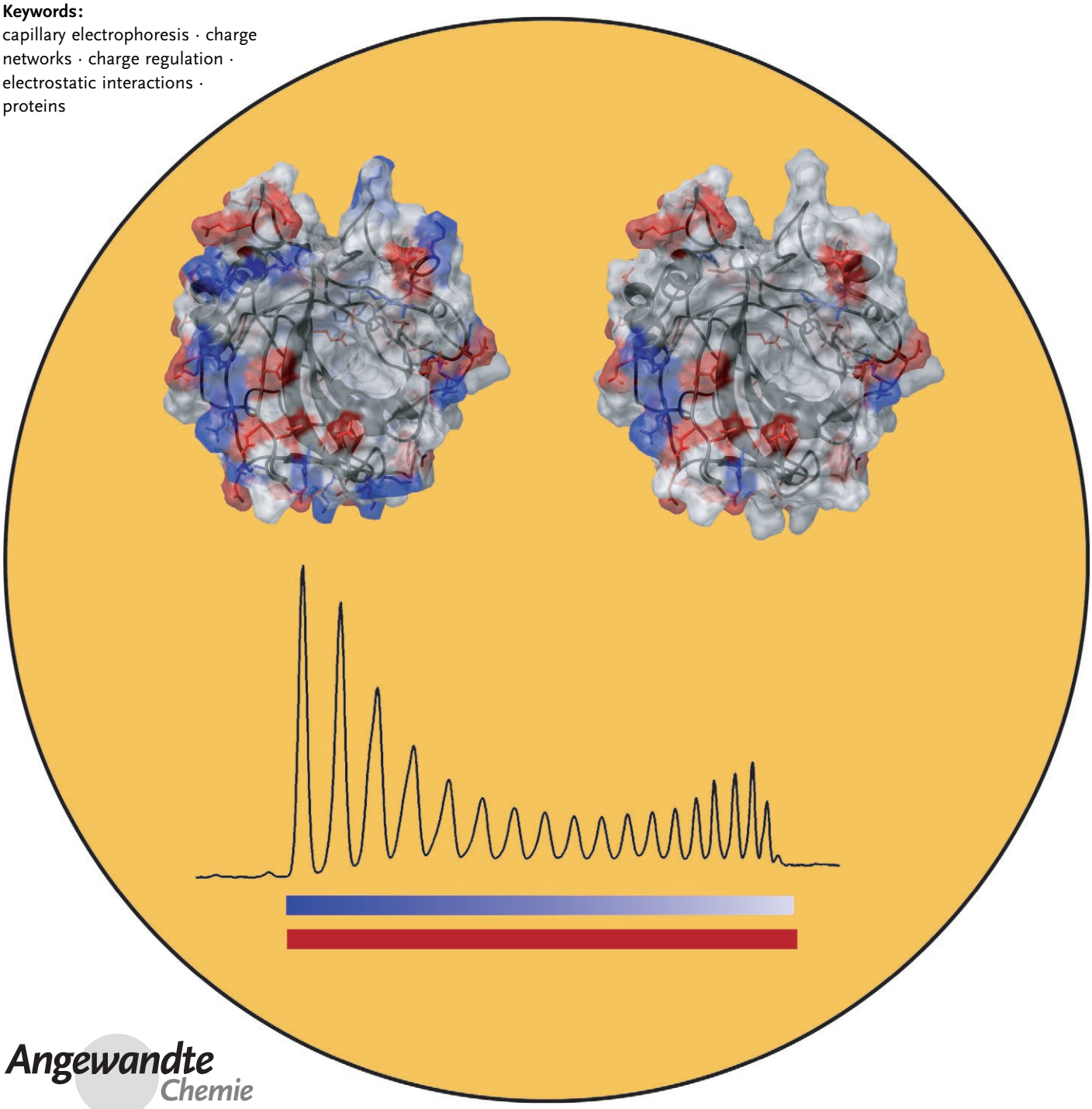
DOI: 10.1002/anie.200502530

Proteins

Why Are Proteins Charged? Networks of Charge–Charge Interactions in Proteins Measured by Charge Ladders and Capillary Electrophoresis

*Irina Gitlin, Jeffrey D. Carbeck, and George M. Whitesides**

Keywords:
capillary electrophoresis · charge networks · charge regulation · electrostatic interactions · proteins



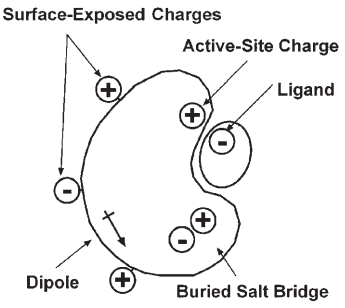
Almost all proteins contain charged amino acids. While the function in catalysis or binding of individual charges in the active site can often be identified, it is less clear how to assign function to charges beyond this region. Are they necessary for solubility? For reasons other than solubility? Can manipulating these charges change the properties of proteins? A combination of capillary electrophoresis (CE) and protein charge ladders makes it possible to study the roles of charged residues on the surface of proteins outside the active site. This method involves chemical modification of those residues to generate a large number of derivatives of the protein that differ in charge. CE separates those derivatives into groups with the same number of modified charged groups. By studying the influence of charge on the properties of proteins using charge ladders, it is possible to estimate the net charge and hydrodynamic radius and to infer the role of charged residues in ligand binding and protein folding.

From the Contents

1. Introduction

Why do proteins include amino acids with charged side chains? Charged groups are often found at the active site of proteins, where they interact electrostatically with substrates, transition states, and products. Most efforts to study the role of charged groups in the biological activity of proteins have focused on these groups, and in many cases the mechanisms of their action are understood qualitatively, if not quantitatively.[1,2]

Charged groups on proteins are not restricted to active sites: they are found over the entire surface of proteins: approximately one third of the amino acids on the protein–water interface of a typical globular protein are charged.[3] Charged functional groups are also present in the interior of proteins (although they are concentrated on the surface; Scheme 1). The influence of "non-active site" charged residues on the structure, stability, and activity of proteins is not well understood. This Review focuses on these non-active site charges.



Scheme 1. Schematic diagram of a protein, showing surface, active site, and buried charges. The shape of the dielectric cavity seen by the charges, and the overall distribution of charge, may depend on whether a ligand is or is not bound at the active site.

Interactions between charged groups are often referred to simply as "electrostatics" in protein biochemistry and biophysics. This term is misleading: the charges are not static, and charged side chains, substrates, and ions in the solution all

[*] Dr. I. Gitlin, Prof. Dr. G. M. Whitesides
Department of Chemistry and Chemical Biology
Harvard University
12 Oxford St., Cambridge, MA 02138 (USA)
Fax: (+1) 617-495-9857
E-mail: gwhitesides@gmwgroup.harvard.edu
Dr. J. D. Carbeck
Department of Chemical Engineering
Princeton University
A-319 Engineering Quadrangle, Princeton, NJ 08544 (USA)

move. The charged groups on proteins are also predominantly acids and bases that exchange protons with water; they are thus in dynamic equilibrium with their uncharged states. The term "electrostatics" is, however, convenient when referring to charge–charge, charge–dipole, and dipole–dipole interactions in proteins, and we use the term in this sense in this Review.

1.1. The Role of Charged Groups in Biochemistry

Charged groups can unquestionably play important roles in protein biochemistry. Processes that are influenced by electrostatic interactions include:

1) association of receptors with charged ligands;[4–6]
2) binding of substrates by enzymes, and catalysis of reactions;[1,7]
3) formation of protein–protein[8–11] and protein–nucleic acid[12] complexes;
4) transfer of electrons between biological reaction centers;[13]
5) selective transport of ions in protein channels;[14]
6) folding and stability of proteins;[15–19]
7) denaturation of proteins at high and low pH values;[20,21]
8) solubilization and precipitation of proteins;[22]
9) incorporation of proteins into amyloid fibers.[23]

The importance of charged groups in protein biochemistry has been studied extensively using site-directed mutagenesis as the primary tool, and with an emphasis on mutations in the active site.[2] This approach measures the contribution of the side chain of a specific amino acid to the stability and/or activity of a protein. Theoretical models of electrostatic interactions, both microscopic and continuum,[24–26] have complemented these experimental studies.

We have developed protein charge ladders as an alternative experimental approach to study charged groups in proteins; this approach is complementary to site-directed mutagenesis. Protein charge ladders clarify the contributions of charges to the stability of proteins, and to their interactions with ligands. Studies with charge ladders provide a different type of information than do those with site-directed mutagenesis, in that they allow the measurement of average contributions of a particular type of charged group (e.g., NH_3^+ or COO^- groups) on the exterior of the protein to stability or binding. This Review focuses on this subject.

Understanding electrostatic interactions will help to understand many processes in which proteins are involved. This understanding may, in turn, allow the design of molecules that bind to proteins (that is, potential drug leads) and the engineering of proteins for improved activity, stability, and processability, and for new functionality. There is also the opportunity to understand more about structure–function relationships in a way that might be used to design new molecules that mimic its properties.

