VOLUME 119, NUMBER 52
DECEMBER 31, 1997

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Formation of Protein Charge Ladders by Acylation of Amino Groups on Proteins

Ian J. Colton, Janelle R. Anderson, Jinming Gao, Robert G. Chapman, Lyle Isaacs, and George M. Whitesides*

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

Received July 14, 1997[⊗]

Abstract: The values of charge and electrophoretic mobility of a protein are changed upon acylation of its α- and Lys ϵ-NH_3^+ groups. Partial acylation of the amino groups of a protein results in a set of derivatives that is often resolved by capillary electrophoresis into a set of distinct peaks—the "rungs" of a protein charge ladder—that differ incrementally in the number of residues modified. Proteins that have values of MW < 50 kD usually form resolved charge ladders when allowed to react with acetic anhydride, while proteins that have values of MW > 50 kD form broad unresolved peaks. Resolved charge ladders of proteins that have values of MW > 50 kD may be formed using acylating agents that introduce several charges upon acylation of each of their Lys ϵ-NH_3^+ groups. As an example, L-lactate dehydrogenase (MW = 147 kD) does not form a resolved charge ladder when modified with acetic anhydride. When it is acylated with either 1,2,4-benzenetricarboxylic anhydride, **3**, or 1,2,4,5-benzenetetracarboxylic dianhydride, **4**, however, it forms charge ladders in which each of the first several pairs of adjacent rungs are separated by approximately 3 or 4 units of charge, respectively. The procedures described in this paper were used to form resolved charge ladders from 25 proteins differing in pI and in MW.

Introduction

This paper evaluates the generality with which protein charge ladders can be formed by acylating the amino groups on proteins, and resolved by capillary electrophoresis (CE). A protein charge ladder is a set of derivatives of a protein that is separable by CE into a set of distinct peaks—the "rungs" of the ladder—that differ from one another in their values of electrophoretic mobility.[1] Each rung of the charge ladder is composed of a mixture of different regioisomeric derivatives of native protein that have the same number of modifications and usually the same value of charge.[2] Protein charge ladders are useful in determining the values of charge of proteins,[3] in obtaining values of pK_a of the N-terminal α-NH_3^+ group of proteins,[4] in evaluating the influence of charge on the free energy of binding of differently charged proteins to a charged ligand,[5] and in estimating the molecular weights of proteins under non-denaturing conditions.[6]

It is convenient to modify the charge of proteins by acylating the ϵ-NH_2 groups of Lys, since Lys is abundant in proteins (its frequency of occurrence is 5.9%, based on 1021 proteins of known sequence), reactive, and mostly distributed on the surface of water-soluble proteins.[7] Modification of the guanidino residues of Arg using α-diketones[8,9] or amidation of the carboxyl residues of Glu and Asp using a dehydrating agent and amines[10] are also useful ways of modifying the surface of a protein. The role of positively charged ϵ-NH_3^+ groups of Lys in stabilizing the structure of proteins is not well understood, although Lys groups are slightly preferred at the C-terminus of an α-helix,[7]

*Corresponding author. Telephone number: (617)-495-9430. Fax number: (617)-495-9857. E-mail address: gwhitesides@gmwgroup.harvard.edu.

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

(1) For an early example of a charge ladder, see: Creighton, T. E. *Nature* **1980**, *284*, 487−489.

S0002-7863(97)02349-4 CCC: $14.00

perhaps to balance helix dipoles.[11] An important function of Lys in many proteins may be to make them soluble. Modification of ϵ-NH_3^+ groups of Lys on the surface of the protein, however, should affect the stability of the protein significantly less than modification of residues buried inside the protein. We have recently demonstrated that each rung of the charge ladder derived from bovine carbonic anhydrase II (BCA II) by acetylation of Lys ϵ-NH_3^+ groups binds neutral sulfonamide ligands with equal affinity.[5] This observation suggests that in this protein, Lys ϵ-NH_3^+ groups are not important in determining the affinity of the protein for neutral ligands, and by inference that charge on these groups does not determine the conformation around the active site. We do not know for which other proteins this conclusion will hold.

Capillary electrophoresis is an analytical technique that provides information simultaneously about the values of charge and coefficient of friction of proteins in solution. The electrophoretic mobility of a protein, as measured by CE, $\mu_{electro}$, is proportional to its charge at a given value of pH, Z_{CE}, and inversely correlated to M^{α}/C_P, its coefficient of friction (eq 1),[12−16] where M is the MW of the protein, C_P is a proportionality constant, and α is a constant that relates to the shape of the protein under the conditions of its analysis.

$$\mu_{electro} = C_P\left(\frac{Z_{CE}}{M^{\alpha}}\right) \tag{1}$$

We define Z_{seq} as the charge of a protein calculated from its amino acid sequence and any charged cofactors (bound metal ions, prosthetic groups, coenzymes), and post-translationally added residues (charged sugars, phosphate groups), using standard values of pK_a and the pH of the solution. Uncertainties in the values of pK_a (especially for His and for ionizable residues with anomalous values of pK_a) are sufficiently large that calculations of Z_{seq} are inexact. The value of Z_{CE} reflects any differences between the actual and standard values of pK_a of the charged residues of the protein and includes the contributions of charge from tightly associated counterions.

Electrophoretic mobilities of analytes are expressed mathematically as the difference in velocities of an analyte peak and a neutral marker peak per unit of electric field strength (eq 2). In eq 2, L_{tot} (m) is the total length of the capillary, L_{det} (m) is the length from the inlet of the capillary to the detector and V

$$\mu_{electro} = \frac{\left[\left(\frac{L_{det}}{t_x}\right) - \left(\frac{L_{det}}{t_{nm}}\right)\right]}{\left(\frac{V}{L_{tot}}\right)} \tag{2}$$

(V) is the voltage applied across the capillary. The parameters t_x (s) and t_{nm} (s) are the times of migration for an analyte peak x and a neutral marker, respectively. Equating the two expressions for electrophoretic mobility (eqs 1 and 2) and solving for t_x (eq 3) and $1/t_x$ (eq 4), respectively, demonstrates that plots in the 1/time domain are linearly related to Z_{CE}/M^{α}

$$t_x = \frac{t_{nm}}{1 + \frac{t_{nm}VC_P}{L_{det}L_{tot}}\left(\frac{Z_{CE}}{M^{\alpha}}\right)} \tag{3}$$

$$\frac{1}{t_x} = \frac{1}{t_{nm}} + \frac{VC_P}{L_{det}L_{tot}}\left(\frac{Z_{CE}}{M^{\alpha}}\right) \tag{4}$$

and consequently to mobility (eq 1), while those in the time domain are not.[17] We plot all primary data as absorbance vs -1/time (s^{-1}) to show trends in the mobility of rungs of protein charge ladders accurately.

