

Accelerated Articles*Anal. Chem.* 1995, 67, 3093–3100**Using Capillary Electrophoresis To Follow the Acetylation of the Amino Groups of Insulin and To Estimate Their Basicities****Jinming Gao, Milan Mrksich, Frank A. Gomez,<sup>†</sup> and George M. Whitesides\****Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138*

Capillary electrophoresis (CE) is an analytical method that is useful for investigating processes that modify the charge of proteins. This paper explores the ability of CE to rationalize charges and electrophoretic mobilities of a simple protein—insulin and its acylated derivatives—as a function of pH. Insulin is a peptide hormone (MW = 5700) that has two  $\alpha$ -amino groups ( $G^\alpha$  and  $F^\alpha$ ) and one  $\epsilon$ -amino group ( $K^\epsilon$ ). Treatment of insulin with acetic anhydride affords seven derivatives that differ in the sites of acetylation of the three amino groups. Analysis of the pH dependence of the electrophoretic mobilities of these derivatives gives  $pK_a$  values for the two N-terminal ammonium groups:  $pK_a(G^\alpha) = 8.4$ ;  $pK_a(F^\alpha) = 7.1$ . Values of the total charge of insulin estimated from electrophoretic mobility differ from those estimated from values of  $pK_a$  for its ionizable groups by less than 0.5 unit for both bovine and human insulins over the range of pH from 5.5 to 9.5. Analysis of the concentration dependence of the electrophoretic mobility of insulin yields a lower limit for the association constant for dimerization of insulin of  $K_D \geq 6 \times 10^3 M^{-1}$  (25 mM Tris and 192 mM Gly, pH 8.4). Studies of electrophoretic mobility as a function of pH and extent of acetylation of amino groups rationalize the charge of insulin in detail. The sensitivity of CE to charge permits the quantitative study of electrostatic properties of proteins in solution. Insulin is a useful small-protein model with which to investigate phenomena in electrophoresis.

Capillary electrophoresis (CE) can analyze small quantities of proteins and peptides with high resolution under nondenaturing

conditions.<sup>1–6</sup> CE has been used for peptide mapping,<sup>7,8</sup> measuring the total charge of a protein in solution,<sup>9</sup> and studying the binding affinities of ligands to proteins.<sup>10–12</sup> Here we describe the use of CE to characterize the products formed upon acetylation of the amino groups of insulin, to estimate the basicities of these amino groups, and to estimate the association constant for dimerization of insulin. These studies illustrate the value of variations in pH in deciphering complex CE data and suggest the accuracy of  $pK_a$ 's derived from CE.

**Basis for Separation in CE.** The rate of migration of a protein in a capillary is characterized by its electrophoretic mobility,  $\mu$  ( $cm^2 kV^{-1} s^{-1}$ ), which is defined as the steady-state velocity per unit of electric field strength. It is calculated by comparing the migration time of the protein ( $t$ ) to that of a neutral marker ( $t_{e0}$ , for mesityl oxide) using eq 1;  $L_{tot}$  is the total length of the capillary,  $L$  is the length of capillary separating the points of sample injection and detection, and  $V$  is the voltage applied across the capillary. Equation 2 is an approximate expression that

- (1) *Capillary Electrophoresis: Theory and Practice*; Grossman, P. D., Colburn, J. C., Eds.; Academic Press: San Diego, CA, 1992.
- (2) Engelhardt, H.; Beck, W.; Kohr, J.; Schmitt, T. *Angew. Chem., Int. Ed. Engl.* 1993, 32, 629–766.
- (3) Kuhr, W. G.; Monnig, C. A. *Anal. Chem.* 1992, 64, 389–407.
- (4) Novotny, M. V.; Cobb, K. A.; Liu, J. *Electrophoresis* 1990, 11, 735–749.
- (5) Karger, B. L.; Cohen, A. S.; Guttman, A. J. *Chromatogr.* 1989, 492, 585–614.
- (6) Gordon, M. J.; Huang, X.; Pentoney, S. L.; Zare, R. N. *Science* 1988, 242, 224–228.
- (7) Mazzeo, J. R.; Martineau, J. A.; Krull, I. S. *Anal. Biochem.* 1993, 208, 323–329.
- (8) Cobb, K. A.; Novotny, M. V. *Anal. Chem.* 1992, 64, 879–886.
- (9) Gao, J.; Gomez, F. A.; Haerter, R.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U.S.A.* 1994, 91, 12027–12030.
- (10) Chu, Y.-H.; Avila, L. Z.; Biebuyck, H. A.; Whitesides, G. M. *J. Med. Chem.* 1992, 35, 2915–2917.
- (11) Avila, L. Z.; Chu, Y.-H.; Blosser, E. C.; Whitesides, G. M. *J. Med. Chem.* 1993, 36, 126–133.
- (12) Gomez, F. A.; Avila, L. Z.; Chu, Y.-H.; Whitesides, G. M. *Anal. Chem.* 1994, 66, 1785–1791.

<sup>†</sup> Present address: Department of Chemistry and Biochemistry, California State University, Los Angeles, 5151 State University Drive, Los Angeles, CA 90032.

$$\mu = \frac{LL_{\text{tot}}}{V} \left( \frac{1}{t_{\text{eo}}} - \frac{1}{t} \right) \quad (1)$$

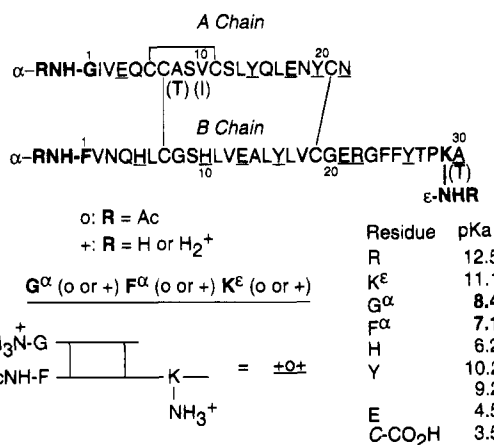
$$\mu \cong C_{\text{eff}} \frac{Z_{\text{eff}}}{M^\alpha} \cong C_P \frac{Z_P}{M^\alpha} \quad (2)$$

relates the electrophoretic mobility of a protein in an electrostatic field to its effective charge  $Z_{\text{eff}}$  or unperturbed charge  $Z_P$  and its mass  $M$ ;  $C_{\text{eff}}$  and  $C_P$  are proportionality constants for  $Z_{\text{eff}}$  and  $Z_P$ , respectively. The mass of the protein is related to the hydrodynamic drag as it moves through the buffer by the term  $M^\alpha$  (typically  $M^{2/3}$  for globular proteins<sup>13,14</sup>). The effective charge is less than the unperturbed charge ( $Z_P$ ) of a protein as modified (screened) by weak interactions with ions in solution.<sup>15–18</sup> The value of  $Z_P$  of a protein is determined by its charged components, including charged amino acid side chains, N-terminal amino and C-terminal carboxyl groups, bound metal ions or charged cofactors, ligands and prosthetic groups, and charged groups introduced by covalent modification (e.g., the sialic acid residues of glycosides). The charge of a protein can be calculated ( $Z_{\text{calc}}$ ) from values of  $\text{pK}_a$  of its ionizable residues; the value of  $Z_{\text{calc}}$  is equal to  $Z_P$ , provided that all ionization constants are known accurately.

