

# Protein Charge Ladders, Capillary Electrophoresis, and the Role of Electrostatics in Biomolecular Recognition

JEFFREY D. CARBECK, IAN J. COLTON,  
JINMING GAO, AND GEORGE M. WHITESIDES\*

*Department of Chemistry and Chemical Biology,  
Harvard University, Cambridge, Massachusetts 02138*

Received September 23, 1997

## Introduction

Life rests on a web of molecular recognition;<sup>1,2</sup> the folding of proteins, the hybridization of nucleic acids, the recognition of signaling molecules by receptors, and the interaction of enzymes with substrates and cofactors all require molecular recognition. Molecular recognition is based on shape-sensitive, noncovalent interactions between molecules or fragments of molecules.<sup>3–6</sup>

The energetics of molecular recognition can reflect van der Waals, hydrophobic, hydrogen bonding, and electrostatic interactions; all include both enthalpic and entropic contributions. The role of van der Waals interactions in molecular recognition in aqueous ionic solutions is well understood; they contribute to molecular recognition primarily through unfavorable steric interactions that *exclude* ligands with noncomplementary shapes from the recognition site of the protein. Hydrophobic interactions, although still a challenge to describe theoretically, can be estimated empirically from the observed partitioning of model compounds between aqueous and hydrocarbon solvents.<sup>7</sup> Hydrogen bonds are best thought of as a specific type of localized electrostatic interaction;<sup>8</sup> they are also well understood qualitatively. Although much theoretical and experimental work has gone into under-

standing the electrostatic properties and interactions that are central to biomolecular recognition events<sup>9–11</sup> (for a review, see ref 12), it has been difficult to measure and evaluate these interactions quantitatively.

One of the central electrostatic properties of proteins is their charge, which is determined, in principle, by the sequence and structure of the protein and the properties of the solvent. Although one can measure the isoelectric point of proteins (the value of pH at which the charge of the protein is zero), there has been no general way of measuring the charge of a protein at other values of pH or as a function of the properties of the solution. Previous attempts to estimate the effects of electrostatic interactions on the affinities of proteins for ligands have focused on the use of site-directed mutagenesis.<sup>10,13–15</sup> This technique, although powerful, is labor intensive, and the number of mutated sites that can be explored practically is limited.

In this Account we describe the use of protein charge ladders in combination with capillary electrophoresis (CE) as a new biophysical tool with which to measure the electrostatic properties and interactions that are central to biomolecular recognition events involving proteins. The partial acetylation of amino groups on a protein changes the charge of the protein and generates a protein charge ladder, that is, a series of derivatives of a protein differing in values of charge but having similar coefficients of friction. The distribution of protein derivatives group, in terms of their values of electrophoretic mobility, into the “rungs” of a charge ladder, which appear in CE as a set of peaks; each successive peak in the charge ladder differs incrementally in the number of acetylated amino groups (Scheme 1). Charge ladders in combination with CE have been used to estimate the charge of the unmodified protein<sup>16</sup> and the values of  $pK_a$  of individual ionizable groups.<sup>17,18</sup> Charge ladders in combination with affinity capillary electrophoresis (ACE) make it possible to quantify the contributions of electrostatics to the free energies of binding of charged ligands to proteins.<sup>19</sup> Although we have developed this approach primarily using charge ladders of bovine carbonic anhydrase II (BCA II, EC 4.2.1.1.) as a model system (Figure 1),<sup>16,19</sup> we have demonstrated the formation of charge ladders with a range of other proteins and biological molecules.<sup>17,18,20–23</sup>

## CE Is an Analytical Technique with Broad Application to Charged Molecules in Aqueous Solutions

CE measures the velocity of charged molecules in an aqueous solution in the presence of an applied electric field. The electrostatic force ( $f_{\text{elec}}$ , N) acting on a molecule is equal to the product of its charge ( $eZ$ , where  $e$  is the fundamental unit of charge) and the magnitude of the

Jeffrey D. Carbeck was born in Seville, Spain, in 1968. He received his B.S. from the University of Michigan in 1990 and his Ph.D. from MIT in 1995. He is currently a postdoctoral fellow with George Whitesides and will be an assistant professor of chemical engineering at Princeton University in 1998.

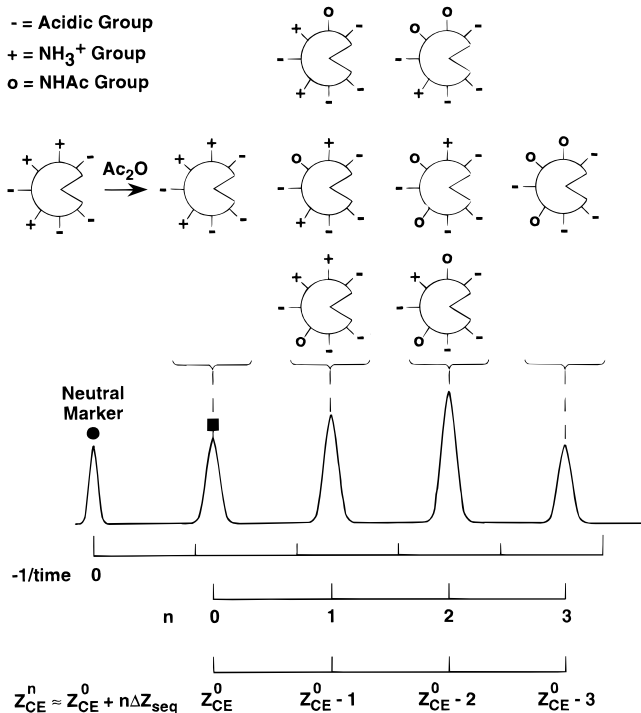
Ian J. Colton was born in Montreal, Quebec, Canada. He received his B.Sc. in 1989, M.Sc. in 1990, and Ph.D. with Romas Kazlauskas in 1995, all from McGill University. He is currently a postdoctoral fellow with George Whitesides.

Jinming Gao was born in Dalian, China, in 1970. He obtained his B.Sc. from Beijing University in 1991 and his Ph.D. with George Whitesides from Harvard University in 1996. He is currently a postdoctoral fellow with Robert Langer at MIT.

