pK_a's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model†

Donald Bashford and Martin Karplus*
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138
Received December 28, 1989; Revised Manuscript Received April 30, 1990

ABSTRACT: A macroscopic electrostatic model is used to calculate the pK_a values of the titratable groups in lysozyme. The model makes use of detailed structural information and treats solvation self-energies and interactions arising from permanent partial charges and titratable charges. Both the tetragonal and triclinic crystal structures are analyzed. Half of the experimentally observed pK_a shifts (11 out of 21) are well reproduced by calculations for both structures; this includes the unusually high pK_a of Glu 35 in the active site. For more than half the pK_a's (13 out of 21), there is a large difference (1–3.3 pK units) between the results from the two structures. Many of these correspond to the titrating groups for which the calculations are in error. Since for an ionic strength of 0.1 M the Debye screening between titratable groups leads to a very high effective dielectric constant (the average value for all pairs of titrating groups is approximately 900), near-neighbor interactions dominate the pK_a perturbations. Thus, the pK_a values are very sensitive to the details of the local protein conformation, and it is likely that side-chain mobility has an important role in determining the observed pK_a shifts.

The pK_a values of titrating groups in proteins play an essential role in their stability and function (Tanford, 1961; Schulz & Schirmer, 1979; Fersht, 1985). For enzymes, in particular, the experimental determination of pK_a's and their relation to the pH dependence of the kinetic partameters is a very active field of research (Fersht, 1985; Knowles, 1976). To fully exploit such data, it is necessary to have an understanding of the electrostatic interactions in proteins including theoretical methods for the calculation and interpretation of pK_a values. Macroscopic continuum dielectric models in which the protein is represented as a low-dielectric sphere in an aqueous high-dielectric medium (Tanford & Kirkwood, 1957; Shire et al., 1974; States & Karplus, 1987) have been employed for this purpose. However, the accuracy of the results was limited (Tanford & Roxby, 1972), and intrinsic pK_a's had to be adjusted to fit the data (Orttung, 1970; States & Karplus, 1987). Recent applications of this type of approach, including an empirical correction for the exposed surface area, are reviewed in Matthew and Gurd (1986). The introduction of finite difference techniques (Warwicker & Watson, 1982) has made possible the use of a more realistic representation of the protein–solvent interface in macroscopic models, and excellent results have been obtained in the prediction of pK_a shifts induced by charge perturbations (Sternberg et al., 1987; Gilson & Honig, 1987, 1988; Bashford & Karplus, 1988).

In this paper, we use a continuum dielectric model and finite difference technique to calculate the pK_a values of the titrating groups of lysozyme. The present calculation is of more general utility and provides a more severe challenge for the continuum model than the charge perturbation analyses. It includes a series of relatively large, approximately canceling terms due to the interaction of charges with the polarization that they induce in the medium, which gives rise to the Born solvation energy (Kortum & Bockris, 1951) and couplings with many nontitrating and partial charges in the protein, as well as the coupling between titrating sites. The previous charge perturbation analyses included only interactions between titrating sites and charges that were perturbed by site-directed mutagenesis (Sternberg et al., 1987; Gilson & Honig, 1987, 1988) or oxidation–reduction (Bashford & Karplus, 1988). Unlike the earlier models (Orttung, 1970; States & Karplus, 1987), the present calculations require no empirical adjustment of intrinsic pK_a's. Moreover, the mean group charge approximation normally used for pK_a estimates (Tanford & Roxby, 1972) is replaced by an accurate approximation for the partition function summed over all titration states.

The polarizability of both the solvent and the protein has an important influence on the pK_a's of ionizable groups. Theoretical methods, in which atoms are assigned point polarizabilities to represent their electronic polarizability, and Monte-Carlo, molecular dynamics, or Langevan calculations used to account for orientational polarizability have been described by Russell and Warshel (1985), Van Belle et al. (1987), and Rullmann et al. (1989). In the present calculations polarization effects, both electronic and orientational, are represented by the continuum dielectric constants. In the aqueous solvent, the chosen experimental value of 80 is dominated by the orientational contribution. For the protein, fixed atomic charges are used for polar groups. Their contribution to the orientational polarizability results from fluctuations about their equilibrium positions and is represented by the use of a protein dielectric constant of 4 (Harvey, 1989), somewhat larger than the electronic polarizability estimate of 2. The assumption of a uniform dielectric response throughout the protein is clearly an approximation (Gilson & Honig, 1986; Nakamura et al., 1988). However, it has the advantage of being well-defined and simple. Most important, it makes possible a self-consistent calculation of the shifts in pK_a values due to desolvation of buried charged sites (loss of Born self-energy) and their interactions with polar and charged groups in the protein.

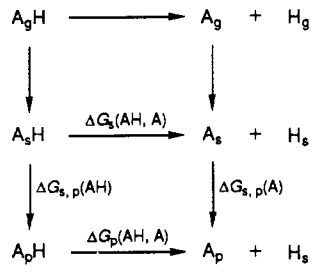
Methods

If A is a model compound (e.g., a lysine side chain in a blocked peptide) with one titratable group analogous to a site in the protein (e.g., a lysine side chain), the titration behavior can be usefully considered in terms of the thermodynamic

† This work was supported in part by grants from the National Science Foundation and the National Institutes of Health.