We chose insulin for this study because of its structural simplicity and medicinal significance.<sup>19–23</sup> Insulin is a peptide hormone (MW = 5700) composed of two chains; these chains are linked by two disulfide bonds, and there is a third disulfide bond within the A chain. Insulin has only three primary amino groups (Figure 1): an  $\alpha$ -amino group from the N-terminal glycine residue of the A chain ( $G^\alpha$ ), an  $\alpha$ -amino group from the N-terminal phenylalanine residue of the B chain ( $F^\alpha$ ), and one  $\epsilon$ -amino group from the lysine residue at position 29 of the B chain ( $K^\epsilon$ ). We used both bovine pancreatic and human recombinant insulins for the studies described here. Bovine and human insulins differ only in three residues; residue A8, Ala or Thr, respectively; residue A10, Val or Ile; residue B30, Ala or Thr (Figure 1).

**Reactivity of Amino Groups.** The  $\alpha$ -amino groups of a protein are less basic than the  $\epsilon$ -amino groups of Lys because they are closer to an amide group. Acylation of an ammonium group ( $\text{RNH}_3^+$ ) occurs only by reaction of its deprotonated neutral form ( $\text{RNH}_2$ ) with the acylating agent. Although neutral  $\epsilon$ -amino groups are more basic and more reactive toward acylation than neutral N-terminal  $\alpha$ -amino groups, the former have a smaller fraction in unionized form at physiological values of pH; eq 3

- (13) Basak, S. K.; Ladisch, M. R. *Anal. Biochem.* **1995**, *226*, 51–58.  
 (14) Rickard, E. C.; Strohl, M. M.; Nielsen, R. G. *Anal. Biochem.* **1991**, *197*, 197–207.  
 (15) Atkins, P. W. In *Physical Chemistry*; Atkins, P. W., Ed.; W. H. Freeman & Co.: New York, 1986; pp 237–243.  
 (16) Grossman, P. D. In *Capillary Electrophoresis: Theory and Practice*; Grossman, P. D., Colburn, J. C., Eds.; Academic Press: San Diego, CA, 1992; pp 114–118.  
 (17) Abramson, H. A.; Moyer, L. S.; Gorin, M. H. In *Electrophoresis of Proteins and the Chemistry of Cell Surfaces*; Abramson, H. A., Moyer, L. S., Gorin, M. H., Eds.; Reinhold Publishing Corp.: New York, 1942; pp 105–172.  
 (18) For discussion of the relationship between charge and mobility, see: Compton, B. J.; O'Grady, E. A. *Anal. Chem.* **1991**, *63*, 2597–2602.  
 (19) Muller-Wieland, D.; Streicher, R.; Siemeister, G.; Krone, W. *Exp. Clin. Endocrinol.* **1993**, *101*, 17–29.  
 (20) Clauser, E.; Leconte, I.; Auzan, C. *Hormone Res.* **1992**, *38*, 5–12.  
 (21) Brange, J.; Owens, D. R.; Kang, S.; Volund, A. *Diabetes Care* **1990**, *13*, 923–954.  
 (22) Brange, J.; Langkjoer, L. *Pharm. Biotechnol.* **1993**, *5*, 315–350.  
 (23) Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D. *Adv. Protein Chem.* **1972**, *26*, 279–402.



**Figure 1.** Primary sequences of bovine and human insulins. The three residues that differ in human insulin are listed in parentheses below the sequence of bovine insulin. The three disulfide bonds are indicated by lines. The two  $\alpha$ -amino groups ( $G^\alpha$  and  $F^\alpha$ ) and the  $\epsilon$ -amino group ( $K^\epsilon$ ) are indicated in boldface type. Other charged residues (e.g., His (H), Glu (E), Arg (R), Tyr (Y), and C-terminal Asn (N) and Ala (A)) are underlined;  $\text{pK}_a$  values for these residues were either determined in this work (bold) or assigned reasonable values (bottom right). The symbols o and + are used to represent acetylated and unmodified amino groups, respectively. Insulin and its seven derivatives are characterized by three such symbols, each representing acetylation at one of the three residues: Gly, Phe, and Lys.

$$\theta = \frac{[\text{RNH}_2]}{[\text{RNH}_2] + [\text{RNH}_3^+]} = \frac{1}{10^{\text{pK}_a - \text{pH}} + 1} \quad (3)$$

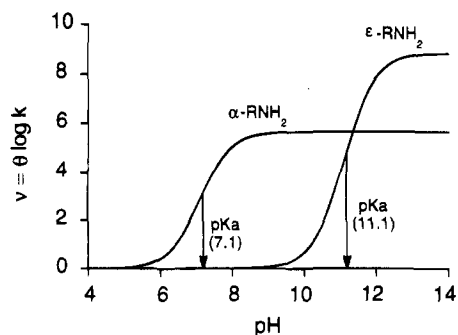
$$\nu = \theta \log k = \frac{\rho \text{pK}_a}{10^{\text{pK}_a - \text{pH}} + 1} \quad (4)$$

expresses this fraction,  $\theta$ , as a function of the values of  $\text{pK}_a$  of the ammonium group and pH of solution. Manipulating  $\theta$  makes it possible to control the relative rates of acylation ( $\nu$ ) of different amino groups (eq 4;  $\rho$  is a parameter that correlates basicity with nucleophilicity and is assumed to be 0.8 in the simulation<sup>24</sup>). At a high pH, for example, both  $\alpha$ - and  $\epsilon$ -amino groups are predominantly in the neutral forms ( $\theta \sim 1$ ), and the more nucleophilic  $\epsilon$ -amino groups can be acylated more rapidly (Figure 2). At pH values between the  $\text{pK}_a$  values of the two ammonium groups, the fraction of  $\alpha$ -amino groups in the neutral form is larger than that of  $\epsilon$ -amino groups ( $\theta_\alpha > \theta_\epsilon$ ) and the  $\alpha$ -amino groups can be acylated selectively. In this study, we used CE to follow the acylation of the three amino groups of insulin with acetic anhydride at different values of pH.