1.2. The Free Energy of Proteins

The net changes in free energy as proteins fold or bind to molecules are often the result of large and opposing contributions,[27–29] and comprised of both enthalpic and entropic terms. For example, the native state of proteins is stabilized primarily by the unfavorable free energy of solvation of nonpolar side chains by water (that is, the "hydrophobic effect"); the hydrophobic effect causes these side chains to form the hydrophobic core of the native state. The less-ordered denatured state is stabilized by the favorable configurational free energy of the peptide backbone and side chains.[30,31]

Hydrophobic interactions are the main driving force for protein folding.[27,28,32] The relative contributions of enthalpy and entropy to the hydrophobic effect vary with temperature, but at 25 °C the free energy of hydrophobic interactions is dominated by the entropy of solvation of nonpolar solutes by water ($\Delta G_{\text{hydrophob}} \approx -T\Delta S_{\text{hydrophob}}$).[28,33] Although hydrophobic interactions are still a challenge to describe theoretically,[34] good empirical correlations exist between the hydrophobic free energy of solvation and solvent-exposed surface area,[35] or number of C–H bonds.[36] These correlations allow reasonable estimates of the overall contribution of hydrophobic interactions to protein stability.[37]

The contributions of configurational free energy to protein stability have been more difficult to quantify.[31] Translational, vibrational, rotational, and configurational entropies of small molecules can be estimated using simple models from statistical mechanics.[38] The configurational entropy and low frequency vibrational modes of proteins

Jeffrey D. Carbeck is a chemical engineer and materials scientist with expertise in biomaterials, biophysical and bio-analytical chemistry, molecular simulations, and theory. He gained his B.S.E from the University of Michigan, and his PhD from MIT. Both degrees were in materials science. He then spent two years as a postdoctoral researcher with G. M. Whitesides at Harvard University. He was a member of the faculty of Princeton University in the Department of Chemical Engineering from 1998 to 2006. In 2005, he co-founded the biomaterials company WMR, Inc.

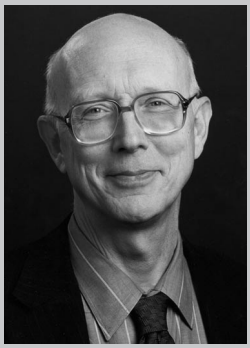
Irina Gitlin received a BS in chemical engineering from the University of Illinois at Urbana-Champaign in 2000. She has recently completed her PhD in chemistry with G. M. Whitesides. Her research in the area of protein chemistry explores the role of electrostatic interactions in protein stability, binding of ligands, and susceptibility to denaturation by surfactants. She has also worked in the area of microfluidics.

are, however, difficult to treat because they are numerous and poorly characterized; they are also difficult to measure experimentally, and molecular dynamics simulations are typically restricted to time-scales too short to capture these modes. Finally, proteins in the denatured state often do not resemble random coils, for which polymer physics has provided useful models of configurational entropy.[39]

The changes in configurational free energy and free energy of solvation that accompany protein denaturation are both on the order of 100 kcal mol^{-1}.[40] In contrast, the net difference in free energy between native and denatured states (that is, the standard state change in Gibbs free energy of denaturation, ΔG°_{D-N}; D: denatured, N: native) under conditions at which most proteins are stable (25 °C, neutral pH, and without chemical denaturants) is of the order of 10 kcal mol^{-1}.[33] This difference is equivalent in free energy to the formation of approximately 10 hydrogen bonds, a net hydrophobic interaction of perhaps 10 CH_3 groups, or electrostatic interaction between two charges in low ($\varepsilon \approx 5$) dielectric separated by a distance of less than 1 nm. Since these large and opposing contributions to the total free energy nearly cancel, other noncovalent interactions influence and may in certain cases determine the net stability of proteins.

Van der Waals interactions, through a combination of repulsion and attraction, provide selectivity to noncovalent interactions. When a pair of molecules has shapes that are complementary, numerous atoms from each of the molecules come in close contact simultaneously; this contact allows the attractive part of the interaction to make a significant contribution to the specificity of the interaction.[41]

Hydrogen bonds stabilize the secondary structure of proteins (e.g., α helices and β sheets). They are a particular type of electrostatic interactions between a weak acid donor group (e.g., N–H, O–H) and an acceptor group with a lone pair of electrons (e.g., N, O). For a hydrogen bond to form, the distance between the donor and acceptor group must be about 3 Å with the donor–H bond preferably pointing along the axis of the lone pair orbital of the acceptor.[42] Although most hydrogen bonds are electrostatic, there is some controversy over the presence and possible importance in enzymatic catalysis of short hydrogen bonds that are characterized by two energy minima with low energy barriers (low-barrier hydrogen bonds) and which may have a covalent component.[43]



George M. Whitesides received his AB degree from Harvard University in 1960, and his PhD from the California Institute of Technology in 1964. He was a Mallinckrodt Professor of Chemistry from 1982–2004, and is now a Woodford L. and Ann A. Flowers University Professor. Prior to joining the Harvard faculty in 1992, he was a member of the chemistry faculty of the Massachusetts Institute of Technology. His research interests include physical and organic chemistry, materials science, biophysics, complexity, surface science, microfluidics, self-assembly, micro- and nanotechnology, and cell-surface biochemistry.

1.3. Measuring and Understanding Electrostatic Interactions in Proteins

Electrostatic interactions between charged groups in proteins in aqueous solution are difficult to understand and quantify for four reasons:

1) Electrostatic interactions are long-range. The electrostatic free energy of a protein includes interactions between charged groups that are separated by distances comparable to the dimensions of the protein.[28] Large numbers of interactions must thus be considered in calculating the total electrostatic free energy.
2) Acid–base equilibria are cooperative. The apparent pK_a value of a particular ionizable group on a protein may depend significantly on its interactions with other charged and polar groups, both on the protein and in solution.[44,45] Thus, a protein must be described as a network of interdependent charges. The quantitative description of a nonlinear network with a number of nodes (that is, charges) is not straightforward.
3) The dielectric properties of protein/solvent systems are heterogeneous. There is no single, well-defined parameter that describes the dielectric properties of proteins and the surrounding solutions.[46] A common simplifying approximation is to consider three separate regions of the system, each with a different dielectric constant: a) the core of the protein has a low dielectric constant ($\varepsilon = 2$–4), b) the bulk solvent (water, buffer, or biological fluid) has a high dielectric constant (typically taken to be that of bulk water: $\varepsilon \approx 80$), and c) the surface of the protein and surrounding layer of solvent has intermediate values of dielectric constant ($\varepsilon \approx 10$–20).[47] The dielectric constant of solvent-exposed regions of the protein is higher than that of the interior as a consequence of the presence and configurational mobility of polar side chains.[47] The dielectric constant of water near the surface of proteins is lower than bulk water because of the reduced mobility—and perhaps enhanced structure—of water in this region. These differences in dielectric properties are important because both the free energy of interaction between charges, and the free energy of solvation of individual charges, depend on the local dielectric properties of the medium the charges experience.[48] Average dielectric constants of the interiors of proteins may vary not only with heterogeneities within a protein but also with their stability to temperature.[49]
4) In biological systems, electrostatic interactions are supplemented by specific ion effects. The background electrolyte can affect the properties of proteins in solutions through forces other than simple electrostatics: the Hofmeister series and related classifications reflect the complexity of interactions involving proteins, ions, and water.[50] These effects are incompletely understood and are omitted in most descriptions of protein-containing solutions. Ninham and co-workers have pointed out that specific ion effects can in part be accounted for by dispersion forces between ions.[51,52] While we do not explicitly review this subject here, we emphasize that a

complete picture of protein–ion or protein–protein interactions must ultimately include those forces.

The most common method used by biologists to explore stability or activity of proteins is site-directed mutagenesis—that is, changing the identity of single amino acids and attempting to measure changes in the free energy or in the activity between the mutant and the wildtype.[2,53] This approach successfully identifies individual amino acids that make major contributions to the process of interest. We demonstrate in this Review that there are important interactions among charged groups on proteins that are best, and perhaps only, described by considering the charged groups as a network, rather than as electrostatically non-interacting charges.

1.4. Protein Charge Ladders

The protein charge ladder method uses chemical modifications that target charged groups on the surface of a protein (e.g. acetylation of a Lys-NH_3^+ residue) to annihilate the charge associated with that group (Figure 1 A).[54] It has the advantage that it produces large numbers of derivatives of a protein—perhaps several thousand—in a single experimental step. It has the disadvantage that these derivatives are characterized only by their net charge, not by knowledge of the detailed location of the charges. Relative to site-specific mutagenesis, it substitutes numbers of derivatives and experimental ease for knowledge of the detailed structure of these mutants.

This approach to producing large numbers of derivatives of a protein was first described by Hollecker and Creighton[19] and further developed by Goto and co-workers.[55] This earlier work was also based on the acylation of amines, and is similar in spirit to the work described in this Review. Hollecker and Creighton used urea-gradient gel electrophoresis to examine effects of reversing the charge of amino groups on protein stability.[19] Conversion of the first several Lys-ε-NH_3^+ groups of cytochrome *c* to negatively charged species (ε-$NHCOCH_2CH_2CO_2^-$) with succinic anhydride produced little effect on stability; succinylation of the final lysine group, however, caused unfolding even in the absence of urea. The limited resolution of gel electrophoresis prevented quantitative measurement of the effects of incremental changes in the net charge on stability.

