A Self-Consistent, Microenvironment Modulated Screened Coulomb Potential Approximation to Calculate pH-Dependent Electrostatic Effects in Proteins

Ernest L. Mehler and Frank Guarnieri
Department of Physiology and Biophysics, Mount Sinai School of Medicine, CUNY, New York, New York 10029

ABSTRACT An improved approach is presented for calculating pH-dependent electrostatic effects in proteins using sigmoidally screened Coulomb potentials (SCP). It is hypothesized that a key determinant of seemingly aberrant behavior in pK_a shifts is due to the properties of the unique microenvironment around each residue. To help demonstrate this proposal, an approach is developed to characterize the microenvironments using the local hydrophobicity/hydrophilicity around each residue of the protein. The quantitative characterization of the microenvironments shows that the protein is a complex mosaic of differing dielectric regions that provides a physical basis for modifying the dielectric screening functions: in more hydrophobic microenvironments the screening decreases whereas the converse applies to more hydrophilic regions. The approach was applied to seven proteins providing more than 100 measured pK_a values and yielded a root mean square deviation of 0.5 between calculated and experimental values. The incorporation of the local hydrophobicity characteristics into the algorithm allowed the resolution of some of the more intractable problems in the calculation of pK_a. Thus, the divergent shifts of the pK_a of Glu-35 and Asp-66 in hen egg white lysozyme, which are both about 90% buried, was correctly predicted. Mechanistically, the divergence occurs because Glu-35 is in a hydrophobic microenvironment, while Asp-66 is in a hydrophilic microenvironment. Furthermore, because the calculation of the microenvironmental effects takes very little CPU time, the computational speed of the SCP formulation is conserved. Finally, results from different crystal structures of a given protein were compared, and it is shown that the reliability of the calculated pK_a values is sufficient to allow identification of conformations that may be more relevant for the solution structure.

INTRODUCTION

The electrostatic interaction between two charges in a dielectric medium decreases rapidly as the separation between the two entities is increased. This is a result of the dielectric medium's intrinsic ability to shield the electrostatic field arising from the charge. There is compelling experimental and theoretical evidence that the general functional form of this screening is sigmoidal. Early experiments on organic acids (Debye and Pauling, 1925; Schwarzenbach, 1936; Webb, 1926) were collected together and presented as mean values in tabular form (Conway et al., 1951) that produced a smooth, sigmoidal screening as a function of separation with an asymptotic value of about 80. Experiments on water as a dielectric medium indicate that the electrostatic field, at loci only two solvation layers ($\leq$6 Å) away from a charge, is dramatically damped, almost approaching the bulk dielectric screening of $\epsilon = 80$ (Harvey and Hoekstra, 1972; Pennock and Schwan, 1969; Takashima and Schwan, 1965). When two charges are in close proximity, however, the dielectric screening rapidly diminishes approaching the vacuum value because there is no intervening dielectric. These conditions are simultaneously satisfied with a radially dependent sigmoidal screening function. Additionally, it has been demonstrated that the generalized Born-surface area continuum solvation model also gives rise to sigmoidal screening (Mehler, 1996b). Thus, the general sigmoidal screening form applies to both high and low dielectric media. Theoretically, the Lorentz–Debye–Sack (LDS) theory of polar solvation also leads to sigmoidally screened electrostatic fields (Debye, 1929; Lorentz, 1952; Sack, 1926; Sack, 1927). LDS theory has been used to develop radial dielectric permittivity profiles to describe ion and dipole field sources (Ehrenson, 1989) and to derive an expression for the hydration energies of spherical ions (Bucher and Porter, 1986) from the integral form of the Born Equation (Born, 1920). Subsequently, Ehrenson (1989) showed that by using Bötcher's analytic expression (Böttcher, 1938) of Onsager's theory (Onsager, 1936), the reaction field corrected LDS theory for both dipolar and asymmetric ion field sources also yielded sigmoidal screening. (For a review of LDS theory and its relation to protein electrostatics, see Mehler, 1996a.) In various applications (Collura et al., 1994; Mehler, 1996a; Mehler and Solmajer, 1991), it has been shown that sigmoidal screening leads to the reliable prediction of properties dependent on electrostatic effects in proteins, and therefore is an attractive formulation for the development of implicit solvent models (Guarnieri et al., 1998; Hassan et al., 1999).

Recently, a self-consistent approach based on using sigmoidally screened Coulomb potentials (SCP) for calculating pH-dependent properties in proteins has been derived and applied to several systems (Mehler, 1996b). The results indicated that the reliability of the method is similar to that

Received for publication 25 June 1998 and in final form 31 March 1999.
Address reprint requests to Dr. Ernest L. Mehler, Department of Physiology and Biophysics, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029. Tel.: 212-241-5852; Fax: 212-860-3369; e-mail: mehler@inka.mssm.edu.
© 1999 by the Biophysical Society
0006-3495/99/07/03/20 $2.00

achieved by the finite difference Poisson–Boltzmann (FDPB), but requires substantially less computing effort. Interestingly, although these various methods predict a majority of the pK_a shifts correctly, they were consistent in the occasional dramatic errors on singularly important residues. The divergent pK_a shifts of the two titratable acids of hen egg white lysozyme, Glu-35 and Asp-66, are particularly problematic. The divergent pK_a shifts cannot be simply explained by the buried fraction (Demchuk and Wade, 1996), since the two titratable groups are nearly completely buried. Thus, it appears more likely that the unique local environment around each residue is a key determinant for the pK_a of Glu-35 to be two pK units higher than its water value, whereas in Asp-66 the pK_a is more than two units lower. Ponnuswamy et al. (1980), showed that at a detailed level, the local environmental hydrophobicity that surrounds each residue is distinct and can be quite different from all other residues, implying that a different screening would be required for each site (for PB calculations, each site would require its own internal permittivity value) as has been pointed out by Warshel (King et al., 1991).

In this work we hypothesize that a key source of single residue aberrations in pK_a is a result of the special protein microenvironment that exists around the singular residue. To test this hypothesis, a strategy has been developed to quantitatively characterize the hydrophilicity or hydrophobicity of the microenvironment around each titratable group, and then use this property to modulate the components of the electrostatic free energy operative in proteins. The advantage of this approach is that by using physicochemical properties independent of the parameters required to calculate the electrostatic effects, one might gain additional insight into the relation between protein structure and the forces that control its function. Accounting for every local microenvironment in the entire protein structure could become prohibitively costly. Therefore, to preserve the computational speed of the SCP, the quantitative calculation of the microenvironments has been implemented in a way that requires virtually no additional CPU time.

In the next section of this paper the method for quantifying the microenvironments is presented, and the dielectric screening function used previously (Mehler, 1996a; Mehler and Eichele, 1984) is reformulated to account for the local variation in the dielectric properties of proteins. The algorithm is applied to calculating the pK_a in bovine pancreatic trypsin inhibitor (BPTI), triclinic hen egg white lysozyme (HEWL), ribonuclease A (RnaseA), ribonuclease T1 (RnaseT), ribonuclease HI (RnaseHI), calbindinD9k (Cab) and the B1 immunoglobulin G-binding domain of protein G (ProtG). These seven proteins provide a data set of 103 measured pK_a values that have been used to develop the parametrization of the algorithm. Subsequently, the method is tested on turkey ovomucoid inhibitor domain 3, HIV protease, and other crystal forms of some of the proteins in the test set. Finally, the results obtained here are compared with other calculations and potential improvements for the quantification of the microenvironments are considered.

THEORETICAL BASIS AND CHARACTERIZATION OF MICROENVIRONMENTS

Calculation of the electrostatic free energy

The calculation of the pK_a of titratable groups in proteins is greatly simplified by considering the thermodynamic cycle (Bashford and Karplus, 1990; Bashford and Karplus, 1991; Warshel, 1981)

$$\begin{array}{ccc} & 2.303RTpK_a^A(s) & \\ AH(s) & \Longleftrightarrow & A^-(s) + H^+(s) \\ \Delta\mu_{s,p}(AH)\ \Uparrow\Downarrow & & \Uparrow\Downarrow\ \Delta\mu_{s,p}(A^-) \\ AH(p) & \Longleftrightarrow & A^-(p) + H^+(s) \\ & 2.303RTpK_a^A(p) & \end{array} \tag{1}$$

where (p) and (s) refer to protein or solvent, respectively. Thus, the pK_a of group A in the protein can be calculated from its pK_a in model solvent (water) and the additional changes in free energy that arise when the titratable group is transferred from water into the protein. Here, only the electrostatic contributions are considered and $pK_a(p)$ is obtained from the relation

$$pK_a^A(p) = pK_a^A(s) + \frac{w_A^{int} + \Delta w_A^{tr}}{2.303RT}, \tag{2}$$

where w_A^{int} is the interaction free electrostatic energy of the charged group in the field of all the other groups in the protein, and Δw_A^{tr} is the change in self-energy on transferring the group from water into the protein.

In Mehler (1996b), it was shown that with an SCP of the form

$$\Phi(\mathbf{r}) = \sum_j \frac{q_j}{D(r_j)r_j}, \tag{3}$$

where q_j is the charge at point $\mathbf{r}_j$, $r_j = |\mathbf{r} - \mathbf{r}_j|$ and $D(r)$ is a screening function, the total electrostatic free energy of a system consisting of M groups could be expressed as the sum of the interaction term and a transfer contribution as

$$w = \sum_{A=1}^{M} w_A^{int} + \Delta w_A^{tr}, \tag{4}$$

where A may refer to subgroups of each amino acid residue and other groups contributing to the total charge of the system. As discussed in the introduction, both experimental results and theoretical considerations indicated that $D(r)$ should have a sigmoidal form (for details, see the appendix) as originally proposed by Mehler and Eichele (1984).

For the calculation of the pK_a, it is not necessary to evaluate the total electrostatic free energy, but only the contribution from each titratable group in the field due to all the other groups in the protein. In most pK_a calculations the full titration charge is placed at one or more fixed positions in the protonatable moiety of the N titratable residues, and

the equilibrium charge state is calculated from the distribution of the charge over the 2^N ionization microstates (Bashford and Karplus, 1991; Beroza et al., 1991; Gilson, 1993). In the procedure developed in Mehler (1996b), the assignment of the ionization charge over the atoms of the protonatable moiety was determined variationally to allow it to respond to the protein environment. To reduce computing time, the direct determination of the 2^N charge microstates was bypassed by coupling the total variational ionization charge of each group, at the given pH, to the Henderson–Hasselbalch equation. In this paper the method presented in Mehler (1996b) is still used, but has been modified; the details of the modified approach are discussed in the appendix.

The calculations reported in Mehler (1996b) were based on a single screening function that had been parametrized earlier on the basis of pK_a shifts in bifunctional organic acids and bases (D_{wds} defined in Mehler and Eichele, 1984). Thus, the only parameter introduced in Mehler (1996b) was a scaling of the interaction energy of nearby charges. To improve and generalize the approach followed in Mehler (1996b), it is hypothesized that the screening for groups deeply buried in the protein should be less than for solvent exposed groups. This reflects the commonly accepted view that the protein interior is more hydrophobic than water. To implement this hypothesis, two new screening curves were constructed by redefining D_0 and λ (see appendix, Eq. A1) so that, for small separations, one curve would provide more screening than D_{wds}, used in Mehler (1996b), whereas the other would provide less screening. These two screening functions, D_1 and D_2, as well as the screening function in water, D_w, are shown in Fig. 1 (D_3 will be discussed below) and the values of D_0 and λ are given in Table I. Although these two screening functions appear to be very similar, for small values of r (see inset, Fig. 1) the energy is quite sensitive to small changes in the screening. Thus, at $r = 2$ Å, the difference between the two curves for unit charges amounts to nearly 4.5 pK units, while, at $r = 3$ Å, the difference is 1.4 pK units. In contrast, at $r = 8$ Å, the difference is only 0.09 pK units and continues to decrease with increasing r. Thus, for small separations between groups, the contributions to the interaction energy are quite sensitive to which screening function is used. This is also the case for the transfer energy, since the Born radii, R_a in Eq. A2, lie between about 1.2 Å and 2.6 Å. Which screening function is used for a given titratable group depends on the fraction of the group's solvent accessible surface area that is buried in the protein, i.e., the buried fraction. When these two screening functions are incorporated into the calculation, there was some improvement in the pK_a values, but the pK_a of Glu-35 in lysozyme was still too small by more than one pK unit.

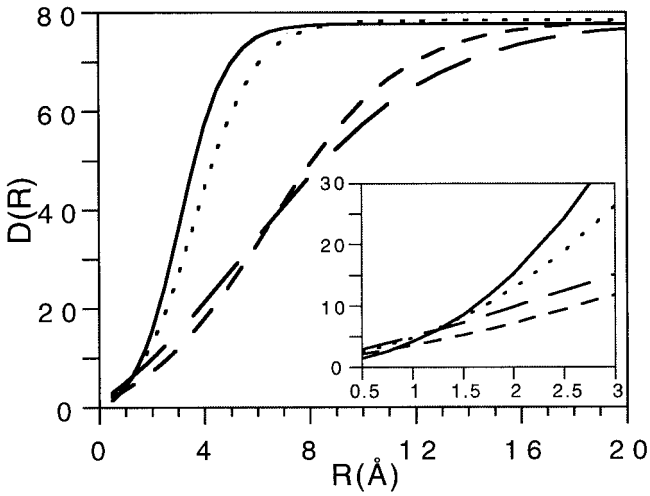


FIGURE 1 Screening functions defined in Table 1. *Solid line*, D_w; *short dashed line*, D_1; *long dashed line*, D_2; and *dots*, D_3. The inset is a detail from 0.5 to 3 Å. (See the appendix for details).