## EXPERIMENTAL SECTION

**CE Equipment.** Isco Model 3140 and Beckman P/ACE system 5500 capillary electrophoresis systems were used in these studies. The capillary tubing (Polymicro Technologies, Inc., Phoenix, AZ) was of uncoated fused silica with an internal diameter of 50  $\mu\text{m}$ . The conditions used for all CE experiments were as follows: voltage, 30 kV; current, 8–90  $\mu\text{A}$  (depending on the ion composition and pH of the buffer); detection, 200 nm; temperature,  $25 \pm 2$   $^\circ\text{C}$ . Samples (8 nL) were introduced into the capillary by vacuum injection. The capillary was flushed with 0.1 N sodium hydroxide for 5 min, water for 5 min, and buffer for 5 min before each experiment.

- (24) Jencks, W. P.; Carriuolo, J. J. *Am. Chem. Soc.* **1960**, *82*, 1778–1786.



**Figure 2.** Rates of acylation of  $\alpha$ - and  $\epsilon$ -amino groups of insulin as a function of the value of pH of the reaction medium (simulated using eq 4). An  $\alpha$ -ammonium group ( $pK_a^\alpha = 7.1$ ,  $\rho = 0.8$ ) can be acylated selectively at pH 8, and an  $\epsilon$ -ammonium group ( $pK_a^\epsilon = 11.1$ ,  $\rho = 0.8$ ) can be acylated selectively at pH 12.

**Reagents.** All chemicals were analytical grade and were used as received. Acetic anhydride and dioxane were purchased from Mallinckrodt. 1,2-Cyclohexanedinitrilotetraacetic acid (CDTA) and zinc sulfate heptahydrate were purchased from Fluka. Bovine pancreatic and human recombinant insulins were purchased from Sigma. Mesityl oxide was purchased from Eastman Organic Chemical (Rochester, NY). Stock solutions of insulin (1.5 mg/mL, 250  $\mu$ M) were prepared by dissolving the lyophilized protein in distilled water.

**Conditions for Studies of Zn(II)-Dependent Aggregation.** Solutions of insulin (25  $\mu$ M) were prepared in buffers containing tris (25 mM), Gly (192 mM), and either CDTA (20, 200, and 2000  $\mu$ M) or  $ZnSO_4$  (20, 50, 100, 200, 500, 1000, and 2000  $\mu$ M) (pH 8.4). Mesityl oxide (MO; 18 mM) was added as a neutral marker, and soybean trypsin inhibitor (50  $\mu$ M) was added as a protein marker. Each sample was analyzed using identical electrophoresis buffers.

**Conditions for Estimation of Dimerization Constants.** Electrophoresis buffers were prepared containing tris (25 mM), Gly (192 mM), and 0, 5, 10, or 20% dioxane (pH 8.4). Solutions of bovine pancreatic insulin (2, 5, 10, 20, 40, 80, 160, 300, and 400  $\mu$ M) were prepared in these buffers with MO (18 mM) added as a neutral marker and soybean trypsin inhibitor (50  $\mu$ M) added as a protein marker.

**Conditions for Acetylation Reactions.** Insulin was acetylated under three different sets of experimental conditions. In the first, 5, 10, and 20 equiv of acetic anhydride (10 mM in dioxane) were added separately to aliquots of a stock solution of insulin (0.5 mL, pH 6–7). After 30 min at room temperature, 10  $\mu$ L of each sample was diluted with 100  $\mu$ L of the electrophoresis buffer. MO was added (18 mM), and the samples were analyzed using CE.

In the second experimental protocol, the pH's of aliquots of a solution of insulin (0.5 mL) were adjusted to pH 12 with 1 N sodium hydroxide. Acetic anhydride (2, 5, or 10 equiv, 10 mM stock solution in dioxane) was added. After 5 min at room temperature, 10  $\mu$ L of a sample was diluted with 100  $\mu$ L of the electrophoresis buffer, MO was added (18 mM), and the samples were analyzed by CE.

In the third protocol, the pH of the solution of the products obtained using the first protocol (with 20 equiv of acetic anhydride) was adjusted to pH 12 using 1 N sodium hydroxide. Acetic anhydride (2, 4, 8, or 15 equiv) was added, and after five min, the samples were diluted with electrophoresis buffer. MO was added (18 mM), and the samples were analyzed by CE.

**Conditions Used for the pH Titration of the Charge Ladder of Insulin.** A solution of insulin and its seven acylated derivatives was prepared at a total concentration of 25  $\mu$ M in electrophoresis buffer (pH 8.4) containing tris (25 mM), Gly (192 mM), and MO (18 mM) added as a neutral marker. Several electrophoresis buffers were prepared having different pH values (9.4, 9.0, 8.7, 8.0, 7.7, 7.4, 7.0, 6.7, 6.3, 6.0, and 5.6) by adding 2 N acetic acid or 1 N sodium hydroxide, as necessary, to the original tris–Gly buffer. The sample was analyzed in each electrophoresis buffer using CE.

## RESULTS AND DISCUSSION

**Nomenclature.** Treatment of insulin with acetic anhydride affords a family of seven derivatives that differ in the site of acetylation of the three primary amino groups. The imidazole rings of the two histidine residues can also be acetylated under these conditions. The resulting *N*-acylimidazoles, however, are not stable under basic conditions and are not observed in these experiments.<sup>25</sup> We use the symbols + and o to represent unmodified and acetylated amino groups, respectively (Figure 1). The modified insulins are characterized by three such symbols, each representing acetylation at one of the three residues;  $G^\alpha$ ,  $F^\alpha$ , and  $K^\epsilon$ . Using this nomenclature, for example, the native insulin is labeled +++, the fully modified insulin is labeled ooo, and the derivative with only  $F^\alpha$  acetylated is labeled +o+ (Figure 1). We employed an electrophoresis buffer containing tris (25 mM) and Gly (192 mM) for all experiments described here: only the pH of the buffer differed.

**Aggregation of the Insulin Monomer.** Insulin dimerizes in aqueous solution, and the dimers can associate further to form tetrameric and hexameric aggregates.<sup>22,23</sup> The association constants for these processes in water depend on the pH, ionic composition and structure, and temperature of the solution: reported values for the dimerization constant ( $K_D = [I_D]/[I_M]^2$ ;  $I_D$  and  $I_M$  are insulin dimer and monomer, respectively) are in the range of  $10^4$ – $10^6$   $M^{-1}$ ; values of the association constants for formation of the tetramer from two dimers in the absence of Zn(II) are in the range of 20–100  $M^{-1}$ ; values for the association constant for formation of the hexamer from the dimer and the tetramer in the absence of Zn(II) are in the range of 200–1000  $M^{-1}$ .<sup>26–30</sup> Because electrophoretic mobility scales linearly with charge and nonlinearly with mass, the mobilities of the aggregates increase with molecular weight (eq 2); we could therefore use CE to establish the relevant state of aggregation of insulin under the conditions of our experiments.