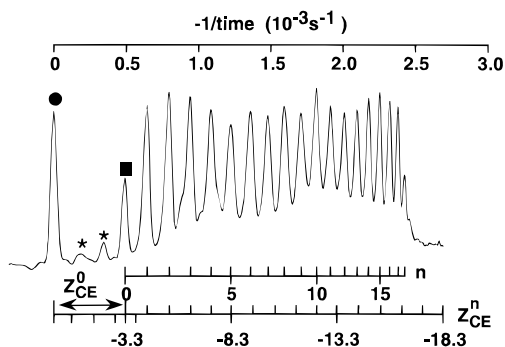
George M. Whitesides was born in Louisville, KY, in 1939. He received his B.A. from Harvard in 1960 and his Ph.D. with John D. Roberts from the California Institute of Technology in 1964. After spending almost 20 years at MIT, he joined the Department of Chemistry at Harvard University in 1982. His research interests include materials science, microfabrication, surface chemistry, rational drug design, polyvalency, molecular virology, molecular self-assembly, and optics.

\* To whom correspondence should be addressed. Phone: (617) 495-9431. Fax: (617) 495-9857. E-mail: gwhitesides@gmwgroup.harvard.edu.

**Scheme 1. Schematic Illustration of the Formation of a Charge Ladder by the Partial Acetylation of Amino Groups on the Surface of a Protein and the Resolution of the Rungs of the Charge Ladder by CE<sup>a</sup>**



<sup>a</sup> The peak marked with a filled circle is a neutral marker of electroosmotic flow, and the peak marked with a filled square is the native protein. The difference in spacing between adjacent peaks is proportional to their difference in mobilities when the data are plotted in the  $-1/\text{time}$  domain.<sup>21</sup>



**FIGURE 1.** Charge ladder of BCA II produced by the partial acetylation of Lys  $\epsilon$ -amino groups. The number of acetylated Lys  $\epsilon$ -amino groups ( $n$ ) and the charge of each rung of the ladder estimated by CE,  $Z_{CE}^n$ , are indicated below the electropherogram. (The native protein is defined as the zeroth rung of the charge ladder.) The  $-1/\text{time}$  scale is directly proportional to the mobility. The peak marked with a filled circle is a neutral marker of electroosmotic flow, and the peak marked with a filled square is the native protein. The peaks marked with an asterisk are impurities.<sup>31</sup>

applied electric field ( $E$ ,  $V\ m^{-1}$ ) (eq 1). The hydrodynamic

$$f_{\text{elec}} = eZE \quad (1)$$

force ( $f_{\text{hydro}}$ , N) acting on a molecule is equal to the product of its coefficient of friction ( $\zeta$ ,  $N\ s\ m^{-1}$ ) and its velocity

( $v$ ,  $m\ s^{-1}$ ) relative to the surrounding solvent (eq 2). The

$$f_{\text{hydro}} = \zeta v \quad (2)$$

electrophoretic mobility ( $\mu$ ,  $m^2\ V^{-1}\ s^{-1}$ , defined as the velocity per unit of electric field strength) is the result of the balance of electrostatic and hydrodynamic forces acting on the charged molecule (eq 3).

$$f_{\text{elec}} = f_{\text{hydro}} \Rightarrow \frac{eZ}{\zeta} = \frac{v}{E} \equiv \mu \quad (3)$$

There are several alternative expressions that relate the electrophoretic mobility of a molecule to its structure;<sup>24</sup> we normally use eq 4 to relate the electrophoretic mobility

$$\mu = (C_p/M^\alpha)Z \quad (4)$$

of a molecule to its charge and mass. In eq 4,  $M$  is the mass of the molecule and  $C_p$  and  $\alpha$  are empirical parameters related to the shape and solvation of the molecule, and to the ionic strength and viscosity of the solution. By combining eqs 3 and 4, we describe  $\zeta$  in terms of  $M$ ,  $\alpha$ , and  $C_p$  (eq 5). The quantities measured by CE are the

$$\zeta = M^\alpha/C_p \quad (5)$$

times of migration ( $t$ , s) of the analyte,  $t_x$ , and of the neutral marker,  $t_{nm}$ . Equation 6 describes the relationship

$$\mu = \frac{l_1 l_2}{V} \left( \frac{1}{t_x} - \frac{1}{t_{nm}} \right) \quad (6)$$

between these times of migration and the electrophoretic mobility of the analyte. The lengths  $l_1$  and  $l_2$  ( $m$ ) are the distance along the capillary between the inlet and the detector, and the total length of the capillary, respectively, and  $V$  (V) is the applied voltage.

## A Central Property of Proteins Is Their Charge

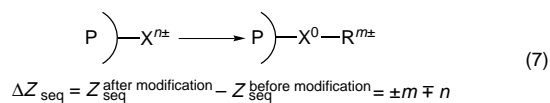
Knowledge of the charge of a protein as a function of both intensive (ionic composition, temperature) and extensive (ionic strength, pH) properties of the solvent may be useful in estimating the stability of the protein and its interaction with other charged species in solution such as ligands, membranes, small ions, and synthetic surfaces. We define the charge of a protein estimated from its sequence,  $Z_{seq}$ , as the sum over all charged components of the protein, where the charge on each component is estimated using a standard value of its ionization constant and the pH of the buffer: these charged components include ionizable amino acid residues, charged groups added posttranslationally (sialic acid groups, phosphates), and charged cofactors (metal ions, prosthetic groups and coenzymes).

The value of  $Z_{seq}$  will not, in general, be the same as the charge of a protein estimated by CE,  $Z_{CE}$ , for a particular protein and buffer due to (i) differences between the actual values of charge of the ionizable groups of the protein and the values of charge estimated from the standard ionization constants, (ii) the association of ions

other than protons with the charged groups of the protein, an effect not included in the determination of  $Z_{\text{seq}}$ , and (iii) effects related to the environment specific to the CE experiment, for example, an induced asymmetry in the distribution of counterions surrounding the protein that reduces the effective electric field acting on the protein.<sup>25</sup> Neither  $Z_{\text{seq}}$  nor  $Z_{\text{CE}}$  is necessarily the actual charge of the protein in solution,  $Z_{\text{protein}}$ , defined as the charge estimated, in principle, from the actual values of charge of the ionizable groups on the protein and ions covalently associated with the protein.  $Z_{\text{protein}}$  has yet to be measured.