TABLE 1 Parameters for the dielectric screening functions*

	D_0	λ	Remarks
D_1	15.0	0.003	Default screening for BF[#] < 0.7
D_2	5.0	0.00475	Default screening for BF ≥ 0.7
D_3	2.0	0.011	Special screening for BF ≤ 0.3
D_w	2.353	0.01443	$D_s = 77.695$; $k = 35.644$[§]

*For the definition of the screening function and the parameters, see the appendix.
[#]BF = buried fraction.
[§]Four parameter fit to D_w (see text and the appendix).

Quantification of the microenvironment

Analysis of the local environments around Glu-35 and Asp-66 in lysozyme shows that for the former the nearest neighbors are largely hydrophobic residues, whereas for the latter residue the nearest neighbors are much more hydrophilic. The increased hydrophobicity around Glu-35 has consequences for both components of the electrostatic free energy: the energy transfer term will become more positive (more unfavorable), which can be effected by a decrease in R^p (see Eq. A6). The screening of the Coulomb potential should also decrease, especially for interactions with nearby atoms so that they are larger in magnitude relative to a less hydrophobic region. However, this will be partially offset by the fairly small partial charges on atoms in hydrophobic groups. Since w^{int} is usually negative, whereas w^{tr} will be positive for hydrophobic regions, the resulting shift in pK_a will depend on the subtle balance between these two quantities. The transfer of Asp-66 from water to a relatively hydrophilic environment, where $D_p(R^p)$ is larger, will result in a smaller transfer energy (less unfavorable; it could be negative if the local environment is more hydrophilic than water), and the Coulombic screening should be greater. These considerations clearly suggest that the relevant quantity for characterizing the local environment, here called the microenvironment, is the hydrophobicity or hydrophilicity (here we use Hpy to refer to these properties and their values in particular cases) of the residues or their fragments near the given titratable group.

There are a large number of Hpy scales available for evaluating the hydrophobicity of amino acid residues (Cornette et al., 1987). Generally these scales have been used to evaluate the hydrophobicity of a particular residue, or at least its side chain. Here, the hydrophobicity of the particular residue is not required, but instead, because the response of the residue to changes in the local environment is the quantity of interest, it is a quantitative measure of the hydrophobicity/hydrophilicity of the microenvironment that is needed. A hydrophobicity scale based on atoms or small fragments is most appropriate for the algorithm proposed here because it allows the microenvironment to be defined in a simple, distance dependent way. In addition, this type of scale can account for the fact that side chains generally do not have a single Hpy value along their entire length, nor do all the atoms of a side chain necessarily belong to a given microenvironment. A further important advantage of a fragment-based scale is that the presence of coenzymes, prosthetic groups, ligands, or nucleic acids can all be accounted for with a consistent scale. From these considerations, the Rekker Fragmental Hydrophobic Constants (Rekker, 1977; Rekker, 1979) were selected to evaluate the hydrophobicities of the microenvironments in the present calculations.

The Rekker scale was the first fragmental system developed and is based on assigning hydrophobicity parameters and correction factors to small functional fragments to calculate log P values (for details, see Rekker, 1977). This scale has been widely used in the pharmaceutical industry as a means to calculate partition coefficients of potential drug candidates. Here, we use it *not* to obtain the hydrophobicity/hydrophilicity of a particular residue, but to obtain this quantity for the local protein environment. This is done by (1) identifying fragments of all the residues that are in the neighborhood of the titratable residue, (2) assigning the Rekker coefficients to these fragments, (3) summing them to determine the local hydrophobicity/hydrophilicity of the region around the titratable residue, and (4) assigning an appropriate screening for this environment. Three other fragmental systems have been proposed, of which two are based on functional fragments (Leo et al., 1975; Suzuki and Kudo, 1990) and one is atom based (Ghose and Crippen, 1986). All the fragmental systems that have been proposed to date are calibrated on the basis of partition coefficients (log P). The parameters are therefore completely independent of protein structural information. Thus, their successful use to describe local protein environments will give some indication of the universality of the hydrophobic concept and the transferability of scales developed from different physical concepts. A recent report compared the four methods and showed that the Rekker system was one of the most accurate available (Mannhold et al., 1995).

In this initial application the original values of the hydrophobic fragmental constants and correction factors were used (Rekker, 1977). For each fragment the values were partitioned to the atoms belonging to that fragment, and to simplify the calculations, net charge on a specific group was not accounted for. Then the microenvironment can be defined on the basis of a distance criterion, and only atoms satisfying it are included in the calculation of the total Hpy around the given group. The initial set of values that have been used in the calculations are tabulated in Table 2. Note that the water value was estimated from the fragmental constants of aliphatic OH and H_{neg} (Rekker, 1977).

The microenvironment of any group is defined as the region that lies within a sphere of radius 4.25 Å from any (nonhydrogenic) atom belonging to the group. This radius was chosen because it essentially includes the first shell around each atom, but does not extend to atoms beyond this. In addition, this value is in reasonable agreement with the criterion used by Ponnuswamy et al. (1980). In analogy with the formula used to calculate log P from the fragmental hydrophobic constants (Rekker, 1977), the Hpy value for each microenvironment of a group is calculated from the formula

$$\mathrm{Hpy_A} = \sum_{a}^{N_A} \sum_{Bb} d_b \,\mathrm{RFHC_b}\ (r_{ab} \le 4.25\ \text{Å}) \quad (B \neq A). \tag{5}$$

In Eq. 5, $\mathrm{RFHC_b}$ is the contribution of atom b to the fragment's fragmental hydrophobic constant as given in Table 2, N_A is the number of atoms in group A and

$$d_b = \begin{cases} 1 & \text{if atom b has not been counted} \\ 0 & \text{if atom b has been counted,} \end{cases}$$

to ensure that each atom in the microenvironment is counted only once. Schematic representations of the microenvironments around Glu-25 and Asp-66 in lysozyme are presented

TABLE 2 Rekker fragmental hydrophobic constants* for amino acid residues

Fragment#	*f* Value§	Remarks¶
CH	0.24	
CH_2	0.53	
CH_3	0.704	
CONH	−1.57	CON −2.03; O: −1.01; N: −1.01; C: 0.0; $H^{\|}_{neg}$
Phenyl	1.886	Each C: 0.314
TRP	2.628	Each C or N: 0.222; H; 0.1
$–NH_2$	−1.428	
$–NH_3^+$	−1.028	
His	−0.14	C,N: −0.068; H: 0.10
HisH	0.04	C; N: −0.032; H: 0.10
COO	−1.29	O: −0.645; C: 0
COOH	−0.954	O: −0.477; C: 0
NH_2CO	−1.970	N: −0.868; O: −1.143; C: 0
SH	0	
S	−0.51	
–S–S–	0.37	
OH	−1.491	
Arg	−4.928	NH: −1.825; NH2: −1.428; C: 0.15
Arg^+	−4.531	
H_2O	−1.029	O: −1.58; $H^{\|}_{neg}$: 0.462; H: 0.09

*R. F. Rekker (1977, 1979).
#Structure of each protonatable fragment is given in Table 3.
§See Table in Appendix of R. F. Rekker (1979).
¶Polar hydrogens are 0.0 unless stated otherwise.
‖See Rekker (1977, p. 108).

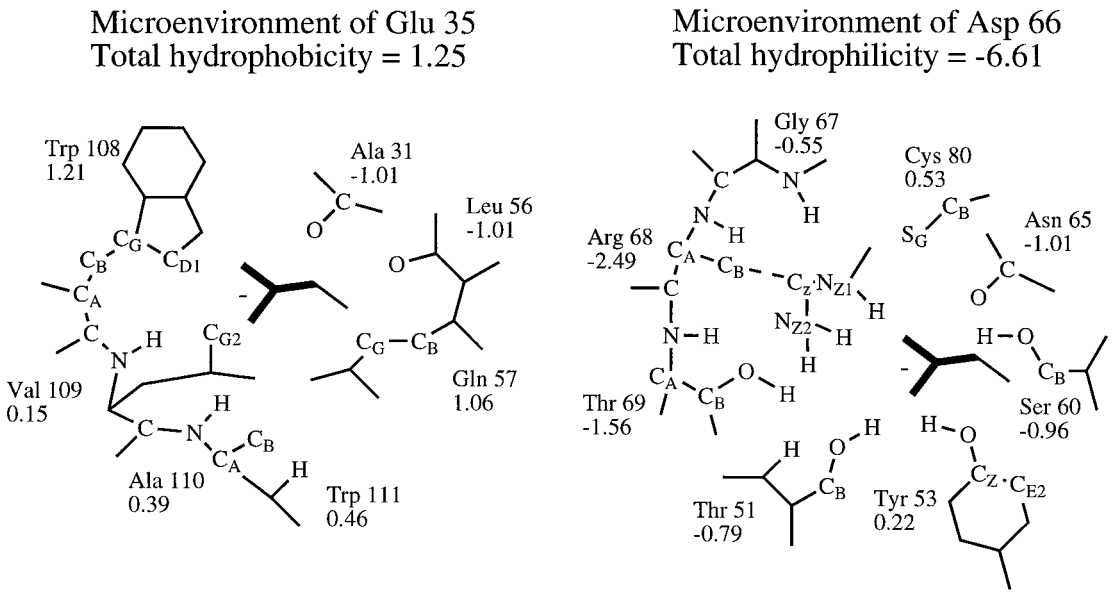


FIGURE 2 Residues defining the microenvironments of Glu-35 and Asp-66 in HEWL. The number under each residue is its contribution to the total hydrophobicity or hydrophilicity. Atoms within 4.25 Å of any atom of the titratable moiety (shown in bold) are labeled. See text for details.

in Fig. 2. These diagrams show that in most cases only some atoms of a given amino acid residue are located within 4.25 Å of one or more atoms of the group in question. Moreover, these arrangements of the atoms in space clearly show how the protein architecture has evolved to create the hydrophobic microenvironment around Glu-35 and the hydrophilic microenvironment around Asp-66.

In addition to the Hpy of the protein microenvironments, the Hpy in the solvent are needed to calculate the gain or loss of hydrophilicity in the solvent to protein transfer process. These were calculated by immersing each titratable residue in an equilibrated water droplet and carrying out 100 ps of dynamics after heating to 300 K. The residues were capped with neutral fragments at the termini. From the interval 70–100 ps, 31 structures were extracted, an Hpy calculated for each structure using the water fragment values given in Table 2, and the final value was obtained from the average of the Hpy values calculated from the 31 structures using Eq. 5. The results for the protonatable residues are given in Table 3 and the atoms belonging to the titratable groups, as defined in the PAR19 parameter set of CHARMM (Brooks et al., 1983), are also shown. It should be noted that in analogy to Rekker's (1977) formula for calculating log P from the hydrophobic fragmental constants, Hpy as defined in Eq. 5 is an extensive quantity. Therefore, in an environment that is uniformly hydrophobic or hydrophilic its magnitude will tend to increase with increasing size of the group. In addition, basic groups will tend to orient the waters with the dipole pointing away from the titration site (oxygen pointing toward the group), which will tend to increase the hydrophilicity of the microenvironment, whereas for the acidic groups the dipole will tend to orient toward the titration site (hydrogens nearer the acidic group) tending to increase the hydrophobicity.

Three measures for characterizing the hydrophobicity of the microenvironments are defined: first is Hpy_A defined in Eq. 5, second is the total hydrophobicity of the microenvironment due to the fraction of the group buried in the protein and the fraction exposed to solvent

$$THpy_A = BF_A\, Hpy_A + (1 - BF_A)Hpy_A^0, \qquad (6)$$

where Hpy_A^0 is the value of Hpy in the solvent, and finally, a measure of the hydrophobicity of a given group's microenvironment relative to what it would be if the group was totally immersed in water, i.e.,

$$rHpy_A = THpy_A/Hpy_A^0. \qquad (7)$$

These quantities have been calculated for all proteins in the data set, and average values, calculated from the seven proteins in the data set for each type of titratable residue, are given in Table 4. It is of interest that the buried fractions and microenvironments are quite different for the different residue types. The reciprocal of *r*Hpy is an indication of the increase in the environmental hydrophobicity on transfer-

TABLE 3 Hydrophilicity of titratable groups in water

Residue	Titratable Group*	Hpy^0
His	C-Imidazole	−18.453
Lys	C-NH_3	−11.420
Arg	C(NH)C(NH_2)$_2$	#
Tyr	C–OH	−8.368
Asp	C–COO	−13.492
Glu	C–COO	−12.832

*Moiety of the titratable residue that carries the charge as defined in CHARMM PAR19 (Brooks et al., 1983).
#Not calculated, set to Lys value.

TABLE 4 Average values of the microenvironment properties for each type of titratable residue

Residue (No)	$\overline{BF}$	$\overline{Hpy}$	$\overline{THpy}$	$\overline{rHpy}$
Tyr (31)	0.76 (0.20)	−0.55 (1.71)	−2.48 (2.08)	0.30 (0.25)
His (13)	0.63 (0.26)	−3.53 (2.90)	−9.51 (3.21)	0.52 (0.18)
Asp (36)	0.56 (0.25)	−3.55 (2.98)	−8.20 (2.83)	0.61 (0.25)
Glu (45)	0.52 (0.25)	−1.35 (2.80)	−7.11 (3.13)	0.55 (0.24)
Arg (32)	0.40 (0.23)	−1.02 (1.68)	−7.50 (1.93)	0.66 (0.17)
Lys (49)	0.33 (0.22)	−0.42 (1.17)	−7.77 (2.51)	0.68 (0.22)

Averages calculated from the seven protein data set; values in parentheses in columns 2–5 are the standard deviations.

ring the titratable group from solvent to its final position in the protein. Thus, for the lysines, this increase is only about 1.5, which is not surprising considering the small value of $\overline{BF}$. Histidine is the second most buried residue, but on the average, its microenvironment appears to be quite hydrophilic, and the increase in hydrophobicity is less than two. The titratable moiety of Tyr is the most buried with the most hydrophobic microenvironments. In an initial more extended study of the microenvironments of all groups in a sample of 13 globular proteins comprising a total of 7132 groups, it was found that $\overline{BF}$ and $\overline{rHpy}$ had values of 0.81 and 0.43, respectively (Mehler, unpublished results). Therefore, the average increase in hydrophobicity is about 2.3, which compares well with the value of 2.6 obtained in an earlier study by Ponnuswamy et al. (1980) using very different methods and assumptions.