0006-2960/90/0429-10219$02.50/0

Scheme I



cycles (Scheme I), where A_g (A_gH), A_s (A_sH), and A_p (A_pH) represent the deprotonated (protonated) site in the gas phase, solvent phase, and protein environment, respectively, and H_g and H_s represent the proton in the gas phase and in solution, respectively. By using model compounds whose pK_a's in solution are known, we avoid the need to refer the results to the gas phase and can restrict the calculations to the lower cycle. It corresponds to

$$\begin{aligned} pK_a(\text{protein}) &= pK_a(\text{model}) + \\ &\quad \frac{1}{2.303k_BT}[\Delta G_{s,p}(A) - \Delta G_{s,p}(AH)] \\ &= pK_a(\text{model}) + \frac{1}{2.303k_BT}[\Delta G_p(AH,A) - \\ &\quad \Delta G_s(AH,A)] \end{aligned} \tag{1}$$

where the free energy differences used in eq 1 are defined in Scheme I. In what follows, pK_a(model) is obtained from experiment, and the second equality in eq 1 is used for the calculations because, as in free energy simulations (Jorgensen, 1989), it is simplest to determine the difference between structures that are most similar (i.e., AH vs A in the protein or model compound rather than AH or A in the protein vs the model). Figure 1 shows the macroscopic idealizations of the protein and model compounds that are used. The protein (or model compound) is a rigid object with a dielectric constant (Tanford & Roxby, 1972) ϵ_m equal to 4 and contains sites i, whose charging or discharging represents protonation or deprotonation (solid circles), and fixed charges k, representing the molecule's permanent charges (open circles). The solvent has a dielectric constant $\epsilon_s = 80$ and, outside the ion-exclusion layer, an ionic strength of 0.1 M. The boundary between the molecule and solvent is defined by the contact and reentrant surfaces (Richards, 1977) of a 1.4-Å probe rolling on the van der Waals surface of the molecule. The ion-exclusion layer accounts for the finite size of the ions. Following Gilson and Honig (1988), we have chosen a 2.0 Å thick ion-exclusion layer. The coordinates for the model compounds, the *N*-formyl *N*-methylamide derivatives of the amino acids, are derived from the protein crystallographic coordinates. The coordinates of the atoms of the residue containing site i, along with the peptide C and O of the previous residue and the N, H, and C_α of the following residue, are taken as the coordinates of the model compound corresponding to site i. This selection of atoms gives overall neutrality for the peptide groups when the CHARMM (Brooks et al., 1983) charge set is used.

Protonation is represented as the addition of a unit charge to a point in the titrating site. It is assumed that the chemical bonds formed in introducing the model compound into the protein are far enough removed from the titrating site that $\Delta G_p(AH,A) - \Delta G_s(AH,A)$ will have significant contributions only from electrostatic interactions. For the calculation of $\Delta G_s(AH,A)$, we use a model compound conformation that is the same as that of the titrating group in the protein and

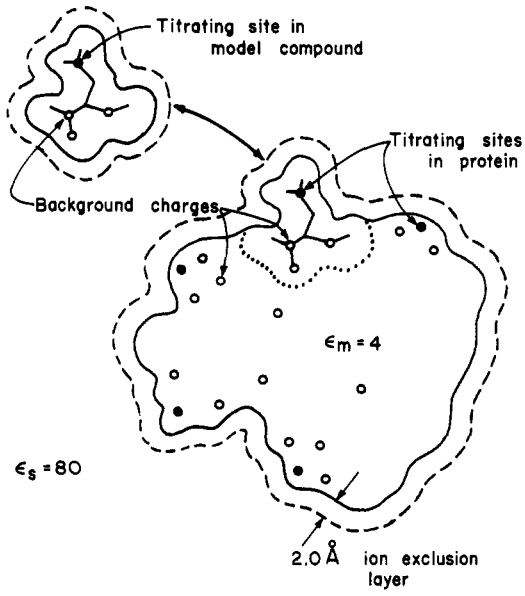


FIGURE 1: Model used for the calculation of electrostatic contributions to titration behavior. The dotted line shows the dielectric boundary of the model compound (see text).

neglect any conformational change on protonation (but see below). In both the protein and the model compound there are interactions with permanent charges (only partial charges in the model compound) and with the induced polarization of the surrounding medium. For the protein, there are also interactions with other titratable sites that depend on their protonation state.

Since $\Delta G_p(AH,A)$ and $\Delta G_s(AH,A)$ are assumed to differ only in their electrostatic interactions, the essential step in determining a pK_a is the calculation of the electrostatic work of charging a site in the protein or model compound from zero to one unit (positive or negative) of electronic charge. The pK_a of site i in a hypothetical protein molecule having the titration state of all other sites fixed can then be written:

$$\begin{aligned} pK_a(\text{protein}) &= pK_a(\text{model}) + \frac{1}{2.303k_BT}[\Delta G_{Born}(\text{protein}) - \\ &\quad \Delta G_{Born}(\text{model}) + \Delta G_{back}(\text{protein}) - \\ &\quad \Delta G_{back}(\text{model}) + \Delta G_{interact}(\text{protein})] \\ &= pK_{intr} + \Delta G_{interact}(\text{protein}) \end{aligned} \tag{2}$$

where the $\Delta G(AH,A)$ terms of eq 1 are represented as the sum of three electrostatic charging terms: ΔG_{Born}, the Born solvation energy of the charge at site i; ΔG_{back}, the interaction of the charge at site i with nontitrating "background" charges (e.g., the peptide group dipoles); $\Delta G_{interact}$, the interaction between site i and other titratable sites in the protein. The signs of these terms depend on whether the site is cationic or anionic. The intrinsic pK_a, pK_{intr}, is a useful intermediate quantity defined as the pK_a of the site when all other titratable sites in the protein are neutralized (Tanford & Kirkwood, 1957). In many previous studies the pK_{intr} values were empirically adjusted (Orttung, 1970; Matthew & Gurd, 1986; States & Karplus, 1987), but here we calculate them explicitly.