## Charge Ladders Are Families of Proteins That Differ by Values of Charge

Charge ladders of proteins—and other molecules—are formed by the modification of charged groups in a way that results in a change in the charge of the protein; the functional group and the reaction are chosen so that this change is (or is assumed to be) known. Equation 7



describes the change in the calculated charge of a protein due to the modification of a charged group,  $\Delta Z_{\text{seq}}$ . (P is a protein, and  $\text{X}^{n\pm}$  is a functional group that is charged: for example, we often use  $\text{X} = \text{NH}_3^+$  and  $\text{X} \text{---} \text{R} = \text{NHCOR}'$ .)  $\Delta Z_{\text{seq}}$  is determined by the charge introduced by the modifying reagent,  $\pm m$ , and the charge of the group that is modified,  $\pm n$ . The magnitudes of both  $m$  and  $n$  are dependent on the values of  $\text{p}K_{\text{a}}$  of the group that is modified and of any ionizable groups on the modifying agent, and the pH of the solution.

Charge ladders of proteins are most conveniently produced by acetylating their Lys  $\epsilon$ -amino groups. The N-terminal  $\alpha$ -amino group is also readily modified, if it is not already naturally acetylated. If the amino group is fully charged at the pH of the electrophoresis buffer (the value of the  $\text{p}K_{\text{a}}$  of the Lys  $\epsilon$ -amino group is  $\sim 10$ , and that of the N-terminal  $\alpha$ -amino group is  $\sim 7\text{--}8$ ),  $\Delta Z_{\text{seq}} \approx -1$  for acetylation; if the amino group is partially charged at the pH of the electrophoresis buffer—as is often the case for the N-terminal  $\alpha$ -amino group— $-1 < \Delta Z_{\text{seq}} < 0$ . Equation 8 describes this relationship among the values of  $\Delta Z_{\text{seq}}$ , pH, and  $\text{p}K_{\text{a}}$ .

$$\Delta Z_{\text{seq}} = -1/1 + 10^{(\text{pH} - \text{p}K_{\text{a}})} \quad (8)$$

For many proteins, each rung of the charge ladder comprises derivatives of the native protein that have the same number of acylated amino groups and the same value of  $Z_{\text{CE}}$ .<sup>20</sup> BCA II has 18  $\epsilon$ -amino groups; partial acetylation of these groups results in an electropherogram showing 19 rungs of the charge ladder (Figure 1). The center rung of the ladder of BCA II (where 9 of the 18  $\epsilon$ -amino groups are acetylated) may, in principle, comprise  $\sim 50\,000$  regioisomers;<sup>26</sup> the total number of protein

derivatives that constitute the charge ladder of BCA II may be greater than 250 000.

The preceding analysis is based on the assumption that all Lys  $\epsilon$ -amino groups have the same reactivities. It is possible that these groups have different reactivities on the native protein, and that the reactivity of any one group may change, depending on the state of acetylation of the other groups on the protein. Such a change could occur through either steric or electrostatic interactions. At this point, we do not know the extent of these effects, nor do we know the precise composition of protein derivatives that make up a charge ladder.

Charge ladders have been produced from proteins other than BCA II and with procedures for modification other than acetylation. We surveyed 32 proteins to examine the generality of the formation of charge ladders;<sup>20</sup> Figure 2 shows representative electropherograms of charge ladders produced by acetylating the  $\epsilon$ -amino groups of Lys. Acylations of Lys  $\epsilon$ -amino groups with succinic anhydride, 1,2,4-benzenetricarboxylic anhydride, 1,2,4,5-benzenetetracarboxylic dianhydride, and benzenhexacarboxylic monoanhydride result in charge ladders with values of  $\Delta Z_{\text{seq}} = -2, -3, -4,$  and  $-6$ , respectively (Figure 3). The formation of charge ladders through the modification of different functional groups such as the guanidino group of arginine and the carboxylate group of aspartic and glutamic acids is also possible. We have also extended the concept of a charge ladder to biological molecules other than proteins, such as oligopeptides<sup>23</sup> and oligoDNA.<sup>22</sup>

## Charge Ladders, in Combination with CE, Make It Possible To Measure the Electrostatic Properties of Proteins

**Estimation of the Charge.** Equation 9, obtained from the combination of eqs 3–6, expresses the value of  $Z_{\text{CE}}$  in terms of the experimentally measured quantities  $t_{\text{nm}}$  and  $t_{\text{x}}$ . For the first several rungs of the charge ladder of BCA

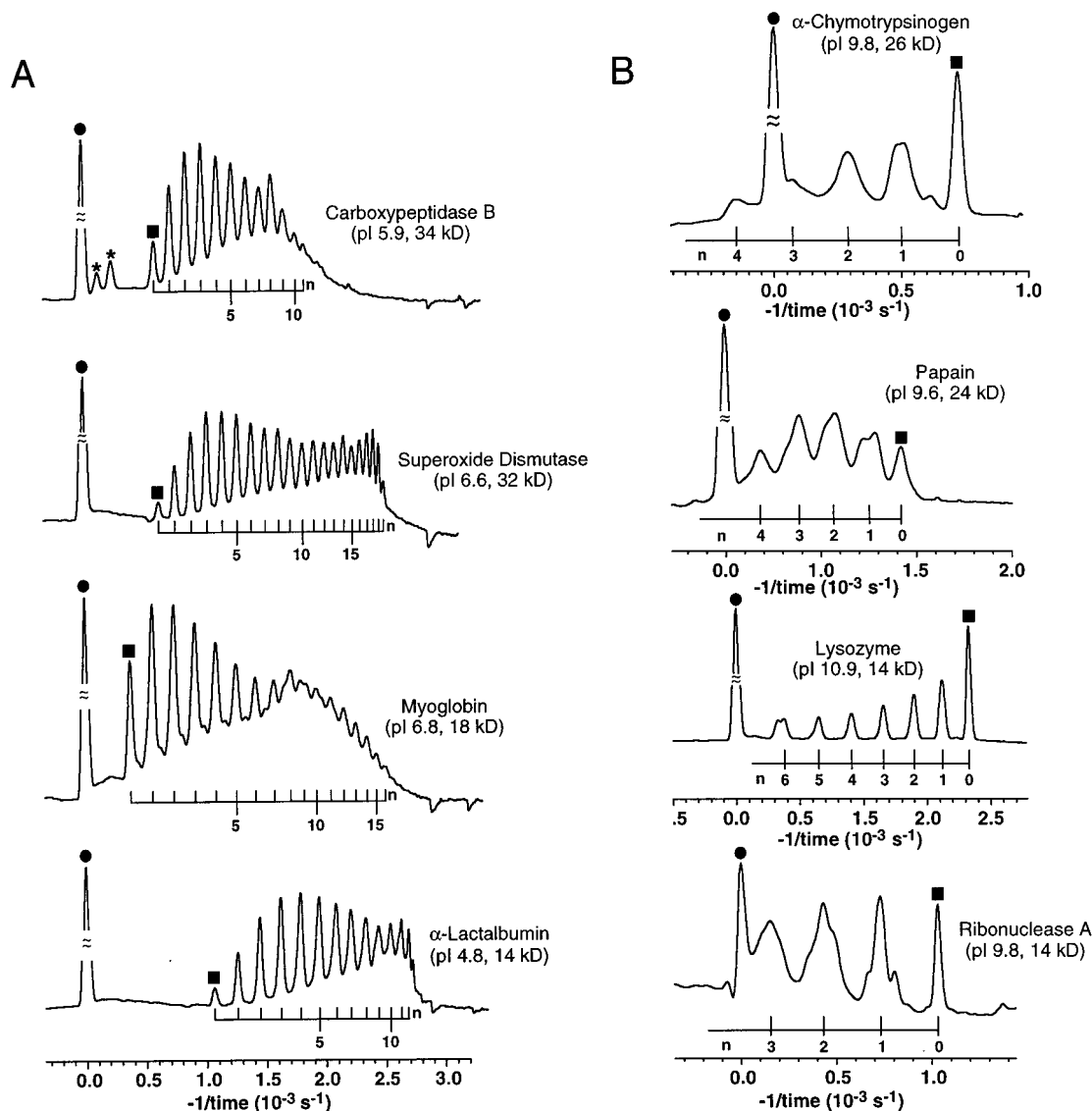
$$Z_{\text{CE}} = \frac{M^{\alpha}}{C_{\text{p}}} \frac{I_1 I_2}{V} \left( \frac{1}{t_{\text{x}}} - \frac{1}{t_{\text{nm}}} \right) \quad (9)$$

