JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

(Registered in U. S. Patent Office)

VOLUME 79 OCTOBER 22, 1957 NUMBER 20

PHYSICAL AND INORGANIC CHEMISTRY

[CONTRIBUTION NO. 1415 FROM STERLING CHEMISTRY LABORATORY, YALE UNIVERSITY]

Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres

BY CHARLES TANFORD[1] AND JOHN G. KIRKWOOD

RECEIVED MAY 8, 1957

For many years the theory of titration curves of impenetrable proteins has been based on a model which represents the protein molecule as a sphere with a continuous and uniform distribution of charge on its surface. In this paper this model is replaced by a more realistic one in which the charges are taken to be discrete unit charges located at fixed positions. General equations are obtained which express the titration curve as a function of the locations of ionizable sites and of their intrinsic properties. It is concluded that the intrinsic properties may themselves be quite sensitive to the location of the dissociable site with respect to the surface of the protein molecule.

Introduction

Hydrogen ion titration curves of proteins can be obtained experimentally with considerable accuracy. Their most prominent feature is a strong electrostatic interaction which results in the fact that the titration curve of a protein containing acidic and basic side chains in any given proportions is considerably flatter than the titration curve of a mixture of corresponding simple acids and bases in the same proportions.

To account theoretically for this effect[2-5] it has been customary to represent a protein molecule by a sphere impenetrable to solvent; to assume all titratable groups independent, except for electrostatic interaction; to consider all of the titratable groups of a given kind (*e.g.*, all phenolic groups) to be intrinsically identical; and to allow the titratable groups to occupy, with equal probability, any position on the surface of the spherical molecule. In addition, the mixture of protein ions, with various integral values of Z, which are ordinarily present in any protein solution, has in the customary treatment been replaced by a single kind of ion with continuously variable charge, $\bar{Z}$.

With these assumptions one obtains the result that the titration curve of a protein is a superposition of the curves for the individual types of groups: the fraction α of dissociated groups of any type being given by the relation

$$p\mathrm{H} - \log\frac{\alpha}{1-\alpha} = pK_{\mathrm{int}} - \frac{1}{2.303kT}\frac{\partial \mathbf{W}}{\partial \bar{Z}} \qquad (1)$$

where K_{int} is an intrinsic dissociation constant characteristic of the type of group and $\mathbf{W}$ is the work done in placing all of the protein charges onto the molecule. Since all of the groups are considered smeared evenly over the surface, positive and negative charges occupy the same space and thus cancel; only the average net charge $\bar{Z}$ remains. In terms of $\bar{Z}$

$$\mathbf{W} = \frac{\bar{Z}^2\epsilon^2}{2D}\left(\frac{1}{b} - \frac{\kappa}{1+\kappa a}\right) \qquad (2)$$

where b is the radius of the sphere, a is the radial distance to which salt ions are excluded (*i.e.*, $a - b$ is of the dimensions of a salt ion radius), κ is the Debye–Hückel parameter proportional to the square root of the ionic strength, ϵ the protonic charge and D the dielectric constant of the *solvent*.

With this value of $\mathbf{W}$, equation 1 becomes

$$p\mathrm{H} - \log\frac{\alpha}{1-\alpha} = pK_{\mathrm{int}} - 0.868w\bar{Z} \qquad (3)$$

where w

$$w = \frac{\epsilon^2}{2DkT}\left(\frac{1}{b} - \frac{\kappa}{1+\kappa a}\right) \qquad (4)$$

(1) Department of Chemistry, State University of Iowa, Iowa City, Iowa. John Simon Guggenheim Memorial Fellow, 1956–1957, Yale University. This work was also supported by research grant RG-2350 from the National Institutes of Health, Public Health Service.

(2) K. Linderström-Lang, *Compt. rend. trav. lab. Carlsberg*, **15**, No. 7 (1924).

(3) J. G. Kirkwood in E. J. Cohn and J. T. Edsall, "Proteins, Amino Acids and Peptides," Reinhold Publ. Corp., New York, N. Y., 1943.

(4) G. Scatchard, *Ann. N. Y. Acad. Sci.*, **51**, 660 (1949).

(5) C. Tanford in T. Shedlovsky, ed., "Electrochemistry in Biology and Medicine," John Wiley and Sons, Inc., New York, N. Y., 1955.

The assumptions inherent in this treatment are clearly invalid. Especially is this true of the assumption that the titratable groups may be considered equally likely to lie anywhere on the molecular surface. The acidic and basic groups of protein molecules lie on side chains known to be arranged in a definite order. Furthermore, protein molecules are believed to be tightly coiled in a specific way, so that the acidic and basic groups must occupy definite fixed positions, not necessarily at the surface, and certainly not evenly distributed.

Accordingly we shall present in this paper a more realistic theory in which the titratable groups are assigned specific fixed locations on the protein molecule. This will markedly alter the expression for **W** (equation 2) and will also affect the term $\log \alpha/(1-\alpha)$ in equations 1 and 3, which occurs there as the result of the random positioning of titratable groups, which places all of them at an equal electrostatic potential and thus equally accessible to protons. The new theoretical treatment will also show when similar groups can be considered intrinsically identical.