Experimental Section

Materials. Acylase I (porcine kidney), aldolase (type X: rabbit muscle), alkaline phosphatase (type VII-NL: bovine intestinal mucosa), anti-DNP IgE (mouse), carbonic anhydrase II (bovine), carbonic anhydrase II (human), calmodulin (bovine brain), α-chymotrypsinogen (type II: bovine pancreas), cytochrome *c* (horse heart), ferritin (type I: horse spleen), *β*-galactosidase (*E. coli*), glucose-6-phosphate dehydrogenase (type XII: torula yeast), growth hormone (human), hemoglobin (bovine), insulin (bovine), α-lactalbumin (bovine milk), L-lactate dehydrogenase (type XI: rabbit muscle), myoglobin (horse heart), ovalbumin (chicken egg), pyruvate kinase (type III: rabbit muscle), superoxide dismutase (bovine erythrocytes), ubiquitin (bovine erythrocytes), sodium benzoate, and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDAC) were purchased from Sigma (St. Louis, MO). Bovine pancreatic trypsin inhibitor, carboxypeptidase B (porcine pancreas), creatine kinase (rabbit muscle), deoxyribonuclease I (bovine pancreas), dextranase (*Penicillium sp.*), hexokinase (yeast), lysozyme (egg white), pancreatic lipase (porcine), papain (papaya latex), peroxidase (horseradish), phospholipase A2 (*crotalus adamanteus* venom), and ribonuclease A (bovine pancreas) were purchased from Worthington (Freehold, NJ). Hexadimethrine bromide (Polybrene), *N*-hydroxy succinimide (NHS), 1,2,4-benzenetricarboxylic anhydride, 1,2,4,5-benzenetetracarboxylic dianhydride, and mellitic acid were purchased from Aldrich (Milwaukee, WI). Succinic anhydride was purchased from Eastman (Rochester, NY). Uncoated fused silica capillaries with an internal diameter of 50 μm were purchased from Polymicro Technologies (Phoenix, AZ). Cholic acid was purchased from Lancaster (Windham, NH). *N*,*N*-Dimethyl formamide (DMF) was purchased from EM Science (Gibbstown, NJ). *N*-Hydroxysuccinimidyl cholate was prepared according to the procedure of Okahata *et al.*[18] The zwitterion 3-quinuclidinopropanesulfonate was synthesized as previously described.[19] NICK Spin columns containing G-50 Sephadex gel were purchased from Pharmacia Biotech (Piscataway, NJ). The ^{1}H-NMR spectra were recorded at 400 MHz on a Bruker spectrometer.

(2) We name each rung of a charge ladder by the degree of modification of the proteins which it comprises, starting with native protein as the zeroth rung. The *n*th rung of the charge ladder of a protein having N Lys ϵ-NH_2 groups may be composed of as many as $(N!/(N-n)!n!)$ regioisomeric derivatives. As an example, the ninth rung of the charge ladder of bovine carbonic anhydrase II, an enzyme that has 18 modifiable Lys ϵ-NH_2 groups, may comprise up to 48620 regioisomeric derivatives. The assumptions made for these calculations are that all Lys ϵ-NH_2 groups have the same reactivity and that reaction at one site does not influence the reactivity at another; both of these assumptions are doubtless incorrect at some presently undefined level of detail.

(3) Gao, J.; Gomez, F. A.; Haerter, R.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12027−12030.
(4) Gao, J.; Mrksich, M.; Gomez, F. A.; Whitesides, G. M. *Anal. Chem.* **1995**, *67*, 3093−3100.
(5) Gao, J.; Mammen, M.; Whitesides, G. M. *Science* **1996**, *272*, 535−537.
(6) Gao, J.; Whitesides, G. M. *Anal. Chem.* **1997**, *69*, 575−580.
(7) Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman and Company: New York, 1993; p 6.
(8) Takahashi, K. J. *J. Biol. Chem.* **1968**, *243*, 6171−6179.
(9) Roberts, C.; Córdova, E.; Whitesides, G. M. Unpublished results.
(10) Anderson, J. R.; Colton, I. J.; Whitesides, G. M. Unpublished results.
(11) Hol, W. G. J. *Prog. Biophys. Mol. Biol.* **1985**, *45*, 149−195.
(12) Grossman, P. D. *Capillary Electrophoresis: Theory & Practice*; Academic Press: San Diego, CA, 1992; pp 111−132.
(13) Compton, B. J.; O'Grady, E. A. *Anal. Chem.* **1991**, *63*, 2597−2602.
(14) Engelhardt, H.; Beck, W.; Kohr, J.; Schmitt, T. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 629−766.
(15) Novotny, M. V.; Cobb, K. A.; Liu, J. *Electrophoresis* **1990**, *11*, 735−749.
(16) Karger, B. L.; Cohen, A. S.; Guttman, A. J. *J. Chromatogr.* **1989**, *492*, 585−614.
(17) Mammen, M.; Colton, I. J.; Carbeck, J.; Bradley, R.; Whitesides, G. M. *Anal. Chem.* **1997**, *69*, 2165−2170.
(18) Okahata, Y.; Ando, R.; Kunitake, T. *Bull. Chem. Soc. Jpn.* **1979**, *52*, 3647−3653.
(19) Gomez, F. A.; Avila, L. Z.; Chu, Y.; Whitesides, G. M. *Anal. Chem.* **1994**, *66*, 1785−1791.

Chemical shifts are reported in parts per million downfield of tetramethylsilane.

Acylation of Amino Groups of Proteins. Proteins of MW < 50 kD were dissolved in water at a concentration of ~0.1 mM, and 10 vol % of 0.1 N NaOH was added to each solution to bring the pH to ~12. Proteins having values of MW > 50 kD were dissolved in 100 mM sodium phosphate, pH 6.8 at a concentration of ~0.1 mM. Five to 20 equiv of acetic anhydride dissolved in dioxane was added to each protein solution, and the reactants were quickly mixed by vortexing. Reactions were usually complete within 1 minute. The sample was diluted in electrophoresis buffer (25 mM Tris-192 mM Gly, pH 8.4) prior to analysis. Reactions using reagents **3**−**5** were carried out in a similar manner except that the reagents were dissolved initially in DMF, and the protein charge ladders were purified on NICK spin columns spun at 2000 rpm for 4 min, prior to analysis by capillary electrophoresis.

Acetylation of Insulin. To a suspension of insulin (100 μL, 1.5 mg/mL in distilled H_2O) was added acetic anhydride (5 μL, 100 mM). After 20 min, the reaction mixture was adjusted to pH 12 with NaOH (20 μL, 0.1 N) and allowed to react with acetic anhydride (9 μL, 10 mM). The sample was diluted in electrophoresis buffer (25 mM Tris-192 mM Gly, pH 7.4 or 8.4) prior to analysis.

N-Hydroxysuccinimidyl Benzoate. To a mixture of sodium benzoate (300 mg, 2.1 mmol) and NHS (239 mg, 2.1 mmol) in 15 mL of DMF at 0 °C was added EDAC (480 mg, 2.5 mmol). The mixture was allowed to reach ambient temperature and was stirred overnight. The solvent was removed *in vacuo*. The residue was extracted with H_2O/CH_2Cl_2 (3 × 40 mL); the organic layers were combined and dried with $MgSO_4$. The solvent was removed *in vacuo*, and the white solid was purified by column chromatography using 19:1 CH_2Cl_2/MeOH yielding 320 mg of a white solid. This product was recrystallized with CH_2Cl_2/hexanes to give 290 mg; 64% yield of *N*-hydroxysuccinimidyl benzoate as a white solid. ^{1}H-NMR ($CDCl_3$, 400 MHz) δ 2.92 (s, 4H), 7.52 (m, 2H), 7.67 (m, 1H), 8.14 (m, 2H). HRMS (FAB^+) calcd for $C_{11}H_{13}N_2O_4$ $(M + NH_4^+)^+$ 237.0875, found 237.0878.

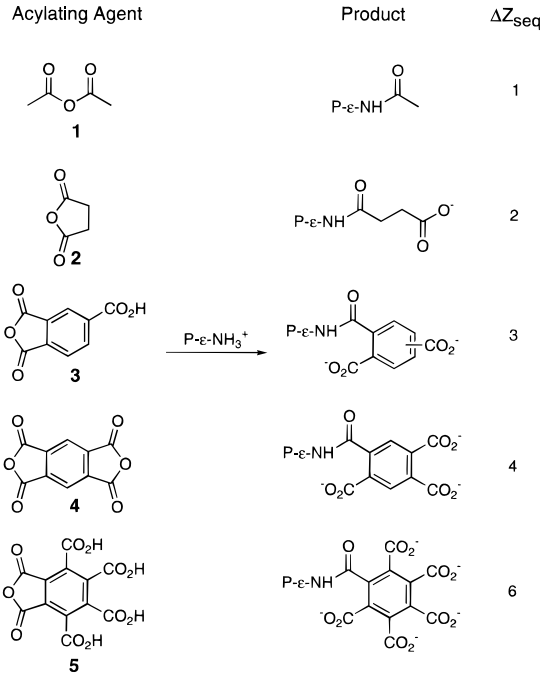
Benzenehexacarboxylic Acid Monoanhydride (5). Mellitic acid (57 mg, 0.17 mmol) was dissolved in 5 mL of DMF with heating. To this solution was added EDAC (58 mg, 0.3 mmol). The color of the solution turned dark violet within a 5-min period. The reaction mixture was stirred for an additional hour at ambient temperature. The crude mixture containing the monoanhydride of mellitic acid was used in acylation reactions directly. HRMS (FAB^-) calcd for $C_{12}H_4O_{11}$ (M^-) 323.9754, found 323.9744.