The mixture of protein derivatives was analyzed by capillary electrophoresis (CE). This analytical technique separates the derivatized proteins into separate bands composed of families of proteins that all have the same number of modified charged groups and, therefore, approximately the same net charge (Figure 1 B, details of the analysis are described in Section 6.2). We call the protein derivatives separated in this way a "protein charge ladder," and each peak of the "ladder" we call a "rung". The resolution of CE allows quantitative measurement of the effects of changes in charge on protein folding[56] and ligand binding[4,5] that result from adding or subtracting a unit of charge by modification of a charged group. The combination of charge ladders and CE also makes it possible to isolate net charge[57] and hydrodynamic drag[58] in physiologically relevant solutions.

We have used three well-characterized proteins—carbonic anhydrase, α-lactalbumin, and lysozyme—for most of our studies. Charge ladders can, however, be generalized to many,[54] although not all, proteins. Charge ladders can also be used to count the number of modifications on a protein, and thus to study other properties of the protein surface: in this Review we describe changes in the hydrodynamic size of the protein and surface hydrophobicity produced by using this same approach of incremental chemical modifications.

The objective of this Review is to describe the use of protein charge ladders to address questions in the physical chemistry of proteins, especially the long-range interactions between charged groups. Section 2 is an overview (or tutorial) of the various contributions to the electrostatic free energy of proteins, with an emphasis on the concept of proteins as networks of charges.

We do not describe in detail theoretical methods for calculating the electrostatic free energies of proteins.[59] We also do not review the literature on experiments carried out using site-directed mutagenesis to measure the role of charged groups.[60] Excellent reviews of these topics exist already.[24,61]

2. Electrostatic Free Energies of Proteins

This section provides a short explanation of electrostatic contributions to the free energies of proteins. We start by describing the different definitions of the net charge of proteins, and discuss why the net charge is a useful parameter for characterizing the electrostatic properties of proteins. We classify individual charged groups on proteins in terms of their location, and of the character of their interactions with other charged groups. We introduce relevant basic ideas concerning proton equilibria, and describe charge regulation. We describe the simplest electrostatic interaction— two charges interacting with one another—and extend this description to the interaction between multiple charges; this description emphasizes that both enthalpy and entropy are important in determining the overall free energy of electrostatic interactions. We describe the contributions of solvation of charged groups to the electrostatic free energy of protein folding or binding to other molecules. Finally, we discuss how the electrostatic interactions contribute to both the strength and the specificity of molecular recognition events involving proteins.

2.1. Definitions of Net Charge of a Protein: Z_{H^+}, Z_{seq}, $Z_{protein}$, and Z_{CE}

Net charge is defined most generally as the sum over all charged groups that are covalently or tightly associated with a protein. We consider four definitions of net charge. Proton titrations measure the total number of protons bound to the protein as a function of the pH value and provide one measure of net charge, Z_{H^+}.[45] Proteins often change struc-

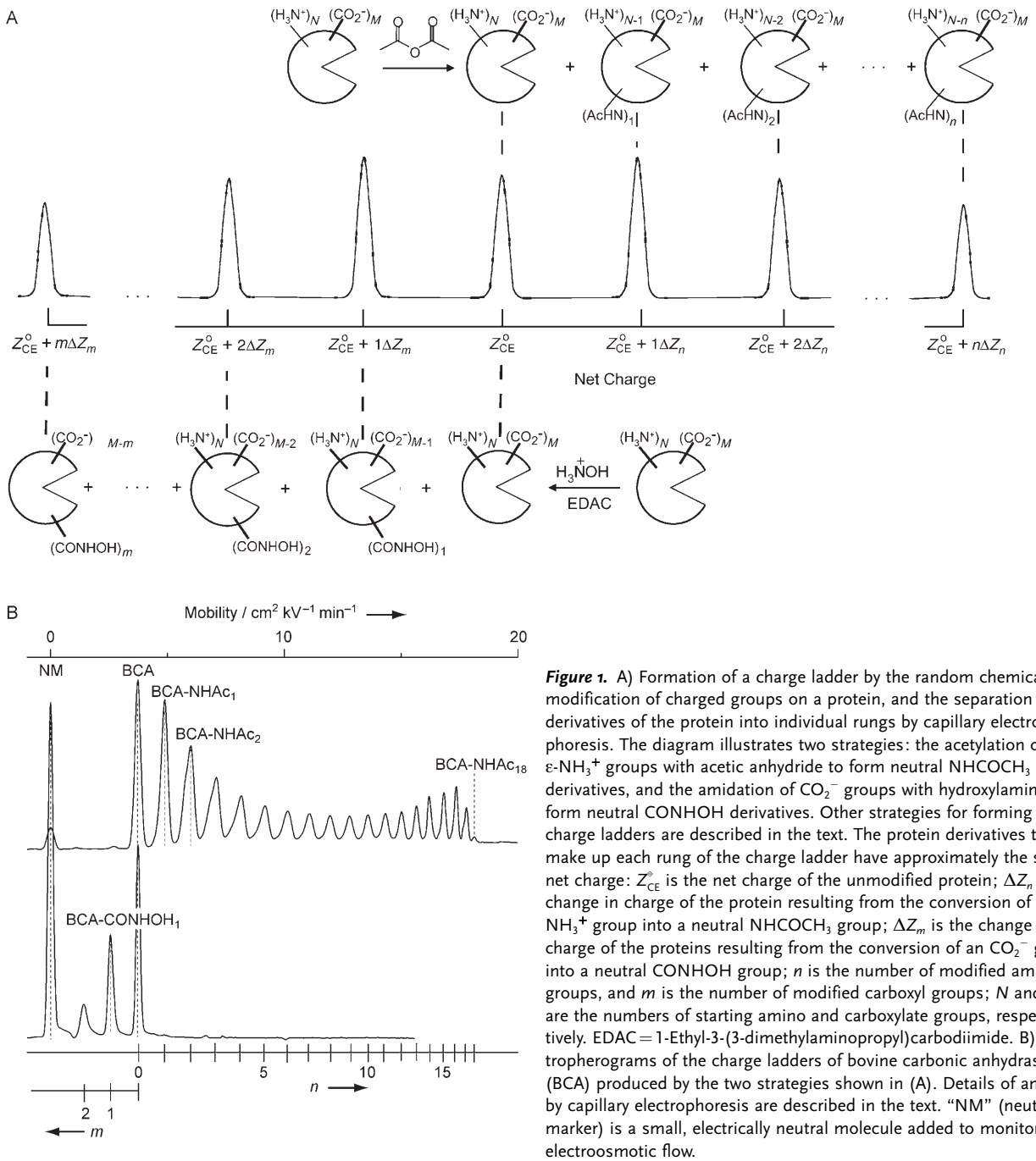


Figure 1. A) Formation of a charge ladder by the random chemical modification of charged groups on a protein, and the separation of the derivatives of the protein into individual rungs by capillary electrophoresis. The diagram illustrates two strategies: the acetylation of Lys ε-NH_3^+ groups with acetic anhydride to form neutral $NHCOCH_3$ derivatives, and the amidation of CO_2^- groups with hydroxylamine to form neutral CONHOH derivatives. Other strategies for forming charge ladders are described in the text. The protein derivatives that make up each rung of the charge ladder have approximately the same net charge: Z°_{CE} is the net charge of the unmodified protein; ΔZ_n is the change in charge of the protein resulting from the conversion of an NH_3^+ group into a neutral $NHCOCH_3$ group; ΔZ_m is the change in charge of the proteins resulting from the conversion of an CO_2^- group into a neutral CONHOH group; *n* is the number of modified amino groups, and *m* is the number of modified carboxyl groups; *N* and *M* are the numbers of starting amino and carboxylate groups, respectively. EDAC = 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide. B) Electropherograms of the charge ladders of bovine carbonic anhydrase II (BCA) produced by the two strategies shown in (A). Details of analysis by capillary electrophoresis are described in the text. "NM" (neutral marker) is a small, electrically neutral molecule added to monitor electroosmotic flow.

ture, aggregate, or denature when the pH value of the solution changes, and so values of Z_{H+} can be difficult (or impractical) to measure experimentally. Proton titrations also require large quantities of proteins. Net charge can also be estimated starting with the sequence of the protein (Z_{seq}), where the charge on each ionizable group is approximated using a standard value of its ionization constant (that is, its pK_a value) and the pH value of the solution. Values of Z_{seq} are usually inaccurate because they neglect both the roles of the local dielectric environment on the pK_a values, and the cooperativity inherent to proton equilibria in proteins.

Values of Z_{H+} and Z_{seq} also ignore the contribution of other charged species (e.g. buffer ions) that are noncovalently associated with the protein. The "true net charge" of a protein ($Z_{protein}$) is defined as the charge determined from the actual values of charge of the ionized groups of the amino acids making up the protein, and—when appropriate—any ionic species tightly associated with the protein; there has, in the past, been no practical method to measure $Z_{protein}$.

The combination of protein charge ladders and CE provides a method for the measurement of the net charge of proteins (Section 8). This approach measures the change in

electrophoretic mobility that results from incremental changes in the number of charged groups present on the surface of a protein and extrapolates to the number n of charged groups that would have to be added to or subtracted from the protein to make its electrophoretic mobility, and therefore its value of net charge, zero. If we know or can estimate the value of the change in charge ΔZ that results from the addition or subtraction of one fully charged group to the protein, then the value of net charge of the unmodified protein Z_{CE} is $-n\Delta Z$. The value of ΔZ depends on the group that is modified and the pH value and ionic strength of the solution. With an accurate estimate of ΔZ, we believe that Z_{CE} is closer to the true value $Z_{protein}$ than Z_{H+} and Z_{seq}. In Section 8, we describe several ways of estimating ΔZ.

