THE ACID–BASE TITRATION OF PROTEINS[1]

R. KEITH CANNAN

Department of Chemistry, New York University College of Medicine, New York, New York

Received January 2, 1942

The more important problems which arise in the determination and the interpretation of the acid–base titration curves of soluble proteins are reviewed. Methods of identifying the number and nature of the ionizing groups are considered. An electrostatic theory, first developed by Linderstrøm-Lang, is applied to the description of the shapes of the curves and to the effect of electrolytes on them. Certain deductions from the theory are outlined.

I. INTRODUCTION

The reaction of a protein with a strong acid or base may be described most comprehensively by a curve having the qualities of a dissociation curve. This curve expresses the relation between the pH of a solution of the protein and a quantity, which we will call h, and which is generally described as the amount of acid or base "bound" to the protein. A series of reaction mixtures are prepared, containing a fixed quantity of protein and varying quantities of a strong acid or base. The pH of each mixture is determined with the aid of an appropriate electrode. The acid or base bound to the protein is computed by subtracting the amount of free hydrogen ion or hydroxide ion in each solution from the amount of acid or base which had been added. We will adopt the convention that h shall be a positive quantity when acid is bound and negative when base is combined. It is the purpose of this paper to discuss the type of information which may be derived from a study of the dissociation curves of soluble proteins and of the effect on them of a number of variables. Some preliminary remarks on the question of standard methods of recording the basic data may be appropriate.

II. THE pH SCALE

The limitations of the pH scale as a scale for the measurement of the relative acidities of acid–base systems are well known. However, there is available no other type of measurement which is comparable in scope and flexibility. In the study of proteins we certainly cannot yet dispense with the pH scale if we are to attempt any consistent correlation of the rapidly accumulating mass of physical and chemical observations on protein solutions. It is proper, therefore, that the reactions of proteins with acids and bases should be described in terms of pH. It is desirable, however, that the scale adopted should be in as close accord with thermodynamic quantities as may be conveniently realizable.

[1] Presented at the Symposium on Physicochemical Methods in Protein Chemistry, which was held under the joint auspices of the Division of Physical and Inorganic Chemistry and the Division of Biological Chemistry at the 102nd Meeting of the American Chemical Society, Atlantic City, New Jersey, September 8–12, 1941.

Two different methods have been advocated for the standardization of the hydrogen–calomel cell. Both seek to establish equality in the ideal relation, $pH = -\log a_{H^+}$, by the assignment of an appropriate value to the potential of the reference electrode. In the simpler method, liquid-junction potentials are assumed to be constant and are implicitly incorporated in the standard potential of the reference electrode. The latter is so chosen as to give, for a standard buffer solution of a weak acid, a pH consistent with the thermodynamic dissociation constant of the acid. MacInnes (19) and others (13) have recently discussed the values of pH which should be assigned to several popular buffer standards. Their recommendations are, undoubtedly, satisfactory when undertaking pH measurements of solutions of weak electrolytes comparable in concentration and in electrolyte pattern with the buffer solution which was chosen as the standard. They admittedly fail in the presence of important concentrations of hydrogen ion and hydroxide ion and, in general, at high ionic strength. In such situations a more laborious procedure has theoretical advantages. In this method, explicit allowance is made for the difference between the potentials of the liquid junction at the standard buffer and at the unknown solution. This is done by subtracting, from the observed potentials of the cell, liquid-junction potentials calculated with the aid of Henderson's equation. The limitations of this equation are admitted. Because this correction is made, the values which are assigned to the standard buffer solutions (11) differ somewhat from those which should be used when liquid-junction potentials are ignored. This method has found favor in studies involving the comparison of pH measurements made at a wide range of ionic strengths. However, there are definite advantages in the adoption of a common practice by all observers. It is suggested, therefore, that all pH measurements should be computed and recorded on the basis of an approved standard, with neglect of liquid-junction potentials. Explicit correction for the latter may then be made by the observer for his own purposes. In the meantime, until a common standard shall have become general, it is desirable that the method adopted in precise studies should always be stated and that the method of establishing the liquid junction should also be given. The purity of protein preparations and the reproducibility of the potentials of their solutions are such as to justify this refinement of definition.

III. THE ISOIONIC POINT

The logical point of origin of the scale of h values is the isoionic point (pI_i), as defined by Sørensen, Linderstrøm-Lang, and Lund (29), rather than the isoelectric point (pI_e) derived from measurements of electrophoretic mobilities or of membrane potentials. The isoelectric point corresponds with the pH at which the net charge is zero, i.e., the number of positive charges on the protein molecule is equal to the number of negative charges. The isoionic point, on the other hand, is the pH at which the number of protons combined on the basic groups is equal to the number dissociated from the acidic groups.[2] The isoelectric

[2] The word "groups" is here understood to refer to uncharged groups such as COOH, NH_2 etc.

and isoionic points are identical only if the protein combines[3] with no ions other than hydrogen ion. In many cases, the value of pI_i may readily be determined without ambiguity. Proteins, such as the globulins, which are prepared by spontaneous precipitation from substantially salt-free solutions probably contain insignificant amounts of non-protein ions. The pH of a solution of such a preparation will be practically identical with pI_i, provided (*a*) the latter falls within the approximate limits of pH 4.5 and 9.5, (*b*) the concentration of protein is not less than 1 per cent, and (*c*) the concentration of salt is low. If pI_i has more extreme values than those indicated, it may be determined in the following way: A series of dilute solutions of a strong acid or base are prepared and their pH values are determined. A solution of the pure protein is added to each and the pH is redetermined. Observations are continued until a solution is found the pH of which is unaffected by addition of the protein. This pH may be identified with the isoionic point, since it may be assumed that there has been no combination of the protein with the anion of the acid, or with the cation of the base, at the very low concentrations in which these are present.

Proteins, such as the albumins, which are prepared by salting-out procedures may be far removed from the isoionic condition. Solutions of these preparations often retain some diffusible ions even after protracted dialysis. In this situation one may resort to electrodialysis or may adopt the method of Sørensen (29) and of Sandor (26). The solution of the protein—assuming that pI_i is below 7—is dialyzed against a dilute solution of ammonia until all diffusible anions have been removed and all cations have presumably been replaced by ammonium ion. The quantity of the latter is determined analytically, an equivalent of strong acid is added, and the pH of the solution is determined. This is identified with pI_i.

