Determination of the effective charge of a protein in solution by capillary electrophoresis

(protein charge ladders)

JINMING GAO, FRANK A. GOMEZ, RALPH HÄRTER, AND GEORGE M. WHITESIDES

Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138

Contributed by George M. Whitesides, August 15, 1994

ABSTRACT   This paper describes two methods to estimate the effective charge of a protein in solution by capillary electrophoresis and demonstrates these methods by using representative proteins. In one method, a "charge ladder"—a series of derivatives of a protein differing by known increments of charge but differing only minimally in hydrodynamic drag—is generated by covalent modification of the ε-amino groups of lysines with 4-sulfophenyl isothiocyanate or acetic anhydride. In the second method, the equivalent of a charge ladder is produced by noncovalent association of a protein with differently charged ligands. Analysis of the electrophoretic mobilities of the protein and its derivatives as a function of added charge allows the effective charge to be estimated for the unmodified protein. This type of analysis permits estimation of the effective charge of a protein without knowing its composition, structure, or amino acid sequence.

Mass and charge are fundamental properties of proteins in solution. Determining mass is straightforward by SDS/polyacrylamide gel electrophoresis; determining charge—especially when the amino acid sequence is unknown—is presently difficult or impossible. Only the isoelectric point—the value of pH at which the charge is zero—is readily accessible experimentally.

The calculation of the net charge of proteins having known sequences of amino acids is straightforward if the pH of the surrounding medium and values of the ionization constants of each polar amino acid residue are known. In practice, the tertiary structure of the protein and the sequence composition may modify the expected values of ionization constants. Experimentally, there are only a few reports concerning the estimation of net charge (1–3). Tanford's method based on isoionic points (1) is straightforward and can be used to estimate the net charge of a protein over a range of values of pH. This method, however, requires substantial quantities (2–10 mg) of pure protein, and the measurement may be influenced by association of the protein with ions in solution. Winzor and Öjteg (2, 3) have estimated net charges of proteins based on Donnan potential measurements. These methods permit the charge of a protein to be determined over a range of ionic composition and values of pH but also require significant quantities (1–10 mg) of a protein.

Here we report two related procedures for estimating the effective charge of a protein in solution at arbitrary values of pH by using capillary electrophoresis (CE) (4–8). The two procedures are complementary, but both rely on the same strategy: to generate and compare the electrophoretic mobilities of a protein with a series of derivatives of the protein that differ in effective charge by simple multiples of a unit charge but that differ only minimally in hydrodynamic drag. Examination of this series of derivatives eliminates the

ambiguities that make it impossible to interpret electrophoretic mobilities of proteins directly in terms of charges. These two methods allow the measurement of the effective charges of many proteins by using small amounts (50–500 μg) of samples in the presence of impurities (so long as they can be separated from the proteins of interest by CE).

MATERIALS AND METHODS

Materials. All chemicals were analytical grade. 4-Sulfophenyl isothiocyanate (sodium salt), compound 1 (see Fig. 2), was purchased from Fluka. Acetic anhydride and dioxane were purchased from Mallinckrodt. Compounds 2–4 were synthesized as described (9–11).

Bovine carbonic anhydrase (BCA, EC 4.2.1.1, containing BCAA and BCAB isozymes, from bovine erythrocytes), human carbonic anhydrase (EC 4.2.1.1, from human erythrocytes), bovine pancreatic insulin (from bovine pancreas), porcine pancreatic trypsin (from porcine pancreas), and bovine α-lactalbumin (type 1, from bovine milk) were purchased from Sigma. Horse heart myoglobin was purchased from United States Biochemical. Mesityl oxide was purchased from Eastman Organic Chemical (Rochester, NY). Stock solutions of proteins were each prepared by dissolving the lyophilized protein in water.

Equipment. The CE system was an ISCO model 3140 system. The capillary tubing (Polymicro Technologies, Phoenix) was of uncoated fused silica with an internal diameter of 50 μm, a total length of 70 cm, and a length from inlet to detector of 45 cm. The conditions used in CE were as follows: voltage, 30 kV; current, 8–20 μA, depending on buffer conditions; detection, 200 nm; temperature, 27 ± 2°C.

