[CONTRIBUTION NO. 1438 FROM STERLING CHEMISTRY LABORATORY, YALE UNIVERSITY]

Theory of Protein Titration Curves. II. Calculations for Simple Models at Low Ionic Strength

BY CHARLES TANFORD[1]

RECEIVED MAY 8, 1957

The equations of the preceding paper are used to calculate titration curves for simple model molecules. It is found that the location of charged sites with respect to the molecular surface is a crucial variable. Only by placing charged sites about 1 Å. below the surfaces can titration curves be obtained with a slope of the order of magnitude observed experimentally. If the depth of charged sites below the surface is kept fixed at 1 Å. considerable variation in calculated titration curves may still arise as a result of the distribution of sites with respect to one another. It is concluded that many of the "anomalies" of experimental titration curves, including apparently anomalous dissociation constants, can be accounted for on the basis of electrostatic interaction alone.

The theory presented in the preceding paper[2] can be used to compute the titration curve of any compactly folded protein with molecules of approximately spherical shape, provided the configuration of the molecule is known in detail. At the present time no such detailed configurations are known, and this paper will therefore be confined to calculations for simple models which show configurational features which one may expect to find in actual proteins.

Titration Curves at Zero Ionic Strength. Method of Calculation

The equilibrium constant $k_\nu^{(i)}$ for the formation of a protein species $PH_\nu^{(i)}$ containing ν dissociable protons, distributed in a particular way, such that ν_j protons are attached to sites of a particular variety, can be calculated from the relation (eq. I-23, I-36, I-38)

$$\log k_\nu^{(i)} = \sum_j \nu_j (pK_{\text{int}})_j + \log \Omega_\nu^{(i)} - \sum_{n=1}^{\infty} \frac{M_n}{2.303 n!} \left(\frac{1}{kT}\right)^n + \frac{W_0}{2.303\, kT} \quad (1)$$

In this equation $(K_{\text{int}})_j$ is the intrinsic dissociation constant for the jth kind of site, $\Omega_\nu^{(i)}$ is the number of possible configurations of $PH_\nu^{(i)}$, W_0 (eq. I-14) is the electrostatic work, exclusive of self-energies, for charging a protein molecule from which all dissociable protons have been dissociated. The M_n are parameters involving the unweighted averages of $W_\nu^{(i)}$, $(W_\nu^{(i)})^2$, etc., over all possible configurations, $W_\nu^{(i)}$ being the electrostatic work, exclusive of self-energies, for charging the protein species $PH_\nu^{(i)}$. The calculation of these averages represents the only real problem in the evaluation of the $k_\nu^{(i)}$ and, hence, of the titration curve.

In this paper we retain only the first two of the M_n in eq. 1. Confining ourselves at first to titration curves *at zero ionic strength*, so that the C_{kl} in eq. I-14 vanish, we need to evaluate the following terms occurring in M_1 and M_2

$$\frac{\langle W \rangle}{kT} = \sum_{k=1}^{m} \sum_{l \neq k} \left(\sum_c \frac{\xi_k \xi_l}{\Omega_\nu^{(i)}} \right) \varphi_{kl} \quad (2)$$

$$\frac{\langle W^2 \rangle}{kT} = \sum_{k=1}^{m} \sum_{l \neq k} \sum_{k'=1}^{m} \sum_{l' \neq k'} \left(\sum_c \frac{\xi_k \xi_l \xi_{k'} \xi_{l'}}{\Omega_\nu^{(i)}} \right) \varphi_{kl} \varphi_{k'l'} \quad (3)$$

where $\sum_c$ represents summation over all possible configurations of $PH_\nu^{(i)}$, and ξ_k is the charge at the kth site in a particular configuration, being always +1 or 0 for a cationic site and −1 or 0 for an anionic site. Also

$$\varphi_{kl} = \frac{\epsilon^2}{2bkT} (A_{kl} - B_{kl}) \quad (4)$$

where b is the radius of the spherical molecule and $A_{kl} - B_{kl}$ is the quantity tabulated in Table I of the preceding paper.

We consider first the calculation of the occupation variables $\sum_c \xi_k \xi_l / \Omega_\nu^{(i)}$ and $\sum_c \xi_k \xi_l \xi_{k'} \xi_{l'} / \Omega_\nu^{(i)}$. These parameters are entirely independent of the model chosen, depending only on the number of sites m_j of each kind of group, and on the number μ_j of these which bear a charge.[3]

In this paper we make calculations for models which consist of only two kinds of sites, one anionic and one cationic. Let k = c, c′, c″, etc., represent sites of one kind ($j = 1$), and k = n, n′, n″, etc., sites of the second kind ($j = 2$). In computing $\sum_c \xi_k \xi_l / \Omega_\nu^{(i)}$ we then have pairs of three types only: k = c, l = c′; k = n, l = n′; k = c, l = n. For the first two types $\xi_k \xi_l$ may be +1 or 0, for the third type it may be −1 or 0. For each particular pair of sites, then, $\sum_c \xi_k \xi_l$ is just numerically equal to the number of configurations with $\xi_k \xi_l$ not equal to zero, and this is just the number of configurations which have charges at each of the two sites k and l.

For instance, for a particular pair c and c′, this number is obtained as the number of ways of distributing $\mu_1 - 2$ charges over $m_1 - 2$ sites and μ_2 charges over m_2 sites, *i.e.*, it is

$$(m_1 - 2)!\, m_2! / (m_1 - 2 - \mu_1)!(\mu_1 - 2)!(m_2 - \mu_2)!\, \mu_2!$$

Combining with eq. I-24 for $\Omega_\nu^{(i)}$ we get at once

$$\sum_c \xi_c \xi_{c'} / \Omega_\nu^{(i)} = \frac{\mu_1(\mu_1 - 1)}{m_1(m_1 - 1)}$$

In the same way we get

$$\sum_c \xi_n \xi_{n'} / \Omega_\nu^{(i)} = \frac{\mu_2(\mu_2 - 1)}{m_2(m_2 - 1)}$$

$$\sum_c \xi_c \xi_n / \Omega_\nu^{(i)} = -\frac{\mu_1 \mu_2}{m_1 m_2}$$

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

(2) C. Tanford and J. G. Kirkwood, THIS JOURNAL, **79**, 5333 (1957). Equations from this paper will be referred to by the prefix I, e.g., I-23 is eq. 23 of the preceding paper.