Capillary Electrophoresis (CE). CE experiments were conducted on a Beckman P/ACE 5500. Reaction products derived from modification of proteins of pI < 8.4 were analyzed at 25 or 37 °C on an uncoated capillary of fused silica (L_{tot} = 47 cm, L_{det} = 40 cm, 50 μm internal diameter) using 25 mM Tris-192 mM Gly buffer, pH 8.4 and an applied voltage of 15 or 30 kV; reaction products derived from modification of anti-DNP IgE were analyzed at 25 °C using 500 mM 3-quinuclidinopropanesulfonate, 10 mM K_2SO_4, 25 mM Tris-192 mM Gly buffer, pH 8.4 and an applied voltage of 15 kV. The acetylated derivatives of insulin were resolved at 25 °C on an uncoated capillary of fused silica (L_{tot} = 77 cm, L_{det} = 70 cm, 50 μm internal diameter) using 25 mM Tris-192 mM Gly buffer, pH 7.4 and an applied voltage of 30 kV. Reaction products from modification of proteins of pI > 8.4 were analyzed at 25 or 37 °C in reverse polarity on a Polybrene-coated capillary (L_{tot} = 47 cm, L_{det} = 40 cm, 50 μm internal diameter), using 25 mM Tris-192 mM Gly buffer, pH 8.4 and an applied voltage of 30 kV.[20]

Results and Discussion

Formation of Protein Charge Ladders of BCA II Using Acylating Agents That Introduce Different Values of Charge. Acylation of BCA II with acetic anhydride, **1**, converts its positively charged Lys ϵ-NH_3^+ groups to neutral Lys ϵ-NHAc groups (Scheme 1).[21] The absolute change in calculated charge, ΔZ_{seq}, due to acetylation of each of these Lys ϵ-NH_3^+ groups at pH 8.4 is equal to one unit of charge, based on a standard pK_a of 10.7 for each amino group. We observe from the charge ladder of BCA II, analyzed at pH 8.4, that each of the first four acetylations changes the $\mu_{electro}$ of the protein by an approximately uniform value (Figure 1). The absolute difference in values of $\mu_{electro}$ of adjacent rungs, $\Delta\mu_{electro}$, then decreases for successive pairs beyond the first four. From these results, we assume that the absolute difference in values of Z_{CE} of adjacent rungs, ΔZ_{CE}, for the first four pairs, is approximately equal to one unit of charge, i.e., $\Delta Z_{CE} \cong \Delta Z_{seq}$. We can therefore use the approximately equivalent values of ΔZ_{CE} for each of the first four pairs of adjacent rungs of the acetamide charge ladder to construct a "charge ruler" that has increments of ΔZ_{seq} of one unit of charge.

Scheme 1. Structure of Differently Charged Acylating Agents and Their Products upon Modification of the Lys ϵ-NH_2 Groups on a Protein (P-ϵ-NH_2)



The value of ΔZ_{seq} for acylation of a Lys ϵ-NH_3^+ group on BCA II is greater when acylating agents are used that impart greater values of negative charge to the protein per modification.[22] We observe, however, that the number of successive pairs of adjacent rungs of charge ladders of BCA II which have approximately equal values of $\Delta\mu_{electro}$ decreases as the value of ΔZ_{seq} increases with reagents **1**−**5**. For example, modification of BCA II with succinic anhydride, **2**, results in a charge ladder in which each of the first two pairs of adjacent rungs are separated by a value of ΔZ_{CE} of approximately two units of charge, on comparison to the charge ruler derived from the acetamide charge ladder. Similar comparisons show that only the first rung of charge ladders of BCA II formed by acylating with 1,2,4-benzenetricarboxylic anhydride, **3**, or 1,2,4,5-benzenetetracarboxylic dianhydride, **4**, is separated from native protein by a value of ΔZ_{CE} of approximately 3 or 4 units of

(20) Córdova, E.; Gao, J.; Whitesides, G. M. *Anal. Chem.* **1997**, *69*, 1370−1379.

(21) The α-NH_2 group of BCA II is naturally acetylated.

(22) The calculated values of ΔZ_{seq} for acylations carried out with reagents **2**−**5** and analyzed using pH 8.4 running buffer are based on the values of pK_a of the corresponding carboxylic acids: succinic (4.19; 5.48), trimellitic (2.52; 3.84; 5.20), pyromellitic (1.92; 2.87; 4.49; 5.63), mellitic (1.40; 2.19; 3.31; 4.78; 5.89; 6.96). These values of pK_a were taken from Jencks, W. P.; Regenstein, J. In *Handbook of Biochemistry*, 2nd ed.; Sober, H. A., Ed; CRC Press: Cleveland, OH, 1970; pp J-194−J-194.

Table 1. Panel of Proteins Surveyed To Examine the Generality of Formation of Charge Ladders

ID	protein	EC	subunits[a]	MW (kD)[a]	pI[a]	capillary charge[b]	CL[c]	reagent[d]	Z^o_{CE} [e]	Z^o_{seq} [a]
1	insulin		1	5.7	5.3	−	yes	**1**	−4.6	−3.5
2	bovine pancreatic trypsin inhibitor		1	7.4	10.4	+	yes	**1**	5.5	4.0
3	ubiquitin		1	8.6	7.0	−	yes	**1**	−0.3	−1.0
4	cytochrome *c*		1	12.4	10.0	+	yes	**1**	7.0	4.9
5	ribonuclease A	3.1.27.5	1	13.7	9.8	+	yes	**1**	2.8	3.0
6	α-lactalbumin		1	14.2	4.8	−	ues	**1**	−5.9	−8.0
7	lysozyme	3.2.1.17	1	14.3	10.9	+	yes	**1**	7.4	7.0
8	calmodulin		1	16.8	4.0	−	yes	**1**	−44	−25
9	myoglobin		1	17.6	6.8	−	yes	**1**	−2.2	−3.9
10	human growth hormone		1	22.2	5.3	−	yes	**1**	−7.0	−6.0
11	papain	3.4.22.2	1	23.5	9.6	+	yes	**1**	5.0	4.7
12	α-chymotrypsinogen	3.4.21.1	1	25.7	9.8	+	yes	**1**	4.1	3.0
13	phospholipase A2	3.1.1.4	2	27.6	5.1	−	yes, pr	**1**	−5.3	−10.1
14	carbonic anhydrase II (bovine)[f]	4.2.1.1	1	29.3	5.9	−	yes	**1, 2, 3, 4, 5**	−3.3	−2.0
15	carbonic anhydrase II (human)	4.2.1.1	1	29.4	7.6	−	yes	**1**	−2.3	−1.2
16	deoxyribonuclease I	3.1.21.1	1	30.5	5.1	−	yes, pr	**1, 3**	−13	−10
17	carbonic anhydrase V (murine)[g]	4.2.1.1	1	30.9	7.1	−	yes	**1**	−4.3	−4.6
18	superoxide dismutase	1.15.1.1	2	31.5	6.6	−	yes	**1**	−4.8	−6.0
19	carboxypeptidase B[h]	3.4.17.2	1	34.3	5.9	−	yes	**1**	−4.7	
20	ovalbumin		1	43.3	5.1	−	yes	**1**	−20	−17
21	dextranase[i]	3.2.1.11		44.0	4.2	−	yes	**1**	−14	
22	peroxidase[j]	1.11.1.7	1	44.0	7.2	−	yes	**1**	0.9	
23	creatine kinase[k]	2.7.3.2	2	86.6	7.1	−	yes	*1, 2*, **3**, **4**, **5**	−7.4	−8.9
24	acylase I	3.5.1.14	2	91.0	5.8	−	no	*1, 4, 5*		
25	alkaline phosphatase	3.1.3.1	2	106	6.0	−	no	*1, 4, 5*		
26	glucose-6-phosphate dehydrogenase[k,l]	1.1.1.49	2	115	6.4	−	yes	*1*, **3**, **4**	−21	−14
27	L-lactate dehydrogenase[k]	1.1.2.3	4	147	8.6	−	yes	*1, 2*, **3**, **4**, **5**	1.5	1.7
28	anti-DNP IgE			150		−	no	*5*		
29	aldolase	4.1.2.13	4	158	8.7	−	no	*3, 4, 5*		
30	pyruvate kinase	2.7.1.40	4	234	6.9	−	no	*4, 5*		
31	*β*-galactosidase	3.2.1.23	4	468	5.0	−	no	*1, 4, 5*		
32	ferritin		24	479	6.0	−	no	*1, 4, 5*		

