

Analysis by Capillary Electrophoresis of the Kinetics of Charge Ladder Formation for Bovine Carbonic Anhydrase

Janelle R. Anderson, Oksana Cherniavskaya, Irina Gitlin, Gregory S. Engel, Leonid Yuditsky, and George M. Whitesides*

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

A series of charge ladders of bovine carbonic anhydrase II were synthesized and the relative abundances of the rungs analyzed by capillary electrophoresis as a function of the quantity of acylating agent used. A simulation that models the kinetics of formation of the members of the charge ladders is described. The observed rate constants decreased as the extent of acylation increased. These rate constants correlated adequately with theoretical rate constants calculated using Debye–Hückel theory. The data are compatible with, but do not demand, a model for the formation of this charge ladder in which all unacetylated amino groups in each rung have indistinguishable reactivity and in which the reactivity of the amines in each rung decreases as the net charge on the protein increases; in this model, decreased reactivity is due to increased extent of protonation. This agreement between experiment and model suggests that the charge shielding that results from an ionic strength of 130 mM is not sufficient to suppress the influence of the increasingly negative charge of the protein with acetylation on the extent of protonation of Lys ϵ -NH₂ groups.

Charge ladders of bovine carbonic anhydrase II (CA; EC 4.2.1.1) are formed by acetylating some (or all) of its 18 Lys ϵ -NH₂ groups; the α -NH₂ of this protein is acylated in its native state.^{1–4} Each of the 19 possible rungs of this charge ladder is an undefined collection of regioisomeric protein derivatives having the same number of acetylated Lys groups; all 19 rungs can be separated by capillary electrophoresis (CE).^{5–7}

In synthesizing charge ladders by acetylating various proteins, we have observed that the first amino groups to be acetylated are more reactive than later amino groups and that it is diffi-

cult to obtain the peracetylated form of the protein. Plausible hypotheses to rationalize decreasing reactivity with increasing extent of acylation is that either the Lys ϵ -NH₂ groups differ in reactivity⁸ or that they have the same reactivity and this reactivity changes with degree of acylation as a result of changes in the electrostatic potential at the surface of protein and subsequent changes in the pK_a of Lys ϵ -NH₂ groups or in the value of local pH.^{5,9} It is probable that both differences in reactivity between individual Lys groups and differences in reactivity between sets of the proteins that make up the rungs are involved in the observed decrease in rate of acetylation with increasing extent of acetylation.

We have used the kinetics of formation of the different rungs of the charge ladders to establish the reactivities of the most reactive rung of the charge ladder as increasing numbers of Lys groups were acetylated. We synthesized a series of charge ladders, analyzed them by CE, and simulated their formation numerically. We then deduced the decrease in observed rate constants for acylation of each rung.

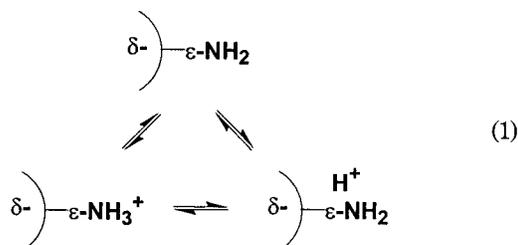
We tested the compatibility of these quantitative data with a theoretical model in which all Lys groups on the protein derivatives in a given rung, x , have the same reactivity. In this model, as the Lys groups are acetylated, the absolute value of the electrostatic potential of the protein—reflecting the contribution from all the charged groups of the protein (from Lys, Glu, Asp, His, Zn(II)·H₂O, Arg, Tyr, and the CO₂H terminus)—increases. A change in the electrostatic potential of a protein is known to affect reactivity: it may lead to a change in the strength of binding of charged ligands², changes in the conformation of CA,¹⁰ changes in the field strength during the CE experiment due to distribution of small ions in the ionic atmosphere surrounding CA,¹¹ and changes interpreted as changes in the value of pK_a of remaining Lys and other charged residues⁵ and in local pH.⁹ Of these effects, only the latter two would lead directly to changes in amino acid ionization and would be capable of decreasing the reactivity of unmodified Lys as the reactions proceed.

We emphasize that this model is an abstraction and oversimplification. Acetylation of Lys amino groups (present as an equilibrium mixture of NH₃⁺ and NH₂ groups) eliminates positive

- (1) Gao, J.; Gomez, F. A.; Härter, R.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12027–12030.
- (2) Gao, J.; Mammen, M.; Whitesides, G. M. *Science* **1996**, *272*, 535–537.
- (3) Gao, J.; Mrksich, M.; Mammen, M.; Whitesides, G. M., Eds. *Affinity Capillary Electrophoresis: Using Capillary Electrophoresis to Study the Interactions of Proteins with Ligands*, John Wiley & Sons: New York, 1998.
- (4) Gao, J.; Whitesides, G. M. *Anal. Chem.* **1997**, *69*, 575–580.
- (5) Carbeck, J. D.; Colton, I. J.; Anderson, J. R.; Deutch, J. M.; Whitesides, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 10671–10679.
- (6) Carbeck, J. D.; Colton, I. J.; Gao, J.; Whitesides, G. M. *Acc. Chem. Res.* **1998**, *31*, 343–350.
- (7) Colton, I. J.; Anderson, J. R.; Gao, J.; Chapman, R. G.; Isaacs, L.; Whitesides, G. M. *J. Am. Chem. Soc.* **1997**, *119*, 12701–12709.

- (8) Zhang, M.; Vogel, H. J. *J. Biol. Chem.* **1993**, *268*, 22420–22427.
- (9) Menon, M. K.; Zydney, A. L. *Anal. Chem.* **2000**, *72*, 5714–5717.
- (10) Aviram, I.; Myer, Y. P.; Schejter, A. *J. Biol. Chem.* **1981**, *256*, 5540–5544.
- (11) Allison, S. A.; Potter, M.; McCammon, J. A. *Biophys. J.* **1997**, *73*, 133–140.

charge (represented in eq 1 by an increase in the (negative)



surface potential δ^-). This negative charge will tend to increase the concentration of positively charged species (H_3O^+ , buffer cations) in the vicinity of the protein. The protons can be accommodated either on water (as H_3O^+) or on Lys (as $\epsilon\text{-NH}_3^+$); the increase in local concentration of H_3O^+ can be interpreted as a change in local pH; the increase in the reactive fraction of Lys $\epsilon\text{-NH}_2$ groups that are protonated (as $\epsilon\text{-NH}_3^+$) can be interpreted as an increase in pK_a . The two effects are not, however, distinct. They are linked as consequences of the change in the local electrostatic potential at the surface of the protein. This electrostatic potential is, of course, heterogeneous: that is, it varies with position on the surface of the protein.

