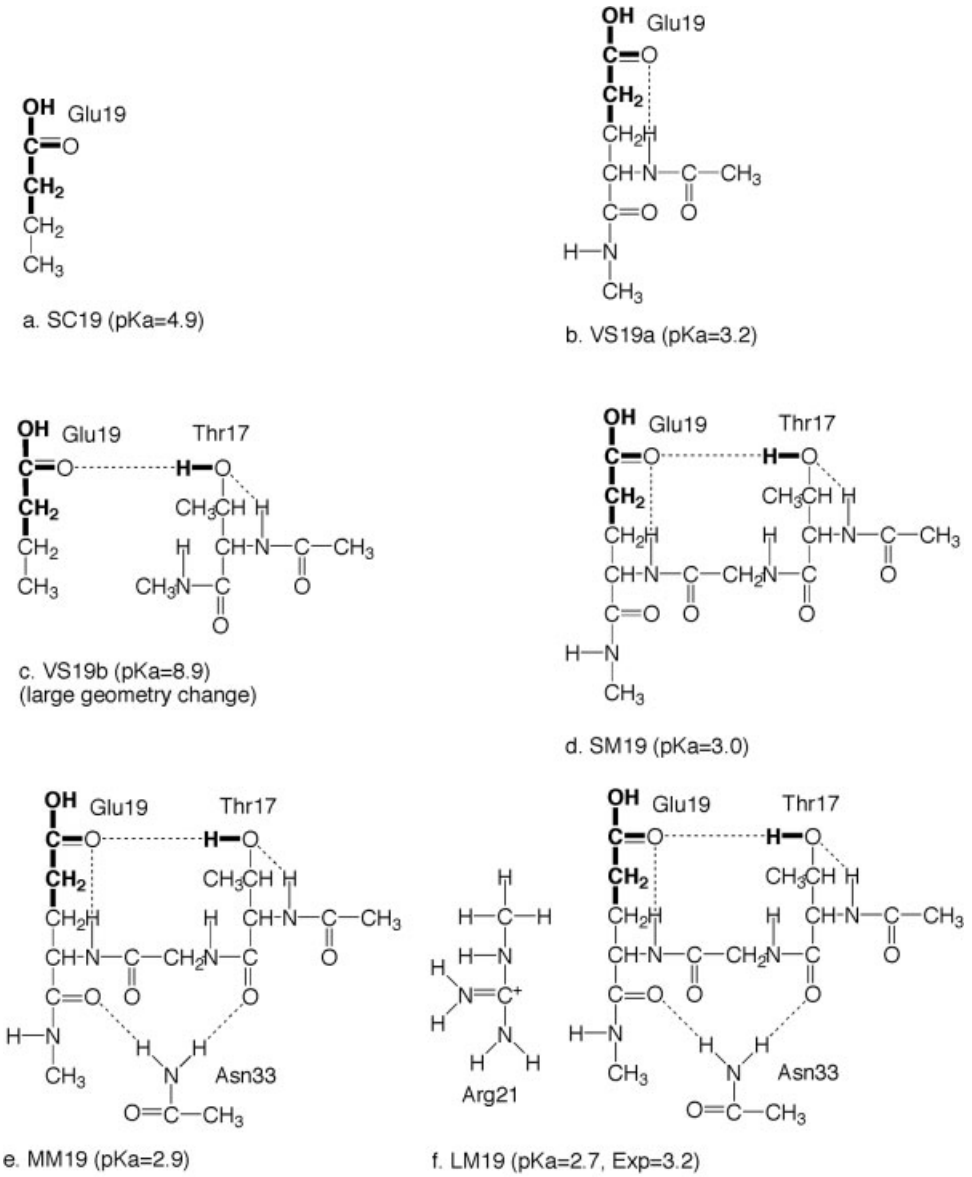


Fig. 3. Model compounds for Glu19 of OMTKY3 and their computed pK_a values. The positions of the atoms in bold were energy minimized. Acid form I is shown.

minants of carboxyl pK_a values in OMTKY3. Fourth, we suggest possible improvements to the more conventional methods for protein pK_a predictions.

COMPUTATIONAL METHODOLOGY

pK_a Calculations

In our approach, the pK_a of a carboxyl group in the protein, HA, is related to the standard free energy change, ΔG, of the following reaction

$$HA(aq) + C_2H_5COO^-(aq) \xrightarrow{\Delta G} A^-(aq) + C_2H_5COOH(aq) \quad (1)$$

by the equation

$$pK_a = 4.87 + \Delta G/1.36 = 4.87 + \{[G(A^-) - G(HA)] - [G(C_2H_5COO^-) - G(C_2H_5COOH)]\}/1.36 \quad (2)$$

Here, 4.87 is the experimentally determined pK_a of propanoic acid at 298 K[64] and 1.36 is RTln10 for T = 298 K in kcal/mol. G(X) is the total free energy (in kcal/mol) of molecule X, which is the sum of the ground state electronic energy (E_{ele}) and solvation energy (G_{sol})

$$G = E_{ele} + G_{sol} \quad (3)$$

E_{ele} is comprised of the potential energy of the electrons and nuclei as well as the kinetic energy of the electrons and is calculated by standard quantum chemical techniques (described in detail in the appendix) using the GAMESS[65] and PQS[66] programs. G_{sol} is calculated by the

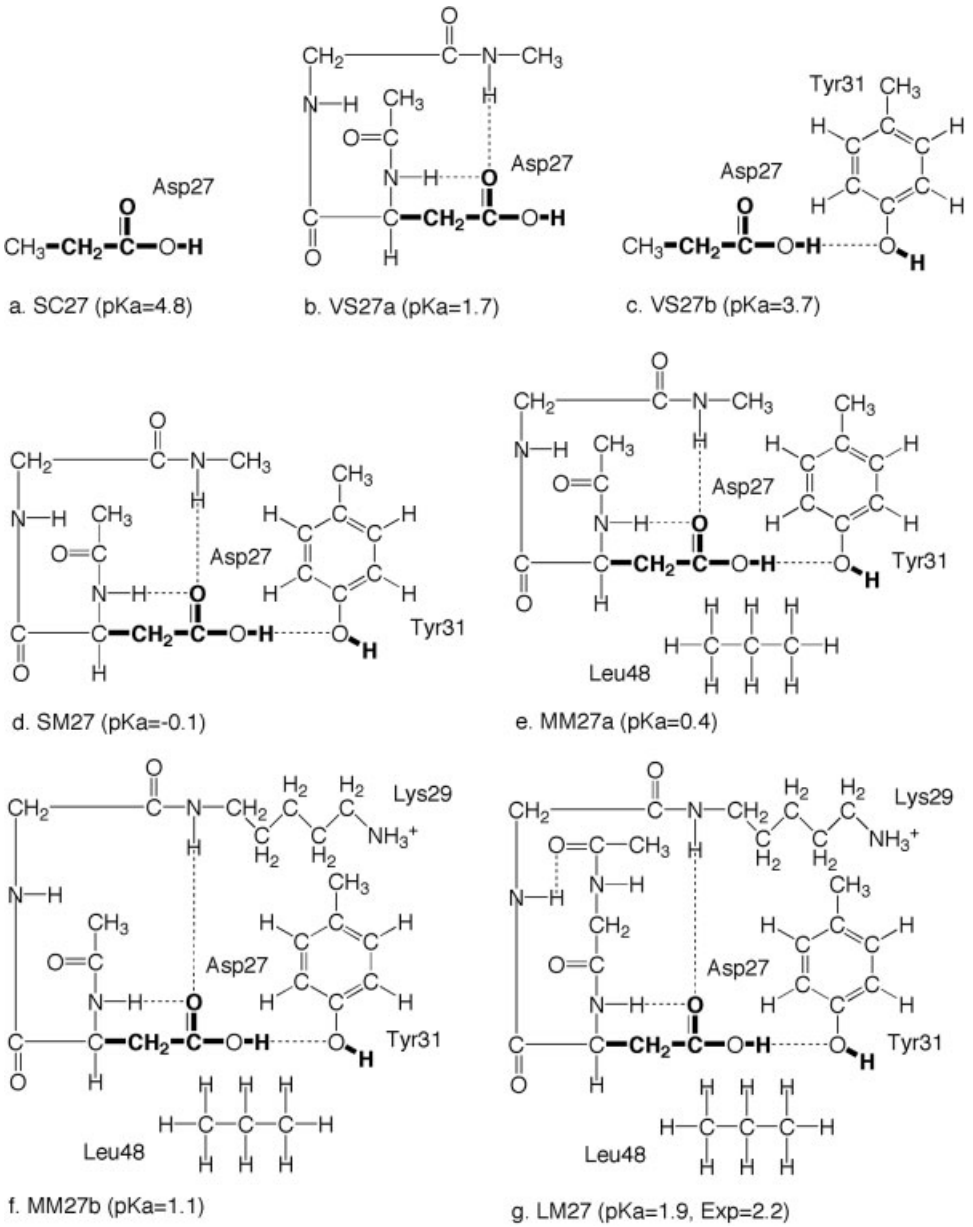


Fig. 4. Model compounds for Asp27 of OMTKY3 and their computed pK_a values. The positions of the atoms in bold were energy minimized. Acid form I is shown.

polarizable continuum model (PCM[67]) as implemented in GAMESS, which represents the solvent as a dielectric continuum surrounding the solute (see the appendix for details of the PCM calculations).