For the present we shall continue to assume that coulombic interaction between their charges is the only type of interaction between the groups. Obviously, the present treatment will not apply where there is appreciable hydrogen bonding between side chain groups.[6] The spherical shape of the protein molecule is also maintained, so that the treatment will not apply to markedly asymmetric proteins such as myosin. However, the assumption of spherical shape should not be a serious one for the common globular proteins even if they deviate somewhat from spherical shape, for electrostatic interaction will depend primarily on the distance between charges and how far within the molecule (*i.e.*, within a region of low dielectric constant) they are placed.

We maintain also the impenetrability of the molecule to solvent. It is probable that this assumption is actually valid for many globular proteins. Any water trapped within such proteins is probably tightly held and probably contains none of the salt ions present in the solvent. In a previous paper[7] the effect of solvent penetration has been discussed on the basis of the smeared site model. If solvent penetration occurs this model is actually not as unrealistic a one as for impenetrable proteins, for it is likely that penetrable proteins are less rigidly coiled than impenetrable ones, and that the side chains are free to assume numerous different positions with equal probability. In the case of serum albumin, which is the only protein for which expanded, penetrable configurations have been established, the greater flexibility of the side chains is indicated by the increased rate of depolarization of the fluorescence of attached dye molecules.[8]

It should be mentioned, finally, that the present treatment will not take into account the interaction between protein molecules, *i.e.*, it applies only to titration curves extrapolated to zero protein concentration. At relatively high ionic strength (0.01 or above), however, titration curves as ordinarily performed (at concentrations of 1% or less) are experimentally independent of the concentration.

(6) M. Laskowski and H. A. Scheraga, THIS JOURNAL, **76**, 6305 (1954).

(7) C. Tanford, *J. Phys. Chem.*, **59**, 788 (1955).

(8) G. Weber, *Biochem. J.*, **51**, 155 (1952); W. F. Harrington, P. Johnson and R. H. Ottewill, *ibid.*, **62**, 569 (1956).

The Work of Charging

The model of a protein molecule used in this study is shown in Fig. 1. The positions of m titratable groups are indicated by points. There is interaction between only those points which bear a charge. If they bear charges, these will be *point charges* embedded in a spherical cavity of dielectric constant D_i. The external dielectric constant is D. Actual charges, of course, are not point charges and their self-energy, as calculated by this model, will be erroneous. The self-energy terms, however, will not appear in the final expressions to be derived. Interaction between charges, provided they do not overlap, is independent of whether they are concentrated at a point or spread over a spherical surface of atomic dimensions.

It should be noted that each titratable group of a protein molecule bears a unit charge either in its acidic or in its basic form, but never in both forms. Thus the charge at each of the m points of Fig. 1 is $\xi_k\epsilon$ where ξ_k may be zero or $+1$ for basic groups and zero or -1 for acidic groups.

The work of charging such a sphere has been evaluated by Kirkwood.[9]

With the energy zero for completely discharged protein

$$\mathbf{W} = \frac{\epsilon^2}{2b}\sum_{k=1}^{m}\sum_{l=1}^{m}\xi_k\xi_l(A_{kl} - B_{kl}) - \frac{\epsilon^2}{2a}\sum_{k=1}^{m}\sum_{l=1}^{m}\xi_k\xi_l C_{kl} \quad (5)$$

where

$$A_{kl} = b/D_i r_{kl} \quad (6)$$

$$B_{kl} = \frac{1}{D_i}\sum_{n=0}^{\infty}\frac{(n+1)(D-D_i)}{(n+1)D+nD_i}\rho_{kl}^n P_n(\cos\theta_{kl}) \quad (7)$$

$$C_{kl} = \frac{1}{D}\left\{\frac{x}{1+x} + \sum_{n=1}^{\infty}\frac{2n+1}{2n-1}\left[\frac{D}{(n+1)D+nD_i}\right]^2 \times \frac{x^2\,\sigma_{kl}^n P_n(\cos\theta_{kl})}{\frac{K_{n+1}(x)}{K_{n-1}(x)} + \frac{n(D-D_i)}{(n+1)D+nD_i}\left(\frac{b}{a}\right)^{2n+1}\frac{x^2}{4n^2-1}}\right\} \quad (8)$$

where

$$x = \kappa a \quad (9)$$

$$\rho_{kl} = r_k r_l/b^2 \quad (10)$$

$$\sigma_{kl} = r_k r_l/a^2 \quad (11)$$

$$K_n(x) = \sum_{s=0}^{n}\frac{2^s\, n!(2n-s)!}{s!\,(2n)!\,(n-s)!}x^s \quad (12)$$

$P_n(\cos\theta_{kl})$ represents ordinary Legendre polynomials and r_k, r_l, r_{kl}, and θ_{kl} are identified by Fig. 1.[10]

The following physical significance may be assigned to the various terms of equation 5: the

(9) J. G. Kirkwood, *J. Chem. Phys.*, **2**, 351 (1934).

(10) In dealing with a single pair of charges, Kirkwood and Westheimer, *J. Chem. Phys.*, **6**, 506 (1938), found it convenient to express **W** in terms of an *effective dielectric constant*, D_E. In their notation (no salt present) $1/D_E$ would be equivalent to our $(r_{kl}/b)(A_{kl} - B_{kl})$.

factor involving the A_{kl} represents the work of charging in an unbounded medium of dielectric constant D_i; the factor involving the B_{kl} represents the modification arising from the fact that the protein molecule is a bounded cavity within a medium of higher dielectric constant D; the factor involving the C_{kl} represents the interaction with the salt ions of the solvent and vanishes at zero salt concentration, when $x = \kappa a = 0$.