The model we test here is the simplest possible. It assumes that the electrostatic potential is homogeneous across the surface of the protein, assumes that all Lys $\epsilon\text{-NH}_2$ groups have the same pK_a , and assumes that any change in reactivity is due to changes in this pK_a (that is, to the fraction of Lys $\epsilon\text{-NH}_2/\text{NH}_3^+$ groups that is present in deprotonated, and thus nucleophilic, form). It is thus a biophysical model of the simplest form. Our objective was to get a sense of the change in pK_a that would be required to reproduce the observed experimental data and to test this simplest model for its compatibility with the experimental data.

In this model of kinetics, the Lys $\epsilon\text{-NH}_2$ groups from all the different rungs form a family of indistinguishable nucleophiles; as such, they would follow a common Brønsted correlation (eq 2), where $k_{\text{acetylation}}$ is the rate of the nucleophilic attack that

$$\log k_{\text{acetylation}} = \beta pK_a + \log G \quad (2)$$

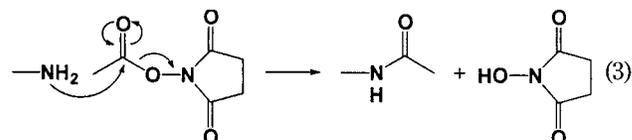
acetylates the amine, $\beta = 0.69\text{--}0.8$ ^{12,13} is a constant characteristic of this type of amine nucleophile in that reaction, and G is a constant for a particular reaction.¹⁴ We used eq 2 with the $pK_{a,\text{eff}}$ of Lys $\epsilon\text{-NH}_2$ calculated using Debye–Hückel theory and the experimental value of β from the literature to predict the decrease in the observed rate constants, $k_{x,\text{obs}}$ as a function of the number x of acetyl groups that had been added to the protein, for each pH. We found good agreement between these predicted values of $k_{x,\text{obs}}$ for each rung and those inferred experimentally.

Studies that have examined the behavior of amino acid residues in changing electrostatic environments based on charge ladders have done so using exclusively the *mobilities* of the proteins.^{2,5,9,15}

We demonstrate that we are able to address the same questions by extracting quantitative data from a different feature of the same type of experiment: the integrated areas of the rungs of charge ladders. We believe that studies of the kinetics of acetylation of reaction will, in some cases, provide an alternative to studying electrophoretic mobilities as a way of exploring the interaction between the behavior (protonation, acetylation) and the overall electrostatic potential at the surface of the protein.

RESULTS AND DISCUSSION

Generation of Charge Ladders. We added an aliquot of the *N*-hydroxysuccinimide ester of acetic acid (AcNHS) in dioxane to native CA (denoted “CA₀”) in borate buffer. The Lys $\epsilon\text{-NH}_2$ groups on CA attacked the ester and formed an amide (eq 3).



We produced 24 charge ladders by performing the acetylation in three borate buffers (pH 9.09, 8.56, and 7.40) at a constant ionic strength using eight concentrations of AcNHS (from 0.024 to 9.7 mM). We analyzed all 24 reactions by CE under conditions that produced electropherograms with almost baseline resolution (Figure 1).¹⁶ We integrated the absorbance peaks ($\lambda = 214$ nm) due to the rungs in each electropherogram to determine the area of each rung; we assume that this area is directly proportional to the concentration of its constituents.¹⁷

The time dependence of the concentration of the x th rung of the charge ladder, $[\text{CA}_x]$, is governed by the second-order rate equation:

$$\begin{aligned}
 d[\text{CA}_x]/dt = (19 - x)k_{x-1,\text{obs}}[\text{CA}_{x-1}][\text{AcNHS}] - \\
 (18 - x)k_{x,\text{obs}}[\text{CA}_x][\text{AcNHS}] \quad (4)
 \end{aligned}$$

where CA_x , from CA_0 to CA_{18} , represents all the regioisomeric CA derivatives that have x modified Lys groups. The terms $(19 - x)$ and $(18 - x)$ adjust the rates statistically for the number of remaining Lys groups (assumed to be of indistinguishable reactivity). The rate constants $k_{x,\text{obs}}$, for $k_{0,\text{obs}}\text{--}k_{17,\text{obs}}$, thus describe the reactivity of all Lys groups on all CA_x derivatives in rung x : that is, $\text{CA}_x \rightarrow \text{CA}_{x+1}$. It is an observed rate constant that is not adjusted for the fraction of Lys groups that are protonated at the pH of the buffer.

The acetylating reagent is consumed both by attack of $\alpha\text{-NH}_2$ and by hydrolysis according to eq 5, where $x = 0\text{--}17$ and k_{hyd} is the rate of hydrolysis of AcNHS, as determined experimentally by UV–visible spectroscopy for each buffer and pH.

(12) Hirata, H.; Nakasato, S. *J. Jpn. Oil Chem. Soc. (Yukagaku)* **1986**, *35*, 438–443.

(13) Jencks, W. P.; Gilchrist, M. *J. Am. Chem. Soc.* **1968**, *90*, 2622–2637.

(14) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover Publications: New York, 1969.

(15) Caravella, J. A.; Carbeck, J. D.; Duffy, D. C.; Whitesides, G. M.; Tidor, B. *J. Am. Chem. Soc.* **1999**, *121*, 4340–4347.

(16) We used 25 mM Tris–192 mM Gly buffer in D_2O , with a value of pD of 8.8, as the running buffer. The deuterium-based buffer improves separation of analytes, compared to the same water-based buffer, because its lower conductivity and higher viscosity are believed to lower Joule heating and electroosmotic flow: Okafo, G. N.; Brown, R.; Camilleri, P. *J. Chem. Soc., Chem. Commun.* **1991**, *13*, 864–866.

(17) Grossman, P. D. *Capillary Electrophoresis: Theory and Practice*; Academic Press: San Diego, CA, 1992.