[a] See SWISS-PROT database. The internet address for SWISS-PROT database: http://expasy.hcuge.ch/sprot/sprot-top.html. Molecular weights are given for the aggregate of all subunits. Values of MW are calculated from the sequences of the proteins and include the MW of any bound metal, heme group, phosphate residue or carbohydrate residue. Values of pI and Z^o_{seq}, the values of Z_{seq} of native proteins, were calculated from the sequences of native proteins using the following standard values of pK_a for the amino acid residues: Arg = 12; Lys = 10.7; Tyr = 10.2; Cys = 9.3; α-NH_3^+ = 7.4; His = 6.5; α-COOH = 4.9; Glu = 4.4; Asp = 4.0. These values are midpoints of the range of values of pK_a given in Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman and Company: New York, 1993; p 6. [b] "Capillary charge" indicates the charge on the surface of the capillaries. An uncoated capillary is negatively charged, and a Polybrene-coated capillary is positively charged. [c] Success in formation of charge ladders. The designation "pr" indicates that the rungs of the charge ladder were only partially resolved. [d] Reagents that were successful in generating charge ladders are shown in bold. In cases where a charge ladder could not be resolved, the reagents shown in italics are those that were tried. [e] Values of Z^o_{CE}, the values of Z_{CE} of native proteins, were determined by extrapolation to the *x*-intercept in plots of $\mu^n_{electro}$ vs $n\Delta Z_{seq}$, using the rungs of lowest overall charge in the corresponding charge ladders. These values were obtained using 25 mM Tris-192 mM Gly running buffer, pH 8.4, 25 °C. [f] The value of pI of bovine carbonic anhydrase II were taken from the following: Deutsch, H. F. *Int. J. Biochem.* **1987**, *19*, 101−113. Holmes, R. S. *Eur. J. Biochem.* **1977**, *78*, 511−520. [g] The value of Z^o_{CE} was estimated using 25 mM Tris-192 mM Gly running buffer, pH 8.3, at 37 °C. [h] The value of MW of carboxypeptidase B was taken from the following: Folk, J. E.; Piez, K. A.; Carroll, W. R.; Gladner, J. A. *J. Biol. Chem.* **1960**, *235*, 2272−2277. The value of pI of carboxypeptidase B was taken from the following: Sasaki, I.; Gotoh, H.; Yamamoto, R.; Hasegawa, H.; Yamahita, J.; Horio, T. J. *Biochem.* **1979**, *86*, 1537−1548. [i] The values of MW and pI of dextranase were taken from the following: Sugiura, A.; Ito, A.; Ogiso, T.; Kato, K.; Asano, H. *Biochim. Biophys. Acta* **1973**, *309*, 357−362. [j] The value of MW of peroxidase was taken from the following: Welinder, K. *Eur. J. Biochem.* **1979**, *96*, 483−502. The value of pI of peroxidase was taken from the following: Maehly, A. In *Methods in Enzymology*; Colowick, S., Kaplan, N., Eds.; Academic Press: New York, 1955; Vol. 2, pp 801−813. The value of Z^o_{seq} of peroxidase was not calculated due to the ambiguity in the number of charged carbohydrate residues on the native protein. [k] The values of Z^o_{CE} of these proteins were estimated using charge ladders formed using reagent **3**. [l] The state of aggregation of glucose-6-phosphate dehydrogenase from *Candida utilis* was taken from the following: Engel, H. J.; Domschke, W.; Alberti, M.; Domagk, G. F. *Biochim. Biophys. Acta* **1969**, *191*, 509−516.

charge, respectively. The first rung of a charge ladder of BCA II made using benzenehexacarboxylic acid monoanhydride, **5**, however, is separated from native protein by a value of ΔZ_{CE} of only 4.6 units of charge.

The observed decreases in values of $\Delta\mu_{electro}$ for successive pairs of adjacent rungs of charge ladders of BCA II may be due to decreases in values of ΔZ_{CE} and/or to increases in the coefficient of friction of the protein with increasing number of modifications. Decreases in values of ΔZ_{CE} may be due to (i) changes in the values of pK_a and affinities for ions of the ionizable groups on the protein as the magnitude of its charge increases and/or to (ii) changes in the distribution of ions surrounding the protein in the presence of an electric field in response to increases in its charge ("ion relaxation").[23] Changes in values of the coefficient of friction may be attributed to (i) contributions of mass and volume from the conjugated acylating agent and/or to (ii) conformational changes resulting from increasingly unfavorable electrostatic interactions on the surface of the protein with increasing number of acylations (the process of "electrostatic denaturation").[24]

We evaluated the contribution of mass and volume from acylating agents **1**−**5** to the coefficient of friction of BCA II by acylating the protein with reagents that introduced different

(23) Allison, S. A.; Potter, M.; McCammon, J. A. *Biophys. J.* **1997**, *73*, 133−140.

(24) Aviram, I.; Myer, Y. P.; Schejter, A. *J. Biol. Chem.* **1981**, *256*, 5540−5544.