Protein Model Construction

The aim of the current study is to elucidate the determinants of the carboxyl pK_a values in OMTKY3 by constructing and analyzing the simplest possible structural model that consistently reproduces the experimental pK_a values. Thus, for each Asp and Glu residue, we construct model compounds that include the ionizable residues and their immediate chemical environments. First a "small model" is designed that includes 1) the side-chain of the ionizable Glu or Asp residue, 2) the two amide groups next to the C^{α} of the Glu or Asp side-chain, and 3) all groups that form hydrogen bonds with the carboxyl group of interest (Figs. 1–6).

The coordinates of the atoms in each model are taken from the PDB file 1PPF.[68] Hydrogen atoms were added to the PDB structure with the WHAT IF program[69,70] at pH = 7. Several new protons were added manually to satisfy the unfilled valences where atoms were removed in constructing the small model. For example, the small model of Glu10 [Fig. 2(b)] consists of $\underline{H_2}$CHNH-Glu-C(O)C(H)$\underline{H_2}$ (i.e., the four underlined hydrogen atoms were added manually, whereas the remaining protons were added by WHAT IF). All of the carboxyl side-chains were originally in the unprotonated form. The acid forms were obtained by adding the acidic protons to the carboxy-

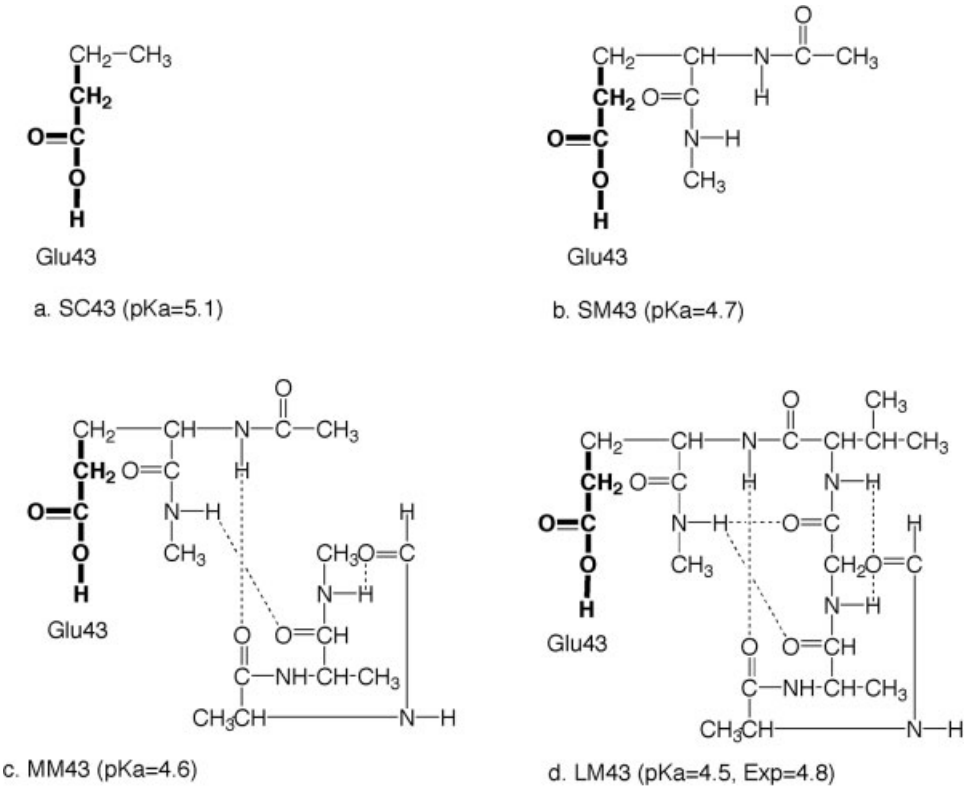


Fig. 5. Model compounds for Glu43 of OMTKY3 and their computed pK_a values. The positions of the atoms in bold were energy minimized. Acid form I is shown.

late groups. Two or three protonation sites (i.e., conformers of the COOH group) were considered for each acid form, whereas only one base form was considered. The total free energy of the acid form is taken to be the "conformational average" of the free energies of each conformer (G_i),[71]

$$G = -RT\ln\Big[\sum_{i}^{\text{conformers}} \exp(-G_i/RT)\Big]$$

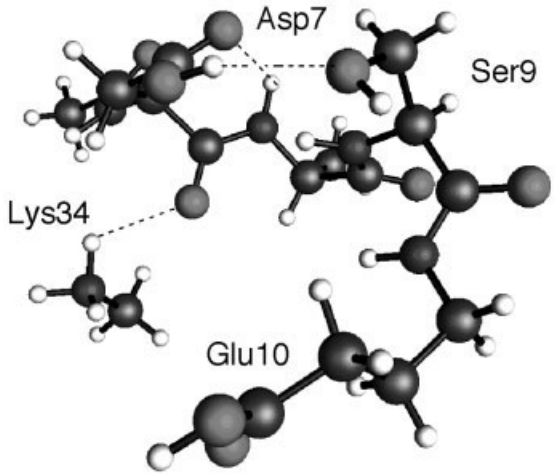
$$= G_0 - RT\ln\Big[1+\sum_{i\neq 0}^{\text{conformers}} \exp(-\Delta G_i/RT)\Big] \qquad (4)$$

where G_0 is the lowest energy conformer and $\Delta G_i = G_i - G_0$. From the latter form of Eq. 4, it can be seen that only low energy conformers ($\Delta G_i < 2RT \approx 1$ kcal/mol at 25°C) contribute significantly to the free energy. If significant, this contribution will lower the free energy of the acid, thereby increasing the pK_a. Physically, this pK_a increase is entropic in nature, because several accessible protonation states increase the protonation probability. In our study, this contribution is always less than ~0.2 pH units.

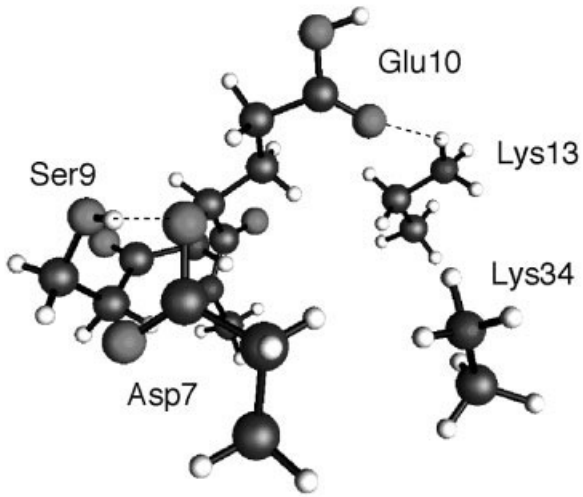
Because we seek the simplest possible computational model that consistently reproduces the experimental pK_a values, we optimize only a few structural parameters. The positions of the atoms in the carboxyl group (CH_2COO^- or CH_2COOH) are optimized by energy minimization, except that the Cartesian coordinates of one of the oxygens are kept fixed (except for Glu43). This allows for the carboxyl bond lengths and angles to adjust to the change in protonation state without greatly altering the overall structure. In addition, for Asp7, Glu19, and Asp27, the positions of the neighboring OH protons of Ser9, Thr17, and Tyr31, respectively, are also optimized because their positions are predicted to depend significantly on the protonation state of the carboxyl group. Optimized atoms for each model are indicated in Figures 1–6.

Often the pK_a predicted using the small model (SM) is quite close to the experimental value, (Table I) indicating that the most important intraprotein interactions are included in the SM. To analyze the interactions further, we construct several small models in which key hydrogen bonds are removed; these are called very small (VS) models (Figs. 1–6). We also construct a side-chain (SC) model, in which the peptide backbone atoms are replaced by a methyl group, to determine the effect of the peptide groups on the pK_a.

