Anal. Chem. **1997,** *69,* 575−580

Using Protein Charge Ladders To Estimate the Effective Charges and Molecular Weights of Proteins in Solution

Jinming Gao and George M. Whitesides*

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

This paper describes the use of capillary electrophoresis (CE) and protein charge ladders to estimate values of effective charge (*Z*) and molecular weight of proteins under nondenaturing conditions. A panel of 14 proteins with a range of charges and shapes was modified by acetylation with acetic anhydride to yield protein charge ladders. A protein charge ladder is a family of derivatives of a protein that differ in integral units of charge, but minimally in hydrodynamic drag; this mixture of proteins appears in electrophoresis as a set of peaks with regular spacings. Analysis of the electrophoretic mobilities of the members of these charge ladders yields values of *Z* and electrophoretic coefficients: for a description of mobility based on the equation $\mu = C_P Z(\mathrm{MW})^{-\alpha}$, $C_P = 6.3$ cm^2 min^{-1} kV^{-1} charge^{-1} kD$^{0.48}$, $\alpha = 0.48$; for $\mu = C_r Z[r(1 + \kappa r)]^{-1}$, $C_r = 55$ cm^2 min^{-1} kV^{-1} charge^{-1} Å (*r* is the spherical radius of the protein and κ is a function of ionic strength). The primary usefulness of charge ladders is in measuring the effective charge, *Z*, of proteins in solution; this information is difficult to obtain by any other procedure. A secondary value of the method is to estimate values of molecular weight. Although less general and convenient than SDS−PAGE, this method allows estimates of molecular weight of nondenatured proteins and is thus applicable to oligomers, noncovalent aggregates, proteins with multiple, non-cross-linked chains, and other systems to which SDS−PAGE is not applicable. The values of molecular weight calculated using the electrophoretic mobilities of proteins in solution and the above constants agreed with literature values to within 20% (with an ambiguous result for ovalbumin). A combination of this technique and SDS−PAGE will be useful in estimating the number of subunits or stage of aggregation of proteins in solution.

This paper describes a method for obtaining values of both effective charge and molecular weight for native proteins by capillary electrophoresis (CE). Capillary electrophoresis is an analytical technique that provides information simultaneously about effective charge and hydrodynamic drag of a protein in solution.[1−7] Hydrodynamic drag is related to the shape and size of a molecule; eq 1 provides one form of an equation commonly

$$\mu = C_P(Z/\mathrm{MW}^{\alpha}) \quad (1)$$

used to relate these properties to electrophoretic mobility (μ); MW is molecular weight, and *Z* is effective charge.[1,8,9] Deconvolution of charge, drag, and shape, based on studies of electrophoretic mobility, has been difficult, since neither the values nor the constancy of the parameters α and C_P has been established. Theoretical values of α range from 0.3 to 1.0, depending on the model used to relate molecular weight to hydrodynamic drag;[1] C_P might also vary with details of protein shape and structure and conditions of the experiments.

Equation 2 represents another form relating the electrophoretic mobility to the size of a protein.[10] It is derived from Debye−

$$\mu = C_r(Z/r(1 + \kappa r)) \quad (2)$$

Hückel−Henry theory: *r* is the radius of a spherical analyte and κ is a function of ionic strength;[11] the reciprocal of κ is the thickness of the double layer.[10,12] Here we analyze experimental mobilities obtained with protein charge ladders to estimate values of the constants C_P, α, and C_r and use these values to estimate the molecular weights of proteins in solution.

A protein charge ladder is a family of derivatives of a protein obtained from modification of charges on the protein.[13−15] If the reagent used in the derivatization is small (typically acetyl group, MW = 42) relative to the molecular weight of the protein, we assume that hydrodynamic radius and hydrodynamic drag are not significantly influenced by the derivatization. If the reagent is