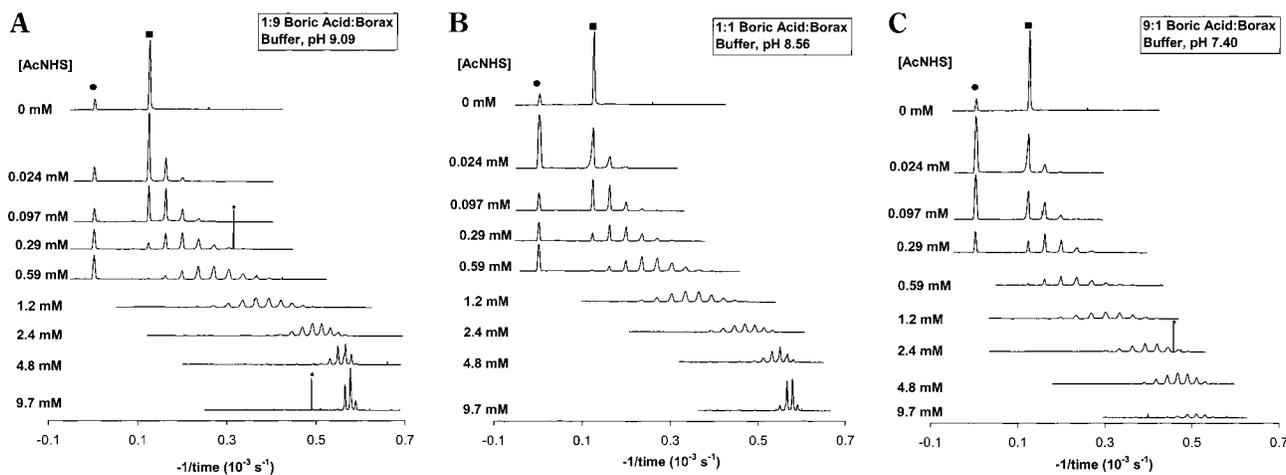


Figure 1. Complete set of raw data: electropherograms of eight reactions in three buffers. Electropherograms are presented without scaling from lowest concentration of AcNHS in the reaction mixture (0.024 mM) to highest concentration (9.7 mM) in the three borate buffers: (A) 1:9 boric acid–Borax, pH 9.09; (B) 1:1 boric acid–Borax, pH 8.56; and (C) 9:1 boric acid–Borax, pH 7.40. The filled circles (●) denote the neutral marker, *p*-methoxybenzyl alcohol, and the squares (■) denote the native protein, CA₀. Very sharp peaks marked with an asterisk (*) are air bubbles, which were not included in the integration of rungs. The reactions took place in H₂O-based borate buffers. The CE separations were done using D₂O-based 25 mM Tris–192 mM Gly buffer. All the products of the reactions with molecular weight less than 10 000 were separated by ultrafiltration prior to performing the CE experiment. We plotted the electropherograms in the inverse time domain because it is directly proportional to electrophoretic mobility (Mammen, M.; Colton, I. J.; Carbeck, J.; Bradley, R.; Whitesides, G. M. *Anal. Chem.* **1997**, *69*, 2165–2170.)

$$d[\text{AcNHS}]/dt = - \sum_x k_x [\text{CA}_x] [\text{AcNHS}] - k_{\text{hyd}} [\text{AcNHS}] \quad (5)$$

We wanted to determine how the rate constants $k_{x,\text{obs}}$ change as x goes from 0 to 17, within the context of our model. For proteins with only a few Lys groups, it is possible to use a recursive method to solve the system of coupled differential equations.¹⁸ The 20 simultaneous differential equations that describe the system (corresponding to the rates of change of the concentrations of 19 rungs of the charge ladder, plus the rate of spontaneous hydrolysis of AcNHS) cannot be solved analytically for $k_{0,\text{obs}} - k_{17,\text{obs}}$. They can, however, be solved by numerical integration.

Modeling. Our modeling of the reaction estimates the rate constants of rung formation using a brute force approach to simulations. We used the Euler–Cauchy method and made the approximation that $d[\text{CA}_x]/dt \approx \Delta[\text{CA}_x]/\Delta t$ for each of the 20 differential equations.¹⁹ One simulation of a set of eight reactions in one buffer consisted of solving for the concentrations of all 19 rungs in all eight experiments at every time step and repeating the calculations until the reagent was exhausted in all reactions. Complete simulations for two reactions are shown in Figure 2.

We performed the search for the rate constants starting from initial guesses of $k_{x,\text{obs}}$ of 0.1 and $10 \text{ M}^{-1} \text{ s}^{-1}$ for all x . From these two starting points, we employed two independent search algorithms that determined subsequent values for $k_{x,\text{obs}}$: (1) we sampled 400 000 combinations of $k_{x,\text{obs}}$, chosen by searching around the $k_{x,\text{obs}}$ that had produced the lowest error in all the previous trials; and (2) we started at the two initial values and searched along the steepest gradient to find the minimum. The

details of the methods used in these searches are described in the Supporting Information to this paper.

A function of merit for comparing the simulated concentrations of protein derivatives from a set of $k_{x,\text{obs}}$ and the experimental data is the sum of squares of residuals due to error (SSE),²⁰ defined in eq 6.

$$\sum_{i=1}^8 \sum_{x=0}^{17} ([\text{CA}]_{x,\text{expt}i} - [\text{CA}]_{x,\text{sim}i})^2 \quad (6)$$

Both search algorithms were directed toward minimum SSE.

The two search algorithms yielded the same rate constants within error, regardless of the initial guesses for $k_{x,\text{obs}}$; each pair of $k_{x,\text{obs}}$ agreed to within 2% on average. They are plotted in Figure 3. The sequence of experiments and simulations is outlined in Scheme 1.

The simulated values for the concentrations of protein derivatives, corresponding to the best rate constants found in the search, can be compared directly to the corresponding areas of the experimental charge ladders by visual inspection of histograms as those shown in Figure 4.

The SSE errors between the eight simulated and experimental charge ladders in buffers at pH 9.09, 8.56, and 7.40 were minimums at 4.3×10^{-10} , 3.4×10^{-10} , and $5.8 \times 10^{-10} \text{ M}^2$, respectively. These errors can alternatively be expressed as the average percent error, defined as the absolute values of the percent difference between the areas of each pair of simulated and experimental rungs, summed over eight reactions, and divided by the total number of rungs in eight charge ladders. For buffers at pH 9.09, 8.56, and 7.40, the average percent error was 6%, 8%, and 12%.

(18) Grzybowski, B. A.; Anderson, J. R.; Colton, I.; Brittain, S. T.; Shakhnovich, E. I.; Whitesides, G. M. *Biophys. J.* **2000**, *78*, 652–661.

(19) Kreysig, E. *Advanced Engineering Mathematics*, 7th ed.; John Wiley & Sons: New York, 1993.

(20) Ott, L. *An Introduction to Statistical Methods and Data Analysis*, 3 ed.; PWS: Kent, U.K., 1988.