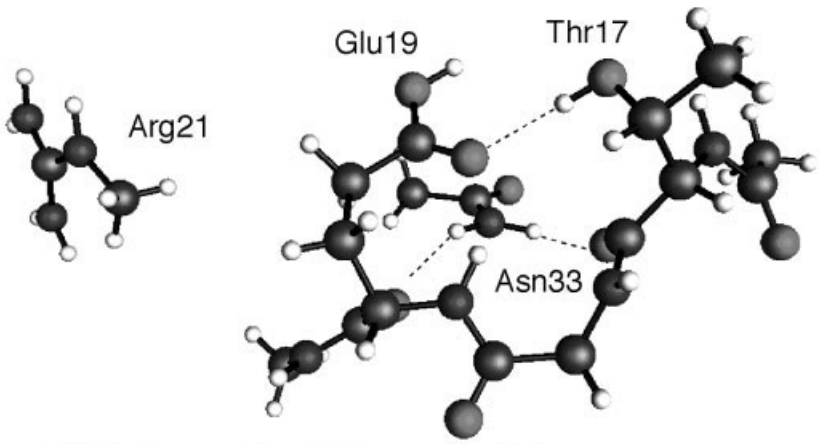
To determine the effect of protein groups not directly hydrogen bonded to the carboxyl group, we construct several medium models (MMs) in which side-chains in the vicinity of the carboxyl group are added one at a time, without geometry reoptimization. The model in which all of the neighboring groups are included is termed the large model (LM), and the pK_a obtained by using this large model is taken to be our best prediction of the experimental value. With the exception of Asp27, we are able to show



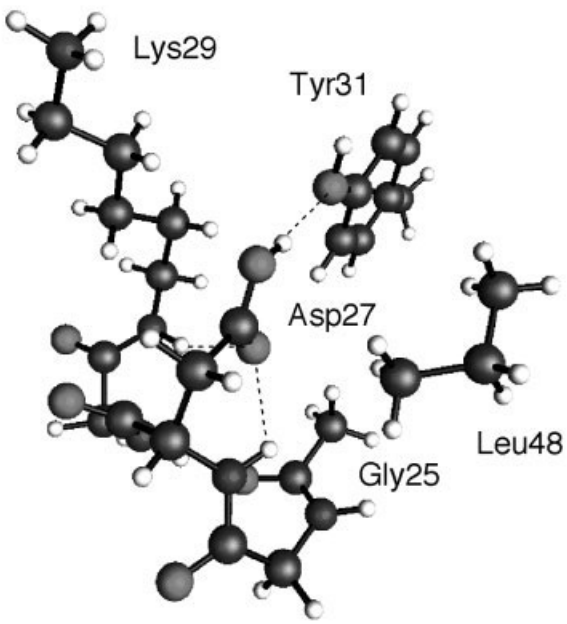
a. LM7, Comp pKa=2.4, Exp pKa=2.5



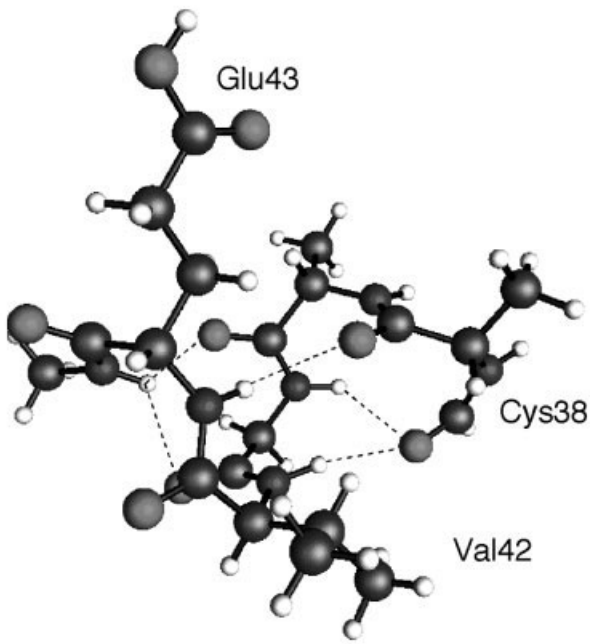
b. LM10, Comp pKa=4.3, Exp pKa=4.1



c. LM19, Comp pKa=2.7, Exp pKa=3.2



d. LM27, Comp pKa=1.9, Exp pKa=2.2



e. LM43, Comp pKa=4.5, Exp pKa=4.8

Fig. 6. Three-dimensional structures of the large model compounds for OMTKY3 and their pK_a values. Acid form I of the large model of each residue is shown (see Figs. 1–5).

TABLE I. Computed pK_a Values of the Model Compounds of the Five Carboxylic Residues in Turkey Ovomucoid Third Domain (OMTKY3)[†]

Models	Asp7	Glu10	Glu19	Asp27	Glu43	rmsd
Experiment[a]	2.5	4.1	3.2	2.2	4.8	
SM	2.1	4.0	3.0	−0.1	4.7	1.05
LM	2.4	4.3	2.7	1.9	4.5	0.31
SC	4.9	4.7	4.9	4.8	5.1	
VS-a	3.1		3.2	1.7		
VS-b	2.8		8.9	3.7		
VS-c	2.8					
MM-a	1.6	4.0	2.9	0.4	4.6	0.92
MM-b	2.9	3.9		1.1		
MM-c		4.1				

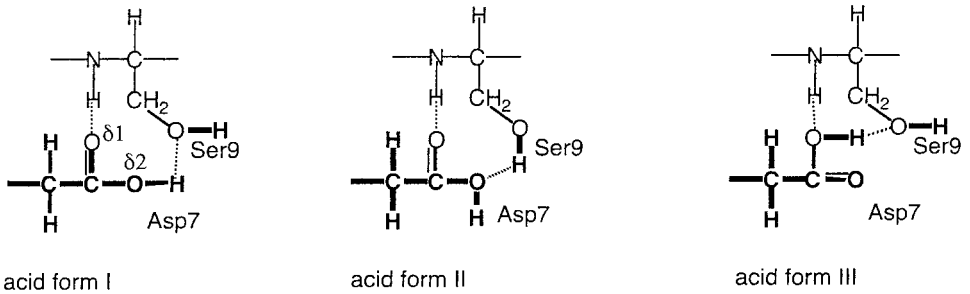
[†]The models are displayed in Figures 1–6. The pK_a values are calculated with the average free energy changes (Eq. 4 in text). The free energy changes and pK_a values for individual conformers are listed in Table II.
[a]Reference 40.

that the pK_a values change very little in going from the SM to the LM. For this reason, we have not considered models larger than the LMs.

RESULTS

Aspartate 7

The crystal structure shows two possible hydrogen bonds to the carboxyl group of Asp7: Ser9-O^{γ}H–$O^{\delta 2}$-Asp7 (O^{γ}-$O^{\delta 2}$ distance = 3.3 Å) and Ser9-NH—$O^{\delta 1}$-Asp7 (N-$O^{\delta 1}$ distance = 3.4 Å). Accordingly, the small model of Asp7 (SM7) includes these interactions in addition to the amide groups of Asp7 and Cys8 as shown in Figure 1(e). For the protonated form, three different proton positions are considered: Asp7-$O^{\delta 2}$H—O^{γ}H-Ser9, Asp7-H$O^{\delta 2}$—HO^{γ}-Ser9, and Asp7-$O^{\delta 1}$H—O^{γ}H Ser9 (acid forms I, II, and III, respectively).



On energy minimization of SM7, Asp7-$O^{\delta 2}$H—O^{γ}H-Ser9 (acid form I) is the lowest in energy, with forms II and III 0.5 and 2.2 kcal/mol higher in energy, respectively. All three contributions to the free energy are included via Eq. 4 and result in a pK_a of 2.1 pH units, in good agreement with the experimental value of 2.5. If only the lowest energy form of the acid (acid form I) is used, the pK_a is 1.9 (Table II).

Removal of the Ser9 side-chain [VS7b, Fig. 1(c)] or the Ser9 amide group [VS7c, Fig. 1(d)], followed by geometry reoptimization, results in pK_a increases of 0.7 pH units. If only the Asp7-$O^{\delta 2}$H—O^{γ}H Ser9 protonation state (acid form I) is used, then the pK_a increases in VS7b and VS7c relative to SM7 are 0.8 and 0.9, respectively. These values are very similar to the conformational average, and we use these results to interpret the molecular basis for the observed pK_a in terms of the electronic energy (ΔE_{ele}) and the change in solvation energy on deprotonation (ΔG_{sol}) listed in Table II. The pK_a increases on removal of the Ser9 side-chain and amide group are due to increases in ΔE_{ele} that are attenuated by smaller changes in solvation energy. Both changes are expected. The hydrogen bonds stabilize the base form, thereby increasing ΔE_{ele} relative to SM7. But formation of the hydrogen bonds leads to desolvation of the carboxyl group, which decreases the absolute value of ΔG_{sol}. It is of interest that the changes in ΔE_{ele} and ΔG_{sol} due to removal of the amide group are roughly twice that due to removal of the Ser9 side-chain. Finally, removal of both groups results in a pK_a of 3.1 [VS7a, Fig. 1(b), indicating that the effects of these hydrogen bonds on the pK_a of Asp7 are nonadditive: removal of the amide group raises the pK_a by 0.7 (SM7 → VS7c) or 0.3 (VS7b → VS7a), depending on whether it is removed in the presence or absence of the Ser9 side-chain, and similarly for the side-chain of Ser9. The data in Table II reveal that this non-additivity is primarily due to the solvation en-

TABLE II. Computed Free Energy Changes (in kcal/mol, see Eq 1–4 in text) and pK_a Values for the Individual Model Compounds of the Five Carboxylic Residues in Turkey Ovomucoid Third Domain (OMTKY3)

