Electrostatic Calculations of Side-Chain pK_a Values in Myoglobin and Comparison with NMR Data for Histidines†

Donald Bashford,* David A. Case,* Claudio Dalvit, Linda Tennant, and Peter E. Wright

Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

Received January 26, 1993; Revised Manuscript Received May 21, 1993

ABSTRACT: Site-specific titration curves for 12 histidine residues in carbon monoxy sperm whale myoglobin (MbCO) have been determined from two-dimensional (2D) double quantum NMR experiments. Eight of these histidine residues are observed to titrate over the accessible pH range, and pK_a values have been determined; bounds on the titration midpoints of the remaining four histidines are also reported. Results for residues 48, 81, and 119 differ significantly from those estimated from earlier, one-dimensional studies, but they are in good agreement with values recently determined for metaquomyoglobin. These experimental values (plus those determined earlier for tyrosine titrations) are compared to predictions from crystal structures of myoglobin using a numerical Poisson–Boltzmann model and a Monte Carlo treatment of the multiple-site titration. An extension of existing models is described that accounts for alternate tautomers for histidines. Calculations are reported using several choices for radii and charges, and for five crystal structures, in order to assess the sensitivity of the results to details of the calculations. In general, the agreement between calculated and observed titration behavior suggests that this theoretical model captures much of the electrostatic behavior in this system, even though it ignores conformational fluctuations and the differences in mean structures that may exist between crystal and solution. Interactions among titrating groups are often important; in general, these interactions lead to more gradual individual site titrations (the mean Hill coefficient is about 0.8), and in several cases the interactions are so strong that two side chains need to be considered as a unit and single residues may participate in two-step titrations. It is suggested that histidines involved in such two-step titrations and carboxylic acid residues with abnormally low pK_a values in the native conformation may be involved in the acid-induced partial unfolding of MbCO.

The importance of electrostatic effects in understanding the behavior of globular proteins has been recognized in many contexts, including enzymatic catalysis, ligand binding, and protein–protein association (Perutz, 1978; Warshel & Russell, 1984; Honig *et al.*, 1986). Among the many manifestations of these interactions are conformation-dependent shifts in titration curves for side chains in proteins, compared to values seen in unstructured peptides. With the advent of site-specific NMR assignment procedures, it has become possible to identify the titration behavior of individual residues, although in most cases reported to date, only a small number of titrating residues in any given protein has been analyzed. One exception is lysozyme, for which p$K_{1/2}$ values have been measured for the non-arginine side chains (Kurimatsu & Hamaguchi, 1980). This data was recently compared to calculations based upon the crystal structure of lysozyme and a continuum electrostatic model (Bashford & Karplus, 1990), and the results were in good general agreement with the observed titration behavior. Here we extend these ideas to another protein, myoglobin, comparing experimental and calculated results for 11 histidine and 3 tyrosine residues. The results provide additional evidence concerning the extent to which these continuum electrostatic models can describe fairly subtle differences in the electrostatic environment of individual residues in a protein.

The titration behavior of histidines in myoglobin was the subject of earlier studies using a modified Tanford–Kirkwood approximation in which the electrostatic equations are solved in a spherical geometry (Shire *et al.*, 1974; Botelho *et al.*, 1978; Friend & Gurd, 1979; Matthew *et al.*, 1985). At that time, unambiguous site-specific assignments were not available, and identification of individual histidine resonances had to be made by *ad hoc* criteria, including comparisons of spectra for myoglobins of different species (Botelho & Gurd, 1978). Recently, two- and three-dimensional (2D and 3D) NMR techniques have been applied to the carbon monoxy form of myoglobin to obtain unambiguous assignments of backbone and side-chain resonances (Dalvit & Wright, 1987; Theriault *et al.*, 1993). Histidine resonances can be followed unambiguously as a function of pH *via* 2D homonuclear double quantum NMR experiments. Similar work on metaquomyoglobin has also been reported (Cocco *et al.*, 1992). This new evidence prompted us to reevaluate the theoretical interpretation of such titration behavior.

During the past decade, it has also become possible to use numerical computer techniques to solve electrostatic problems in geometries that are more realistic than those assumed in the Tanford–Kirkwood (TK) model. In addition to assuming a spherical geometry, the TK model and its variants assume that each titrating site is the same distance below the molecular surface. This has the effect of assuming that the Born energy (the work required to move a change from the solvent into the protein interior) is the same for all titrating groups. The results reported below, along with earlier studies (Bashford & Karplus, 1990), suggest that in this respect the TK model is missing an important feature of the energetics. In addition, through comparisons with appropriate model compounds one can now estimate absolute titration behavior in proteins, eliminating the need to assign intrinsic pK values to particular residues.

† This work was supported by NIH Grants HL40453 (D.A.C.), GM45607 (D.B.), and DK34909 (P.E.W.).

0006-2960/93/0432-8045$04.00/0

EXPERIMENTAL METHODS

Sperm whale myoglobin (Sigma) was purified using previously described methods (Mabbutt & Wright, 1985). Samples of MbCO were prepared for NMR by flushing metmyoglobin under a stream of carbon monoxide, reducing with a small amount of solid sodium dithionite, and then removing excess dithionite by passage through a column of Sephadex G-15 equilibrated with a CO-saturated D_2O buffer containing potassium phosphate (50 mM) and EDTA (0.1 mM). The pH titration was carried out by preparing 13 samples of MbCO (*ca.* 2 mM) in this phosphate buffer at 13 pH values between pH 4.92 and 8.80. This was mostly achieved by equilibration of the Sephadex column with buffers of varying pH. The pH of some MbCO samples was changed by dialysis against the phosphate buffer or by mixing samples of different pH. A further sample at pH 4.67 was prepared by diluting a sample of self-buffered MbCO in D_2O with 500 mM deuteriosuccinate/phosphate buffer to give a final buffer concentration of 60 mM. All pH values quoted are uncorrected meter readings recorded at 35 °C after the NMR experiments.

^{1}H NMR spectra were recorded at 35 °C on a Bruker AM500 spectrometer equipped with digital phase-shifting hardware. Two-dimensional phase-sensitive double quantum spectra were acquired and processed as described previously (Dalvit & Wright, 1987) using the standard pulse sequence (Braunschweiler *et al.*, 1983). Spectra were recorded with double quantum excitation periods of 60 and 100 ms to enhance the histidine $C^{\delta}H$–$C^{\epsilon}H$ cross peaks (Dalvit & Wright, 1987). The pK values were determined by a nonlinear least-squares fit of the titration data.

THEORY

Calculation of Electrostatic Terms. The main methods used to calculate the electrostatic contributions to titration in proteins have been described previously (Bashford & Gerwert, 1992) and are outlined briefly here. It is assumed that the difference in the titration behavior of an ionizable group in a protein and in a model compound can be accounted for by calculating the difference in the electrostatic work of altering the charges from the unprotonated to the protonated state in the protein and the work of making the same alteration in the model compound. The electrostatic potential is assumed to be given by a linearized Poisson–Boltzmann equation in which the protein interior has a low dielectric value, the solvent has a high dielectric value, and the counterion (Boltzmann) terms are excluded from a region near the protein surface. By using finite difference methods to solve the Poisson–Boltzmann equation, we are able to take into account the detailed shape of the protein surface as defined by the atomic coordinates of the protein.

There are three kinds of electrostatic work terms: $\Delta\Delta G_{\text{Born}}$, the interaction of the titrating charges with the polarization that these changes themselves induce in their surroundings; $\Delta\Delta G_{\text{back}}$, the interaction of the titrating charges with nontitrating (background) charges, such as the peptide dipoles; and W_{ij}, the interaction of the charges of one titrating group with the charges of another titrating group. These terms are calculated by the following formulae:

$$\Delta\Delta G_{\text{Born}} = \frac{1}{2}\sum_i Q_i^{(\text{p})}[\phi^{(\text{p})}_{\text{protein}}(\mathbf{r}_i) - \phi^{(\text{p})}_{\text{model}}(\mathbf{r}_i)] - \frac{1}{2}\sum_i Q_i^{(\text{u})}[\phi^{(\text{u})}_{\text{protein}}(\mathbf{r}_i) - \phi^{(\text{u})}_{\text{model}}(\mathbf{r}_i)] \quad (1)$$

$$\Delta\Delta G_{\text{back}} = \sum_j q_j[\phi^{(\text{p})}_{\text{protein}}(\mathbf{r}_j) - \phi^{(\text{u})}_{\text{protein}}(\mathbf{r}_j)] - \sum_j q_j[\phi^{(\text{p})}_{\text{model}}(\mathbf{r}_j) - \phi^{(\text{u})}_{\text{model}}(\mathbf{r}_j)] \quad (2)$$

$$W_{\mu\nu} = \sum_i [Q^{(\text{p})}_{\mu,i} - Q^{(\text{u})}_{\mu,i}][\phi^{(\text{p})}_{\text{protein},\nu}(\mathbf{r}_i) - \phi^{(\text{u})}_{\text{protein},\nu}(\mathbf{r}_i)] \quad (3)$$

where Q and q denote titrating and nontitrating charges, respectively, ϕ_{protein} and ϕ_{model} are potential fields calculated by the finite difference method in the protein and model compound, respectively, and the superscripts (p) and (u) indicate the protonated and unprotonated charge states, respectively. The fields are calculated with only the charges of the titrating site under consideration present. At this point, it is convenient to introduce a hypothetical quantity, pK_{intr}, or intrinsic pK (Tanford & Kirkwood, 1957). This is defined as the pK_a that a titrating group in the protein would have if all other groups were held in their neutral protonation states. Given the known model compound, pK, pK_{model}, pK_{intr} can be calculated by

$$\text{p}K_{\text{intr}} = \text{p}K_{\text{model}} - \frac{1}{2.303k_{\text{B}}T}[\Delta\Delta G_{\text{Born}} + \Delta\Delta G_{\text{back}}] \quad (4)$$

The Poisson–Boltzmann calculations were carried out using the MEAD (Macroscopic Electrostatics with Atomic Detail) suite of programs. It is necessary to solve the potential generated by charge sets corresponding to the protonated and deprotonated sets of each titrating site, one site at a time. For each charge set and site, an initial finite difference solution is calculated using a coarse outer cubic lattice having 81^3 nodes and a lattice spacing of 1.0 Å with its center on the protein center. This is followed by a fine inner lattice having 81^3 nodes and a lattice spacing of 0.25 Å. The corresponding model compound calculations used lattices having 61^3 nodes and the same lattice spacings. All calculations assumed an ionic strength of 0.1 M. A more complete description of the computational scheme has been presented elsewhere (Bashford & Karplus, 1990; Bashford & Gerwert, 1992).