II, we observe a linear correlation between the value of  $\mu_n$  and the change in the charge relative to the native protein,  $n\Delta Z_{\text{seq}}$ , where  $n$  is the number of acetylated Lys  $\epsilon$ -amino groups.

If we assume that the change in charge measured by CE,  $\Delta Z_{\text{CE}}$ , is equal to  $\Delta Z_{\text{seq}}$  for these first several rungs of the ladder, then we can use eq 10 to describe  $\mu_n$ . The

$$\mu_n \cong \frac{C_{\text{p}}}{M^{\alpha}} (Z_{\text{CE}}^0 + n\Delta Z_{\text{seq}}) = \frac{C_{\text{p}}}{M^{\alpha}} (n\Delta Z_{\text{seq}}) + \mu_0 \quad (10)$$

slope of the linear regression analysis of  $\mu_n$  vs  $n\Delta Z_{\text{seq}}$  yields the value of  $C_{\text{p}}/M^{\alpha}$ . The  $x$ -intercept gives the charge of the native protein in solution,  $Z_{\text{CE}}^0$ ; the  $y$ -intercept is the electrophoretic mobility of the native protein,  $\mu_0$  (Figure 4A). The charge of the  $n$ th rung of the ladder,  $Z_{\text{CE}}^n$ , is estimated from eq 9.



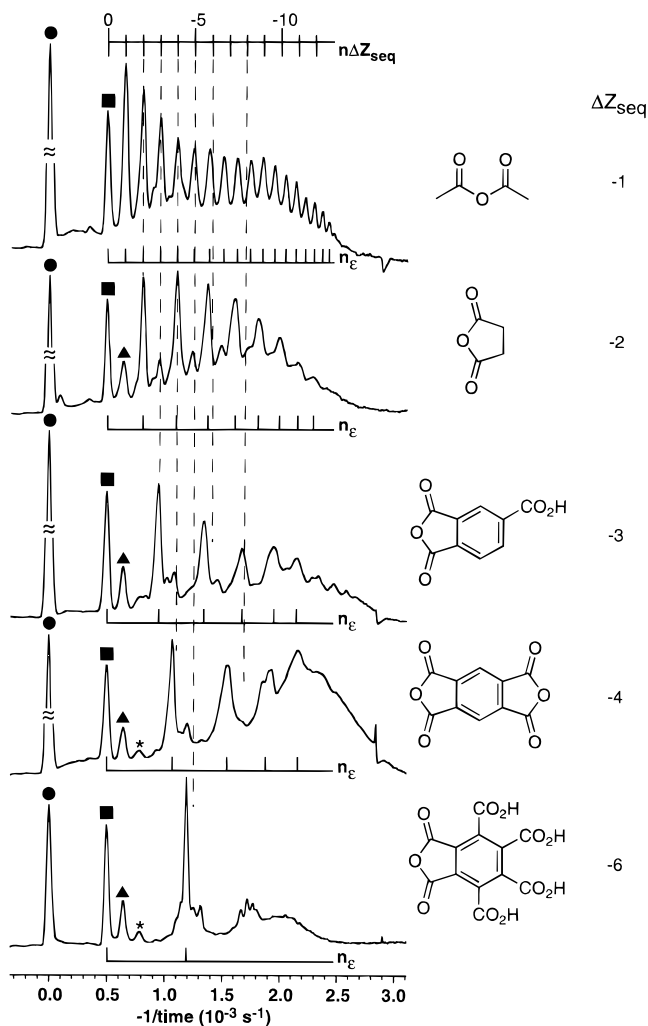
**FIGURE 2.** Electropherograms of charge ladders of proteins varying in  $M$  and  $pI$  that were produced by acetylating the  $\epsilon$ -amino groups of Lys residues: (A) acidic proteins separated on an uncoated fused silica capillary; (B) basic proteins separated on a silica capillary coated with a cationic polymer, Polybrene.<sup>31</sup> Reprinted with permission from ref 20. Copyright 1997 American Chemical Society.

In utilizing this strategy to measure values of  $Z_{CE}^n$ , we assume that the value of  $C_p/M^n$  is independent of  $n$ . We have verified this assumption is approximately true for charge ladders of BCA II by varying the molecular weight of the acylating agent;<sup>20</sup> we find that, for acetic anhydride, the  $N$ -hydroxysuccinimidyl ester of benzoic acid and iodoacetic anhydride—that is, agents that introduce 42, 104, and 164 Da of mass per acylation, respectively—the values of  $\mu_n$  are approximately the same, regardless of the value of  $M$  of the acylating agent. We infer that the acetylation of the Lys  $\epsilon$ -amino groups of BCA II has little effect on the value of  $C_p/M^n$ .