The buried fractions (BF) and microenvironment properties of each titratable moiety in lysozyme are given in Table 5. It is seen that only a few values of Hpy are greater than zero. However, as will be discussed below, it was found that titratable groups with a BF less than 0.7 should be considered as solvent exposed for the pK_a calculations. Considering groups with BF > 0.7, only Glu-35 has a positive value of Hpy. The values of THpy or *r*Hpy provide additional insight into the change in the hydrophobicity of the microenvironment when the group is transferred from water to protein. In HEWL, THpy is always negative and the highest value is −0.59 for Glu-35; the next largest value is −1.99 for Tyr-53. Comparison of the microenvironment of Glu-35 with the average values and standard deviations for Glu (Table 4) also clearly exhibits the unusual nature of the local environment around Glu-35.

Considering *r*Hpy, the smaller the value the more hydrophobic the protein microenvironment of the group relative to water. Thus, the value of Glu-35 is the smallest (it implies that its protein microenvironment is around 20 times more hydrophobic than water!) in HEWL, and the next value is five times larger. Comparison of the microenvironments of Glu-35 and Asp-66 shows how the structural characteristics (see Fig. 2) translate into the quantitative description of the differences in hydrophobicity between the two microenvironments provided by the three hydrophobicity parameters. In HEWL, only Asp-52 has an Hpy value slightly more negative than Asp-66. Thus, the protein contribution to the

TABLE 5 Buried fractions, screening function assignments, and properties of the microenvironments of HEWL

Residue No.	Residue	BF*	D[#], tr, int	Hpy[§]	THpy[¶]	*r*Hpy[‖]
1	NTERM	0.493	1,1	−1.22	−6.39	0.56
1	LYS	0.441	1,1	1.88	−5.55	0.49
5	ARG	0.597	1,1	−3.67	−6.79	0.59
7	GLU	0.267	3,1	−1.36	−9.77	0.76
13	LYS	0.310	1,1	−0.65	−8.08	0.71
14	ARG	0.270	1,1	−0.38	−8.44	0.74
15	HIS	0.711	2,1	−1.96	−6.73	0.36
18	ASP	0.561	1,1	−0.09	−5.97	0.44
20	TYR	0.769	2,2	−0.14	−2.04	0.24
21	ARG	0.223	1,1	−0.77	−9.04	0.79
23	TYR	0.547	1,1	1.04	−3.22	0.38
33	LYS	0.404	1,1	−1.14	−7.27	0.64
35	GLU	0.869	2,2	1.25	−0.59	0.05
45	ARG	0.289	3,1	−2.22	−8.76	0.77
48	ASP	0.497	1,1	−4.06	−8.80	0.65
52	ASP	0.786	1,1	−6.95	−8.35	0.62
53	TYR	0.937	2,2	−1.56	−1.99	0.24
61	ARG	0.607	1,1	−1.78	−5.57	0.49
66	ASP	0.931	1,2	−6.70	−7.17	0.53
68	ARG	0.336	1,1	−3.52	−8.77	0.77
73	ARG	0.423	1,1	0.67	−6.31	0.55
87	ASP	0.391	1,1	−4.15	−9.84	0.73
96	LYS	0.709	2,1	−0.62	−3.76	0.33
97	LYS	0.167	0,3	−1.01	−9.68	0.85
101	ASP	0.648	1,1	−2.77	−6.54	0.48
112	ARG	0.377	1,1	−0.48	−7.29	0.64
114	ARG	0.282	1,1	1.26	−7.84	0.69
116	LYS	0.170	0,3	−0.87	−9.62	0.84
119	ASP	0.393	1,1	−2.38	−9.13	0.68
125	ARG	0.321	1,1	−0.96	−8.06	0.71
128	ARG	0.023	0,3	0.00	−11.16	0.98
129	CTERM	0.197	1,1	0.53	−10.20	0.80

*Buried fraction of the titrating group.
[#]Screening function assignments to transfer energy (tr), interaction energy (int).
[§]See Eq. 5.
[¶]See Eq. 6.
[‖]See Eq. 7.

microenvironments of these two residues are the most hydrophilic in HEWL, while that of Glu-35 is the most hydrophobic. It should finally be noted that the values of *r*Hpy can be negative, indicating extreme hydrophobicity, or greater than one, indicating a microenvironment more hydrophilic than water.

Parametrization and the microenvironment

The hydrophobicity measures of the microenvironments in lysozyme (Fig. 2, Table 5) reveal a complex mosaic of different dielectric regions that present determinants of electrostatic properties that are quite independent of the buried fraction. In several important cases a totally buried residue, which usually is considered to exist in a low dielectric region, is shown to actually be found in a quite hydrophilic microenvironment. Thus, transfer of such a residue from water to protein is, in fact, only a small perturbation because

the screening in such a microenvironment of the protein will be close to that in water. Nevertheless the microenvironments around the titratable groups of proteins show that they are generally more hydrophobic than in water (compare Tables 3, 4, and 5).

The values of D_0 and λ used to define D_1 and D_2 were determined as described above to account for the expected differences in dielectric behavior depending on the extent of burial, but the explicit values were not calibrated against the pK_a of the protein data set. To incorporate these two screening functions into the algorithm, a switching value had to be determined, which was accomplished by assigning D_1 to all protonatable groups with a BF below a given threshold and D_2 for the others. A value of 0.7 was found to give the lowest root mean square deviation (rmsd) between calculated and experimental pK_a values in the data set and is used for switching between D_1 ($BF < 0.7$) and D_2 ($BF \geq 0.7$) for all subsequent calculations. The calculations carried out in this way do not consider the explicit effects of the microenvironments and therefore are labeled as "menv-free" calculations in the following discussions.

From Fig. 2 and Tables 3, 4, and 5, it is apparent that for individual residues there can be considerable variation around the mean values of the properties characterizing the microenvironments. It seems reasonable, therefore, to make use of this to identify microenvironments that exhibit especially large deviations from the mean, and use the physical implications of these deviant regions to modulate the electrostatic properties of the titratable groups in question. To incorporate this idea into the algorithm, it is necessary to define quantitatively when the properties of a microenvironment deviate enough from the mean values to require modification of the menv-free assignments of the screening functions given above. These parameters were determined in the following way. The microenvironments of groups for which the calculated pK_a showed the largest errors were analyzed to rationalize on physical-chemical grounds the sources of the errors and the changes in screening that might lead to improvement. However, only reduction in overall rmsd was used to decide if a particular combination of screening functions was to replace the default assignments depending only on the BF value. This approach was taken because of the observation from the comparison of results using different crystal forms that a particular large error may be due to a conformation that is not relevant for the solution structure (see below) or that the sources of a particular error may be due to one or more inadequacies of the method. For these reasons, and because the results indicate that the quantification of the microenvironments will need further extension and refinement, exhaustive optimization was not carried out. Modification of the default screening was applied to titratable groups with $BF \geq 0.7$, where the hydrophobicity or hydrophilicity of the microenvironments was far from the mean value; the border region with $0.7 \leq BF \leq 0.8$ was treated differently from groups with $BF > 0.8$. For groups with very small buried fractions ($BF < 0.3$) and microenvironments very close to water, a third screening function was introduced (D_3 in Table 1 and Fig. 1) that is much closer to D_w than D_1 or D_2. Table 6 gives the thresholds for which the default conditions are modified for calculating the electrostatic energies, and they are discussed in the following paragraphs.

The protonatable moiety of Glu-35 in lysozyme is a buried titratable group in a very hydrophobic environment (see Fig. 2 and Table 5). According to Rekker (1977), Table IX, 1, p. 301), amino acid residues for which the fragmental hydrophobic constants sum to a relative hydrophobicity greater than one, are considered lipophilic, and this value has been used here to define an extremely hydrophobic microenvironment. As mentioned above, only a few buried tyrosines are in more hydrophobic microenvironments than Glu-35, but their pK_a values have not been measured. The effect of very hydrophobic environments is to reduce screening and to make the transfer of the titratable group from water to the protein more unfavorable. Since the pK_a has been measured only for Glu-35, any parameter adjustment is arbitrary, so that in these cases (see Table 6) Δw^{tr}

TABLE 6 Buried fraction and microenvironment-dependent assignments of screening functions and scaling factors

BF*	D_0[#]	Hyp[§]	rHpy[¶]	D[‖]	Remarks
$0.3 \leq \cdots \leq 0.7$	1	—	—		Always D_1 for w^{int} and Δw^{tr}
$\cdots \geq 0.7$	2	$\cdots > 1.0$	$\cdots < 0.25$	2; 2	Δw^{tr} Scaled by 1.75
$0.7 \leq \cdots \leq 0.8$	2	$-4.0 < \cdots < -0.50$	—	1; 2	Asp; Glu**
$\cdots > 0.8$	2	$\cdots < -4.0$	—	1; 2	Asp; Glu**
$0.7 \leq \cdots \leq 0.8$	2	$\cdots \leq -4.0$	—	1; 1	All residues[##]
$\cdots < 0.3$	1	—	> 0.8	–; 3	Δw^{tr} Set to 0.0
$\cdots < 0.3$	1	$\cdots < -1.0$	$0.7 < \cdots \leq 0.8$	3; 1	All residues

*Buried fraction.

[#]Default screening functions for calculating Δw^{tr} and w^{int} are used for all regions not explicitly defined in the Table.

[§]Hydrophobicity of the protein microenvironment.

[¶]Relative hydrophobicity (see text).

[‖]Changes in assignment of screening functions due to effects from the microenvironments. The first number in the pair is the screening function used for Δw^{tr} and the second number is used for w^{int}.

**For Lys, Tyr, Arg, and His, the screening assignments are reversed: D_2 is used for Δw^{tr} and D_1 is used for w^{int} (see text.)

[##]If $BF > 0.8$ and $THpy < -10.0$ or $rHpy > 0.8$ both w^{int} and Δw^{tr} are calculated with D_1.

has been scaled to increase the energy penalty for transferring a charge from solvent to the protein interior.

The average value of Hpy for the microenvironments surrounding the titratable residues in the seven proteins of the data set is about −1.5, but, as noted (see Table 4), the variations around this mean are considerable. For Asp and Glu when BF > 0.7, the microenvironments of a fairly large number of them are considerably more hydrophilic than −1.5. For such cases, the screening D_2 may imply a microenvironment that is too hydrophobic so that the large, unfavorable transfer energy results in pK_a values that are too high for the acids and too low for the bases. For 14 of the 53 aspartic and glutamic acids in the data set, this was the case, and the overestimation of these pK_a values was corrected by using D_1 instead of D_2 for calculating Δw^{tr}, because this implies transfer to a more hydrophilic microenvironment. For tyrosine and the bases the number of cases meeting the conditions of burial with BF greater than 0.7 and surrounded by very hydrophilic microenvironments was too small to assess any trends. It was found empirically, however, that using D_1 instead of D_2 for calculating w^{int} gave a small improvement in the overall rmsd. This result may be fortuitous because it appears to be counterintuitive. At the same time the response of any titratable group to its environment is determined by the balance between w^{int} and Δw^{tr}, so that, as the differences in properties given in Table 4 suggest, this response may depend on residue type. Finally, the third screening function, D_3, is only used when the properties of the microenvironment indicate conditions that are very close to pure solvent.

Computational details

The calculations reported below were carried out using protein coordinates obtained from the protein data bank (Bernstein et al., 1977) as follows: BPTI:4PTI* (Marquart et al., 1983), HEWL:2LZT* (Ramanadham et al., 1981) and 1HEL (Wilson et al., 1992), RnaseA:3RN3* (Howlin et al., 1989), RnaseT:3RNT* (Kostrewa et al., 1989), RnaseHI: 2RN2* (Katayanagi et al., 1992) and 1RNH (Yang et al., 1990), Cab:3ICB* (Szebenyi and Moffat, 1986), protG: 1PGA* and 1PGB (Gallagher et al., 1994), turkey ovomucoid inhibitor domain 3(OMT3):1PPF (Bode et al., 1986), and hiv protease (1HIV) (Thanki et al., 1992), where the starred coordinate sets were used in the optimization procedure. The preparation of the coordinates was carried out in the same way as in I, and the values of the model pK_a used here are *N*-term, 7.5; C-term, 3.8; His, 6.3; Glu, 4.4; Asp, 4.0; Tyr, 10.0; Lys, 10.4; and Arg, 12.0. The partial charges and group structures of the amino acid residues are those defined in the PAR19 topology file of CHARMM (Brooks et al., 1983), whereas the partial charges of the neutral forms of the titratable groups were taken from the PAR22 (MacKerell et al., 1992; MacKerell et al., 1995) topology files. The $(BF)_a$ of Eq. A6 were evaluated from the solvent-accessible surface areas calculated with CHARMM using a probe radius of 1.0 Å, and the group average (defined by CHARMM) was used to scale all the Δw_a^{tr} for the given group (see Table 5 for the HEWL values). The Born radii were determined in the same way as described in the appendix of Mehler (1996b) except that atomic values were calculated, and the value obtained for D_w was used in both self energy terms in Eq. A6. For charges closer than 1.9 Å, w^{int} was scaled by a factor of 0.35. This scaling of nearby charges is similar to that used in Mehler (1996b) and compensates for the breakdown of the point charge model for nearby charges. However, due to the modified variational procedure (see Appendix) the cutoff distance could be reduced from 2.8 Å in Mehler (1996b) to the present value. Therefore, the scaling is essentially only used where uncompensated steric clashes leave two charges too close together. In addition, ionic strength is taken into account using a Debye screening as in Mehler (1996b). Finally, the polar protons of histidine are treated symmetrically in assigning the charge of the neutral species.