A protonation microstate of a protein with N titratable sites can be specified by an N-element vector, $\mathbf{x}$, whose elements, x_i, each correspond to one site and whoses values are zero or one according to whether the corresponding site, i, is deprotonated or protonated, respectively. The relative free energy for forming the microstate, $\mathbf{x}$, is

$$\Delta G(x) = \sum_{i=1}^{N} x_i 2.303k_{\text{B}}T(\text{pH} - \text{p}K_{\text{intr}}) + \frac{1}{2}\sum_{i,j}^{N} W_{ij}(q_i^0 + x_i)(q_j^0 + x_j) \quad (5)$$

The fraction of molecules having site i protonated at a particular pH is found by a Boltzmann weighted average using eq 5 for the energy:

$$\theta_i = [\sum_{\{\mathbf{x}\}} x_i \exp(\sum_{\mu} x_\mu 2.303(\text{p}K_{\text{intr},\mu} - \text{pH}) - \beta\sum_{\mu,\nu}(x_\mu + q_\mu^0)(x_\nu + q_\nu^0)W_{\mu\nu})] \Big/ [\sum_{\{\mathbf{x}\}} \exp(\sum_{\mu} x_\mu 2.303(\text{p}K_{\text{intr},\mu} - \text{pH}) - \beta\sum_{\mu,\nu}(x_\mu + q_\mu^0)(x_\nu + q_\nu^0)W_{\mu\nu})] \quad (6)$$

where the $\{\mathbf{x}\}$ indicates summation over all 2^N possible protonation states (Bashford & Karplus, 1991). This summation is not possible in practice; therefore, an approximate Monte Carlo procedure is used instead (Beroza *et al.*, 1991).

We calculated the protonation state of all titratable residues from pH −4 to +15 in 0.1 pH unit increments using 10 000 cycles each of equilibration and sampling. Each cycle consists of N attempts to change the protonation state of a randomly chosen site, where N is the total number of sites (73 for myoglobin). For 6% of the attempted protonation flips, the program simultaneously flipped the site most strongly coupled to the site initially being flipped; for 4% of the trials, the simultaneous flip involved the second most strongly coupled site. These protonation "swaps" have been shown to accelerate convergence of the Monte Carlo sampling significantly (Beroza *et al.*, 1991). Reported p$K_{1/2}$ values are simply the midpoints of these titration curves.

Inclusion of Tautomers. Previous titration calculations have generally assumed that a single neutral tautomeric form is present for residues such as His, Arg, Asp, and Glu, even though it is clear that alternatives are possible. Since one form may be favored over the other by the protein environment, this could have consequences for the resulting energetics. Earlier studies have used a symmetrical charge distribution for the protonated forms of aspartic and glutamic acids, effectively averaging the two symmetrical protonated forms. We have followed this idea in the present calculations, but note that it is most likely to break down for relatively buried side chains that can form specific hydrogen bonds with acceptor groups; these are of course the groups most likely to have abnormal values for p$K_{1/2}$.

A more significant problem is likely to occur with histidine, since the two potential protonation sites are chemically distinct, and since steric interactions are more likely (than with Asp or Glu) to prevent side-chain rotations that would average the charge distributions. In effect, histidine should be considered as a three-state system, with a doubly protonated imidazolium form and two neutral forms: one with a proton on N^{ϵ} and a second with N^{δ} protonated. (The negative ion, with both sites deprotonated, is not expected to make a significant contribution under most conditions.) Since earlier calculations considered only two-state models, we outline here a way in which this tautomeric freedom can be incorporated into the usual framework of protein titration calculations.

In our approach, a histidine side chain is represented by two "titration" sites with appropriate corrections made in the energy formulas to prevent the negative ion from appearing and to account for the fact that other charges in the protein see only a single histidine, and not two sites. In this way, titration programs that expect two-state models (Bashford & Karplus, 1991; Beroza *et al.*, 1991) can be used directly to handle this more complex behavior.

In practice, this means that we carry out continuum electrostatic calculations as described earlier (Bashford & Karplus, 1990) for a model in which all histidines are assumed to titrate to an N^{δ} tautomer, followed by a calculation in which all histidines titrate to N^{ϵ}. Finally, the interaction energies among the histidines are evaluated by considering interactions between the N^{δ} form of each histidine with the N^{ϵ} form of every other histidine. Since the time-consuming step of the electrostatic calculations lies in the calculation of the fields arising from the protonation of each residue, the computational burden of this procedure is not as great as it might sound: basically, two electrostatics calculations need to be done for each histidine residue (one for each tautomer), and the remainder of the interactions are very quickly evaluated.

Choice of Parameters and Structures. A main goal of this study is to determine the sensitivity of the results to (inevitable) uncertainties arising from choices of charge models, radii, and coordinates. We have performed two major sets of calculations: one set in which the coordinates were kept fixed and various "logical" choices of charges and radii were made, and a second set in which the charges and radii were kept fixed and coordinates from a variety of structure determinations were used.

Table I: Representative Born Radii[a]

atom	Bondi[b]	OPLS[c]	OPLS[d]	CHARMM[e]	CHARMM[f]	LBK[g]	RN[h]	GH[i]
H	1.20	1.15		0.54	0.60	1.20	1.16	1.20
C	1.70	1.79	2.01	1.87	2.10	2.10	2.46	
N	1.55	1.54	1.73	1.43	1.60	1.60	1.50	1.50
O	1.50	1.39	1.56	1.43	1.60	1.60	1.50	1.60

[a] Carbon and oxygen radii for carboxylic acid and nitrogen and hydrogen for charged amine groups. [b] From Bondi (1964). [c] From Jorgensen and Tirado-Rives (1988), using $\sigma/2$; hydrogen value is that used in this work. [d] Same as *c*, using $R_{min}/2$. [e] From Brooks et al. (1983), using $\sigma/2$. [f] Same as *e*, using $R_{min}/2$. [g] From Lim et al. (1991), "adjusted" for acetic and formic acids, imidazolium, and methylammonium. [h] From Rashin and Namboodiri (1987). [i] From Gilson and Honig (1988); value for carbon not given.

The five parameter sets used are denoted as follows. Amber/Bondi: The charges are taken from the AMBER empirical potential function (Weiner *et al.*, 1986), and the radii are "standard" van der Waals radii from Bondi (1964). C19/Bondi: Charges are from the CHARMM19 empirical potential function (Brooks *et al.*, 1983), with "standard" radii as above. C19/Rmin: Charges are from the CHARMM19 empirical potential function, and radii are from the minimum of the CHARMM19 van der Waals function. C19/σ: Charges are from the CHARMM19 empirical potential function, and radii are from the $\sigma/2$ (the zero potential point) of the CHARMM19 van der Waals function. OPLS: Charges are from the OPLS potential energy function (Jorgensen & Tirado-Rives, 1988), and radii are derived from the OPLS σ parameter according to 0.5σ–0.09, as recommended by Jean-Charles *et al.* (1991).

The OPLS and CHARMM charges have been used in a number of previous studies of solvation energies (Jean-Charles *et al.*, 1991; Mohan *et al.*, 1992) and pK behavior in proteins (Bashford & Karplus, 1990; Soman *et al.*, 1989). The AMBER charges were derived from fitting atom-based charges to electrostatic potentials determined from quantum mechanical calculations on amino acid fragments; these should give a good model for the fields generated at distant points.

The choice of dielectric radii can also have an important influence on calculations of ionization behavior, particularly when side chains are partially buried, since the extent of burial can depend strongly upon the radii chosen. A comparison of various proposals for Born atomic radii is given in Table I. R_{min} and σ values in empirical force fields are generally chosen as a compromise to approximately fit many types of data (such as torsional barriers, hydrogen bonds, liquid densities, and so on) that are not of direct importance to electrostatic calculations, so that their use in such calculations may be disadvantageous. The "standard" van der Waals radii, however, are close to those used in earlier studies of molecular association (Gilson & Honig, 1988) and close to the "optimized" values determined by Lim *et al.* (1991) for a variety of small fragments of amino acid side chains.

The crystallographic structures of five different forms of myoglobin used are denoted as follows. COx: Sperm whale carbon monoxy myoglobin 1.5-Å X-ray structure (Kuriyan *et al.*, 1986) (Brookhaven structure 1MBC). COn: A recently refined neutron diffraction structure of the same crystal as

Table II: Model Compound pK_a Values

residue type	pK_{model}	residue type	pK_{model}
Arg	12.0	His ϵ tautomer	6.6
Asp	4.0	Lys	10.4
C-terminus	4.0	N-terminus	7.5
Glu	4.4		
heme propionate	4.4		
His δ tautomer	7.0	Tyr	9.6

COx (Cheng & Schoenborn, 1990, 1991) (Brookhaven structure 2MB5). oxy: Sperm whale oxy myoglobin 1.8-Å X-ray structure at pH 8.4 (Phillips, 1980) (Brookhaven structure 1MBO). deoxy: Deoxy myoglobin 2.0-Å crystal structure (Takano, 1977) (Brookhaven structure 3MBN). MetMut: Metmyoglobin sperm whale sequence expressed in bacteria with initiator Met and Asp122 → Asn, which crystallizes in a crystal form different from that of the native protein (Phillips *et al.*, 1990; Phillips, 1990) (Brookhaven structure 1MBW). Only the first two of these structures exactly match the protein for which we have experimental evidence of $pK_{1/2}$ values for histidines. The primary purpose of the comparison is to establish the extent to which the relatively small changes in coordinates affect our results.