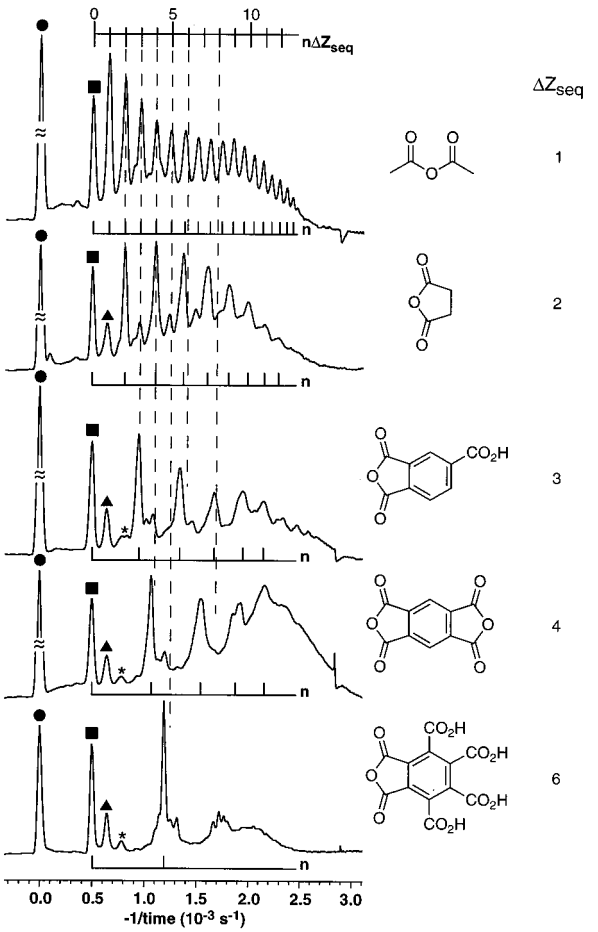


Figure 1. Formation of charge ladders of BCA II at pH 12 using acylating agents **1**−**5**. The scale (*n*) below each electropherogram indicates the number of modified Lys ϵ-NH_2 groups on the protein. The $n\Delta Z_{seq}$ scale at the top of the electropherograms has increments of one unit of charge. The two isozymes of BCA II differ by one unit of charge. Gln in the minor isozyme (pI = 5.4), labeled with a filled triangle (▲), is replaced by Arg in the major isozyme (pI = 5.9) as labeled with a filled square (■). The neutral marker, *p*-methoxybenzyl alcohol, is indicated by a filled circle (●). Modification of BCA II by acetic anhydride yields derivatives of the major isozyme that overlap with those of the minor isozyme. Modification of BCA II by acylating agents **2**−**5** distinguished the two isozymes of BCA II. The asterisks indicate impurities from the protein sample. The −1/time scale at the bottom of the set of electropherograms is proportional to the mobilities of the peaks in each electropherogram relative to the neutral marker. The electropherograms were collected at 25 °C using 25 mM Tris-192 mM Gly, pH 8.4 and an applied voltage of 15 kV.

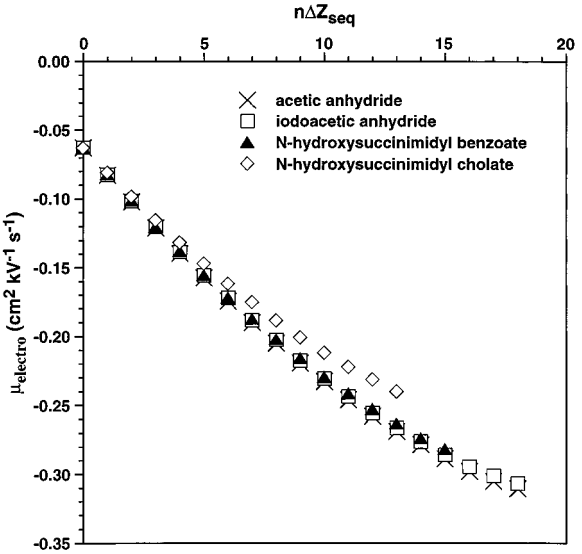


Figure 2. Plot of the mobilities of the rungs of charge ladders of BCA II formed with acetic anhydride (×), iodoacetic anhydride (□), *N*-hydroxysuccinimidyl benzoate (▲), or *N*-hydroxysuccinimidyl cholate (◇) as a function of $n\Delta Z_{seq}$. Some of the more heavily acylated rungs of the charge ladders formed with the *N*-hydroxysuccinimidyl esters of benzoic or cholic acid could not be resolved. The electropherograms were collected at 25 °C using 25 mM Tris-192 mM Gly, pH 8.4 and an applied voltage of 15 kV.

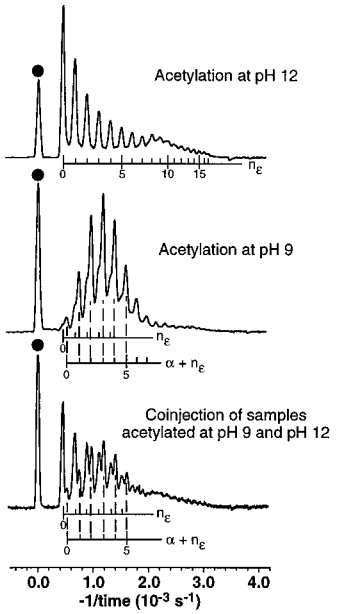


Figure 3. Acetylation of the amino groups of myoglobin at pH 12 and 9. At pH 12, acetylation occurs selectively on the Lys ϵ-NH_2 groups of the protein. At pH 9, modification is not selective, and acetylation occurs on both the α- and Lys ϵ-NH_2 groups. Coinjection experiments established the identity of the derivatives of the protein. The scales (n_ϵ and $\alpha + n_\epsilon$) below each electropherogram indicate the number of modified Lys ϵ-NH_2 groups on the native protein and α-NH_2 acetylated protein, respectively. The filled circle (●) indicates the neutral marker, *p*-methoxybenzyl alcohol. The electropherograms were collected at 37 °C using 25 mM Tris-192 mM Gly, pH 8.3 electrophoresis buffer.

values of mass and volume but only changed the value of Z_{seq} of the protein by one unit of charge per modification. Acylation of BCA II with *N*-hydroxysuccinimidyl benzoate and iodoacetic anhydride contribute 104 and 164 D of mass per acylation, respectively. We observe that the values of $\Delta\mu_{electro}$ for successive pairs of adjacent rungs of the charge ladders formed with these reagents are approximately the same as those of the corresponding pairs from the acetamide charge ladder (Figure 2); when the larger cholic acid residue (389 D) is conjugated to BCA II, however, only the values of $\Delta\mu_{electro}$ of the first three pairs of adjacent rungs of the resulting charge ladder are similar. These results suggest that the derivatives of BCA II that constitute the rungs of the acetamide and iodoacetamide charge ladders do not have significant contributions of mass and volume from their respective conjugated acylating agents. Any differences in the coefficient of friction among the derivatives that make up successive rungs of each ladder may then be attributed to differences in their conformation due to the process of "electrostatic denaturation".[24] We conclude from these studies that decreases in values of $\Delta\mu_{electro}$ for the successive pairs of adjacent rungs of the charge ladders of BCA II formed with reagents **1** and **2** are caused by electrostatic effects that influence values of ΔZ_{CE} and/or the coefficient of friction of the protein. The values of $\Delta\mu_{electro}$ of the pairs of adjacent rungs of charge