Isoionic points may also be located on the dissociation curves by comparing the curves for different concentrations of protein. Provided the solutions contain few diffusible ions, the slopes of the curves will vary with the concentration of protein but will all intersect at pI_i (*cf.* table 3). It has been suggested that the isoionic point should also be independent of salt concentration. This view received much support from the observation (29) that the curves of egg albumin in solutions of ammonium chloride (0.05–3 *M*) all intersected at pI_i. Unfortunately the situation is not general. It does seem to be true that pI_i is less sensitive than pI_e to changes in salt content, but definite effects have been observed in some cases (4) at concentrations of potassium chloride and of other salts exceeding 0.1 *M*. The change in apparent isoionic point depends on the nature both of the salt and of the protein.

IV. THE CALCULATION OF h

When h is measured from pI_i, it represents the difference between the amount of acid or base which must be added to a protein solution to bring it from pI_i

[3] The anions or cations which establish electrical neutrality with the protein ions are sometimes loosely described as "bound" to the protein. This, of course, is not to be confused with the binding referred to here, as a result of which the charge on the protein is modified in proportion to the number of ions bound.

to the final pH and the amount of free hydrogen ion or hydroxide ion appearing in the process. The latter is usually determined by measuring the amount of acid or base which must be added to the same quantity of protein-free solvent to effect the same change in pH. The protein-free solution should simulate the electrolyte content of the solution of protein as closely as possible. This procedure involves the assumption that the protein ions do not significantly affect γ_{H^+} or the liquid-junction potential. Errors due to failure of these assumptions are likely to be smaller in solutions containing moderate amounts of a neutral uni-univalent electrolyte than in those which contain protein and acid or base alone. Over the greater part of the dissociation curve, $[H^+]$ and $[OH^-]$ are small, relative to the acid or base added, and small errors in the values assigned to them do not significantly affect the absolute values of h. At extremes of pH, on the other hand, h becomes a small difference between two large quantities and cannot be evaluated accurately. This is one of the considerations which fix the dependable limits of titration of dilute solutions of proteins at about pH 1.5 and 12. Unfortunately, maximum acid- and base-binding capacities of proteins are attained only in these limiting regions. It has been our experience that the most reproducible observations at the extremities of the curves are obtained with solutions containing high concentrations of both protein and potassium chloride.

Values of h are usually recorded in equivalents per gram of protein, although it may sometimes be appropriate to express them in terms of some acceptable molecular weight of the protein. Now, experimental values of the weight of dry protein and of protein nitrogen in a solution are known to depend, in some degree, upon the method of determination. It is important, therefore, that the method of determining the protein content of solutions should be stated.

V. THE NATURE OF THE REACTION OF PROTEIN WITH ACID AND BASE

Irreversible reactions

All proteins are irreversibly affected by a sufficient concentration of hydrogen ion or of hydroxide ion. Some proteins, such as hemoglobin, are denatured rapidly at pH values not far removed from the isoionic point. Others appear to react reversibly with acid or base over a range of pH which may extend from 1.5 to 11 or 12. Denaturation by acid or alkali is generally associated with the formation of a product—a metaprotein—insoluble in the isoelectric region, but the chemical changes involved are not known. They probably vary from protein to protein. In some cases there is evidence of a change in the number of acid- and base-binding groups. Now, some denaturation is to be anticipated at the extremes of pH at which maximum combining capacities are measured. It is evident that the results obtained cannot be taken to correspond with the numbers of ionizing groups in the original protein, unless irreversible reactions with hydrogen ion or hydroxide ion have been excluded.

As evidence of the reversibility of the reactions with acid and base, it has often been held sufficient to demonstrate that the potentials observed showed no consistent drifts with time. It is advisable also to show that a solution, after

submission to an extreme pH, retraces an unmodified curve when titrated back to pI_i. This serves, also, to disclose the presence of any metaprotein. In critical situations it may be desirable to show that the protein has suffered no change in some characteristic physical or biological property. On the other hand, such tests may be unduly severe. Many changes in the physical and biological characteristics of proteins are accompanied by no demonstrable change in amphoteric properties.

Too much confidence should not be placed in the absence of drifting potentials. In well-buffered regions significant changes in dissociation may not reveal themselves in noticeable drifts. Moreover, the rate of denaturation may vary with a high power of $[H^+]$ or $[OH^-]$. In such case, a small addition of acid or base may be sufficient to carry the solution from a pH at which the rate is very small to one at which the reaction is complete in the period allowed for establishment of equilibrium at the electrode. It would, then, only be by chance that a reaction mixture would be encountered in which the irreversible reaction was proceeding at such a rate as to give a persistent drift of potential. In a study (25) of the denaturation of crystalline egg albumin by sodium hydroxide at 25°C., it was found that the reaction followed a roughly bimolecular course, the rate of which was proportional to $[OH^-]^4$. At initial pH values of 12.50, 12.33, 12.02, and 11.84, it required 3 min., 15 min., 4.5 hr., and 20 hr., respectively, to denature 33 per cent of the protein. The pH fell steadily during the reaction. Dissociation curves of solutions of the partially and completely denatured protein showed that the curve, above pH 8, was progressively displaced from that of the native protein as denaturation proceeded. The changes were consistent with a liberation of weak acid groups, but were not due to carbon dioxide or other diffusible acids. It may be added that similar studies of the denaturation of egg albumin in dilute hydrochloric acid led to the conclusion that this reaction was accompanied by little, if any, change in acid-binding capacity.