In the factors containing A_{kl} and B_{kl} the terms with $k = l$ (*i.e.*, A_{kk} and B_{kk}) are self-energy terms, while those with $k \neq l$ represent pair-wise interaction between the sites. In the ionic strength-dependent factor the terms with $k = l$ represent the excess chemical potential of individual charges due to their interaction with salt ions; terms with $k \neq l$ represent the effect of the salt ions on the pair-wise interactions. A_{kk} and, when $\rho_{kk} = 1$, also B_{kk}, are infinite because of the assumption that all charges are assumed concentrated at a point.

It will be found convenient to separate the infinite self-energy terms from the rest and to write

$$\mathbf{W} = \frac{\epsilon^2}{2b}\sum_{k=1}^{m}\xi_k^2(A_{kk} - B_{kk}) + W \qquad (13)$$

$$W = \frac{\epsilon^2}{2b}\sum_{k=1}^{m}\sum_{l\neq k}\xi_k\xi_l(A_{kl} - B_{kl}) - \frac{\epsilon^2}{2a}\sum_{k=1}^{m}\sum_{l=1}^{m}\xi_k\xi_l C_{kl} \qquad (14)$$

Since

$$r_{kl} = b\left[\rho_{kl}\left(\frac{r_k}{r_l} + \frac{r_l}{r_k} - 2\cos\theta_{kl}\right)\right]^{1/2} \qquad (15)$$

A_{kl} may be rewritten as

$$A_{kl} = 1/D_i\left[\rho_{kl}\left(\frac{r_k}{r_l} + \frac{r_l}{r_k} - 2\cos\theta_{kl}\right)\right]^{1/2} \qquad (16)$$

In many instances, sites k and l will be located the same distance r_k from the center: under these conditions A_{kl} becomes

$$A_{kl} = 1/D_i[2\rho_{kl}(1 - \cos\theta_{kl})]^{1/2} \qquad (17)$$

Since $D_i << D$ it is possible to expand the expression B_{kl} in increasing powers of $\delta = D_i/D$, retaining only terms up to δ^2. One may also use the following relation involving Legendre polynomials[11]

$$\sum_{n=0}^{\infty}\rho_{kl}^n P_n(\cos\theta_{kl}) = 1/(1 - 2\rho_{kl}\cos\theta_{kl} + \rho_{kl}^2)^{1/2} \qquad (18)$$

and by integration with respect to ρ_{kl}

$$\sum_{n=0}^{\infty}\frac{1}{n+1}\rho_{kl}^n P_n(\cos\theta_{kl}) = \frac{1}{\rho_{kl}}\ln\left[\frac{(1 - 2\rho_{kl}\cos\theta_{kl} + \rho_{kl}^2)^{1/2} + \rho_{kl} - \cos\theta_{kl}}{1 - \cos\theta_{kl}}\right] \qquad (19)$$

whence

$$B_{kl} = \frac{1 - 2\delta + 2\delta^2}{D_i(1 - 2\rho_{kl}\cos\theta_{kl} + \rho_{kl}^2)^{1/2}} + \frac{\delta - 3\delta^2}{D_i\rho_{kl}}\ln\left[\frac{(1 - 2\rho_{kl}\cos\theta_{kl} + \rho^2{}_{kl})^{1/2} + \rho_{kl} - \cos\theta_{kl}}{1 - \cos\theta_{kl}}\right] + \frac{\delta^2}{D_i}\sum_{n=0}^{\infty}\frac{\rho_{kl}^n P_n(\cos\theta_{kl})}{(n+1)^2} \qquad (20)$$

(11) E.g., J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., New York, N. Y.

The series occurring in equation 20 converges rapidly, so that both A_{kl} and B_{kl} are readily computed.

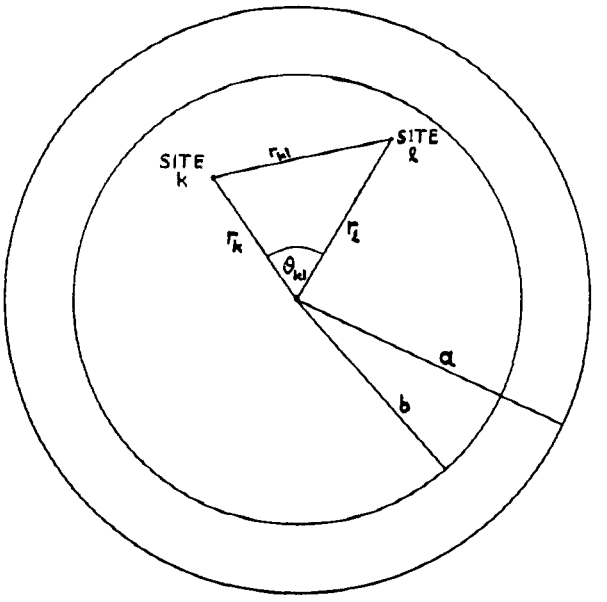

Fig. 1.—Model of protein molecule. The points k and l represent the sites of two of the titratable groups. The dielectric constant is D_i within the radius b, and D outside it. Salt ions in the solvent cannot penetrate within the radius a.