**Aggregation Dependent on Zn(II).** Formation of the tetramer and hexamer of insulin (but not dimer) is enhanced in the presence of Zn(II).<sup>22,23</sup> We measured the electrophoretic mobility of insulin with Zn(II) present in the sample and electrophoresis buffer ( $[Zn(II)]$  ranged between 20  $\mu$ M and 2 mM). The electrophoretic mobility of insulin was constant over this range of concentrations of Zn(II); these data suggested that the state of aggregation of insulin did not change with increasing concentrations of Zn(II). Since bovine insulin obtained from Sigma contains

(25) Fife, T. H. *Acc. Chem. Res.* **1993**, *26*, 325–331.

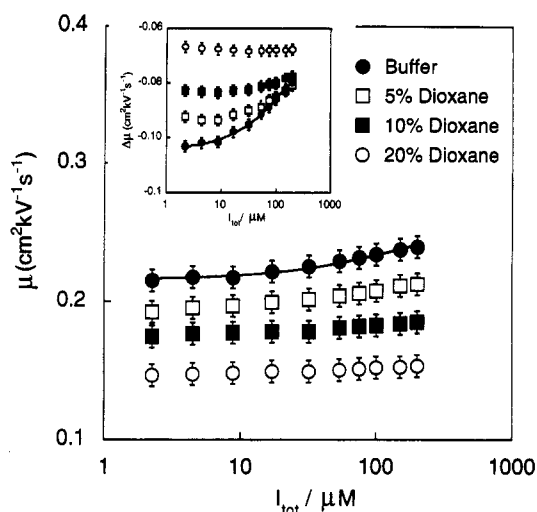
(26) Roy, M.; Lee, R. W. K.; Brange, J.; Dunn, M. F. *J. Biol. Chem.* **1990**, *265*, 5448–5452.

(27) Strazza, S.; Hunter, H.; Walker, E.; Darnall, D. W. *Arch. Biochem. Biophys.* **1985**, *238*, 30–42.

(28) Pocker, Y.; Biswas, S. B. *Biochemistry* **1981**, *20*, 4354–4361.

(29) Jeffrey, P. D. *Biochemistry* **1974**, *13*, 4441–4447.

(30) Carpenter, F. H.; Goldman, J. *Biochemistry* **1974**, *13*, 4566–4574.



**Figure 3.** Electrophoretic mobility of bovine insulin as a function of the concentration of insulin, in several different buffers: (●) 25 mM tris–192 mM Gly, pH 8.4 (A); (□) 5% dioxane in buffer A; (■) 10% dioxane in buffer A; (○) 20% dioxane in buffer A. Soybean trypsin inhibitor was used as a protein marker to correct for changes in electroosmotic flow; the inset shows plots of corrected mobilities after subtracting the mobility of STI. A least-squares fit of the uncorrected  $[\mu, I_{\text{tot}}]$  data to eq 6 gave a value of  $K_D = 6(\pm 1) \times 10^3 \text{ M}^{-1}$  in buffer A.

0.5% Zn(II) by weight (0.5 equiv of Zn(II) per insulin monomer), there was the possibility that the insulin was present exclusively as the hexamer. We therefore examined the mobility of insulin using an electrophoresis buffer containing CDTA (20  $\mu\text{M}$  to 2 mM); CDTA coordinates Zn(II) with an association constant of  $10^{18} \text{ M}^{-1}$ . The electrophoretic mobility of insulin did not change with the concentration of CDTA. Since the aggregation of insulin was not dependent on the concentration of Zn(II), and it is unlikely that insulin at a concentration of 25  $\mu\text{M}$  was present exclusively as a tetramer or hexamer, we concluded that insulin was present as a monomer or dimer.

**Dimerization of Insulin.** We measured the electrophoretic mobility of insulin as a function of its concentration over the range from 2  $\mu\text{M}$  to 400  $\mu\text{M}$ . Soybean trypsin inhibitor (STI) was added to each sample as a marker to make it possible to correct for changes in electroosmotic flow. The electrophoretic mobility of insulin increased with the concentration of insulin in the sample (Figure 3), while the mobility of STI remained constant. If we assume that monomeric and dimeric forms of insulin exist at equilibrium and in rapid exchange in the migrating plug, the increase in mobility of insulin is consistent with a larger fraction of dimeric insulin in the migrating plug at higher concentrations of insulin. Fredericq reported that dioxane, when added to an aqueous solution of insulin, suppresses dimerization of the protein.<sup>31</sup> We repeated the experiment that measured the concentration dependence using electrophoresis buffers containing 5, 10, and 20% dioxane. With the buffer containing 20% dioxane, the electrophoretic mobility of insulin remained constant over the entire range of concentrations examined: the dimerization of insulin was completely inhibited in this buffer. The inhibition of dimer formation is consistent with the observation that solvents having low dielectric constants destabilize the hydrophobic interactions between the two monomers.<sup>23,31</sup> Buffers

containing 5 and 10% dioxane were less effective at preventing the dimerization of insulin. These data, together with the observation that the electrophoretic mobility of the STI remained constant throughout all experiments, suggested that the changes in electrophoretic mobility of insulin with concentration were due to dimerization.

**Measurement of the Dimerization Constant.** We assume that the measured electrophoretic mobility of insulin ( $\mu$ ) is a weighted average of the electrophoretic mobilities of the monomer ( $\mu_M$ ) and dimer ( $\mu_D$ ) ( $\mu = \mu_M(1 - \theta_D) + \mu_D\theta_D$ ), where  $\theta_D$  is the mole fraction of dimer (eq 5). Equation 6 relates the measured

$$\theta_D = \frac{2[I_D]}{[I_{\text{tot}}]} = 1 + \frac{1 - \sqrt{8K_D[I_{\text{tot}}] + 1}}{4K_D[I_{\text{tot}}]} \quad (5)$$

$$\mu = (\mu_D - \mu_M) \left( 1 + \frac{1 - \sqrt{8K_D[I_{\text{tot}}] + 1}}{4K_D[I_{\text{tot}}]} \right) + \mu_M \quad (6)$$