RESULTS AND DISCUSSION

Calculation of pK_a

The calculated pK_a for the test suite of seven proteins are presented in Tables 7–13. To evaluate the effect on the pK_a of introducing the dependence of the electrostatics on the microenvironment, the $pK_{1/2}$ calculated with the screening functions D_1 and D_2 only, i.e., menv-free are also tabulated. Calculations were carried out in steps of 0.25 pH units and the $pK_{1/2}$ of each titratable group was estimated by linear

TABLE 7 pK_a values for BPTI

Residue No.	Residue	$pK^0_{1/2}$*	Error#	$pK_{1/2}$§	Error#
1	NTERM	7.95	0.06	8.11	0.22
3	ASP	3.70	0.13	3.81	0.24
7	GLU	3.97	0.08	3.97	0.08
10	TYR	9.84	0.38	9.86	0.39
15	LYS	10.51	0.08	10.45	0.02
21	TYR	10.06	0.12	10.07	0.13
23	TYR	10.11	−0.89	10.68	−0.31
26	LYS	10.44	0.34	10.42	0.32
35	TYR	10.31	−0.29	10.35	−0.25
41	LYS	10.61	0.01	10.62	0.02
46	LYS	10.27	0.40	10.36	0.49
49	GLU	4.16	0.16	4.21	0.21
50	ASP	2.37	−0.81	2.37	−0.81
58	CTERM	3.05	−0.00	3.01	−0.04
	rmsd		0.379		0.326

*pK_a values using default assignments of the screening functions (see Table 6).

#Error = pK(calc) − pK(exp). The experimental data were taken from March et al. (1982). The experiments were carried out at 41°C and the effective pK_a values corrected to 25°C. The pK_a values used here are averages over values reported from NMR shifts on different carbons; uncertain assignments were excluded from the average. See March et al. (1982).

§pK_a values using microenvironment-modulated screening function assignments.

interpolation between the two pH values where the fraction of charged species crossed the 0.5 value. The ionic strength was adjusted to be as close as possible to the experimental conditions and is 0.1 except where stated otherwise. For most of the proteins in the data set the experimental pK_a values to be used for calculating rmsd are clear from the quoted reference, but in some cases more than one choice is possible, and these will be noted as required for clarity. Finally, the Null model introduced by Antosiewicz et al., (1994) assumes that the pK_a values in the protein are the same as in the model (solvent), and it was noted that the rmsd for this model is usually quite small, since most of the pK_a values do not shift much from their model values. Because of this the rmsd resulting from the calculations provide only a partial measure of reliability. Of equal or greater importance are the number and size of the large errors, and these will also be considered in assessing the reliability of the approach.

TABLE 8 pK_a values for CabD9K

Residue No.	Residue	$pK^0_{1/2}$*,#	Error§	$pK_{1/2}$*,¶	Error§
1	LYS	10.97	0.37	10.99	0.39
7	LYS	10.95	−0.40	10.85	−0.50
12	LYS	11.63	0.63	11.41	0.41
16	LYS	10.71	0.62	10.68	0.59
25	LYS	11.75	0.06	11.75	0.06
29	LYS	11.63	0.64	11.63	0.64
41	LYS	10.52	−0.37	10.54	−0.35
55	LYS	11.61	0.22	11.15	−0.24
71	LYS	10.50	−0.22	10.56	−0.16
72	LYS	10.95	−0.02	10.96	−0.01
	rmsd		0.413		0.390

*The pK_a values were calculated at I = 5 mM.
#pK_a values using default assignments of the screening functions (see Table 6).
§Error = pK(calc) − pK(exp). Experimental values from Kesvatera et al. (1996).
¶pK_a values using microenvironment-modulated screening function assignments.

Bovine pancreatic trypsin inhibitor

The titratable groups in BPTI are mostly solvent exposed, thus the pK_a values are not shifted far from the null values (see Table 7). For only three residues (Arg-20, Tyr-23, and Tyr-35) are the BFs in the protein greater than 0.7. Tyr-23 is most noteworthy because the microenvironment is very hydrophilic (Hpy = −5.4), and D_1 was used to screen the SCP (see Table 6) causing an increase of about 0.5 pK units in $pK_{1/2}$, which reduces the error to about ⅓ of its menv-free value. The rmsd for BPTI is quite small in agreement with other calculations (Antosiewicz et al., 1994; Demchuk and Wade, 1996; Juffer et al., 1997).

Calbindin D_{9k}

In Cab 9 of the 10 lysine residues are exposed to solvent, so that the microenvironments are less significant in modulating their electrostatic effects. The results given in Table 8 show that the calculated pK_a values of all 10 lysines are in good agreement with experiment, and the rmsd is slightly smaller than that calculated from the menv-free conditions. Inspection shows that small shifts in several p/Ca values lead to the decrease in rmsd with the largest shift of 0.2 pK units exhibited by Lys-12.

The authors of the NMR measurements (Kesvatera et al., 1996) also used a Monte Carlo method (Svensson et al., 1990; Svensson et al., 1991) to calculate the pK_a of the lysine residues. However, the resulting rmsd of 0.86 is larger than the null model value of 0.74. Juffer et al. (1997) calculated the pK_a values using a boundary element method to solve the Poisson–Boltzmann equation. In agreement with other studies (Antosiewicz et al., 1996) they found that with a low internal dielectric constant (ϵ_i) the results deviated substantially from the observed values. With an ϵ_i = 20, agreement was much better (rmsd = 0.51). Using the value of the solvent's dielectric constant for ϵ_i gave an rmsd of 0.36 for both zero and 0.1 ionic strength. Except for Lys-55, the other lysine residues are exposed to solvent so that the results of Juffer et al. (1997), that high values of ϵ_i yield better results, is not unexpected. It is noteworthy that the pK_a values of Lys-25 and Lys-55 are shifted upward one or more pKunits. The BF of these two residues is about 0.5 and 0.8, respectively, so that the upward shift in pK_a shows that the interresidue interactions can be as important as the transfer terms in accounting for large changes in pK_a from the Null model values.

Hen egg lysozyme

The main problem with HEWL is proper calculation of the pK_a of Glu-35, and, at the same time, to obtain reasonable values for the pK_a of the other titratable residues. The source of the difficulty is the high hydrophobicity of the microenvironment of Glu-35 as discussed above. Quantification of the hydrophobic characteristics of the microenvironments reveals why solution of the PB equation requires a low ϵ_i to give a reasonable result for the pK_a of Glu-35, but for many other pK_a such a low value causes large deviations from the experimental results. With a larger value of ϵ_i, this trend is reversed (Antosiewicz et al., 1994).

The pK_a values for HEWL are given in Table 9, and from the rmsd values it is seen that introduction of the microenvironments leads to substantial improvement. A large portion of this improvement is due to Glu-35, but it is seen that the errors in several other pK_a values are also reduced, e.g., Tyr-53, Lys-96, and His-15. Other residues show smaller improvements, and a few pK_a values show slightly larger errors. Glu-7, which deviates by more than one pK unit from the experimental value, is only 27% buried and not influenced by the microenvironment. This problem is further discussed in the next section.

Demchuk and Wade (1996) attempted a local characterization for identifying a subset of residues described by a

TABLE 9 pK_a values* of hen egg lysozyme

Residue No.	Residue	$pK^0_{1/2}$#	Error§	$pK_{1/2}$¶	Error§	
1	NTERM	8.66	0.76	8.66	0.76	
1	LYS	10.42	−0.18	10.41	−0.19	
7	GLU	3.85	1.12	3.74 (3.12)	1.01	(0.34)
13	LYS	10.42	0.12	10.42 (11.88)	0.12	(1.58)
15	HIS	6.00	0.42	5.68	0.10	
18	ASP	3.52	0.74	3.52	0.74	
20	TYR	10.99	0.69	11.00 (9.50)	0.70	(−0.80)
23	TYR	9.75	−0.05	9.74	−0.06	
33	LYS	11.39	0.99	11.39	0.99	
35	GLU	4.92	−1.23	5.87	−0.28	
48	ASP	3.11	−0.29	3.11	−0.29	
52	ASP	3.95	0.28	3.49	−0.18	
53	TYR	11.00	−1.10	11.56	−0.54	
66	ASP	2.63	0.63	2.62	0.62	
87	ASP	3.14	0.30	3.14	0.30	
96	LYS	11.23	0.53	10.68	−0.02	
97	LYS	10.65	0.55	10.55	0.45	
101	ASP	4.18	0.01	4.18	0.01	
116	LYS	10.12	−0.08	10.24 (11.38)	0.04	(1.18)
119	ASP	3.11	0.26	3.11	0.26	
129	CTERM	2.67	−0.14	2.64 (1.62)	−0.18	(−1.2)
	rmsd		0.622		0.486	

*The trigonal structure (Ramanadham et al., 1981) was used for all calculations except the values in parentheses that were obtained from the tetragonal crystal structure (Wilson et al., 1992).

#pK_a values using default assignments of the screening functions (see Table 6).

§Error = pK(calc) − pK(exp). The experimental pK_a values used here were derived from Kuramitsu et al. (1980) for the lysines and tyrosines, and, for aspartate, glutamate, and histidine, averages were calculated from the results reported by Kuramitsu et al. (1980) and Bartik et al. (1994).

¶pK_a values using microenvironment-modulated screening function assignments.

high internal dielectric constant (HD) and a set to be described by a low value (LD). For HEWL, their overall rmsd was 0.63 as compared to 0.49 obtained here. The HD subset (see Demchuk and Wade, 1996) gave an rmsd of only 0.37, while here it was 0.50.

The titration curves of Glu-35 obtained from the menv-free calculation and with inclusion of the microenvironments are shown in Fig. 3. Both curves titrate over a fairly extended pH range. The net effect of the hydrophobic microenvironment (scaling of w^{tr} by 1.75) is to decrease the slope of the titration curve thereby shifting $pK_{1/2}$ to a larger value and further extending the titration range, so that between a pH of 4 and 6 the equilibrium charged fraction changes from about 0.25 to 0.55. This extended titration range of Glu-35 may have functional significance because it ensures that an ample concentration of lysozyme with protonated Glu-35 is available over a fairly large pH range around the optimal value (Fukamizo et al., 1983). The dependence of the interaction energy and transfer energy on pH is presented in Fig. 4. The effect of scaling is clearly seen in w^{tr} for pH values greater than about eight. In this region, w^{tr} continues to increase, only starting to level off at very high pH. In contrast, the transfer energy from the

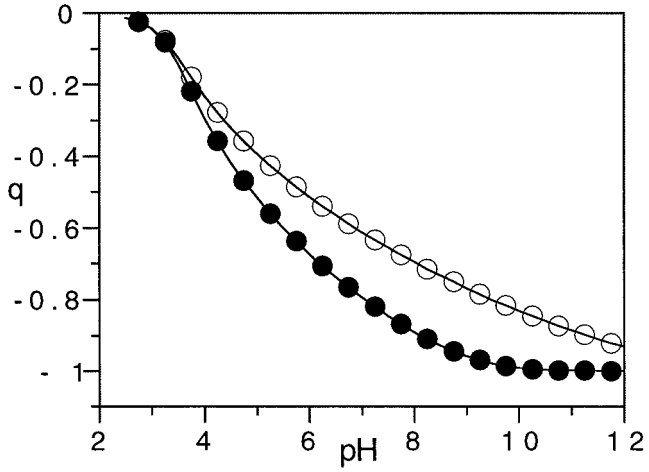


FIGURE 3 Glu-35 of HEWL titration curve. *Closed circles*, menv-free calculation; *open circles*, includes microenvironment.

menv-free calculation levels off around a pH of 10. The behavior of the energy components in the important pH range 4 to 6 is surprising. Here w^{tr} and w_0^{tr} are nearly identical, in spite of the scaling factor, but w^{int} is less negative than w_0^{int}. From Eq. A6, w^{tr} ($q = 1$) has a value of 6.8 Kcal/mol and scaling by 1.75 increases this to 12.0 Kcal/mol. At a pH of 6 the menv-free calculation gives $q_{35} = -0.673$, yielding 3.1 Kcal/mole for the transfer free energy. Interestingly, the equilibrium value of the scaled Δw^{tr} is 3.2 Kcal/mol, which is achieved by reducing the menv-free value of q_{35} to -0.514. In this case the variational procedure lowers the net charge to minimize the unfavorable transfer contribution with the consequence that the $pK_{1/2}$ value is increased. The effect on the interaction energy is a small increase of 0.3 Kcal/mol (see Fig. 4).