The choice of dielectric parameters and model compounds is similar to that in previous work (Bashford & Karplus, 1990): The protein or model compound is assumed to have an internal dielectric constant of 4 and the surrounding water a dielectric constant of 80 and an ionic strength of 0.1. The boundary between protein and solvent is defined by the contact and reentrant surfaces (Richards, 1977) of a 1.4-Å sphere rolling on the atomic radii of the atoms in the molecule. The counterion exclusion zone extends 2.0 Å beyond the atomic radii. The model compounds are the *N*-formyl-*N*-methyl amide derivatives of the amino acid, and their coordinates are taken as the coordinates of the corresponding amino acid residue in the protein along with its flanking peptide groups. The pK_{model} values used are given in Table II. For most residues they are taken from data on the titration behavior of denatured proteins (Nozaki & Tanford, 1967).

For calculating histidine tautomerism by the technique described above, it is necessary to choose pK_{model} values for each tautomer. These correspond to the microscopic equilibrium constants for going from a specific tautomer to the doubly protonated state. Tanokura (1983) has made NMR studies of imidazole compounds with their imidazole nitrogen atoms selectively blocked by methyl or carboxymethyl groups. From these data he estimates the microscopic p*K* values to be 6.92 and 6.53 for the $N^{\delta}H$ and $N^{\epsilon}H$ tautomers, respectively, for N^{α}-acetylhistidine methyl amide. Since Tanokura finds that these values give macroscopic pK_a values that are slightly less than the observed values, we have rounded them up to 7.0 and 6.6, respectively.

EXPERIMENTAL RESULTS

Both 1D proton NMR spectra and homonuclear 2D double quantum spectra were recorded for MbCO at every pH point. The increased resolution in the 2D spectra allowed the pH dependence of both the $C^{\delta}H$ and $C^{\epsilon}H$ chemical shifts to be followed unambiguously over the entire pH range. A region of a 2D double quantum spectrum of MbCO containing the $C^{\delta}H$–$C^{\epsilon}H$ connectivities for some of the histidine residues is shown in Figure 1. Sequence-specific assignments of the histidine resonances have been made using homonuclear 2D methods and heteronuclear 2D and 3D experiments with

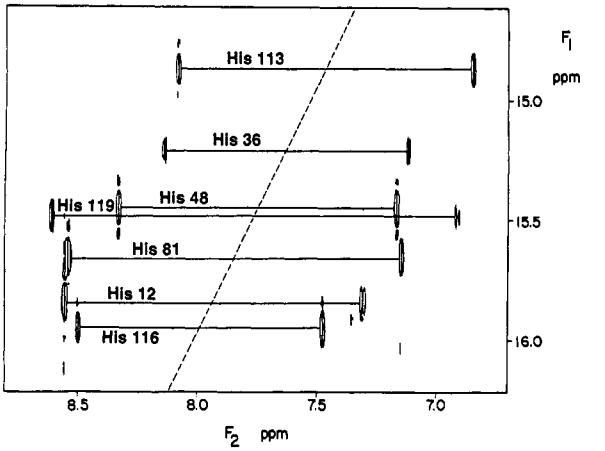

FIGURE 1: Region of homonuclear phase-sensitive double quantum spectrum of MbCO in D_2O at 35 °C, pH 5.56, and 50 mM potassium phosphate buffer, obtained with a multiple quantum excitation period of 100 ms. The direct connectivities between the $C^{\epsilon}H$ and $C^{\delta}H$ resonances of some of the histidines are shown. The broken line indicates the double quantum pseudodiagonal.

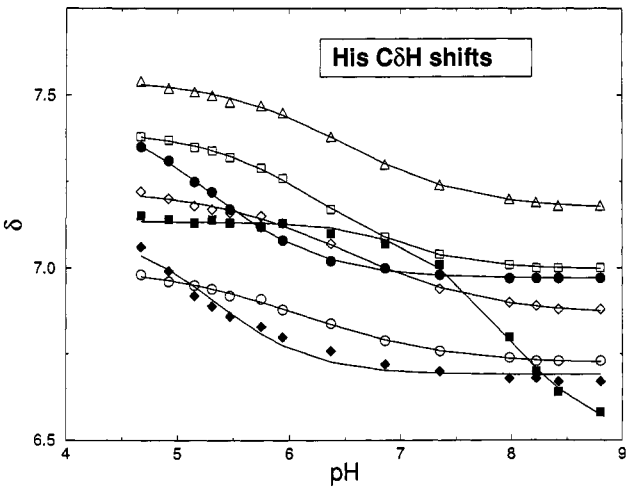

FIGURE 2: Experimental chemical shifts (symbols) and fitted data (curves) for the H^{δ} protons of histidine side chains in MbCO. Histidine: 12, □; 36, ■; 48, ●; 81, ◇; 97, ▲; 113, ◆; 116, △; 119, ○. See also Table III.

uniformly ^{15}N-labeled protein (Dalvit & Wright, 1987; Theriault *et al.*, 1993). Experimental data and fitted curves for the H^{δ} and H^{ϵ} shifts are shown in Figures 2 and 3, and a summary of the results is given in Table III. Four histidines (24, 64, 82, and 93) showed no appreciable shift changes in the pH range 5–9 and are presumably protonated only below pH 5. Of the remaining side chains, 48, 97, and 113 are the most acidic, with $pK_{1/2}$ values between 5 and 6; one histidine (36) has a $pK_{1/2}$ value above 7, and the remaining side chains titrate between pH 6 and 7.

Both the histidine assignments and the p*K*'s reported here differ substantially from those determined earlier by Botelho *et al.* (1978) and by Carver and Bradbury (1984) using 1D NMR methods; the differences are substantial for residues 48, 81, and 119. However, the $pK_{1/2}$ values for histidines in MbCO are all within 0.1 pH unit of those determined recently for metaquomyoglobin, also by 2D methods (Cocco *et al.*, 1992). The Hill coefficients determined here, which range from 0.7 to 0.9, are also in good agreement with those reported for metaquomyoglobin, except for residue 48 where we find a value of *ca.* 0.8 for MbCO, compared to unity for the metaquo protein.

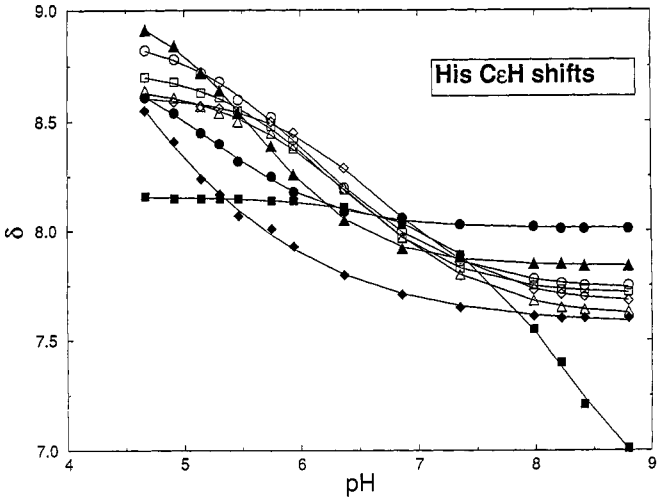

FIGURE 3: Experiment chemical shifts (symbols) and fitted data (curves) for the H^{ϵ} protons of histidine side chains in MbCO. Symbols are the same as for Figure 2.

Table III: Experimental Titration Parameters for MbCO

	$C^{\epsilon}H$				$C^{\delta}H$			
histidine	p$K_{1/2}$	δ^{neut}	$\Delta\delta$	n	p$K_{1/2}$	δ^{neut}	$\Delta\delta$	n
12	6.29	7.71	1.04	0.82	6.26	6.99	0.41	0.84
24	(6.2)[a]	7.75	(0.25)[a]		(<4.5)[a]	6.31		
36	8.06	6.82	1.32	(1)[c]	7.91	6.50	0.63	(1)[c]
48	5.25	8.01	0.70	0.82	5.30	6.97	0.50	0.84
64	<5[b]	7.07			<5[b]	4.88		
81	6.68	7.67	0.95	0.86	6.53	6.87	0.36	0.67
82	<5[b]	7.60			<5[b]	6.70		
93	<5[b]	1.66			<5[b]	1.13		
97	5.63	7.84	1.22	0.90				
113	5.44	7.63	1.01	(1)[c]	5.36	6.69	0.42	(1)[c]
116	6.49	7.60	1.07	0.75	6.50	7.17	0.38	0.75
119	6.13	7.73	1.18	0.74	6.13	6.72	0.28	0.65

[a] Reflects coupled behavior with His119; see text. [b] Upper limit on p$K_{1/2}$ is shown for histidines that do not titrate appreciably. [c] Data was insufficient to fit a value for n, so unity was assumed.