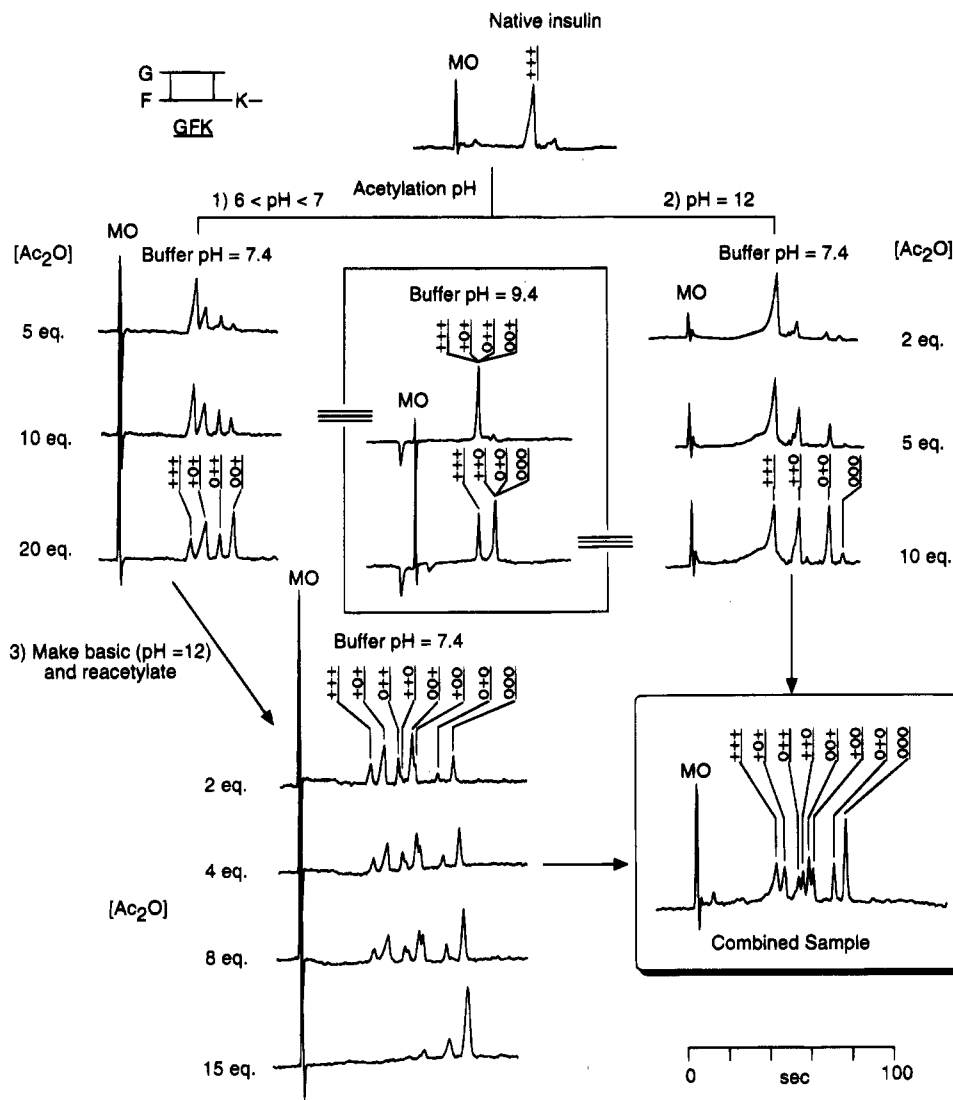
mobility,  $\mu$ , to  $\mu_M$ , to  $\mu_D$ , to the dimerization constant ( $K_D$ ), and to the concentration of insulin in the sample ( $I_{\text{tot}}$ ). A least-squares fit of the  $[\mu, I_{\text{tot}}]$  data to eq 6 using  $\mu_D$ ,  $\mu_M$ , and  $K_D$  as adjustable parameters gave values of  $K_D$  of  $6(\pm 1) \times 10^3$  and  $4.7(\pm 0.5) \times 10^2 \text{ M}^{-1}$  at pH values of 8.4 and 10.0, respectively (25 mM tris, 192 mM Gly) (Figure 3). These dimerization constants are lower limits since the concentration of insulin in the capillary is reduced relative to that in the sample vial due to broadening of the analyte zone. The best-fit values of  $\mu_D$  were a factor of 1.22 (at pH 8.4) and 1.14 (at pH 10) greater than  $\mu_M$ , in good agreement with the expected value of 1.25 if the coefficient  $C_{\text{eff}}$  remains constant in eq 2 (where  $\alpha = 2/3$ ). We were unable to measure the dimerization constant at pH 6.0 because of the low solubility of insulin at this value of pH (the  $pI$  of insulin is  $\sim 5.5$ ; see below). The dimerization constants for insulin in dioxane–water buffers were too low to be measured.

The CE experiments described in subsequent sections used solutions of insulin and its derivatives at a total concentration of 25  $\mu\text{M}$  and analyzed in an aqueous electrophoresis buffer containing tris (25 mM) and Gly (192 mM). We believe insulin was present predominantly as a monomer at these conditions.

**Selective Acetylation of  $\alpha$ - and  $\epsilon$ -Amino Groups of Insulin.** CE of pure insulin, with MO as a neutral marker to establish the rate of electroosmotic flow, in electrophoresis buffer (pH 7.4) showed a single peak (Figure 4). This sample was divided and acetylated in buffers having two different pH values. In the first, the pH of aqueous solutions of insulin was adjusted to values between 6 and 7 with 1 N sodium hydroxide and different amounts of acetic anhydride (10 mM in dioxane) were added to aliquots of the insulin-containing solutions (no other salts were added). After 30 min, the samples were diluted with electrophoresis buffer and analyzed by CE. In the second set of experiments, the pH of insulin-containing solutions was adjusted to 12 with 1 N sodium hydroxide and different amounts of acetic anhydride were added to the samples. After 5 min, the solutions were diluted with electrophoresis buffer and analyzed by CE.

**Identity of the Products of the Acetylation Reaction at Low pH.** When acetylations were carried out at the lower pH, four peaks were observed in the electropherogram (buffer pH 7.4); these peaks indicated that at least three different derivatives of insulin had been formed (Figure 4). Electrophoresis of the same

(31) Fredericq, E. *J. Am. Chem. Soc.* **1957**, *79*, 599–601.



**Figure 4.** Selective acetylation of the three amino groups of bovine insulin. Electropherograms are shown for families of acetylated insulins that were generated using acetic anhydride (at three or four different concentrations) under three different sets of conditions: (1) selective acetylation of  $G^\alpha$  and  $F^\alpha$  at a pH between 6 and 7 (top left); (2) selective acetylation of  $K^\epsilon$  and  $G^\alpha$  at a pH of 12 (top right); and (3) basification of the solution resulting from reaction 1 to a pH of 12 and subsequent acetylation (bottom left). Analysis of the reaction products from conditions 1 and 2 using an electrophoresis buffer adjusted to a pH 9.4 identifies modified peptides that have  $K^\epsilon$  acetylated (top middle). Combining the reaction products from conditions 2 and 3 affords a mixture containing insulin and its seven derivatives (bottom right). Samples were analyzed in 25 mM tris–192 mM Gly buffer (pH 7.4 or 9.4). MO was used as a neutral internal standard to establish the rate of electroosmotic flow: its migration time varies with pH and buffer but is typically in the range of 120–140 s. The time scale (lower right) applies to all electropherograms.