The main task is to evaluate these electrostatic terms for the protein and model compounds by finite difference methods. The electrostatic fields are assumed to be governed by the linearized Poisson–Boltzmann equation (McQuarrie, 1976):

$$\nabla\epsilon(\mathbf{r})\nabla\phi(\mathbf{r}) - \epsilon(\mathbf{r})\kappa^2(\mathbf{r})\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{3}$$

$$\epsilon\kappa^2 = \frac{8\pi e^2 N_a I}{k_BT}$$

where ϕ is the electric potential, ϵ is the dielectric constant, κ is the inverse salt screening length, e is the electronic charge, N_a is the Avogadro number, and I is the ionic strength. The shape of the dielectric boundary and the exclusion of salt ions from the protein interior are introduced by the $\mathbf{r}$ dependence of ϵ and κ (see Figure 1). In regions where salt ions are excluded, $\kappa = 0$; otherwise, the value in eq 3 is used. Equation 3 can be solved on a cubic lattice by standard finite difference methods (Press et al., 1986). Our implementation of the method used an initial 80-Å cube with a 1.0-Å lattice spacing centered on the protein followed by a 20-Å cube with a 0.25-Å spacing centered on the titrating site to improve the accuracy of the potential near the titrating sites. To check the accuracy of the finite difference methods, several calculations were done in which the lattice was rotated with respect to the protein. The deviations produced by such rotations should be representative of the errors introduced by the finite lattice spacing. The deviations in calculations of ΔG_{Born}, ΔG_{back}, and the site–site interactions were <5%, <3%, and <10%, respectively.

For the ith site in the protein (see eq 2), the work of adding a charge q_i is

$$\Delta G_i = \Delta G_{Born} + \Delta G_{back} + \sum_j q_i q_j W_{ij} \quad (4)$$

where ΔG_{Born} and ΔG_{back} are the solvation and background terms, respectively, and the final term is $\Delta G_{interact}$, the interaction between titrating sites in which q_j is the charge on the jth titrating site (also, see eq 2). The corresponding expression for the model compound contains only ΔG_{Born} and ΔG_{back} terms. The first term, ΔG_{Born}, is due to the interaction of the charge at $\mathbf{r}_i$ with the polarization that the charge itself has induced in the surroundings. It is calculated by a charging process at $\mathbf{r}_i$. The potential $\Phi(\mathbf{r}_i,\mathbf{r}_j)$ is defined as the potential produced at $\mathbf{r}_j$ by a unit charge at $\mathbf{r}_i$. If $\mathbf{r}_i$ and $\mathbf{r}_j$ both lie in the $\epsilon = \epsilon_m$ (interior) region, Φ can be written as the sum of a Coulomb term and a term, Φ^*, due to the dielectric interface (Jackson, 1975), $\Phi(\mathbf{r}_i,\mathbf{r}_j) = 1/(\epsilon_m|\mathbf{r}_i - \mathbf{r}_j|) + \Phi^*(\mathbf{r}_i,\mathbf{r}_j)$. Since ΔG_{Born} is an interaction of a charge at $\mathbf{r}_i$ with a field at $\mathbf{r}_i$, it is necessary to omit the Coulomb term to avoid the singularity as $\mathbf{r}_j$ approaches $\mathbf{r}_i$. Physically, this term corresponds to the infinite self-energy of compressing a finite charge into a point and should not appear in solvation free energy calculations. The finite difference procedure used in the present study to solve eq 1 does not give Φ^* directly. If a single unit charge is placed on a lattice point corresponding to $\mathbf{r}_i$, the finite difference procedure gives a lattice field corresponding to Φ, but in place of the Coulomb term's $1/r$ singularity at $\mathbf{r}_i$, there is a finite contribution at $\mathbf{r}_i$ which depends on the lattice spacing, the interior dielectric constant, ϵ_m, and the way in which $\mathbf{r}_i$ is positioned relative to the lattice. By keeping these factors the same in both the protein and model compound calculations, the Coulomb contribution can be made to cancel out when the differences in ΔG_{Born} are calculated. The infinitesimal contribution to ΔG_{Born} in a charging process is $\delta G_{Born} = \delta q q \Phi^*(\mathbf{r}_i,\mathbf{r}_i)$, whose integral gives $\Delta G_{Born} = (q_i^2/2)\Phi^*(\mathbf{r}_i,\mathbf{r}_i)$. It can be shown that the ΔG_{Born} term defined here reduces to the Born model for the electrostatic solvation free energy in the case of spherical monoatomic ions.

The second term in eq 4, ΔG_{back}, the interaction with the background of nontitrating partial charges, q_k, is

$$\Delta G_{back} = \sum_k q_i q_k \Phi(\mathbf{r}_i,\mathbf{r}_k) \quad (5)$$

where $\Phi(\mathbf{r}_i,\mathbf{r}_k)$ is defined as the potential at $\mathbf{r}_i$, due to a unit charge at $\mathbf{r}_j$. The third term in eq 4 is due to the interactions with the other titrating sites, j, in the protein and depends on their protonation state; W_{ij} is given by $W_{ij} = \Phi(\mathbf{r}_i,\mathbf{r}_j)$ where $\mathbf{r}_j$ are the other titrating sites. The calculation of ΔG_{back} and W_{ij} by the finite difference method is straightforward (Press et al., 1986). It should be noted that since eq 3 is linear in ϕ and ρ, the $\Phi(\mathbf{r}_i,\mathbf{r}_j)$ functions are charge independent and all terms in eq 4 are either linear or quadratic in the charges.

Once the pK_{intr} for each site and the W_{ij} have been obtained from the finite difference method, the titration behavior can be determined. In principle, one should calculate the titration curve by taking a Boltzmann-weighted sum over all possible protonation states at each pH; i.e., each site is either protonated or not. The fraction of molecules having site i protonated is then

$$\theta_i = \sum_{\{\mathbf{x}\}} x_i \exp(\sum_\mu x_\mu (2.303)(\mathrm{p}K_{intr,\mu} - \mathrm{pH}) - (1/2)\beta \sum_{\mu,\nu} (x_\mu + q_\mu^\circ)(x_\nu + q_\nu^\circ) W_{\mu\nu}) / \sum_{\{\mathbf{x}\}} \exp(\sum_\mu x_\mu (2.303)(\mathrm{p}K_{intr,\mu} - \mathrm{pH}) - (1/2)\beta \sum_{\mu,\nu} (x_\mu + q_\mu^\circ)(x_\nu + q_\nu^\circ) W_{\mu\nu}) \quad (6)$$