RESULTS OF CALCULATIONS

We begin with an assessment of the results of the calculations, showing the general trends, the complexities arising for particular sites or sets of strongly interacting sites, and the sensitivity of the results to variations in radius and charge parameters and conformation. Tables IV and V give the basic results for p$K_{1/2}$, and the supplementary material provides information on the breakdown of these values into $\Delta\Delta G_{Born}$, $\Delta\Delta G_{back}$, and pK_{intr}, respectively. The pK_{intr} values quoted below in the text are for the tautomer with the lowest pK_{intr} value. This is done because this tautomer is the one that would be preferred if there were no site–site interactions. The statistics cited in subsequent sections are gathered from the results tabulated in Tables IV and V and the supplementary material.

General Features. The $\Delta\Delta G_{Born}$ terms always tend to increase the pK_{intr} values of anionic sites such as carboxylate groups and decrease the pK_{intr} values of cationic groups such as imidazolate and ammonium groups, because the titrating groups are less well solvated in the protein than in the model compound, and this loss of solvation disfavors the charged form. The $\Delta\Delta G_{back}$ terms are usually smaller in magnitude than the $\Delta\Delta G_{Born}$ terms and show no consistent direction. Interactions with the charged forms of other titrating sites (the W_{ij} terms) tend to offset the effects of large $\Delta\Delta G_{Born}$ terms for carboxylic acid groups and lysine and arginine side groups. Large W_{ij} terms are also responsible for several cases of complex multisite titration behavior, as discussed below. This is in contrast to the results of previous calculations on lysozyme (Bashford & Karplus, 1990) in which $\Delta\Delta G_{back}$ terms tended to compensate for the $\Delta\Delta G_{Born}$ terms and interactions between titrating sites were of less importance. Since $\Delta\Delta G_{back}$ is a summation of the effects of many dipolar groups, its sign may differ between calculations, and we see such changes most frequently between the Amber/Bondi parameter set and the others. We also find that the magnitude of $\Delta\Delta G_{back}$ tends to be greater in Amber/Bondi calculations compared to calculations using other parameters. This may be due to a larger peptide dipole in the Amber charge set and the inclusion of all hydrogen atoms, and thus more dipoles, in Amber.

Comparison of the results of calculations using the same charges but different radii shows that the $\Delta\Delta G_{Born}$ terms are sensitive to the radius parameters. This is to be expected since the Born solvation of a simple ion is inversely proportional to its radius (Born, 1920). In Table I we compare the radii used here with those of several other studies. The largest variations are in the radii chosen for polar hydrogens. The CHARMM19 parameters use small, polar hydrogens ($\sigma/2$ = 0.535 Å and R_{min} = 0.6 Å for hydrogen atoms in ammonium groups; $\sigma/2$ = 0.713 Å and R_{min} = 0.8 Å for hydrogen atoms in imidazole groups) as part of a scheme to eliminate the need for specific hydrogen-bonding terms in the empirical energy function (Brooks *et al.*, 1983). The OPLS and Bondi hydrogen radii are substantially larger: 1.15 and 1.2 Å, respectively. These differences have a large effect on the $\Delta\Delta G_{Born}$ terms for lysine and arginine side chains, as discussed below, with larger values appearing to give more realistic results. It should be noted in this connection that Lim *et al.* (1991) found that it was necessary to increase the ammonium hydrogen atom radii from their CHARMM19 values to 1.2 Å to reproduce the correct solvation energy of methylammonium. Other workers have also found it necessary to use a large Born radius for polar hydrogens in order to get good results (Gilson & Honig, 1988; Rashin & Namboodiri, 1987).

Histidines. Histidines 24, 64, and 82 have large $\Delta\Delta G_{Born}$ terms favoring the neutral protonation states because they are not well exposed to solvent. As a result, these side chains should not be protonated without a significant structural change, and this prediction is in agreement with experiment: These three histidines do not titrate above pH 4.7 in the present experiments, and the titration of His64 is thought to be coupled to a local conformational change in the heme pocket of MbCO near pH 4.4 (Morikis *et al.*, 1989). The heme-coordinated proximal His93 has not been included as a titrating site. The His119–His24, His97–PropA, and His36–Glu38 pairs of side chains have complex interactions that are discussed separately below. The remaining histidines, which we shall refer to as the "ordinary" histidines, all follow a common pattern: The Born solvation terms shift pK_{intr} values down by as much as 2.1 pK units. There is no persistent trend in the sign of the background terms, and their magnitudes are significantly less than those of the Born terms. The interactions with the charged forms of other titrating groups are large—cross terms ($W_{ij}/2.303kT$) are often greater than 1 pK unit. Sites whose pK_{intr} is in the neutral to mildly acidic range usually have p$K_{1/2}$ values in the same range, not because cross terms are small but because interactions with negative and positive charges tend to cancel or produce a modest net pK increase.

Table IV: $pK_{1/2}$ Values and Tautomerism by Structure

	structure					
site	COx	COn	oxy	deoxy	MetMut	expmtl[a]
His12[b]	5.57ε(66%)	5.76ε	6.51ε	5.35ε(57%)	5.55ε	6.3
His24[c]						
His36[d]	6.24δ(63%)	6.29ε(52%)	5.58δ(65%)	6.12δ(61%)	7.47ε	8.0
His48[b]	4.79ε	5.03ε	4.83ε	5.00ε	4.96ε	5.3
His64[c]	–2.90ε	–3.19ε		–3.51ε	–2.59ε	
His81[b]	6.87δ(65%)	7.06ε(68%)	6.87δ(49%)	6.22ε(53%)	6.36ε(59%)	6.6
His82[c]	–2.29ε	–1.64ε	–2.29ε	–2.24ε	–0.28ε	
His97[d]	6.73ε	6.62ε	5.73ε	7.15ε	5.37ε	5.6
His113[b]	4.26ε(72%)	5.24ε(68%)	4.35ε(70%)	2.93ε	4.28ε(70%)	5.4
His116[b]	6.22ε(55%)	6.60ε(69%)	5.95ε(53%)	6.17ε(62%)	6.28ε(57%)	6.5
His119[d]	3.34ε	3.58ε	4.36ε	3.09ε	1.98ε	6.1
Tyr103	14.06	13.44	14.55	14.31	14.08	10.3
Tyr146						
Tyr151	11.84	11.82	12.10	11.74		10.5
Asp20	1.48	1.76	0.84	1.24	2.55	
Asp27	1.40	2.97	2.49	3.33	1.08	
Asp44	1.16	2.92	1.62	2.96	2.36	
Asp60	1.90	1.41	1.06	0.91	–0.91	
Asp122	–0.44	–3.16	1.56	1.86		
Asp126	3.63	3.72	4.25	3.52	3.69	
Asp141	5.71	3.20	2.04	1.77	–3.39	
Glu4	1.46	3.52	1.60	–3.46	2.85	
Glu6	1.63	2.24	1.81	3.02	2.68	
Glu18	2.82	3.67	2.67	3.02	3.61	
Glu38	3.54	3.17	4.00	3.22	3.37	
Glu41	3.73	3.83	4.89	3.72	4.28	
Glu52	1.18	3.31	0.19	2.08	2.66	
Glu54	3.93	3.63	3.78	4.54	3.85	
Glu59	4.44	4.36	4.36	4.50	2.82	
Glu83	3.55	4.49	4.42	3.43	3.19	
Glu85	2.73	2.51	3.81	10.03	4.40	
Glu105	0.88	1.58	1.58	1.86	3.79	
Glu109	4.43	3.68	3.51	3.86	1.05	
Glu136	3.54	4.35	4.22	4.25	3.41	
Glu148	3.14	3.63	3.91	3.77	4.12	
Lys16	9.60	14.34	11.01	8.69	10.22	
Lys34	11.81	10.26	10.99	10.61	10.68	
Lys42	9.09	11.56	11.43	9.39	12.26	
Lys47	13.82	11.55	13.40	11.70	11.61	
Lys50	10.74	11.29	10.94	10.69	10.89	
Lys56	11.24	10.68	11.11	10.51	11.47	
Lys62	11.00	11.27	11.56	11.51	10.15	
Lys63	11.00	10.77	10.78	10.98	10.38	
Lys77	11.99	12.52	12.21	12.09	9.87	
Lys78	11.94	11.56	12.65	9.17	12.16	
Lys79	13.66	11.43	13.56		11.90	
Lys87	10.57	10.75	10.80	10.56	10.71	
Lys96	10.82	9.93	10.75	10.88	10.70	
Lys98	10.57	10.73	11.28	10.81	11.35	
Lys102	12.47	13.31	12.82	11.79	11.06	
Lys133	13.24	10.73	13.00	12.25	12.69	
Lys140	11.98	11.02	11.06	11.85	13.24	
Lys145	10.88	11.30	11.07	11.16	11.12	
Lys147	11.04	10.48	10.13	10.94	10.44	
NTVal1	6.32	7.59	7.19	6.97	6.16	
CTGly153	7.33	4.16	4.30	6.00	6.22	
PropA	0.94	2.18	–0.07	1.63	–0.24	
PropD	2.50	1.41	0.96	1.02	1.11	
Arg31	12.53	12.19	12.31	13.37	12.44	
Arg45	14.69	14.62	14.17	14.41	14.13	
Arg118	14.29	12.60		13.30		
Arg139	11.46	13.50	10.79	13.18	11.36	

[a] Experimental values for histidine are from carbon monoxy myoglobin as described in the text. Values for tyrosine are from Wilber and Allerhand (1976). [b] Ordinary histidines. [c] Nontitrating histidines. [d] Histidines involved in coupled pair titrations.