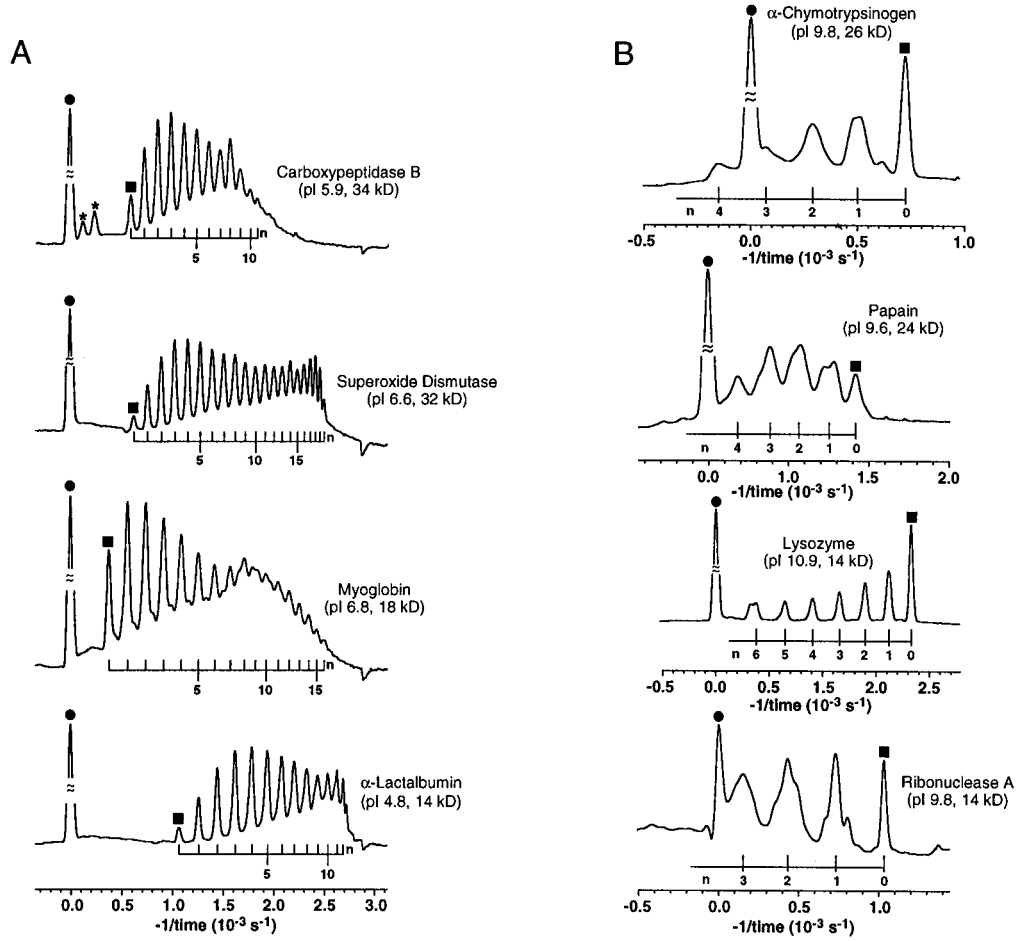


Figure 4. Representative charge ladders of proteins formed by selective acetylation of Lys ϵ-NH_2 groups at pH 12. The neutral marker, *p*-methoxybenzyl alcohol, is indicated by the filled circles (●) and native proteins are indicated by the filled squares (■). The number of acetylated Lys ϵ-NH_2 groups (n) is indicated below each electropherogram. (A) Charge ladders of proteins having values of pI < 8.4, analyzed using uncoated capillaries. (B) Charge ladders of proteins having values of pI > 8.4, analyzed using Polybrene-coated capillaries. The −1/time scale on the bottom of the electropherograms is proportional to the mobilities of the peaks in each electropherogram relative to the neutral marker.

ladders formed with reagents **3**−**5**, however, may in addition be influenced by differences in the coefficient of friction of the proteins that constitute successive rungs of each charge ladder due to appreciable contributions of mass and volume from the conjugated acylating agents.

Selective Acetylation of the ϵ-Amino Groups of Lys at High Values of pH. The α-NH_3^+ group on the N-terminus of a protein has a lower value of pK_a (pK_a ∼ 7−9) than do the ϵ-NH_3^+ groups of Lys (pK_a ∼ 11); the nucleophilicity of the α-NH_2 group is also lower than that of Lys ϵ-NH_2 groups.[4,20] Acylation of an ammonium group (RNH_3^+) occurs only by reaction of its neutral form (RNH_2) with the acylating agent. Since a neutral ϵ-NH_2 group is more basic and more nucleophilic than a neutral α-NH_2 group, reaction at a high value of pH where both amino groups are deprotonated will occur selectively on the ϵ-NH_2 group. At lower values of pH, where the α-NH_3^+ is partially deprotonated but the ϵ-NH_3^+ is essentially entirely protonated, it is possible to achieve selective acylation on the α-NH_2 position.

Figure 3 shows electropherograms collected at pH 8.3 of acetylation reactions of myoglobin carried out at pH 12 and 9. At pH 8.3, the value of ΔZ_{CE} from native protein is greater for a protein that is acetylated on one of its Lys ϵ-NH_3^+ groups than for one that has its α-NH_3^+ group acetylated, due to differences in the values of pK_a (and, consequently, of fractional protonation and average charge) of the two residues. Partial acetylation of myoglobin at pH 12 resulted in a charge ladder in which each of the first few pairs of adjacent rungs were separated by a value of ΔZ_{CE} of approximately one unit of charge. Acetylation carried out at pH 9 resulted in two overlapping charge ladders; the rungs of the minor charge ladder had values of $\mu_{electro}$ that overlapped with those of the rungs of the charge ladder made at pH 12. Acetylation of myoglobin at pH 7 gave results similar to those obtained from the modification at pH 9. A coinjection of reaction mixtures from pH 12 and 9 confirmed that two distinct charge ladders were formed by acetylation of successive ϵ-NH_2 groups at pH 9: one derived from native myogobin, and the other derived from myoglobin acetylated on the N-terminus. The results also confirmed that the ϵ-NH_2 groups on native myogobin are selectively acetylated at pH 12. The difference between the two charge ladders—approximately 1/3 of a unit of charge—is apparent in coinjection experiments and reflects the fact that at pH 8.3, the pH of the CE experiments, the α-NH_3^+ group of myoglobin is approximately 66% deprotonated (pK_a = 8.0 for Gly α-NH_3^+).

Successful Examples of Charge Ladders of Proteins. We explored the generality of formation of charge ladders by acylating proteins having a range of values of MW and pI (Table 1) and analyzing the products by CE. Some of the proteins we studied contain multiple subunits, metal ions, heme groups, glycosylated residues, and/or phosphorylated residues. Figure 4A shows several successful examples of protein charge ladders

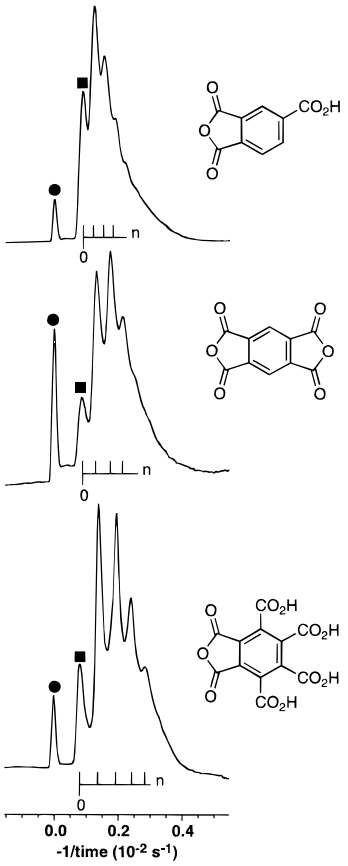


Figure 5. Electropherograms of the reaction products from modification of creatine kinase (MW = 87 kD) at pH 6.8 using acylating agents **3**−**5**. The scale of the −1/time axis is proportional to the mobilities of peaks in each electropherogram relative to the neutral marker, *p*-methoxybenzyl alcohol (●). The filled square (■) is native creatine kinase. The electropherograms were collected at 25 °C using 25 mM Tris-192 mM Gly, pH 8.4 and an applied voltage of 30 kV.