(3) In terms of the ν_j of eq. 1, $\mu_j = \nu_j$ for cationic sites and $\mu_j = m_j - \nu_j$ for anionic sites.

In computing $\sum_c \xi_k\xi_l\xi_{k'}\xi_{l'}/\Omega_\nu^{(i)}$ we note from eq. 3 that $k \neq l$ and $k' \neq 1'$. Each set of possible values of k, l, 1', 1' thus represents two pairs of sites. The two pairs may be identical (k' and 1' the same as k and l) in which case the configurations which have $\xi_k\xi_l\xi_{k'}\xi_{l'} = \pm 1$ are the same as those which have $\xi_k\xi_l = \pm 1$, and $\sum_c \xi_k\xi_l\xi_k\xi_{l'} = \sum_c \xi_k\xi_l$. If the pairs are different, but share one site then

$$\sum_c \xi_c\xi_{c'}\xi_c\xi_{c''}/\Omega_\nu^{(i)} = \frac{\mu_1(\mu_1-1)(\mu_1-2)}{m_1(m_1-1)(m_1-2)}$$

$$\sum_c \xi_c\xi_{c'}\xi_c\xi_n/\Omega_\nu^{(i)} = -\frac{\mu_1(\mu_1-1)\mu_2}{m_1(m_1-1)m_2}, \text{ etc.}$$

or, if all four sites are different

$$\sum_c \xi_c\xi_{c'}\xi_{c''}\xi_{c'''}/\Omega_\nu^{(i)} = \frac{\mu_1(\mu_1-1)(\mu_1-2)(\mu_1-3)}{m_1(m_1-1)(m_1-2)(m_1-3)}$$

$$\sum_c \xi_c\xi_{c'}\xi_n\xi_{n'}/\Omega_\nu^{(i)} = \frac{\mu_1(\mu_1-1)\mu_2(\mu_2-1)}{m_1(m_1-1)m_2(m_2-1)}, \text{ etc.}$$

Consider, for example, the titration of four carboxyl groups on a molecule also containing four amino groups. The latter may be supposed positively charged throughout the titration and we are therefore concerned with five different forms of the protein: $\nu_1 = 0, 1, 2, 3, 4$; the corresponding values of μ_1 being 4, 3, 2, 1, 0. For the form with $\mu_1 = 4$ all occupation variables are equal to unity—there is a single configuration and all sites are charged. For the form with $\mu_1 = 0$ there is also a single configuration, all carboxyl groups being uncharged and all amino groups charged, *i.e.*, $\sum_c \xi_n\xi_{n'}/\Omega$ and all $\sum_c \xi_n\xi_{n'}\xi_{n''}\xi_{n'''}/\Omega$ are equal to unity (regardless of where n'' or n''' are equal to n or n') and any occupation variable involving a carboxyl group, *i.e.*, containing a ξ_c, must be zero. The occupation variables for all possible combinations are shown in Table I. The left-hand column of this table shows

TABLE I

OCCUPATION VARIABLES FOR 4 COOH GROUPS TITRATED ON A MOLECULE ALSO CONTAINING 4 NH_3^+ GROUPS

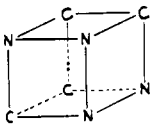
		ν_1	1	2	3
		μ_1	3	2	1
			$\sum \xi_k\xi_l/\Omega_\nu$		
CC			1/2	1/6	0
NN			1	1	1
CN			−3/4	−1/2	−1/4
			$\sum \xi_k\xi_l\xi_{k'}\xi_{l'}/\Omega_\nu$		
CC × CC	cc'cc'		1/2	1/6	0
	cc'cc''		1/4	0	0
	cc'c''c'''		0	0	0
NN × NN	All terms		1	1	1
CN × CN	c = c'		3/4	1/2	1/4
	c ≠ c'		1/2	1/6	0
CC × NN			1/2	1/6	0
CC × CN	cc'cn		−1/2	−1/6	0
	cc'c''n		−1/4	0	0
NN × CN	All terms		−3/4	−1/2	−1/4

the type of site pair (or product of pairs) being considered. The small letters c, c', c'', etc., refer to different carboxyl sites, *i.e.*, the symbol CC × CC cc' cc' means that the term considered is a product of two identical pairs of carboxyl group (k' and 1' equal to k and l), the symbol CC × CC cc' cc'' represents a product of two carboxyl pairs with k' = k or l, etc. Since all $\xi_n = 1$ it is not necessary in this example to distinguish between the NN × NN products in this way. Occupation variables in which $\xi_k\xi_l$ or $\xi_k\xi_l\xi_{k'}\xi_{l'}$ is negative in sign are so indicated.

It should be emphasized that the occupation variables of Table I are independent of the locations chosen for the sites involved. By contrast, the φ_{kl} of eq. 4 depend *only* on location, being independent of the values of ν or μ.

For a simple model there will be just a limited number of possible values of φ_{kl}. For instance, if we choose a model with eight sites at the corners of a cube, there will be 56 terms in the sum over k and l in eq. 2, but only three different values of φ_{kl} are required. There will be 24, 24 and 8 terms, respectively, with φ_{kl} equal to φ_1, φ_2 and φ_3, where φ_1 represents the value of φ_{kl} for nearest neighbors, φ_2 for sites at the ends of face diagonals, and φ_3 for sites at the ends of cube diagonals. For the same model the sum over k, l, k' and 1' in eq. 3 will have $56^2 = 3136$ terms, but there will be only six different values of $\varphi_{kl}\varphi_{k'l'}$. There will be 576 terms in φ_1^2, 576 in φ_2^2, 64 in φ_3^2, 1152 in $\varphi_1\varphi_2$, 384 in $\varphi_1\varphi_3$ and 384 in $\varphi_2\varphi_3$.