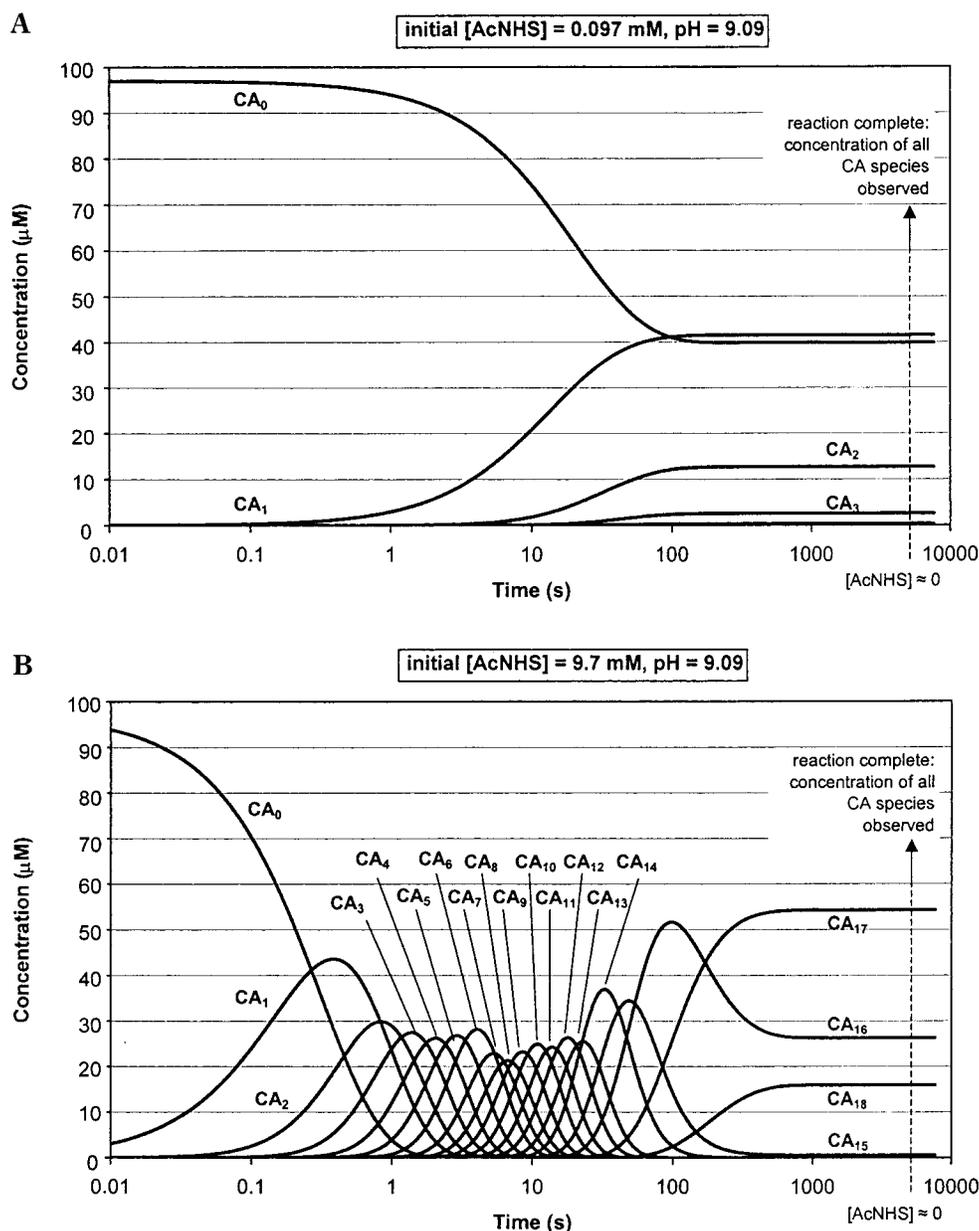


Figure 2. Simulations of two reactions showing the evolution of all rungs, CA_x as a function of time. The corresponding experiments are the charge ladders formed by (A) reacting 0.097 mM AcNHS and 0.097 mM CA_0 in 1:9 boric acid–Borax at pH 9.09, and (B) reacting 9.7 mM AcNHS and 0.097 mM CA_0 in 1:9 boric acid–Borax at pH 9.09. The simulations used a k_{hyd} of $7.9 \times 10^{-3} \text{ s}^{-1}$ for hydrolysis of AcNHS in this buffer. The concentration of all rungs are shown as they evolve over small time increments according to the rate equations described in eqs 4 and 5; the concentration of AcNHS is not shown. The set of $k_{x,\text{obs}}$ used in the calculations for these simulations were those found to produce the set of charge ladders at pH 9.09. Because some of the values of $k_{x,\text{obs}}$ deviate from a smooth line along x , the maximums of $[CA_x]$ also show some deviations from a smooth line. The dotted line indicates the point at which the AcNHS is completely consumed; the reaction stops at this point. The concentrations of CA_x at this time are the simulated charge ladder; these values are compared with the integrated peaks of the experimental charge ladder in the histogram shown in Figure 4.

Uncertainty. We believe that there are two principle sources of error in generating the matrix of experimental results. One source is human error introduced during sample preparation, which we estimate to be $\pm 5\%$ of volumes and weights measured. To offset the effect of these errors, we performed more reactions than would be required to cover all the rungs of the charge ladders; the rungs of eight reactions overlapped such that every rung was produced in two to five of the charge ladders in each buffer with the exceptions of rungs $x = 16$ – 18 , which could not be formed at lower values of pH before the reagent had hydrolyzed. There is no set of rate constants that fits all the inte-

grated experimental rungs perfectly, because peak area in replicate reactions in the same buffer differ slightly and because there are small irreconcilable discrepancies between the rate constants necessary to form the same rungs within different reactions.

Error is also introduced in the CE experiment and the integration of rungs of the charge ladders. We observe occasional aberrations, such as air bubbles, in the CE experiments; we ran every reaction twice to distinguish these sharp peaks due to random errors from the peaks of the charge ladder. Running every sample twice also provided a measure of the error generated in the CE experiments and integration. We found an average