			Acid form I			Acid form II			Acid form III		
	Model	Neighboring groups	ΔE_{ele}	ΔG_{sol}	pK_a	−107.2	ΔG_{sol}	pK_a	ΔE_{ele}	ΔG_{sol}	pK_a
Asp7	SC	None	−0.4	−0.1	4.5	−5.2	3.8	3.8	−0.2	0.1	4.8
	VS-a	2 Amides	−11.8	9.3	3.0	−16.8	13.3	2.3	−13.2	10.1	2.6
	VS-b	3 Amides	−23.0	20.1	2.7	−27.5	23.9	2.2	−25.6	21.6	1.9
	VS-c	2 Amides, Ser9	−18.3	15.5	2.8	−26.3	22.3	1.9	−25.6	20.0	0.7
	SM	3 Amides, Ser9	−28.5	24.5	1.9	−35.6	31.1	1.6	−35.4	29.2	0.3
	MM-a	SM, Lys34	−84.6	79.8	1.3	−96.3	91.5	1.3	−92.3	85.3	−0.3
	MM-b	SM, Ser9, Glu10	−31.8	29.0	2.8	−39.6	35.9	2.1	−38.8	33.7	1.1
	LM	SM, Ser9, Lys34, Glu10	−87.0	83.5	2.3	−99.3	95.3	1.9	−94.8	88.9	0.5
Glu10	SC	None	−0.3	0.0	4.6	−1.0	−0.1	4.1			
	SM	2 Amides	−5.8	4.6	4.0	−6.9	4.6	3.2			
	MM-a	SM, Lys13	−68.5	67.1	3.8	−76.5	74.8	3.6			
	MM-b	SM,Lys13,Ser9,Lys34	-116.0	114.4	3.7	−127.0	125.0	3.4			
	MM-c	SM, Lys13, Ser9, Asp7	−18.7	17.5	4.0	−26.7	25.1	3.7			
	LM	SM, Lys13, Ser9, Asp7, Lys34	−67.5	66.4	4.1	−78.2	77.0	4.0			
Glu19	SC	None	−0.6	0.5	4.8	−1.3	0.6	4.3		Same as II	
	VS-a	2 Amides	−12.2	9.9	3.2	−15.9	11.0	1.3		Same as II	
	VS-b	2 Amides, Thr17	−9.9	15.4	8.9	−18.0	17.2	4.3	−21.6	15.6	0.4
	SM	4 Amides, Thr17	−25.6	23.0	2.9	−29.7	24.8	1.3	−32.2	23.4	−1.6
	MM	SM, Asn33	−31.6	28.9	2.9	−36.2	31.0	1.0	−38.9	29.6	−2.0
	LM	SM, Asn33, Arg21	−69.4	66.4	2.7	−71.7	65.9	0.6	−74.8	65.0	−2.3
Asp27	SC	None	0.2	−0.5	4.6	−4.6	3.7	4.2			
	VS-a	3 Amides	−30.6	25.8	1.3	−33.8	29.1	1.4			
	VS-b	Tyr31	−16.2	14.6	3.7	−23.3	18.3	1.2			
	SM	3 Amides, Tyr31	−41.3	34.6	−0.1	−51.1	40.8	−2.7			
	MM-a	SM, Leu48	−41.9	35.9	0.4	−51.3	41.6	−2.3			
	MM-b	SM, Leu48, Lys29	−95.5	90.4	1.1	−107.8	99.1	−1.5			
	LM	SM, Leu48, Lys29, Gly25	−95.5	91.4	1.8	−107.2	99.5	−0.8			
Glu43	SC	None	0.2	0.0	5.0	−0.2	−0.2	4.6			
	SM	2 Amides	−5.3	4.9	4.6	−6.5	5.5	4.1			
	MM	SM, Cys38, Asn39, Ala40	0.6	−1.2	4.4	−1.0	−0.2	4.0			
	LM	SM, Cys38, Asn39, Ala40, Val41-42	−0.3	−0.4	4.3	−2.0	0.8	4.0			

ergy. Removal of the amide group changes ΔE_{ele} and ΔG_{sol} by 5.5 and −4.4 kcal/mol (SM7 → VS7c) or 6.5 and −6.2 (VS7b → VS7a), depending on whether it is removed in the presence or absence of the Ser9 side-chain, and similarly for the side-chain of Ser9.

The low pK_a of VS7a suggests that interactions with the two neighboring amide groups have a significant effect on the pK_a: a decrease of 0.9 units relative to the usual model value[72] for Asp of 4.0 and 1.8 units relative to the side-chain model SC7 [Fig. 1(a)]. The decrease is presumably due to an interaction with the amide proton of Cys8. Given the (N)H—$O^{\delta 1}$ distance of 2.94 Å and N-H-O angle of 91°, the interaction with the charged carboxyl group is best classified as an ion-dipole interaction rather than a hydrogen bond. Gunner and coworkers[73] previously discussed the importance of backbone dipoles in depressing carboxyl pK_a values.

Larger, medium-sized models are created from SM7 by the addition of either the Lys34 side-chain [MM7a, Fig. 1(f)] or the Glu10 side-chain [MM7b, Fig. 1(g)]. The addition of Lys34 (N − $O^{\delta 2}$ distance = 5.3 Å) decreases the pK_a by 0.5 pH units relative to SM7. Conversely, addition of Glu10, which is neutral at pH < 3, increases the pK_a by 0.8 pH units, due to desolvation. Comparison with a large model LM7 [Fig. 1(h)], in which both groups are present, show that the effects are additive and result in a final pK_a of 2.4 pH units, in very good agreement with the experimental value of 2.5. This close agreement is somewhat fortuitous, because mutation of Lys34 to Gln or Thr has no measurable effect on the pK_a of Asp7,[7] whereas our model predicts a 0.5 unit decrease. This may indicate that the position of the Lys34 side-chain in the crystal structure does not reflect that in solution. Alternatively, the pK_a may be affected by interactions with the Gln and Thr side-chain at position 34 in the mutant protein, which are not included in the current study. In any case, our model indicates that the very low pK_a value of Asp7 is not primarily due to Lys34, in agreement with the experiment.

On the basis of our model of Asp7, the primary contributions to the observed 1.5 unit pK_a decrease relative to the standard Asp pK_a value of 4.0 are 1) a 1.0 unit decrease due to hydrogen bonds with the amide and OH groups of Ser9 and 2) a 0.9 unit decrease due to interactions with the two amide groups connected to the Asp7 C^{α}. Secondary

contributions to the pK_a value are 1) a 0.8 unit pK_a *increase* due to desolvation by (neutral) Glu10 and 2) a 0.5 unit decrease due to charge–charge interactions with Lys34.

Glutamate 10

The crystal structure of OMTKY3 does not reveal any intramolecular hydrogen bonds involving the Glu10 carboxyl group. Thus, the small model of Glu10 (SM10) was chosen to be the Glu10 side-chains and the two neighboring amide groups of Glu10 and Tyr11 [Fig. 2(b)]. For the acid form, protonation of both $O^{\epsilon 1}$ and $O^{\epsilon 2}$ were considered.

The pK_a of model SM10 is predicted to be 4.0 pH units, in excellent agreement with the experimental value of 4.1. Using only one acid form in which $O^{\epsilon 2}$ is protonated also results in a pK_a of 4.0. The decrease of 0.9 units in the pK_a relative to the side-chain model SC10 [Fig. 2(a)] is thus due to the two neighboring amide groups, presumably mostly due to stabilization of the base by the Glu10 amide NH, for which the (N)H—$O^{\epsilon 1}$ distance is 4.4 Å and the N-H-O angle is 106°.

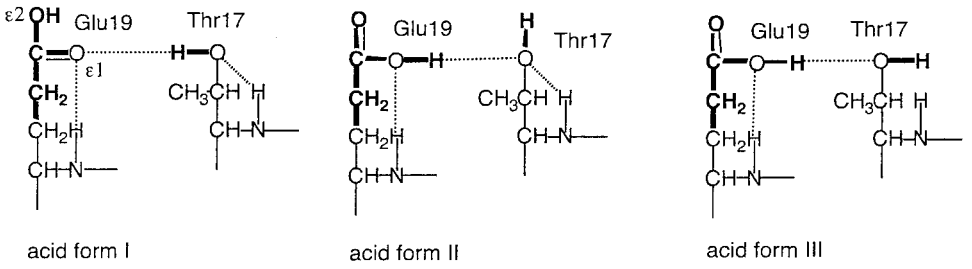
Further investigation of the crystal structure shows that the N^{ζ} of Lys13 is 3.9 Å from the nearest carboxyl oxygen of Glu10. It is surprising that including this residue as an ethyl ammonium group [MM10a, Fig. 2(c)] does not affect the pK_a. Inspection of Table II reveals an interesting explanation: the pK_a values of the two different acid forms are 3.8 and 3.6 units for $O^{\epsilon 2}H$ and $O^{\epsilon 1}H$, respectively. Thus, one effect of Lys13 is to make the two protonation states nearly isoenergetic. However, this increases the conformational entropy of the acid slightly (cf. Eq. 4), thus raising the pK_a back to 4.0, leading to a net zero effect.