When a pair of charges is located at the same distance r_k from the center, $A_{kl} - B_{kl}$ depends only on ρ_{kl}, θ_{kl} and δ. Table I shows calculations applicable to this condition. The values apply to aqueous solutions at 25° since D_i has been placed equal to 78.5 δ. At any other temperature (or for another solvent), with dielectric constant D, the values must be multiplied by (78.5/D). For interpolation it is convenient to plot $1/(A_{kl} - B_{kl})$ *versus* cos θ_{kl}, a plot which shows relatively little curvature.

Where all charges are not located the same distance from the center the sum $A_{kl} - B_{kl}$ must be corrected by the difference in A_{kl} between equations 16 and 17.

It will be noted from Table I that $A_{kl} - B_{kl}$ and, hence, the work of charging, is markedly sensitive to ρ_{kl}, and when $\rho_{kl} < 1$, markedly sensitive to the choice of δ (*i.e.*, D_i).

It is not possible to effect much simplification to equation 8 for C_{kl}. It should be noted, however, that the term $x/(1 + x)D$ is the most important one at low ionic strength. Since $\sum_{k=1}^{m}\sum_{l=1}^{m}\xi_k\xi_l = Z^2$, this means that, at very low ionic strength, the contribution of this term to W becomes $-Z^2\epsilon^2\kappa a/2Da\,(1+\kappa a)$. This is identical with the ionic-strength dependent term of equation 2, *i.e.*, at very low ionic strength the ionic strength effect is the same as for the smeared charge model.

Figure 2 shows a plot of C_{kl} (with $D = 78.5$) *versus* x, for values of x up to 1.0. Calculations for $x > 1.0$ are laborious and have not been made for the present. (A value of $x = \kappa a = 1.0$ corresponds, for the smallest proteins, to an ionic strength of about

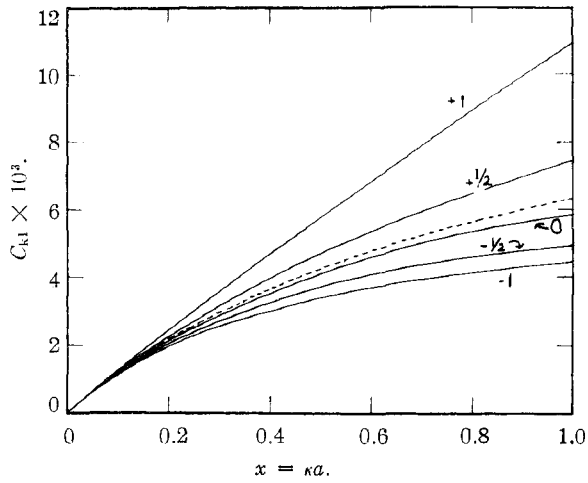

Fig. 2.—C_{kl} as a function of ionic strength in water at 25°. The solid lines are labeled to indicate the values of cos θ_{kl} to which they apply. The dotted line represents the function, $x/D(1 + x)$.

0.04.) The calculations are based on $\delta = 0.0025$, $b/a = 0.85$ and $\rho_{kl} = 1$. The result is quite insensitive to the values of δ and b/a, and depends only slightly on ρ_{kl} (note that $\sigma_{kl}^{1/2} = \rho_{kl}^{1/2} b/a$). The dependence is in such direction as to bring C_{kl} even closer to $x/(1 + x)D$ when $\rho_{kl} < 1$.

One may conclude from Fig. 2 that, where there is a more or less uniform distribution of the values of cos θ_{kl} over the possible range, the substitution of $x/(1 + x)D$ for C_{kl} is a good approximation up to $x \simeq 0.5$ with charges near the surface, and to even higher values of x for charges appreciably below the surface.

It should be noted further that, as the ionic strength increases to high values, the major contribution to the ionic strength-dependent term will come from terms with cos θ_{kl} close to unity. In most instances only the self-energy terms (cos $\theta_{kl} = 1$) will have cos θ_{kl} close to unity, *i.e.*, at high ionic strength the effect of ionic strength becomes largely that caused by the excess chemical potential of the individual charges.

TABLE I

VALUES OF $A_{kl} - B_{kl}$ IN WATER AT 25°, FOR $r_k = r_l$

$\rho_{kl}^{1/2} = \frac{r_k}{b}$	cos θ_{kl}	$A_{kl} - B_{kl}$: $\delta = 0.025$ ($D_l \simeq 2$)	$\delta = 0.050$ ($D_l \simeq 4$)	$\delta = 0.125$ ($D_l \simeq 10$)
	−1	0.0040	0.0041	0.0043
1	−0.5	.0050	.0051	.0054
	0	.0069	.0070	.0072
	+0.5	.0116	.0116	.0118
	+0.9	.0351	.0349	.0342
	−1	0.0064	0.0057	0.0054
	−0.5	.0084	.0073	.0068
0.9	0	.0126	.0105	.0094
	+0.5	.0258	.0197	.0161
	+0.9	.1580	.0972	.0598
	−1	0.0135	0.0097	0.0076
	−0.5	.0189	.0131	.0098
0.8	0	.0307	.0201	.0140
	+0.5	.0710	.0429	.0261
	+0.9	.4519	.2416	.1148
	−1	0.1109	0.0600	0.0296
	−0.5	.1539	.0821	.0392
0.5	0	.2383	.1252	.0574
	+0.5	.4683	.2415	.1054
	+0.9	1.6450	.8314	.3432

The Hypothetical Discharged State.—In the following discussion we shall have occasion to refer to an hypothetical discharged state of any protein molecule having positive and negative charges at specified locations. This state is defined such that all bond energies remain the same, all protons remain attached to their original positions. Only the electrostatic charges are removed.[12] Going from the actual state of the protein molecule to the hypothetical discharged state at the same ionic strength clearly involves an amount of work equal to **W** as given in equation 13.