As the number,  $n$ , of acetylated Lys  $\epsilon$ -amino groups on BCA II increases, the relationship between  $\mu_n$  and  $n\Delta Z_{seq}$  becomes nonlinear (Figure 4B); that is, the change in mobility between adjacent rungs of the charge ladder,  $\Delta\mu$ , decreases with increasing values of  $n$ . As more amino groups are acetylated, the electrostatic potential at the surface of the protein becomes more negative. This

increasingly negative electrostatic potential may reduce the absolute value of  $\Delta\mu$  in several ways. First, it may decrease the charge of the protein by increasing the values of  $pK_a$  of the ionizable groups (i.e., by increasing the basicity of the conjugate bases), or by increasing the affinity of anions on the surface of the protein for positively charged ions in solution. Second, it may increase the coefficient of friction by causing the protein to unfold. Third, it may decrease the field acting on the protein by increasing the effects of ion relaxation. At this point, we do not know which of these effects dominate the observed nonlinearity.

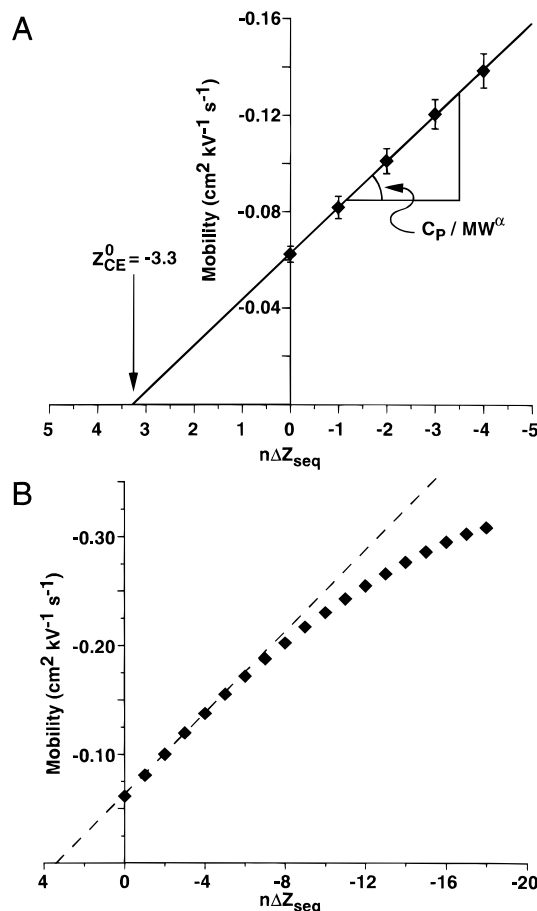
**Determination of the Values of  $pK_a$  of Ionizable Groups.** The charge of a protein is determined, in part, by the charge of the individual ionizable groups on the protein. We have used CE and charge ladders of proteins to estimate the values of  $pK_a$  of the  $\alpha$ -amino groups of insulin<sup>17</sup> and lysozyme.<sup>18</sup> Insulin has three primary amino groups: the  $\alpha$ -amino groups of the A chain ( $G^A$ ) and the



**FIGURE 3.** Electropherograms of charge ladders of BCA II formed using reagents that impart values of  $\Delta Z_{\text{seq}}$  that range from  $-1$  for acetic anhydride to  $-6$  for benzenehexacarboxylic monoanhydride. The peaks marked with a filled circle are neutral markers of electroosmotic flow, and the peaks marked with a filled square are the native protein. The peaks marked with an asterisk are impurities.<sup>31</sup> Reprinted with permission from ref 20. Copyright 1997 American Chemical Society.

B chain ( $F^\alpha$ ) and the single Lys  $\epsilon$ -amino group ( $K^\epsilon$ ). We use the symbols “+” and “o” to represent the unmodified and acetylated amino groups, respectively. Each derivative is characterized by three such symbols, representing the acetylation of  $G^\alpha$ ,  $F^\alpha$ , and  $K^\epsilon$ , respectively. (For example, the native insulin is labeled “+++”, the fully acetylated insulin is labeled “ooo”, and the derivative with only  $G^\alpha$  acetylated is labeled “o++”.) We determined the ionization constants of the  $\alpha$ -ammonium groups of insulin by analyzing the electrophoretic mobilities of the protein derivatives produced by the partial acetylation of the primary amino groups using electrophoresis buffers having different values of pH (Figure 5). Equation 11 relates

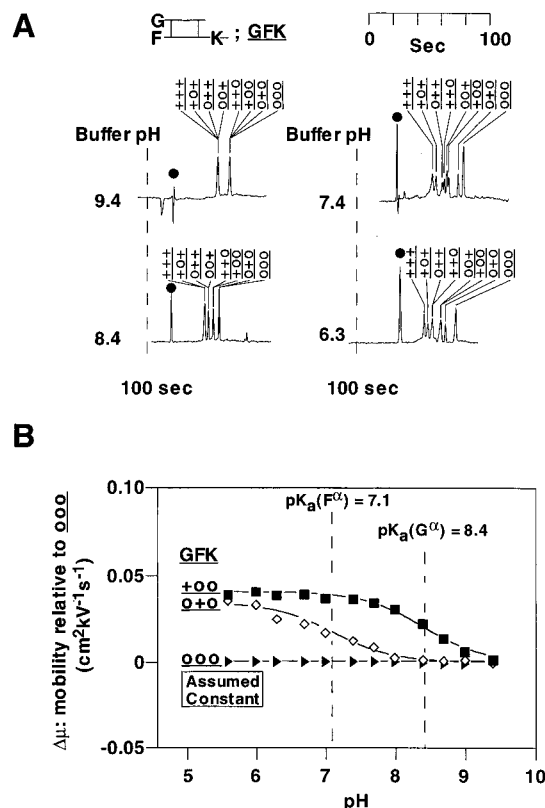
$$\Delta\mu = \frac{C_p}{M^\alpha} \left( \frac{1}{1 + 10^{(\text{pH} - \text{p}K_a)}} \right) \quad (11)$$



**FIGURE 4.** (A) Electrophoretic mobilities ( $\mu_n$ ) of the first five rungs of the charge ladder of BCA II plotted as a function of  $n\Delta Z_{\text{seq}}$ . The slope of the linear regression analysis of  $\mu_n$  versus  $n\Delta Z_{\text{seq}}$  yields the value of  $C_p/M^\alpha$ ; the x-intercept gives the charge of the native protein in solution,  $Z_{\text{CE}}^0$  (see eq 10 in the text). (B) Values of  $\mu_n$  of all the rungs of the charge ladder of BCA II plotted as a function of  $n\Delta Z_{\text{seq}}$ . The dashed line is the result of the linear regression analysis for the first five rungs of the ladder.<sup>31</sup>

the change in electrophoretic mobility ( $\Delta\mu$ ) of a species resulting from the acetylation of a basic residue to the values of  $C_p/M^\alpha$ , the pH of the electrophoresis buffer, and the  $\text{p}K_a$  of the ionizable residue.