For the "ordinary" histidines, the average across structures of the pK shift due to the $\Delta\Delta G_{Born}$ term ranges from –1.2 (for His48) to –0.5 (His81). The rms deviation of the Born shift across structures for any particular side chain ranges from 0.05 to 0.4 pK unit. Similar figures are found for the variation of $\Delta\Delta G_{Born}$ across parameter sets. Calculations using the C19/σ or C19/Rmin for their radii give $\Delta\Delta G_{Born}$ terms with 30–40% greater magnitude than those using standard van der Waals radii. The average across structures of the pK shift of "ordinary" histidines due to the $\Delta\Delta G_{back}$ terms ranges from –0.5 to 1.4 pK units. Across the various structure/parameter combinations, the pK shift due to site–site interactions (that is, $pK_{1/2} - pK_{intr}$) ranges from –4.9 to 0.8 for the "ordinary" histidines. For individual side chains, the variation caused by parameter/structure variation is typically around 1 pK unit, but can be as much as 5.1 pK units (His113). More variation

Table V: p$K_{1/2}$ Values and Tautomerism by Parameter set

site	parameter set					expmtl[a]
	C19/σ	C19/Rmin	C19/Bondi	OPLS	Amber/Bondi	
His12[b]	6.79ϵ(76%)	6.32ϵ(78%)	5.64ϵ(64%)	5.36ϵ(62%)	5.57ϵ(66%)	6.3
His24[c]						
His36[d]	5.74ϵ(59%)	5.95ϵ(53%)	5.75ϵ(52%)	3.81[e]δ(53%)	6.24δ(63%)	8.0
His48[b]	4.39ϵ	3.83ϵ	4.96ϵ	4.94ϵ	4.79ϵ	5.3
His64[c]	−1.54ϵ(73%)	−3.62ϵ(51%)	−1.50ϵ	−1.45ϵ(71%)	−2.90ϵ	
His81[b]	5.68ϵ(72%)	6.32ϵ(56%)	6.74δ(54%)	6.65ϵ(49%)	6.87δ(65%)	6.6
His82[c]	−3.14ϵ		−1.69ϵ	−2.27ϵ	−2.29ϵ	
His97[d,f]	−0.27ϵ(66%)	−0.07ϵ	2.02ϵ	−2.55ϵ(66%)	6.73ϵ	5.6
His113[b]	3.88ϵ(79%)	3.14ϵ	4.30ϵ(71%)	4.58ϵ(76%)	4.26ϵ(72%)	5.4
His116[b]	5.10ϵ	5.51ϵ(56%)	6.24ϵ(59%)	6.18ϵ(62%)	6.22ϵ(55%)	6.5
His119[d]	3.14ϵ	3.13ϵ	3.45ϵ	2.77ϵ	3.34ϵ	6.1
Tyr103	13.53	14.04	13.63	12.98	14.06	10.3
Tyr146						
Tyr151	11.51	11.56	11.51	11.38	11.84	10.5
Asp20	0.73	−0.53	0.23	0.42	1.48	
Asp27	2.41	2.41	2.44	2.60	1.40	
Asp44	1.77	1.05	1.17	1.51	1.16	
Asp60	2.80	2.20	2.48	3.19	1.90	
Asp122	4.62	1.94	−0.54	1.46	−0.44	
Asp126	3.23	3.06	2.96	3.33	3.63	
Asp141	5.60	4.98	5.07	6.36	5.71	
Glu4	2.08	1.19	1.52	2.01	1.46	
Glu6	0.64	−1.19	−0.07	0.07	1.63	
Glu18	3.57	2.81	2.94	3.67	2.82	
Glu38	3.98	3.39	3.97	5.97	3.54	
Glu41	4.30	4.23	4.03	4.17	3.73	
Glu52	1.66	0.40	1.16	1.54	1.18	
Glu54	3.09	2.69	2.92	2.58	3.93	
Glu59	4.52	4.66	4.50	4.61	4.44	
Glu83	3.37	3.22	3.27	3.69	3.55	
Glu85	3.49	2.77	3.13	3.72	2.73	
Glu105	1.45	0.69	1.22	1.62	0.88	
Glu109	4.71	4.62	4.01	4.56	4.43	
Glu136	3.35	3.12	3.30	3.73	3.54	
Glu148	3.33	3.11	3.32	3.37	3.14	
Lys16	−2.39	4.35	10.31	8.93	9.60	
Lys34	9.16	10.27	11.53	11.63	11.81	
Lys42	6.92	7.09	9.66	9.27	9.09	
Lys47	11.15	12.23	13.27	13.67	13.82	
Lys50	8.96	9.03	10.92	10.93	10.74	
Lys56	9.79	8.99	11.14	11.13	11.24	
Lys62	11.58	11.27	10.85	10.95	11.00	
Lys63	11.41	11.10	10.85	10.91	11.00	
Lys77	9.66	10.88	11.88	12.25	11.99	
Lys78	11.52	11.78	12.10	11.98	11.94	
Lys79	11.14	12.40	13.53	13.93	13.66	
Lys87	11.15	11.02	10.63	10.66	10.57	
Lys96	10.88	10.79	10.79	10.73	10.82	
Lys98	11.26	11.16	10.53	10.68	10.57	
Lys102	9.98	11.88	12.33	12.32	12.47	
Lys133	9.46	11.13	12.72	13.13	13.24	
Lys140	10.42	10.50	11.28	11.69	11.98	
Lys145	8.59	8.91	10.76	10.66	10.88	
Lys147	11.21	11.22	11.02	11.06	11.04	
NTVal1	6.84	6.46	6.67	6.50	6.32	
CTGly153	7.44	7.12	6.25	7.96	7.33	
PropA	5.58	4.98	3.71	3.24	0.94	
PropD	3.42	3.06	2.67	3.03	2.50	
Arg31	13.10	13.10	12.84	12.66	12.53	
Arg45	10.54	12.69	14.28	13.50	14.69	
Arg118	8.95	12.61	14.84		14.29	
Arg139	6.87	9.40	12.04	11.07	11.46	

[a] Experimental values for histidine are from carbon monoxy myoglobin as described in the text. Values for tyrosine are from Wilber and Allerhand (1976). [b] Ordinary histidines. [c] Nontitrating histidines. [d] Histidines involved in coupled pair titrations. [e] The low p$K_{1/2}$ value for His36 in the OPLS calculation is caused by proton sharing with Glu38 (see text). [f] The low p$K_{1/2}$ values for His97 reflect the first (lowest pH) titration of the His97–PropA pair. A second titration of a proton shared between His97 and PropA occurs at higher pH (see text).

in p$K_{1/2}$ − pK_{intr} is caused by structure variation than by parameter variation. In many cases, the pK shift due to site–site interactions is negative, reinforcing rather than compensating for the negative shifts due to $\Delta\Delta G_{Born}$.

Comparison of calculated p$K_{1/2}$ values to experimental values is given in Tables IV and V. The correct prediction that histidines 24, 64, and 82 do not titrate has been noted above. For the "ordinary" histidines, the ordering of p$K_{1/2}$ values is in agreement with experiment for most structure/parameter combinations. In all cases, the calculations correctly distinguish 48 and 113 as more acidic and 12, 116, and 81 as less acidic, although in some cases the ordering

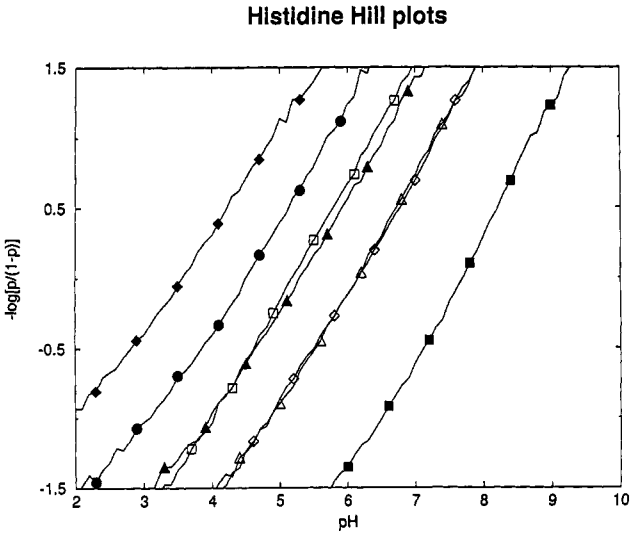

FIGURE 4: Hill plots for the Amber/Bondi/MetMut calculation. Symbols for histidines are the same as in Figure 2. Mean slopes of the curves averaged over 1 pH unit near the midpoints are the following: His12, 0.86; His36, 0.94; His48, 0.86; His81, 0.76; His97, 0.81; His113, 0.78; His116, 0.84; His119, 0.64.

within this latter group is not in agreement with experiment. Calculated $pK_{1/2}$ values for these five "ordinary" histidines are generally lower than the experimental values. The disagreement is smaller for histidines 12, 81, and 116, which have experimental pK_a values at or slightly below the pK_{model} values for histidine, and the disagreement is larger for histidines 113 and 48, which have lower experimental pK_a values. This suggests a tendency of the calculations to exaggerate the downward pK shifts of the "ordinary" histidines. The standard deviation across different structures for calculated $pK_{1/2}$ values for particular sites is less than 0.4 for all "ordinary" histidines except 113, for which it is 0.7. The standard deviation across parameter sets is less than 0.6 for all "ordinary" histidines. For the histidines participating in titrating pairs, 36, 97, and 119, the calculated $pK_{1/2}$ values are more varied and are less in agreement with experiment than those for the "ordinary" histidines. These residues will be discussed individually below.

The calculated Hill plots for the Amber/Bondi/MetMut calculation are shown in Figure 4. These are nearly linear over the range $-1.5 < \log[p/(1-p)] < 1.5$, in accord with the observation that the experimental curves are well fit by a single protonation event with an adjustable Hill coefficient. Those histidines at the left of the figure, which titrate at lower pH, tend to have smaller Hill coefficients (given in the caption to Figure 4). These more acidic histidines interact more with partially protonated Asp and Glu residues than do the more neutral histidines, and these site–site interactions are anti-cooperative, *i.e.*, protonation of one side chain will make it more difficult to protonate a nearby residue. In general, the computed and observed Hill coefficients are in good agreement, with most values ranging from 0.7 to 0.9 in both cases.