General Acylation Procedures. An aqueous solution of a protein (0.1 mM, 0.5 ml) was adjusted to pH 12 by using 0.1 M NaOH, and 4-sulfophenyl isothiocyanate (sodium salt) 1 (10 mM in water, 25 μl) was added at room temperature. After 5 min, the solution was exchanged with electrophoresis buffer by centrifugation using a microdialysis tube at 4°C (Centricon-3 for insulin and Centricon-10 for other proteins). The residue was diluted with electrophoresis buffer (1 ml) and analyzed. Acylation using acetic anhydride (10 mM in dioxane) was conducted in similar fashion. At pH 12, acylation resulted in significant preference for Lys ε-amino groups over N-terminal α-amino groups (12).

Procedures for CE. The sample for injection into the electrophoresis capillary was prepared by diluting 50 μl of a sample of acylated protein with 50 μl of electrophoresis buffer. Mesityl oxide at a concentration of 20 μM was used as an electrically neutral marker. The sample solution (8 nl) was introduced into the capillary by vacuum injection. The charge ladders obtained from acylation of amino groups on proteins were observed directly by electrophoresis. The

Abbreviations: CE, capillary electrophoresis; BCA, bovine carbonic anhydrase.

charge ladders inferred from binding a series of charged ligands to BCA were analyzed using appropriate concentration of an arylsulfonamide ligand (0.5–1 mM, depending on the binding affinity of charged ligand) in the electrophoresis buffer to ensure saturation of the binding site (9–11).

RESULTS

Generation of Charge Ladders. The required series of charged derivatives were generated either by covalent modification of the protein or by noncovalent association of charged ligands with it (Eq. 1; in this equation, we assume a single active site). For simplicity, only ammonium and car-

[1]
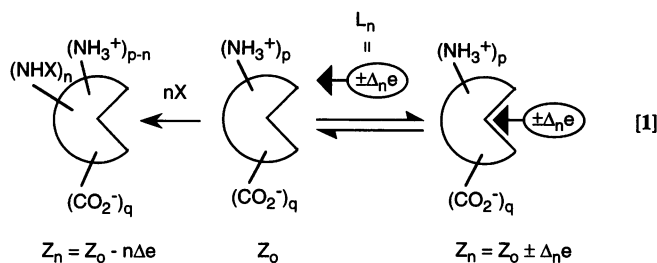
boxyl groups are shown in Eq. 1, but other charged or reactive groups are included in the analysis as follows: X is the derivatizing group (for example, acetyl), Δe is the change in charge on derivatization of a single functional group, and $\Delta_n e$ is the unit charge on the inhibitor L_n. Covalent derivatization using small reagents generates a series of modified proteins with different values of charge Z_n but with very similar values of hydrodynamic drag. These modified proteins appear in the electropherogram as a set of peaks with uniform spacing: that is, as a charge ladder. Analysis of the spacing of the ladder yields the effective charge Z_0 of the unmodified protein (Fig. 1).

We illustrate the formation of charge ladders using the reagents acetic anhydride and 1, which are selective for amino groups (Fig. 2), although reagents with selectivities for other functional groups are also useful in some circumstances. In a CE buffer with a pH between 6 and 8, Lys amino groups (pKa ≈ 10.3) are essentially fully protonated, so that
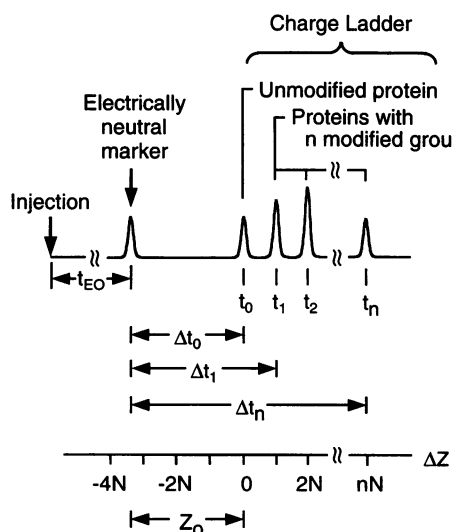