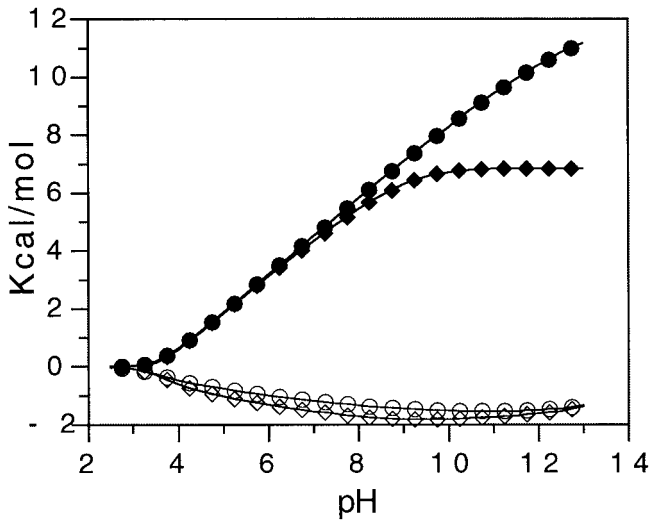


FIGURE 4 Energy components of Glu-35 as a function of pH. *Closed symbols*, transfer energy; *open symbols*, interaction energy; *diamonds*, menv-free calculation; *circles*, complete calculation.

Immunoglobulin G-binding domain B1 of Protein G

The pK_a values of this small protein (56 amino acid residues) were recently reported by Khare et al. (1997), and the calculated results are reported in Table 10. The microenvironments have very little effect on the calculated pK_a because only 3 of the 21 titratable residues are buried (Tyr-3, Tyr-45, and Glu-56). Hpy of Glu-56 is 0.93 and *r*Hpy is 0.10, so that the microenvironment of this residue is quite hydrophobic, but not as hydrophobic as for Glu-35 of HEWL. Comparison with the calculated $pK_{1/2}$ given by Khare et al. (1997) shows reasonably good agreement in the overall trends of the deviations.

Ribonuclease A and Ribonuclease T1

Results for RnaseA are given in Table 11. Introduction of the microenvironments leads to overall improvement in the calculated pK_a. His-48 shows the biggest improvement when the environmental effects are incorporated while the error in the pK_a of Glu-2 also decreases substantially. The overall rmsd obtained here was 0.55, whereas Demchuk and Wade (1996) obtained 0.63. For the subset labeled HD, they obtained an rmsd of 0.44, whereas here, an rmsd of 0.60 was obtained. These results from HEWL and RnaseA suggest that the HD selection criterion proposed by Demchuk and Wade (1996) can successfully select a subset of titratable residues with high ϵ_i, but that a single smaller ϵ_i for the low dielectric residues is not sufficient for discriminating the differences in the microenvironment of these residues.

TABLE 10 pK_a values* of the B1 immunoglobulin G-binding domain of protein G

Residue no.	Residue	$pK^0_{1/2}$#	Error§	$pK_{1/2}$¶	Error§
10	LYS	10.84	−0.17	10.63	−0.37
15	GLU	3.95	−0.45	3.95 (3.12)	−0.45 (−1.28)
19	GLU	3.86	0.16	3.86	0.16
22	ASP	2.87	−0.03	2.87	−0.03
27	GLU	3.11	−1.39	3.11	−1.39
28	LYS	11.63	0.73	11.63 (10.56)	0.73 (−0.33)
33	TYR	10.22	−0.78	10.22	−0.78
36	ASP	4.41	0.61	4.41	0.61
40	ASP	4.15	0.15	4.04	0.04
42	GLU	4.53	0.13	4.46	0.06
46	ASP	3.81	0.21	3.81	0.21
47	ASP	2.63	−0.78	2.62	−0.78
56	GLU	4.13	0.13	4.14	0.14
	rmsd		0.583		0.588

*pK_a values were calculated from the crystal structure 1pga, except for the values in parentheses that were calculated from 1pgb (Gallagher et al., 1994).
#pK_a values using default assignments of the screening functions (see Table 6).
§Error = pK(calc) − pK(exp); experimental pK_a values from Khare et al. (1997).
¶pK_a values using microenvironment-modulated screening function assignments.

TABLE 11 pK_a values of RnaseA

Residue No.	Residue	$pK^0_{1/2}$*	Error#	$pK_{1/2}$§	Error#
1	NTERM	8.12	0.52	8.13	0.53
2	GLU	4.31	1.52	3.84	1.03
9	GLU	4.55	0.55	4.56	0.56
12	HIS	6.28	0.08	5.83	−0.37
14	ASP	2.63	0.63	2.62	0.62
38	ASP	3.72	0.62	3.81	0.71
48	HIS	7.11	1.11	6.44	0.44
49	GLU	4.33	−0.37	4.34	−0.36
53	ASP	3.95	0.05	3.95	0.05
83	ASP	3.15	−0.35	3.11	−0.39
86	GLU	4.47	0.37	3.76	−0.34
105	HIS	7.09	0.40	7.10	0.40
111	GLU	4.37	0.87	4.37	0.87
119	HIS	6.25	0.15	6.25	0.16
121	ASP	3.86	0.76	3.92	0.82
124	CTERM	2.42	0.02	2.42	0.02
	rmsd		0.651		0.554

*pK_a values using default assignments of the screening functions (see Table 6); calculations were carried out at I = 0.15 M.
#Error = pK(calc) − pK(exp); Experimental values from Antosiewics et al. (1994), Table 5.
§pK_a values using microenvironment-modulated screening function assignments.

The results of the pK calculations for RnaseT1 are presented in Table 12. The titratable moiety of His-92 exists in a fairly hydrophobic microenvironment, which shifts the pK_a by about 0.6 pK units, reducing the error by about half.

Ribonuclease HI

As shown in Table 13, incorporation of the microenvironment improves the pK_a values of Glu-32 and Asp-148. Glu-32 is about 71% buried, but in a fairly hydrophilic environment, with Hpy and *r*Hpy having values of −3.45 and 0.48, respectively. Thus, the menv-free screening of D_2 for the transfer term is changed to D_1, but D_2 is still used for the interaction term. The effect of this change in screening is to decrease both transfer and interaction energies leading to a decrease in $pK_{1/2}$. Similarly, Asp-148 is 97% buried, but with an Hpy of −4.4 is still in a hydrophilic environ-

TABLE 12 pK_a values of RnaseT$_1$

Residue No.	Residue	$pK^0_{1/2}$*	Error#	$pK_{1/2}$§	Error#
27	HIS	7.38	0.08	7.12	−0.18
40	HIS	7.37	−0.53	7.37	0.53
58	GLU	4.75	0.45	4.73	0.43
92	HIS	6.81	−0.99	7.38	−0.42
	rmsd		0.603		0.410

*pK_a values using default assignments of the screening functions (see Table 6); calculations carried out at I = 0.15 M.
#Error = pK(calc) − pK(exp); Experimental values from Inagaki et al. (1981) and Shirley et al. (1989).
§pK_a values using microenvironment-modulated screening function assignments.

TABLE 13 pK_a values* for RnaseHI

Residue No.	Residue	$pK^0_{1/2}$ #	Error§	$pK_{1/2}$ ¶	Error§	
6	GLU	4.31	0.21	4.31	0.21	
10	ASP	5.36	−0.16	5.88	0.36	
32	GLU	4.11	0.61	3.58	0.08	
48	GLU	3.99	−0.21	4.00	−0.20	
57	GLU	2.75	−0.92	2.75	−0.92	
61	GLU	3.35	−0.68	3.62	−0.41	
62	HIS	6.98	−0.02	6.98	−0.02	
64	GLU	4.24	−0.23	4.25	−0.22	
70	ASP	4.31	0.94	4.33	0.96	
83	HIS	5.67	0.17	5.67	0.17	
94	ASP	3.47	0.20	3.49	0.22	
102	ASP	3.24	1.24	3.27	1.27	
108	ASP	3.04	−0.51	3.61 (2.44)	0.06	(−1.11)
114	HIS	5.24	0.24	5.25	0.25	
119	GLU	3.93	−0.54	3.85	−0.62	
124	HIS	5.85	−1.25	5.85 (6.86)	−1.25	(−0.24)
127	HIS	7.56	−0.34	7.56 (6.03)	−0.34	(−1.87)
129	GLU	3.04	−0.66	3.04	−0.66	
131	GLU	4.57	0.11	4.51	0.04	
134	ASP	4.02	−0.10	3.85	−0.27	
135	GLU	4.48	−0.02	4.39	−0.11	
147	GLU	4.39	0.17	4.43	0.20	
148	ASP	4.23	2.23	3.18	1.18	
154	GLU	3.97	−0.38	3.98	−0.37	
155	CTERM	3.74	0.37	3.75	0.38	
	rmsd		0.708		0.574	

*Calculations carried out using the 2RN2 structure (Katayanagi et al., 1992), except for values in parentheses that are from the 1RNH (Yang et al., 1990) structure. For 1RNH, the coordinates of residues 1 and 153–155 were not reported. To carry out the calculation, the missing residues were added to the peptide chain and their conformation modeled by simple minimization.

#pK_a values using default assignments of the screening functions (see Table 6).

§Error = pK(calc) − pK(exp); For the carboxylic acids, Oda et al. (1994) reported pK_a values derived from ^{13}C and proton chemical shifts. The titration curves derived from the proton chemical shifts showed more complicated behavior, which the authors suggested may be due to the susceptibility of the proton shifts to other factors (Oda et al., 1994). For this reason, we have counted the pK_a values determined from the ^{13}C shifts double in calculating the experimental values of the acid pK_a values; for the histidines, only proton chemical shifts were measured (Oda et al., 1993).

¶pK_a values using microenvironment-modulated screening function assignments.

ment, so that the screening used to calculate w^{tr} is set to D_1. The effect of the increased screening is a reduction in w^{tr} leading to a decrease of 1 pK unit in the calculated $pK_{1/2}$. Nevertheless, the remaining error for this residue, as well as Asp-102, is still greater than 1 pK unit. Oda et al. (1994) pointed out that these two residues did not titrate in the pH range of 2 through 8. Both residues form ionic H-bonds with basic residues (Katayanagi et al., 1992). Oda et al. (1994) have also calculated the pK_a values of the acidic groups using the FDPB method (Klapper et al., 1986; Warwicker and Watson, 1982). Their rmsd of about 1.6 was obtained with an internal dielectric constant of 10; their calculation would probably have been considerably improved if they had used an ϵ_i of 20 (Antosiewicz et al., 1996).

The calculated pK_a of His-124 is too low by more than 1 pK unit. However, as Oda et al. (1994) note, comparison of the 2RN2 (Katayanagi et al., 1992) crystal structure (used in the present calculations) and the 1RNH (Yang et al., 1990) crystal structure shows a very large conformational change for His-124 in the two structures. In 2RN2, His-124 is far away from both catalytic residues, Asp-10 and Asp-70, but close to a Lys from a neighboring molecule (Oda et al., 1993), which would decrease the $pK_{1/2}$ of His-124 as found here. In contrast, in 1RNH His-124 is close to Asp-10 and Asp-70, and in this configuration His-124 would be in an acidic environment that would increase its pK_a as observed; this point will be discussed in a following section.

Turkey ovomucoid third domain and HIV protease

As a test of the microenvironment selection criteria developed using the above proteins, the pK_a values of the 15 titratable groups in OMT3 have been calculated. The experimental titrations (Schaller and Robertson, 1995) were carried out at a nominal ionic strength of 10 mM, but the authors point out that, at the end of the titration, the ionic strength is closer to 50 mM. pK_a calculations were done at ionic strengths of 10 mM, 25 mM, and 50 mM. The smallest rmsd was obtained at 25 mM and these results are reported in Table 14. Results calculated at 1 M ionic strength are also included since Schaller and Robertson (1995) also measured the pK_a values at this ionic strength.

The remarkable aspect of this protein is that, in spite of its small size, the pK_a values of four of the carboxy groups are shifted downward by one or more pK units, although all the acidic groups except Asp-27 are solvent exposed. The source of these large pK shifts is attributed to H-bond

TABLE 14 pK_a values of turkey ovomucoid third domain

		I = 0.025M		I = 1M	
Residue No.	Residue	$pK_{1/2}$*	Error#	$pK_{1/2}$*	Error#
1	NTERM	7.64		7.76	
7	ASP	2.87	0.20	3.57	0.58
10	GLU	4.06	−0.04	4.30	−0.03
11	TYR	10.54		11.02	
13	LYS	11.39		10.40	
19	GLU	3.58	0.37	4.14	0.16
20	TYR	9.44		10.01	
21	ARG	12.32		12.08	
27	ASP	3.26	0.98	3.92	1.21
29	LYS	10.63		10.46	
31	TYR	12.50		12.32	
34	LYS	11.63		10.98	
43	GLU	4.42	−0.39	4.42	−0.30
52	HSC	7.38		7.11	
55	LYS	10.83		10.51	
56	CTERM	2.84	0.47	3.09	0.60
	rmsd		0.504		0.617

*pK_a values using microenvironment-modulated screening function assignments.

#Error = pK(calc) − pK(exp); Experimental values from Schaller and Robertson (1995).

formation and other electrostatic interactions (Swint-Kruse and Roberson, 1995). From Table 14 it is seen that the overall rmsd agrees well with the values obtained for the other proteins. Only Asp-27 is in error by a larger value, which is in agreement with the trend found by Antosiewicz et al. (1996). At 1 M ionic strength the rmsd is somewhat greater, but the overall trend to larger $pK_{1/2}$ values as observed from the measurements is preserved. The larger rmsd may be the result of using a simple Debye screening to account for ionic strength. This approximation is probably reasonable in the range up to 200 mM, but for 1 M probably is no longer valid.