Since each term is to be multiplied by the corresponding occupation variable, it will be necessary further to subdivide these terms so as to count separately those terms of the sum with different values of the occupation variables. It is at this stage that the positions of the individual sites must be specified. Suppose we continue to consider four carboxyl groups titrated in the presence of four cationic groups which bear a positive charge throughout the *p*H region in which the carboxyl groups are being titrated. Let the eight sites be placed at the corners of a cube as shown by model A, on which N represents a cationic site and C a carboxyl group. (The points, of course, represent the location of point charges if $\xi_k = \pm 1$.) The subdivision of site pairs and their products, as required by the occupation variables of Table I, is then that shown in Table II.



Model A

It is now a simple matter to compute M_1/kT and $M_2/(kT)^2$ by multiplying each term in Table II by the appropriate occupation variable in Table I. The result of the calculation for this particular model, with suitable dimensions, is shown in Table III.[4]

The four equilibrium constants of the present model can now be calculated at once from eq. 1, and, hence, the curve for the titration of carboxyl groups can be obtained from eq. I-40.

(4) It should be noted that the maximum contribution of M_2 to any $\log k_\nu$ (*cf.* eq. 1) is about 0.17. This is small compared to the maximum contribution of all of the electrostatic terms, which reaches 3.65. The term in M_2 may *seem* to make a large contribution in the case of $\log k_1$, but this is not so because the total electrostatic effect on $\log k_1$ is very small.

TABLE II

NUMBER OF TERMS OF DIFFERENT ENERGY FOR MODEL A

(a) In evaluation of M_1

	φ_1	φ_2	φ_3
CC	6	4	2
NN	6	4	2
CN	12	16	4

(b) In evaluation of M_2

		φ_1^2	φ_2^2	φ_3^2	$\varphi_1\varphi_2$	$\varphi_1\varphi_3$	$\varphi_2\varphi_3$
CC × CC	cc'cc'	12	8	4	0	0	0
	cc'cc''	16	0	0	48	16	16
	cc'c''c'''	8	8	0	0	8	0
NN × NN	All terms	36	16	4	48	24	16
CN × CN	c = c'	40	64	8	96	16	32
	c ≠ c'	104	192	8	288	80	96
CC × NN		72	32	8	96	48	32
CC × CN	cc ' cn	64	64	0	144	64	48
	cc'c''n	80	64	16	144	32	48
NN × CN	All terms	144	128	16	288	96	96

TABLE III

ELECTROSTATIC FREE ENERGY TERMS FOR MODEL A[a,b]

ν_1	μ_1	$\frac{M_1}{kT} = \frac{<W>}{kT}$	$\frac{M_2}{(kT)^2} = \frac{<W>^2 - <W^2>}{(kT)^2}$	$\frac{M_1}{kT} + \frac{1}{2}\frac{M_2}{(kT)^2} - \frac{W_0}{kT}$
0	4	−2.648[c]	0	0
1	3	−1.987	−0.787	0.268
2	2	−0.362	− .670	1.951
3	1	+2.224	+0.105	4.767
4	0	+5.773	0	8.421

[a] For a sphere of radius 8 Å., with charges 0.8 Å. below the surface of the molecule. [b] By equation 1 the last column of this table is the contribution of all electrostatic energy terms to the various log k_ν. It will be noted that there are no sub-species in this particular model so that the $k\nu$ do not need to be sub-divided into various $k_\nu^{(i)}$. [c] Since we are considering the titration of carboxyl groups only we are in effect considering the cationic protons as not dissociable. Hence this term is W_0/kT for this model.

In reporting the results of this and similar calculations we shall not in general present the titration curve itself, but shall show instead a plot of pH − log $\alpha/(1-\alpha)$ *vs.* average net charge $\bar{Z}$. Here α is the average degree of dissociation of the type of group under consideration, *i.e.*[5]

$$\alpha = (m_i - \bar{\nu}_i)/m_i \qquad (5)$$

We have chosen this method of presentation because experimentally such plots are often close to linear: we shall find that pH − log $\alpha/(1-\alpha)$ is also close to a linear function of $\bar{Z}$ when computed by the methods here given. We therefore write

$$p\mathrm{H} - \log \alpha/(1-\alpha) = pK' - 0.868w\bar{Z} \qquad (6)$$

where pK' and w are arbitrary parameters representing the intercept (at $\bar{Z} = 0$) and the slope of such a plot.

The "smeared site" model of Linderstrøm-Lang, described in the introduction to paper I, which has been the model ordinarily used in the past for the interpretation of titration curves of proteins, predicts, for impenetrable proteins at zero ionic strength, that

$$pK' = pK_{\mathrm{int}}$$
$$w = w_0 = \epsilon^2/2DbkT \qquad (7)$$

(5) The function pH − log $\alpha/(1-\alpha)$ is the negative logarithm of an *apparent* equilibrium constant at a particular net charge and is sometimes called pK'. We are reserving the symbol pK' for the parameter defined by eq. 6.

These relations follow from eq. I-3 and I-4. Experimental results agreeing quite closely with the prediction of eq. 7 are sufficiently common so that these equations provide a natural reference to which to compare the results of calculations by the more exact equations of this paper.[6]

Calculated Results at Zero Ionic Strength

All of the models used here have their acidic and basic groups located on spheres concentric with the sphere representing the molecule so that r_k/b is the same for all sites. In order to reduce the number of energy parameters, the sites have been placed on symmetrical polyhedra inscribed within the spheres. The cube already has been discussed in the first section. To obtain a greater variety of locations for a small number of sites we have used the dodecahedron, which allows five possible inter-site distances.[7] The models have been chosen, as Table IV shows, so as to possess approximately the same ratio of available surface area to number of sites as is found in actual proteins of interest.

TABLE IV

APPROXIMATE SITE DENSITIES IN PROTEINS AND SIMPLE MODELS

(Expressed in reciprocal units of Å.² total molecular surface per ionizable site.)

Actual proteins[a]

	If spherical	If prolate ellipsoids of axial 3:1
Ribonuclease	120	145
Ovalbumin, β-lactoglobulin	80	95
Serum albumin	55	65

Spherical models of this paper

	b = 8 Å.	b = 10 Å.	b = 15 Å.
8 sites	101	157	
12 sites		105	
20 sites		63	142

[a] In making calculations for actual proteins about 20% by weight of hydration has been assumed. In counting ionizable sites the phenolic groups have been omitted because most of the model calculations are for pH ranges in which phenolic groups are not ionized, so that their presence would have no effect.