Intrinsic Free Energies

In the following section we shall compute the free energy change resulting from the addition of protons to basic sites of a protein molecule. Part of this free energy will clearly be electrostatic and related to **W** of the preceding section. Another part, however, must be a chemical free energy change closely related to the free energy change of similar reactions in model compounds, *i.e.*, reactions such as R–NH_2 + H^+ → R–NH_3^+; R–COO^- + H^+ → R–COOH; where R represents a group of atoms approximating as closely as possible the immediate vicinity of the corresponding group on the protein molecule.

Accordingly we define for each site on the protein an *intrinsic standard free energy change*, $(\Delta F^0_{int})_k$ as the standard free energy change at zero ionic strength for the reaction: basic form of kth site attached to a protein molecule *with all other sites discharged* + H^+ → acid form of the kth site attached to a protein molecule *with all other sites discharged*. Thus the reaction for cationic sites is

$$P'N + H^+ \longrightarrow P'NH^+ \qquad (I)$$

where P′ represents the otherwise discharged protein molecule, while for anionic sites

$$P'O^- + H^+ \longrightarrow P'OH \qquad (II)$$

The standard states for the protein species in these reactions are the usual "hypothetical ideal one molal" states. That for H^+ is the state of unit activity as defined by the *p*H measurement. Since all other sites are discharged, and since our model allows none but coulombic interaction, $(\Delta F^0_{int})_k$ must be independent of whether other sites are in their acidic or basic form.

The definition of $(\Delta F^0_{int})_k$ here given has been chosen because it can be expected to approximate most closely the standard free energy changes in model compounds. It should be noted that the change in charge type on proton addition is different in model cationic and anionic compounds, and that

(12) If the present treatment were not confined to infinite dilution of the protein solute we should have to worry at this stage about counter-ions which would have to be discharged along with the protein unless the net charges were zero. An infinite dilution these counter-ions are infinitely dispensed and spend only an infinitesimal fraction of the time in the vicinity of the protein molecule.

this difference is preserved in the definition of $(\Delta F^0_{int})_k$.

We next define a quantity ΔF_k^* as the difference between the free energy of a protein molecule having a proton on site k, in its hypothetical, completely discharged form, and a similarly discharged molecule not having a proton on site k, *i.e.*, in terms of reactions I or II, $\Delta F_k^* = F(P'NH) - F(P'N)$ or $\Delta F_k^* = F(P'OH) - F(P'O)$. We can express $(\Delta F^0_{int})_k$ in terms of ΔF_k^* by discharging the protein on the right side of reaction I or that on the left side of reaction II. Using equation 13 and recalling that reactions I and II occur at zero ionic strength we get, for a cationic site

$$(\Delta F^0_{int})_k = \Delta F_k^* + \frac{\epsilon^2}{2b}(A_{kk} - B_{kk}) - \mu^0_{H^+} \quad (21)$$

and for an anionic site

$$(\Delta F^0_{int})_k = \Delta F_k^* - \frac{\epsilon^2}{2b}(A_{kk} - B_{kk}) - \mu^0_{H^+} \quad (22)$$

where $\mu^0_{H^+}$ is the standard chemical potential of an hydrogen ion, which is part of $(\Delta F^0_{int})_k$ but not part of ΔF_k^*.

Since our model takes into account no interaction other than coulombic interaction between charged sites, and between charged sites and external salt ions, we can conclude that ΔF_k^* must have identically the same value for all sites of a particular chemical kind. ΔF_k^* must also be independent of ionic strength.[13]

It is apparent from equation 21 and 22 that, although ΔF_k^* is the same for all sites of the same chemical kind, $(\Delta F^0_{int})_k$ is not necessarily the same for all such sites. The term A_{kk} is the same, but B_{kk} as given by equation 7 or 20, with $\cos\theta_{kl} = 1$ clearly depends on the depth of the site within the molecule, as given by ρ_{kk}.[14]

In all of the calculations based on the present paper we shall place all sites of the same kind at the same distance from the center of the sphere. In that event $(\Delta F^0_{int})_k$ will be the same for all sites of the same kind. If we suppose that of the m sites on the molecule m_j ($j = 1, 2, \ldots$) are of a particular kind (*e.g.*, phenolic groups), then each of these m_j sites will have the same value of $(\Delta F^0_{int})_k$, which we shall call $(\Delta F^0_{int})_j$.

Model compounds are generally characterized by their acid dissociation constants. Since $(\Delta F^0_{int})_j$ refers to the corresponding *association* reaction, we define an intrinsic *dissociation* constant $(K_{int})_j$ for the jth kind of group by the relation

$$(\Delta F^0_{int})_j = -2.303kT(pK_{int})_j \quad (23)$$

(13) The principal non-coulombic factor likely to occur in proteins, which might in fact produce differences in ΔF^* between different sites of the same kind, is hydrogen bonding; *cf.* ref. 6.