2.2. Why is Net Charge Important?

The net charge of a protein is a fundamental physical property,[62] and its value directly influences the solubility,[22] aggregation,[63] and crystallization[64,65] of the protein. These influences can all be described in terms of the value of the second virial coefficient (B_{22}) of proteins; B_{22} is a measure of two-body interactions between solutes in a solution,[66] and can be determined experimentally by static light scattering, membrane osmometry, or self-interaction chromatography.[67] Positive values of B_{22} represent net repulsive interactions between proteins, while negative values represent net attractive interactions.

The net charge of proteins modulates the second virial coefficient.[62,68] It is known empirically that crystallization of proteins occurs most readily over a narrow range of negative values of B_{22}, that is, when there is a net attractive force between proteins in solution.[65,69,70] If the attraction becomes too strong (that is, the value of B_{22} is too negative), the proteins aggregate and precipitate, rather than crystallize. If the proteins have a net repulsion in solution (a positive B_{22} value), the solution is stable and crystallization does not occur. Doye et al.[71] argued that proteins have evolved with a negative design principle that avoids crystallization, because crystallization or aggregation would compromise the viability of the cell and cause disease.[72] Studies of the effects of mutations on the crystallizability of proteins indicate that Lys residues, which present positively charged primary ammonium groups, when present on the surface of proteins may play an important role in inhibiting crystallization.[64] It is not known at this time whether or not this inhibition of crystallization is primarily an effect of electrostatic interactions, or of some other property of the side chain of this amino acid.

The connection between net charge and second virial coefficient is not fully understood. Simple models of charged colloids that can predict values of the second virial coefficient as a function of net charge do not always accurately predict the behavior of proteins in solution.[73] This occurs in part because the models treat small ions in solution by using continuum electrostatic theory, which neglects the excluded volume and dispersion interactions of the ions.[51,74,75] The relationship between net charge, the second virial coefficient, the structure of the surface of a protein, and its tendencies toward solubility, aggregation, or crystallization is an area of active interest in structural biology and bioengineering.[70,76]

Both experiments and theoretical models suggest that net charge can influence the rates of protein–protein association[10,11,77,78] and enzymatic catalysis.[79,80] Long-range electrostatic interactions between the two molecules involved are thought to enhance their rate of association; this effect can result in rates of association that exceed the limits of diffusion.[80] The net charge of proteins can also influence their equilibrium properties: that is, the free energy of their interactions with ligands,[4,5] and their thermal stability.[56,81]

Net charge values are also important in the processes that lead to the separation and purification of proteins. Filtration,[82–85] chromatography,[86,87] and two-phase extraction[88] are all techniques that can be used to separate proteins on the basis of net charge. The net charge also influences the electrophoretic mobility of proteins in solution, a fact which we take considerable advantage of in the work described in this Review.

2.3. Classification of Charged Groups on Proteins

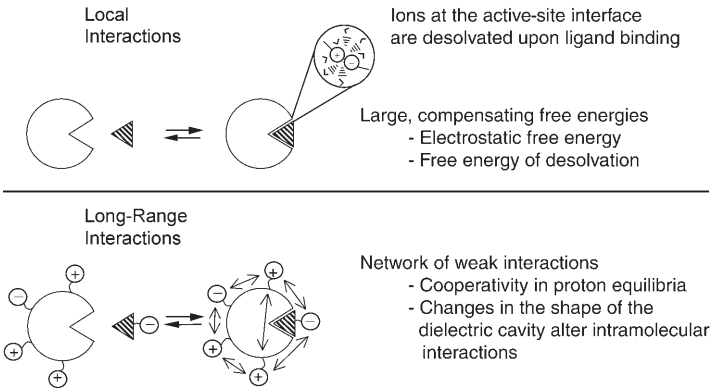
The charges on proteins result from the reversible exchange of protons with water and other acids or bases in solution. Table 1 lists ionizable groups encountered in proteins. Cofactors (e.g., Zn^{2+}, ATP, NADPH) or modifications of proteins (e.g., phosphorylation) must also be considered.

Table 1: Average and anomalous values of pK_a of ionizable residues encountered in proteins.

Residue	Average pK_a[a]	Anomalous residue in a protein	pK_a of that residue	Reason	Ref.
Asp	4.0	RNase T1: Asp76	0.5	buried, 3 intramolecular H bonds, 1 H bond to buried water	[91]
Glu	4.5	Staph.[c] nuclease: Glu66[b]	8.8	buried	[92]
His	6.4	T4 lysozyme: His31	9.1	involved in salt bridge with Asp70	[93]
Cys	9.1	papain Cys25	4.0	interacts with imidazole group of His159	[94]
Lys	10.4	Staph. nuclease: Lys66[b]	5.7	buried	[92]
Tyr	9.7	glutathione *S*-transferase A1-1	8.1	stabilized by conserved Arg and by electropositive ring edge of Phe10	[95]
Arg	12.0	n/a			
N-term. NH_3^+	7.8	n/a			
C-term. COOH	3.6	n/a			

[a] Ref. [2]. [b] Introduced by site-directed mutagenesis (Val66Glu and Val66 Lys); n/a: examples not found. [c] Staph. = Staphylococcal.

The type and location of the functional group is important (Scheme 1). Charged groups are located in the active site, on the surface, or in the interior of the protein. Charges may exist in pairs or groups, or may be isolated. We categorize interactions between charged groups in macromolecules as either local or long-range (Scheme 2), based on the distance between them. If two charged groups are sufficiently close (≤ 4 Å) that a water molecule cannot fit between them, we consider the interaction between those charges as local. If two oppositely charged groups with local interactions are in sufficient proximity that they can form a hydrogen bond, they are referred to as a salt bridge.[15] If two charges can accommodate water or other species between them, their interaction is long-ranged.



Scheme 2. Schematic diagram illustrating the difference between local and long-range interactions among charged groups on proteins.

The location of charged groups plays a role in determining the types of interactions in which they participate, and the contributions they make to the electrostatic free energy of the protein. For example, a buried salt bridge that is surrounded by the low dielectric interior of the protein has a very different electrostatic free energy than two charged groups on the surface of the protein that are exposed to water and dissolved ions. Table 1 illustrates the effects of the environment on the ionization of amino acids through examples of anomalous values of pK_a encountered in proteins; they can differ by as much as three units from the average values.

Another classification of charged groups is based on whether the interactions among them are pair-wise additive or cooperative. In general, the charge on one ionizable group influences the charge on other ionizable groups in its vicinity by shifting proton equilibria.[44] We describe cooperativity in proton equilibria in greater detail in Sections 2.4 and 2.5. This interdependence means that we cannot, in general, represent the total electrostatic free energy of a protein as the sum of pair-wise charge–charge interactions.

There are conditions, however, when this interdependence can be neglected: for example, when the charged groups are separated by large distances (e.g., >20 Å in water). This distance, of course, depends on the local dielectric environment as well as on the concentration of salts and other charged species in solution. Also, if the pK_a values of two groups differ both from each other and from the pH value of the solution by several (ca. 3) pH units, cooperativity is normally not important. For example, at pH 7, interactions between Arg groups on proteins (p$K_a \approx 12$) and phosphate groups on the backbone of DNA (p$K_a \approx 2$) can be treated as occurring between groups of fixed charge. Both groups are fully ionized at this pH value, and even if interactions between them were to shift their apparent pK_a values, their extent of ionization would be largely unaffected.

In addition to charges associated with ionized groups, the total electrostatic free energy of proteins must also account for the fact that many covalent bonds have significant dipole moments. In modeling studies of the interactions between charges and bond dipoles, it is often convenient to treat the dipole as comprising equal and opposite point charges (either whole units of charge or fractions) separated by the length of the bond. By using this convention, polar bonds and groups may be treated within the same system as charged groups. This approach is embodied in the development of a number of different sets of charge parameters[89,90] used in molecular simulations and continuum electrostatic models.

2.4. Proton Equilibria

This section starts with the simplest picture of proton equilibria: acid–base equilibria involving individual monovalent acids and bases. It then describes the simplest system—glycine—that demonstrates cooperativity of proton equilibria. Finally it extends this picture to describe the essential elements of proton equilibria of ionizable groups on proteins where there is cooperativity within a network of ionizable groups.