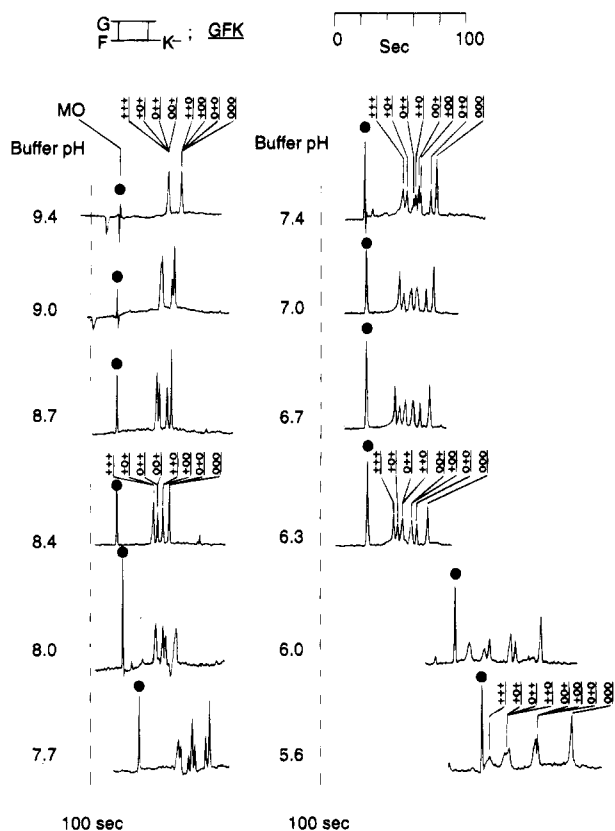
sample in a buffer at pH 9.4 showed a single peak. At this pH, the  $\alpha$ -amino groups ( $G^\alpha$  and  $F^\alpha$ ) were not protonated, and differences in charge, and hence, electrophoretic mobility, arose only from acetylation of  $K^\epsilon$ . These four species (at pH 7.4) were thus native insulin (+++), two derivatives that were acetylated at either  $G^\alpha$  or  $F^\alpha$  (o++ or +o+), and one derivative that was acetylated at both  $\alpha$ -amino groups (oo+). A co-injection experiment verified that the first peak was the native insulin; since the four species should migrate with mobilities proportional to their effective charge, the last peak was assigned as the oo+ derivative. Assignments of the peaks for the two singly modified derivatives were made on the basis of the fact that the value of the  $pK_a$  of  $G^\alpha$  is higher than that of  $F^\alpha$ .<sup>32</sup>

**Identity of the Products of the Acetylation Reaction at High pH.** Electrophoresis of the reaction products obtained from acetylation of insulin at a pH of 12 showed three major peaks and

one minor peak (Figure 4). Electrophoresis of these four derivatives in a buffer at a pH of 9.4 showed two peaks: the mixture therefore comprised at least one derivative containing an unmodified  $K^\epsilon$  group and at least one derivative containing an acetylated  $K^\epsilon$  group. The four species were assigned as +++, ++o, o+o, and ooo. The identities of native insulin (+++) and the peracetylated derivative (ooo) were proved by co-injection experiments (for assignment of ooo, see the next paragraph). The assignments of the ++o and o+o derivatives were supported by co-injection experiments with a sample containing all of the acetylated derivatives of insulin (described in subsequent sections).

**Preparation of a Sample Containing Insulin and its Seven Acetylated Derivatives.** Solutions of products from the acetylation reaction at a pH of 6–7 were adjusted to pH 12 using 1 N sodium hydroxide and treated again with different amounts of acetic anhydride. After 5 min, an equal volume of a solution

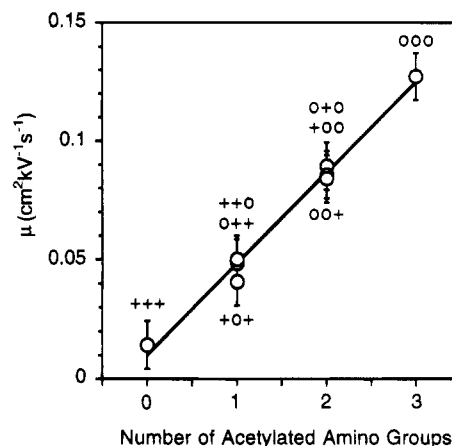
(32) Hefford, M. A.; Oda, G.; Kaplan, H. *Biochem. J.* 1986, 237, 663–668.



**Figure 5.** Electropherograms of a mixture of bovine insulin and its seven acetylated derivatives at several pH values of the electrophoresis buffer (25 mM tris and 192 mM Gly). The time scale applies to all experiments, and the vertical dashed lines provide a common reference point for direct comparison of absolute migration times. The peaks for the neutral marker (MO) are indicated by filled circles.

resulting from the acetylation reaction at the high pH value was added and the mixture was analyzed by CE. All eight derivatives were resolved using an electrophoresis buffer at a pH of 7.4 (Figure 4). The size of the peak with the slowest migration time increased with the amount of acetic anhydride in the reaction and was assigned as the peracetylated derivative (ooo). The identities of the other peaks were assigned by comparison with the mixtures prepared by the procedures described in previous sections: these assignments were further supported by pH titration experiments described in the next paragraph.

**Electrophoretic Mobilities of Acetylated Derivatives of Insulin in Buffers with Different Values of pH.** The electrophoretic mobilities of insulin and its acetylated derivatives depend on the pH of the buffer (Figure 5). Analysis of a mixture of the eight species in a buffer at pH 9.4 showed two peaks: each peak corresponds to the families of derivatives containing either an acetylated or unmodified (protonated) K<sup>ε</sup> group. When the pH of the buffer was lowered to 8.7, two new peaks emerged from the two parent peaks: these peaks correspond to derivatives that contain an unmodified G<sup>α</sup> and were resolved due to partial protonation of G<sup>α</sup>. At pH's of 7.7 and 7.4, four additional peaks appeared in the electropherogram due to protonation of F<sup>α</sup>; all eight derivatives of insulin were resolved by CE at this pH. Four groups of peaks were observed using a buffer at a pH of 5.6, where each group contained derivatives with an equal number of modified amino groups. At this low pH, acetylation of each amino group results in a change in the charge of the protein by  $\sim -1$ . A plot relating the electrophoretic mobility of each derivative to the



**Figure 6.** Relationship between electrophoretic mobility and the number of acetylated amino groups for insulin and its seven derivatives at a pH of 5.6. The electrophoretic mobilities of insulin and its acetylated derivatives are related linearly to the number of NH<sub>3</sub><sup>+</sup> groups.

number of modified amino groups gives a straight line: this correlation demonstrates that the differences in electrophoretic mobility of insulin and its acetylated derivatives depend only on differences in charge (Figure 6).