where $\mathbf{x}$ is an N element protonation state vector whose elements, x_μ, are zero or one according to whether site μ is protonated or unprotonated, q_μ° is the charge of the unprotonated state of site μ, and the $\{\mathbf{x}\}$ indicates summation over all 2^N possible protonation states. In a protein with 20 titrating sites, this summation would contain 2^{20} terms and would be very expensive to compute. A common approximation (Tanford & Roxby, 1972; Matthew & Gurd, 1986) has been to iterate to a self-consistent set of *partial* charges on the titrating sites at a given pH. However, we have found that this method can lead to significant errors and have developed an alternative approximation method. It takes advantage of the fact that at a given pH many of the sites will be either almost entirely protonated or entirely deprotonated, so the Boltzmann summation can be taken over a reduced set of sites. At each pH point, we make a precalculation of the maximum and minimum possible θ_i values for each site based on the calculated pK_{intr} and W_{ij} values. If θ_{max} is less than 0.05, the site is regarded as entirely deprotonated, and if θ_{min} is greater that 0.95, the site is regarded as entirely protonated. In tests on a series of statistically generated 10-site molecules, titration curves generated by this "reduced site" approximation differed from those from the more rigorous eq 6 by less than 0.03 proton at the titration midpoints of each site (Bashford and Karplus, unpublished results).

Results and Discussion

We have carried out calculations for both the triclinic and tetragonal crystal forms of lysozyme. The heavy-atom coordinates for the triclinic form of hen egg white lysozyme were taken from the data of Hodsdon et al. (unpublished) deposited in the Brookhaven databank [see Bernstein et al. (1977)]. Coordinates for the tetragonal form were kindly provided by D. C. Phillips. Polar hydrogen positions were generated by the HBUILD facility of the CHARMM computer program (Brünger & Karplus, 1988). Atomic radii and nontitrating partial charges were taken from the polar hydrogen parameter set of CHARMM (Brooks et al., 1983). For titrating groups, it was assumed that the entire formal charge was localized on one atom. The single sites for the titrating group charges are taken to be the carboxyl carbon atom for Asp, Glu, and the C-terminus; the N_{ϵ_2} atom for His; the hydroxyl oxygen atom of Trp; and the amino nitrogen atom of Lys and the N-terminus. Arginine side chains are treated as part of the nontitrating background since they titrate beyond the pH range in which lysozyme is stable. We have made trial calculations in which the titrating charge is distributed over several atoms,

Table I: Calculation of Intrinsic pK_a for Tetragonal and Triclinic Lysozyme[a]

site	pK_{mod}	Born terms	background terms	pK_{intr}
		(A) Tetragonal Lysozyme		
N term	7.5	−4.0	1.6	5.1
His 15	6.3	−3.7	−0.3	2.3
Glu 7	4.4	2.3	−1.6	5.1
Glu 35	4.4	4.6	−3.1	5.9
Asp 18	4.0	2.7	−3.5	3.2
Asp 48	4.0	2.3	−4.6	1.6
Asp 52	4.0	3.6	−1.3	6.3
Asp 66	4.0	6.4	−8.3	2.1
Asp 87	4.0	0.8	−2.8	2.0
Asp 101	4.0	1.5	−0.5	5.0
Asp 119	4.0	1.6	−4.1	1.5
Tyr 20	9.6	2.8	−0.8	11.6
Tyr 23	9.6	3.4	−3.3	9.8
Tyr 53	9.6	5.7	−3.8	11.4
Lys 1	10.4	−2.8	−0.6	6.9
Lys 13	10.4	−1.8	−0.7	7.9
Lys 33	10.4	−1.6	−1.3	7.5
Lys 96	10.4	−2.5	0.6	8.6
Lys 97	10.4	−2.5	−0.4	7.5
Lys 116	10.4	−1.9	0.8	9.3
C term	3.8	0.9	−1.0	3.6
		(B) Triclinic Lysozyme		
N term	7.5	−3.2	2.3	6.6
His 15	6.3	−3.6	1.2	3.9
Glu 7	4.4	1.1	−2.8	2.7
Glu 35	4.4	4.4	−2.8	6.0
Asp 18	4.0	2.1	−2.6	3.5
Asp 48	4.0	2.2	−5.1	1.1
Asp 52	4.0	3.2	−1.6	5.6
Asp 66	4.0	5.7	−8.0	1.7
Asp 87	4.0	0.9	−2.3	2.7
Asp 101	4.0	2.0	2.3	8.3
Asp 119	4.0	0.9	−1.6	3.3
Tyr 20	9.6	3.9	−0.2	13.2
Tyr 23	9.6	3.3	−1.8	11.2
Tyr 53	9.6	7.2	−4.5	12.3
Lys 1	10.4	−1.1	−0.4	8.9
Lys 13	10.4	−0.3	−0.5	9.6
Lys 33	10.4	−1.6	0.4	9.1
Lys 96	10.4	−2.0	1.7	10.1
Lys 97	10.4	−0.7	0.2	10.0
Lys 116	10.4	−0.4	−0.4	9.7
C term	3.8	0.5	−0.7	3.6

[a] pK_{mod} are the model compound pK_a's taken from data on the titration behavior of denatured proteins (Nozaki & Tanford, 1967).

but the results differ only slightly from those obtained with single charges, and one encounters uncertainty as to which oxygen atom in the carboxyl group should receive the proton.