2.4.1. Proton Equilibria in Molecules with a Single Ionizable Site: Monovalent Acids and Bases

The average extent of protonation v [Eq. (1)] of an acid

$$v = \frac{[HA]}{[HA]+[A^-]} = \frac{K_a[H^+]}{1+K_a[H^+]} = \frac{1}{1+10^{(pH-pK_a)}} \quad (1)$$

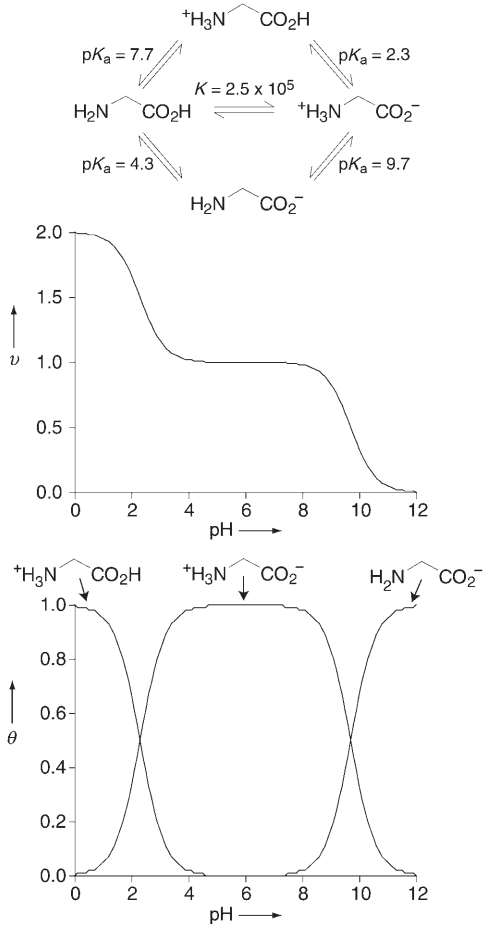
(HA) is determined by the dissociation constant $K_a = [H^+][A^-]/[HA]$[90] of the ionizable group, and by the pH value of the medium. The same equations describe proton equilibria for a monovalent base by replacing [HA] with [HB^+] and [A^-] with [B]. The charge associated with a particular ionizable group i as a result of proton equilibria is $Z = q - v$, where $q = 1$ if the molecule is a base, or $q = 0$ if it is an acid. The rate at which proton equilibrium is reached is normally fast relative to the rates of protein folding, ligand binding, and electrophoresis;[2] thus we assume that ionizable groups are always at thermodynamic equilibrium with protons (hydronium ions) in solution.

The value of v, and therefore the value of Z, depends on the difference between the pH and pK_a values. Changes in Z can be the result of a shift in either of these values.

2.4.2. *Proton Equilibria in Molecules with Two Ionizable Sites: Glycine*

Glycine contains both a carboxylic acid group ($pK_a = 2.3$) and an amine ($pK_a = 9.7$). For pH values between 4 and 8, the carboxylic acid group is assumed to be completely deprotonated and the amine fully protonated; under these conditions glycine is a zwitterion. Scheme 3, however, shows that in this range of pH values there are in principle four (and in practice, three) different species in equilibrium.[96]

At a fixed pH value, the extent of deprotonation of the carboxylic acid group is greater when the amino group is protonated than when it is deprotonated; analogously, the extent of protonation of the amino group is lower when the carboxylic acid group is protonated than when it is deprotonated. These changes in proton equilibria can be interpreted in two ways: 1) Differences in the extent of protonation are due to shifts in the pK_a values of the molecule in response to local values of electrostatic potential, which reflect the other charges on the molecule.[97,98] 2) Differences in extent of protonation result from local pH values that differ from bulk values, because electrostatic fields caused by the molecule locally concentrate or deplete protons. Both interpretations lead to the same prediction of the effects of cooperativity in proton equilibria, since local pH values also reflect the local value of electrostatic potential.



Scheme 3. Cooperativity of proton equilibria in glycine. The pK_a value of the CO_2H group is lower when the NH_2 group is protonated than when deprotonated. The plot of extent of protonation ν versus pH value shows two transitions. The plot of the relative populations θ of the different species as a function of the pH value shows that the transitions in the proton titration curve correspond primarily to ^+H_3N-CH_2-$CO_2H \rightleftharpoons {}^+H_3N$-$CH_2$-$CO_2^-$ (centered at pH ≈ 2.3), and ^+H_3N-CH_2-$CO_2^- \rightleftharpoons H_2N$-$CH_2$-$CO_2^-$ (centered at pH ≈ 9.7). Favorable electrostatic interactions between the NH_3^+ and CO_2^- groups increases the difference in the pK_a values between these groups, and the neutral species H_2N-CH_2-CO_2H is never populated significantly.

Scheme 3 implies there are five separate constants that describe proton equilibria in glycine. In fact, when we distinguish the two sites for protonation, there are only three independent equilibrium constants. If we do not distinguish the sites and just count the number of protons associated (as is done with proton titration), there are only two. To demonstrate these two cases, we use two different approaches to derive expressions for the extent of protonation υ.[48] In doing so, we also clarify the idea of cooperativity.

First we consider the approach that distinguishes the two sites for protonation: the carboxylate group and the amine. There are three possibilities for protons (H^+) to equilibrate with glycine (G): protonation only of the carboxylate (site a) with equilibrium constant K_a [Eq. (2)], protonation only of

$$G + H^+ \rightleftharpoons G_aH;\ K_a = \frac{[G_aH]}{[G][H^+]} \tag{2}$$

the amine (site b) with equilibrium constant K_b [Eq. (3)], or

$$G + H^+ \rightleftharpoons G_bH;\ K_b = \frac{[G_bH]}{[G][H^+]} \tag{3}$$

simultaneous protonation at both sites with equilibrium constant K_c [Eq. (4)].

$$G + 2H^+ \rightleftharpoons G_{ab}H_2;\ K_c = \frac{[G_{ab}H]}{[G][H^+]^2} \tag{4}$$

The extent of protonation (the average number of protons associated with a molecule of G as a function of $[H^+]$) is given by Equation (5). If the two sites are independent, then $K_c = K_aK_b$, and there is no cooperativity in the proton equilibria, and υ is defined completely by two equilibrium constants [Eq. (6)].

$$\upsilon = \frac{(K_a + K_b)[H^+] + 2K_c[H^+]^2}{1 + (K_a + K_b)[H^+] + K_c[H^+]^2} \tag{5}$$

$$\upsilon = \frac{K_a[H^+]}{1 + K_a[H^+]} + \frac{K_b[H^+]}{1 + K_b[H^+]} \tag{6}$$

When $K_c \neq K_aK_b$, the equilibria are cooperative: when $K_c > K_aK_b$, the cooperativity is positive and when $K_c < K_aK_b$, it is negative (sometimes referred to as anti-cooperative). A free energy that describes the interactions between sites (that is, the cooperativity) ΔG_c° is given in Equation (7). For the

$$\Delta G_c^\circ = -RT\ln\left(\frac{K_c}{K_aK_b}\right) \tag{7}$$

case of glycine, a model based on Coulomb's law may provide a reasonable estimate of ΔG_c°.

A shortcoming of the site-based approach is that values of K_a and K_b cannot be determined uniquely by fitting Equation (5) to titration data. The terms involving K_a and K_b in Equation (5) always appear as sums. Without an independent measure of K_a, K_b, and K_c (e.g. by NMR experiments that distinguish the extent of protonation of individual sites on the molecules), the proton equilibria can only be described by two equilibrium constants.

These two independent equilibrium constants can be obtained by using an approach based on the stoichiometry of the proton equilibria, where two protons can bind to glycine with equilibrium constants K_1 and K_2 [Eqs. (8) and (9)],

$$G + H^+ \rightleftharpoons GH;\ K_1 = \frac{[GH]}{[G][H^+]} \tag{8}$$

$$GH + H^+ \rightleftharpoons GH_2;\ K_2 = \frac{[GH_2]}{[GH][H^+]} = \frac{[GH_2]}{K_1[G][H^+]^2} \tag{9}$$

consistent with the stoichiometry. The extent of protonation as a function of $[H^+]$ is given by Equation (10).

$$\upsilon = \frac{K_1[H^+] + 2K_1K_2[H^+]^2}{1 + K_1[H^+] + K_1K_2[H^+]^2} \tag{10}$$

The stoichiometric approach, which does not distinguish between sites, requires only two parameters to describe the titration data of glycine. This approach yields values of $K_1 = 2\times10^2\,\text{M}^{-1}$ ($pK_1 = 2.3$) and $K_2 = 5\times10^9\,\text{M}^{-1}$ ($pK_2 = 9.7$).[96] Even though it is tempting to do so, it is not appropriate to assign these values explicitly to the carboxylic acid and amine groups. By comparing Equations (5) and (10), we see that $K_1 = K_a + K_b$ and $K_2 = cK_aK_b/(K_a+K_b)$ where $c = K_c/(K_aK_b)$. Scheme 3 gives $K_a = 5\times10^9\,\text{M}^{-1}$ and $K_b = 2\times10^4\,\text{M}^{-1}$. Therefore, $K_1 \approx K_a$ and $K_2 \approx cK_b$. By comparing the values of K_a, K_b, K_1, and K_2, we infer that $c = 10^{-2}$ and, from Equation (7), $\Delta G_c^{\circ} = 2.7\,\text{kcal mol}^{-1}$ at 25 °C. This value of free energy indicates a negative cooperativity that results from a repulsive electrostatic interaction between the sites (the positive charge associated with the ammonium group creates a potential at the carboxylate group that makes the protonation of this group less favorable than protonation in the absence of charge). These ideas can be extended to a molecule with an arbitrary number of sites for association by using the concept of binding polynomials, as pioneered by Wyman and Gill.[48,96]

2.4.3. Proton Equilibria in Molecules with Networks of Ionizable Groups: Proteins

When there are multiple ionizable groups on a molecule, the interdependence of proton equilibria described for the simple case of glycine in Section 2.4.2 results in a picture of the molecule as a network of ionizable groups. The quantitative model developed for glycine can be used to understand proton equilibria in proteins qualitatively.