We analyzed the electrophoretic mobilities of the oo and o+o derivatives of insulin to determine the value of the  $\text{p}K_a$  of  $G^\alpha$  and that of  $F^\alpha$ , respectively; the two other amino groups in each species are acetylated and do not contribute to changes in the total charge as the value of the pH is changed. The effects of other ionizable residues (e.g., His) were eliminated by subtracting the electrophoretic mobilities of the peracetylated derivative from those of the oo and o+o derivatives ( $\Delta\mu = \mu - \mu_{\text{ooo}}$ ): in this way, only the extent of protonation of either  $G^\alpha$  or  $F^\alpha$  contributed to changes in charge and, hence, to  $\Delta\mu$ . Plots of  $\Delta\mu$  versus pH for the oo and o+o derivatives of insulin are shown in Figure 5. Equation 11 was fitted to the [ $\Delta\mu$ , pH] data using values of  $C_p/M^\alpha$  and the value of the  $\text{p}K_a$  of  $G^\alpha$  and that of  $F^\alpha$ , respectively, as adjustable parameters: the least-squares best-fit values of  $\text{p}K_a$  were  $8.4(\pm 0.1)$  for  $G^\alpha$  and  $7.1(\pm 0.1)$  for  $F^\alpha$ .

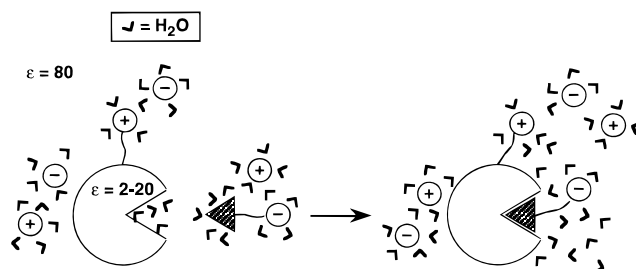


**FIGURE 5.** (A) Electropherograms of the charge ladders of bovine insulin, produced by partial acetylation of the primary amino groups, at several values of the pH of the electrophoresis buffer. The peaks marked with a filled circle are neutral markers of electroosmotic flow. (See the text for the definitions of the symbols.) (B) Dependence of the electrophoretic mobility relative to that of the fully acetylated insulin ( $\Delta\mu$ ) on the pH of the electrophoresis buffer for the  $+oo$  and  $o+o$  derivatives of the insulin charge ladder. Values of  $\Delta\mu$  were obtained by subtracting the value of  $\mu_{ooo}$  from the value of  $\mu$  measured for each derivative: the mobility of  $ooo$  is assumed constant. The curves superimposed on the data for the  $+oo$  and  $o+o$  derivatives were obtained by performing nonlinear least-squares fits using eq 11.

## Charge Ladders and Affinity Capillary Electrophoresis Can Be Used To Estimate the Contributions of Electrostatics to the Free Energy of Protein–Ligand Binding

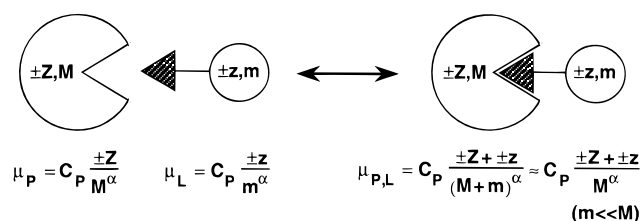
Electrostatic interactions in a vacuum, or in a simple (homogeneous) dielectric medium, are described completely by Coulomb's law. Electrostatic interactions in biologically relevant environments are more difficult to model (Scheme 2). Electrostatic interactions are long range, and their quantification involves summing over large volumes of solution and large numbers of atoms. Furthermore, electrostatic interactions in aqueous electrolytes are not purely enthalpic: the desolvation of ions, the orientation of molecular dipole moments, and the distribution of small ions in solution in response to the local electric field all involve both enthalpic and entropic contributions. As a result, electrostatic interactions in a biological environment are difficult to measure and rationalize.

### Scheme 2. Schematic Illustration of the Complexities of Electrostatic Interactions in Biomolecular Recognition<sup>a</sup>



<sup>a</sup> These environments are characterized by high concentrations of ions, and by a heterogeneous dielectric constant ( $\epsilon$ ); the dielectric constant of bulk water ( $\epsilon \approx 80$ ) is very different from that of water near the surface of the protein and of the interior of proteins ( $\epsilon \approx 2-20$ ).<sup>32</sup>

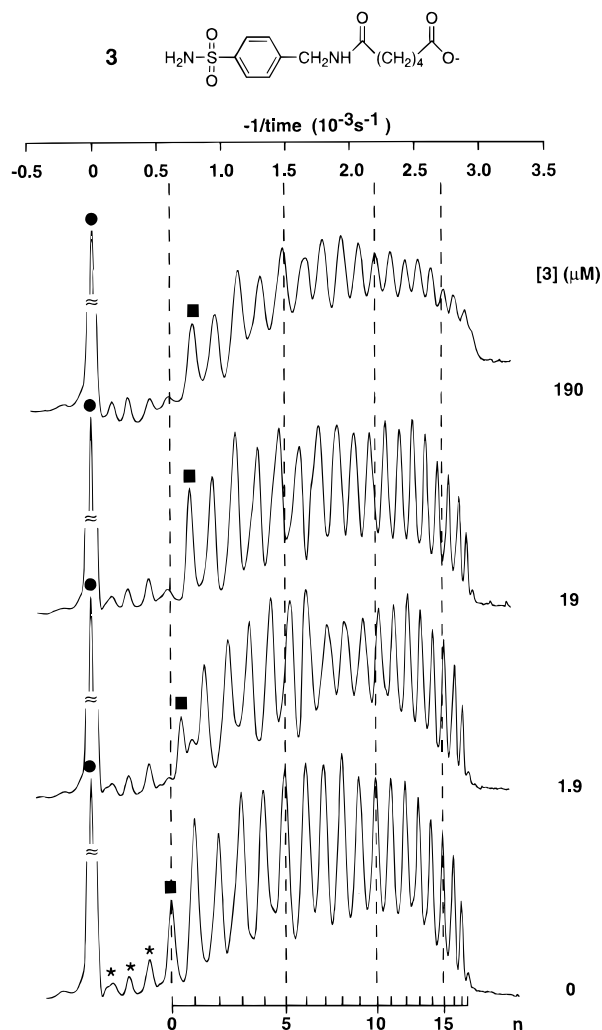
### Scheme 3. Schematic Illustration of the Binding of a Ligand (L) (of Charge $\pm z$ and Mass $m$ ) to a Protein (P) (of Charge $\pm Z$ and Mass $M$ ) and the Effects on Electrophoretic Mobility, $\mu$