Tyrosine Residues. The titration behavior of tyrosines in myoglobin has been studied by ^{13}C NMR with CO and CN^- as ligands (Wilbur & Allerhand, 1976) and by resonance Raman spectroscopy for the metaquo form (Asher *et al.*, 1991). The metaquo titration results include effects arising from the (pH-dependent) replacement of water by hydroxide at the iron (and hence are less simply related to the present calculations). The NMR results for MbCO are listed in the experimental column of Tables IV and V, but all of the experimental results agree that Tyr146 has an abnormally high $pK_{1/2}$, probably around 13.

The pK_{intr} of Tyr146 is raised substantially by the loss of Born solvation due to the burial of the side chain, and this residue is predicted not to titrate in the normal range, in agreement with experiment. The pK shifts of tyrosines 103 and 151 due to Born terms range from 0.7 to 3.7, with Amber/Bondi parameters generally giving smaller shifts than other parameters and the largest structural deviation occurring for MetMut Tyr151. The pK shift due to background terms for these residues ranges from –0.6 to 1.9, with Amber/Bondi parameters generally giving more positive values. The resulting pK_{intr} values range from 11.1 to 15.2. At the high pH values corresponding to these pK_{intr} values, most of the lysine and arginine residues have lost their positive charges so that the negative charges of the carboxylate groups are uncompensated. The result is that the $pK_{1/2}$ values are 1.5 or more units higher than the pK_{intr} values for Tyr103 and 0.3 or more unit higher for Tyr151. This leads to predicted $pK_{1/2}$ values that are substantially above the values of 10.3–10.5 seen experimentally. It would seem that either the calculations are overestimating the penalty paid to deprotonate these partially buried side chains or substantial conformational changes occur upon titration. Compared to other force fields, the charge model in the Amber force field leads to larger (unfavorable) interactions between the deprotonated tyrosines and nearby groups, and hence to higher $pK_{1/2}$ values, but all of the results tend to be higher than those seen experimentally. As with histidines, this may reflect a general tendency toward overestimates of the protein-induced shifts from model compound results.

Carboxylic Acid Groups. The pK shift of carboxylic acid groups due to $\Delta\Delta G_{Born}$ terms is always positive, and the average shift over the 24 carboxylic acid residues in myoglobin is 1.2 in the COx/Amber/Bondi calculations and ranges from 1.7 to 2.2 in calculations using other parameter sets. For different structures, the average pK shift due to $\Delta\Delta G_{Born}$ ranges from 1.0 to 1.2. Comparing $\Delta\Delta G_{Born}$ shifts of individual groups across parameter sets and structures, we find that parameter and structure variation each cause variability on the order of 0.5 pK unit. The pK shifts due to background terms are generally about one-half of the magnitudes of the Born terms and are variable in sign. There are strong, pK-lowering interactions with positively charged titrating sites, and these are somewhat larger with the Amber/Bondi parameters than with the other parameters. Roughly one-half of the calculated $pK_{1/2}$ values are below 3.0, and some are less than 1.0. Only a handful are greater than 4.5. Heme propionic acid A, Glu38, and Asp122 are involved in complex pair titrations and will be discussed below. No experimental data are available on the titration of specific carboxylic acid groups in myoglobin. Carbon monoxy myoglobin undergoes a global structural transition with a pK_a of 3.9 (Sage *et al.*, 1991), so that calculated $pK_{1/2}$ values below or near 3.9 should not be considered realistic.

Lysine and Arginine Residues. For the 19 lysine residues in myoglobin, the average pK shift due to Born terms in calculations *other than* those using the C19/σ or C19/Rmin parameter sets ranges from –1.2 to –0.9, while in the C19/σ and C19/Rmin calculations the average lysine Born shifts are –2.5 and –2.4, respectively. This striking difference suggests that the unusually small hydrogen radii in these parameters sets could lead to overestimates of Born terms. The lysine pK shifts due to background terms tend to have magnitudes on the order of 0.3 and variable signs. Site–site interactions increase the lysine $pK_{1/2}$ values in most cases. The magnitude of the increase is typically on the order of 1

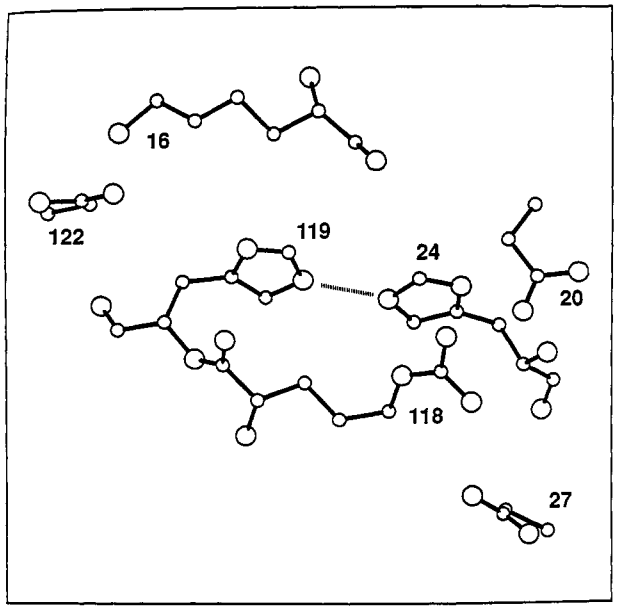

FIGURE 5: His24–His119 pair and nearby residues. Side chains of Lys16, Asp20, Asp27, Arg118, and Asp122 are shown. Coordinates are for the COx structure.

or 2 pK units, but there is considerable variation by site, structure, and parameters used. Most of the resulting lysine p$K_{1/2}$ values are in the 10–13 range. Lys16 participates in a complex coupled pair titration with Asp122, leading to a wide range of p$K_{1/2}$ values. The general pattern of results for the four arginine residues in myoglobin is similar to that for the lysines, including the observation that the parameter sets with small hydrogen radii produce Born shifts that are roughly twice as large as those produced by other parameter sets.

Coupled Pairs. Strongly coupled protonation sites are likely to exist in many proteins and often have functional roles, as in many proteases. It is of interest to study the behavior of such proton sharing within a model that goes beyond the mean field approximation, since important features of the protonation behavior are not captured in the simpler theories. For this reason, we examine here in some detail three examples from myoglobin that illustrate some of the varieties of behavior that can exist.

(*1*) *The His24–His119 Pair.* In the crystal structure the imidazole groups of His24 and His119 are hydrogen bonded—the ϵ nitrogens of each imidazole are 2.71 Å apart, apparently sharing a single proton between them (Figure 5). The simple electrostatic model used here is probably inadequate for quantitative predictions of the energetics of this system, but one can hope for qualitative understanding. We denote the protonation state of the His24–His119 pair by pairs of letters representing the states of His24 and His119, respectively. The letters ϵ, δ, and p represent the neutral tautomer with the proton bound to N^{ϵ}, the tautomer with the proton bound to N^{δ}, and the +1 charged protonated state, respectively. Of the nine possible states, $\epsilon\epsilon$, ϵp, pϵ, and pp are expected to be forbidden without significant conformational change, since protonating the ϵ positions of both imidazoles would involve a near-superposition of hydrogen atoms. Of the three remaining neutral tautomers, one would expect $\delta\delta$ to have a higher free energy since the negatively charged, deprotonated N^{ϵ} atoms of the pair are in close proximity. The Born terms have relatively little influence on tautomer preferences, but the background terms make $\delta\epsilon$ the most strongly favored neutral state by at least 1.5 pH units. The state with the next lowest free energy is either $\delta\delta$ or $\epsilon\delta$, depending on the structure. Of the two possible +1 charged states (δp and pδ), δp is preferred by between 0.6 and 2.0 pH units (depending on the parameter set), mainly because His24 is more deeply buried than His119 and therefore is more difficult to protonate.

The titration of the pair from its preferred neutral state, $\delta\epsilon$, to its preferred +1 charged state, δp, has a pK_{intr} value ranging from 4.0 to 4.8, depending on the parameter set used. As Figure 5 shows, the pair has several charged near neighbors which exert strong influence. For example, in the COx/Amber/Bondi calculation, the strongest influences are from the positively charged side-chains Lys16 and Arg118, which lower the pK by 1.6 and 0.8, respectively, and Asp122, Asp20, and Asp27, which raise the pK by 1.1, 0.3, and 0.2, respectively. The net result of these and other smaller interactions is that the pair's pK_{intr} value of 4.8 is shifted down to a p$K_{1/2}$ value of 3.3. The calculations with other structure/parameter combinations follow a similar pattern, and the resulting p$K_{1/2}$ values range from 2.0 to 4.4 for all structure/parameter combinations. All of the predicted His119 p$K_{1/2}$ values are much lower than the experimental value (6.2). Most likely, this reflects the difficulties the theory has treating this "shared proton" using only model compound information from isolated histidines and a classical electrostatic model.

(*2*) *The His97–PropA Pair.* At first sight, His97 seems to be an extraordinarily sensitive site, since Tables IV and V show that the p$K_{1/2}$ values vary from –2.6 to 7.2, depending on the parameters and structures. In fact, this is another example of potential proton sharing between two residues, in this case between His97 and one of the heme propionate groups. For the Amber/Bondi calculations on all structures, the His97 pK shift due to $\Delta\Delta G_{Born}$ ranges from –3.5 to –2.7 pK units and the shift due to $\Delta\Delta G_{back}$ ranges from 0.8 to 1.8. The resulting pK_{intr} values for His97 range from 4.0 to 5.3 and favor the ϵ tautomer. Interactions with negative charges, principally the deprotonated heme propionate A (PropA), raise the p$K_{1/2}$ values to the 5.4–7.1 range. The calculated titration curves have the normal sigmoidal form, and the δ tautomer never has a significant population. In the same calculations, the PropA pK shift due to $\Delta\Delta G_{Born}$ ranges from 1.7 to 1.9, and the shift due to $\Delta\Delta G_{back}$ ranges from 0.0 to 1.3 pK units. Strong interactions with positive charges, principally His97, then lower the p$K_{1/2}$ values to the range –0.2 to 2.2.