(14) It is possible by varying ρ_{kk} to change B_{kk} by as much as 10,000 to 100,000 calories per mole of reaction. This would correspond to a change in *intrinsic* pK of from 7 to 70! It is, of course, an oversimplification to consider reaction II as the disappearance of a pair of discrete charges. Instead, the reaction should be considered as the formation of an ^-O–H^+ dipole. However, the work released in forming such a dipole also depends critically on the depth below the surface. The fact that intrinsic pK values of ionizable groups on protein molecules normally correspond closely to those to be expected from model compounds suggests that these groups in proteins are located at the same depth below the cavity surface as in model compounds.

Total Free Energy

A protein molecule can occur in numerous forms, depending on how many protons are bound to it, and on where they are. The titration curve is a measure of the equilibrium between these forms as a function of pH, and we therefore need to know the relative total free energy of the various forms.

The symbol P will be used to represent a protein molecule from which all dissociable protons have been removed. It will bear a net negative charge equal to the number of anionic sites. Similarly PH_m represents a molecule with no dissociated protons, bearing a net positive charge equal to the number of cationic sites. Both P and PH_m represent unique forms with unique configurations.

In general, $PH\nu$ will be used to represent a molecule containing ν dissociable protons. These ν protons may be attached to different kinds of sites; we shall use ν_j as the number of protons on a given kind of site, and shall distinguish between various species $PH_\nu^{(i)}$ ($i = 1, 2, \ldots$) which differ in the various ν_j, but which have the same value of $\nu = \Sigma\nu_j$.[15]

It must further be recognized that each form $PH_\nu^{(i)}$ can exist in a large number $\Omega_\nu^{(i)}$ of possible configurations, arising from the fact that each set of ν_j protons may be distributed over m_j sites of the same kind. $\Omega_\nu^{(i)}$ is given by the relation

$$\Omega_\nu^{(i)} = \prod_j \frac{m_j!}{\nu_j!(m_j - \nu_j)!} \quad (24)$$

To calculate the titration curve at any ionic strength we need to know the relative total free energies of all of the possible species $PH_\nu^{(i)}$ at that ionic strength. It will be convenient to calculate all free energies $F_\nu^{(i)}$ relative to the free energy F_0 of the form P, at the same ionic strength. In so doing we must recognize that the work of charging of $PH_\nu^{(i)}$ will be different for each of the possible configurations, so that $F_\nu^{(i)}$ will also depend on the configuration. (To clarify the subsequent discussion it may be worthwhile to repeat here that the various ν_j define the form $PH_\nu^{(i)}$; a specific set of values of the ξ_k compatible with the chosen ν_j defines a configuration.)

We first convert P to its hypothetical discharged state at the same ionic strength. By the equation 13, the resulting free energy change is

$$-\frac{\epsilon^2}{2b}\sum_{k=1}^{m}(\xi_k^2)_0(A_{kk} - B_{kk}) - W_0$$

the subscript zero indicating that the values of ξ_k to be used are those appropriate to the form P (*i.e.*, $\nu = 0$).

We next compute the difference in free energy between $PH_\nu^{(i)}$ and P, both in their hypothetical discharged state. This is clearly equal to $\sum_j \nu_j \Delta F_j^*$.

Next we convert $PH_\nu^{(i)}$ to its actual charged form with the free energy change

$$\frac{\epsilon^2}{2b}\sum_{k=1}^{m}(\xi_k^2)_\nu^{(i)}(A_{kk} - B_{kk}) + W_\nu^{(i)}$$

(15) In practice we can ignore the vast majority of possible species $PH_\nu^{(i)}$. For instance, any form having protons on carboxyl groups and not on amino groups or phenolic groups can be neglected because its stability is far too low to allow it to be present to an appreciable extent.

with the values of ξ_k appropriate to the particular configuration of $PH_\nu^{(i)}$, which we are considering. Adding these contributions we get

$$F_\nu^{(i)} - F_0 = \sum_j \nu_j \Delta F_j^* + \frac{\epsilon^2}{2b} \sum_{k=1}^{m} [(\xi_k^2)_\nu^{(i)} - (\xi_k^2)_0](A_{kk} - B_{kk}) + W_\nu^{(i)} - W_0 \quad (25)$$

Next we wish to convert the ΔF_j^* into the corresponding $(\Delta F^0_{int})_j$ by equation 21 or 22, and we observe that, in so doing, the term $(\epsilon^2/2b)(A_{kk}-B_{kk})$ enters in a different way for cationic and anionic sites. For cationic sites (all of which have $(\xi_k)_0 = 0$) such a term is to be subtracted from $(\Delta F^0_{int})_j$ for each site which has $\xi_k = 1$ in the chosen configuration of $PH_\nu^{(i)}$. For anionic sites (all of which have $(\xi_k)_0 = -1$) such a term is to be added to $(\Delta F^0_{int})_j$ for each site which has $\xi_k = 0$ in the chosen configuration of $PH_\nu^{(i)}$. This can be achieved by adding this term for each anionic site in the form P, and subtracting it again for each site still having $\xi_k = 1$ in the chosen configuration of $PH_\nu^{(i)}$. Thus

$$\sum_j \nu_j \Delta F_j^* = \sum_j \nu_j (\Delta F^0_{int})_j - \frac{\epsilon^2}{2b} \sum_{k=1}^{m} [(\xi_k^2)_\nu^{(i)} - (\xi_k^2)_0](A_{kk} - B_{kk}) + \nu\mu^0_{H^+} \quad (26)$$