Affinity capillary electrophoresis (ACE) is an extension of affinity gel electrophoresis that uses the change in electrophoretic mobility of a protein upon the binding of a ligand to quantify the free energies of binding (Scheme 3). It has been the subject of a previous review.<sup>27</sup> The combination of ACE and protein charge ladders allows the direct measurement of the influence of electrostatic interactions on the binding affinities of proteins for ligands in biologically relevant environments. A strength of the combination of ACE and protein charge ladders is the ability to measure the binding affinity of all the protein derivatives that make up the charge ladder for a common ligand simultaneously. This large number of directly comparable data allows one to measure small changes in the binding affinity of ligands for protein derivatives that differ in their charge with high accuracy.

Using ACE, we estimated the affinity constants of the protein derivatives that constitute the rungs of the charge ladder of BCA II for charged and neutral inhibitors.<sup>19</sup> The affinity constants were estimated from the change in mobility of the rungs of the charge ladder as the concentration of ligand in the running buffer was increased (Figure 6). Most or all the regioisomers that constitute each rung of the charge ladder seem to have a similar interaction with the ligand.<sup>28</sup>

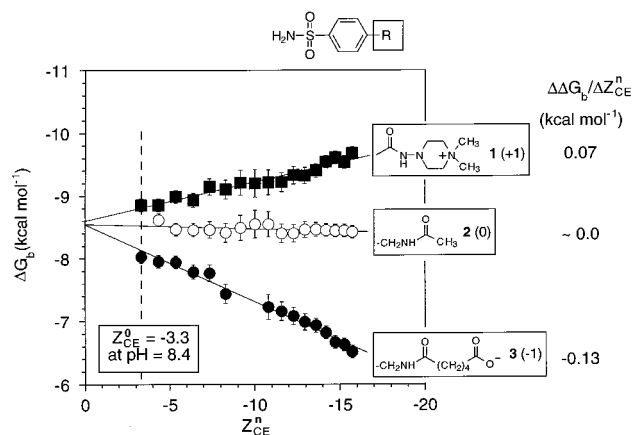
The free energies of binding ( $\Delta G_b$ ,  $\text{kcal mol}^{-1}$ ), measured by ACE, of the protein derivatives that make up the charge ladder of BCA II to several charged and neutral inhibitors were plotted as a function of the value of  $Z_{CE}^n$  (Figure 7). For the neutral inhibitor, there is no measurable change in  $\Delta G_b$  across the rungs of the charge ladder. From this result, we infer that the conformation around the active site of BCA II does not change upon acetylation of its Lys  $\epsilon$ -amino groups.<sup>29</sup>



**FIGURE 6.** Electropherograms demonstrating the changes in the electrophoretic mobility of the rungs of the charge ladder of BCA II due to the binding of ligand **3**.<sup>31</sup> The peaks marked with a filled circle are neutral markers, and the peaks marked with a filled square are the native protein. The peaks marked with an asterisk are impurities. The number of acetylated Lys  $\epsilon$ -amino groups ( $n$ ) is indicated below the electropherogram. The  $-1/\text{time}$  scale shown at the top of the figure applies to all the electropherograms.

For charged ligands,  $\Delta G_b$  is proportional to the charge of the proteins that comprise each rung of the ladder. Quantitatively, the magnitude of the dependence of  $\Delta G_b$  on the charge of the proteins that constitute the charge ladder is 0.07 and 0.13 kcal mol<sup>-1</sup> per unit charge for the charged ligands **1** and **3**, respectively. If we describe the distribution of charges of the protein in terms of its charge, dipole, and higher multipole moments, then the  $y$ -intercept of the linear regression analysis of a plot of  $\Delta G_b$  vs  $Z_{CE}^n$  is an estimate of the free energy of protein–ligand binding in the absence of any charge on the protein.

The crystal structure of BCA II is not known, but the crystal structures of several other isozymes from several species, including human CA II (HCA II), are known; all these enzymes have similar configurations of the peptide backbone. The active site of HCA II, defined by the location of the tetrahedral complex between a Zn(II) ion and three imidazole rings, sits at the bottom of a conical



**FIGURE 7.** Dependence of the free energy of binding ( $\Delta G_b$ ) on the charge ( $Z_{CE}^n$ ) of the rungs of the charge ladder of BCA II, and on the charge on the ligands. The binding affinity of each rung of the charge ladder to ligands **1–3** was measured by ACE in 25 mM Tris–192 mM Gly (pH 8.4). Due to slight broadening of peaks near the center of the charge ladder ( $n = 6–12$ ), the uncertainties in the values of  $\Delta G_b$  for these derivatives are larger than those for the other parts of the charge ladder (for some values of  $n$ , the peak broadening resulted in missing data). The slopes ( $\Delta\Delta G_b/\Delta Z_{CE}^n$ ) from the linear regression analyses of  $\Delta G_b$  vs  $Z_{CE}^n$  yielded the magnitudes of influence of charges on CAII–ligand interactions. The observation that the lines for **2** and **3** extrapolate to the same value of  $\Delta G_b$  at  $Z_{CE}^n = 0$  suggests that the groups attached to the benzenesulfonamide moiety make a very similar contribution to binding through nonelectrostatic interactions. (This plot differs from the original plot in ref 19: here, the values of  $\Delta G_b$  are plotted as a function of  $Z_{CE}^n$ ; in ref 19 the data were plotted as a function of  $Z_{CE}^0 + n\Delta Z_{seq}$ .)

binding pocket approximately 15 Å in depth. Assuming that BCA II has the same backbone structure as human CA II, we propose that all the Lys  $\epsilon$ -amino groups on BCA II are situated outside of the conical binding pocket. We conclude that electrostatic interactions between the charges on a pendant group attached to the ligand and charges on the protein that are outside the conical binding pocket can influence the free energy of binding.