As a second test the pK_a values of HIV protease have been calculated, but only the results for the Aspartyl dyad are discussed. The experimental data show a range for both pK_{a1} (4.9–6.8) and pK_{a2} (3.1–3.7) (Hyland et al., 1991; Ido et al., 1991). The calculation was carried out at zero ionic strength and yielded the values 5.45 and 3.32 for pK_{a1} (Asp-25A) and pK_{a2} (Asp-25B), respectively. Thus the calculated values appear to fall within the experimentally determined range, and they are also in reasonable agreement with the recent calculations reported by Luo et al. (1998). The fraction surface area buried in the protein for Asp-25A is 0.88, while for Asp-25B it is 0.85. Thus, as in lysozyme, the separation of the two pK_a values cannot be due to differences in the degree of burial. However, because of small differences in conformation of the two monomers, the microenvironments of the two groups are different with values for Hpy of −1.72 (Asp-25A) and −3.97 (Asp-25B). With these values, w^{int} and Δw^{tr} of both groups are calculated with D_2 (see Table 6). Nevertheless, the lower Hpy value for Asp-25B suggests additional interactions with polar groups leading to enhanced stabilization of the deprotonated form. Finally, in the 1HIV structure, Asp-25B is able to form a better H-bond with Gly-27B [R(OD1 − *N*) = 2.81 Å] than is the case for Asp-25A [R(OD1 − *N*) = 3.23 Å].

Assessment of the results

Table 15 summarizes the reliability of the SCP approach using the description of microenvironments developed in this paper. Incorporation of the modulating effects of the titration site microenvironments on the electrostatic free energy leads to an improvement of about 15% in the rmsd relative to the menv-free calculation for the entire data set of 103 measured pK_a values. It should be realized, however, that with two different screening functions some accounting of environment, based on the buried fraction only, has already been included in the menv-free calculations. Therefore, the rmsd is smaller than that obtained by Antosiewics et al. (1994) from their FDPB calculations, but closer to the overall rmsd obtained by Demchuk and Wade (1996). More significant is the decrease in the larger errors due to inclusion of the environmental effects: at the error threshold of 0.5 pK unit, the decrease is 17% and increases to 33% at a threshold of 1.0 pK unit. A scatter plot of the pK_a values from both menv-free and microenvironment calculations is given in Fig. 5. The regression line has a slope of 0.98 and an intercept of 0.21. The correlation coefficient was 0.99. The scatter plot shows that accounting for the microenvironment induces shifts in the points so that they lie closer to the ideal line $pK_{calc} = pK_{exp}$.

An error analysis according to residue type is given in Table 16. With the present data set the largest errors were found in the pK_a values of Asp and Glu, with the bases having smaller errors; for Tyr and the termini not enough data is available to be statistically meaningful. The error trends from the present calculations are somewhat different from the trends noted in Antosiewicz et al. (1996), where the rmsd from all the residues were about the same.

The mean error of Asp shows that the calculated pK_a are too large. This trend is clearly seen in Fig. 6, where the experimental (Tanford and Roxby, 1972) and calculated titration curves of lysozyme have been plotted. The overestimation of the acidic pK_a values results in the slope of the calculated curve being too steep in the pH region 2–4. In the pH interval 4–6, the titration curve is primarily controlled by His-15 and Glu-35 that both titrate in this interval. The $pK_{1/2}$ points of the calculated titration curves of these two residues (see inset, Fig. 6) are shifted relative to the measured values (Bartik et al., 1994) such that their equilibrium charges approximately cancel in the pH interval 4.5–5.5 and

TABLE 15 Summary of results

Prot (*N*)*	rmsd	Error > 0.5[#]	Error > 0.75[#]	Error > 1.0[#]	Maximum error[§]
BPTI (14)	0.33	1	1	0	0.81
CabD9k (10)	0.39	3	0	0	0.63
Lysozyme (21)	0.49	7	3	1	1.01
ProtG (13)	0.59	5	3	1	1.39
RnaseA (16)	0.55	7	3	1	1.27
RnaseT1 (4)	0.41	1	0	0	0.53
RnaseHI (25)	0.57	8	5	3	1.27
Total (103)	0.50 (0.60)[¶]	32 (40)	15 (20)	6 (9)	1.39 (2.23)

*N = number of experimental values used in the summary.
[#]Number of calculated pK_a values in error by more than the given threshold value.
[§]Magnitude of maximum error.
[¶]Values in parentheses are from menv-free calculations using default assignments of microenvironments (see Table 6).

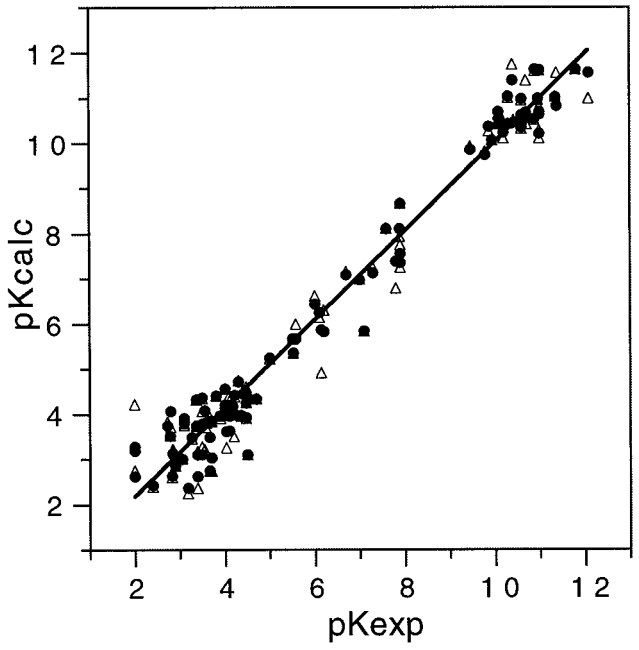


FIGURE 5 Scatter plot of calculated versus experimental pK_a values. *Open triangles*, menv-free results; *closed circles*, complete calculation.

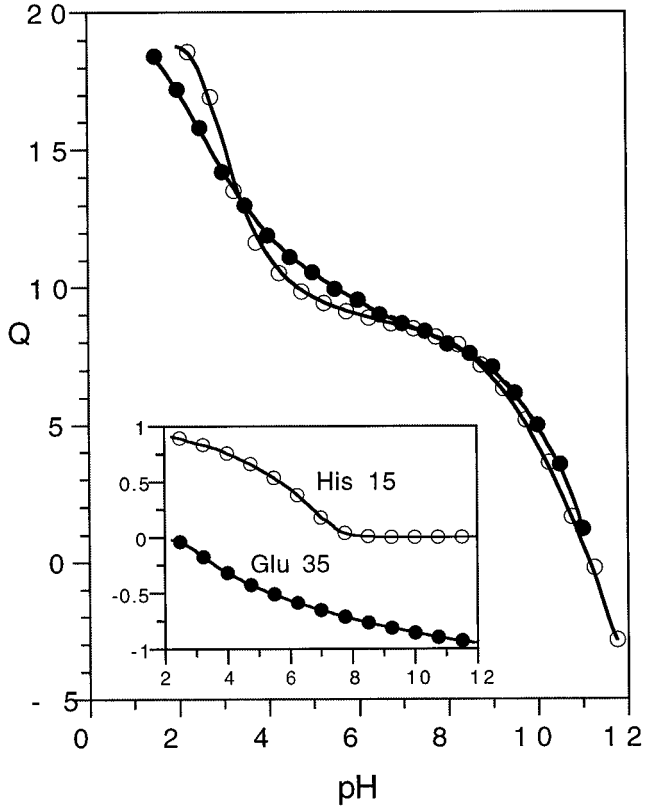


FIGURE 6 Calculated (*open circles*) and experimental (*closed circles*) titration curves of lysozyme. Inset shows calculated titration curves of His-15 and Glu-35.

lead to a flatter slope in this portion of the calculated titration curve. Inspection of the slope of the His-15 experimental titration curve suggests that it is steeper than the calculated slope (Bartik, K., C. Redfield, and C. M. Dobson, private communication). Finally, from a pH of 6 to 11, the experimental and calculated curves are in good agreement.

Limitations of pK_a calculations using a single structure

In their analysis of pK_a values calculated with the FDPB algorithm, Antosiewicz et al. (1996) included average values obtained from NMR structures, and they noted a trend toward improvements in the results. The use of simulation to incorporate protein flexibility (relaxation effects) into the calculation of protein electrostatics has been addressed by Warshel and collaborators using linear response theory (Sham et al., 1997; Sham et al., 1998). Other approaches have been reported (Alexov and Gunner, 1997; Beroza and Case, 1996), and these authors also found some improvements in the calculated values. However, at present not enough systems have been tested to come to any general conclusions. Variability in the conformation of side chains was already pointed out by Bashford and Karplus (1990) and others (Yang et al., 1993; You and Bashford, 1995) in pK calculations using the trigonal and tetragonal crystal forms of HEWL. These differences are already reflected in the buried fractions of the titratable residues and in their microenvironments, which can show variations that, in some cases, are large enough to change the screening assignments for the calculation of the electrostatic free energy components.

Oda et al. (1993) have pointed out the large conformational difference of His-124, which in the 1RNH structure is in an acidic environment (see above) that would lead to an increased $pK_{1/2}$ as observed. They therefore suggest that the conformation of His-124 in the 1RNH structure is probably its conformation in solution. The $pK_{1/2}$ calculated from this structure (see Table 12) reproduces the experimental value with only a small error and supports the contention of Oda et al. (1993). However, the pK_a of His-127 calculated from the 1RNH structure is in error by nearly 2 pK units. Comparison of the two structures shows that His-127 also is in two different conformations. In the 2RN2 structure His-127 is H-bonded to Glu-119 (NE-OE1 distance of 3.4 Å), increasing the pK_a, whereas in the 1RNH structure the separation between these two residues is greater (4.1 Å) leading

TABLE 16 Statistical summary by residue

Residue	No.	Mean error	σ*	rmsd
Asp	26	0.257	0.537	0.586
Glu	27	−0.015	0.578	0.567
His	13	−0.120	0.462	0.460
Lys	22	0.149	0.418	0.425
Tyr	8	−0.088	0.494	0.471
N-term	3	0.496		0.543
C-term	4	0.047		0.212

*Standard deviation.

to a weaker interaction and a lower pK_a. Asp-108 also is in error by more than 1 pK unit when calculated from the 1RNH structure. This error is apparently due to the small separation of OD from the amino nitrogen of Lys-86 (2.4 Å in 1RNH and 3.4 Å in 2RN2). In the 1RNH structure, HBUILD, Brooks et al. (1983) has placed the proton only 1.8 Å from OD, whereas, in 2RN2, the distance is 2.4 Å. It is clear that, due to the strong nonlinearity of the screening functions (Fig. 1) in regions of small *r*, the response of the electrostatic interactions to small changes in distance will be very large.

In Tables 9 and 10 several pK_a are reported that were calculated from alternative crystal structures of HEWL and ProtG, respectively, and differed substantially from the values calculated with the first crystal structure. In essentially all cases, the trends exhibited by these differences are in agreement with results calculated with a variety of theoretical approaches (Antosiewicz et al., 1994; Bashford and Karplus, 1990; Khare et al., 1997; Sham et al., 1997; Yang et al., 1993; You and Bashford, 1995). Moreover, comparison of the crystal structures shows that, as in the case of Rnase HI, the differences in the pK_a values have their origin in conformational variations of the residues in question as discussed by the above authors.

CONCLUSIONS

In this paper a novel approach has been proposed for calculating the modulating effects that the local environment has on the electrostatic free energy controlling pH-dependent properties. This approach, combined with a modified variational procedure for calculating the titration charge distribution has been applied to calculating the pK_a values in seven proteins with a combined data set of over 100 measured values. The method has yielded results that are among the most reliable obtained to date while conserving the computational economy of the original algorithm (Mehler, 1996b). The overall rmsd are small, and larger shifts are correctly reproduced in a majority of cases. Thus, for the first time, the measured upward pK_a shift of Glu-35 in lysozyme has been correctly calculated while, at the same time, the pK_a values of the other titratable groups were also calculated with acceptable errors so that the overall rmsd was less than 0.5.

The quantification of the microenvironments of titratable groups reveals a complex mosaic of dielectric regions that are usually more hydrophobic than water, but with an ambient level that may be less hydrophobic than expected. In this background, pockets of extreme hydrophobicity or hydrophilicity are easy to identify; more difficult are the regions that differ less from the average, where the modulating effect on the interaction and transfer free energies is less clear to specify. Nevertheless, using local Hpy in the manner proposed here introduces a description that is based on a component of the physical chemistry of the system that is independent of the electrostatic component, but that clearly affects it. Thus, the hydrophobicity of the microenvironment is an attribute of the macromolecule that can be used to improve the electrostatic description at the given site, and also to exhibit a potential source of abnormal behavior leading to large pK shifts. Moreover, because of the generality and importance of local hydrophobicity or hydrophilicity, it is expected that the microenvironment will also be useful in defining implicit solvent models for use with simulations on biological macromolecules (Guarnieri et al., 1998; Hassan et al., 1999).

It is rather remarkable that the Rekker fragmental hydrophobic constants (Rekker, 1977) used to quantify the microenvironments in the approach presented here work so well. This is gratifying because these quantities were derived from water/octanol partition functions of a large number of organic and pharmaceutically important molecules that, for the most part, have little in common with amino acids or protein structure. It clearly demonstrates the robustness of the concept of hydrophobicity.