When discussing proteins, it is useful to assign an average extent of protonation υ_i to each residue. Residues of a particular kind (e.g. Lys or Glu) in a protein do not necessarily all have the same values of υ_i at a given value of pH, or a value of υ_i equal to that of model compounds. These differences in the υ_i values are usually interpreted in terms of a model that postulates differences in pK_a values. The pK_a values are also influenced by solvent accessibility and by the presence of other charged and polar residues on the protein.

There are three important implications of this picture of proteins as networks of ionizable groups:

1) The pK_a values of ionizable groups in proteins can differ from model compounds by several pH units (see Table 1).
2) Not all ionizable groups of a particular kind (e.g., carboxylic acid groups of Glu) on a protein have the same pK_a value.
3) The pK_a values of of charged groups will, in general, be different in folded and unfolded protein,[99] as well as in free protein and protein bound to other proteins.[11]

These dependencies mean that the free energies of folding and binding can be linked thermodynamically to the pH value of the solution.

Lysozyme provides a good example of the link between proton equilibria and protein stability. This protein has a maximum in stability at about pH 7. As the solution is made more acidic, the protein becomes less stable, and denatures spontaneously when the pH value is less than 2 (at 25 °C).[97] Denaturation under either acidic or alkaline conditions has been documented for many proteins, with detailed studies conducted with lysozyme,[93,100] various ribonucleases,[101] α-lactalbumin,[102] myoglobin,[18,103] and cytochrome *c*.[103,104]

In all of these cases, differences in proton equilibria between the native and denatured states influence (and may determine) the dependence of the stability on the pH value.[21] Section 10 shows how charge ladders and CE can determine the difference in the number of protons associated with α-lactalbumin in the native and denatured states, and thus can predict the dependence of the stability of this protein on the pH value of the solution.

2.5. Charge Regulation Is a Useful Model for the Network of Ionizable Groups on a Protein

A change in the extent of the protonation of one group of a molecule can cause an adjustment in the extent of protonation of the other ionizable groups on this molecule. This process is called "charge regulation"[105–107] and reflects the fact that the charges comprise a network. Linderstrøm-Lang first modeled this phenomenon in the context of pH titrations of proteins.[108] Linderstrøm-Lang's model treats a protein as a sphere with the charge distributed uniformly over its surface. A change in the extent of protonation of one residue changes the density of the charge on the sphere, and thereby the electrostatic potential at the surface of the sphere. This change in potential shifts the state of protonation of the other ionizable groups. In this model (Section 8.1), the net charge is determined by iteratively calculating the protonation state of all the residues on the protein sphere. This process makes the Linderstrøm-Lang model a mean-field model. Despite its lack of molecular detail, this model is

qualitatively useful for understanding charge regulation in proteins.

As in glycine, the shift in the protonation state of one group because of another can be interpreted equivalently by either of the two mechanisms: 1) a change in the local proton concentration, or 2) a shift in the pK_a values of ionizable residues. From the expression for equilibrium of a single acidic group, $K_a/[H^+] = [A^-]/[HA]$, the change in the ratio of ionized to unionized acid can be interpreted either as a change in K_a or in the local value $[H^+]$.

Other, more complex models of proton equilibria exist that use more realistic representations of the protein.[109] These representations explicitly include the positions of ionizable groups, based on the crystal structure of the protein, and calculate the net charge of ionizable groups by averaging over the ensemble of different possible protonation states of the protein, rather than by using the mean-field approach of Linderstrøm-Lang.

Charge regulation is relevant to the analysis of protein charge ladders. Acetylation of a charged ε-NH_3^+ group of Lys removes a positive charge, and thus perturbs the network of charges on the protein. Other ionizable residues will adjust their average extent of protonation in response to this perturbation. For quantitative analysis of experiments with charge ladders, we need to know the change in the net charge of the protein ΔZ that results from the annihilation of charge on a particular group on the protein. Our initial work on charge ladders[4,110,111] did not consider charge regulation, and assumed $\Delta Z = -1$: that is, we assumed the change in net charge of the protein to be just the change in the charge of the particular group that was modified. Menon and Zydney pointed out that this assumption was wrong.[105] Estimates by Menon and Zydney, and subsequently by us,[106,107] showed that values of ΔZ for the acetylation of Lys ε-amino groups ranged from −0.6 to −0.95 (depending on pH value, ionic strength, and position of acetylation) and suggested that the error in net charge associated with earlier analyses (assuming $\Delta Z = -1$ at pH 8.4) was approximately 10%.

2.6. Enthalpic and Entropic Contribution to the Electrostatic Free Energy

The aim of this section is to describe the electrostatic free energy of proteins, starting from a given distribution of charges. We begin with a simple description of the interactions between two point charges and extend this discussion to proteins as arrays of point charges.

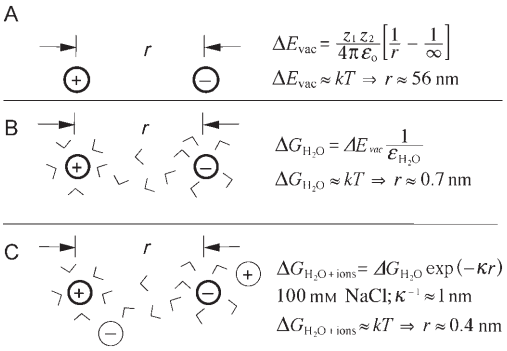
The models used in this section to predict electrostatic free energy, enthalpy, and entropy of charge–charge interactions are highly simplified. We use Coulomb's law and Debye–Hückel theory to calculate the work done in taking two point charges from infinity to some finite distance of separation. Other researchers have developed more rigorous approaches to describe the interactions of charges in solution by using statistical mechanics.[112,113] These approaches, which are part of a more general theory of noncovalent binding,[114] predict equilibrium association constants between ions, from which standard-state free energies can be determined. Although our models are simplified relative to these more rigorous approaches, the conclusions reached are essentially the same:[113] that is, both enthalpy and entropy contribute in unexpected ways to the free energy of interaction between charges in water.

2.6.1. Interactions between Two Charges

Coulomb's law in a vacuum [Eq. (11)] describes the work

$$\Delta E_{vac} = \frac{e^2 z_1 z_2}{4\pi\varepsilon_0(1/r_0 - 1/\infty)} \tag{11}$$

required to bring two point charges from infinity to a finite separation r_0 in a vacuum (ε_0 is the dielectric permittivity of a vacuum; Scheme 4A). Equation (11) gives the difference in energy (in this case, enthalpy; Coulomb's law in a vacuum does not describe entropy) between charges at $r = \infty$ and $r = r_0$.



Scheme 4. Diagram describing differences in electrostatic energies in vacuum (A), water (B), and in the presences of dissolved ions (C). The free energy of interaction between two, singly charged groups in water is smaller than between isolated charges in a vacuum because of the presence of water with a large dielectric constant and ions that tend to shield interactions. κ is the inverse Debye screening length.

There is no temperature dependence to this interaction; it is not a free energy. Values of ΔE_{vac} are favorable (negative) if the two charges are opposite in sign, and unfavorable if they have the same sign. The value of ΔE_{vac} approaches the value of thermal energy (RT) when $r \approx 56$ nm at 298 K. Clearly, electrostatic interactions are long-range in a vacuum relative to the dimensions of a protein; a typical protein has a diameter of 5 nm.

In a uniform dielectric with dielectric constant ε, the Coulombic interaction energy is just $\Delta E_{vac}/\varepsilon$ (Scheme 4B). Since values of ε are always greater than 1, the effect of the medium is to reduce the magnitude of the interaction. The value of ε is dependent on temperature, and thus the energy of interaction is now a free energy: for example, in water $\Delta G_{H_2O} = \Delta E_{vac}/\varepsilon_{H_2O}$, where $\varepsilon_{H_2O} \approx 80$ at 25 °C. The temperature dependence of ε reflects the fact that enthalpic and entropic contributions are included in the work of polarizing the medium by the charges. In this discussion, we ignore all the

molecular details of water. We treat it as a uniform continuum described by a single (temperature-dependent) parameter ε: the larger the ε value of the medium, the smaller in magnitude is the free energy of interaction for a particular value of r. For water, ΔG_{H_2O} approaches the value of thermal energy (RT) at 298 K when $r \approx 0.7$ nm. The large dielectric constant of water plays a significant role in establishing the magnitude of the electrostatic free energy of interactions between charged groups in biological systems.

A charge in a solution containing other dissolved ions tends to repel ions of the same charge, and to attract ions of the opposite charge. Therefore, the local concentration of oppositely charged ions in the vicinity of a charged group is higher than that in the bulk and the concentration of like-charged ions is lower than in the bulk. The electrostatic potential produced by the charge is therefore weakened, or "screened" (Scheme 4C).