Reversible reactions

It is customary to assume that the reversible reaction of proteins with acids and bases is entirely a process of proton transfer. If so, we may expect h to be identical with the net charge and pI_i with pI_e. There are, however, well-established cases in which the isoelectric and isoionic points differ materially from one another. In some cases pI_e has been shown to be very sensitive to ionic strength and to the nature of the ions present (32). These effects may be interpreted either as the result of a combination of the protein with specific ions or on certain rather elusive assumptions respecting the individual activity coefficients of the variously charged protein ions which constitute the isoionic assembly of ions (28). Observations of the osmotic pressures, membrane potentials (1), and silver chloride electrode potentials (12) of protein solutions have also been held to be consistent with the specific combination of protein with simple ions[4] such as Cl^-, $C_2H_3O_2^-$, and $H_2PO_4^-$. If such combination

[4] Striking specific anion effects have been observed in the reaction of insoluble proteins with acids (30).

does occur, it is presumably an ion association such as that discussed by Bjerrum (3). In so far as the interpretation of dissociation curves is concerned, it will be sufficient, for the present, to acknowledge that combination with ions other than hydrogen ion will affect the relation of h to the net charge and, presumably, will modify the slopes of the curves. On the other hand, it will not affect deductions which are based on the identification of stoichiometric points on the curves and which lead to estimates of the number and nature of the ionizing groups in the protein. In the discussion which follows, h will be identified with the charge on the protein, leaving open the question whether such anomalies as arise may be ascribed to specific ion combination.

VI. CHARACTERISTICS OF THE DISSOCIATION CURVES OF PROTEINS

Many examples of dissociation curves of proteins may be found in the literature (6). The proteins the curves of which have been studied most intensively are gelatin, casein, edestin, hemoglobin, egg albumin, β-lactoglobulin, and clupein. Clupein (24) is a very interesting but atypical protein. The work on hemoglobin is distinctive, not only as a contribution to the chemistry of the reaction of hemoglobin with oxygen (35), but also because it included the first careful study of the effect of temperature on the dissociation curve of a protein (34). The investigations on gelatin, casein, and edestin are notable for the fact that the acid-combining capacities were checked by observations employing a cell without liquid junction (12). The properties of many of the proteins mentioned make them unsuitable for investigation over the wide range of conditions to which the method of titration is adaptable. Egg albumin (4) and β-lactoglobulin (5) have proved to be more satisfactory proteins for a study of the characteristics of dissociation curves. They are both reproducible crystalline products, the solutions of which exhibit some, at least, of the criteria of molecular homogeneity. They are soluble over the whole experimental pH range and appear to react reversibly over the greater part of it. Many of their physical properties have been investigated as a function of pH—in particular, their electrophoretic mobilities. Since the latter are, so obviously, a function of charge, the relation of mobility to h is of great interest (5, 18).

The curves of individual proteins have much in common. Indeed, they differ only in the relative slope and span of specific regions of the curves. These differences are believed to reflect differences in the relative numbers of certain types of ionizing groups in the individual proteins. The curves in figure 1 will serve to illustrate some of the characteristics which will be discussed in this paper. The very considerable effect of [KCl] on the slope of the curve, which is seen in the figure, will be discussed in a later section. In the main, the change in slope seems to be a function of ionic strength, although small specific ion effects have been observed (4). Two well-defined inflexions are evident in the curves of figure 1. Their positions on the h-axis are independent of μ. One occurs at low pH and corresponds to the establishment of a definite maximum cation charge (maximum acid-binding capacity). A second is observed in the neighborhood of pH 8.5. The curves of some proteins are almost horizontal in this region.

In others there is no clear end point, but the value of h at the point of minimum slope can still be determined with fair precision. There have been many attempts to establish experimental values for the maximum anion charge at high pH. The curves of β-lactoglobulin and of some other proteins do appear to be approaching an end point somewhere beyond pH 12. However, they certainly do not attain it within the zone of reversibility of the curves, nor, indeed, at a pH at which dependable values of h may be computed. We are doubtful if measurements of the maximum base bound by proteins may be accepted with the same confidence as those of the maximum acid-combining capacities.

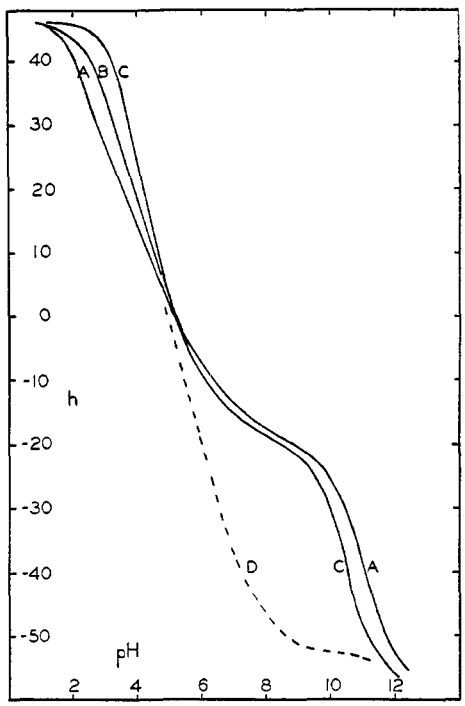

FIG. 1. Dissociation curves of β-lactoglobulin. Curve A, 0.019 M potassium chloride, 0.5 per cent protein; curve B, 0.135 M potassium chloride, 0.5 per cent protein; curve C, 0.67 M potassium chloride, 0.5 per cent protein; curve D, 1 M formaldehyde, 2 per cent protein.

The stoichiometric point near pH 8.5 has been confirmed and a new one, close to pH 6, has been revealed in studies of the effects of temperature on dissociation curves. Wyman (34) found that the displacement with temperature, $\Delta pH/\Delta t$, of the curve of hemoglobin was discontinuous. The curve relating $\Delta pH/\Delta t$ to $pH_{25°}$ showed a sharp inflexion close to pH 8.5 and another near pH 5.5. A similar situation has been observed in the curves of all other proteins the temperature coefficients of which have been examined. In the curves of cytochrome c (31) the inflexions appear at about the same levels as in hemoglobin. In β-lactoglobulin, hen albumin, and duck albumin the lower point is close to pH 6.5.