(1) Grossman, P. D. *Capillary Electrophoresis: Theory and Practice*; Academic Press: San Diego, CA, 1992; pp 111−132.
(2) Compton, B. J.; O'Grady, E. A. *Anal. Chem.* **1991**, *63,* 2597−2602.
(3) Engelhardt, H.; Beck, W.; Kohr, J.; Schmitt, T. *Angew. Chem., Int. Ed. Engl.* **1993**, *32,* 629−766.
(4) Kuhr, W. G.; Monnig, C. A. *Anal. Chem.* **1992**, *64,* 389−407.
(5) Novotny, M. V.; Cobb, K. A.; Liu, J. *Electrophoresis* **1990**, *11,* 735−749.
(6) Karger, B. L.; Cohen, A. S.; Guttman, A. *J. Chromatogr.* **1989**, *492,* 585−614.
(7) Gordon, M. J.; Huang, X.; Pentoney, S. L.; Zare, R. N. *Science* **1988**, *242,* 224−228.
(8) Basak, S. K.; Ladisch, M. R. *Anal. Biochem.* **1995**, *226,* 51−58.
(9) Rickard, E. C.; Strohl, M. M.; Nielsen, R. G. *Anal. Biochem.* **1991**, *197,* 197−207.
(10) Abramson, H. A.; Moyer, L. S.; Gorin, M. H. *Electrophoresis of Proteins and the Chemistry of Cell Surfaces*; Reinhold Publishing Corp.: New York, 1942; pp 105−172.
(11) The typical range of ionic strength that is compatible with CE analysis is between 1 mM and 1 M.
(12) $\kappa = (2000N_A e^2 I/\epsilon\epsilon_o kT))^{1/2}$, where N_A is Avogadro's number, *e* is the magnitude of the elementary charge, *I* is ionic strength in solution, ϵ is dielectric constant, ϵ_o is the permitivity of free space, *k* is the Boltzmann constant, and *T* is the absolute temperature. See ref 1.
(13) Gao, J.; Mammen, M.; Whitesides, G. M. *Science* **1996**, *272,* 535−537.
(14) Gao, J.; Mrksich, M.; Gomez, F. A.; Whitesides, G. M. *Anal. Chem.* **1995**, *34,* 3093−3100.
(15) Gao, J.; Gomez, F.; Haerter, R.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91,* 12027−12030.

S0003-2700(96)00807-4 CCC: $14.00

selective for functional groups that carry a full unit of charge (for example, the lysine ϵ-NH_3^+ group[16]), the modification (ϵ-$NH_3^+ \rightarrow \epsilon$-$NHCOCH_3$) will give a mixture of modified proteins that differ by integral values in charge. This mixture of proteins appears in electrophoresis as a set of peaks with regular spacings. The intervals in μ establish the influence of effective charge on mobility and allow the effective charge of the native protein to be determined directly.[15] Recently we demonstrated that the charge ladder of a protein is useful in estimating the influence of charge on protein−ligand interactions.[13]

In this paper, we extend our analysis of protein charge ladders to estimate values for α and C_P for a set of 14 proteins chosen to represent low and medium molecular weight enzymes, receptors, and binding proteins. This analysis suggests that a common pair of values of these two constants predicts the electrophoretic mobilities of these proteins with a standard deviation of 0.6 $cm^2\ kV^{-1}\ min^{-1}$ (that is, <10% of μ). Analysis using eq 2 gave the same value of standard deviation, with a similar correlation coefficient. Using these values, we show that information concerning the mobilities of the members of a charge ladder makes it possible to estimate both the effective charge and the molecular weight of a protein in its native conformation simultaneously.

This procedure has several useful applications. Its simplicity makes it an attractive method to determine effective charge (a parameter that is not readily available by another technique). It can estimate the molecular weights of proteins in nondenaturing native conditions, and therefore supplements information from SDS−PAGE. It allows the study of the fundamental electrophoretic behavior of proteins in solution, that is, the influence of shapes, dipoles, and charge distributions on their electrophoretic mobilities.

EXPERIMENTAL SECTION

CE Equipment. Beckman P/ACE system 5510 capillary electrophoresis systems were used in these studies. The capillary tubing (Polymicro Technologies, Inc., Phoenix, AZ) was of uncoated fused silica with an internal diameter of 50 μm. Some capillaries were coated with polyethylenimine: coating was performed according to literature procedures.[17] The conditions used for all CE experiments were as follows: voltage, 20 kV; detection wave length, 214 nm; total length of capillary, 47 cm; length of capillary from injecting end to detector, 40 cm; temperature, 37 °C. Samples (5 nL) were introduced into the capillary by pressure injection. All the protein samples were analyzed in the buffer of 25 mM Tris−192 mM Gly (pH 8.3).

Reagents. All chemicals were analytical grade and were used as received. Acetic anhydride and dioxane were purchased from Aldrich. Mesityl oxide (MO) was purchased from Eastman Organic Chemical. Insulin, bovine carbonic anhydrase II, human carbonic anhydrase II, horse heart myoglobin, cytochrome *c*, ovalbumin, and α-chymotrypsinogen A were purchased from Sigma. Carboxypeptidase B, lysozyme, superoxide dismutase, α-lactalbumin, ribonuclease A, bovine pancreatic trypsin inhibitor, and peroxidase were purchased from Worthington.

(16) Modification of Arg residues on a protein by phenylglyoxal will also yield a charge ladder with integral unit of charge difference: Takahashi, K. *J. Biol. Chem.* **1968**, *243*, 6171−6179. Roberts, C.; Córdova, E.; Whitesides, G. M., unpublished results.

(17) Córdova, E.; Gao, J.; Whitesides, G. M. *Anal. Chem.*, submitted.