#### Determination of Values of pK<sub>a</sub> of G<sup>α</sup> and F<sup>α</sup> of Insulin.

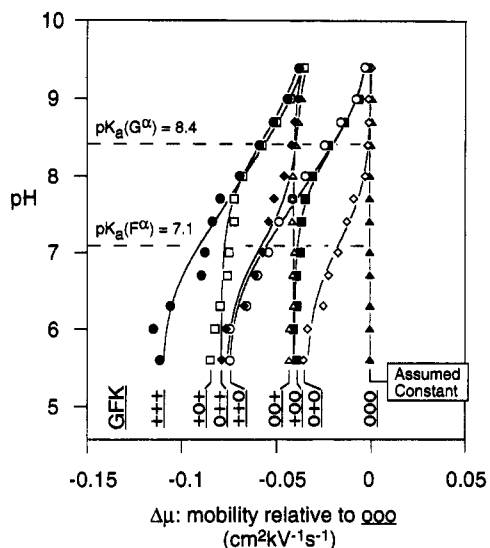
We used CE to estimate the ionization constants of the α-ammonium groups of insulin.<sup>33</sup> The electrophoretic mobility ( $\mu$ ) of a protonated base (BH<sup>+</sup>) containing no other acidic protons is related to the electrophoretic mobilities and mole fractions of the charged, protonated ( $\mu_{\text{BH}^+}$ ,  $\theta_{\text{BH}^+}$ ) and the neutral, unprotonated ( $\mu_{\text{B}}, 1 - \theta_{\text{BH}^+}$ ) forms by eq 7. Rearrangement of eqs 3 and 7 gives

$$\mu = (1 - \theta_{\text{BH}^+})\mu_{\text{B}} + \theta_{\text{BH}^+}\mu_{\text{BH}^+} \quad (7)$$

$$\Delta\mu = \mu - \mu_{\text{B}} = (\mu_{\text{BH}^+} - \mu_{\text{B}}) \frac{1}{1 + 10^{\text{pH} - \text{pK}_a}} \quad (8)$$

eq 8, which relates the change in electrophoretic mobility of a species resulting from protonation of its basic residue to  $\mu_{\text{BH}^+}$ ,  $\mu_{\text{B}}$ , the pH of the electrophoresis buffer, and the pK<sub>a</sub> of the ionizable residue. The analysis based on eq 8, however, cannot be applied directly to the determination of pK<sub>a</sub> of a unique ammonium group of insulin because ionization of several other residues also contributes to changes in the charge. We analyzed the electrophoretic mobilities of the +oo and o+o derivatives to determine the pK<sub>a</sub> values of G<sup>α</sup> and F<sup>α</sup>, respectively; the two other amino groups in each species are acetylated and do not contribute to changes in total charge as the pH is changed. The effects of other ionizable residues (e.g., His) were eliminated by comparing the electrophoretic mobilities of the +oo and o+o derivatives relative to that of the peracetylated derivative ( $\Delta\mu = \mu - \mu_{\text{ooo}}$ ): in this way, only the extent of protonation of either G<sup>α</sup> or F<sup>α</sup> contributed to changes in effective charge and, hence, to  $\Delta\mu$ . Plots of  $\Delta\mu$  vs pH for insulin and its seven derivatives are shown in Figure 7. Equation 8 was fitted to the [ $\Delta\mu$ , pH] data for the +oo and o+o derivatives using values of ( $\mu_{\text{BH}^+} - \mu_{\text{B}}$ ) and pK<sub>a</sub>'s of G<sup>α</sup> and F<sup>α</sup>, respectively, as adjustable parameters: the least-squares best-fit values of pK<sub>a</sub> were 8.4(±0.1) for G<sup>α</sup> and 7.1(±0.1) for F<sup>α</sup>.

(33) For estimation of values of pK<sub>a</sub> of simple organic acids, see: Smith, S. C.; Khaledi, M. G. *Anal. Chem.* **1993**, *65*, 193–198.



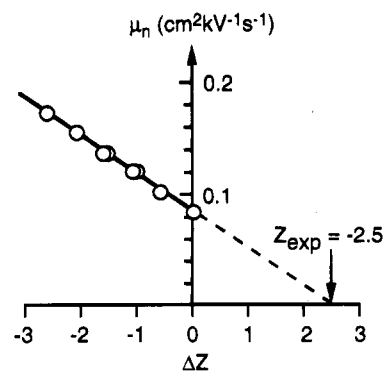
**Figure 7.** Dependence of  $\Delta\mu$  on the pH of the electrophoresis buffer for insulin and its seven derivatives. Values of  $\mu$  for each derivative were computed from the electropherograms in Figure 5 using eq 1. Values of  $\Delta\mu$  were obtained by subtracting the value of  $\mu(\text{ooo})$  from the value of  $\mu$  measured for each derivative: the mobility of ooo is assumed constant. The data are represented as follows: +++ (●), +o+ (□), o++ (◆), ++o (○), oo+ (△), +oo (■), o+o (◇), and ooo (▲). Overlapping points are offset slightly for clarity. The curves superimposed on the data for the +oo and o+o derivatives were obtained by performing nonlinear least-squares fits using eq 8; these fits yielded  $pK_a$  values of  $8.4(\pm 0.1)$  for  $G^\alpha$  and  $7.1(\pm 0.1)$  for  $F^\alpha$ . The other curves were simulated on the basis of these values of  $pK_a$  and a value of 11.1 for  $K^\epsilon$  (see text).

Using these  $pK_a$  values, and a value of 11.1 for  $K^\epsilon$ ,<sup>34</sup> the dependence of  $\Delta\mu$  on pH was simulated for insulin and its five other derivatives using eq 8. The contribution to  $\Delta\mu$  from  $G^\alpha$ ,  $F^\alpha$ , and  $K^\epsilon$  at each pH was determined by calculating the fractional protonation of each amino group using eq 3. The partial charges were summed and fitted to the experimental data using eq 8 and  $(\mu_{\text{BH}^+} - \mu_{\text{B}})$  as an adjustable parameter. These simulated curves agree well with the experimental data (Figure 7).