When the titrations of a protein are performed in solutions containing formaldehyde (1–8 per cent), the shape of the curve is greatly modified (figure 1). Above pH 5, the values of $-h$ rise progressively above those found in the absence of the aldehyde, until the curve attains a maximum at about pH 8.5. Thereafter it may remain horizontal up to pH 11, which is the feasible limit of titration in the presence of formaldehyde. An analysis of the curves has shown (14) that the formaldehyde has displaced a definite segment of the curve from the region of pH 10 to about pH 7. This segment is believed to correspond with the free amino groups of the protein.

Other devices for isolating definite segments of a curve have been useful in special cases. The maximum cation charge of a prolamine has been determined by titration in 80 per cent alcohol (8). Titrations of iodized (21) and of deaminized (10) proteins have also been illuminating.

VII. THE ACIDIC AND BASIC GROUPS OF PROTEINS

In table 1 are listed the acidic and basic groups the presence of which in the protein molecule may be anticipated from our knowledge of the composition and structure of proteins. The majority of these groups are those which occur in the side chains of the trivalent amino acids, the names of which are given in the table. In addition to these, such α-amino and α-carboxyl groups as are not combined in peptide or amide linkage will contribute to the dissociation of the protein. In a simple polypeptide, one pair of alpha groups will remain free. If, however, the polypeptide contains cystine, additional alpha groups will be present. Their number will depend on whether the disulfide links occur in a single peptide chain or as cross linkages between two or more chains. In general, any departure from a simple peptide structure may liberate additional alpha groups. Finally, the possible presence of ionizing groups in the non-protein constituent of a complex protein must always be considered. Estimates of the values of the logarithm of the intrinsic dissociation constant, pK'_0, and of the temperature coefficient of this constant, $\Delta pK'_0$ (Δt = 10°C.), are included in table 1. These are based upon the dissociation constants (7) and heats of dissociation (34) of simple molecules containing the group to which they refer. The intrinsic dissociation constant of a group may be taken to represent the constant of an isolated representative of that group in the absence of interaction with other charges on the molecule.

Let us assume that the magnitudes quoted in table 1 do represent, in a general way, the acid–base characteristics of the ionizing groups of proteins. The dissociation of a protein may, then, be expected to be concentrated in three zones centered about the pH values 4, 7, and 10–12, respectively, and significant changes in $\Delta pH/\Delta t$ should be observed on passing from one zone to the next. The discontinuities in the temperature coefficients of the experimental curves are entirely consistent with these anticipations, and the sharpness with which $\Delta pH/\Delta t$ has been observed to change in the regions of the two transition points suggests that there is no serious overlapping of adjacent zones. The actual experimental values of $\Delta pH/\Delta t$ indicate that the segment of the curve which lies below pH 6 must be due almost entirely to the dissociation of carboxyl groups,

the segment between pH 6 and 8.5 chiefly to that of imidazole groups, and the arm of the curve immediately above pH 8.5 to that of amino groups. Theoretically, α-amino groups should also contribute to the central region of the curve, and the alkaline segment should be a composite of the contributions of the amino, guanidine, phenolic, and sulfhydryl groups. The guanidine groups, however, are so strongly basic[5] that they should exist entirely as cations below pH 11. Their only effect on the curve will be the addition of a fixed positive charge to the net charge which is measured by h. There is evidence, also, that the phenolic groups of proteins do not dissociate much below pH 11. The dissociation curves of several proteins containing considerable amounts of tyrosine have been determined in the presence of formaldehyde. They are almost horizontal from pH 8.5 to 11. That is to say, no buffering remains in this region after that of the amino groups has been removed by reaction with formaldehyde. Since phenolic groups are unlikely to be completely dissociated below pH 8.5, we may conclude

TABLE 1

Ionizing groups of proteins

GROUP	AMINO ACID	pK'_0	Q'	$\frac{\Delta pK'}{(\Delta t = 10°)}$
			calories	
α-Carboxyl		3.5	0	0
Carboxyl	Aspartic acid	4.0	0	0
	Glutamic acid			
Imidazole	Histidine	7.0	7,000	−0.17
α-Amino		8.0	10,500	−0.26
Amino	Lysine	10.0	10,500	−0.26
Phenolic	Tyrosine	10.0	6,000	−0.15
Sulfhydryl	Cysteine	10.0	?	?
Guanidine	Arginine	12.5	12,500	−0.31

that they contribute nothing to the curve or to the charge below pH 11 (*cf.* 21). The situation with respect to sulfhydryl groups is obscure. It is probable that most soluble proteins contain few, if any, active SH groups. When they do occur it is likely that they will be found to behave as weak acids similar in strength to the phenolic groups. This leaves only the carboxyl, amino, and imidazole groups to determine the whole sweep of the curve between pH 2 and pH 11. The span of the curve from the end point at low pH to the first transition point close to pH 6 should correspond with the number of carboxyl groups, and the span between this point and the point of inflexion near pH 8.5 with the number of imidazole groups[6]. The number of amino groups can also be deter-

[5] Simms (27) has suggested that arginine may be present in some proteins as a cyclic structure—prearginine—in which the guanidine group is replaced by a weakly basic group. The suggestion was made in order to reconcile the dissociation curves of edestin and of gelatin with analyses of the amino acids which they contained. No chemical evidence for the existence of such a structure has yet appeared.

[6] Some justification will be given later for ignoring the possibility that α-amino groups may contribute to the imidazole span.

mined with the aid of dissociation curves of the protein in solutions of formaldehyde. Since the sum of the amino, imidazole, and guanidine groups should be equal to the maximum cation charge, the number of guanidine groups can be obtained by difference.

The significance of a series of differential titrations, such as those which have been outlined, depends upon the precision with which the several end points can be located on the h-axis of the curves. Expressed in terms of a protein with a molecular weight of about 40,000, we believe that the probable error in fixing a particular point should not exceed 1 equivalent. Proteins of this size may contain more than 100 ionizing groups and, consequently, the over-all error is small. The error, however, is greatly magnified if it is concentrated in the estimate of a group, such as the imidazole group, which is usually present in proteins in small amount.