FIG. 1. Schematic illustration of capillary electropherogram (detector response vs. time) of a mixture containing an unmodified native protein with effective charge Z_0 and a ladder of derivatives differing from the unmodified protein by increments of charge $N = \Delta e$ or $\Delta_n e$ (Eq. 1). Z_0 can be read directly from a scale calibrated in units of N or evaluated using Eq. 5.
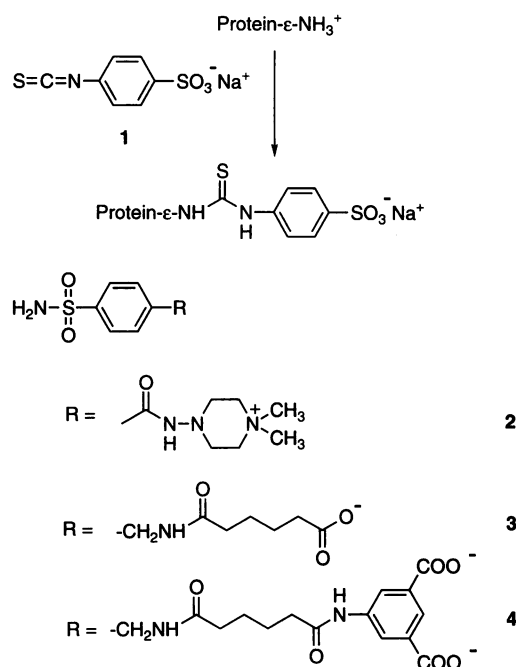

FIG. 2. Acylation of ε-NH$_3^+$ of proteins by reagent 1; structure of the charged inhibitors for BCA.

proteins with n modified amino groups differ in charge from the native protein by nN units ($N = -2$ for 1, $N = -1$ for Ac$_2$O). For the N-terminal ammonium groups with pKa values ranging from 7 to 9 (13), the situation is more complicated. In a buffer with pH near the pKa of an α-amino group, covalent modification will contribute a fractional charge to the protein. Although analysis of this effect is straightforward and provides useful information about the acidity and reactivity of the N-terminal amino group, in this paper we show only data obtained using conditions for the acylation that are selective for the ε-amino groups of Lys.

Equations for the Analysis. The basis of the analysis is a generic approximate expression for the electrophoretic mobility (μ, cm$^2 \cdot$kV$^{-1} \cdot$sec^{-1}) of a protein (Eq. 2) (14).

$$\mu = C_P \frac{Z}{M^\alpha}.$$

[2]

The mobility of a protein in an electrostatic field of a value defined by the applied voltage and the length of the capillary is proportional (with the proportionality constant C_P) to its effective charge Z (which determines the force on it in an electric field) and inversely related to its hydrodynamic drag (which resists movement under the influence of the field). The drag is often related to the mass of the protein through the function M^α (typically $M^{2/3}$ for globular proteins), but both the relationship between structure and drag and the value of the proportionality constant C_P depend on the details of shape, charge, and charge distribution and are not known *a priori*. This dependence makes it impossible, at present, to relate the electrophoretic mobility of a protein analytically to its charge, even if the amino acid sequence and other details of the structure have been determined.

If the charge on a protein is modified from Z_0 to Z_n in such a way that the hydrodynamic drag does not change significantly, the ratio of the mobilities of modified and unmodified protein is given by Eq. 3.

$$\mu_0/\mu_n = Z_0/Z_n.$$

[3]

This equation assumes that two unknown constants, C_P and α, do not change significantly on modification of the protein with small reagents and eliminates both these constants and M. Eq. 3 forms the basis for our method of determining effective charge.

The time of emergence t_n of the peaks in a charge ladder can be expressed by Eq. 4 in terms of the velocity of electroosmotic flow V_{EO}, the distance from injection to detector in the electrophoresis capillary d, and the electrophoretic mobility of the protein (μ_n) with n modified Lys groups. The time of emergence of an electrical neutral marker specie (here, mesityl oxide) is t_{EO} and establishes the magnitude of V_{EO} experimentally. Algebraic manipulation of Eq. 4 gives Eqs. 5–7 ($\Delta t_n = t_n - t_{EO}$).