Identical Sites.—Figures 1 and 2 show calculated logarithmic plots for eight identical sites distributed in various ways over a spherical protein molecule of radius 10 Å. The sites have been taken to be carboxyl groups with an assumed pK_{int} of 4.60. For any other type of site identical curves would have been obtained except for displacement of the ordinate so that the new pK_{int} appears in place of 4.60.

(6) Experimental titration curves are, of course, measured at finite ionic strength: the lowest ionic strength reached has generally been about 0.01. The data at zero ionic strength can, however, be obtained by extrapolation with the aid of eq. I-3 and I-4. A fuller discussion of the effect of ionic strength will be given below.

(7) In actual proteins this distance (or the value of θ_{kl}) will be different for each pair. However, the interaction energy decreases rapidly with the distance between sites. Thus the calculation can be simplified by using the exact value of θ_{kl} only for near neighbors of each site. Those at greater distance can be grouped together. For instance all pairs with $0.16 > \theta_{kl} > -0.16$ can be assigned $\theta_{kl} = 0$; all pairs with $-0.16 > \theta_{kl} > -0.5$ can be assigned $\theta_{kl} = -0.33$, etc. The error introduced by such a procedure should be less than experimental uncertainty.

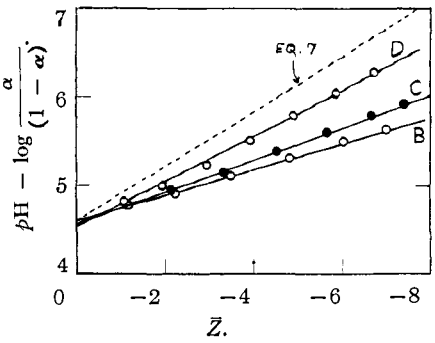
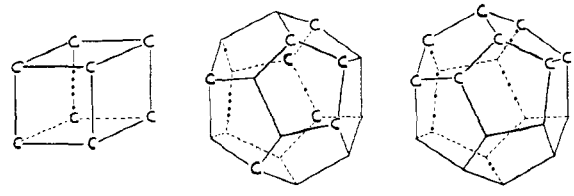


Fig. 1.—Logarithmic plots for eight -COOH groups *at the surface* of a sphere of radius 10 Å. Sites distributed according to models B. C and D.

There are several reasons for choosing models with sites at or near the surface. These are the ready accessibility of most acidic sites in the common small proteins; the fact that these groups have intrinsically about the same dissociation properties as similar groups in small molecules; and the fact that eq. 7, which, among other assumptions, require the charges to be on the surface, accounts at least moderately well for many experimental results. Thus Fig. 1 shows calculated curves for eight carboxyl groups on the protein surface. Three distributions were chosen: a regular distribution (model B), a random distribution (model C), and a crowded distribution (model D). (The cube and



Model B Model C Model D

dodecahedra are inscribed in a sphere equal in radius to that of the protein molecule.) In each case a linear plot is obtained,[8] as described by eq. 6, with $pK' \simeq pK_{int}$ as predicted by eq. 7. However, the observed values of w depend on the distribution of the sites, and, moreover, are all smaller than w_0. For model B, $w < 0.5\ w_0$; for model C, $w \simeq 0.6\ w_0$; for model D, $w \simeq 0.8\ w_0$.

Figure 3 will show that slopes of similar magnitude (relative to w_0) are obtained when a larger number of sites is considered, and also when the site density is varied. Calculations for a surface distribution of carboxyl groups in the presence of positively charged cationic sites also lead to essentially the same result. It is therefore clear that we cannot account for the "typical" slopes of protein titration curves by placing the sites at the very surface of the molecules, unless we assume configurations even more crowded than that of model D, which surely could not represent typical distributions of sites.[9]

(8) Actually straight lines are not the best curves through the calculated points. The best curves would be very shallow S-shaped curves extrapolating to the true intrinsic pK of 4.60. However, the deviations from linearity are of the order of the experimental error in experimental curves: if points such as those of Fig. 1 were obtained in an experiment, a straight line would be drawn through them, with a slope determined by giving greatest weight to the points in the middle of the portion of the curve. In most cases the ends of the curve are not readily accessible experimentally in any event. In titrating carboxyl groups, for instance, the alkaline end of the curve is obscured by overlapping with imidazole groups, and the acid end occurs at low pH, where the calculation of $\bar{\nu}$ becomes relatively inaccurate.

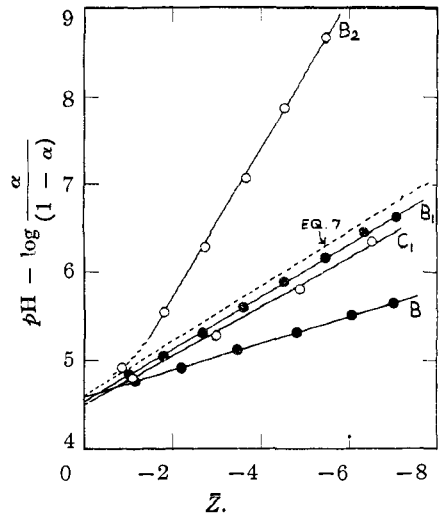


Fig. 2.—Logarithmic plots for eight -COOH groups (models B and C) at various depths within a sphere of radius 10 Å. Curve B is for sites at the surface, curves B_1 and C_1 for sites 1 Å. below the surface, curve B_2 for sites 2 Å. below the surface.

Accordingly, in Fig. 2, we investigate the effect of placing charges below the surface. Shown are curves calculated on the basis of model B: curve B with charges at the surface; curve B_1 with charges 1 Å. below the surface ($\rho_{kl}^{1/2} = 0.9$; $\delta = 0.025$, *i.e.*, $D_i \simeq 2$); curve B_2 with charges 2 Å. below the surface ($\rho_{kl}^{1/2} = 0.8$, $\delta = 0.025$). Curve C_1 is calculated on the basis of model C with charges 1 Å. below the surface ($\rho_{kl}^{1/2} = 0.9$, $\delta = 0.050$, *i.e.*, $Di \simeq 4$: with $D_i \simeq 2$ the curve would fall slightly above curve B_1).