The Debye–Hückel model describes the reduction in the free energy of interaction between charges resulting from the presence of dissolved ions. In this model, the free energy of interaction is $\Delta G_{H_2O+ions} = \Delta G_{H_2O}\exp(-\kappa r)$, where κ^{-1} is the characteristic distance of the screening of the potential, known as the Debye length. Values of κ are temperature dependent, and reflect the work done (involving both enthalpic and entropic contributions) in the creation of a localized concentration gradient of ions surrounding the two charges (that is, the creation of an "ion atmosphere"). As in the case of the dielectric constant, Debye–Hückel theory ignores all the molecular details of the ions—they are treated as a continuum charge density described by a single (temperature-dependent) parameter κ. A typical ionic strength of a biological solution is 150 mM. Under these conditions, $\kappa^{-1} \approx 1$ nm and $\Delta G_{H_2O+ions}$ approaches the value of the thermal energy (RT) at 298 K when $r \approx 0.4$ nm.

2.6.2. Enthalpy and Entropy of Interaction between Two Ions

The free energy of interaction between charged groups in a vacuum is entirely enthalpic: that is, this value is independent of temperature. However, the free energy of interaction between charged groups in water is often dominated by entropy. Changes in entropy reflect the way in which water and ions in solution reconfigure in response to the presence of the charges. These effects are described by the dielectric constant (ε) and by the Debye length (κ^{-1}).

Two charges interact in a uniform dielectric with a free energy described by Equation (12). Since the dielectric

$$\Delta G = \frac{e^2 z_1 z_2}{4\pi\varepsilon_0 \varepsilon r} \tag{12}$$

constant depends on temperature, the interaction described by Equation (12) is a free energy. The entropy is defined in Equation (13), and the enthalpy of interaction by Equation (14). Values of $(\partial\varepsilon/\partial T)$ are typically less than zero:

$$\Delta S = -\left(\frac{\partial \Delta G}{\partial T}\right)_{p,N} = \Delta G\left(\frac{1}{\varepsilon}\frac{\partial \varepsilon}{\partial T}\right) \tag{13}$$

$$\Delta H = \Delta G + T\Delta S = \Delta G\left(1 + \frac{T}{\varepsilon}\frac{\partial \varepsilon}{\partial T}\right) \tag{14}$$

$(T/\varepsilon)(\partial\varepsilon/\partial T) \approx -0.26$ for hexane, and ≈ -1.37 for water at 25 °C.[115] Therefore, $\Delta H \cong 0.74\Delta G$ for hexane, but $\Delta H \cong -0.37\Delta G$ for water: that is, at all distances the enthalpy of interaction between oppositely charged ions in the lower dielectric hexane is favorable ($\Delta H < 0$), while the enthalpy of interaction in water is unfavorable ($\Delta H > 0$). For both cases the entropy of interaction is favorable ($\Delta S > 0$).

This result is surprising: we usually think about a statement "unlike charges attract" in terms of the favorable enthalpic component of the interaction. In fact, in water, unlike charges do attract—but not because of enthalpic interactions between ions; rather, they attract because the entropy of ion–solvent interactions (as described by the temperature-dependence of the dielectric constant of water) becomes increasingly more positive (favorable) as ions approach one another. This result also indicates that bringing two ions of the *same* charge together in water releases heat ($\Delta H < 0$) even though the free energy increases.[116] These characteristics are the result of four unique properties of water: its large ε value, its large and negative value of $\partial\varepsilon/\partial T$, its small molar volume, and its ability to form a network of hydrogen bonds in the liquid state.

Equation (15) gives the free energy of interaction of two

$$\Delta G = \frac{e^2 z_1 z_2}{4\pi\varepsilon_0 \varepsilon r}\exp(-\kappa r) \tag{15}$$

charges in a uniform dielectric that contains dissolved ions, as predicted by the Debye–Hückel model with the concentration of ions described by κ. Equations (16) and (17) give the

$$\Delta S = \Delta G\left(\frac{1}{\varepsilon}\frac{\partial \varepsilon}{\partial T}\right) \tag{16}$$

$$\Delta H = \Delta G\left(1 + \frac{T}{\varepsilon}\frac{\partial \varepsilon}{\partial T} + rT\frac{\partial \kappa}{\partial T}\right) \tag{17}$$

entropy and enthalpy of interaction predicted by this model. The values of $(\partial\kappa/\partial T)$ are also less than zero, and therefore also contribute to the favorable entropy of interaction between oppositely charged groups.

Figure 2 compares the contributions to the free energy of interaction between two singly charged groups as they are brought from infinite to finite separation distance r in pure hexane (a low dielectric medium: $\varepsilon = 2$) and in water ($\varepsilon = 80$) containing 100 mM NaCl. In hexane, it is the favorable enthalpy of interaction that makes the largest contribution to the favorable electrostatic free energy of interaction of opposite charges as they approach the point of contact. In water with dissolved ions, the enthalpy of interaction between opposite charges is actually unfavorable (the enthalpy includes the contributions of the polarized water and ion atmospheres surrounding the charges, as well as the interaction between the charges themselves). It is instead the

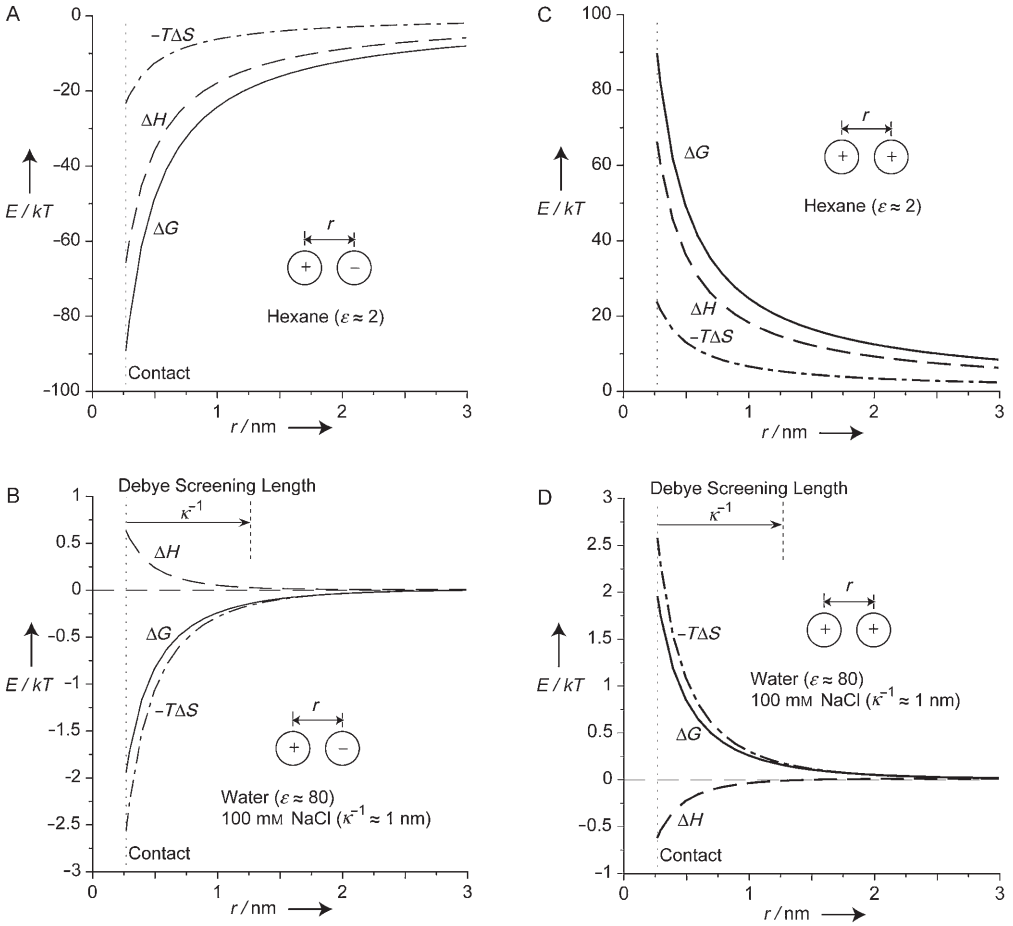


Figure 2. The electrostatic free energy, enthalpy, and entropy of interaction between two ions of the opposite (A, B) and of the same (C, D) charge in hexane (A, C) and water (B, D) containing 100 mM NaCl depends on the distance between the ions. Values of the free energy of interaction in this simplified, isothermal, one-dimensional system (ΔG, in units of kT) were calculated at 25 °C using Equations (12)—(14) for hexane, and Equations (15)–(17) for water with salt.

entropic contribution to the free energy that makes the interactions between opposite charges favorable.

This conclusion—that plus and minus charges do attract in aqueous solutions, but for entropic and not enthalpic reasons—is one of the almost endless, counterintuitive properties of water. The answer to the question "Why?" is "because the value of the dielectric constant decreases steeply with increasing temperature," which is more a statement of experimental fact than molecular-level explanation. The dielectric constant reflects the static polarization of water, both between the plates of a capacitor and around an ion. The structure of the water, which influences this polarizability, becomes disordered as the temperature increases, and releases water molecules as ions approach one another.

The nature of electrostatic interactions in water, and the partitioning of these interactions between enthalpic and entropic terms, is one surprise about water. There are others. For example, Ninham and co-workers have shown that electrostatic interactions do not always dominate the total free energy of interactions between ions. As the concentrations of ions increases, dispersion forces among ions and water have to be included in accurate thermodynamic descriptions of solutions of electrolytes.[75] A striking, counterintuitive result of the study by Ninham and co-workers is that the net interaction between some ions of like charge is attractive when dispersion forces are included.[75] For example, in a 1 M aqueous solution of $NaNO_3$, the ion–ion pair distribution function shows that NO_3^- ions prefer to be next to other NO_3^- ions rather than Na^+ ions; Na^+ ions, however, do not show an analogous preference.