On the other hand, for the non-Amber calculations (all on the COx structure), the His97 pK shift due to $\Delta\Delta G_{back}$ ranges from –4.0 (OPLS) to –1.3 pK units (C19/Bondi), which is *opposite in sign* to the Amber/Bondi results. The $\Delta\Delta G_{Born}$ shifts, ranging from –3.7 to –2.5, are similar to the Amber/Bondi results. The resulting pK_{intr} values range from 0.2 to 3.1, favoring the δ tautomer—the *opposite tautomer* to the Amber/Bondi results. The pK shift for the propionate due to $\Delta\Delta G_{Born}$ ranges from 1.7 to 2.1, and the shift due to $\Delta\Delta G_{back}$ ranges from –0.1 to 1.3. The resulting PropA pK_{intr} values range from 6.2 to 7.8. The strong interaction between His97 and PropA gives rise to complex titration behavior in the non-Amber/Bondi titration curve calculations. As in the Amber/Bondi calculations, there are two titrations, one in the pH range from –2.5 to 1.0 and a second one at a higher pH. But the first titration is not a simple titration of PropA; it is the loss of a proton either entirely from His97 (C19/σ and OPLS) or partly from His97 and partly from PropA (C19/Bondi and C19/Rmin). In the intermediate pH range the δ tautomer of His97 is preferred. The second titration is either PropA (C19/σ and OPLS) or the shared proton (C19/Bondi and C19/Rmin). The second titration also induces a change in

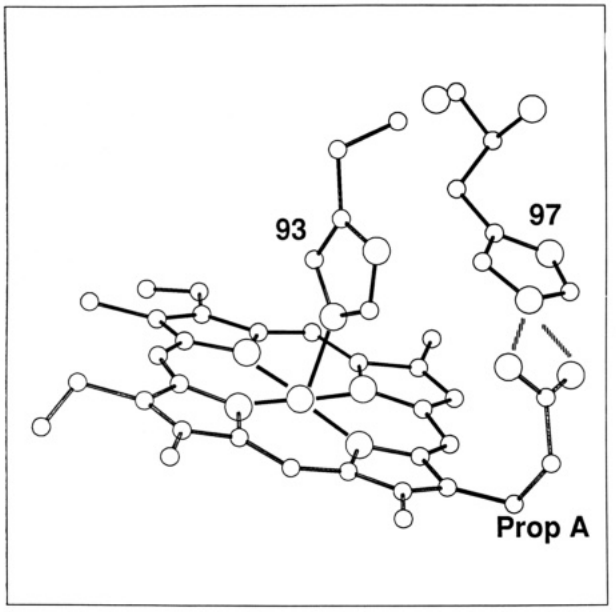

FIGURE 6: His97–PropA pair and the heme group. Coordinates are from the COx structure.

the His97 tautomer from δ to ϵ. For the C19/OPLS parameters, the second titration occurs around pH 3.0, but for C19/σ, C19/Rmin, and C19/Bondi, the second titration occurs in the pH range 5.3–5.5. Thus the Amber/Bondi and non-Amber calculations are not as different as a simple inspection of $pK_{1/2}$ values would suggest. With the exception of the C19/OPLS results, they have in common the high-pH tautomeric state of His97 and the rough positions of two titration points, particularly the second titration in the pH range 5.3–7.2, but they differ in the identity of the groups titrating at these points. The difficulties of interpreting $pK_{1/2}$ values for such two-step titrations have been discussed in greater detail for the case of bacteiorhodopsin (Bashford & Gerwert, 1992).

The present results for the His97–PropA pair in the pH range between the two titrations can be understood in terms of the relative stability of two possible one-proton states of the pair: *up*, the neutral state in which His97 is unprotonated and PropA is protonated, and *pu*, the ion-pair state. (We are neglecting tautomerism to simplify the discussion.) If only pK_{intr} terms are considered, the neutral *up* state is favored since the pK_{intr} of His97 is always lower than that of PropA for *all* structure/parameter combinations, largely because of the $\Delta\Delta G_{Born}$ terms (desolvation). However, the strong interaction between the charged forms will tend to favor the ion-pair state, *pu*, when the site–site interaction is included. If $W_{ij}/2.303kT$ is greater than the difference in pK_{intr}, *pu* will be more stable, neglecting interactions with other titrating sites. In the Amber/Bondi calculations, the pK_{intr} differences are always smaller than the ion interaction terms by 1 pK unit or more, so that the *pu* state is favored and the *up* state never appears. In the non-Amber calculations, the pK_{intr} differences are substantially larger than those in the Amber/Bondi calculations because of the opposite sign of $\Delta\Delta G_{back}$. These pK_{intr} differences are similar in magnitude to the ion interaction terms, so that the result is a mixture of the *up* and *pu* states.

That the distinction between these states should be sensitive to parameters is not surprising after examination of the structure (Figure 6). The N^{ϵ} atom of His97 and one of the oxygen atoms of PropA are close enough to share a proton in a hydrogen-bonding interaction. The distinction between *up* and *pu* is then a question of which group is the "donor" and which is the "acceptor". For purposes of comparison with experiment, the calculations for the full range of structures and parameters (except OPLS) should be interpreted as predicting that a titration occurs with a pK_a between 5.3 and 7.2 in which the His97–PropA pair gives up a proton and goes to the *uu* state with His97 in the ϵ tautomer. One would see a chemical shift change at His97 whether the proton that titrates "belongs" to the histidine or the propionate. However, the Amber/Bondi calculations (which generally seem reliable) in all cases predict that it is the histidine that titrates. This is consistent with the experimental pK_a of 5.6 for His97, although the range of predictions is very wide.

(*3*) *The His36–Glu38 Pair*. The pK shift of His36 due to $\Delta\Delta G_{Born}$ ranges from –2.4 to –1.2, and the shift due to $\Delta\Delta G_{back}$ ranges from –0.3 to 0.7 pK unit. The resulting pK_{intr} values range from 4.7 to 5.3 across all structure/parameter combinations and favor the ϵ tautomer, although often by only a few tenths of a pK unit. The pK shift of Glu38 due to $\Delta\Delta G_{Born}$ ranges from 1.2 to 2.7, while the shift due to $\Delta\Delta G_{back}$ ranges from –1.4 to –0.2 pK unit. The resulting pK_{intr} values range from 5.1 to 6.1. Variation due to structural differences is greater than variation due to parameter differences in the results for both groups.

In all of the structures, His36 and Glu38 are in close proximity, although they do not form a hydrogen bond. This results in a strong electrostatic interaction that usually tends to raise the $pK_{1/2}$ of His36 and lower the $pK_{1/2}$ of Glu38. But with the two residues having pK_{intr} values in similar ranges, proton sharing is possible. This occurs in the pH 3–5 range in the calculations for most structure/parameter combinations, but the proton stays mainly on the His36, so that the final $pK_{1/2}$ values reflect the second step of the two-step titration process. Exceptions are the COx/OPLS calculation where there is also proton sharing in this range, but the proton stays mostly on Glu38 so that His36 and Glu38 have $pK_{1/2}$ values of 3.8 and 6.0, respectively—a reversal of the usual titration order and the MetMut structure, where the titrations are not coupled. In the MetMut structure, the interactions between His36 and positively charged lysines, arginines, and histidines are weaker than those in other structures, resulting in a high $pK_{1/2}$ value, 7.5, and an ϵ tautomer. This gives better agreement with experiment (8.0) than any of the other calculated $pK_{1/2}$ values.

DISCUSSION

The results reported here add to our knowledge of the strengths and weaknesses of a more elaborate model that incorporates many of the same continuum dielectric notions as simpler models [*e.g.*, Tanford and Kirkwood (1957)], but makes fewer numerical approximations. We have seen that the basic titration results are reasonably insensitive to the parameters or structures assumed, except where very small hydrogen radii are used or where strong site–site interactions make the net protonation behavior a sensitive function of the characteristics of the individual sites. The general classification of histidine side chains into nontitrating ($pK_{1/2} < 5$), acidic ($pK_{1/2}$ between 5 and 6), normal (between 6 and 7), and slightly basic (above 7) is generally well reproduced in the calculations, especially for the P6 crystal form. Computed Hill coefficients are also in good general agreement with those observed. Variations in the structures or parameters used produce variations in the calculated $pK_{1/2}$ values on the order of 1 pK unit for "ordinary" histidines. The larger variations that occur in some other histidines are understandable in terms of smaller energetic variations affecting proton sharing in

coupled pairs of titrating sites. It is not known whether such proton sharing and two-step titration actually occurs in myoglobin, but if it does it could play a significant role in acid-induced conformational changes (see below). It should be noted that predictions of such effects are impossible in simpler models of protein titration (Tanford & Kirkwood, 1957; Tanford & Roxby, 1972). The predicted p$K_{1/2}$ values for tyrosines are significantly higher than those observed, which suggests that the current model may be overestimating the energy penalty to partially bury a deprotonated tyrosine side chain or may be ignoring what are energetically important structural changes that accompany the ionization of these side chains.