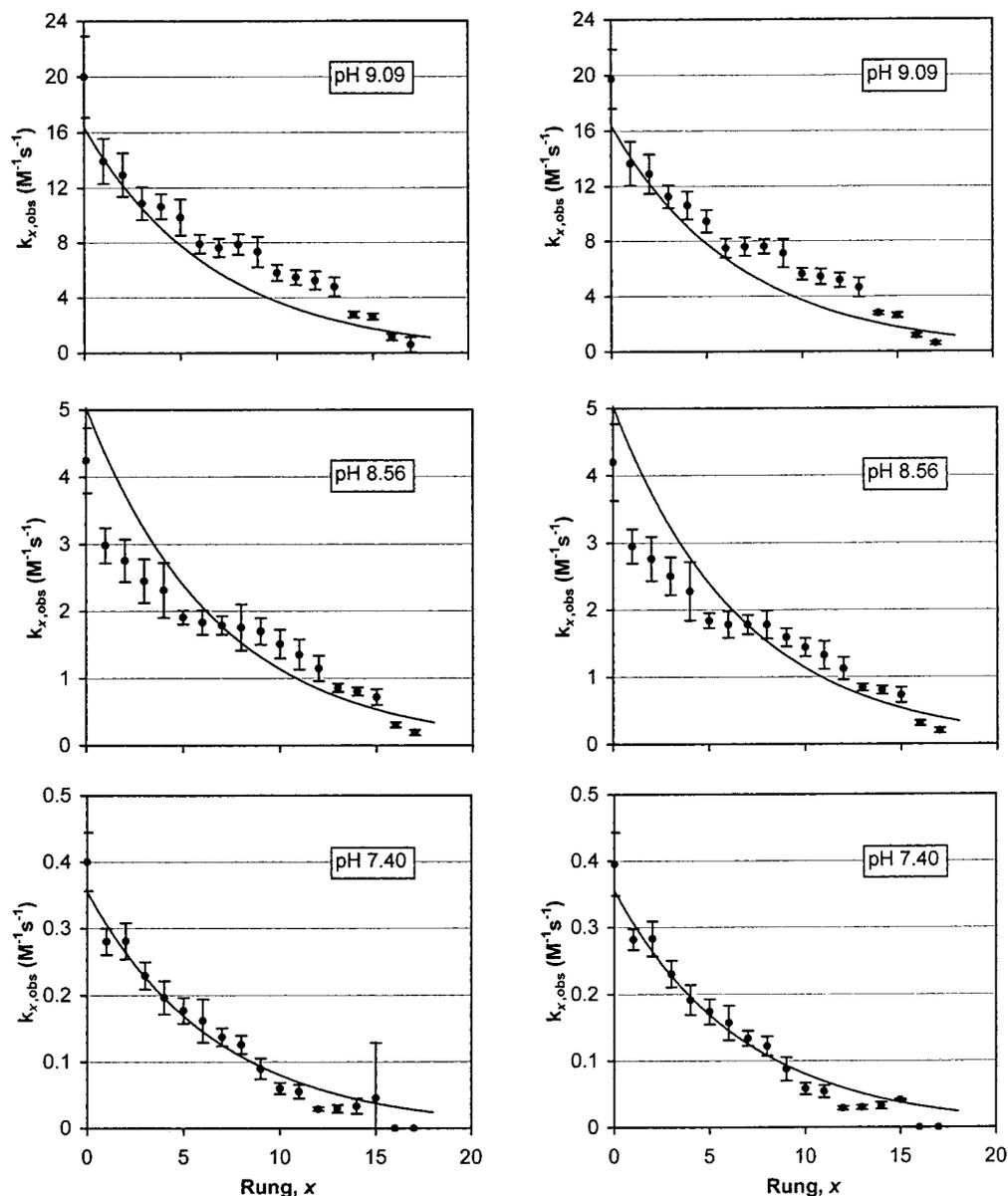


Figure 3. Decreasing reactivity with acetylation. The data on the left were generated using the first search algorithm (programmed by the authors in C and described in the text), and those on the right were generated using the second search algorithm (generalized reduced gradient method). The filled circles (●) are the observed rate constants, $k_{x,obs}$, determined from modeling the kinetics of rung formation from a series of charge ladders formed at three values of pH. The line shows the theoretical estimation of observed rate constants at these three values of pH, based on eq 10. We first solved iteratively for the values of Z_x . We started with a value of 0.8 for β and decreased it by 0.006 for every acylation. The parameter G was calculated by minimizing the difference between the theoretical and experimental values of k_x . The error was assigned based on rate constants determined for input data that had been perturbed by a random amount between -10% and $+10\%$. Each error bar indicates two standard deviations for that rate constant found from five perturbed data sets (96% confidence limit).

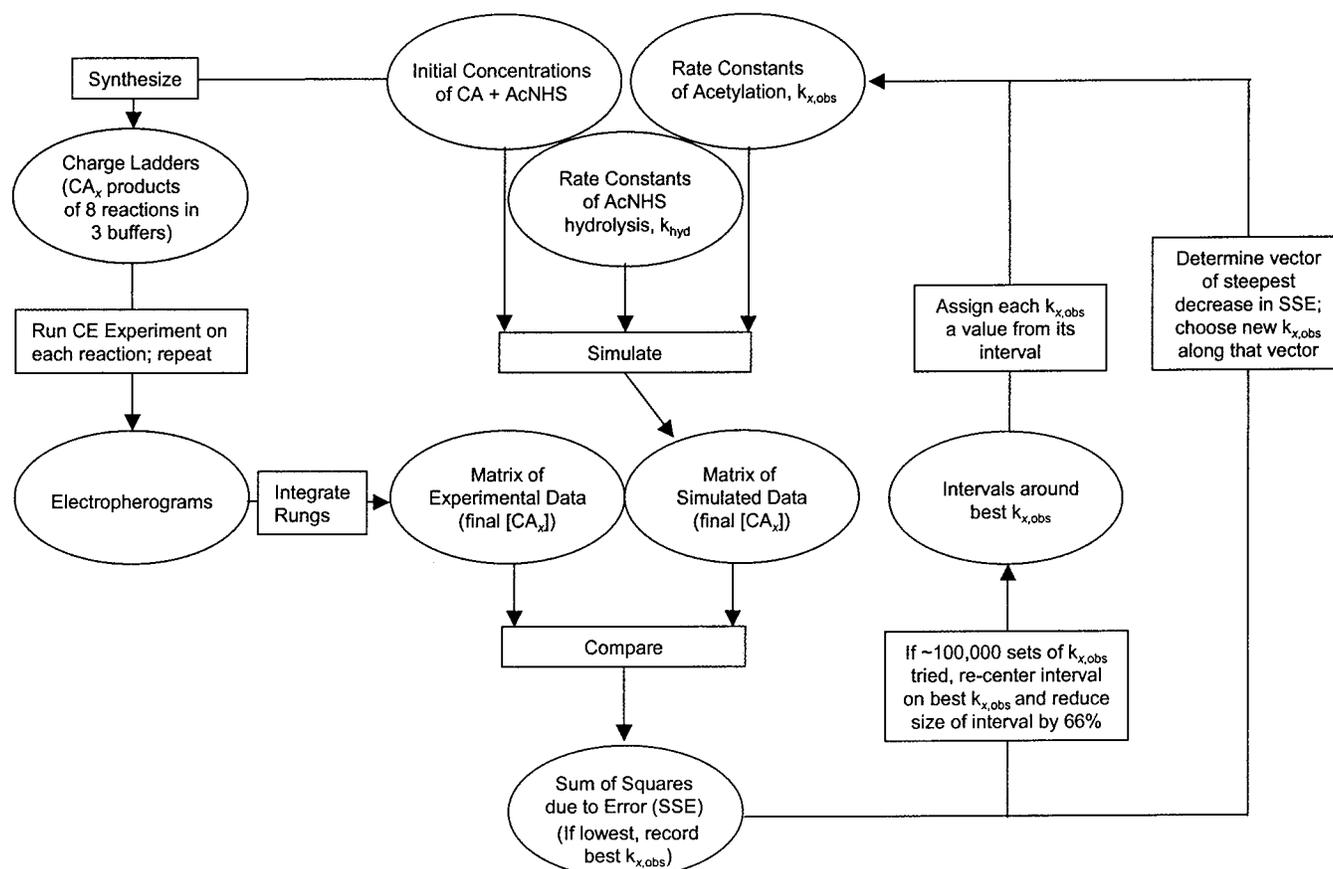
difference between the two runs of 6% in peak areas of all rungs of all charge ladders in the three buffers.