The results of the calculation for the triclinic and tetragonal crystal structures of lysozyme are given in Tables I and II. Table I shows that the Born term lowers the pK_a of cationic sites and raises the pK_a of anionic sites. This is expected since the charged states of the sites are less well solvated (more buried in a low-dielectric medium) in the protein than in the model compound. In most cases, the background term has a sign opposite to that of the Born term, so that loss of solvation is approximately compensated by favorable interactions with the permanent charges of the protein. For the lysines, there is no clear trend in the background terms; the shifts are relatively small since the lysines tend to be well solvated, and the interpretation is complicated by the fact that the 11 arginines in lysozyme have been excluded from consideration as titratable sites, so their "permanent" positive charge is included in ΔG_{back}. Table II shows that 11 of the 21 values of pK_{app} calculated for the triclinic form are within 1 pK unit of the experimental values; for the tetragonal calculation, 10 of the values agree to within 1 pK unit. The direction of the shift between model compound pK_a's and calculated pK_{app}'s is in agreement with experiment for 15 sites in the tetragonal form and 12 sites in the triclinic form. Errors are usually due to overestimates of the magnitude of the shift. It is striking that the calculated pK_{app} values for the triclinic and tetragonal forms are within 1 pK unit of each other for only 8 of 21 sites and differ by as much as 3.3 pK units. This indicates that conformational effects make significant contributions. (The root-mean-square variation in atomic coordinates between the two crystal forms is 0.616 Å for backbone atoms and 2.293 Å for side-chain atoms.) In many cases, the difference can be localized to the interactions of certain side chains that have different positions in the two structures, but there are others where side-chain–main-chain or more distributed interactions are involved (see below). The difference between the model compound pK_a and the calculated pK_{intr} is generally larger than the difference between pK_{intr} and pK_{app}; that is, the effect of the Born and background terms in eqs 2 and 4 is generally more important than site–site interactions. Analysis in terms of effective dielectric constants, ϵ_{eff} (defined by $\epsilon_{eff} = q_i q_j / r\Phi$), shows the extent to which most site–site interactions are screened by solvent and salt effects, especially at long distances. In the triclinic calculations, the 210 pairs of sites have an average ϵ_{eff} of 895; this perhaps surprisingly large value for ϵ_{eff} arises from the exponential screening for distances larger than the Debye screening parameter κ^{-1} (eq 3), which is approximately 10 Å for an ionic strength of 0.1 M. Only 23 pairs have $\epsilon_{eff} < 80$, and all of these are less than 12 Å apart. The strongest interaction, that between Asp 66 and Tyr 53 (see below), has $\epsilon_{eff} = 10$ at a distance of 3.1 Å. These results are consistent with previous finite difference calculations (Sternberg et al., 1987; Gilson & Honig, 1987, 1988; Bashford & Karplus, 1988) and with the interpretation of certain trends in experimental measurements of pK_{eff} (Rees, 1980; Mehler & Eichele, 1984).

The results for His 15 provide a striking example of conformational effects. In the tetragonal calculations (Table IA) the ΔG_{Born} term lowers the pK_a by 3.7 units, indicative of a buried residue, and the ΔG_{back} term, rather than compensating, lowers the pK_a by an additional 0.3 unit. The His 15 side chain is buried by the guanidinium group of Arg 14, which lies parallel to and at a distance of about 4.6 Å from the imidazole group (Figure 2, top). Both the loss of solvation and the noncompensating background term are apparently due to this conformation. In the triclinic structure, His 15 has a similar conformation, but Arg 14 has moved so that the imidazole group is exposed to solvent and the guanidinium group is 8.3 Å away (Figure 2, bottom). As seen in Tables I and II, the ΔG_{Born} term is similar, but it is partially compensated by the ΔG_{back} term, so that pK_{app} is much closer to experiment. The change in ΔG_{back} is due to Arg 14; Arg residues are considered as part of the background term (see above). Conformational effects are also evident in the case of Asp 101. In the triclinic structure, the ΔG_{Born} and ΔG_{back} terms each increase the pK_a by about 2.0 units, resulting in a pK_{app} much higher than experiment. Asp 101 is in a turn at the C-terminal end of an α-helix, and in the triclinic structure the Asp carboxyl group is, somewhat surprisingly, in close proximity to two peptide oxygen atoms from the helix backbone and a third from the turn. In the tetragonal structure the turn and side-chain conformations are such that the Asp carboxyl group points out into the solvent and is farther from the peptide oxygen atoms. A slight increase in solvation lowers the Born term by 0.5 unit, and the negative background and site–site interactions result in a near-normal pK_{app}, in agreement with experiment. A third example of a conformational effect is

Table II: Results of Titration Calculation for Tetragonal and Triclinic Lysozyme[a]

site	pK_{mod}	pK_{intr} tetr	pK_{intr} tric	pK_{app} tetr	Δ	pK_{app} tric	Δ	exptl pK_a[b]	Δ
N term	7.5	5.1	6.6	5.0	−2.5	6.4	−1.1	7.8–8.0*	0.4
His 15	6.3	2.3	3.9	2.4	−3.9	4.0	−2.3	5.8*	−0.5
Glu 7	4.4	5.1	2.7	1.2	−3.2	2.1	−2.3	2.6*	−1.8
Glu 35	4.4	5.9	6.0	6.2	1.8	6.3	1.9	6.1*	1.7
Asp 18	4.0	3.2	3.5	2.6	−1.4	3.1	−0.9	2.8–3.0**	−1.1
Asp 48	4.0	1.6	1.1	1.6	−2.4	1.0	−3.0	4.3*	0.3
Asp 52	4.0	6.3	5.6	8.5	4.5	7.0	3.0	3.5–3.7**	−0.4
Asp 66	4.0	2.1	1.7	2.2	−1.8	1.7	−2.3	1.5–2.5**	−2.0
Asp 87	4.0	2.0	2.7	0.8	−3.2	1.2	−2.8	3.5–3.75**	−0.4
Asp 101	4.0	5.0	8.3	4.3	0.3	7.9	3.9	4.0–4.25**	0.1
Asp 119	4.0	1.5	3.3	1.3	−2.7	3.2	−0.8	2.2–2.8*	−1.5
Tyr 20	9.6	11.6	13.2	12.1	2.5	14.0	4.4	10.3*	0.7
Tyr 23	9.6	9.8	11.2	10.1	0.5	11.7	2.1	9.8*	0.2
Tyr 53	9.6	11.4	12.3	18.8	9.2	20.8	11.2	12.1*	2.5
Lys 1	10.4	6.9	8.9	10.8	0.4	9.6	−0.8	10.7–10.9*	0.4
Lys 13	10.4	7.9	9.6	10.1	−0.3	11.6	1.2	10.4–10.6*	0.1
Lys 33	10.4	7.5	9.1	7.7	−2.7	9.6	−0.8	10.5–10.7*	0.2
Lys 96	10.4	8.6	10.1	8.9	−1.5	10.4	0.0	10.7–10.9*	0.4
Lys 97	10.4	7.5	10.0	8.4	−2.0	10.6	0.2	10.2–10.4*	−0.1
Lys 116	10.4	9.3	9.7	9.7	−0.7	9.9	−0.5	10.3–10.5*	0.0
C term	3.8	3.6	3.6	2.2	−1.6	2.3	−1.5	2.7–2.8**	−1.0