$$(V_{EO} + \mu_n)t_n = d. \qquad [4]$$

$$\Delta Z = Z_n - Z_0 = Z_0\left[\frac{(\Delta t/t)_n}{(\Delta t/t)_0} - 1\right]. \qquad [5]$$

$$\frac{\Delta t_n}{Z_n} \approx \frac{\Delta t_0}{Z_0}. \qquad [6]$$

$$Z_0 \approx \frac{nN\Delta t_0}{t_n - t_0}. \qquad [7]$$

Eq. 5 indicates that a plot of the form shown in Fig. 3C should be linear with slope Z_0: that is, the slope should be equal to the charge of the unmodified protein. Eqs. 6 and 7 are useful for approximation: they indicate that as long as $t_n/t_0 \approx 1$, the peaks in the charge ladder should be spaced linearly, with spacing proportional to Z_n. This ladder establishes a scale of emergence times for a set of proteins whose differences in electrophoretic mobility is due, to a good approximation, only to differences in their charge; extrapolation of this scale (in units of $N = \Delta e$) between t_{EO} and t_0 gives the value of Z_0 directly.

Determination of Effective Charges of Proteins. We have tested two procedures—one based on covalent modification of Lys ε-amino groups on a protein and the second based on modifying charge by binding a series of ligands with different values of charge to a protein—using BCA as an example (Fig. 3). This laboratory has used (9–11) BCA previously as a model protein for studies in affinity CE; it has two isozymes, BCAA and BCAB. Treatment of BCA with reagent 1 con-

verts positively charged ammonium groups to negatively charged groups and, therefore, changes Z in units of $N = -2$; similarly, acylation with acetic anhydride changes Z in units of $N = -1$.

After modification with 1, analysis of the charge ladders for BCAB and BCAA yields estimates of the effective charges at pH 8.3 directly (Fig. 3C; $Z_{BCAB} \approx -3.5$ and $Z_{BCAA} \approx -4.5$). Noncovalent modification of charge was accomplished by obtaining the emergence times of complexes of BCA with the arylsulfonamides 2–4. Analysis of the data using Eq. 5 indicates that $Z_0 \approx -3.7$ for BCAB and $Z_0 \approx -5.0$ for BCAA. The disagreement between two independent measurements is 0.5 pKa units and suggests an uncertainty to these methods of approximately this value. BCAB and BCAA differ by the replacement of Arg by Gln (Arg-56 in BCAB and Gln-56 in BCAA) (15); the difference of one unit of charge between them observed experimentally is that expected from the amino acid sequences.

DISCUSSION

Quantity of a Protein Needed for the Analysis. The quantity of a protein required in these analysis is dependent on the detection sensitivity of the CE and convenient volumes of sample reservoirs for injection. A good signal-to-noise ratio is obtained with a concentration of 0.5 mg/ml of a protein with an injected volume of 8 nl by using the commercial UV detection of our instrument. We typically prepare sample volumes of 50–100 μl at a protein concentration of 0.5 mg/ml; with these samples, less than 50 μg of the protein is required for the analysis. The amount of protein can be further reduced by employing more sensitive detection methods (i.e., fluorescence) or by working with smaller sample volumes, although manipulating very small samples can be experimentally difficult.

Estimation of the Net Charge of a Protein from Its Amino Acid Sequence. By using the amino acid sequences of proteins and corrected ionization constants for polar amino acids (16, 17), we estimated the net charges of the proteins examined here. For example, the net charge of BCAB (18) at pH 8.3 was calculated using the following values of pKa for charged functional groups (pK$_a$ of a group, fractional charge per group at pH 8.3): 18 Lys (10.3, +0.99), 9 Arg (12.5, +1), 11 His (6.2, +0.01), 16 Asp (3.5, -1), 10 Glu (4.5, -1), 7 Tyr (10.3, -0.01), N-terminal Ser (7.3, +0.1), C-terminal Lys (3.2, -1), 3 Asx (Asp or Asn, charge either -1 or 0), 1 Glx (Glu or Gln, charge either -1 or 0), Zn^{2+} (+2), and (Zn^{2+})-H$_2$O (7.0,