We noted that rungs with larger areas have a percentage error much lower than that of the smaller rungs. We found that the average error in peaks that are larger than the median of nonzero peaks in a charge ladder was 4%. The average error in peaks that have nonzero areas smaller than this median value was 30%. The charge ladders of CA (Figure 1)—though the best resolved that have been obtained for a protein charge ladder—do not have perfect baseline resolution. Due to noise and random drift in the CE experiments, the smallest rungs on either end of some charge ladders are so small that they are interpreted as noise in one CE

run and as a peak in another. This error is a significant contributor to the average error in peak area. We took into account the fact that larger peaks had a greater relative accuracy than smaller peaks when we chose the SSE as the measure of fit between our simulated and experimental results. The search was effectively biased toward matching the larger peaks because SSE is a measure of absolute error, not relative error, in peak area.

Determining the uncertainty in $k_{x,obs}$ based on the propagation of error is not straightforward. We can, however, assess the uncertainty in each $k_{x,obs}$ by rerunning the simulations on data (areas of rungs) that we perturbed randomly within an interval of -10% to $+10\%$ and noting the response in the $k_{x,obs}$. The

Scheme 1



distribution of perturbations was flat across the interval; i.e., perturbations were not weighted around the “real” values of areas of rungs as with a Gaussian distribution. We generated five sets of perturbed data for each buffer and used both search algorithms to find the set of best $k_{x,obs}$. The standard deviation in a set of $k_{x,obs}$ from this sensitivity test was 5%–12% of the “correct” peak areas, on average. The error bars in Figure 3 show two standard deviations, representing the 96% confidence limit for individual $k_{x,obs}$. The first search algorithm produces slightly higher error on average (9%) than the second algorithm (6%) and particularly in the later rate constants. Analysis of the evolution of best $k_{x,obs}$ indicates that additional sets of simulations would reduce this error.

Analysis of Data. We analyzed the rate constants generated by these procedures on the assumption that all Lys groups in a rung have the same reactivity and pK_a but that this reactivity decreases as Lys groups are progressively acetylated and the surface of the CA_x becomes more negatively charged. The possible physical justification of this assumption is that all Lys groups in CA are found on its surface and none are in the active site or buried in the hydrophobic interior.^{15,21}

We compared the observed rate constants $k_{x,obs}$ from our modeling with $k_{x,obs}$ that we estimate from theory that we derive as follows. Each $k_{x,obs}$ in a set of 18 represents the estimated reactivity for a residual unacetylated Lys group on CA_x and does

not differentiate between $\epsilon\text{-NH}_2$ and $\epsilon\text{-NH}_3^+$. The rate constant $k_{\text{acetylation}}$ given in eq 2 refers only to the reactivity of the fraction of the Lys that is unprotonated and therefore capable of nucleophilic reaction. These two rate constants are related by the fraction of unprotonated amine (eq 7).

$$k_{\text{obs}} = k_{\text{acetylation}} / (1 + 10^{pK_a - \text{pH}}) \quad (7)$$

The electrostatic potential, φ_x , at the surface of CA_x becomes increasingly negative with acetylation, as described by Debye–Hückel theory. This theory applies to values of φ_x of up to ~ 25 mV and is applicable to all rungs for these experiments. As the potential of CA_x becomes more negative, its residues are more likely to be protonated, according to a Boltzmann distribution. This effect has been interpreted in two different, but equivalent, ways (eq 1). Carbeck et al. cited the Tanford–Roxby equation that relates the increasingly negative potential of CA_x to an upward shift in pK_a of its residues by a value of $e\varphi_x/2.303kT$ where e is electronic charge, k is the Boltzmann constant, and T is absolute temperature.⁵ This shift reflects the requirement for increasing association of protons to minimize the free energy of the system. Menon and Zydney derived the same term under the theory of “charge regulation” as a downward shift in the local pH.⁹ That is, protons are attracted to the vicinity of the negatively charged protein such that the residues experience a pH that is lower than that of the bulk solution. Regardless of whether the additional protons attracted to the protein are associated with water

(21) Dodgson, S. J., Tashian, R. E., Gros, G., Carter, N. D., Eds. *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*; Plenum: New York, 1991.