Combining equations 25 and 26, the self-energy terms vanish and

$$F_\nu^{(i)} - F_0 = \sum_j \nu_j (\Delta F^0_{int})_j + W_\nu^{(i)} - W_0 + \nu\mu^0_{H^+} \quad (27)$$

Of the terms on the right-hand side of equation 27 only $W_\nu^{(i)}$ is configuration dependent. To compute the average over all configurations of $F_\nu^{(i)} - F_0$, it is necessary only to evaluate the average contribution of $W_\nu^{(i)}$ to the free energy. To do this we first evaluate the partition function

$$f_\nu^{(i)} = \sum_c e^{-W_\nu^{(i)}/kT} \quad (28)$$

where the summation extends over all configurations. The average contribution to the free energy is then

$$\overline{W_\nu^{(i)}} = -kT \ln f_\nu^{(i)} = -kT \ln \sum_c e^{-W_\nu^{(i)}/kT} \quad (29)$$

If each of the $\Omega_\nu^{(i)}$ configurations had the same energy then $\overline{W_\nu^{(i)}}$ would simply reduce to $-kT \ln \Omega_\nu^{(i)} + W_\nu^{(i)}$. Since this will not in general be true, some other way must be found to simplify the sum on the right-hand side of equation 29.

We shall use here a procedure first employed by Kirkwood in dealing with coöperative phenomena in metal alloys.[16] We define an energy $W_\nu^{(i)\prime}$ such that

$$-kT \ln \sum_c e^{-W_\nu^{(i)}/kT} = -kT \ln \Omega_\nu^{(i)} + W_\nu^{(i)\prime} \quad (30)$$

We next expand the exponentials in equation 30 so that the left-hand side becomes

$$-kT \ln \left\{ \sum_c 1 - \sum_c W_\nu^{(i)}/kT + (1/2!) \sum_c (W_\nu^{(i)})^2/(kT)^2 - \ldots \right\}$$

Since

$$\sum_c 1 = \Omega_\nu^{(i)}$$

we may add $kT \ln \Omega_\nu^{(i)}$ to each side, and obtain

$$W_\nu^{(i)\prime} = -kT \ln \left\{ 1 - \frac{\langle W \rangle}{kT} + \frac{1}{2!} \frac{\langle W^2 \rangle}{(kT)^2} - \ldots \right\} \quad (31)$$

where

$$\langle W \rangle = \sum_c W_\nu^{(i)}/\Omega_\nu^{(i)}$$
$$\langle W^2 \rangle = \sum_c (W_\nu^{(i)})^2/\Omega_\nu^{(i)}, \text{ etc.} \quad (32)$$

Expanding the logarithm occurring in equation 31 in ascending powers of $(1/kT)$ we finally obtain

$$W_\nu^{(i)\prime} = kT \sum_{n=1}^{\infty} \frac{M_n}{n!} \left(\frac{1}{kT}\right)^n \quad (33)$$

where

$$M_1 = \langle W \rangle$$
$$M_2 = \langle W \rangle^2 - \langle W^2 \rangle \quad (34)$$
$$M_3 = \langle W^3 \rangle - 3\langle W^2 \rangle \langle W \rangle + 2\langle W \rangle^3, \text{ etc.}$$

and, by combination with equations 29, 30 and 27, this gives for the average free energy over all states

$$\overline{F_\nu^{(i)}} - F_0 = \sum_j \nu_j (\Delta F^0_{int})_j + \nu\mu^0_{H^+} - kT \ln \Omega_\nu^{(i)} + kT \sum_{n=1}^{\infty} \frac{M_n}{n!} \left(\frac{1}{kT}\right)^n - W_0 \quad (35)$$

Essentially, then, the calculation of the free energy reduces to the calculation of the M_n of equation 34, and, hence, the averages of equation 32. How difficult this calculation is depends on the number of dissociable sites of the protein molecule and on the number of terms retained in equation 33. If one retains only the first term the calculation becomes quite simple, and still represents an approximation considerably superior to the smeared model discussed in the introduction. In the calculations of the following paper we shall retain two terms in the expansion; most of the time the second term will be found to make a relatively small contribution, so that neglect of terms higher than the second is probably hardly ever of any importance.

The Titration Curve

It is now possible to write down the standard free energy change $\Delta F^0_\nu{}^{(i)}$ for the reaction

$$P + \nu H^+ \longrightarrow PH_\nu^{(i)} \quad (III)$$

for

$$\Delta F^0_\nu{}^{(i)} = \overline{F_\nu^{(i)}} - F_0 - \nu\mu^0_{H^+}$$

i.e.

$$\Delta F^0_\nu{}^{(i)} = \sum_j \nu_j (\Delta F^0_{int})_j - kT \ln \Omega_\nu^{(i)} + kT \sum_{n=1}^{\infty} \frac{M_n}{n!} \left(\frac{1}{kT}\right)^n - W_0 \quad (36)$$

We now introduce $k_\nu^{(i)}$, the association constant for reaction III, *i.e.*

$$k_\nu^{(i)} = (PH_\nu^{(i)})/(P)a^\nu_{H^+} \quad (37)$$

where brackets represent relative concentrations as the total protein concentration approaches zero. This constant is clearly related to $\Delta F^0_\nu{}^{(i)}$

$$\Delta F^0_\nu{}^{(i)} = -kT \ln k_\nu^{(i)} \quad (38)$$

(16) J. G. Kirkwood, *J. Chem. Phys.*, **6**, 70 (1938). A detailed account of this method is given by T. der Haar, "Elementary Statistical Mechanics," Rinehart and Co., New York, N. Y., 1954.