[a] The apparent pK_a of a site, pK_{app}, is defined as the point in the calculated titration curve where the site is half-protonated. The three columns labeled Δ give the shifts relative to pK_{mod} for the tetragonal pK_{app}, the triclinic pK_{app}, and the experimental pK_a, respectively. [b] (*) data from Kurimatsu and Hamaguchi (1980); (**) data from Dobson et al. (personal communication).

provided by Glu 7 and Lys 1. There is a significant increase in pK_{intr} of Glu 7 in the tetragonal versus the triclinic structure due to the differences in the Born and Background terms. However, in the tetragonal structure Glu 7 and Lys 1 are only 3.3 Å apart (Figure 2, top), and their interaction is equivalent to 3.4 pK units. This contributes to a large downward shift in the pK_a of Glu 7 and an upward shift for Lys 1, relative to the pK_{intr} values. The Glu 7–Lys 1 salt bridge has been observed in solution NMR studies of lysozyme with methylated lysines (Gerken et al., 1982). In the triclinic structure, Glu 7 is pointing away from Lys 1 (Figure 2, bottom) so that the distance is 11.5 Å and the interaction is weaker by a factor of 10.

The results for the two active site carboxyl groups, Glu 35 and Asp 52, make an interesting contrast in solvation effects. For Glu 35, both the triclinic and tetragonal calculations show a loss of solvation that is incompletely compensated by the background with the result that the elevated experimental pK_a is reproduced. However, for Asp 52 the calculated pK_{app} values are too high, also due to uncompensated loss of solvation. It is of interest that Mehler and Eichele (1984) found that esterification of Asp 52 shifts the pK_a of Glu 35 by 1.1 pK units; the calculated W_{ij} value between Glu 35 and Asp 52 gives a shift of 0.8 pK unit.

The largest difference between the calculated pK_{app} and the experimental pK_a occurs for Tyr 53. The calculated pK_{intr} is near the experimental pK_a, but a strong interaction with Asp 66 raises the pK_{app} value far higher. It is likely that the difference between theory and experiment can be explained by a significant conformational change upon deprotonation. There is experimental evidence that the unfolding of lysozyme is associated with the ionization of this residue (Imoto et al., 1972).

The overall agreement with experiment obtained with the present calculations is rather good considering the difficulty of calculating the effects of solvation and interaction with nontitrating protein polar groups. Warshel et al. (1984) have estimated shifts of 25 pK units for the Born term with interior dielectric constant $\epsilon_{int} = 2$ and concluded that macroscopic models with a low ϵ_{int} cannot give correct results. Their calculation is based on the complete immersion of a small charged sphere inside a large sphere representing the protien, and the formula used is truncated at the dipole term, so that it is valid only when the smaller sphere is deeply buried inside the larger one. In the present calculations the shifts due to ΔG_{Born} are relatively small, only 1–7 pK units. This occurs because many of the surfaces that are exposed to solvent in the model compound are correspondingly exposed in the protein site. Even the more buried sites, such as Asp 66 and Glu 35, are sufficiently near the protein surface to have significant favorable solvation energies. Microscopic models have been applied to the calculation of protonation free energy differences (Russell & Warshel, 1985; Warshel et al., 1986; Rullmann et al., 1989), but these methods are computationally expensive and have problems of accuracy and convergence. Solvation energies of around −70 kcal/mol in model compounds and in the protein must be subtracted to give energy differences on the order of 1 kcal/mol if the pK_a shifts observed experimentally are to be reproduced (Russell & Warshel, 1985). The macroscopic method presented here achieves this level of accuracy for many sites. The deviation of some of the present results from experiment is not necessarily due to the inherent limitations of macroscopic electrostatic theory. The differences between the calculations for the two different crystal structures, particularly in the cases of His 15 and Asp 101, indicate that the neglect of conformational variability may be the major source of error. Another possible source of deviations between experimental and calculated results is the binding of counterions to specific sites on the protein. To model this effect, it would be necessary to identify ion-binding sites and include the ion charges explicitly in the ΔG_{back} term. Bound ions have been included in a Tanford–Kirkwood-type calculation for ribonuclease A (Matthew & Richards, 1982).