It is at once apparent that the depth below the surface is a more crucial variable than the distribution of sites at a given depth. Thus for sites at the very surface (curve B), $w < 0.5\ w_0$, while for sites 2 Å. below the surface (curve B_2) $w > 3\ w_0$. To obtain $w \simeq w_0$ it is clearly necessary to place the sites potentially bearing a charge at a distance of about 1 Å. below the surface.

In order to make certain that this result is not an artifact, due to the rather small number of sites or to the particular site density of the model, we show in Fig. 3 that an essentially similar result is obtained for a larger number of sites, with two very different site densities.

From the results just presented one can draw the conclusion that any experimental value of w between about $0.5\ w_0$ and $w \gg w_0$ can be accounted for by varying the depth of sites below the surface.

(9) Why does the "smeared site" model yield approximately the right result *with sites at the very surface?* Clearly because it allows with equal probability all inter-site distances from $r_{kl} = 0$ to $r_{kl} = 2b$, whereas a discrete site model excludes all values of r_{kl} less than, say, 5 Å. Thus the mutual electrostatic free energy of eight charges smeared over the surface of a sphere (radius b) in an infinite medium of dielectric constant D is $5.6\epsilon^2/Db$. For eight discrete charges at the corners of a cube inscribed in a similar sphere the interaction energy is $39.5\epsilon^2/Db$.

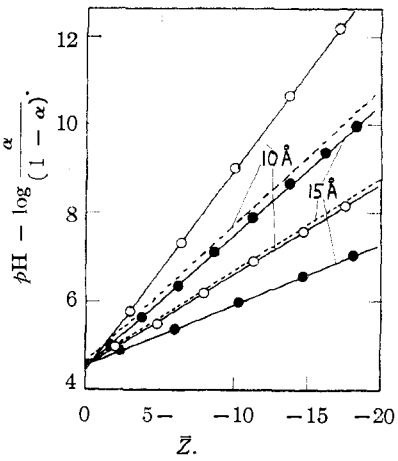


Fig. 3.—Logarithmic plots for twenty –COOH groups, uniformly distributed, on spheres of radius 10 and 15 Å., respectively. For each value of the radius the upper solid line has the sites 1 Å. below the surface, the lower solid line has them at the surface, and the dashed line represents a plot according to eq. 7.

Without some assumption about this depth, the experimental slopes cannot be used, for instance, as evidence for the physical dimensions of the protein molecule. Even with the assumption which we shall introduce in the following paragraph, the distribution of sites relative to one another allows considerable variation in w.[10,13]

We have shown that the typical protein titration curve ($w \simeq w_0$) can be accounted for theoretically only if the acidic sites, when charged, are close to 1 Å. from the molecular surface. Experimentally, deviations from the typical result are often observed. The most striking deviation occurs for lysozyme,[11] the carboxyl groups of which have a markedly flat titration curve, requiring that w be at least 2 w_0 or even greater. A more minor deviation occurs in ribonuclease,[16] where $w > w_0$ for the carboxyl groups, while $w < w_0$ for the phenolic groups. Are these and similar deviations to be explained on the basis of variation in depth below the surface? Are the carboxyl groups of lysozyme perhaps 2 Å. rather than 1 Å. below the surface, and the phenolic groups of ribonuclease say 0.5 Å. below the surface? In principle, there is no reason whatever to prevent such variation. For the remainder of this paper, however, we shall make the assumption that such variation does not occur. The following reasons (all of them purely empirical) may be cited in support of such an assumption. (1) All known deviations from the typical result, including that for lysozyme, can be explained on the basis of the distribution of sites at constant depth. (2) According to the preceding paper variation in depth can lead to large changes in pK_{int}. Ordinarily only small changes in pK' of equation 6 are observed, and these can again be accounted for on the basis of distribution of sites at constant depth. (3) We have used the well-known electrostatic effects affecting the acidity of low molecular weight organic acids[17] to calculate the depth of charges in these molecules. Once again most of them are strikingly close to 1 Å. It appears quite possible that this distance is a constant inherent in the Kirkwood model used as the basis for all these calculations.[18,19]

Two Kinds of Sites.—We consider next the titration of one kind of site on a protein molecule also containing a second kind of site. We first take up the simple case where the second kind of site is titrated in a different pH region from the first, so that its charges can be considered as fixed charges during the titration of the first kind of site.

Figure 4 shows the results of calculations for model molecules containing four carboxyl groups and four fixed sites of positive charge, which could be amino or guanidyl groups. The eight sites are placed at the corners of a cube inscribed in a sphere of radius 8 Å., in three different configurations, illustrated by models A, E and F. We have placed $\rho^{1/2} = 0.9$, so that the sites of charges fall 0.8 Å. below the surface.

Figure 4 shows that the value of w for each of the chosen models is essentially the same as that for the same configuration of carboxyl groups in the ab-

(10) Two earlier conclusions made by me are clearly unjustified: (1) The fact that titration of the amino groups of lysozyme[11] gives for w a value perhaps a little larger than w_0 cannot be used as evidence that this protein is more sparingly hydrated than other small proteins. (2) Although the shape of the logarithmic plot for the titration of the carboxyl groups of insulin[12] is corroborative evidence for reversible association of the molecules in the acid pH range, an absolute estimate of the molecular weight cannot be made. The titration curve cannot be used as evidence to exclude a molecular weight of 6000 in aqueous solution.

(11) C. Tanford and M. L. Wagner, THIS JOURNAL, **76**, 3331 (1954).

(12) C. Tanford and J. Epstein, *ibid.*, **76**, 2163 (1954).

(13) The present model does not, however, except for as few as three or four titratable sites (*cf.* Fig. 4), allow w to be much smaller than $0.5w_0$. The fact that w for serum albumin[14] falls to less than $0.2w_0$ in acid solutions can therefore be used as proof that the molecule can no longer be an impenetrable sphere. For myosin[15] w is also much smaller than w_0, presumably because it is a long rod-shaped molecule.

(14) C. Tanford, S. A. Swanson and W. S. Shore, THIS JOURNAL, **77**, 6414 (1955).