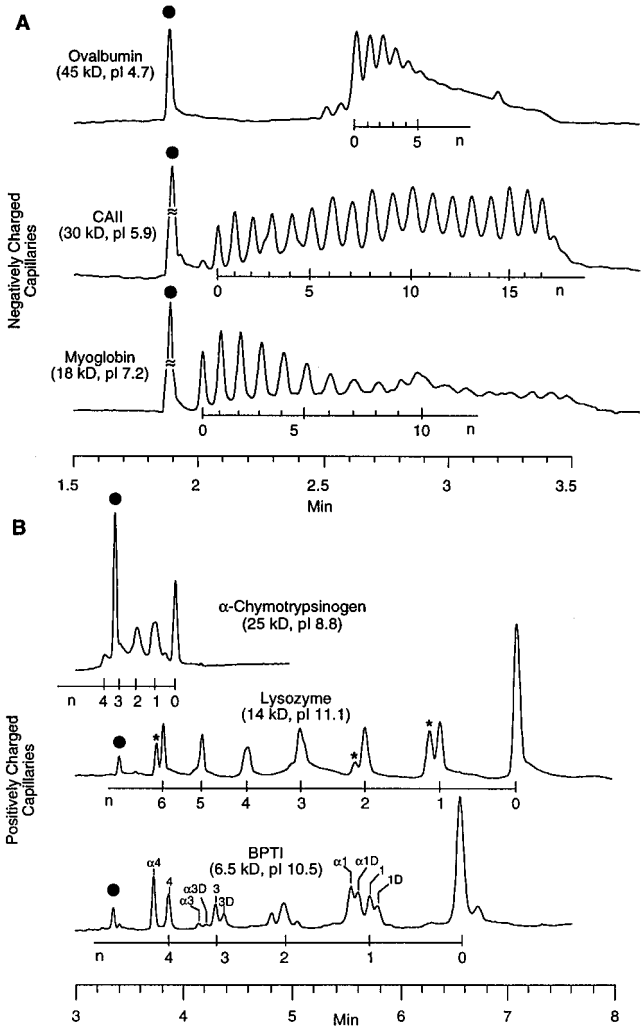

Figure 1. Representative electropherograms of charge ladders of proteins. The electropherograms are ordered so that the molecular weight of the protein increases from bottom to top in both (A) and (B). The charge ladders were formed by acetylation directed predominantly toward the ϵ-amino groups of Lys. Mesityl oxide was used as a neutral marker and is indicated by the filled circles in the electropherograms. The electrophoresis buffer is 25 mM Tris−192 mM Gly buffer (pH 8.3). (A) Charge ladder of proteins having values of p*I* ≤ 7.4, analyzed using uncoated capillaries. (B) Charge ladder of proteins having values of p*I* ≥ 8.8, analyzed using polyethyleniminecoated capillaries (see the text for an explanation of the labels on the electropherograms of lysozyme and BPTI). The number of acetylated Lys ϵ-amino groups (*n*) is indicated below each electropherogram. The time scale on the bottom of each figure applies to all the electropherograms.

Preparation of Charge Ladders of Proteins. The general procedure for the preparation of a charge ladder of a protein follows. The pH of aliquots of a solution of a protein (~3 mg/mL, 0.2 mL) was adjusted to pH 12 with 0.1 N sodium hydroxide. Acetic anhydride (5, 10, or 20 equiv; 100 mM stock solution in dioxane) was added. After 5 min at room temperature, 10 μL of a sample was diluted with 90 μL of the electrophoresis buffer, MO was added (final concentration 18 mM), and the samples were analyzed by CE. Different amounts of lightly and heavily acetylated samples were combined and diluted in the electrophoresis buffer to yield a charge ladder that contained balanced quantities of differently charged derivatives (with a total concentration of proteins of ~0.3 mg/mL). Figure 1 shows the electropherograms of some of these samples.

Table 1. Panel of Proteins

ID	proteins	EC	subunits[a]	p*I*[a]	MW[a] ($\times 10^{-3}$)	volume[b] (Å^3)	r^c (Å)	capillary charge[d]
1	insulin		1	5.5	5.7	6 817	11.8	−
2	bovine pancreatic trypsin inhibitor		1	10.5	6.5	7 774	12.3	+
3	cytochrome *c*		1	9.3	12.5	14 950	15.3	+
4	ribonuclease A	3.1.27.5	1	9.5	13.7	16 385	15.8	+
5	lysozyme	3.2.1.7	1	11.1	14.0	16 744	15.9	+
6	α-lactalbumin		1	4.8	14.2	16 983	15.9	−
7	myoglobin		1	7.2	18.0	21 528	17.3	−
8	α-chymotrypsinogen	3.4.21.1	1	8.8	25.0	29 900	19.3	+
9	bovine carbonic anhydrase	4.2.1.1	1	5.9	30.0	35 880	20.5	−
10	human carbonic anhydrase	4.2.1.1	1	7.4	30.0	35 880	20.5	−
11	superoxide dismutase	1.15.1.1	2	5.0	32.5	38 870	21.0	−
12	carboxypeptidase B	3.4.17.2	1	6.0	34.3	41 023	21.4	−
13	peroxidase	1.11.1.7	1	7.2	40.0	47 840	22.5	−
14	ovalbumin		1 or 2	4.7	45.0	53 820	23.4	−

[a] See refs 18 and 19. [b] The volumes of these proteins were calculated using their molecular weight and 0.72 mL/g as the average value of partial volumes for the proteins.[23] [c] Spherical radius, calculated by assuming that all the proteins are spherically shaped. The radii of these proteins were calculated from their volumes using $r = (3V/4\pi)^{1/3}$. [d] Capillary charge indicates the charge on the surface of the capillaries. An uncoated capillary is negatively charged, and a polyethylenimine-coated capillary is positively charged.