The further addition of Lys34 and Asp7 [represented by methyl ammonium and propionate, Fig. 2(f), which are 5.4 Å and 6.1 Å from Glu10] raises the pK_a by 0.3 pH units. Removal of either Asp7 [Fig. 2(d)] or Lys34 [Fig. 2(e)] results in respective pK_a values of 3.9 and 4.1, respectively, indicating that the 0.3 pK_a increase is primarily due to Asp7. Our findings are consistent with the mutagenesis study of Forsyth and Robertson,[7] which showed that the pK_a for Glu10 in the K34T and K34Q mutants of OMTKY3 is within 0.2 pH units of the wild-type value.

Thus, the large model [Fig. 2(f)] yields a pK_a of 4.3, which is in good agreement with the experimental value of 4.1. The main determinant of this pK_a is interactions with the two neighboring amide groups in the protein backbone.

Glutamate 19

The crystal structure shows two possible hydrogen bonds to the carboxyl group of Glu19: Thr17-$O^{\gamma}H$—$O^{\epsilon 1}$-Glu19 (O^{γ}-$O^{\epsilon 1}$ distance = 3.0 Å) and Glu19-NH—$O^{\epsilon 1}$-Glu19 (N-$O^{\epsilon 1}$ distance = 2.76 Å). Accordingly, the small model of Glu19 (SM19) includes these interactions in addition to the amide groups of Leu18 and Thr17 as shown in Figure 3(d). For the acid form, three different proton positions were considered: Glu19-$HO^{\epsilon 1}$—HO^{γ}-Thr17, Glu19-$O^{\epsilon 1}H$—$O^{\gamma}H$-Thr17, and Glu19-$O^{\epsilon 1}H$—$O^{\gamma}H$-Thr17 (acid forms I, II, and III, respectively).



On energy minimization of SM19, acid form I is the lowest in energy, with the other two 2.3 (form II) and 6.2 (form III) kcal/mol higher in energy, respectively. All three contributions to the free energy are included via Eq. 4 and result in a pK_a of 3.0 pH units, in good agreement with the experimental value of 3.2. If only the lowest energy form of the acid (acid form I) is used, the pK_a is 2.9 pH units (Table II).

Removal of the Thr17 side-chain and neighboring amide groups [VS19a, Fig. 3(b)] results in only two acid forms and a slight pK_a increase of 0.2 pH units. This is in reasonable agreement with the 0.4 units increase in the pK_a of Glu19 in a T17V variant of OMTKY3, recently reported by Song et al.[41] Both acid forms are used in the pK_a prediction, although acid form I remains the lowest energy form. The data in Table II show that removal of Thr17 significantly increases ΔE_{ele} (by 13.4 kcal/mol), but that this increase is canceled by a decrease in the solvation energy (of 13.1 kcal/mol). Thus, the main source of the pK_a decrease relative to SC19 [Fig. 3(a)] is primarily due to the neighboring amide groups and the Glu19-NH—$O^{\epsilon 1}$-Glu19 hydrogen bond in particular.

Removal of the amide group from SM19 [VS19b, Fig. 3(c)] results in a very large pK_a of 8.9 pH unit. This is partially due to a relatively large change in the geometry on energy reminimization in the absence of the amide group.

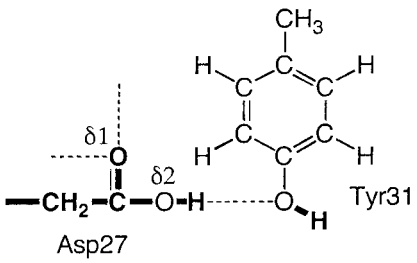
A medium model MM19 [Fig. 3(e)] and a large model LM19 [Fig. 3(f)] are created from SM19 by addition of the Asn33 side-chain and, subsequently, the Arg21 side-chain. These changes perturb the pK_a of Glu19 by −0.1 and 0.2 units, respectively. The latter pK_a shift is in good agreement with

the 0.2 unit increase in the pK_a of Glu19 in a N33A variant of OMTKY3, recently reported by Song et al.[41]

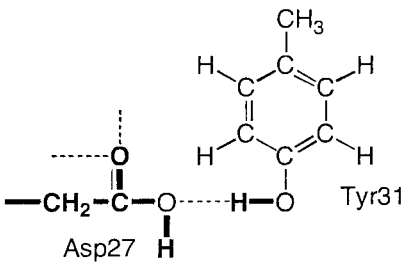
On the basis of our model of Glu19, the primary contribution to the observed 1.2 unit decrease relative to the standard Glu pK_a value[72] of 4.4 is a 1.2 unit decrease due to a hydrogen bond with the Glu19 amide NH. Secondary contributions to the pK_a value are 1) a 0.2 unit decrease due to a hydrogen bond with Thr17, 2) a 0.2 unit pK_a decrease due to charge–charge interactions with Arg21, and 3) a 0.1 unit decrease due to charge-dipole interactions with Asn33. The lack of a significant charge–charge interaction is surprising given that the pK_a of Glu 19 is increased from 3.2 to 4.0 in going from 10 mM to 1 M of KCl.[40] Such an ionic strength dependence is usually ascribed to screening of charge–charge interactions, whereas our calculations and the experiments of Song et al. show that the closest charged group (Arg21) has little (0.2 units) effect on the pK_a. We discuss this point further in the next section.

Aspartate 27

The crystal structure shows three possible hydrogen bonds to the carboxyl group of Asp27: Tyr31-$O^{\eta}H$—$O^{\delta 2}$-Asp27 (O^{η}-$O^{\delta 1}$ distance = 2.5 Å), Asp27-NH—$O^{\delta 1}$-Asp27 (N-$O^{\delta 1}$ distance = 2.9 Å), and Lys29-NH—$O^{\delta 1}$-Asp27 (N-$O^{\delta 1}$ distance = 2.9 Å). Accordingly, the small model of Asp27 (SM27) includes these interactions in addition to the amide group of Asn28 as shown in Figure 4(d). For the acid form, two different proton positions were used: Asp27-$O^{\delta 2}$H—O^{η}H-Tyr31 and Asp27-$HO^{\delta 2}$—HO^{η}-Tyr31 (acid forms I and II, respectively).



acid form I



acid form II

On energy minimization of SM27, acid form I is the lowest in energy, with form II 3.5 kcal/mol higher in energy. Both contributions to the free energy are included via Eq. 4 and result in a pK_a of −0.1 pH units, significantly less than the experimental value of 2.2. We show below that the pK_a is increased significantly in larger models.

Removal of the Tyr31 side-chain [VS27a, Fig. 4(b)] followed by geometry reoptimization results in a pK_a increase of 1.8 pH units. Comparison of the VS27a pK_a (1.7) to that of the SC27 [4.8, Fig. 4(a)] indicates that the amide hydrogen bonds are primarily involved in lowering the pK_a relative to propanoic acid.

A similar conclusion is reached by first removing the amide hydrogen bonds [to form VS27b, Fig. 4(c)], which increases the pK_a by 3.8 pH units. Subsequent removal of the Tyr31 side-chain (to form SC27) increases the pK_a further by 1.1 pH units. Thus, as for Asp7, the effects of hydrogen bonds on the pK_a are not additive.

Larger models are constructed from SM27 by adding the Leu48 side-chain [MM27a, Fig. 4(e)], followed by the Lys29 side-chain [MM27b, Fig. 4(f)], and finally part of the Ser26 and Gly25 main-chain [LM27, Fig. 4(g)]. These additions increase the pK_a by 0.5, 0.7, and 0.7 pH units, respectively, so that the pK_a of LM27 is 1.9, in good agreement with the experimental value of 2.2 pH units. All groups contain aliphatic protons that are within 3.0 Å of the carboxyl oxygens of Asp27: Leu48-$C^{\gamma 1}$H'—$O^{\delta 1}$ (2.84 Å), Leu48-$C^{\gamma 1}$H—$O^{\delta 2}$ (2.81 Å), Lys29-C^{β}H—$O^{\delta 1}$ (2.87 Å), Lys29-C^{δ}H—$O^{\delta 2}$ (2.98 Å), Gly25-C^{α}H—$O^{\delta 1}$ (2.80 Å). These rather weak interactions effectively desolvate the carboxyl group of Asp27, thereby raising the pK_a. It is especially interesting to note that the aliphatic portion of a Lys residue can increase the pK_a of a neighboring carboxyl group.

Glutamate 43

The crystal structure of OMTKY3 does not reveal any intramolecular hydrogen bonds to the Glu43 carboxyl group. Thus, the small model of Glu43 (SM43) was chosen to be the Glu43 side-chains and the two neighboring amide groups [Fig. 5(b)]. For the acid form, protonation of both $O^{\epsilon 1}$ and $O^{\epsilon 2}$ was considered.

The pK_a of model SM43 is predicted to be 4.7 pH units, in excellent agreement with the experimental value of 4.8. Because there are no strong interactions of the carboxyl group with the backbone, the pK_a is shifted only modestly from the SC43 [Fig. 5(a)]. Addition of the amide H-bonding partners [MM43, Fig. 5(c)] and additional nearby fragments of the main-chain [LM43, Fig. 5(d)] have only small (0.1 pH units) effects on the pK_a.