There have been many attempts to interpret the curves of proteins in terms of their constituent amino acids (6, 21, 22). The majority of these have been content to show that there is a similarity between the total basic amino acids present and the maximum acid-combining capacity. The more detailed analysis described above has, as yet, been applied only to egg albumin and to β-lactoglobulin. In both cases there is quite satisfactory agreement between the numbers of imidazole and guanidine groups in the intact protein and the moles of histidine and of arginine, respectively, which have been isolated from hydrolysates of the proteins. The numbers of titratable amino groups, however, exceed by 20 to 50 per cent the amounts of lysine isolated from the proteins. Moreover, the titratable carboxyl groups of egg albumin also exceed by 50 per cent the groups which would correspond[7] with the amounts of the dicarboxylic acids which have been isolated. No determinations of the dicarboxylic acids of β-lactoglobulin are available for comparison with the titrations. In considering these discrepancies it is significant that the amino groups measured by titration have been found to be equal to the groups in the intact protein which yield nitrogen with nitrous acid. Furthermore, we have recently obtained evidence (15) that the dicarboxylic acids present in a crude fraction of the hydrolysates of pure egg albumin greatly exceed the amounts of these acids which have been isolated from the hydrolyzed protein. Indeed, the amounts which are present in this fraction are almost sufficient to account for the titratable carboxyl groups of the native protein. It is our belief that the discrepancies which have been referred to are due, in the main, to incomplete isolation of the acidic and basic amino acids. Theorell and Åkesson (31) have suggested that the excess of free amino groups may represent α-amino groups. This may be the situation in some cases, but we doubt if more than one or two of these groups are present in egg albumin or in β-lactoglobulin. Typical α-amino groups should make their contribution to the dissociation curve close to pH 8. This is precisely the region which, in many proteins, is almost horizontal. When, moreover, imidazole groups are measured by titration between the two transition points, the major proportion of the α-amino groups

[7] It is assumed that the free carboxyl groups are equal to the difference between the total dicarboxylic acids and the amide groups.

should be included and the titration should exceed the histidine content correspondingly. This discrepancy has not been observed. The alternative is to assume that such α-amino groups as are present are atypical in that they do not contribute to the curve below pH 9. In that case there is no way in which they may be distinguished from the amino groups of the lysine side chains. Dicarboxylic and diamino acids other than those listed in table 1—e.g., hydroxyglutamic acid and hydroxylysine (33)—may be partly responsible for the discrepancies to which we have drawn attention. However, the evidence available at present suggests that the distribution of these amino acids in proteins is limited.

VIII. THEORETICAL DISSOCIATION CURVES

The dissociation curve of a polyvalent ampholyte can, as von Muralt (20) has pointed out, be simulated by that of an appropriate mixture of univalent acids. A number of reconstructions of protein curves have been made on this principle (7, 9). There is, however, considerable latitude in the choice of the number of different univalent acids which shall be employed, the proportions in which they shall be mixed, and the dissociation constants which shall be assigned to them. The choices made in a successful reconstruction, therefore, bear no necessary relation to the ionizing groups in the protein. Moreover, this method of analysis affords no rational basis for a description of the important effect of ionic strength on the shape of the curve. A more promising approach is that first outlined by Linderstrøm-Lang (17). He derived equations for a hypothetical polyvalent ampholyte which were particularly designed to describe the effect of μ on the dissociation of a protein. Unfortunately, data adequate for a comprehensive test of the usefulness of the theory were not available at the time that it was developed. Quite recently the argument has been adapted and applied to the dissociation curves of egg albumin and of β-lactoglobulin (4, 5) and some aspects of the theory have been reëxamined by Kirkwood (16).

Linderstrøm-Lang based his treatment on Bjerrum's theory (2) of electrostatic interaction in the dissociation of polyvalent acids and on the Debye–Hückel theory of interaction with the ion atmosphere. In the application of these theories to proteins, some simplifying assumptions are made. These are as follows: (*a*) that the protein molecule may be regarded as a sphere, of radius r, on the surface of which there is a random distribution of the charged groups; (*b*) that all groups of a kind may be described by a single intrinsic dissociation constant, K_0; and (*c*) that the several values of K_0, corresponding to the several kinds of ionizing groups, are such that the dissociations of the different kinds of groups do not significantly overlap.

In Kirkwood's treatment the latter assumption is not made. It is, however, a condition which seems to be approximately met in the dissociation of proteins and will be adopted in the discussion which follows. Its adoption allows the dissociation of each kind of group to be treated separately, as that of a symmetrical polyvalent acid. The separate curves may then be combined in a single curve which will deviate only slightly from the theoretical curve in those narrow regions in which significant overlapping occurs.

Let there be m acid groups of the "A" kind in a protein molecule and let the charge on the molecule be $(n - x)$, when x protons have dissociated from the "A" groups. The dissociation of the m protons may be described by m constants (K'), which are related to the m dissociation constants (K) as follows:

$$K'_x = a_{H^+} \cdot [P_{n-x}]/[P_{n-x+1}] = K_x \cdot \gamma_{n-x+1}/\gamma_{n-x} \tag{1}$$

where [P] represents the concentration and γ the activity coefficient of the protein ion the charge of which is denoted by the subscript.

On the basis of Bjerrum's theory of the dissociation of polyvalent acids, we may define K_x in terms of K_0 and a quantity $b = e^2/2\epsilon rkT$, in which e is the charge on the electron, ϵ is the dielectric constant, k is the molecular gas constant, and T is the absolute temperature. The exact relation depends on whether the acid groups of the "A" kind are uncharged or are positively charged. For an uncharged acid group, e.g., COOH, we have

$$K_x = K_0 \cdot e^{2b(n-x+1)} \cdot (m - x + 1)/x \tag{2}$$

and for a positively charged acid, e.g., NH_3^+,

$$K_x = K_0 \cdot e^{2b(n-x)} \cdot (m - x + 1)/x \tag{3}$$

The Debye–Hückel expression for the activity coefficient of a polyvalent ion is now introduced to provide a relation between K_x and K_x. We have

$$-\ln \gamma_1 = \frac{e^2}{2\epsilon kT} \cdot \frac{\kappa}{1 + \kappa a} \tag{4}$$

$$\gamma_{n-x} = e^{-(n-x)^2 \ln \gamma_1} \quad \text{and} \quad \frac{\gamma_{n-x+1}}{\gamma_{n-x}} = e^{2(n-x+0.5) \ln \gamma_1} \tag{5}$$