Analysis Using Polyethylenimine-Coated Capillary. Each new fused-silica capillary was flushed (velocity 20 cm/min) with 0.1 N NaOH for 15 min and then with deionized water for 15 min. After preconditioning, the capillary was coated with polyethylenimine by flushing the capillary with a 7.5% (w/v) polymer solution, prepared in 25 mM Tris−192 mM Gly buffer (pH 8.3), for 15 min. Finally, the capillary was flushed with the electrophoresis buffer (absence of polymer) for 5 min. Recoating of the capillary with the polymer was accomplished using a similar method. The previously coated capillary was flushed with 0.1 N acetic acid (pH 5.0) for 15 min and then with deionized water for 15 min. After rinsing, the capillary was recoated with the polymer by flushing the capillary with the 7.5% polymer solution for 15 min. Finally, the capillary was flushed with the electrophoresis buffer (absence of polymer) for 5 min before analysis.

RESULTS AND DISCUSSION

Panel of Proteins (Table 1**).** The panel was chosen to represent proteins of low and medium molecular weight with a broad range of charges and shapes. The molecular weight for this set of proteins ranges from 5700 to 45 000, and p*I* ranges from 4.7 to 11.1.[18,19] Except for superoxide dismutase and perhaps ovalbumin (see below), all the proteins have a single subunit. Superoxide dismutase has two identical subunits that are connected by a Cys−Cys disulfide bond. Peroxidase and ovalbumin are glycosylated.[18]

Electropherograms of Charge Ladders of Proteins Formed by Acetic Anhydride (Figure 1). We formed the protein charge ladders by modifying the ϵ-NH_3^+ groups on the proteins to neutral *N*-acetyl groups by acetylation with acetic anhydride. Each modification changed the net charge on the protein by subtraction of one unit of positive charge. Figure 1A shows the electropherograms of derivatives of negatively charged proteins, analyzed using uncoated capillaries. A scale below each electropherogram indicates the number of ϵ-amino groups being modified. Since positively charged proteins tend to stick to the negatively charged surface of the uncoated capillaries, we coated the surface of the capillary with a positively charged polymer, polyethylenimine.[17] Polyethylenimine reverses the charges on the surface of the capillary, and also the direction of electroosmotic flow. The reversed electroosmotic flow results in the reverse ordering of emergence times for the differently charged derivatives of proteins compared to the uncoated capillaries (Figure 1B).

For some protein samples, we also modified the α-NH_2 groups, still using acetic anhydride. The α-NH_3^+ groups of proteins have lower values of pK_a than do the ϵ-NH_3^+ groups of Lys. When the pH of the solution is close to the value of pK_a of the α-NH_3^+ group, the *average* charge on that group is a fraction of a charge. Acetylation of this mixture of α-NH_2 and α-NH_3^+ changes the charge by less than 1.[20] For example, the charge ladder of lysozyme contains derivatives in which the α-NH_2 group is modified (Figure 1B, peaks labeled with asterisk in the charge ladder of lysozyme). The fractional charge of these derivatives can be calculated using the value of pK_a of the α-NH_3^+ group and the pH in the medium. The value of pK_a of a charged group on a protein can be determined using CE by analyzing the differences of mobility between two derivatives of the protein (one has the unmodified charged group, and the other neutralizes the charge on this group) as a function of pH.[14,17]

In the charge ladder of bovine pancreatic trypsin inhibitor (BPTI), we also observed multiple peaks for derivatives having the same number of modifying acetyl groups. The difference of electrophoretic mobility for derivatives from modification of α-amino group (for example, 4 and α4 in Figure 1B) is sensitive to the pH in the buffer; analysis of $\Delta\mu$ ($\Delta\mu = \mu_4 - \mu_{\alpha 4}$) as a function of pH establishes the value of pK_a of the α-NH_3^+ group to be 8.0 ± 0.1.[14,17] The difference in electrophoretic mobilities for 1D and 1 (peaks as indicated in the BPTI charge ladder in Figure 1B) might, in principle, be due to the change in hydrodynamic drag between the two derivatives or to an ϵ-NH_3^+ group with an abnormal value of pK_a. We found that the difference of electrophoretic mobilities between 1D and 1 is insensitive to the change of pH in solution and conclude that the more probable of the two explanations is that at least two derivatives with the same number

(18) Worthington, V. *Worthington Enzyme Manual*; Worthington Biochemical Corp.: Freehold, NJ, 1993.