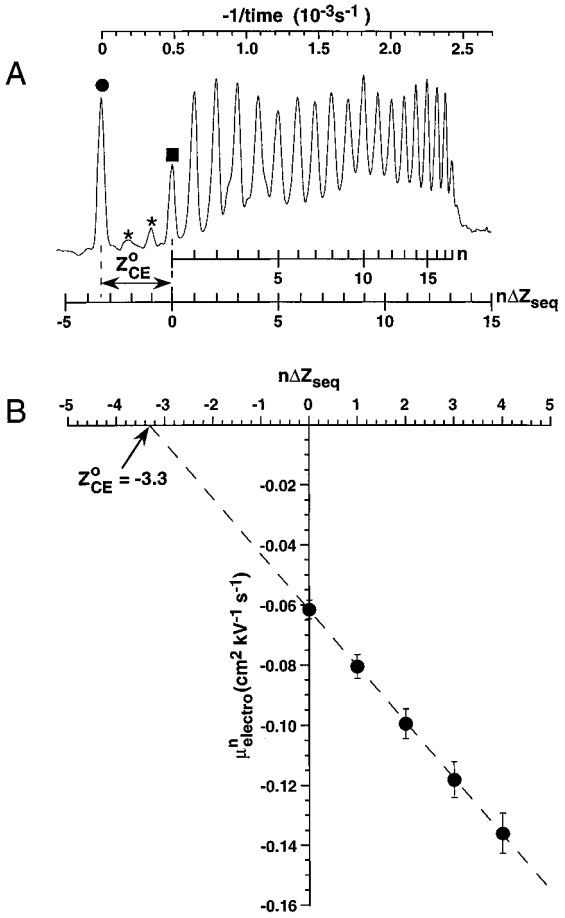


Figure 6. (A) The values of ΔZ_{CE} between each of the first four pairs of adjacent rungs of the acetamide charge ladder of BCA II are approximately uniform and can be used as a "charge ruler" to estimate Z^{o}_{CE}. The filled square (■) is native BCA II. The scale of the −1/time axis is proportional to the mobilities of peaks in the electropherogram relative to the neutral marker, *p*-methoxybenzyl alcohol (●). The asterisks indicate impurities from the protein sample. The electropherogram was collected at 25 °C using 25 mM Tris-192 mM Gly, pH 8.4. (B) The value of Z^{o}_{CE} of −3.3 of native BCA II is estimated by plotting the mobilities of the first five rungs of the acetamide charge ladder as a function of $n\Delta Z_{seq}$ (ΔZ_{seq} = one unit of charge) and extrapolating to the *x*-intercept. The assigned errors in measured electrophoretic mobilities are ±5%.

formed from proteins having values of pI < 8.4, the pH of the running buffer.

Proteins having values of pI > 8.4 have a net positive charge at pH 8.4 and tend to stick to the negatively charged surface of uncoated capillaries. To overcome this problem, we used Polybrene-coated capillaries to reverse the charges on the surface of the capillary. Creating a positively charged surface reverses the direction of electroosmotic flow (EOF).[25] Analysis of samples must therefore be done in reverse polarity and results in the grouping of rungs of charge ladders in decreasing order of modification (Figure 4B).

We assume that each rung of a charge ladder comprises proteins that have approximately the same values of Z_{CE} and coefficient of friction.[26] We have observed, however, that the more heavily acylated rungs of some charge ladders are superimposed on a broader background. We do not presently understand the nature of this broad heterogeneous peak; its characterization will be the subject of later work.

Charge Ladders of High MW Proteins with Acylating Agents 3−5. Proteins that have values of MW >50 kD do not form resolved charge ladders when modified with acetic anhydride. The lack of resolution is due to the low value of the ratio of terms $\Delta Z_{CE}/M^{\alpha} \cong 1/M^{\alpha}$ (eq 1) that accounts for the separation between adjacent rungs of the charge ladder. In order to overcome this problem, we tested the ability of acylating agents **3**−**5** to yield resolved charge ladders on modification of high MW proteins. Modifications performed with these reagents result in higher values of $\Delta Z_{CE}/M^{\alpha} \cong 3/M^{\alpha}$ to $6/M^{\alpha}$ between each of the first several pairs of adjacent rungs of the corresponding charge ladders—and consequently better resolution of these products by CE. We were able to form resolved charge ladders of creatine kinase (MW = 87 kD) (Figure 5), glucose-6-phosphate dehydrogenase (MW = 115 kD), and L-lactate dehydrogenase (MW = 147 kD) by using reagents **3**−**5**.

Estimating the Charge of Proteins in Solution by CE Using Protein Charge Ladders. Estimations of the charge of proteins over a range of values of pH have been performed using methods based on isoionic points[27] and Donnan potential

(25) Electroosmotic flow (EOF) is the bulk flow of solution in a capillary that results from the application of an electric field. The value of EOF is typically expressed as the time required for a neutral marker to migrate from the inlet of a capillary to the detector window.

(26) The poor peak shape for the rungs of some of the charge ladders shown in Figure 4B may be due to slight differences in the values of Z_{CE} and/or coefficient of friction for the different regioisomers that make up these rungs.

(27) Nozaki, Y.; Tanford, C. *Methods Enzymol.* **1967**, *11*, 715−734.

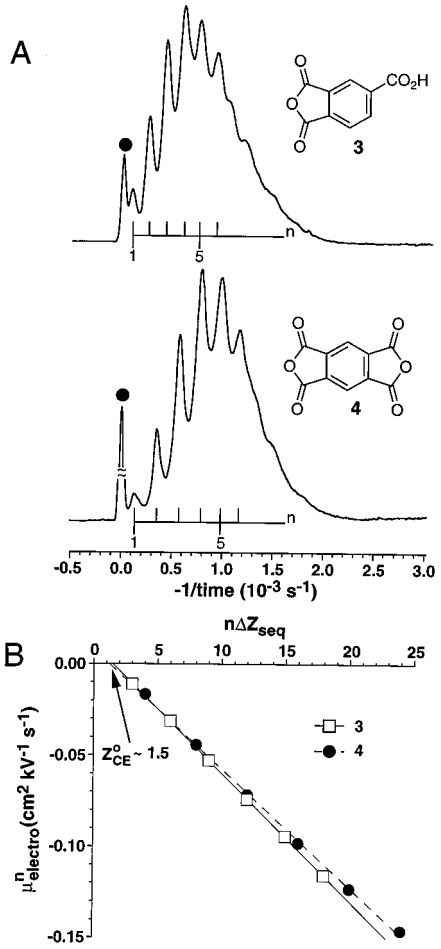


Figure 7. (A) Electropherograms of the charge ladder of L-lactate dehydrogenase (MW = 147 kD) formed at pH 6.8 using acylating agents **3** and **4**. Each unit in *n* (the number of modified Lys ϵ-NH_3^+ residues) corresponds to values of ΔZ_{seq} of approximately 3 and 4 units of charge, respectively. The neutral marker, *p*-methoxybenzyl alcohol, is indicated by the filled circles (●). The electrophoresis buffer is 25 mM Tris-192 mM Gly, pH 8.4. The −1/time scale on the bottom of the figure applies to both the electropherograms. (B) A plot of the mobilities of the rungs of the charge ladder of L-lactate dehydrogenase vs $n\Delta Z_{seq}$ used to estimate the value of Z^o_{CE} of ~1.5.

measurements.[28,29] A major drawback of these procedures, however, is that they require significant amounts (1−10 mg) of pure proteins. We have developed a method based on charge ladders that is used to estimate the charge of proteins in solution of different values of pH using only μg quantities of proteins;[3] pure samples are not required in this method provided that any impurities can be identified and separated from the rungs of the charge ladder.

The values of ΔZ_{CE} are approximately uniform for the pairs of adjacent rungs of lowest overall charge of a charge ladder and are used as a charge ruler to estimate the value of Z_{CE} of native protein, Z^o_{CE}.[3] For example, as previously described, the value of ΔZ_{CE} at pH 8.4, for each of the first four pairs of adjacent rungs of the acetamide charge ladder of BCA II, is assumed to be approximately equal to one unit of charge, the value of ΔZ_{seq} (Figure 6). To estimate the value of Z^o_{CE} of BCA II at pH 8.4, we first plot the mobilities of the first five rungs as a function of $n\Delta Z_{seq}$, where *n* is the number of acetylations. These data are then fit by the method of linear least squares to a line which is defined by eq 5. In this equation, $\mu^n_{electro}$ refers to the mobility of the *n*th rung of the charge ladder, $\mu^o_{electro}$ is the mobility of native protein, $-C_P/M^\alpha$ is the slope of the line, and $n\Delta Z_{seq}$ is approximately equal to the difference between values of Z^o_{CE} and Z^n_{CE}, the charge of the *n*th rung of the charge ladder (eq 6). Extrapolation to the *x*-intercept (where $\mu^n_{electro} = \mu^{nm}_{electro} = 0$ and $Z^n_{CE} = Z^{nm}_{CE} = 0$, nm = neutral marker) of the fit line gives the value of Z^o_{CE} of BCA II of −3.3 (eq 7). Values of Z^o_{CE} were similarly calculated for high MW proteins, such as L-lactate dehydrogenase (Figure 7) using acylating agents **3**−**5**, where ΔZ_{seq} is equal to approximately 3−6 units of charge.