Sites that have different pK_a values in different conformational forms of the protein can lead to pH-dependent conformational changes (Tanford, 1970). For example, the acid-induced denaturation of myoglobin could be explained if the histidines in myoglobin that have extremely low pK_a values in the native structure (due to large $\Delta\Delta G_{Born}$ terms) become exposed in an unfolded form. Indeed, experimental studies of the acid denaturation of metMb have led to the hypothesis that six histidines that are inaccessible to titration in the native state are exposed upon denaturation (Breslow & Gurd, 1962; Puett, 1973). More recent work has shown that, at pH 4.4, MbCO undergoes a local conformational change in the heme distal region accompanied by the titration of His64 (Morikis *et al.*, 1989). At a slightly lower pH (3.9), MbCO undergoes a global conformational transition to a partially unfolded form, and this transition is coupled to the uptake of protons at approximately six sites; in the case of metMb and deoxyMb the transition occurs at pH 4.3 and also involves six titrating sites (Sage *et al.*, 1991). The proximal histidine–heme bond remains intact after the above transition in MbCO, but it is broken in the case of metMb and deoxyMb. Considering the present results, there do not seem to be enough histidines to account for the conformational change and proton uptake by approximately six sites of MbCO at pH 3.9: all histidines in MbCO except for 24, 64, 82, and 93 titrate well above the transition point according to the experimental results presented here. His93, the proximal histidine, is not available for the transition since it remains ligated to the heme. His64 titrates at 4.4 as noted above and so is not available. This leaves only histidines 24 and 82. On the other hand, the calculations presented here suggest that histidines 36, 97, and 119 titrate in two steps, so that the titrations observed experimentally may only be partial and may be followed by further proton uptake at a lower pH, possibly in connection with conformational changes. The present calculations also show a number of carboxylic acid groups with unusually low p$K_{1/2}$ values in the native form. This suggests that these sites, as well as histidines, could be involved in the conformational transition. Calculations pursuant to this suggestion at a level similar to the present calculations would require an atomic model, or set of models, for the partially unfolded, low-pH forms of myoglobin. The generation of such models is a topic of current research (C. Brooks, personal communication).

There are several possible sources of error in these calculations. We do not allow for conformational variation since we model titration energetics by the electrostatic work done in adding or removing charges from a rigid dielectric system. If the flexibility of the real system allows a large range of electrostatic environments to a titrating site, and particularly if conformational changes are coupled to titration, our model may lead to significant errors although, as noted above, such errors may be diagnostic for pH-dependent conformational effects. The crystallographically determined conformation will tend to be the one most favorable to the protonation state prevailing under the pH and other solution conditions under which the structure is determined. Subsequent calculations using this structure may then be biased toward that protonation state. In the present case, the substantial conformational changes known to occur in myoglobin at low pH (see above) mean that our calculations on carboxylic acid groups may be biased toward the deprotonated state, that is, the p$K_{1/2}$ values may be too low. Indeed, we do find unusually low values for these residues. Another source of error that has been explored here is the choice of charge and radius parameters. In particular, we find that results can be sensitive to the choice of hydrogen radii and that the small radii sometimes used in molecular mechanics empirical energy functions to reproduce hydrogen bonding may not be suitable for the present type of electrostatic calculations. This may be indicative of a more general problem: that errors arise when parameters optimized for molecular mechanics with no polarizability, a dielectric constant of 1.0, and radii not intended for use in a Born-like formula are used in MEAD-type calculations such as those presented here. We have modeled the dielectric response of the protein interior by a single constant, 4.0, which is obviously a great simplifcation of effects arising from small conformational changes and electronic polarizability of chemical groups ranging from aliphatic to aromatic. Recently, Sharp *et al.* (1992) have proposed a more complex model of solute dielectric effects. Finally, there may be errors due to the representation of water as a continuum, which would neglect, for example, a protein-bound water molecule acting as a hydrogen-bonding bridge between two titratable groups. In spite of these uncertainties, the results suggest that many useful features of electrostatic behavior in proteins can be described by the present level of theory.

Computer Programs and Data. Some of the computer programs and data files used for the electrostatic calculations can be obtained electronically *via* the InterNet by opening an FTP connection to the host, ftp.scripps.edu, using the login name, "anonymous", and obtaining files from the directory, "pub/electrostatics".

ACKNOWLEDGMENT

We thank Pat Jennings and Dimitrios Morikis for useful discussions. We thank P. Beroza, D. Fredkin, M. Y. Okamura, and G. Feher for supplying us with a copy of their Monte Carlo titration program.

SUPPLEMENTARY MATERIAL AVAILABLE

Tables listing Born energy contributions, background energy contributions, and intrinsic pK values (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Asher, S. A., Larkin, P. J., & Teraoka, J. (1991) *Biochemistry 30*, 5944–5954.

Bashford, D., & Karplus, M. (1990) *Biochemistry 29*, 10219–10225.

Bashford, D., & Karplus, M. (1991) *J. Phys. Chem. 95*, 9556–9561.

Bashford, D., & Gerwert, K. (1992) *J. Mol. Biol. 224*, 473–486.

Beroza, P., Fredkin, D. R., Okamura, M. Y., & Feher, G. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5804–5808.

Bondi, A. (1964) *J. Chem. Phys. 64*, 441.

Born, M. (1920) *Z. Phys. 1*, 45–48.

Botelho, L. H., & Gurd, F. R. N. (1978) *Biochemistry 17*, 5188–5196.

Botelho, L. H., Friend, S. H., Matthew, J. B., Lehman, L. D., Hanania, G. I. H., & Gurd, F. R. N. (1978) *Biochemistry 17*, 5197–5205.
Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.
Breslow, E., & Grud, F. R. N. (1962) *J. Biol. Chem. 237*, 371–381.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.
Carver, J. A., & Bradbury, J. H. (1984) *Biochemistry 23*, 4890–4905.
Cheng, X., & Schoenborn, B. P. (1990) *Acta Crystallogr., Sect. B. 46*, 195.
Cheng, X., & Schoenborn, B. P. (1991) *J. Mol. Biol. 220*, 381–399.
Cocco, M. J., Kao, Y.-H., Phillips, A. T., & Lecomte, J. T. J. (1992) *Biochemistry 311*, 6481–6491.
Dalvit, C., & Wright, P. E. (1987) *J. Mol. Biol. 194*, 313–327.
Friend, S. H., & Gurd, F. R. N. (1979a) *Biochemistry 18*, 4612–4619.
Friend, S. H., & Gurd, F. R. N. (1979b) *Biochemistry 18*, 4620–4630.
Gilson, M. K., & Honig, B. (1988) *Proteins: Struct., Funct., Genet. 4*, 7–18.
Honig, B., Hubbell, W., & Flewelling, R. (1986) *Annu. Rev. Biophys. Biophys. Chem. 15*, 163–193.
Jean-Charles, A., Nicholls, A., Sharp, K., Honig, B., Tempczyk, A., Henderson, T. F., & Still, W. C. (1991) *J. Am. Chem. Soc. 113*, 1454–1455.
Jorgensen, W. L., & Tirado-Rives, J. (1988) *J. Am. Chem. Soc. 110*, 1657–1671.
Kurimatsu, S., & Hamaguchi, K. (1980) *J. Biochem. 87*, 1215–1219.
Kuriyan, J., Wilz, S., Karplus, M., & Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133–154.
Lim, C., Bashford, D., & Karplus, M. (1991) *J. Phys. Chem. 95*, 5610–5620.
Mabbutt, B., & Wright, P. E. (1985) *Biochim. Biophys. Acta 832*, 175–185.
Matthew, J. B., Gurd, F. R. N., Garcia-Moreno E., B., Flanagan, M. A., March, K. L., & Shire, S. J. (1985) *CRC Crit. Rev. Biochem. 18*, 91–197.
Mohan, V., Davis, M. E., McCammon, J. A., & Pettitt, B. M. (1992) *J. Phys. Chem. 96*, 6428–6431.
Morikis, D., Champion, P. M., Springer, B. A., & Sligar, S. G. (1989) *Biochemistry 28*, 4791–4800.
Nozaki, Y., & Tanford, C. (1967) *Methods Enzymol. 11*, 715–734.
Perutz, M. F. (1978) *Science 201*, 1187–1191.
Phillips, G. N., Jr. (1990) *Biophys. J. 57*, 381–383.
Phillips, S. E. V. (1980) *J. Mol. Biol. 142*, 531.
Phillips, G. N., Jr., Arduini, R. M., Springer, B. A., & Sligar, S. G. (1990) *Proteins 7*, 358.
Puett, D. (1973) *J. Biol. Chem. 248*, 4623–4634.
Rashin, A. A., & Namboodiri, K. (1987) *J. Phys. Chem. 91*, 6003–6012.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151–176.
Sage, J. T., Morikis, D., & Champion, P. M. (1991) *Biochemistry 30*, 1227–1237.
Sharp, K., Jean-Charles, A., & Honig, B. (1992) *J. Phys. Chem. 96*, 3822–3828.
Shire, S. J., Hanania, G. J. H., & Gurd, F. R. N. (1974) *Biochemistry 13*, 2967–2974.
Soman, K., Yang, A.-S., Honig, B., & Fletterick, R. (1989) *Biochemistry 28*, 9918–9926.
Takano, T. (1977) *J. Mol. Biol. 110*, 569–584.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1–95.
Tanford, C., & Kirkwood, J. G. (1957) *J. Am. Chem. Soc. 79*, 5333–5339.
Tanford, C., & Roxby, R. (1972) *Biochemistry 11*, 2192–2198.
Tanokura, M. (1983) *Biochim. Biophys. Acta 742*, 576–585.
Theriault, Y., Pochapsky, T. C., Chiu, M., Sligar, S. G., & Wright, P. E. (1993) *J. Biomol. NMR* (submitted for publication).
Warshel, A., & Russell, S. T. (1984) *Q. Rev. Biophys. 17*, 283–422.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230–252.
Wilbur, D. J., & Allerhand, A. (1976) *J. Biol. Chem. 251*, 5187–5194.