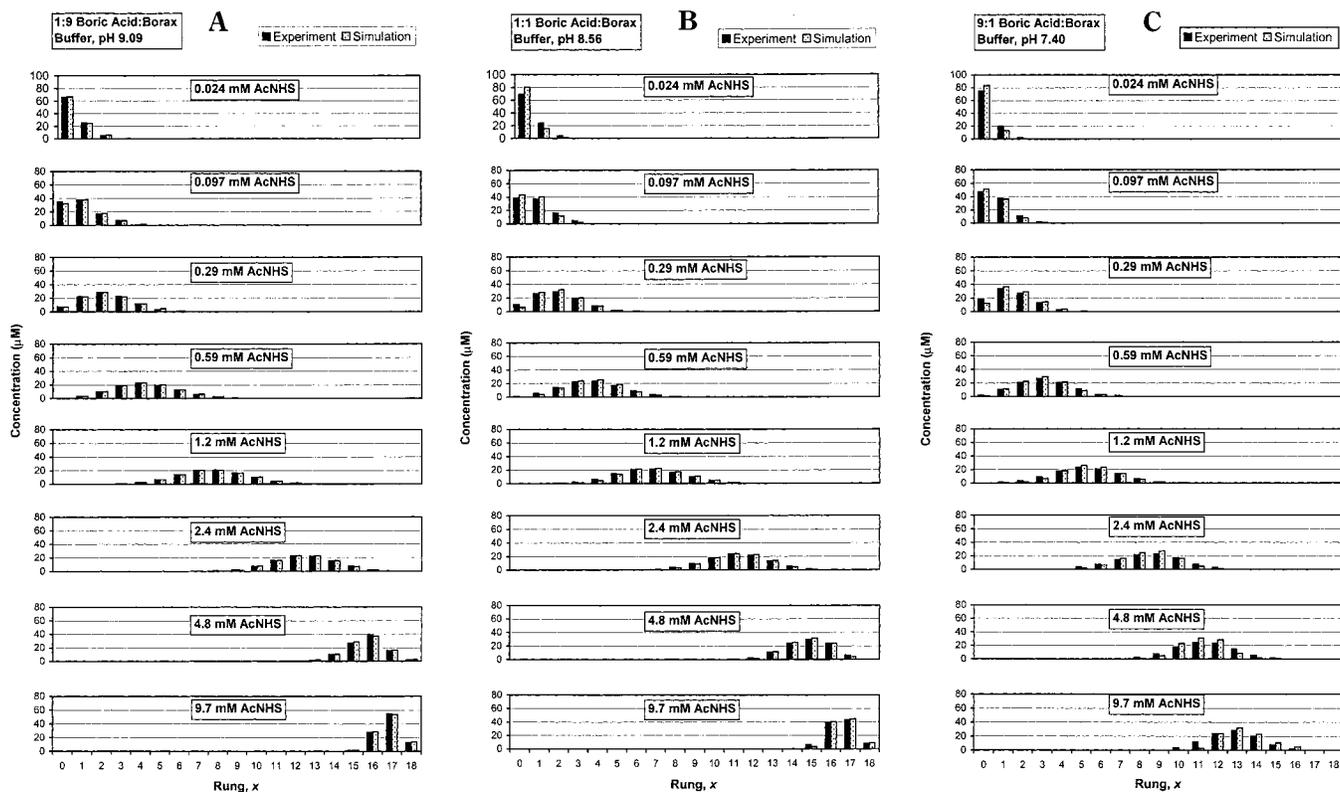


Figure 4. Comparison of data from experiment and simulation in (A) 1:9 boric acid–Borax, pH 9.09; (B) 1:1 boric acid–Borax, pH 8.56; and (C) 9:1 boric acid–Borax, pH 7.40. The rungs of charge ladders from CE experiments were integrated and are shown as the “experimental” bars in the histograms. The best rate constants $k_{x,\text{obs}}$ found in each buffer were used in a simulation (similar to that shown in Figure 2) that produced the final concentrations of rungs shown here as the “simulation” bars.

molecules in the buffer (as a lower local pH), or with the residues of CA_x (as a higher $\text{p}K_a$), the term contributes to the fraction of Lys groups that are dissociated and might best be thought of as an adjustment that gives an effective $\text{p}K_a$ ($\text{p}K_{a,\text{eff}}$, eq 8). Incorporat-

$$\text{p}K_{a,\text{eff}} = \text{p}K_a + e\varphi_x/2.303kT \quad (8)$$

ing eqs 2 and 8 into eq 7 gives eq 9 upon simplification. From the

$$k_{x,\text{obs}} = 10^{\beta(\text{p}K_a + e\varphi_x/2.303kT) + G} / (1 + 10^{\text{p}K_a + e\varphi_x/2.303kT - \text{pH}}) \quad (9)$$

linearized Poisson–Boltzmann equation, $\varphi_x = eZ_x/4\pi\epsilon\epsilon_0a(1 + \kappa a)$, where ϵ is the dielectric constant of the buffer, ϵ_0 is the permittivity of free space, κ is the inverse Debye length, a is the radius of the protein, and Z_x is the charge of CA_x (unitless, until multiplied by the charge of a proton). In the three buffers at constant ionic strength, κ^{-1} is ~ 10 nm and ϵ is ~ 80 . We used a protein radius of 2.05 nm^4 to derive eq 10, where c is a unitless constant of

$$k_{x,\text{obs}} = 10^{\beta(\text{p}K_a + cZ_x) + \log G} / (1 + 10^{\text{p}K_a + cZ_x - \text{pH}}) \quad (10)$$

~ 0.00667 for this system. We solved eq 10 to estimate a theoretical value of $k_{x,\text{obs}}$ as a function of x at values of pH of 9.09, 8.56, and 7.40. We first determined Z_x for each x by summing the contribution of all charged residues. Because the charge on each residue depended on its $\text{p}K_{a,\text{eff}}$, which in turn depended on Z_x , we solved iteratively for Z_x in the manner of previous work.⁹ That is,

Table 1. Number of Ionizable Residues and $\text{p}K_a$ Values for CA Used in Calculations of Charge of a Rung

| residue | n | $\text{p}K_a$ |
|-----------------------------------|------------------|---------------|
| Lys | 18 ²² | 10.3 |
| His | 8 ²³ | 6.2 |
| Arg | 9 | 12.5 |
| Zn ^{II} -OH ₂ | 1 | 7.0 |
| Asp | 19 | 3.5 |
| Glu | 11 | 4.5 |
| C-terminus | 1 | 3.2 |

we started with an estimated value of Z_x , used it to calculate the $\text{p}K_{a,\text{eff}}$ for each ionizable residue according to eq 8, and determined the charge on each residue from $n_{\text{res}}/(1 + 10^{\text{p}K_{a,\text{eff}} - \text{pH}}) = Z_{\text{res}}$, where n_{res} is the number of a particular ionizable residue present in CA and the $\text{p}K_a$ of that residue is given in Table 1.¹ The sum of Z_{res} for the charged residues shown in Table 1 gives Z_x , which we solved numerically to be self-consistent with the initial guess.

For these experiments, performed at a higher ionic strength than previous studies (130 mM as opposed to 10–25 mM^{5,9}), the $\text{p}K_{a,\text{eff}}$ of Lys was calculated to increase only 0.006–0.007 unit per acetylation at all three values of pH. Similarly, ΔZ_x between the pairs of rungs varies between 0.93 and 0.98, depending on x and the pH.

Hirata and Nakasato found a value of β for the reactions of AcNHS with amino acids of 0.69.¹² Jencks and Gilchrist showed a value of 0.8 is typical for β of reactions of amines with esters,

except for reactions between the most basic amines and most reactive esters.¹⁴ When the value of pK_a for the conjugate acid of the leaving group is 4–5 units below that of the amine, the value of β begins to decrease with increasing pK_a of the amine, dropping to ~ 0.15 unit over ~ 3 units of increasing pK_a of the nucleophiles.¹³ The reactions between the Lys ϵ -NH₂ ($pK_a = 10.3$ – 10.5 over $x = 0$ – 17) and AcNHS (pK_a of NHS = 6.1)²⁴ fall in the region where we estimate β to be 0.8 – $0.006x$. The value of β that we used to solve eq 10 decreased from 0.8 to 0.69 over the full range of x , thereby incorporating both of the literature trends. The parameter G , a constant for a particular reaction, was determined by minimizing the difference between experimental and theoretical rate constants for all pH values. We plotted the values for $k_{x,obs}$ we calculated from eq 10 and those we derived from simulations of the formation of experimental charge ladders as a function of x in Figure 3. The theoretical values for $k_{x,obs}$ are in general agreement with the three sets of $k_{x,obs}$ found from analysis of the kinetics of formation of charge ladders (Figure 3).