Although the results presented in this paper show that the microenvironments are important for controlling properties, as already recognized earlier by Ponnuswamy et al. (1980), the remaining errors in the calculations clearly show that substantial room for improvement remains. These are many and only a few are mentioned here. First, in the algorithm developed here, the microenvironment was defined in terms of atoms, not whole functional groups. In some cases this meant that the value of the fragmental constant had to be distributed over the constituent atoms of the functional groups, and this necessarily was to some extent arbitrary. Second, it would be more correct to include the Hpy values of both neutral and charged forms of the titratable groups. Third, the values used in the present calculations are from the original publication (Rekker, 1977), and these have been updated and extended, and finally the other fragmental schemes (Ghose and Crippen, 1986; Leo et al., 1975; Suzuki and Kudo, 1990) should also be tested for their effectiveness in describing the properties of the microenvironments.

It is of interest that this approach led to great enough improvement in reliability to help identify conformations of side chains that might be less relevant for the solution structure, e.g., His-124 in RnHI (see also Oda et al., 1993), Glu-7 in HEWL, or Glu-15 in ProtG. In other cases, larger errors remained in the pK_a values calculated from both crystal structures, e.g., Glu-27 in ProtG or Asp-102 and Asp-148 in RnHI and show that the approach still needs considerable further improvement.

APPENDIX

Sigmoidal screening functions

A sigmoidal screening function, D(*r*), can be described by a differential equation of the form (Mehler and Eichele, 1984)

$$dD(r)/dr = \lambda(D_o + D)(D_s - D) \qquad \text{(A1a)}$$

with the solution

$$D(r) = B/[1 + k\exp(-\lambda Br)] - D_o, \qquad \text{(A1b)}$$

where $B = D_s + D_0$, k is a constant of integration, D_0 and λ are parameters, and D_s is the dielectric constant of water with a value of 78.4. For the present applications, it is convenient to set $D(0) = 1$ by taking $k = (D_s - 1)/(D_0 + 1)$. Then the two parameters, D_0 and λ, define the explicit form of the solutions of Eq. A1a.

The screening function in water, D_w, can be obtained from the analytic form of the radial permitivity (Ehrenson, 1989) in the following way: In Guarnieri et al. (1998), it was shown that for a charge q_a and Born radius, R_a, the self energy is given by

$$2w^{\text{self}} = q_a^2 \int_{R_a}^{\infty} \frac{dr}{\epsilon(r) r^2} = \frac{q_a^2}{R_a D(R_a)}. \tag{A2a}$$

Introducing the change in variable $u = r/R_a$, Eq. A2a becomes

$$2w^{\text{self}} = \frac{q_a^2}{R_a} \int_1^{\infty} \frac{du}{\epsilon(uR_a) u^2}, \tag{A2b}$$

so that the integral in Eq. A2b gives $1/D(R_a)$ directly. The integral is evaluated numerically and the results are fitted to Eq. A1b. Here the radial permittivity obtained from LDS theory without the reaction field correction (Ehrenson, 1989) has been used, with the ionic partial charge set to a value of 0.55, which is representative of the magnitude of the average partial ionization charge on the atoms of the protonatable groups [cf. CHARMM, PAR19 (Brooks et al., 1983)]. To fit D_w all four parameters, D_0, D_s, λ, and k were used and resulted in excellent agreement with the analytic result obtained from LDS theory, except in the region 4.5–7 Å where the fitted curve disagreed with the values calculated from Eq. A2b by 1–2 units (out of around 70). However, in the present calculations D_w is only used to calculate the self energy of the charges in water, and Born radii greater than 2.6 Å do not occur.

Modified self-consistent variational procedure

In Mehler (1996b), the distribution of the ionization charge was written in the form

$$q_A = \sum_a f_a q_a^0 \tag{A3}$$

where the q_a^0 are fixed initial partial charges ($\neq 0$, see below), and the f_a are scaling factors that are subject to variation to determine a stationary point of w. For convenience, the group subscript A is omitted from Eq. A3 and will be dropped in all further equations. It is assumed that "a" refers to the atoms in group A, "b", the atoms in group B, etc. To allow the total ionization charge to be coupled to the Henderson–Hasselbalch equation, the variations of the scaling factors were constrained so that the total ionization charge remained constant at a given pH. This constrained variational procedure yielded a set of self-consistent equations for the f_a, but with the disadvantage that, because of the way the transfer energy component was expressed, it dropped out of the variational equations and had to be included in an ad hoc way [see Eq. 15 and the discussion following in Mehler (1996b) for details]. A second disadvantage was that the neutral forms of the titratable residues could not be included simultaneously with the charged form as the residue passed through its titration point.

Here, both of these disadvantages are resolved in the following way. First, the partial charge of an atom a in group A is defined as

$$q_a = (1 - \theta_A) q_a^n + f_a q_a^0, \tag{A4}$$

where q_a^n is the partial charge from the neutral group, and θ_A is the fraction of A in the charged state (note that, for neutral groups, θ_A is zero because $q_a^0 = 0$ for all a, and since $\Sigma q_a^n = 0$, Eq. A3 is still valid. Thus the first term on the r.h.s. of Eq. A4 is the contribution to the charge q_a from the neutral form of the protonatable group, while the second term represents the contribution from the ionization charge. The electrostatic interaction free energy of the ionization charge of atom a in group A is

$$w_a^{\text{int}} = \tfrac{1}{2} f_a q_a^0 \Phi_a(\mathbf{r}_a) \tag{A5}$$

$$\Phi_a(\mathbf{r}_a) = \sum_{B,b} \frac{(1 - \theta_B) q_b^n + f_b q_b^0}{D(r_{ab}) r_{ab}}, \quad (B \neq A)$$

where $\Phi_a(\mathbf{r}_a)$ is the potential at $\mathbf{r}_a$ from all the other groups in the system, and $D(r)$ is the radially dependent screening function defined by Eqs. A1a and A1b. It is seen that, as in I, w_a^{int} only includes the contributions to the total electrostatic free energy arising from the ionization charge (see below). It is assumed that all intraresidue interactions are included in the pK_a(s) values so that none of these terms are included in w_a^{int}, even when A is a subgroup of the titratable residue.

The transfer energy term is expressed as the difference of two self-energy terms derivable from the integral form of the Born equation (Guarnieri et al., 1998)

$$\Delta w_a^{\text{tr}} = (f_a q_a^0)^2 \tfrac{1}{2} \left\{ \frac{1}{D_p(R_a^p) R_a^p} - \frac{1}{D_s(R_a^s) R_a^s} \right\} (BF)_a. \tag{A6}$$

In the following, a subscript on D is used to denote the dielectric medium; also, where the radial dependence is denoted in lower case, i.e., $D(r_{ab})$, it indicates a distance between the two charges q_a and q_b, whereas an upper case symbol refers to a (Born) radius, and a superscript, e.g., R^p, is used to denote the medium where p and s refer to protein and model solvent, respectively, and $(BF)_a$ is the fraction of the surface of atom a buried in the protein.

The variational problem is expressed in the form

$$\delta w = \sum_{A,a} \delta w_a^{\text{int}} + \delta \Delta w_a^{\text{tr}} = 0, \tag{A7}$$

with the constraints that q_A remain constant at a given pH, i.e.,

$$\delta q_A = \sum_a q_a^0 \delta f_a = 0. \tag{A8}$$

Incorporating the constraints into the variational equations with the help of Lagrangian multipliers yields

$$\delta w = \sum_{A,a} \left(\frac{\partial w_a^{\text{int}}}{\partial f_a} + \frac{\partial \Delta w_a^{\text{tr}}}{\partial f_a} - \lambda_A q_a^0 \right) \delta f_a = 0. \tag{A9}$$

Carrying out the operations with the help of Eqs. A5 and A6 yields

$$\sum_{A,a} \left(\frac{w_a^{\text{int}}}{f_a} + \frac{2 \Delta w_a^{\text{tr}}}{f_a} - \lambda_A q_a^0 \right) \delta f_a = 0. \tag{A10}$$

Since the δf_a constitute arbitrary variations, it is necessary that

$$w_a^{\text{int}} + 2 \Delta w_a^{\text{tr}} = \lambda_A f_a q_a^0. \tag{A11}$$

To evaluate the Lagrangian multipliers both sides of Eq. A11 are summed over a to yield

$$\lambda_A = \frac{1}{q_A} (w_A^{\text{int}} + 2 \Delta w_A^{\text{tr}}), \tag{A12}$$

and substituting into Eq. A11 yields

$$f_a q_a^0 = q_A \frac{w_a^{\text{int}} + 2 \Delta w_a^{\text{tr}}}{w_A^{\text{int}} + 2 \Delta w_A^{\text{tr}}} \quad \text{(for all titratable groups, A)}. \tag{A13}$$

Equations A13 state that the optimal distribution of the ionization charge on any atom is given by the equilibrium ionization charge scaled by a factor related to the fractional energy contribution from the charge ($f_a^o q_a^0$) on atom a to group A. Equations A13 must be solved iteratively because the energy terms are dependent on the scaling factors. Thus, the procedure starts with an initial guess of the ionization charges ($f_a^o q_a^o$) and their distribution over the atoms of the titratable groups. At each step, a new set of scaling factors is obtained until, at the nth step, $f_a^n = f_a^{n-1}$, for all a, to some preselected threshold, at which point the system of Eqs. A13 have been solved and are said to be self-consistent. It is important to realize that only at self-consistency does a solution exist, because only then are the Eqs. A13 valid. The initial guess, $f_a^0 q_a^0$, generally does not solve Eqs. A13 and, therefore, does *not* constitute an initial charge assignment from which the charges redistribute. It is also noted that inclusion of the first term on the r.h.s. of Eq. A4 in w_a^{int} (Eq. A5) has no effect on Eqs. A13 because of the constraints (Eqs. A8). Thus, any term involving θ_A (or equivalently, q_A) drops out of Eqs. A13.

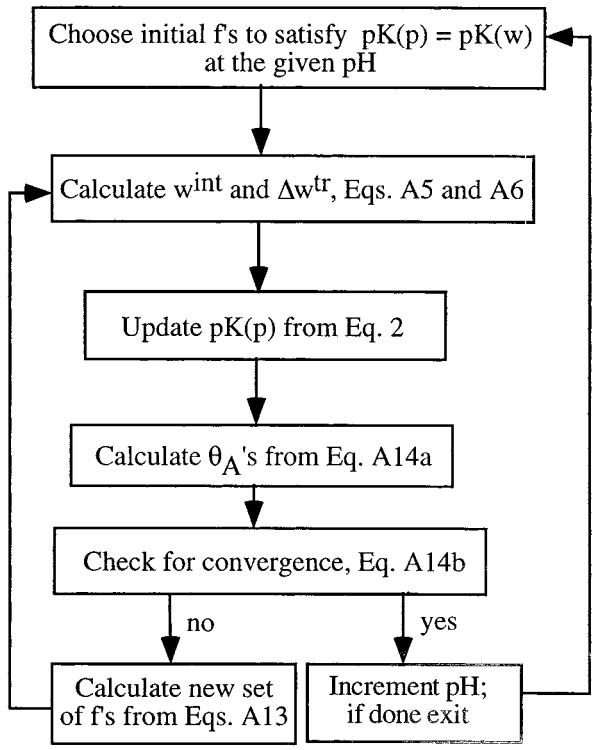
The pH dependence of the equilibrium charge state is coupled to the variational Eqs. A13 by the relation

$$\theta_A = \begin{cases} \left(1 + \dfrac{[H^+]}{K_a^A(p)}\right)^{-1} & \text{for acids} \\ \left(1 + \dfrac{K_a^A(p)}{[H^+]}\right)^{-1} & \text{for bases} \end{cases} \qquad \text{(A14a)}$$

and requiring that at self consistency

$$q_A = \sum_a q_a^0 f_a = Z_A \theta_A, \qquad \text{(A14b)}$$

where Z_A is the formal charge on group A. It was found that Eq. A14b is sufficient to ensure convergence and is used instead of testing all the scaling factors. The method of solution is outlined in scheme A1 (for additional details, see Mehler, 1996b). It should be noted that coupling of the variational procedure to Eqs. A14 is a matter of convenience. Coupling to a more accurate statistical mechanical treatment (Beroza et al., 1991) would also be possible. However, given the accuracy of the present results, it seems unlikely that a more accurate treatment of the statistical problem would yield enough improvement to justify the additional computational cost. This is especially so, since, as we discuss below, the method apparently resolves the problems associated with strongly interacting charges in the mean field approximation (Bashford and Karplus, 1991).



SCHEME A1 Flow diagram for calculating self-consistent ionization charge scaling factors.

Discussion

By coupling the variational Eqs. A13 to Eqs. A14 it is not necessary to evaluate the 2^N microstates explicitly. The nature of the approximations made in this approach can be understood by using the relation $\partial/\partial f_a = (\partial/\partial q_A)(\partial q_A/\partial f_a)$, and Eq. A3. Substituting into Eq. A9 and reducing to a single titration site ($a = 1$) one easily obtains the relation

$$w_A^{int} + 2\Delta w_A^{tr} - \lambda_A q_A = 0. \qquad \text{(A15)}$$

Equation A15 is analogous to Eq. 16 in Bashford and Karplus (1991), and by an appropriate choice of λ_A the mean field approximation (Tanford and Roxby, 1972) could be recovered. In the present approach this is not done; rather, by using Eq. A12 the Lagrangian multipliers are incorporated into the variational equations and they never need to be evaluated explicitly. This way of handling Lagrangian multipliers is similar to the coupling or screening operator techniques developed for applying constrained variational methods in quantum mechanics (Birss and Fraga, 1963; Mehler, 1977). It is also seen from summing both sides of Eq. A13 over a that the constraints (Eqs. A8) are automatically enforced, which suggests that they are essential for preventing excessive or unphysical scaling factors.