Conclusions

The present calculations provide an instructive partitioning of the effects of the protein environment into Born solvation, which depends strongly on solvent accessibility; background interaction, which depends on nearby nontitrating charges in the protein; and the more familiar site–site interactions. Extension of these techniques to include local conformational mobility, such as averaging over multiple side-chain confor-

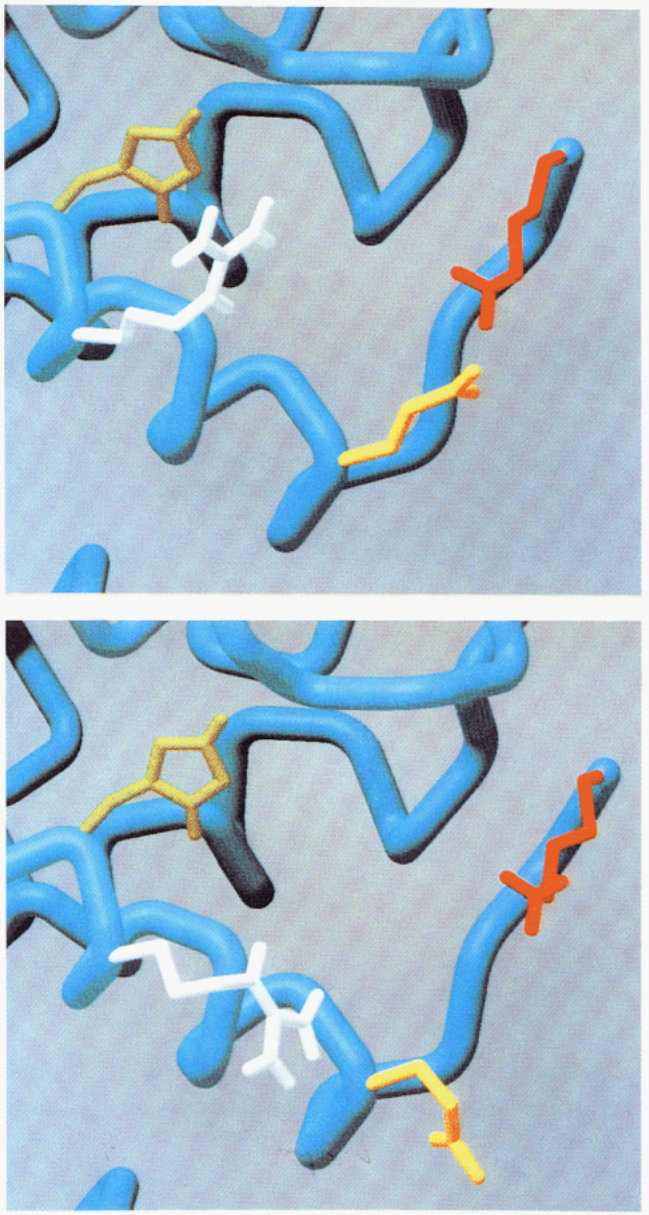


FIGURE 2: Crystal structures of hen egg white lysozyme showing the variability in positions of surface residues: (top) tetragonal structure (Phillips et al., unpublished results); (bottom) triclinic (Hodson et al., unpublished results). His 15 (green), Arg 14 (white), Glu 7 (yellow), and Lys 1 (red) are highlighted, and the protein backbone in this region is shown in blue. (Photographs by Michael Pique, Research Institute of Scripps Clinic, using MCS software by Michael Connolly.)

mations weighted by an energy determined with a combination of empirical energy functions and continuum electrostatic models, may make reliable predictions of protein pK_a possible for sites where large conformational changes upon protonation or large differences between solution and crystal structure do not occur. Conversely, deviations between measured and calculated pK_a values may be a useful indicator of conformational effects. Experimental data on the mobility of titrating sites and protein conformational changes upon protonation would be of great interest in this context. Preliminary NMR data (C. M. Dobson, C. Redfield, L. J. Smith, and M. J. Sutcliffe, personal communication) show considerable mobility in exterior lysozyme side chains. Corresponding techniques may also be applicable for estimating the effect of protein solvation on the charged species generated in electron transfer systems (Treutlein et al., 1988).

Acknowledgments

We thank C. M. Dobson for discussions and for providing data in advance of publication. We thank Mike Sommer for helpful discussions and calculations that checked some of the results reported in this paper. We thank the Minnesota Supercomputer Center and the National Science Foundation for grants of Cray-2 time. Calculations were done at the Minnesota Supercomputer Center and at NASA Ames.

Registry No. L-His, 71-00-1; L-Glu, 56-86-0; L-Asp, 56-84-8; L-Tyr, 60-18-4; L-Lys, 56-87-1; lysozyme, 9001-63-2.

References

Bashford, D., & Karplus, M. (1988) *J. Mol. Biol. 203*, 507–510.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput, Chem. 4*, 187–217.

Brünger, A. T., & Karplus, M. (1988) *Proteins 4*, 148–156.

Fersht, A. (1985) *Enzyme Structure and Mechanism*, W. H. Freeman, New York.

Gerken, T. A., Jentoft, J. E., Jentoft, N., & Dearborn, D. G. (1982) *J. Biol. Chem. 257*, 2894–2900.

Gilson, M. K., & Honig, B. H. (1986) *Biopolymers 25*, 2097–2119.

Gilson, M. K., & Honig, B. H. (1987) *Nature 330*, 84–86.

Gilson, M. K., & Honig, B. H. (1988) *Proteins 3*, 32–52.

Harvey, S. (1989) *Proteins 5*, 78–92.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., & Rupley, J. A. (1972) in *The Enzymes* (Boyers, P. D., Ed.) pp 665–868, Academic, New York.

Jackson, J. D. (1975) *Classical Electrodynamics*, Wiley, New York.

Jorgensen, W. L. (1989) *Acc. Chem. Res. 22*, 184–189.

Knowles, J. R. (1976) *CRC Crit. Rev. Biochem. 4*, 165–173.

Kortum, G., & Bockris, J. O. (1951) *Textbook of Electrochemistry*, Elsevier, New York.

Kurimatsu, S., & Hamaguchi, K. (1980) *J. Biochem. 87*, 1215–1219.

Matthew, J. B., & Richards, F. M. (1982) *Biochemistry 21*, 4989–4999.

Matthew, J. B., & Gurd, F. R. N. (1986) *Methods Enzymol. 130*, 413–436.

McQuarrie, D. A. (1976) *Statistical Mechanics*, Harper and Row, New York.

Mehler, E. L., & Eichele, G. (1984) *Biochemistry 23*, 3887–3891.

Nakamura, H., Sakamoto, T., & Wada, A. (1988) *Protein Eng. 2*, 177–183.