We suggest that there are at least two ways that the annihilation of charges on the protein outside of the binding pocket can influence the electrostatic contributions to the free energy of binding of charged ligands: (i) the annihilation may contribute *directly* to changes in the electrostatic potential in the binding pocket; (ii) the annihilation influences the values of the charge of the other ionizable groups on the protein (results in a change in their values of  $pK_a$ ) and thereby may contribute *indirectly* to changes in the electrostatic potential in the binding pocket. At present, we cannot distinguish between these two mechanisms.

## Conclusions

A protein charge ladder represents a large number of derivatives of a protein that differ in the number and location of charged groups. CE separates these protein derivatives into the “rungs” of a charge ladder. The charge ladder gives an estimate of the coefficient of friction of

proteins that make up the rungs of the ladder that allows the direct determination of the values of charge of proteins from values of electrophoretic mobility.

Charge ladders in combination with CE provide a new biophysical tool for the direct measurement of the electrostatic properties of biomolecules ( $Z_{CE}$ ) and of the influence of electrostatic interactions on proteins (protein–ligand binding) that are central to biochemistry. The precision that this combination offers makes it possible to estimate the contributions of electrostatic interactions that have previously been difficult to investigate.

A surprising finding is the homogeneity in the electrostatic properties and interactions of the protein derivatives that constitute a rung of the ladder: for example, the binding affinity of the protein derivatives that make up the charge ladder of BCA II for a common charged ligand depends largely on a single property of these derivatives—their charge—and not on the specific location of the sites that are modified.

Although much of our work has focused on the charge ladders of carbonic anhydrase, we have demonstrated that charge ladders can be formed from many proteins. We have also extended these techniques through the modification of other functional groups—guanidino, carboxylate,<sup>30</sup> and phosphate—and of other biological molecules—small peptides<sup>23</sup> and oligonucleotides.<sup>22</sup> Work exploiting these ladders is just beginning.

*This work was supported by the National Institutes of Health (Grants GM51559 and GM30367).*

## References

- Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 3rd ed.; Garland Publishing, Inc.: New York, 1994.
- Whitesides, G. M.; Mathias, J. P.; Seto, C. T. *Science* **1991**, *254*, 1312–1319.
- Lehn, J.-M. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 89–112.
- Cram, D. J. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1009–1020.
- Lehn, J.-M. *Science* **1993**, *260*, 1762–1763.
- Philp, D.; Stoddart, J. F. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1154–1196.
- Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147–153.
- Simanek, E. E.; Mathias, J. P.; Seto, C. T.; Chin, D.; Mammen, M.; Gordon, D. M.; Whitesides, G. M. *Acc. Chem. Res.* **1995**, *28*, 37–44.
- Gilson, M. K.; Honig, B. H. *Nature* **1987**, *330*, 84–86.
- Schreiber, G.; Fersht, A. R. *Nat. Struct. Bio.* **1996**, *3*, 427–431.
- Sternberg, M. J. E.; Hayes, F. R. F.; Russell, A. J.; Thomas, P. G.; Fersht, A. R. *Nature* **1987**, *330*, 86–88.
- Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.
- Russell, A. J.; Thomas, P. G.; Fersht, A. R. *J. Mol. Biol.* **1987**, *193*, 803–813.
- Russell, A. J.; Fersht, A. R. *Nature* **1987**, *328*, 496–500.
- Thomas, P. G.; Russell, A. J.; Fersht, A. R. *Nature* **1985**, *318*, 375–376.
- Gao, J.; Gomez, F. A.; Haerter, R.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12027–12030.
- Gao, J.; Mrksich, M.; Gomez, F. A.; Whitesides, G. M. *Anal. Chem.* **1995**, *67*, 3093–3100.
- Córdova, E.; Gao, J.; Whitesides, G. M. *Anal. Chem.* **1997**, *69*, 1370–1379.
- Gao, J.; Mammen, M.; Whitesides, G. M. *Science* **1996**, *272*, 535–537.
- Colton, I. J.; Anderson, J. R.; Gao, J.; Chapman, R. G.; Isaacs, L.; Whitesides, G. M. *J. Am. Chem. Soc.* **1997**, *119*, 12701–12709.
- Mammen, M.; Colton, I. J.; Carbeck, J. D.; Bradley, R.; Whitesides, G. M. *Anal. Chem.* **1997**, *69*, 2165–2170.
- Carbeck, J. D.; Bradley, R.; Whitesides, G. M. Unpublished results.
- Rao, J.; Colton, I. J.; Whitesides, G. M. *J. Am. Chem. Soc.* **1997**, *119*, 9336–9340.
- Grossman, P. D.; Colburn, J. C. *Capillary Electrophoresis Theory and Practice*; Academic Press: San Diego, 1992.
- Allison, S. A.; Potter, M.; McCammon, J. A. *Biophys. J.* **1997**, *73*, 133–140.
- Assuming equal reactivities of the  $\epsilon$ -amino groups, the number of regioisomers in the  $n$ th rung of a charge ladder is given by the binomial coefficient  $N!/n!(N-n)!$ , where  $N$  is the total number of  $\epsilon$ -amino groups on the protein.
- Chu, Y.-H.; Avila, L. Z.; Gao, J.; Whitesides, G. M. *Acc. Chem. Res.* **1995**, *28*, 461–468.
- The resolution of the peaks is not sufficient to investigate the behavior of each rung of the charge ladder in detail. The broadening of the middle rungs of the ladder may indicate heterogeneity in the interaction of the protein derivatives with the ligand.
- If a Lys  $\epsilon$ -amino group is part of the active site, or involved in a specific interaction that aids in the formation of structure (e.g., a salt bridge), then acetylation of this group may disrupt the conformation and/or binding properties of the protein.
- Anderson, J. R.; Colton, I. J.; Whitesides, G. M. Unpublished results.
- Separations were performed at 25 °C on a 47 cm silica capillary (40 cm from inlet to detector; i.d. 50  $\mu$ m) using a running buffer of 25 mM Tris–192 mM Gly (pH 8.4). Detection was at 214 nm.
- Antosiewicz, J.; McCammon, J. A.; Gilson, M. *J. Mol. Biol.* **1994**, *238*, 415–436.

AR970255Q