Other Interpretations of Decreasing Reactivity. There is another explanation for the apparent decrease in the reaction rate with acetylation that is not based on the increasing association of protons with Lys groups. This interpretation would assume that the Lys groups have a distribution of reactivities. The most reactive Lys groups would be acetylated first. The calculations of Caravella et al. suggest that a range of reactivities is plausible for CA.¹⁵

CONCLUSION

This work uses the kinetics of acetylation to provide information about the rungs of a charge ladder. We detected the response of residues on the surface of CA to incremental changes in the surface potential at a fairly high ionic strength and found that the reactivity of residual Lys groups decreases appreciably as the surface of the protein becomes more negative. The decrease is consistent with a model that quantifies reactivity lost in terms of an increasing tendency of protons to associate with unacetylated Lys amino groups. The change in proton association balances the charges in the system and suggests that acetylated and unacetylated Lys groups on the surface “communicate” with each other through long-range electrostatic interactions, even in this environment of significant charge shielding.

The technique used in this work combines syntheses, CE experiments, and mathematical modeling. The response of the residues to changing potential is reflected in the mixture of products from multiple reactions. As such, the reactions can be carried out under a variety of conditions suited to testing different hypotheses—e.g., different values of pH and ionic strength—while the CE experiments can be carried out under completely different conditions suited to optimal separation—e.g., a deuterated buffer of low ionic strength. By contrast, other studies that have probed the response of protein charge to incremental changes in potential

using the electrophoretic mobility have not had this flexibility. We have shown the modeling to be robust in producing values of $k_{x,obs}$ that are independent of our search algorithm but responsive to changes in the input.

This technique is limited to proteins that form well-resolved charge ladders (generally <100 kDa),⁷ and the complexity of its modeling component increases with the number of Lys groups on the protein.

EXPERIMENTAL SECTION

Reactions. We made stock solutions of 200 mM boric acid and 50 mM Borax ($Na_2B_4O_7$) in deionized water and combined them in different ratios to make buffer solutions at three values of pH. The buffers were formed by mixing stock solutions of boric acid and Borax in the ratios 1:9, 1:1, and 9:1, yielding values of pH of 9.09, 8.56, and 7.40, respectively, as measured by an Accumet model 15 pH meter (Fisher Scientific, Pittsburgh, PA). We added 69 and 114 mM NaCl to the latter two buffers, respectively, to ensure all buffers had the same ionic strength of 0.13 M.

We made solutions of 3 mg of carbonic anhydrase II (from bovine erythrocytes, $pI \sim 5.9$, Sigma Chemical Co., St. Louis, MO, C-2522) in 1 mL of each of buffers of pH 9.09, 8.56, and 7.40. We made a stock solution of 32 mg of acetic acid *N*-hydroxysuccinimide ester (Sigma Chemical Co., A-9153) in 1 mL of dioxane. This stock solution was then diluted 400, 100, 33.3, 16.7, 8, 4, 2, and 1 times with dioxane. These reagents were made immediately before they were used. We performed 24 reactions by quickly adding 10 μ L of each dilution of AcNHS to a 200- μ L aliquot of the BCAII solution in each borate buffer and agitating the reaction immediately using a vortex mixer. The reactions were left overnight to ensure completion. We separated the reaction products by adding 75 μ L–1.5 mL of H₂O in a Centricon 10 tube (Millipore, Burlington, MA) and centrifuging for 20 min at 6500 rpm, followed by adding 1.5 mL of H₂O to the concentrate and centrifuging for 20 additional min at 6500 rpm. The concentrate contained only molecules larger than 10 kDa.

CE Experiments. The reactions were analyzed by capillary electrophoresis experiments using a Beckman P/ACE 5500. The running buffer used was 25 mM Tris–192 mM Gly in D₂O. To prepare samples of each reaction, we mixed 50 μ L of each purified reaction with 50 μ L of 50 mM Tris–384 mM Gly in D₂O and 2 μ L of 0.5 vol % *p*-methoxybenzyl alcohol in 25 mM Tris–192 mM Gly in D₂O. Each sample was pressure injected for 15 s into a 50- μ m fused-silica capillary. The capillaries measured 117 cm in total length and 110 cm in length from injection to the detector. The applied voltage was 30 kV; detection used UV absorbance at 214 nm. Each sample was run twice.

Determination of Rates of Hydrolysis in Buffers. Values of rates of hydrolysis of reagent in each buffer were measured by a time series of UV absorbance measurements using a Hewlett-Packard 8453 UV/visible spectrophotometer. We added 10 μ L of ~ 25 mM AcNHS in dioxane to 990 μ L of each of the three buffers and monitored the absorbance of NHS at 260 nm, sampling every 10 s over 190–1100 nm during the first 1200 s of reaction. Based on the time range 30–500 s, values of 7.9×10^{-3} , 2.6×10^{-3} , and 4.6×10^{-4} s⁻¹ were calculated for buffers at pH 9.09, 8.56, and 7.40, respectively.

(22) The α -NH₂ Lys is naturally acetylated and, therefore, not included here: Colton, I. J.; Anderson, J. R.; Gao, J.; Chapman, R. G.; Isaacs, L. Whitesides, G. M. *J. Am. Chem. Soc.* **1997**, *119*, 12701–12709.

(23) Three additional His residues are coordinated to the Zn²⁺ atom in the active site; they are not ionizable and not included here: Carbeck, J. D.; Colton, I. J.; Anderson, J. R.; Deutch, J. M.; Whitesides, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 10671–10679.

(24) Koppel, I.; Koppel, J.; Leito, I.; Pihl, V.; Grehn, L.; Ragnarsson, U. *J. Chem. Res., Synop.* **1994**, *6*, 212–213.

Integration of CE peaks. Using PeakFit software, version 4.0, the peaks were approximated using the AutoFit Peaks I Residuals algorithm. The baseline was determined by detecting the second-derivative minimums below 2% of the amplitude of the highest peak. The baseline was subtracted from the data of the electropherogram. The data from the electropherogram were not smoothed. Peaks with an amplitude above 0.5% of the amplitude of the highest peak were fit to a Gaussian peak shape with variable width. The total peak areas were normalized to the initial concentration of native protein, 9.7×10^{-5} M, for each experiment.

Simulations. Simulations were written in C and executed using an Intel Pentium III 800-MHz processor.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health Grants GM 51559 and GM 30367. O.C. acknowledges a Harvard University Undergraduate Research Award. I.G. acknowledges a National Science Foundation Graduate Research Fellowship. J.R.A. acknowledges Steven J. Metallo for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Additional data as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review September 11, 2001. Accepted January 29, 2002.

AC0109927