(19) Righetti, P. G.; Caravaggio, T. *J. Chromatogr.* **1976**, *127*, 1−28.

(20) The peaks due to modification on the α-amino group of a protein are assigned by the difference of chemical reactivity of the α-amino group toward acetic anhydride compared to that of an ϵ-amino group. See ref 14.

of acetylated groups differ significantly in drag.[21] In this study, we used the set of derivatives indicated by the scale below the electropherogram to estimate the hydrodynamic drag of BPTI (Figure 1B).

Generality of the Procedure To Form Protein Charge Ladders. The proteins listed in Table 1 (p*I* from 4.7 to 11.1, MW from 5700 to 45 000) all gave interpretable charge ladders. In our initial survey, several proteins (for example, phospholipase *C*, *β*-casein, and ferritin) did not give interpretable and resolvable charge ladders upon acetylation. We have not analyzed these failures but note several possible contributions to them. First, commercial samples of proteins may contain impurities that are peptide based and can also be modified by acetic anhydride. The peaks due to impurities or derivatives of impurities may complicate the assignment of the charge ladder of interest. Second, modification of a protein may cause denaturation or a change in the hydrodynamic drag of the protein. The derivatives of the protein would yield broader peaks than that for the native protein. Third, high molecular weight proteins (for example, ferritin, p*I* 4.3, MW 440 000) have large hydrodynamic drag (i.e., large values of M^{α}, eq 1), which reduces the change in mobility that is observed per unit change in *Z*. For high molecular weight proteins, modification of charge by acetic anhydride may result in an insufficient charge increment for separation of the members of the charge ladders to be achieved by CE. In such cases, a polycharged modifying agent that gives larger charge difference upon modification is required to yield a resolvable charge ladder of the protein.

Determination of the Electrophoretic Coefficient $C = \Delta\mu/\Delta n$ from Protein Charge Ladders. The electrophoretic mobilities (μ_n) are proportional to the total charge of a protein [$Z_P(n)$, where *n* represents the number of ϵ-NH_3^+ groups modified to neutral *N*-acetyl derivatives] in solution (eqs 1 and 2). A plot of the electrophoretic mobility vs the added charge (*n* in eq 3, also

$$\mu_n \approx CZ_P(n) = CZ_P(0) - nC \tag{3}$$

where, for the relationship of eq 1

$$C = C_P/\mathrm{MW}^{\alpha} \tag{4}$$

and, for that of eq 2

$$C = C_r/r(1 + \kappa r) \tag{5}$$

shown under each electropherogram in Figure 1) yielded the electrophoretic coefficient, $C = \Delta\mu/\Delta n$, as the slope of the line, and the effective charge [$Z_P(0)$] of the native protein as the *x* intercept (eq 3). For each protein charge ladder, we used the first five charged derivatives of the protein in the analyses to minimize the influence of possible changes of hydrodynamic drag originating in heavy modifications of the protein. All the linear analyses gave high correlation coefficients ($R > 0.99$). The values of *C* so obtained were further correlated to the drag of the protein according to eqs 4 and 5 to establish the values of constant C_P, α, or C_r in eq 1 or 2, respectively. The values of $Z_P(0)$ were compared to the theoretical charges that are calculated from the amino acid sequences of the proteins (see below).

(21) The suggestion of differences in drag is a very much unproved hypothesis. If it is correct, it might be caused by the change of conformation of the protein upon modification.