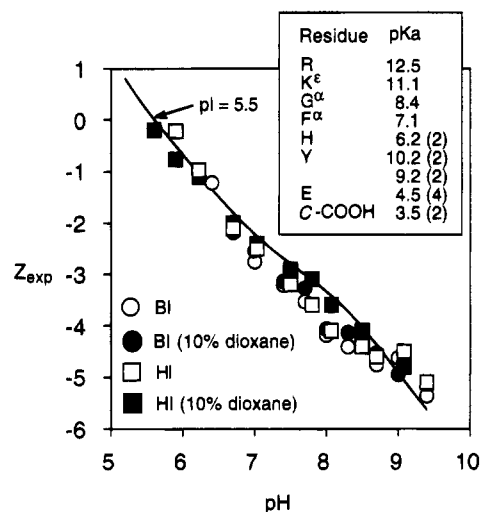
**Influence of pH on the Total Charge of Insulin.** We described previously a method to measure the total charge of a protein ( $Z_{\text{exp}}$ ) using CE by analyzing a protein charge ladder—a series of derivatives of a protein that differ by known increments of charge but differ only minimally in hydrodynamic drag—generated by acylation of the  $\epsilon$ -amino groups of Lys residues of the protein.<sup>9,35</sup> The method is based on the relationship between the electrophoretic mobilities of peaks in a charge ladder ( $\mu_n$ ) and their charges relative to the unmodified protein ( $\Delta Z_P = Z_P(n) - Z_P(0)$ , where  $n$  indicates the number of modified amino groups, eq 9). A plot of the electrophoretic mobilities of peaks

$$\mu_n - (C_P/M^\alpha)Z_P(0) = (C_P/M^\alpha)\Delta Z_P \quad (9)$$

( $\mu_n$ ) in the charge ladder vs the values of  $\Delta Z_P$  gives a linear correlation where the intercept at the  $x$ -axis (where  $\mu_n = 0$ ) is equal to the total charge of the native protein (eq 9). Using this analysis, we determined values of  $Z_{\text{exp}}$  of insulin at several values of pH of the buffer (Figure 8). The measured values of  $C_P/M^\alpha$



**Figure 8.** Determination of the total charge  $Z_{\text{exp}}$  of insulin at pH 7.0 using eq 9. The calculation of increments of charge ( $\Delta Z_P$ ) is described in the text. Extrapolation of the best-fit line to  $\mu_n = 0$  gives a value of  $-2.5$  for the charge of insulin at a pH of 7.0 (regression coefficient  $r = 0.999$ ).



**Figure 9.** Influence of pH on the total charge of insulin. The measured charge of insulin ( $Z_{\text{exp}}$ ) at each pH was determined by analyzing the charge ladder from bovine or human insulin using different electrophoresis buffers: bovine (○) and human (□) insulin in tris-Gly buffer, and bovine (●) and human (■) insulin in tris-Gly buffer containing 10% dioxane. The curve represents calculated values of charge of insulin ( $Z_{\text{calc}}$ ) as a function of pH. The inset shows the values of  $pK_a$  of the ionizable residues of insulin used to calculate  $Z_{\text{calc}}$ .

(e.g., at pH 7.0, from the slope in Figure 8) were constant at different pH's in this experiment:  $C_P/M^\alpha = 0.031 \pm 0.001 \text{ cm}^2 \text{ s}^{-1} \text{ kV}^{-1} \text{ charge}^{-1}$ . Assuming  $\alpha = 2/3$ , and using MW = 5700 for insulin,  $C_P$  is equal to  $9.9 \text{ cm}^2 \text{ s}^{-1} \text{ kV}^{-1} \text{ charge}^{-1} \text{ D}^{2/3}$ . We analyzed charge ladders generated from both bovine and human insulins using either an aqueous electrophoresis buffer or an aqueous buffer containing 10% dioxane.

Figure 9 shows plots of the measured total charge of insulin vs the pH of the buffer. The curve that is superimposed on the data was generated by calculating the charge of insulin as a function of pH. The charge carried by each ionizable residue was calculated using eq 3, the pH of the buffer, and its  $pK_a$  shown in Figure 9: these charges were added to yield the total charge of insulin ( $Z_{\text{calc}}$ ). Extrapolation of the data from the four experiments to  $Z_{\text{exp}} = 0$  provides an estimate of  $5.5 \pm 0.2$  for the  $pI$  of insulin; this value is in good agreement with the literature value of 5.3.<sup>36</sup>

Although the experimental data agree well with the simulated curve, there remain slight differences in the range of pH from 7

(34) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285–297.

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

(36) Tanford, C.; Epstein, J. J. *Am. Chem. Soc.* **1954**, *76*, 2163–2169.

to 9. In principle, the  $pK_a$  values for the ionizable residues used in the simulation may have been inaccurate or altered upon acetylation of amino groups. Nonetheless, this method to measure the electrostatic properties of insulin provides data that are well rationalized by its known structural characteristics.

## CONCLUSIONS

Insulin is an excellent model system with which to study the influence of chemical modification of charged groups on the electrophoretic mobility of proteins. At pH 7.4, derivatives of insulin with 0.1 unit charge difference (o++ and ++o, oo+ and +oo) were well separated. For larger proteins, the resolution of partial charges will be much more difficult or impossible. Since the Lys ammonium groups of large proteins have similar, high values of  $pK_a$ , their selective modification<sup>9</sup> will only give derivatives that differ in full units of charge. Use of polycharged modifying agents to increase the increments in mobility in a protein charge ladder makes it easier to measure the charges of high molecular weight proteins using capillary electrophoresis, but the resolution in these systems will be less than that with insulin.

These studies demonstrate that CE is useful for the characterization of electrostatic properties of proteins in their native conformations. By analyzing a family of derivatives of a protein generated by acylation of its amino groups, partial charges of individual residues (and their  $pK_a$  values) and the total charge of the protein can be determined. The unique sensitivity of CE to the charge of a protein makes it appropriate for use in studies that involve protein electrostatics: for example, the influence of the charge of a protein on the binding of charged ligands to it.

## ACKNOWLEDGMENT

This work was supported by the NIH (Grants GM 51559 and GM 30367). M.M. is grateful to the American Cancer Society for a postdoctoral fellowship and F.A.G. is grateful to the Damon Runyon–Walter Winchell Cancer Research Fund for a postdoctoral fellowship. We also thank Mathai Mammen for helpful discussions.

## GLOSSARY

Frequently used abbreviations and symbols are listed in the order of appearance

CE	capillary electrophoresis
$\mu$	electrophoretic mobility
$t_{eo}$	migration time of neutral marker
MO	mesityl oxide, as neutral marker
$Z_{eff}$	effective charge
$Z_P$	unperturbed charge
$C_{eff}$ and $C_P$	proportionality constants for $Z_{eff}$ and $Z_P$
$Z_{calc}$	calculated charge
$G^\alpha$	$\alpha$ -amino group from N-terminal Gly of A chain
$F^\alpha$	$\alpha$ -amino group from N-terminal Phe of B chain
$K^\epsilon$	$\epsilon$ -amino group from Lys of B chain
$\theta$	molar fraction
$\nu$	rate of acylation
$K_D$	dimerization constant
$I_D$	dimer of insulin
$I_M$	monomer of insulin
$\mu_D$	electrophoretic mobility of dimer of insulin
$\mu_M$	electrophoretic mobility of monomer of insulin
+++	native insulin
o++, +o+, ++o	monoacetylated insulin (see text for details)
+oo, o+o, oo+	diacetylated insulin (see text for details)
ooo	peracetylated insulin
$Z_{exp}$	charge determined by CE

Received for review May 30, 1995. Accepted July 17, 1995.\*

AC9505177

\* Abstract published in *Advance ACS Abstracts*, August 15, 1995.