(15) E. Mihalyi, *Enzymologia*, **14**, 224 (1950); L. B. Nanninga, *Arch. Biochem. Biophys.*, **56**, 334 (1955).

(16) C. Tanford, J. D. Hauenstein and D. G. Rands, THIS JOURNAL, **77**, 6409 (1955); C. Tanford and J. D. Hauenstein, *ibid.*, **78**, 5287 (1956).

(17) J. G. Kirkwood and F. H. Westheimer, *J. Chem. Phys.*, **6**, 506, 513 (1938).

(18) A full account of this work is given in the next paper (p. 5348). It should be noted that all of these calculations, both for proteins and for the smaller molecules, are based on the assumption of an internal dielectric constant $D_i = 2$. If a different internal dielectric constant had been used, essentially the same result would have been obtained, except that the numerical value for the distance below the surface would have been different. However, it would be necessary to have $D_i \simeq 15$ in order for the empirical depth below the surface to increase from 1 to 2 Å.

(19) Another way of stating the assumption here made is that all protein sites accessible to titration are in direct contact with the solvent (water). The 1 Å. distance then represents perhaps the van der Waals radius of the atom on which the charge is located, or it may be an effective distance resulting from electric saturation of the solvent near a charged site. In either case, it is reasonable that the distance should be the same in proteins as in organic acids of low molecular weight. Furthermore, it would be difficult to assign any meaning to a change in this distance of a few tenths of an Å. except in so far as cationic sites and anionic sites may differ in their contact with solvent. One could interpret a change in depth of *several* Å.: this would clearly mean that another atom, or a methylene group, etc., had been placed between the charged site and the solvent. However, no reversibly titrated sites on protein molecules have titration curves which could conceivably allow a depth of several Å. below the surface. Presumably all sites so located will be among the *inaccessible* sites often observed in titration curves, *e.g.*, in hemoglobin.[20]

(20) J. Steinhardt and E. M. Zaiser, *Advances in Protein Chem.*, **10**, 180 (1955).

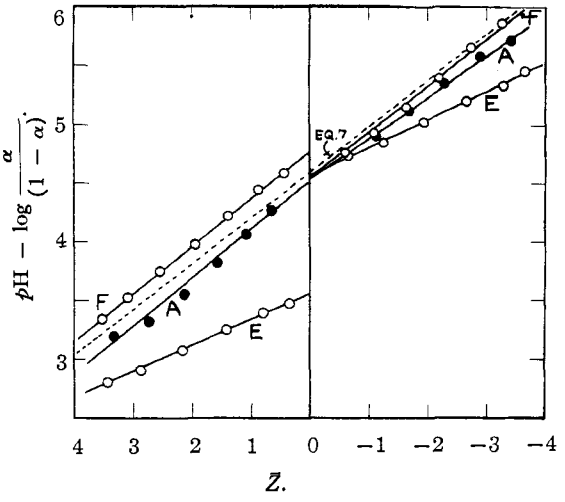
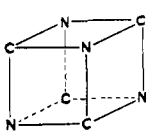
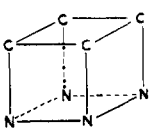
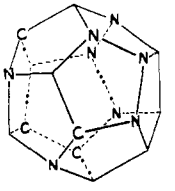


Fig. 4.—Logarithmic plots for four –COOH groups in the presence of four $-NH_3^+$ groups, at a depth of 0.8 Å. in a sphere of radius 8 Å. The left-hand side shows calculations for models A, E and F. The right-hand side applies to the same models with the $-NH_3^+$ groups removed.

sence of the positively charged sites. The data for the carboxyl groups alone are shown on the right side of Fig. 4. It is noted that a symmetrical arrangement of as few as four sites gives $w < w_0$ even at a depth approximately 1 Å. below the surface. For a larger number of groups, as we have seen, the most symmetrical arrangement does not lead to so different a result.


Model E


Model F


Model G

The most striking feature of Fig. 4 is the pronounced variation in pK', the intercept at $\bar{Z} = 0$. The same value of pK_{int} (4.60) was assumed for each model, and it is quite obvious therefore that pK' is not necessarily equal to pK_{int}. Where all of the groups are identical, then $\bar{Z} = 0$ corresponds in fact to an uncharged molecule, and the resulting disappearance of all major electrostatic effects. Under these circumstances $pK' \simeq pK_{int}$. Where both cationic and anionic sites are present, however, $\bar{Z} = 0$ corresponds to a zwitterion form with equal numbers of positive and negative charges. The distribution of these may be such as to cause their effects to cancel. On the other hand, in an arrangement such as that of model E, their effects do not cancel. As a result, the carboxyl groups at $\bar{Z} = 0$ are not equivalent to carboxyl groups on a discharged molecule, and pK' can then differ considerably from pK_{int}.

The result just given can account for most observed variations of pK' in actual proteins. For carboxyl groups, for instance, one finds $pK' = 4.6$ in β-lactoglobulin,[21] $pK' = 4.3$ in ovalbumin[22] and $pK' = 3.95$ in serum albumin.[14] It is obviously unnecessary (though not demonstrably wrong) to postulate elaborate schemes of hydrogen bonding[23] to account for such variations.

Figure 5 shows data similar to Fig. 4, for cationic and anionic groups titrated in the same pH region.

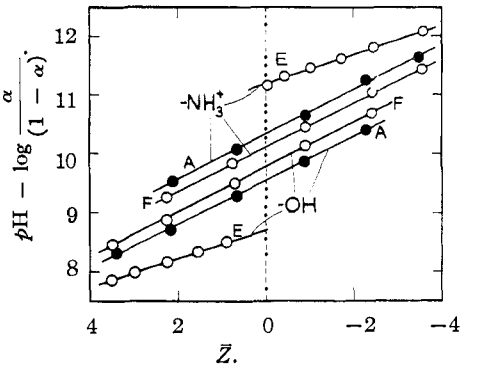


Fig. 5.—Logarithmic plots for four $-NH_3^+$ groups and four $-C_6H_4OH$ groups, at a depth of 0.8 Å. in a sphere of radius 8 Å. Distribution according to models A, E and F.