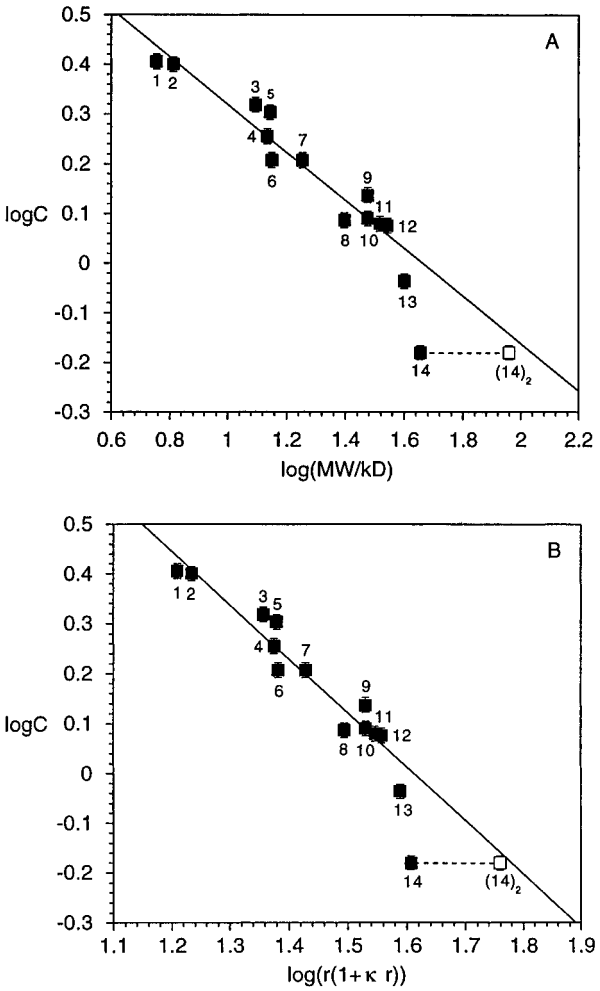

Figure 2. Correlation of log *C* with values of log MW (A) and log(*r*(1 + *κr*)) (B) for 13 proteins (except ovalbumin) in Table 1. The correlation in Figure 2A yields α = 0.48 and C_P = 6.3 cm^2 min^{-1} kV^{-1} charge^{-1} kDa$^{0.48}$ with a correlation coefficient R = 0.94. The correlation in (B) yields 1.0 for the slope and C_r = 55 cm^2 min^{-1} kV^{-1} charge^{-1} Å (R = 0.96). We excluded monomeric ovalbumin in the linear analysis due to its obvious deviation, but note that a possible dimeric form indicated by open squares fits the analysis very well. The error bars indicate the uncertainties in the experiments.

Determination of Values of C_P, α, and C_r. Equation 4 gives the relationship between the electrophoretic coefficient *C* and the values of C_P and α. A plot of log *C* vs log MW yields a linear relationship (Figure 2A); the slope of the line, 0.48, represents the value of α. The value of C_P was obtained from the *y* intercept and is equal to 6.3 cm^2 min^{-1} kV^{-1} charge^{-1} kDa$^{0.48}$ with correlation coefficient R = 0.94 for 13 protein samples (we exclude ovalbumin from the analysis).[22] The datum for ovalbumin was omitted due to its obvious deviation (as a monomer) from those for other proteins. We attribute this deviation to the dimerization of ovalbumin in solution (see discussion below).

To determine the values of C_r, we assumed a spherical shape for each protein. The molecular volume of the protein was calculated from its molecular weight using 0.72 mL/g as the average partial volume (the reciprocal of density) of proteins,[23] and the radius of the protein was calculated from the molecular

(22) The value of α is only mildly dependent on the ionic strength in solution. Over a range in ionic strength from 2 to 200 mM, the value of α changes from 0.46 to 0.49. The value of C_P, however, changes over this range in ionic strength from 4.2 to 6.8 cm^2 min^{-1} kV^{-1} charge^{-1} kDa$^{\alpha}$.

volume using the equation $r = (3V/4\pi)^{1/3}$ (Table 1). The calculated values of molecular volume and radius were consistent with those calculated by using the Quanta program from the crystal structures of the proteins. The value of $1/\kappa$ (the thickness of double layer) was calculated to be 31 Å using a value for the ionic strength of 0.01 M for 25 mM Tris−192 mM Gly buffer (pH 8.3). A plot of log C vs $\log(r(1 + \kappa r))$ is linear ($R = 0.96$) with a slope of 1.0; the correlation suggests that the expression $r(1 + \kappa r)$ may be useful in correcting the influence of ionic strength on electrophoretic mobilities. The value of C_r from the y intercept is 55 cm^2 min^{-1} kV^{-1} charge^{-1} Å.

Analysis using eqs 1 and 2 showed similar patterns of deviation for the individual proteins, indicating that the assumptions in both equations are similar (for example, both equations ignore shapes, densities, electric dipoles and quadrupoles, and charge distributions on the proteins). Further molecular modeling studies will be useful to correlate these factors to rationalize the second-order influences of these characteristics on the electrophoretic mobilities of the proteins.

Estimation of Charge and MW Based on Charge Ladders of Proteins. Figure 3A compares effective charges of proteins measured experimentally from charge ladders to those calculated from their amino acid sequences. The effective charges agree well with the calculated charges except for ovalbumin, considered as a monomer. We did not expect closer agreement: the values of pK_a used in these calculations are standard values and do not take into account nonstandard values of pK_a. The effective charge of ovalbumin measured from the charge ladder experiment is nearly twice that calculated from its amino acid sequence and again suggests that ovalbumin is a dimer in solution.

Using the electrophoretic mobilities of proteins, and the values of C_P and α obtained previously, we calculated the molecular weight of proteins directly by eq 1 (Figure 3B). As a comparison, using the electrophoretic mobilities of proteins and the value of C_r, we solved a quadratic equation (eq 2) to obtain the values of hydrodynamic radius (r) of proteins. The hydrodynamic volumes were calculated from the hydrodynamic radius by using $V = 4\pi r^3/3$ and assuming that all the proteins are spherically shaped. The molecular weights of proteins were calculated by dividing the hydrodynamic volume by the partial volume of proteins (0.72 mL/g). Figure 3B compares the values of molecular weight of proteins estimated from charge ladders with those from literature. These experiments confirm the consistency of the selected values of the constants α, C_P, and C_r in estimating the values of molecular weights of proteins.