κ being the thickness of the ion atmosphere and a the distance of closest approach of the ions of the atmosphere to the protein ion. Introducing equation 5 into equation 1 and combining with equations 2 and 3, we have

$$K'_x = K_0^0 \cdot e^{2w(n-x+0.5)} \cdot (m - x + 1)/x \tag{6}$$

which, on conversion to common logarithms, becomes

$$pK'_x = pK_0^0 - 0.868w(n - x + 0.5) - \log \frac{m - x + 1}{x} \tag{6a}$$

In equations 6 and 6a,

$$K_0^0 = K_0 e^{\pm b}$$

$$w = b + \ln \gamma_1 = b\left[1 - \frac{\kappa r}{1 + \kappa a}\right] \tag{7}$$

The sign of the exponent in equation 7 is *plus* for the case of an uncharged acid group, and *minus* for the case of a positively charged acid group.[8]

[8] In the paper on egg albumin (4) this distinction was not made; consequently, the k_0 of that paper, in the case on carboxyl groups, differs from the K_0 defined above. On the other hand, k'_0 and K_0^0 are identical constants.

The m constants defined by equation 6a are symmetrically distributed about their mean, which is identical with $pK'_{\frac{m+1}{2}}$ and with the mid-point of the dissociation curve ($pH_{mid.}$).

$$\sum_{x=1}^{m} pK'_x/m = pK_0^0 - 0.868w(n - m/2) = pH_{mid.} \tag{8}$$

Let C_p be the bulk concentration of protein. The m equations 1 may be combined to give

$$\frac{[P_{n-x}]}{C_p} = \frac{K'_1K'_2 \cdots K'_x \cdot a_{H^+}^{m-x}}{a_{H^+}^m + \sum_{x=1}^{m} K'_1K'_2 \cdots K'_x \cdot a_{H^+}^{m-x}} \tag{9}$$

and

$$h = \frac{na_{H^+}^m + \sum_{x=1}^{m} (n - x)K'_1K'_2 \cdots K'_x \cdot a_{H^+}^{m-x}}{a_{H^+}^m + \sum_{x=1}^{m} K'_1K'_2 \cdots K'_x \cdot a_{H^+}^{m-x}} \tag{10}$$

When m and n are known, data for the construction of a theoretical curve corresponding to selected values of K_0^0 and w may be computed by introducing equation 6 into equation 10 and solving for h at a suitable range of values of pH. Now the values of w which are compatible with the sizes of protein molecules are quite limited. If we take 20 Å. to represent a minimum radius and set ϵ equal to the dielectric constant of water, then the value of w, at $\mu = 0$, will be about 0.17. As μ increases, w diminishes and approaches zero at high ionic strength. For values of w of this order, the curves are smooth and are practically linear in the region of the mid-point. Linderstrøm–Lang (17) has derived an expression for this limiting slope which may be written as

$$-\Delta pH/\Delta h = 0.868(w + 2/m) \tag{11}$$

Equation 11 provides a convenient graphic method of obtaining a value for w from any well-defined segment of an experimental curve, i.e., a segment corresponding with a fairly large value of m. The mid-point, $pH_{mid.}$, of the segment can be identified with the point at which $h = n - m/2$. By introducing this into equation 8, a value for K_0^0 may be obtained. In this way, all quantities necessary for the solution of equations 6 and 10 are made available, and a comparison can be made of the experimental curve with the theoretical curve which corresponds with these experimental values of K_0^0, w, m, and n. This procedure has been applied to the curves of egg albumin and of β-lactoglobulin. Only the carboxyl regions of the curves are sufficiently broad to permit a full analysis. It has been found that a single value of w does suffice to describe practically the whole span of dissociation of the carboxyl groups at constant μ. Moreover, w has been found to vary with μ in a manner qualitatively consistent with the theory. However, the values of w which best described the curves were all somewhat smaller than the values calculated from equation 7, employing $\epsilon = 78$ and computing r from

the diffusion constant of the protein. The ratio of w (theory) to w (found) was about 1.25 for egg albumin and between 1.05 and 1.1 for β-lactoglobulin. So many dubious assumptions are involved in the calculation of w that larger discrepancies than those found might well have been anticipated.

The values of w which were found to describe the dissociation of the carboxyl groups were used to calculate theoretical curves for the imidazole and amino groups of the two proteins. In the case of egg albumin they did fit the experimental curves quite satisfactorily. The theory, however, failed to account for the slopes and positions of the amino segments of the curves of β-lactoglobulin. The values of pK_0 which were found to apply to egg albumin were 4.2 (carboxyl), 6.7 (imidazole), and 10.0 (amino), while those of β-lactoglobulin were 4.5 (carboxyl) and 6.7 (imidazole).

IX. AN APPROXIMATION TO EQUATIONS 6 AND 10

The computations involved in the solution of equation 10 are so tedious that we have sought a simpler relation which would approximate equation 10 for a representative range of values of w and m. Now, when $w = 0$, the curve obtained (*cf.* 20) is identical with that of m equivalents of a univalent acid for which $K = K_0^0$. That is to say, if α represents the fraction of the m groups which have dissociated a proton, then

$$\mathrm{pH} = \mathrm{p}K_0^0 - \log (1 - \alpha)/\alpha \tag{12}$$

A series of curves for a range of positive values of w and a fixed value of m were then constructed from equation 10. When these were compared with the curve for equation 12, it was found that they were displaced from it by an amount approximately equal to $0.868wh$. That is to say,

$$\mathrm{pH} = \mathrm{p}K_0^0 - \log (1 - \alpha)/\alpha - 0.868wh \tag{13}$$

That this relation should hold in the linear segments of the curve is implicit in equation 11. It was encouraging to find that it applied with fair precision over the whole course of a family of curves corresponding to a range of values of w wider than that to be expected in proteins.