Nozaki, Y., & Tanford, C. (1967) *Methods Enzymol. 11*, 715–734.

Orttung, W. H. (1970) *Biochemistry 9*, 2394–2402.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986) *Numerical Recipes. The Art of Scientific Computing*, Cambridge University Press, Cambridge, U.K.

Rees, D. C. (1980) *J. Mol. Biol. 141*, 323–326.

Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151–176.

Rullmann, J. A. C., Bellido, M. N., & van Duijnen, P. Th. (1989) *J. Mol. Biol. 206*, 101–118.

Russell, S. T., & Warshel, A. (1985) *J. Mol. Biol. 185*, 389–404.

Schulz, G. E., & Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer-Verlag, New York.
Shire, S. J., Hanania, G. I. H., & Gurd, F. R. N. (1974) *Biochemistry 13*, 2967-2974.
States, D. J., & Karplus, M. (1987) *J. Mol. Biol. 197*, 122-130.
Sternberg, M. J. E., Hays, F. R. F., Russell, A. J., Thomas, P. G., & Fersht, A. R. (1987) *Nature 330*, 86-88.
Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Wiley, New York.
Tanford, C., & Kirkwood, J. G. (1957) *J. Am. Chem. Soc. 79*, 5333-5339.
Tanford, C., & Roxby, R. (1972) *Biochemistry 11*, 2192-2198.
Treutlein, H., Schulten, K., Deisenhofer, J., Michel, H., Brünger, A., & Karplus, M. (1988) in *The Photosynthetic Bacterial Reaction Center* (Vermeglio, A., Ed.) pp 101-118, Plenum Press, London.
Van Belle, D., Couplet, I., Prevost, M., & Wodak, S. J. (1987) *J. Mol. Biol. 198*, 721-735.
Warshel, A., & Russell, S. T. (1984) *Q. Rev. Biophys. 17*, 283-422.
Warshel, A., Russell, S. T., & Churg, A. K. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4785-4789.
Warshel, A., Sussman, F., & King, G. (1986) *Biochemistry 25*, 8368-8372.
Warwicker, J., & Watson, H. C. (1982) *J. Mol. Biol. 157*, 671-679.

Recombinant Expression, Biochemical Characterization, and Biological Activities of the Human MGSA/*gro* Protein†

Rik Derynck,*,‡ Eddy Balentien,‡,§ Jin Hee Han,§ H. Greg Thomas,§ Duanzhi Wen,‡ Ajoy K. Samantha,∥ Claus O. Zachariae,∥ Patrick R. Griffin,⊥ Rainer Brachmann,‡ Wai Lee Wong,‡ Kouji Matsushima,∥ and Ann Richmond§,#

Departments of Developmental Biology and Protein Biochemistry, Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, Veterans Administration Medical Center, Atlanta, and Emory University School of Medicine, 1670 Clairmont Road, Decatur, Georgia 30033, and Laboratory of Molecular Immunoregulation, Biological Response Modifiers Program, Division of Cancer Treatment, National Cancer Institute, Frederick, Maryland 21701

Received April 27, 1990; Revised Manuscript Received July 9, 1990

ABSTRACT: Melanoma growth stimulatory activity (MGSA) is a mitogenic protein secreted by Hs294T melanoma cells that corresponds to the polypeptide encoded by the human *gro* gene. The MGSA/*gro* cDNA has been expressed in mammalian cells and the secreted recombinant factor has been purified. Biochemical and biological characterization shows that the recombinant protein is identical with the natural protein and is devoid of posttranslational glycosylation, sulfation, and phosphorylation. The two C-terminal amino acids are proteolytically removed from the mature recombinant MGSA, indicating a length of 71 instead of the predicted 73 amino acids. The recombinant MGSA is mitogenically active on the Hs294T melanoma cells. The purified MGSA competes with interleukin 8 for binding to neutrophil receptors and exhibits neutrophil chemotactic activity equivalent to that of interleukin 8.

A variety of factors modulate the proliferation of mammalian cells in vitro and in vivo. These growth regulators can either stimulate or inhibit cellular proliferation depending on the nature of the factor, the cell type, and the physiological circumstances. The biological and structural studies on growth factors have largely concentrated on a limited set of growth factors, which by now have been relatively well-defined. However, there are still a variety of growth regulators that are less well-defined, often due to a lack of purified protein. One of these factors is known as "melanoma growth stimulatory activity" (MGSA). MGSA was initially identified as a factor secreted by the human melanoma cell line Hs294T with the ability to mitogenically stimulate these same melanoma cells (Richmond et al., 1982, 1983, 1985; Richmond & Thomas, 1986). Recent isolation and characterization of purified protein (Thomas & Richmond, 1988) and cDNAs (Richmond et al., 1988) have revealed that human MGSA is a single-chain polypeptide with a predicted length of 73 amino acids that is proteolytically derived from a 107 residue long precursor. The determination of the derived polypeptide sequence also established that human MGSA is identical with the gene product of the human *gro* gene. The cDNAs for *gro* were isolated and sequenced by Anisowicz et al. (1987), who proposed that the expression of this gene was growth-related. However, they did not define a specific activity or function for the gene product. Northern hybridization has established that the MGSA/*gro* gene is expressed by a variety of normal and transformed cells from different origins, such as fibroblasts, melanoma cells, epithelial cells, and endothelial cells (Anisowicz et al., 1987; Bordoni et al., 1989; Richmond et al.,

† This work was supported in part by NCI Grant CA34590 (A.R.), VA Merit Awards (H.G.T. and A.R.), and the NCI BRM program (K.M.).
* Author to whom correspondence should be addressed.
‡ Department of Developmental Biology, Genentech Inc.
§ VA Medical Center and Emory University School of Medicine.
∥ National Cancer Institute.
⊥ Department of Protein Biochemistry, Genentech Inc.
Present address: Department of Cell Biology, Vanderbilt University Medical School, Nashville, TN 37232.

0006-2960/90/0429-10225$02.50/0