Protein charge ladders provide estimates of the molecular weights of proteins (1−12) that agree within 20% with literature values, except for monomeric ovalbumin and peroxidase (Figure 3B). We believe that the obvious deviation of ovalbumin is due to the formation of a dimer of the protein under the experimental conditions, as also indicated from the deviation of the effective charge of the protein (Figure 3A). This hypothesis is supported by dynamic light scattering experiments that demonstrate ovalbumin can undergo irreversible conformational changes and form both a dimer and further aggregation products in solution upon thermal perturbation[24] and by results from rheological and small-angle X-ray scattering experiments.[25,26] We speculate that the deviation for peroxidase may be due to the glycosylation of the protein: the extended oligosaccharide unit may cause the hydrodynamic drag to be greater than expected for a compact polypeptide sequence of the same molecular weight and cause the molecular weight to be overestimated.

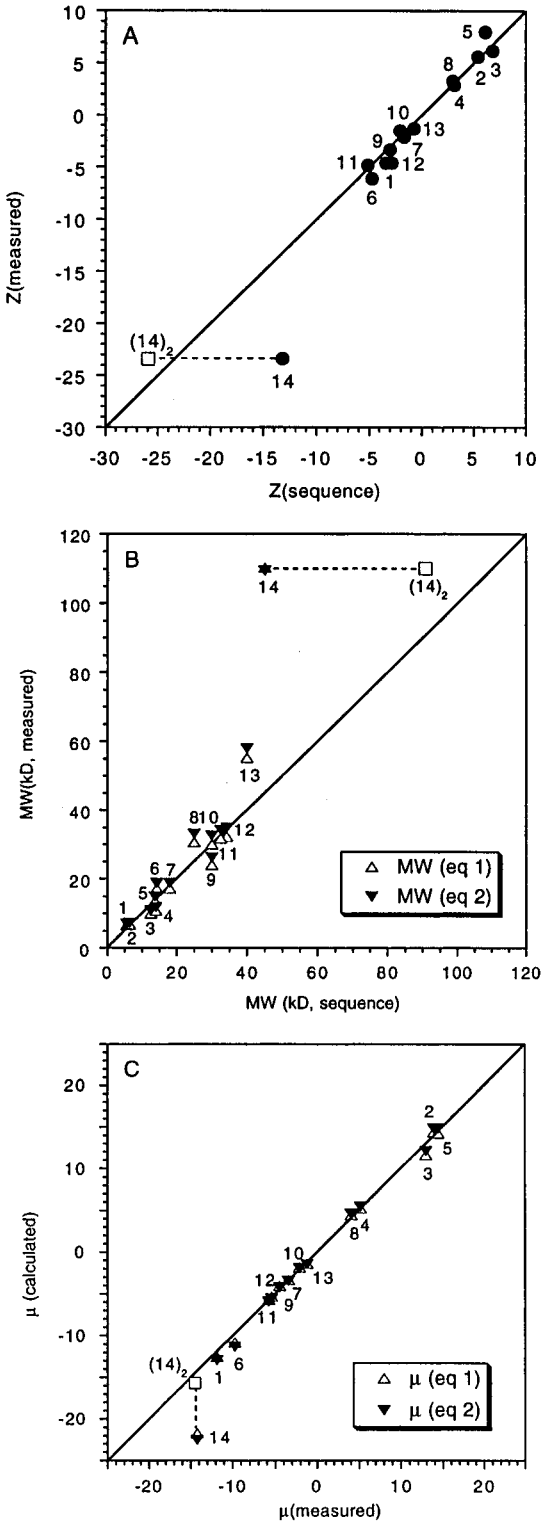

Figure 3. Comparison of values of effective charge (A) and molecular weight (B) of proteins from charge ladder experiments with those calculated based on amino acid sequences. (C) Comparison of electrophoretic mobility estimated using eq 1 or 2 with that determined from experiment. The open square in each figure indicates the theoretical values of $Z_P(0)$, molecular weight, and electrophoretic mobility for the dimer of ovalbumin. In all three figures, the scales for the x and y axes were the same, and a diagonal line was drawn to help the visual comparison.

(23) Creighton, T. E. *PROTEINS Structures and Molecular Properties*; W. H. Freeman and Co.: New York, 1993; p 266.