At the mid-point ($\mathrm{pH_{mid.}}$) of the curves, $\alpha = 0.5$. Consequently, since $\alpha = (n - h)/m$,

$$\mathrm{pH_{mid.}} = \mathrm{p}K_0^0 - 0.858w(n - 0.5m)$$

and

$$\mathrm{pH} = \mathrm{pH_{mid.}} - \log (1 - \alpha)/\alpha - 0.868wm(\alpha - 0.5) \tag{14}$$

According to equation 14, the slope of the curve relating pH to α should be uniquely determined by the value of the product wm. This offers a simple means of exploring the magnitude of the deviations of the approximate equation 14 from equation 10. The situation is illustrated in table 2, wherein the true values of α at constant wm, but variable m, are compared with the corresponding values, α', calculated from equation 14. It will be seen that α' approximates more and

more closely to α as m increases. However, as is seen in the last column, when the deviations are expressed in equivalents rather than in terms of α, they are substantially independent of m. Moreover they exceed 0.1 equivalent at no point on the curves. It is doubtful if the experimental points on the curve of a protein can be reproduced with a smaller probable error than this.

TABLE 2

Comparison of equations 10 and 14

$pK_0^0 = 4.00$; $n = m/2$

				$wm = 2.82$			
m.....	4	6	10	20	50	α' (EQUATION 14)	$m(\alpha - \alpha')$
w.....	0.705	0.470	0.282	0.141	0.0564		
pH	α (EQUATIONS 6 AND 10)						
4.00	0.500	0.500	0.500	0.500	0.500	0.500	
4.3	0.576	0.574	0.573	0.572	0.572	0.571	0.02
4.6	0.651	0.648	0.645	0.643	0.642	0.642	0 04
4.9	0.723	0.718	0.715	0.712	0.710	0.709	0.06
5.2	0.794	0.786	0.781	0.776	0.774	0.773	0.08
5.5	0.852	0.845	0.839	0.835	0.832	0.831	0.08
5.8	0.907	0.897	0.890	0.885	0.882	0.881	0.09
6.1	0.944	0.937	0.930	0.925	0.923	0.921	0.09
6.4	0.970	0.964	0.961	0.957	0.954	0.952	0.08
6.7	0.986	0.981	0.977	0.975	0.973	0.972	0.05
				$wm = 1.212$			
w.....	0.303	0.202	0.1212	0.0606	0.0242	α'	$m(\alpha - \alpha')$
pH	α						
4.3	0.614	0.611	0.609	0.607	0.606	0.606	0.03
4.6	0.720	0.715	0.712	0.709	0.707	0.707	0.05
4.9	0.811	0.806	0.802	0.799	0.797	0.796	0.06
5.2	0.885	0.878	0.874	0.870	0.867	0.867	0.07
5.5	0.933	0.929	0.926	0.922	0.921	0.920	0.05
5.8	0.964	0.961	0.958	0.956	0.955	0.954	0.04
6.1	0.983	0.980	0.978	0.976	0.976	0.975	0.03
6.4	0.992	0.990	0.989	0.988	0.987	0.987	0.02

X. SOME DEDUCTIONS FROM THE THEORY

A problem of some interest is that of the effect of molecular size on the dissociation curve of protein. This may be approached by considering how the polymerization of a protein would be reflected in equation 14. The value of m would, obviously, increase in proportion to the increase in molecular weight. On the other hand, w varies with $1/(r + \kappa r^2)$, assuming that $a = r$. Consequently, wm will be proportional to $r^2/(1 + \kappa r)$. In the case of a molecule of the size of the smaller proteins, κr will exceed unity at ionic strengths as low as 0.1. Under these conditions, the limiting relation, in which wm varies with r

(i. e., with $\sqrt[3]{\text{mol. wt.}}$), will be approached. Consequently the span of the curve on the pH-axis should increase with increase in size of the molecule, but large changes in size may be necessary to bring the effect clearly into evidence.

The size of a molecule may also increase through hydration. In this case, since m remains constant, the curve should contract as r increases.

In studies of protein solutions, it is often desirable to estimate the contribution of the protein ions to the ionic strength. A common, but empirical, practice has been to assume that it is equal to one-half the equivalent concentration of the net charge, i.e., to $hC_p/2$. In a protein solution, for example, in which the concentrations of added hydrochloric acid and sodium chloride were [HCl] and [NaCl], we would compute μ from

$$2\mu = [H^+] + [HCl] + 2[NaCl] + h \cdot C_p$$

TABLE 3

The ionic strength of protein ions (β-lactoglobulin)

h	$C_p \times 10^3$, 1.23 [NaCl] $\times 10^3$, 9.0				$C_p \times 10^3$, 0.61 [NaCl] $\times 10^3$, 4.5			
	pH	$\frac{2w}{2.3}$	μ'	μ	pH	$\frac{2w}{2.3}$	μ'	μ
40	2.35	0.033	0.070	0.063	2.24	0.037	0.050	0.035
35	2.78	0.034	0.066	0.054	2.63	0.038	0.045	0.029
30	3.10	0.036	0.055	0.046	2.95	0.041	0.034	0.024
25	3.42	0.038	0.045	0.039	3.25	0.044	0.026	0.021
20	3.70	0.041	0.034	0.033	3.56	0.048	0.018	0.017
15	3.98	0.045	0.024	0.026	3.88	0.052	0.013	0.013
10	4.33	0.048	0.018	0.020	4.25	0.056	0.010	0.010
5	4.74	0.048	0.018	0.014	4.70	0.056	0.010	0.007
0	5.19				5.19			

μ = ionic strength calculated from equation 15.
μ' = ionic strength calculated from w and equation 4.
C_p calculated from molecular weight = 40,000.

Since electrical neutrality requires that $[HCl] = [H^+] + h \cdot C_p$, this reduces to

$$\mu = [H^+] + [NaCl] + h \cdot C_p \qquad (15)$$

If [NaCl] is low, μ will increase with h as well as with C_p. We suggest that this may be sufficient explanation for the striking difference in shape between a curve at constant μ and the curve of a protein solution in the substantial absence of salt. In table 3 data from two curves of β-lactoglobulin are recorded. A very small amount of sodium chloride was added to prevent precipitation of the protein at the isoelectric point. With the aid of equation 14 a value of w has been calculated for each point on each curve and the corresponding value of μ has been calculated from equation 7. This is a rather precarious calculation, and the values obtained are to be considered only as coarse approximations. They are compared, in table 3, with values of μ calculated directly from equa-

tion 15. It is evident that the effect of the concentration of protein on the dissociation curve is not inconsistent with equation 15.