The titration curve itself is a plot of the average number of bound protons, $\bar{\nu}$ *versus* pH, or, more often, the average number of dissociated protons, $m - \bar{\nu}$, *versus* pH. It is obtained at once from equation 37

$$\bar{\nu} = \frac{\sum_{\nu=0}^{m} \sum_{(i)} \nu(PH_\nu^{(i)})}{\sum_{\nu=0}^{m} \sum_{(i)} (PH_\nu^{(i)})} = \frac{\sum_{\nu=1}^{m} \left(\sum_{(i)} k_\nu^{(i)} \right) \nu a^\nu_{H^+}}{1 + \sum_{\nu=1}^{m} \left(\sum_{(i)} k_\nu^{(i)} \right) a^\nu_{H^+}} \quad (39)$$

where $\sum_{(i)}$ represents summation over all the species giving the same value of ν. As previously noted, only a few $k_\nu^{(i)}$ normally make an appreciable contribution to $\sum_{(i)} k_\nu^{(i)}$.

In many experiments (*e.g.*, spectrophotometric titration of the phenolic groups) one obtains $\bar{\nu}_j$ for a particular kind of group rather than $\bar{\nu}$. This is obtained as

$$\bar{\nu}_j = \frac{\sum_{\nu=1}^{m} \left(\sum_{(i)} k_\nu^{(i)} \nu_j \right) a^\nu_{H^+}}{1 + \sum_{\nu=1}^{m} \left(\sum_{(i)} k_\nu^{(i)} \right) a^\nu_{H^+}} \quad (40)$$

It should be noted that whereas ν in equation 39 is the same for each form $PH_\nu^{(i)}$, ν_j will in general be different for each species $PH_\nu^{(i)}$.

Discussion

The general approach of this paper has been used recently in a number of investigations by Hill.[17-19] In two of these[17,19] are tabulated values of the effective dielectric constant D_E for the interaction between a pair of charges. These are simply related[10] to the functions A_{kl}, B_{kl} and C_{kl} of the present paper.

Two of Hill's papers discuss, among other matters, the protein titration curves. In one of these[18] it was shown that closely spaced pairs of sites would be expected to lead to large deviations from the simple model discussed in the introduction. In the other[19] Hill calculated titration curves for a model which assumed random distribution of the various kinds of dissociable sites over fixed positions. In the model used in this paper each fixed position has been assigned to a particular kind of dissociable site. Hill's calculation takes into account the interaction of each site with nearest neighbors only, which, for proteins, is probably a poor approximation. Both of Hill's studies, in so far as they apply to impenetrable spherical proteins, could be obtained as special results of the treatment of this paper.

Linderstrøm-Lang[20] recently has calculated the effect of ionic strength on the chemical potential of spherical multipolar ions, using the same procedure we have used. This means, in the terminology of this paper, that he has calculated the sum

$$(\epsilon^2/2a) \sum_k \sum_l \xi_k \xi_l C_{kl}$$

His most interesting application is to ions with a net charge of zero. Here the contributions of the first term in C_{kl} (equation 8), *i.e.*, $x/(1+x)\ D$, cancel, and the remainder of equation 8 becomes important. In contrast, we shall generally be interested in ions for which the net charge is not zero, and here $x/(1+x)D$ is at low ionic strength the predominant term in C_{kl}.

Harris and Rice[21] have applied to flexible polyelectrolytes essentially the same ideas as are here applied to proteins. Their treatment is simpler than that given here in that they have assumed values for the effective dielectric constant for pairwise interaction. It is also more complicated in that the charges of flexible polyelectrolytes are not at fixed distances with respect to one another.

It would be desirable to extend the treatment of this paper to ellipsoidal ions. To do this one needs to know **W** for such ions. In the absence of added electrolyte the expression for **W** given by Hill[22] is applicable. There is, however, no general relation applicable to finite ionic strength. Linderstrøm-Lang[20] has worked out in great detail, with numerical tabulations, the ionic strength dependent portion of **W** for ellipsoidal ions with charges located along the major axis of the ellipsoid, but this result is inapplicable to globular protein ions.

It should be noted, finally, that the results of the present paper can be extended without additional difficulty to the reaction of proteins with any kind of small ion. In this connection we should mention the calculations made several years ago by Schellman,[23] which showed that the intrinsic association constants for small ion binding should depend on the depth of the binding sites below the molecular surface. His arguments follow in principle the same lines used in this paper in discussing ΔF^0_{int} for hydrogen ion dissociation.

NEW HAVEN, CONN.

(17) T. L. Hill, *J. Phys. Chem.*, **60**, 253 (1956).
(18) T. L. Hill, THIS JOURNAL, **78**, 3330 (1956).
(19) T. L. Hill, *ibid.*, **78**, 5527 (1956).
(20) K. Linderstrøm-Lang, *Compt. rend. trav. Lab. Carlsberg*, **28**, 281 (1953).
(21) F. E. Harris and S. A. Rice, *J. Phys. Chem.*, **58**, 725, 733 (1954); *J. Chem. Phys.*, **24**, 326 (1956); **25**, 955 (1956).
(22) T. L. Hill, *ibid.*, **12**, 147 (1944).
(23) J. A. Schellman, *J. Phys. Chem.*, **57**, 472 (1953).