We believe that protein charge ladders will be useful in determining the molecular weights of proteins under nondenaturing conditions. For example, superoxide dismutase is composed of two identical subunits that are connected by a Cys−Cys disulfide bond. Under the normal conditions for SDS−PAGE, 2-mercaptoethanol reduces the disulfide bond. In the current study, the molecular weight of superoxide dismutase obtained from the charge ladder experiment is twice that of the monomeric subunit of the protein, and it suggests that the native protein comprises two subunits in solution. In combination with SDS−PAGE, the analysis of charge ladders will be useful in estimating the number of subunits and the stage of aggregation of a protein in solution.

Estimation of Electrophoretic Mobilities of Proteins Using Values of C_P, α, and C_r in Eq 1 and Eq 2, Respectively. Based on the measured values of C_P, α, and Cr, we have estimated values of the electrophoretic mobilities of our set of proteins. The estimated values agree well with the experimental values, with a standard deviation of 0.6 $cm^2\ kV^{-1}\ min^{-1}$ (that is, <10% of μ) (Figure 3C). The ability to predict the electrophoretic mobilities of proteins accurately is useful in optimizing conditions for protein separation in capillary electrophoresis and in identifying which peak in a complex mixture may be the one of interest.

CONCLUSIONS

We believe that the major uses of the protein charge ladder technique will be to estimate values of the effective charge *Z* (values that are not readily available by other techniques).[13−15] An added benefit of forming and analyzing charge ladders is to explore the extent of aggregation of proteins in solution and to provide correlated values of molecular weight and *Z*. This technique suffers from the limitations of CE applied to proteins: not all proteins can be observed by CE (since adsorption on the walls of the capillary can be a problem, especially for high molecular weight proteins), and it may be necessary to coat the walls of capillaries to observe positively charged proteins. Moreover, the procedure requires acetylation under alkaline conditions (even though the protein is exposed to these conditions for only a short period of time) to achieve significant modification of lysine ϵ-NH_3^+ groups. For high molecular weight proteins, acetylation may not be sufficient to generate a charge ladder with well-resolved members, and modification with acylating agents that introduce multiple negative charges may be required.

The analysis carried out here defines values of *C* and α that fit the test proteins well. These proteins cover a wide range of values of p*I* and molecular weight and are probably representative of soluble, globular proteins; whether they will apply to other classes of proteins (for example, proteins involved in the extracellular matrix, in biological rheology control, and in nonenzymatic and receptor functions and proteins that are heavily glycosylated) and biologically derived macromolecules remains to be tested experimentally; we believe that the correlations shown in Figure 2 are sufficiently good that large deviations from them will be interpretable structurally.

Both eq 1 and eq 2 can predict the electrophoretic behavior of proteins in solution semiquantitatively. Equation 1 is simple and easily applied. Analysis using eq 2 is less straightforward as a method to determine the molecular weights of proteins; it includes, however, the influence of ionic strength on the electrophoretic mobility, assuming Debye−Hückel theory to be obeyed. Further modification of eq 2 may give a more quantitative correlation between the structures of proteins and their electrophoretic behavior than that based on eq 1.[10] In combination with molecular modeling studies, analysis by either treatment may help to understand the influence of shapes, dipoles, and asymmetry of proteins on their electrophoretic mobilities in solution[10] and thus to clarify the role of electrostatics in biochemistry.

Although the primary usefulness of charge ladders is to obtain *Z*, estimation of molecular weight comes without significant additional effort. Using the value of *Z*, eq 1 or 2, and the values of C_P and α obtained here, we believe that this information will be useful in detecting aggregation, in characterizing noncovalent protein−protein and protein−ligand complexes, and perhaps more broadly in examining the diffusion and hydrodynamic drag of native proteins moving in an applied electrostatic field.

ACKNOWLEDGMENT

This work was supported by the NIH Grant GM 51559.

Received for review August 8, 1996. Accepted December 4, 1996.[⊗]

AC9608073

[⊗] Abstract published in *Advance ACS Abstracts,* January 15, 1997.

(24) Nemoto, N.; Koike, A.; Osaki, K.; Koseki, T.; Doi, E. *Biopolymers* **1993**, *33*, 551−559.

(25) Matsumoto, T.; Chiba, J. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 2877−2882.

(26) Results from ref 25 demonstrate that native ovalbumin forms a dimer at concentrations above 1 mg/mL. To measure the value of C_P at a concentration far below this value, we modified ovalbumin using the succinimidyl ester of 5-carboxyfluorescein to allow the use of laser-induced fluorescence detection and improved sensitivity. We obtained the same value of C_P at a concentration of 3 μg/mL as we did using acetylation with acetic anhydride, UV detection, and a concentration of 1 mg/mL ovalbumin. Since dioxane has been shown to suppress hydrophobic interactions between monomeric insulin in solution (see ref 14), we examined the value of C_P of ovalbumin as a function of concentration of dioxane in solution. The result demonstrates that added dixoane in the buffer did not change the state of aggregation of ovalbumin. These results suggest that the commercial sample of ovalbumin (obtained from Sigma) may already be irreversibly modified to exist primarily as a dimer in solution.