Models A, E and F were used, with the same dimensions as before, with C now representing phenolic groups with an assumed pK_{int} of 10.2. The calculations in this case are quite laborious because 25 different subspecies $PH_\nu^{(i)}$ can occur and, hence, 24 different equilibrium constants must be calculated. The results of the calculation resemble those of Fig. 4, both as regards slope and pK'. It should be noted that for model E the pK' values for both kinds of groups are again shifted by about 1.0, and that the shift for the amino groups is in the opposite direction from that for the phenolic groups, as is to be expected. Figure 6 shows the actual titration curves for models E and F.

In the models just discussed all of the sites of any one kind have been similarly situated with respect to the sites of the second kind. The change in interaction energy with ν has therefore been principally a function of the locations of the sites being titrated with respect to one another, and it is for this reason that w was found to be not very different from its value in the absence of a second type of site.

Another possibility presents itself, however, when the sites being titrated are relatively small in number compared to the sites of the second kind. When this happens the slope may be determined largely by the position of the sites being titrated with respect to the second kind. In model G, for instance, one carboxyl group has two cationic sites (fixed positive charge) as nearest neighbors at a distance of 6.4 Å.; two carboxyl groups each have one cationic site at a distance of 6.4 Å.; the fourth carboxyl group has no charged sites at this distance at all. These carboxyl groups would show a different affinity for hydrogen ions even in the absence of mutual interaction, and this is reflected in

(21) R. K. Cannan, A. H. Palmer and A. Kibrick, *J. Biol. Chem.*, **142**, 803 (1942).

(22) R. K. Cannan, A. Kibrick and A. H. Palmer, *Ann. N. Y. Acad. Sci.*, **41**, 243 (1941).

(23) For example, G. I. Loeb and H. A. Scheraga, *J. Phys. Chem.*, **60**, 1633 (1956).

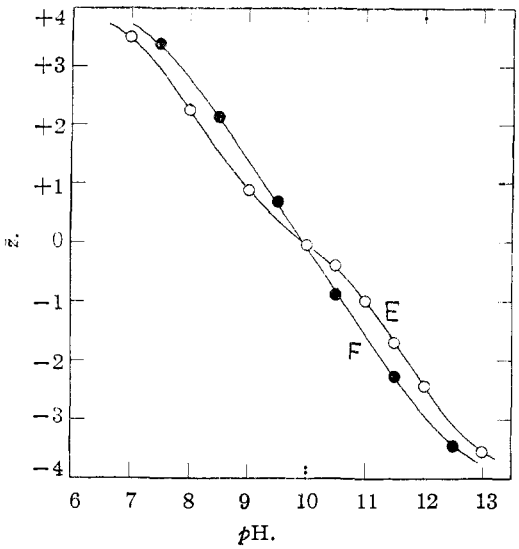


Fig. 6.—Actual titration curves for two of the models of Fig. 5. Curves E and F differ only in the distribution of sites. The same pK_{int} values have been used.

Fig. 7 by a value of w considerably *larger* than w_0 when $\rho_{kl}^{1/2} = 0.9$. Moreover, it should be noted that the titration of the carboxyl groups here occurs far from the point of $\bar{Z} = 0$. If now one draws a reasonable straight line and extrapolates to $\bar{Z} = 0$ one obtains a pK' value of about 6.75, although the usual value of 4.6 has been assumed for pK_{int} in making the calculations.

There is a striking similarity between Fig. 7 and the experimental plot for the carboxyl groups of ly-

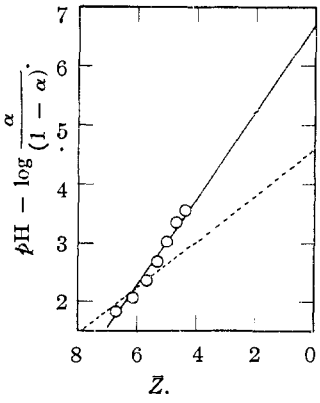


Fig. 7.—Logarithmic plot for the four –COOH groups of model G (circles and solid line). The dashed line is drawn according to eq. 7.

sozyme,[11] shown in Fig. 8. It should be noted that lysozyme is precisely the kind of protein represented by model G, for it possesses 19 sites of positive charge and only 10 carboxyl groups. Its isoionic point is at pH 11.1. The same kind of effect has been observed to a much lesser degree in another basic protein, ribonuclease.[16]

The Effect of Ionic Strength

The effect of ionic strength is most simply computed in terms of activity coefficients $\gamma_\nu^{(i)}$ for the various protein species $PH_\nu^{(i)}$. If we let $K_\nu^{(i)}$ designate the equilibrium constant at any ionic

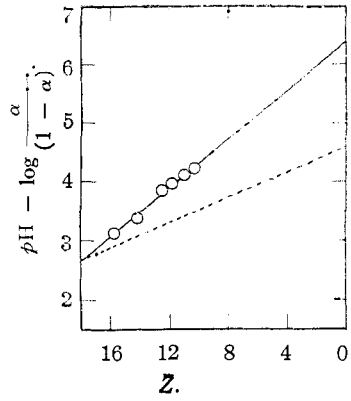


Fig. 8.—Experimental logarithmic plot for the carboxyl groups of lysozyme (ref. 11), at ionic strength 0.03. The dashed line is drawn according to eq. 7 (pK_{int} assumed equal to 4.6), with w corrected for ionic strength by eq. 10.

strength for the reaction $P + \nu H^+ - PH_\nu^{(i)}$, retaining the symbol $k_\nu^{(i)}$ for the equilibrium constant of the same reaction at zero ionic strength, then

$$K_\nu^{(i)} = k_\nu^{(i)}\gamma_0/\gamma_\nu^{(i)} \qquad (8)$$

We can evaluate these activity coefficients by calculating $\bar{F}_\nu^{(i)}$ of eq. I-35 first at zero ionic strength and then at the desired ionic strength. It will be recalled from paper I that at *low ionic strength* (up to close to $x = \kappa a = 1$) the predominant term in the ionic-strength dependent portion of the energy is $(Z^2\epsilon^2/2Da)x/(1+x)$ which is independent of configuration. Under these circumstances it is not necessary in calculating activity coefficients to extend the expansion in eq. I-35 to as many terms as are used in computing $k_\nu^{(i)}$. Since two terms have been used for the latter we need to consider the effect of ionic strength on one term only (*i.e.*, on M_1 of eq. I-35). We thus obtain for the difference between $\bar{F}_\nu^{(i)}$ at any ionic strength and $\bar{F}_\nu^{(i)}$ at zero ionic strength, *i.e.*, for kT in $\gamma_\nu^{(i)}$

$$kT \ln \gamma_\nu^{(i)} = -\frac{\epsilon^2}{2a}\sum_{k=1}^{m}\sum_{l=1}^{m}\left(\frac{\sum_c \xi_k\xi_l}{\Omega_\nu^{(i)}}\right)C_{kl} \qquad (9)$$

where C_{kl} is given by eq. I-8 and some values at low ionic strength are shown in Fig. 2 of paper I.