One final deduction from the theory may be of interest. It touches the question of the nature of the assembly of protein ions in a protein solution. The quantity h represents only the time average of the net charge. It is evident that significant numbers of ions differing materially in charge from h may be present in a given solution. The average concentration of an ion of any particular charge may be calculated from equation 9. To illustrate the situation we have chosen an isoelectric solution of β-lactoglobulin. In figure 2, curves are reproduced giving the distribution of ions, according to equation 9, in this solution at three different ionic strengths.

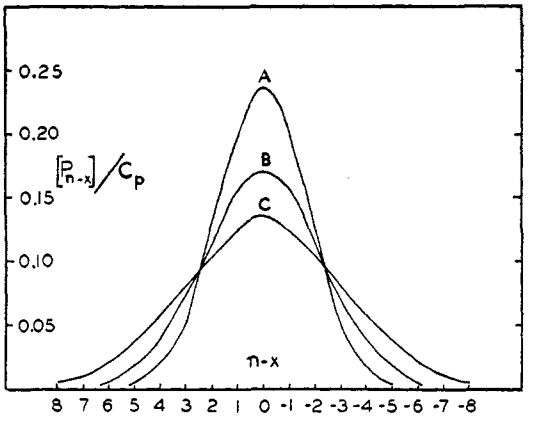

FIG. 2. Isoionic point of β-lactoglobulin. Theoretical distribution of ions. Curve A, $\mu = 0.0$; curve B, $\mu = 0.069$; curve C, $\mu = 2.1$.

REFERENCES

(1) ADAIR, G. A., AND ADAIR, M. E.: Biochem. J. **38,** 199, 1230 (1934); Trans. Faraday Soc. **36,** 23 (1940).
(2) BJERRUM, N.: Z. physik. Chem. **106,** 219 (1923).
(3) BJERRUM, N.: Kgl. Danske Videnskab. Selskab, Math.-fys. Medd. **7,** 9 (1926–27).
(4) CANNAN, R. K., KIBRICK, A. C., AND PALMER, A. H.: Ann. N. Y. Acad. Sci. **41,** 243 (1941).
(5) CANNAN, R. K., PALMER, A. H., AND KIBRICK, A. C.: J. Biol. Chem. **142,** 803 (1942).
(6) COHN, E. J.: Physiol. Rev. **5,** 349 (1935).
(7) COHN, E. J.: Cold Spring Harbor Symposia Quant. Biol. **6,** 8 (1938).
(8) COHN, E. J., EDSALL, J. T., AND BLANCHARD, M. H.: J. Biol. Chem. **105,** 319 (1934).
(9) COHN, E. J., GREEN, A. A., AND BLANCHARD, M. H.: J. Am. Chem. Soc. **59,** 509 (1937).
(10) GREENBERG, D. M., AND SCHMIDT, C. L. A.: Proc. Soc. Exptl. Biol. Med. **21,** 281 (1924).
(11) GUGGENHEIM, E. A., AND SCHINDLER, J. D.: J. Phys. Chem. **38,** 533 (1934).
(12) HITCHCOCK, D. I.: J. Gen. Physiol. **16,** 357 (1932–33).
(13) HITCHCOCK, D. I., AND TAYLOR, A. C.: J. Am. Chem. Soc. **59,** 1812 (1937).
(14) KEKWICK, R. A., AND CANNAN, R. K.: Biochem. J. **30,** 235 (1936).
(15) KIBRICK, A. C.: Unpublished observations.
(16) KIRKWOOD, J. G.: Ann. N. Y. Acad. Sci. **41,** 321 (1941).
(17) LINDERSTRØM-LANG, K.: Compt. rend. trav. lab. Carlsberg **15,** No. 7 (1923–25).

(18) LONGSWORTH, L. G.: Ann. N. Y. Acad. Sci. **41**, 267 (1941).
(19) MACINNES, D. A.: *The Principles of Electrochemistry*, Chap. XV. Reinhold Publishing Corporation, New York (1939).
(20) MURALT, A. L. VON: J. Am. Chem. Soc. **52**, 3518 (1930).
(21) NEWBERGER, A.: Biochem. J. **28**, 1982 (1934).
(22) PAULI, W., AND VALKÓ, E.: *Kolloidchemie der Eiweisskörper*. Steinkopff, Dresden (1933).
(23) PEDERSEN, K. O.: Biochem. J. **30**, 948 (1936).
(24) RASMUSSEN, K. E., AND LINDERSTRØM-LANG, K.: Compt. rend. trav. lab. Carlsberg **20**, No. 10 (1933-35).
(25) SALK, J., AND CANNAN, R. K.: Unpublished observations.
(26) SANDOR, G.: Bull. soc. chim. biol. **18**, 877 (1936).
(27) SIMMS, H. S.: J. Gen. Physiol. **14**, 87 (1931).
(28) SMITH, E. R. B.: J. Biol. Chem. **113**, 473 (1936).
(29) SØRENSEN, S. P. L., LINDERSTRØM-LANG, K., AND LUND, E.: Compt. rend. trav. lab. Carlsberg **16**, No. 5 (1926).
(30) STEINHARDT, J.: Ann. N. Y. Acad. Sci. **41**, 287 (1941).
(31) THEORELL, H., AND ÅKESSON, A.: J. Am. Chem. Soc. **63**, 1818 (1941).
(32) TISELIUS, A., AND SVENSSON, H.: Trans. Faraday Soc. **36**, 16 (1940).
(33) VAN SLYKE, D. D., HILLER, A., AND MACFAYDEN, D. A.: J. Biol. Chem. **141**, 681 (1941).
(34) WYMAN, J., JR.: J. Biol. Chem. **127**, 1 (1939).
(35) WYMAN, J., JR.: J. Biol. Chem. **127**, 581 (1939).