DISCUSSION

The single biggest contributor to low carboxyl pK_a values in OMTKY3 is backbone amide hydrogen bonding to the carboxyl oxygens. Table III lists the pK_a values together with the distances from the carboxyl oxygens to

TABLE III. Experimental pK_a Values of the Five Carboxylic Residues in Turkey Ovomucoid Third Domain (OMTKY3) and the Distances Between the Carboxyl Oxygens and the Nearest Amide Protons[†]

				Chemical shift changes (ppm)			
		Distances (Å)		Experimental[a]		Theoretical	
Residue	pK_a[a]	Amide 1	Amide 2	Amide 1	Amide 2	Amide 1	Amide 2
Asp27	2.2	1.99	2.33	−0.90	−0.85	−1.08	−0.56
Asp7	2.5	2.58	2.94	−0.39	NA	−0.42	+0.02
Glu19	3.2	1.87	3.84	−0.63	NA	−1.74	+0.14
Glu10	4.1	4.38	6.37	−0.20	NA	−0.08	+0.07
Glu43	4.8	5.08	5.16	NA	NA	NA	NA

[†]The chemical shift changes, both experimental and theoretical values of the neighboring amide protons on deprotonation of the carboxylic residues are also listed.
[a]Reference 40.

the nearest amide proton (based on crystal structures and proton positions determined using WHAT IF) and shows a good correlation between the number and strengths, as judged by the hydrogen bond distances, of these hydrogen bonds and the pK_a.

It is of interest that the chemical shifts of some of these amide protons are affected by the deprotonation of the neighboring carboxyl group,[40] as shown in the last column of Table III. Table III also lists changes in chemical shift on deprotonation, calculated by using the respective small models and the lowest energy acid form (see the appendix for computational details). With the exception of Glu19, the agreement with experiment is good (with errors < 0.3 ppm), especially considering the amide NH bond length is not optimized. It is not clear why the error for Glu19 is so much larger (1.11 ppm), but it does coincide with the largest error in the predicted pK_a (0.5 pH units) and shortest amide NH—carboxyl hydrogen bond considered here. Overall, the computed proton chemical shift changes further validate our model.

Gunner and coworkers[73] suggested that the electrostatic potential due to amide bonds will tend to lower pK_a values of Asp and Glu residues. This suggestion is consistent with our analysis, because hydrogen bonding has a significant eletrostatic component.[74] However, specific hydrogen bonds involving amide protons are rarely implicated when rationalizing pK_a shifts. Direct experimental study of the effects of these hydrogen bonds is significantly more difficult than side-chain–side-chain interactions, which can be probed by using mutagenesis. A recent survey of 250 nonhomologous protein X-ray structures found hydrogen bonds to the main chain for 59% and 35% of all Asp and Glu carboxyl groups, respectively.[75] Thus, such hydrogen bonds are likely to be key pK_a determinants for most Asp residues and many Glu residues.

Hydrogen bonds from side-chain hydroxyl groups of serine and tyrosine to the carboxyl groups of Asp7 and Asp27 have effects on the pK_a values that are similar to those of amide NH hydrogen bonds. In contrast, the carboxyl—HO-Thr hydrogen bond to Glu19 has a significantly smaller effect on the pK_a, in agreement with the experiment.[41] In silico mutation of Thr17 to Ser has little effect on the pK_a (data not shown), indicating that it is the hydrogen bond geometry, rather than the methyl group in the Thr side-chain, which is the main determinant of the pK_a shift. Other mutagenesis studies[76–78] have shown that hydrogen bonding to neutral and charged residues can affect the pK_a of ionizable residues by up to 1.6[76] and 2.4[78] units, respectively.

Analysis of the pK_a of Asp27 shows that neighboring hydrophophic regions can raise the pK_a by as much as 0.7 pH units. It is very interesting that the aliphatic part of the Lys29 side-chain is predicted to raise the pK_a of Asp27 by 0.5 pH units. The combined effect of the hydrophobic interactions on the pK_a of Asp27 is predicted to be 2.0 pH units. The importance of hydrophobic environments in determining the pK_a has been emphasized previously, in particular by Mehler, Warshel, and Garcia-Moreno.[9,10,18,23,27]

Neighboring charged residues such as Lys (for Asp7, Glu10, and Asp27) or Arg (in Glu19) are predicted to have more modest effects ($\leq$0.5 units) than hydrogen bonding on the OMTKY3 carboxyl pK_a values. In Glu19, this observation is supported by experiment:[41] the pK_a of Glu19 is increased by 0.2 and 0.4 units in R21A and T17V mutants of OMTKY3. Other mutagenesis studies[7,8,79] have shown that the neutralization of a charged residue (outside hydrogen bonding range) can change a pK_a by up to 0.7 units and that the effect tends to decrease significantly (to $\leq$ 0.2 units) for mutations of residues $>$ 10 Å from the ionizable residue. There are two notable exceptions to this statement. One exception is the mutation of Asp99 to Ser in *B. amyloliquefaciens* subtilisin, which decreases the pK_a of the distant (12 Å) His64 by 0.4 units.[5] This unusually large long-range effect may be due to the fact that His64 is largely screened from the solvent. Another exception is the mutation of Glu78 to Gln in *B. circulans* xylanase, which decreases the pK_a of nearby (5.3 Å) Glu172 by 2.5 units.[80] This unusually large effect may be due to a disruption of a hydrogen-bonding network that connects the two Glu residues.[80] Charge-reversal mutations introduce pK_a changes that are roughly twice the size of the pK_a changes introduced by the corresponding charge neutralization mutations. Because the charge–charge interactions appear to be roughly additive, multiple charge-reversal mutations within the same protein can produce pK_a shifts similar in magnitude to those commonly induced by a single hydrogen bond.[32,79,81] In the most extreme example, Laurents et al.[32] changed three Asp and

TABLE IV. Comparison Between the pK_a Values Predicted for OMTKY3 in the Current Study and Previous Studies

Residue	Experimental[a]	Current study[b]	Forsyth et al.[c]	Nielsen et al.[d]	Mehler et al.[e]	Havranek et al.[f]
Asp7	2.5	2.4	2.9	2.7	2.9	2.1
Glu10	4.1	4.3	3.4	3.6	4.1	4.0
Glu19	3.2	2.7	3.2	2.7	3.6	3.1
Asp27	2.2	1.9	4.0	3.4	3.3	2.9
Glu43	4.8	4.5	4.3	4.3	4.4	5.6
rmsd		0.3	0.9	0.7	0.6	0.5
Max. error		0.5	1.8	1.2	1.1	0.8

[a]Reference 40.
[b]The predicted pK_a values based on the large models (LMs) and B3LYP instead of MP2 are 2.2(Asp7), 4.4(Glu10), 2.2(Glu19), 2.2(Asp27), and 4.9(Glu43), with an RMSD = 0.5 and the maximum error = 1.0
[c]Reference 43.
[e]Reference 30.
[e]Reference 27.
[f]Reference 39.

two Glu residues to Lys residues in ribonuclease Sa, thereby introducing pK_a shifts of up to 2.2 units in the remaining 11 ionizable sites. However, the average pK_a shift was only 0.6 units, and six of the sites showed pK_a shifts of <0.3 units.

In general, we conclude the prime determinants of the Asp and Glu pK_a values in OMTKY3 are local interactions within ~4–5 Å of the ionizable residue. It is interesting that an ionic strength dependence of 0.8 units has been observed[40] for the pK_a of Glu19, which is traditionally ascribed to screening of long-range interactions. Given that hydrogen bonding appears to be the prime determinant of this pK_a, it is possible that an ionic strength increase may alter the hydrogen-bonding pattern around the residue, perhaps by ion binding at a specific site close to the ionizable residue.[34,82] The ionic strength dependence of the pK_a values of Asp7 and Asp27, for which we also predict that hydrogen bonding is the main determinant, varied by 0.0–0.6 and 0.4–0.5 units, respectively, depending on which proton resonance was used to determine the titration curve.[40]

Comparison With Classical Electrostatic Models

The pK_a values of OMTKY3 have been predicted by using a variety of classical electrostatic approaches.[27,30,39,43] Table IV lists the pK_a values obtained by using these methods along with the pK_a values predicted by using the large models in this study. Before comparing the results, we note that the classical methods are intended to predict pK_a values with relatively little user intervention, whereas our method is used to interpret pK_a values and in Asp7/Glu10 relies on the experimentally determined pK_a values to determine protonation states. However, it is clear that despite using relatively small protein models, our results are at least as good as the current literature values. Unlike the current approach, the classical methods all predict a relatively high pK_a for Asp27. In general, Schutz and Warshel[31] have noted that these methods tend to underestimate pK_a shifts.

We hope that this and future studies will help improve the more conventional pK_a prediction methods. Here we highlight some findings that may help guide such improvements.