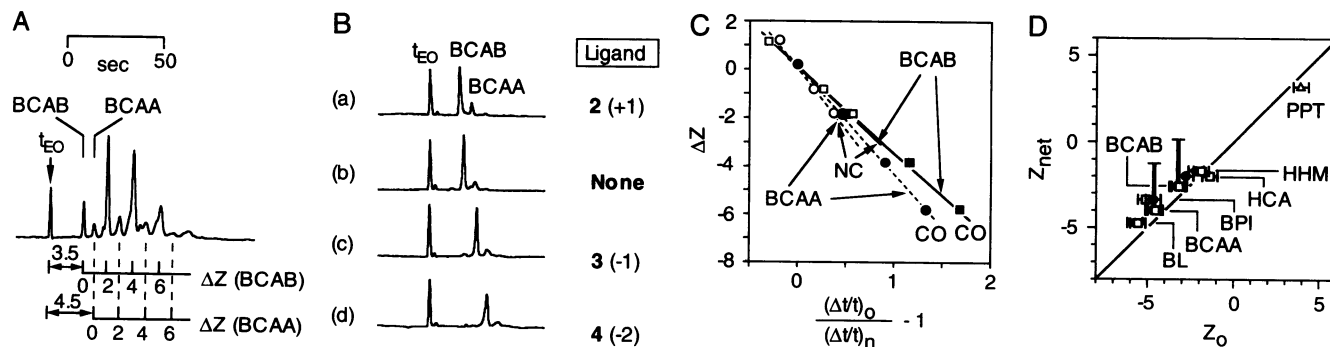

Fig. 3.   Measurement of effective charge of BCA. The sample contained two isozymes BCAB and BCAA in an $\approx$10:1 ratio; these isozymes differ by one unit of charge. (*A*) Covalent modification of Lys ε-amino groups on BCA by 1 generated two charge ladders of BCAA and BCAB (buffer = 25 mM Tris·HCl/200 mM Gly, pH 8.3; $t_{EO} \approx$ 170 sec). The emergence times of the unmodified proteins are indicated and provide origins of scales in ΔZ. (*B*) Ladder generated by complexation of BCA with charged inhibitors (Fig. 2). Traces: a, [2] = 1.0 mM; b, buffer; c, [3] = 1.0 mM; d, [4] = 0.5 mM. (*C*) Determination of effective charge from the analysis of charge ladders by using Eq. 5. The data shown for BCAB and BCAA were estimated using ladders generated by covalent (CO) and noncovalent (NC) methods. (*D*) Comparison of effective charges (Z_0) of various proteins determined experimentally (by covalent modification) and net charges (Z_{net}) estimated from their amino acid sequences. PPT, porcine pancreatic trypsin; HHM, horse heart myoglobin; HCA, human carbonic anhydrase; BPI, bovine pancreatic insulin; BL, bovine α-lactalbumin. Buffers: ●, 25 mM Tris·HCl, pH 8.3; □, 25 mM Tris·HCl/200 mM Gly, pH 8.3; △, 20 mM sodium phosphate/2 M betaine, pH 7.5.

−0.95). The net charge of BCAB is calculated to be between −3.0 and 1.0; this uncertainty is due to the ambiguity in the amino acid sequence (Asx and Glx) of BCAB. Based on the results obtained here, we believe that this ambiguity should be resolved in favor of Asp and Glu and that the effective charge of BCAB at pH 8.3 is −3.5 (and that of BCAA is −4.5). The imperfect agreement between the net charge calculated from the sequence and the effective charge that estimated experimentally can be rationalized by errors in the experimental method or uncertainties in the values of pKa used (especially for the Zn-OH$_2$ group and the α-amino group of the N-terminal Ser).

Net Charge vs. Effective Charge. The net charge of a protein (Z_{net}) is the summation of charges carried by its electrostatic components, including charged amino acid residues, metal ions, and charged cofactors. The effective charge of a protein moving in an electrolyte solution, however, is less than Z_{net}. This difference is due to the screening of the charge by association of the protein with counterions in solution. The degree of counterion shielding depends on the nature and concentration of the counterions and on the characteristics of the protein. A general description of this phenomenon is complicated and will not be addressed here (14).