In the applications presented here and elsewhere (Luo et al., 1999) pathological scaling factors have not been observed. Table A1 lists the ionization charge assignments for several typical cases in lysozyme at different pH values. In many cases the entire titration charge is assigned to a single atom as shown by Asp-101. This type of situation seems to be closest to the mean field approximation where the charge is usually assigned to a single titration site (Bashford and Karplus, 1991; Beroza et al., 1991; Tanford, 1962). Note however, that, because the electrostatic field changes with changes in pH, the location of the charge can shift as shown by Arg-45. In contrast, the ionization charge in Asp-66 is distributed over the carboxy moiety at pH 4 and 8 and on OD1 at a pH of 12. Such distributions appear to be more typical for cases where the titratable groups can interact strongly with nearby, ionizable residues. As seen from Fig. 2, Asp-66 interacts with titratable groups Arg-68 and Tyr-53, and with several polar groups. The phenol group of Tyr forms an H-bond with OD2 leading to favorable interactions at pH below the $pK_{1/2}$ of Tyr-53. However, when Tyr-53 deprotonates a strong repulsive interaction would develop between the phenoxylate oxygen and OD2. This is compensated by the Asp-66

TABLE A1 Assignment of equilibrium titration charge of titratable groups in HEWL

Residue	Atom	q (pH = 4)	q (pH = 8)	q (pH = 12)
Asp-66	CG	−0.10	−0.10	0.0
	OD1	−0.48	−0.49	−1.0
	OD2	−0.42	−0.41	0.0
Asp-101	CB	−0.43	−1.00	−1.00
Arg-45	HH11	0.0	0.0	0.54
	NH2	1.0	1.0	0.0
Arg-68	CZ	1.0	1.0	0.03
	NH2	0.0	0.0	0.14
	HH21	0.0	0.0	0.72

titration charge moving to OD1. At lower pH, OD1 and OD2 can also interact favorable with the titration charge on CZ of Arg-68.

The ionization charge distributions of Arg-45 and Arg-68 have been included to illustrate how the variational procedure distributes charge to minimize unfavorable interactions. Beroza et al. (1991) reported some problems with this pair, apparently because the fixed assignment of the titration charge to a single site led to very strong interactions between them that required special handling of the Monte Carlo procedure and gave an incorrectly depressed pK_a for one of the arginines. Here, no such problems arose, and, as indicated by the results given in Table A1, the ionization charge distributions are quite reasonable; moreover, the pK_a for these two groups were between 12 and 12.5. At pH 4 and 8, the titration charge is assigned to NH2 and CZ of Arg-45 and Arg-68, respectively. These two atoms are about 7 Å apart, so that their interaction is fairly small. Moreover, CZ of Arg-68 is about 4 Å from OD1 and OD2 of Asp-66, which yields favorable interactions. At pH 12, both arginines are titrating. The charge on Arg-45 shifts to HH11, which is about 4 Å from HH21 of Arg-68. However, HH21 is about 4 Å from OD1 of Asp-66, with which it can interact favorably.

The results given in Table A1 show that the variational approach distributes the ionization charge in a way that is in reasonable accord with the physical situation. However, it is also clear that the physical significance of the individual values should not be overinterpreted because the constrained variational procedure formulated here does not require that the individual partial charges can be used in any other context. It is only required that the partial charges optimize w, and that their sum satisfy the constraints. This limitation should not be surprising. Partial charges can, at best, have only limited physical significance because they lack a rigorous definition as physical observables, and therefore, the associated quantum mechanical operator does not exist. Hence, a well defined expectation value cannot be calculated. Because of this, a given set of partial charges, determined by fitting to a selected set of physical data, will not perform with uniform reliability when applied to different systems (Mehler and Solmajer, 1991), nor can it be a priori assumed that a given set of charges will necessarily give reliable results when applied outside the context of its definition.

The authors would like to thank Dr. Christina Redfield for providing unpublished data from the NMR titration of HEWL and for critical discussions concerning the titration of HEWL. Computational support was provided by the Pittsburgh Supercomputer Center (sponsored by the National Science Foundation), the Cornell National Supercomputer Facility, the Advanced Scientific Computing Laboratory at the Frederick Cancer Research Facility of the National Cancer Institute (Laboratory for Mathematical Biology), and the University Computer Center of the City University of New York. Partial support of the work by NSF grant DBI9732684 is gratefully acknowledged.

REFERENCES

Alexov, E. G., and M. R. Gunner. 1997. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* 74:2075–2093.

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

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pKas in proteins. *Biochemistry.* 35:7819–7833.

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

Bashford, D., and M. Karplus. 1990. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochem.* 29: 10219–10225.

Bashford, D., and M. Karplus. 1991. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* 95:9556–9561.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rogers, D. Kennard, T. Simanouchi, and M. Tatsumi. 1977. The protein data bank: a computer-based archival file for macromolecular structure. *J. Mol. Biol.* 112:535–542.

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

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

Birss, F. W., and S. Fraga. 1963. Self-consistent-field theory. I. General treatment. *J. Chem. Phys.* 38:2552–2557.

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

Born, M. 1920. Volumen und Hydrationswärme der Ionen. *Z. Phys.* 1:45–48.

Böttcher, C. J. F. 1938. The dielectric constant of dipole liquids. *Physica.* 5:635–639.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

Bucher, M., and T. L. Porter. 1986. Analysis of the Born model for hydration of ions. *J. Phys. Chem.* 90:3406–3411.

Collura, V. P., P. J. Greaney, and B. Robson. 1994. A method for rapidly assessing and refining simple solvent treatments in molecular modelling. Example studies on the antigen-combining loop H2 from FAB fragment McP603. *Prot. Eng.* 7:221–233.

Conway, B. E., J. O. M. Bockris, and I. A. Ammar. 1951. The dielectric constant of the solution in the diffuse and Helmholtz double layers at a charged interface in aqueous solution. *Trans. Faraday Soc.* 47:756–767.

Cornette, J. L., K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky, and C. DeLisi. 1987. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.* 195: 659–685.

Debye, P. 1929. Polar Molecules, Dover, New York.

Debye, P., and L. Pauling. 1925. The inter-ionic attraction theory of ionized solutes. IV. The influence of variation of dielectric constant on the limiting law for small concentrations. *J. Am. Chem. Soc.* 47: 2129–2134.

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

Ehrenson, S. 1989. Continuum radial dielectric functions for ion and dipole solution systems. *J. Comp. Chem.* 10:77–93.

Fukamizo, T., T. Torikata, T. Nagayama, T. Minematsu, and K. Hayashi. 1983. Enzymatic activity of avian egg-white lysozymes. *J. Biochem. (Tokyo).* 94:115–122.

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

Ghose, A. K., and G. M. Crippen. 1986. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships I. Partition coefficients as a measure of hydrophobicity. *J. Comp. Chem.* 7:565–577.

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

Guarnieri, F., A. B. Schmidt, and E. L. Mehler. 1998. A screened Coulomb potential based implicit solvent model: formulation and parameter development. *Int. J. Quantum Chem.* 69:57–64.

Harvey, S. C., and P. Hoekstra. 1972. Dielectric relaxation spectra of water adsorbed on lysozyme. *J. Phys. Chem.* 76:2987–2994.

Hassan, S. A., F. Guarnieri, and E. Mehler. 1999. A screened Coulomb potential based implicit solvent model: parametrization and prediction of structures of small peptides. *Biophys. J.* 76:A198.

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

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

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

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

Juffer, A. H., P. Argos, and H. J. Vogel. 1997. Calculating acid-dossociation constants of proteins using the boundary element method. *J. Phys. Chem.* 101:7664–7673.

Katayanagi, K., M. Miyagawa, M. Matsushima, M. Ishikawa, S. Kanaya, H. Nakamura, M. Ikehara, T. Matsuzaki, and K. Morikawa. 1992. Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* 223:1029–1052.

Kesvatera, T., B. Jonsson, E. Thulin, and S. Linse. 1996. Measurement and modelling of sequence-specific pKa values of lysine residues in calbindin D_{9k}. *J. Mol. Biol.* 259:828–839.

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

King, G., F. S. Lee, and A. Warshel. 1991. Microscopic simulations of macroscopic dielectric constants of solvated proteins. *J. Chem. Phys.* 91:3647–3661.

Klapper, I., R. Hagstrom, R. Fine, K. Sharp, and B. Honig. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins Struct. Func. Genet.* 1:47–59.

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

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

Leo, A., P. Y. C. Jow, C. Silipo, and C. Hansch. 1975. Calculation of hydrophobic constant (log P) from π and f-constants. *J. Med. Chem.* 18:865–868.

Lorentz, H. A. 1952. Theory of Electrons. Dover, New York.

Luo, N., E. Mehler, and R. Osman. 1999. Specificity and catalysis of uracil DNA glycosylase. A molecular dynamics study of reactant and product complexes with DNA. *Biochemistry*. In press.

Luo, R., M. S. Head, J. Moult, and M. K. Gilson. 1998. pKa shifts in small molecules and HIV protease: electrostatics and conformation. *J. Amer. Chem. Soc.* 120:6138–6146.

MacKerell, A. D. J., D. Bashfod, M. Bellot, R. L. J. Dunbrack, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, J. Wiorkiewicz-Kuczera, and M. Karplus. 1992. Self-consistent parametrization of bio molecules for molecular modeling and condensed phase simulations. *Biophys. J.* 6:A143.

MacKerell, A. D. J., J. Wiorkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical Emergy function for the simulation of nucleic acids. *J. Amer. Chem. Soc.* 117:11946–11975.

Mannhold, R., R. F. Rekker, C. Sonntag, A. M. ter Laak, K. Dross, and E. E. Polymeropouos. 1995. Comparative evaluation of the predictive power of calculation procedures for molecular lipophilicity. *J. Pharm. Sci.* 84:1410–1419.

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

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

Mehler, E. L. 1977. Self-consistent nonorthogonal group function approximation for polyatomic systems. I. Closed shells. *J. Chem. Phys.* 67: 2728–2739.

Mehler, E. L. 1996a. The Lorentz–Debye–Sack theory and dielectric screening of electrostatic effects in proteins and nucleic acids. *In* Molecular Electrostatic Potential: Concepts and Applications, J. S. Murray and K. Sen, editors. Elsevier Science, Amsterdam, The Netherlands. 371–405.

Mehler, E. L. 1996b. A self-consistent, free energy based approximation to calculate pH dependent electrostatic effects in proteins. *J. Phys. Chem.* 100:16006–16018.

Mehler, E. L., and E. Eichele. 1984. Electrostatic effects in water-accessible regions of proteins. *Biochemistry.* 23:3887–3891.

Mehler, E. L., and T. J. Solmajer. 1991. Electrostatic effects in proteins: comparison of dielectric and charge models. *Prot. Eng.* 4:903–910.

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

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

Onsager, L. 1936. Electric moments of molecules in liquids. *J. Amer. Chem. Soc.* 58:1486–1493.

Pennock, B. D., and H. P. Schwan. 1969. Further observations on the electrical properties of hemoglobin-bound water. *J. Phys. Chem.* 73: 2600–2610.

Ponnuswamy, P. K., M. Prabhakaran, and P. Manavalan. 1980. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. *Biochim. Biophys. Acta.* 623:301–316.

Ramanadham, M., L. C. Sieker, and L. H. Jensen. 1981. Structure of triclinic lysozyme and it Cu(2+) complex at 2 angstroms resolution. *Acta Crys. A.* 37:33.

Rekker, R. F. 1977. The Hydrophobic Fragmental Constant. Elsevier, Amsterdam, The Netherlands.

Rekker, R. F. 1979. The hydrophobic fragmental constant: an extension to a 1000 data point set. *Eur. J. Med. Chem.* 14:479–488.

Sack, V. H. 1926. The dielectric constant of electrolytes. *Phys. Z.* 27: 206–208.

Sack, V. H. 1927. The dielectric constants of solutions of electrolytes at small concentrations. *Phys. Z.* 28:199–210.

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

Schwarzenbach, G. 1936. Der Einfluss einer Ionenladung auf die Acidität einer Säure. *Z. Physik. Chem. A.* 176:133–153.

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

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

Shirley, B. A., P. Stanssen, J. Steyaert, and C. N. Pace. 1989. Conformational stability and activity of ribonuclease T1 and mutants. *J. Biol. Chem.* 264:11621–11625.

Suzuki, T., and Y. Kudo. 1990. Automatic log P estimation based on combined additive modeling methods. *J. Comp.-Aided Mol. Design.* 4:155–198.

Svensson, B., B. Jönsson, and C. Woodward. 1990. Electrostatic contributions to the binding of Ca^{2+} in calbindin mutants: a Monte Carlo study. *Biophys. Chem.* 38:179–183.

Svensson, B., B. Jönsson, C. E. Woodward, and S. Linse. 1991. Ion-binding properties of calbindin D_{9k}: a Monte Carlo simulation study. *Biochem.* 30:5209–5217.

Swint-Kruse, L., and A. D. Roberson. 1995. Hydrogen bonds and the pH dependence of ovomucoid third domain stability. *Biochemistry.* 35: 4724–4732.

Szebenyi, D. M. E., and K. Moffat. 1986. The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. *J. Biol. Chem.* 261:8761–8777.

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

Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* 84: 4240–4247.

Tanford, C., and R. Roxby. 1972. Interpretation of protein titration curves. Application to lysozyme. *Biochemistry.* 11:2192–2198.

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

Warshel, A. 1981. Calculations of enzymatic reactions: calculations of Pk_a, proton transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry.* 20:3167–3177.

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

Webb, T. J. 1926. The free energy of hydration of ions and the electrostriction of the solvent. *J. Am. Chem. Soc.* 48:2589–2603.

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

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

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

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