1) Inspection of Table II shows that the placement of the acidic proton in the acid form of the carboxyl group can affect the pK_a by several pH units. This has also been observed for some residues in OMTKY3, BPTI, and lysozyme by Antosiewicz et al.[22] More generally, Nielsen and Vriend[30] showed that optimizing the hydrogen bond network for each protonation state tends to improve pK_a predictions. Here we note that the lowest energy acid form in the large models for all five residues has a C-C-O-H dihedral angle close to 180° (acid form I).

2) Our approach describes intraprotein interactions ab initio, using second-order perturbation theory (MP2) to describe electron correlation and a relatively large basis set [6-31+G(2d,p)]. The use of B3LYP density functional theory, rather than MP2 to describe electron correlation, leads to reasonable pK_a predictions except for Glu19 where the error increases to 1.0 pK_a units (see footnote b in Table IV). Neglecting electron correlation or using smaller basis sets leads to significantly poorer pK_a predictions (data not shown). Thus, significant improvements in pK_a predictions may be gained by using more detailed, and presumably more accurate, force fields as they become generally available.[83,84] For example, it is well known that using a detailed charge model allows for the use of a lower (and perhaps more physically realistic?) protein dielectric constant.[22]

Future Directions

Clearly, the generality of the conclusions presented here must be tested by the study of many more systems. We are currently performing a similar study of the carboxyl pK_a values in ubiquitin, for which a wealth of mutagenesis data are available.[8] Other studies on the pK_a of active site residues in lysozyme and xylanase are planned in the near future. The significance of different protonation states in calculating pK_a values, as described above, highlights the importance of sampling and including reasonable alternative conformations in the calculations. In this regard, additional studies will also be directed at extending the sampling of alternative conformations for the interacting groups.

ACKNOWLEDGMENTS

This work was supported by a Research Innovation Award from the Research Corporation, a grant from the National Science Foundation to JHJ, and a grant from the National Institutes of Health to ADR. HL gratefully acknowledges a predoctoral fellowship from the Center for Biocatalysis and Bioprocessing at the University of Iowa. The calculations were performed on IBM RS/6000 workstations and an eight-node QuantumStation obtained through NSF grants CHE 9974502 and MCB 0209941, respectively, and on computers at the Advanced Biomedical Computing Center at the National Cancer Institute at Frederick.

REFERENCES

1. Forsyth WR, Antosiewiez JM, Robertson AD. Empirical relationships between protein structure and carboxyl pK(a) values in proteins. Proteins 2002;48:388–403.
2. Edgcomb SP, Murphy KP. Variability in the pKa of histidine side-chains correlates with burial within proteins. Proteins 2002; 49:1–6.
3. Harris TK, Turner GJ. Structural basis of perturbed pK(a) values of catalytic groups in enzyme active sites. IUBMB Life 2002;53:85–98.
4. Nielsen JE, Borchert TV, Vriend G. The determinants of alpha-amylase pH-activity profiles. Protein Eng 2001;14:505–512.
5. 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.
6. 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.
7. 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.
8. 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;41:7586–7596.
9. Mehler EL, Fuxreiter M, Garcia-Moreno B. Role of microenvironments in modulating pKa shifts in proteins. Biophys J 2002;82: 1748.
10. Mehler EL, Fuxreiter M, Simon I, Garcia-Moreno EB. The role of hydrophobic microenvironments in modulating pK(a) shifts in proteins. Proteins 2002;48:283–292.
11. 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.
12. Lambeir AM, Backmann J, Ruiz-Sanz J, Filimonov V, Nielsen JE, Kursula I, Norledge BV, Wierenga RK. The ionization of a buried glutamic acid is thermodynamically linked to the stability of Leishmania mexicana triose phosphate isomerase. Eur J Biochem 2000;267:2516–2524.
13. Wada A. The alpha-helix as an electric macrodipole. Adv Biophys 1976;9:1–63.
14. Hol WGJ. The role of the alpha-helix dipole in protein function and structure. Prog Biophys Mol Biol 1985;45:149–195.
15. Richardson JS, Richardson DC. Amino-acid preferences for specific locations at the ends of alpha-helices. Science 1988;240:1648–1652.
16. Sancho J, Serrano L, Fersht AR. Histidine-residues at the N-termini and C-termini of alpha-helices—perturbed Pkas and protein stability. Biochemistry 1992;31:2253–2258.
17. 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.
18. Warshel A. Electrostatic basis of structure-function correlation in proteins. Acc Chem Res 1981;14:284–290.
19. Bashford D, Karplus M. pkas of Ionizable groups in proteins—atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
20. Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B. On the calculation of pkas in proteins. Proteins 1993;15:252–265.
21. Antosiewicz J, McCammon JA, Gilson MK. Prediction of Ph-dependent properties of proteins. J Mol Biol 1994;238:415–436.
22. Antosiewicz J, Briggs JM, Elcock AH, Gilson MK, McCammon JA. Computing ionization states of proteins with a detailed charge model. J Comput Chem 1996;17:1633–1644.
23. Sham YY, Chu ZT, Warshel A. Consistent calculations of pK(a)'s of ionizable residues in proteins: semi-microscopic and microscopic approaches. J Phys Chem B 1997;101:4458–4472.
24. Delbuono GS, Figueirido FE, Levy RM. Intrinsic pkas of ionizable residues in proteins—an explicit solvent calculation for lysozyme. Proteins 1994;20:85–97.
25. Warshel A, Sussman F, King G. Free-energy of charges in solvated proteins—microscopic calculations using a reversible charging process. Biochemistry 1986;25:8368–8372.
26. Kollman P. Free-energy calculations—applications to chemical and biochemical phenomena. Chem Rev 1993;93:2395–2417.
27. Mehler EL, Guarnieri F. A self-consistent, microenvironment modulated screened Coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. Biophys J 1999;77:3–22.
28. Sandberg L, Edholm O. A fast and simple method to calculate protonation states in proteins. Proteins 1999;36:474–483.
29. Matthew JB, Gurd FRN. Calculation of electrostatic interactions in proteins. Methods Enzymol 1986;130:413–436.
30. Nielsen JE, Vriend G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. Proteins 2001;43:403–412.
31. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 2001; 44:400–417.
32. Laurents DV, Huyghues-Despointes BMP, Bruix M, Thurlkill RL, Schell D, Newsom S, Grimsley GR, Shaw KL, Trevino S, Rico M, Briggs JM, Antosiewicz JM, Scholtz JM, Pace CN. Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI = 3.5) and a basic variant (pI = 10.2). J Mol Biol 2003;325:1077–1092.
33. Nielsen JE, McCammon JA. On the evaluation and optimization of protein X-ray structures for pKa calculations. Protein Sci 2003;12:313–326.
34. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. Biophys J 2002;83:1731–1748.
35. Koumanov A, Karshikoff A, Friis EP, Borchert TV. Conformational averaging in pK calculations: improvement and limitations in prediction of ionization properties of proteins. J Phys Chem B 2001;105:9339–9344.
36. van Vlijmen HWT, Schaefer M, Karplus M. Improving the accuracy of protein pK(a) calculations: conformational averaging versus the average structure. Proteins 1998;33:145–158.
37. Zhou HX, Vijayakumar M. Modeling of protein conformational fluctuations in pK(a) predictions. J Mol Biol 1997;267:1002–1011.
38. You TJ, Bashford D. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. Biophys J 1995;69:1721–1733.
39. Havranek JJ, Harbury PB. Tanford-Kirkwood electrostatics for protein modeling. Proc Natl Acad Sci USA 1999;96:11145–11150.
40. Schaller W, Robertson AD. Ph, ionic-strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid 3rd domain. Biochemistry 1995;34:4714–4723.
41. Song JK, Laskowski M, Qasim MA, Markley JL. NMR determination of pK(a) values for Asp, Glu, His, and Lys mutants at each variable contiguous enzyme-inhibitor contact position of the turkey ovomucoid third domain. Biochemistry 2003;42:2847–2856.
42. Swintkruse L, Robertson AD. Hydrogen-bonds and the pH-dependence of ovomucoid 3rd domain stability. Biochemistry 1995;34:4724–4732.
43. Forsyth WR, Gilson MK, Antosiewicz J, Jaren OR, Robertson AD. Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. Biochemistry 1998;37:8643–8652.
44. Li H, Hains AW, Everts JE, Robertson AD, Jensen JH. The prediction of protein pK(a)'s using QM/MM: the pK(a) of lysine 55 in turkey ovomucoid third domain. J Phys Chem B 2002;106:3486–3494.