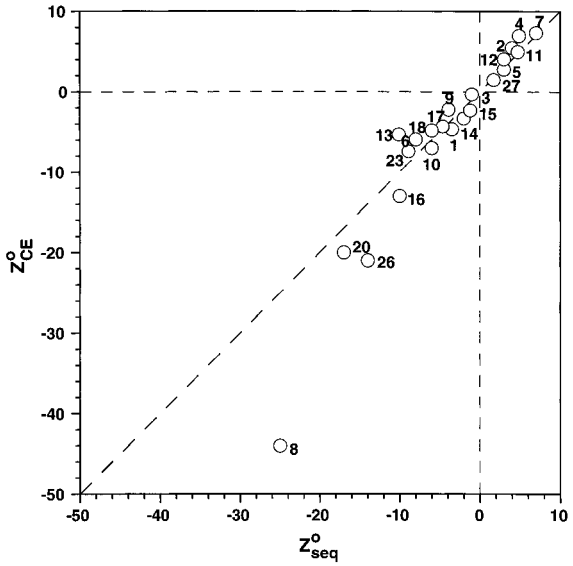


Figure 8. Comparison of values of Z^o_{CE} of proteins to the corresponding values of Z^o_{seq}. The values of Z^o_{CE} for creatine kinase (23), glucose-6-phosphate dehydrogenase (26), and L-lactate dehydrogenase (27) were estimated using charge ladders made using reagent **3**. The values of Z^o_{CE} for all other proteins represented in this figure were estimated from their respective acetamide charge ladders. The dashed diagonal line is added as a visual aid to represent a 1:1 relation between values of Z^o_{CE} and Z^o_{seq}. Peaks are numbered according to their ID numbers from Table 1.

$$\mu^n_{electro} = -\frac{C_P}{M^\alpha}(n\Delta Z_{seq}) + \mu^o_{electro} \quad (5)$$

$$n\Delta Z_{seq} \cong Z^o_{CE} - Z^n_{CE} \quad (6)$$

$$Z^o_{CE} = \frac{\mu^o_{electro}}{\left(\frac{C_P}{M^\alpha}\right)} \quad (7)$$

Figure 8 compares the values of Z^o_{CE} of different proteins to the corresponding values of Z^o_{seq}, the values of Z_{seq} of native proteins. We observed good agreement between values of Z^o_{CE} and Z^o_{seq} when Z^o_{seq} was ≥ -10. For values of $Z^o_{seq} < -10$, however, we found that absolute values of Z^o_{CE} were overestimated by the charge ladder method. We assume that the values of ΔZ_{CE} for the pairs of adjacent rungs of lowest overall charge of a charge ladder are approximately equal to the corresponding value of ΔZ_{seq}; if these values of ΔZ_{CE} are actually lower than that of ΔZ_{seq}, then the value of Z^o_{CE} will be overestimated. The value of Z^o_{CE} of calmodulin is estimated to be −44 units of charge when we assume that the value of ΔZ_{CE} of each of the first few pairs of adjacent rungs of its acetamide charge ladder

(28) Ford, C. L.; Winzor, D. J. *Biochim. Biophys. Acta* **1982**, *703*, 109−112.
(29) Öjteg, G.; Lundahl, P.; Wolgast, M. *Biochim. Biophys. Acta* **1989**, *991*, 317−323.

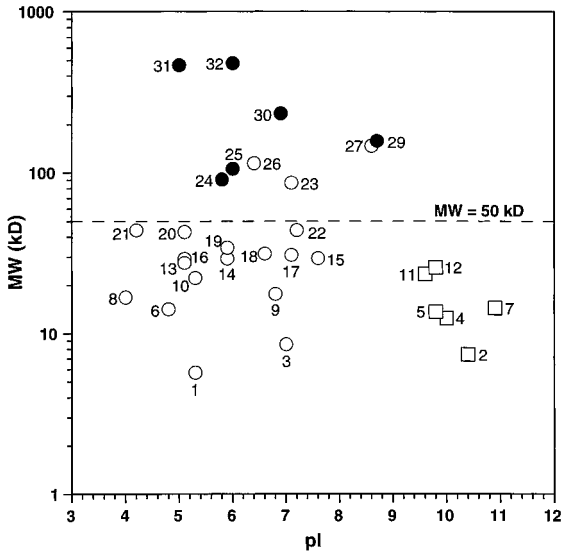


Figure 9. Summary of the results of experiments designed to form protein charge ladders. The values of pI of the proteins were plotted against their values of MW. Proteins from which we were unable to form charge ladders are indicated by filled circles (●). Proteins from which we successfully formed charge ladders are represented by open circles (○) if they were analyzed using uncoated capillaries or by open squares (□) if they were analyzed using Polybrene-coated capillaries. Peaks are numbered according to their ID numbers from Table 1.

is approximately equal to one unit of charge, the corresponding value of ΔZ_{seq}; its value of Z^o_{seq}, however, is -25 units of charge. We obtained better agreement between the values of Z^o_{CE} and Z^o_{seq} of native calmodulin when we assumed that each value of ΔZ_{CE} was equal to $0.6\Delta Z_{seq}$, i.e., 0.6 of a unit of charge, to estimate the value of Z^o_{CE}. This low value of ΔZ_{CE} for even the first several pairs of adjacent rungs of the charge ladder of calmodulin is reasonable based on the magnitude of the value of Z^o_{seq} of calmodulin.

Conclusions

The combination of charge ladders of proteins and CE is a new biophysical tool that can be used to estimate the values of charge of proteins and to explore values of pK_a, ionic association, interaction of charged ligands with proteins, and a range of other phenomena involving charged species interacting with proteins that are not always accessible by other methods. This work demonstrates that charge ladders can be made from a broad array of proteins differing in MW and pI; among the 32 proteins we examined, 25 yielded resolved charge ladders (Figure 9). We conclude that charge ladders can be formed reliably from proteins that have values of MW < 50 kD. Proteins that have values of MW > 50 kD may form charge ladders when allowed to react with reagents **3**−**5**, which impart approximately −3, −4, and −6 units of negative charge, respectively, for each ϵ-NH_3^+ group acylated. A few of the high MW proteins that we examined did not form resolved charge ladders, even with reagents **3**−**5**, instead broad unresolved peaks were observed.[30] We have not currently identified the reasons for these failures. In the cases where charge ladders of proteins were not formed, the value of the ratio of terms, $\Delta Z_{CE}/M^{\alpha}$, between different acylated rungs of these charge ladders may have to be > $6/M^{\alpha}$ to obtain sufficient separation by CE to resolve a charge ladder.

Acknowledgment. Murine carbonic anhydrase V was a gift from Dr. David Christianson (UPenn). This work was supported by the National Institute of Health Grants GM51559 and GM30367. The NMR facilities at Harvard were supported by NIH Grants 1-S10-RR04870-01 and CHE-8814019. FAB-MS were obtained by Dr. A. Tyler at the Harvard University Mass Spectrometry Facility supported by Grants from NSF (CHE-9020043) and NIH (SIO-RR06716).

JA9723491

(30) We did not include certain commercial proteins that gave ambiguous results upon acylation. Native hexokinase and pancreatic lipase from Worthington and hemoglobin from Sigma revealed multiple peaks by CE. The impurities present in these samples made assignment of the some of the rungs of the charge ladders of these proteins difficult due to the overlapping of peaks.