Fig. 3*D* compares effective charges of other proteins calculated from ladders of covalently modified proteins and net charges estimated from their amino acid sequences. For this set of proteins, the experimental values of effective charge agree with values of net charge calculated from sequence data to within 0.5–1 unit.

In conclusion, the procedures outlined here provide a method of estimating the effective charge of a protein in the absence of its sequence or structural information. The key to the method is the generation of a charge ladder—a series of derivatives of a protein of interest that differ by known integral quantities of charge but differ minimally in hydrodynamic drag. Analysis of the charge ladder eliminates ambiguities in relating effective charge to electrophoretic mobility and makes it possible to estimate charge directly from the emergence times of the members of a charge ladder. There are obvious cautions and limitations associated with the method: modification of the protein or association of the protein with a ligand must not change the charge in unexpected ways; nonspecific association of a protein with other species in solution may influence its effective charge and cause effective charge and net charge to differ; impurities must not confuse interpretation of the charge ladder; modification of the ε-amino groups of Lys must produce changes in charge of whole units. We suggest that these conditions will often be met and believe that these procedures will be useful in characterizing proteins, in measuring their charge as a function of pH, in following modifications that change charge (for example, hydrolysis of Asn to Asp and of Gln to Glu, sialylation or desialylation, phosphorylation, or alkylation of Cys thiol with charged reagents such as iodoacetate), in examining interaction of proteins with charged ligands and substrates, in detecting isozymes, and in studying the interactions of proteins with metallic and organic ions and with charged surfactants.

1. Nozaki, Y. & Tanford, C. (1967) *Methods Enzymol.* **11**, 715–734.
2. Ford, C. L. & Winzor, D. J. (1982) *Biochim. Biophys. Acta* **703**, 109–112.
3. Öjteg, G., Lundahl, P. & Wolgast, M. (1989) *Biochim. Biophys. Acta* **991**, 317–323.
4. Engelhardt, H., Beck, W., Kohr, J. & Schmitt, T. (1993) *Angew. Chem. Int. Ed. Engl.* **32**, 629–766.
5. Kuhr, W. G. & Monnig, C. A. (1992) *Anal. Chem.* **64**, 389–407.
6. Grossman, P. D., Colburn, J. C., Lauer, H. H., Nielsen, R. G., Riggin, R. M., Sittampalam, G. S. & Rickard, E. C. (1989) *Anal. Chem.* **61**, 1186–1194.
7. Novotny, M. V., Cobb, K. A. & Liu, J. (1990) *Electrophoresis* **11**, 735–749.
8. Grossman, P. D. (1992) in *Capillary Electrophoresis: Theory and Practice*, eds. Grossman, P. D. & Colburn, J. C. (Academic, San Diego).
9. Chu, Y.-H., Avila, L. Z., Biebuyck, H. A. & Whitesides, G. M. (1992) *J. Med. Chem.* **35**, 2915–2917.
10. Avila, L. Z., Chu, Y.-H., Blossey, E. C. & Whitesides, G. M. (1993) *J. Med. Chem.* **36**, 126–133.
11. Gomez, F. A., Avila, L. Z., Chu, Y.-H. & Whitesides, G. M. (1994) *Anal. Chem.* **66**, 1785–1791.
12. Stark, G. R. (1970) *Adv. Protein Chem.* **24**, 261–308.
13. Tanford, C. (1962) *Adv. Protein Chem.* **17**, 69–165.
14. Grossman, P. D. (1992) in *Capillary Electrophoresis: Theory and Practice*, eds. Grossman, P. D. & Colburn, J. C. (Academic, San Diego), pp. 111–132.
15. Gulian, J. M., Limozin, N., Charrel, M., Laurent, G. & Derrien, Y. (1974) *C.R. Acad. Sci. Ser. D* **278**, 1123–1128.
16. Rickard, E. C., Strohl, M. M. & Nielsen, R. G. (1991) *Anal. Chem.* **197**, 197–207.
17. Bundi, A. & Wüthrich, K. (1979) *Biopolymers* **18**, 285–297.
18. Hewett-Emmett, D. & Tashian, R. E. (1991) in *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*, eds. Dodgson, S. J., Tashian, R. E., Gros, G. & Carter, N. D. (Plenum, New York), Chap. 2.