45. Lim C, Bashford D, Karplus M. Absolute Pka calculations with continuum dielectric methods. J Phys Chem 1991;95:5610–5620.
46. Chen JL, Noodleman L, Case DA, Bashford D. Incorporating solvation effects into density-functional electronic-structure calculations. J Phys Chem 1994;98:11059–11068.
47. Richardson WH, Peng C, Bashford D, Noodleman L, Case DA. Incorporating solvation effects into density functional theory: calculation of absolute acidities. Int J Quantum Chem 1997;61:207–217.
48. Topol IA, Burt SK, Rashin AA, Erickson JW. Calculation of substituent effects on pK(a) values for pyrone and dihydropyrone inhibitors of HIV-1 protease. J Phys Chem A 2000;104:866–872.
49. Topol IA, Nemukhin AV, Dobrogorskaya YI, Burt SK. Interactions of azodicarbonamide (ADA) species with the model zinc finger site: theoretical support of the zinc finger domain destruction in the HIV-1 nucleocapsid protein (NCp7) by ADA. J Phys Chem B 2001;105:11341–11350.
50. Topol IA, Tawa GJ, Burt SK, Rashin AA. Calculation of absolute and relative acidities of substituted imidazoles in aqueous solvent. J Phys Chem A 1997;101:10075–10081.
51. Topol IA, Tawa GJ, Caldwell RA, Eissenstat MA, Burt SK. Acidity of organic molecules in the gas phase and in aqueous solvent. J Phys Chem A 2000;104:9619–9624.
52. Jang YH, Sowers LC, Cagin T, Goddard WA. First principles calculation of pK(a) values for 5-substituted uracils. J Phys Chem A 2001;105:274–280.
53. Schuurmann G, Cossi M, Barone V, Tomasi J. Prediction of the pK(a) of carboxylic acids using the ab initio continuum-solvation model PCM-UAHF. J Phys Chem A 1998;102:6706–6712.
54. da Silva CO, da Silva EC, Nascimento MAC. Ab initio calculations of absolute pK(a) values in aqueous solution I. Carboxylic acids. J Phys Chem A 1999;103:11194–11199.
55. Silva CO, da Silva EC, Nascimento MAC. Ab initio calculations of absolute pK(a) values in aqueous solution. II. Aliphatic alcohols, thiols, and halogenated carboxylic acids. J Phys Chem A 2000;104: 2402–2409.
56. Liptak MD, Gross KC, Seybold PG, Feldgus S, Shields GC. Absolute pK(a) determinations for substituted phenols. J Am Chem Soc 2002;124:6421–6427.
57. Liptak MD, Shields GC. Experimentation with different thermodynamic cycles used for pK(a) calculations on carboxylic acids using complete basis set and gaussian-n models combined with CPCM continuum solvation methods. Int J Quantum Chem 2001;85:727–741.
58. Liptak MD, Shields GC. Accurate pK(a) calculations for carboxylic acids using complete basis set and gaussian-n models combined with CPCM continuum solvation methods. J Am Chem Soc 2001;123:7314–7319.
59. Chipman DM. Computation of pK(a) from dielectric continuum theory. J Phys Chem A 2002;106:7413–7422.
60. Klicic JJ, Friesner RA, Liu SY, Guida WC. Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. J Phys Chem A 2002; 106:1327–1335.
61. Zheng F, Zhan CG, Ornstein RL. Theoretical determination of two structural forms of the active site in cadmium-containing phosphotriesterases. J Phys Chem B 2002;106:717–722.
62. Ullmann GM, Noodleman L, Case DA. Computing macroscopic and microscopic pK(a) vaules of the Rieske iron-sulfur cluster using a combined electrostatic-density functional method. J Inorg Biochem 2001;86:464.
63. Ullmann GM, Noodleman L, Case DA. Density functional calculation of pK(a) values and redox potentials in the bovine Rieske iron-sulfur protein. J Biol Inorg Chem 2002;7:632–639.
64. CRC Handbook of Chemistry and Physics, 83 ed. Cleveland, OH: CRC Press, 2003.
65. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, et al. General atomic and molecular electronic structure system. J Comput Chem 1993;14:1347–1363.
66. PQS version 2.4 PQS, Fayetteville, AR, http://www.pqs-chem.com; sales@pqs-chem.com. z.pqs-chem.com; sales@pqs-chem.com.
67. Cances E, Mennucci B, Tomasi J. A new integral equation formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics. J Chem Phys 1997;107:3032–3041.
68. Bode W, Wei AZ, Huber R, Meyer E, Travis J, Neumann S. X-ray crystal-structure of the complex of human-leukocyte elastase (Pmn Elastase) and the 3rd domain of the turkey ovomucoid inhibitor. EMBO J 1986;5:2453–2458.
69. Hooft RWW, Sander C, Vriend G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins 1996;26:363–376.
70. http://www.cmbi.kun.nl/gv/servers/WIWWWI/.
71. McQuarrie DA.Statistical thermodynamics. Mill Valley, CA: University Science Books, 1973.
72. Nozaki Y, Tanford C. Examination of titration behavior. Methods Enzymol 1967;11:715–734.
73. 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.
74. Jeffrey GA. An introduction to hydrogen bonding. New York, NY: Oxford University Press, 1997.
75. Eswar N, Ramakrishnan C. Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. Protein Eng 2000;13:227–238.
76. 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.
77. Wang PF, McLeish MJ, Kneen MM, Lee G, Kenyon GL. An unusually low pK(a) for Cys282 in the active site of human muscle creatine kinase. Biochemistry 2001;40:11698–11705.
78. Tishmack PA, Bashford D, Harms E, VanEtten RL. Use of H-1 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.
79. Lee KK, Fitch CA, Garcia-Moreno B. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. Protein Sci 2002;11:1004–1016.
80. 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 C-13-NMR study of Bacillus circuluns xylanase. Biochemistry 1996;35:9958–9966.
81. Fisher BM, Schultz LW, Raines RT. Coulombic effects of remote subsites on the active site of ribonuclease A. Biochemistry 1998;37: 17386–17401.
82. Waldron TT, Modestou MA, Murphy KP. Anion binding to a protein-protein complex lacks dependence on net charge. Protein Sci 2003;12:871–874.
83. Ren PY, Ponder JW. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. J Comput Chem 2002;23:1497–1506.
84. Kaminski GA, Stern HA, Berne BJ, Friesner RA, Cao YXX, Murphy RB, Zhou RH, Halgren TA. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. J Comput Chem 2002;23: 1515–1531.
85. Jensen F. Introduction to computational chemistry. West Sussex, England: John Wiley & Sons, Ltd, 1999.
86. Barone V, Cossi M, Tomasi J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. J Chem Phys 1997;107:3210–3221.
87. Miertus S, Scrocco E, Tomasi J. Electrostatic interaction of a solute with a continuum—a direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem Phys 1981;55:117–129.
88. Pierotti RA. A scaled particle theory of aqueous and nonaqueous solutions. Chem Rev 1979;76:717–726.
89. Floris FM, Tomasi J, Ahuir JLP. Dispersion and repulsion contributions to the solvation energy—refinements to a simple computational model in the continuum approximation. J Comput Chem 1991;12:784–791.
90. Zhan CG, Chipman DM. Reaction field effects on nitrogen shielding. J Chem Phys 1999;110:1611–1622.

APPENDIX

Here we provide the details of the computational methodology necessary to reproduce the presented results using the GAMESS program[65] and the PQS program.[66]

The electronic energies were computed at the MP2/6-31+G(2d,p) level of theory,[85] based on structures that are

fully optimized (only for propanoic acid and propionate) or partially optimized (for the model compounds of OMTKY3) in at the RHF/6-31G(d) level of theory.

The solvation energies were computed at the IEF-PCM[67]/RHF/6-31G(d) level of theory, based on the same geometries. Diffuse functions (L shell) were added to the four heavy atoms (two carbon and two oxygen atoms) in the negatively charged carboxylate groups ($-CH_2COO^-$) under consideration. The UAHF radii[86] were used to define the molecular cavities. RET = 100 was selected to prevent the use of additional spheres; 240 initial tesserae were used for each sphere. ICOMP = 2 was selected for surface charge renormalization.[87] The cavitation energy was computed with the method of Pierotti.[88] For propanoic acid/propionate and the small model compounds, the dispersion-repulsion energies were computed with the method of Floris et al.[89] For some of the very small, medium, and large model compounds, the solute is contained in two separate cavities, and the dispersion-repulsion energies could therefore not be calculated. So, for all the medium and large model compounds, the dispersion-repulsion energies of the corresponding small model compounds were used. For the very small models with two cavities, the dispersion-repulsion energies were not included in the pK_a calculations. This introduces virtually no error because the changes in the dispersion-repulsion energies on deprotonation are small (~0.2 kcal/mol) and virtually identical for the small, medium, and large model compounds of a given residue.

The NMR proton chemical shifts were computed at the B3LYP/6-311++G(d,p) level of theory using the PQS program. The effects of the solvent on the proton chemical shifts are modeled by the method proposed by Zhan and Chipman.[90] First, an IEF-PCM energy was calculated for a particular structure at the B3LYP/6-311++G(d,p) level of theory using GAMESS, 60 initial tesserae per atom, and ICOMP = 2. The resulting apparent surface charges and their positions were printed out in a format appropriate for PQS, scaled by one half (again following Zhan and Chipman), and included in the PQS B3LYP/6-311++G(d,p) chemical shielding calculations as static charges.

Computational Programs

Geometry optimization, some MP2 single-point energy calculation, and IEF-PCM computation were performed with the GAMESS program.[65] The PCM option of 240 initial tesserae for each sphere was implemented into a local version of GAMESS. The code will be released in the July 2003 of GAMESS. MP2 single-point energy calculations for the largest systems were performed with the PQS program[66] on a QS8 QuantumStation.