Since the major portion of $\ln \gamma_\nu^{(i)}$ at low ionic strength is independent of configuration, the calculated activity coefficients differ little from those computed by the smeared-site procedure, and, for 8 or more sites, the effect of ionic strength then differs little from that predicted by the value of w given in eq. I-4, *i.e.*

$$w = \frac{\epsilon^2}{2DbkT} - \frac{\epsilon^2}{2DakT}\frac{x}{1+x} = w_0 - w_1 \qquad (10)$$

even if the calculated value of w at zero ionic strength is quite different from w_0 as given by eq. 7. (If the number of sites is as small as four we again see some deviations, corresponding to those observed in Fig. 4.)

The important result is that all the characteristics of the various models computed for zero ionic strength *are maintained as the ionic strength is increased.* For instance, Fig. 9 shows curves B, B_1 and B_2 of Fig. 2, but at ionic strength 0.038, while

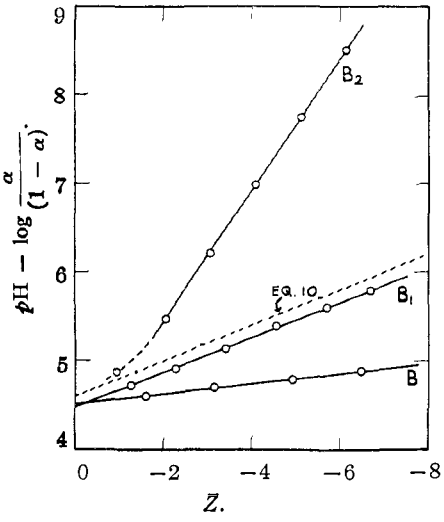


Fig. 9.—Recalculation of the data of Fig. 2 at an ionic strength of 0.038.

Fig. 10 shows the effect of ionic strength on curve E of Fig. 4.

In Fig. 11 we show that the ionic-strength dependence of the slopes of these curves, and compare it with the factor $(0.868\ \epsilon^2/2DakT)\ x/(1+x)$ predicted by eq. 10. In the two examples involving eight titratable sites there is good agreement with eq. 10; in the case of the data of Fig. 10, involving the titration of only four sites, there is some difference, as already mentioned.

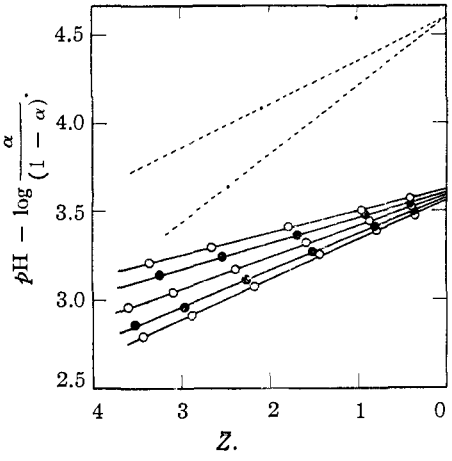


Fig. 10.—The effect of ionic strength on the logarithmic plot for model E. The solid lines (reading from top to bottom) are for ionic strengths 0.06, 0.025, 0.0065, 0.001 and zero, respectively. The dashed lines represent equations 10 and 7, at ionic strengths 0.06 and zero. Depth and radius the same as for Fig. 4.

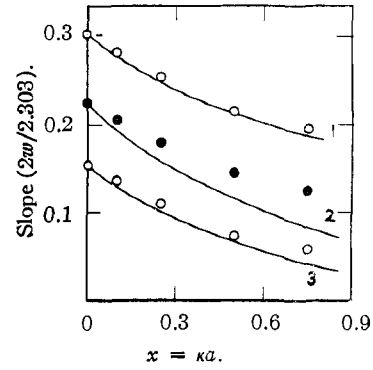


Fig. 11.—The effect of ionic strength on the slopes of logarithmic plots. The points represent slopes calculated at various ionic strengths for models B and E. The solid lines are drawn through these points at zero ionic strength, but the *change in slope* with ionic strength (*i.e.*, w_1) is computed according to eq. 10. Curves 1 and 3 are for model B (*cf.* Fig. 9) with sites, respectively, 1 Å. below the surface and at the surface. Curve 2 is for model E (*cf.* Fig. 10).

Figure 9 shows that pK' decreases slightly with increasing ionic strength. This is due to the self-energy terms in C_{kl} and is the same effect which causes a decrease in the pK of a simple carboxylic acid with increasing ionic strength. Of greater interest is the change in pK' in Fig. 10, where the value at zero ionic strength is far below pK_{int}. We see that there is now an increase in pK' as the ionic strength decreases the effect of the neighboring amino groups in Model E. However, the amount of change is very small, *i.e.*, the abnormal pK' values calculated at zero ionic strength, for proteins with the pertinent configurational features, are maintained as the ionic strength is increased.

It is of interest that proteins which show pK' values different from pK_{int} do in fact exhibit ionic strength effects in the same direction and of similar magnitude as those shown by Fig. 10. Thus, for the carboxyl groups of serum albumin,[14] pK' changes from 3.92 to 4.02 between ionic strength 0.01 and 0.15 (pK_{int} being presumably 4.6); for the phenolic groups of ribonuclease[16] it changes over the same range of ionic strength from 9.99 to 9.92 (pK_{int} presumably about 9.6).

NEW HAVEN, CONN.