2.7. From a Pair of Charges to Proteins

There are two ways of describing the distribution of charges on a protein mathematically: One, which can be applied to whole proteins or to individual side chains, is through a multipole expansion. In this approach, the electrostatic free energy of interaction between two molecules is represented as a sum of charge–charge, charge–dipole, dipole–dipole, and higher-order interactions. The multipole expansion requires evaluations of large numbers of terms, and is cumbersome when describing large molecules such as proteins.

An alternative to using the multipole expansion focuses on the density of charge $\rho(\mathbf{r})$ [$C\,m^{-3}$]. If all the charge is localized on individual atoms, then $\rho(\mathbf{r})$ corresponds to an array of point charges, with z_i as the charge associated with atom i. Values of z_i can be whole or fractional units of charge. The total electrostatic free energy of a particular distribution of charge is the work done in assembling an array of charges from a reference state of infinite separation [Eq. (18)], where

$$\Delta G = \sum_i z_i \varphi(\mathbf{r}_i) \tag{18}$$

the sum is over all charges in the array, and $\varphi(\mathbf{r}_i)$ is the electrostatic potential at the position $\mathbf{r}_i$ ar a consequence of the other charges in the array.

The problem of calculating the electrostatic free energy is that of calculating the electrostatic potential $\varphi(\mathbf{r}_i)$. This calculation can be done for proteins by using the Poisson–Boltzmann equation, which is described briefly in Section 3.2.

2.7.1. The Free Energy of Solvation of Charges

Interactions between charges during protein folding, ligand binding, and protein–protein interaction often involve changes in their solvation. The free energy of transferring a charged group from a medium with a larger value of ε to one with a smaller value of ε is always positive (that is, unfavorable);[48,117] this effect is often referred to as desolvation in biophysics, as it typically occurs when a charged group moves from a more-solvated to a less-solvated state.

The free energy of solvation of a charged group is the result of the polarization induced in the dielectric medium surrounding the group.[48] The polarization produces a local electric field (called the "reaction field", or sometimes the "self field"). The larger the value of ε, the greater is the magnitude of the reaction field. The free energy of interaction between a charged group and its reaction field is always negative (that is, favorable). This free energy of solvation is a major driving force in the dissolution of ionic solids in water.

The transfer of a charged group from a region with a large ε value to a region with a smaller ε value reduces the magnitude of the reaction field of the ion, and results in a positive free energy of transfer. These changes in solvation can have interesting consequences on protein stability and activity. For example, in the absence of a ligand, charges on the surface of the active site of a protein are exposed to solvent. The binding of a ligand to the protein expels solvent from the interface; the charged group is transferred from an environment with a high dielectric constant ($\varepsilon \approx 20$–80) to the interior of the complex where the dielectric constant is much lower ($\varepsilon \approx 4$). The consequence of this transfer is that charged groups in active sites do not always contribute to the strength of the interaction between the protein and the ligand. Davis et al.[8] argued in a study of the binding of two cell-surface proteins CD2 and CD48 that charged residues on CD2 contribute little to the *affinity* of binding of CD48 because of the trade-off between the unfavorable energy of desolvation of charges and the favorable energy of forming electrostatic contacts. Instead, these charged residues contributed primarily to the *specificity* of interactions because of electrostatic complementarity of the two surfaces.

Desolvation plays a major role in the energetics of buried charges. The formation of a buried salt bridge can actually be destabilizing if the favorable free energy of electrostatic interactions between two oppositely charged groups is smaller than the unfavorable free energy of desolvation of the individual charges. Hendsch and Tidor estimated this effect in protein folding.[17] Calculations based on continuum electrostatic theory indicated that the majority of buried salt bridges are destabilizing in the folded state, relative to interactions between amino acids that are the same size, form hydrogen bonds, but are not charged; this destabilization by salt bridges is due to the large penalty of desolvation.

From the work of Hendsch and Tidor[17] as well as others,[15,118] it is clear that the free energy of desolvation is similar in magnitude to the free energy of electrostatic interactions in the formation of buried salt bridges. The net free energy for the formation of buried salt bridges is still uncertain, because these interactions offset each other. For example, Baker and co-workers[118] have argued that the free energy of electrostatic interactions between charges more than compensates for the free energy of desolvation in most cases. Kumar and Nussinov[15] surveyed the free energy of formation for a large number of salt bridges predicted by continuum electrostatic calculations, and suggested, based on their calculations, that only a minority (34%) of buried salt bridges were destabilizing. Kumar and Nussinov attributed the discrepancy between their work and the work of Hendsch and Tidor[17] to differences in the distance of charge–charge interactions used to define a salt bridge: while the study of Hendsch and Tidor included many salt bridges with separation distances greater than 4 Å, Kumar and Nussinov defined salt bridges as having separations of ≤ 4 Å.

2.7.2. The Influence of Ionic Strength on Electrostatic Interactions in Proteins

It is a common belief that increasing the ionic strength of a solution minimizes the free energy of the electrostatic interactions by shielding charged groups by the Debye layer. In proteins, however, the electrostatic free energy is not necessarily reduced by ionic strength for three reasons:

1) The energy of desolvation is only weakly dependent on ionic strength. If desolvation dominates the contribution of the electrostatic interactions to the free energy, then the net free energy will be fairly insensitive to the ionic strength.[17]
2) The distances between many charged groups in proteins (especially those involved in salt bridges) are significantly smaller than the Debye screening lengths ($\kappa(I{=}100\text{ mM}) \approx 1$ nm; I = ionic strength); the effect of changing the screening length on the free energy of the interaction is therefore insignificant for these groups.
3) Charges on the surface of a protein also interact through the low dielectric interior of the protein; free ions in solution should have little effect on this type of interaction.

2.7.3. Conclusions

The charged groups on a protein constitute a network with interdependent, cooperative proton equilibria. The variety of environments in which charged groups are found—on the surface or in the interior, proximal to other charges or not—make it impossible to generalize the contributions of electrostatic interactions to many processes in proteins. As a result, experimental work and new experimental methods are required.

3. Methods of Studying Electrostatic Interactions in Proteins

3.1. Experimentally, by Site-Directed Mutagenesis

The most common strategy for exploring the contribution of charged groups to the stability and activity of proteins is site-directed mutagenesis. This approach identified key amino acids that contribute substantially to stability or activity.[119,120] Double mutant cycles have also been used to estimate the free energy of interactions between two charged groups on a protein[121,122] or between two proteins.[11] This approach focuses on local interactions, and identifies the roles of particular charged residues—often the ones in the active site, at the area of contact with another macromolecule or a ligand, or forming a salt bridge.

Fersht and co-workers pioneered the use of site-directed mutagenesis to examine the role of charged groups in proteins with studies on barnase and subtilisin.[2,121,123] A classic example is the work of Anderson et al. on the role of charged groups on the stability of T4 lysozyme.[93] This study demonstrated that a specific salt bridge (Asp70-His31) contributed 3–5 kcal mol^{-1} to the stability of the protein. Dao-pin et al.[122] measured the stability of 13 different single, double, triple, and quadruple mutants of T4 lysozyme produced by site-directed mutagenesis, and concluded that long-range electrostatic interactions contributed, on average, little to stability. In another example, which reached a different conclusion, Perl et al. replaced a single negatively charged amino acid (Glu) on the surface of a mesophilic cold-shock protein by a positively charged amino acid (Arg); this mutation increased the thermal stability of the protein by more than 2.8 kcal mol^{-1}, and rendered the mutated mesophilic protein nearly equal in stability to its thermophilic counterpart.[119] Analysis of the structure of these proteins showed this change in stability was not due to the formation of an ion pair involving Arg and a specific negatively charged group on the protein, but a result of many long-range, but generally favorable interactions. Significant changes in stability can thus result from the change in the electrostatic free energy of interactions on mutation of a single charged amino acid, but understanding and predicting these changes may not be simple.

Benkovic and Hammes-Schiffer used site-directed mutagenesis to explore the mechanisms of the catalytic activity of dihydrofolate reductase.[124] This elegant, extensive work demonstrated that residues that are neither in direct contact with ligands in the active site, nor able to make contact after substantial conformational rearrangements, are still able to influence catalytic activity. The authors concluded that these conserved residues act as a coupled network that determines protein structure and, perhaps more importantly, coordinates intramolecular motions of the protein central to its catalytic activity.

Site-directed mutagenesis has two major limitations. First, the expression and purification of protein mutants is time consuming. Second, the individual or pair-wise interactions targeted by this technique can be too weak to detect or quantify. As we describe later, CE and charge ladders circumvent some of these limitations (while introducing others).

3.2. Theoretically, by Calculations of Continuum Electrostatics

The Poisson–Boltzmann equation is a second-order, nonlinear differential equation that relates values of electrostatic potential to the density of charge that is embedded in a nonuniform dielectric continuum. In this equation, values of electrostatic potential, charge density, and dielectric constant are all functions of position. Numerical solutions of the Poisson–Boltzmann equation, first applied to proteins by Warwicker and Watson[125] and then by Honig and co-workers,[59,126] are the method of choice for calculating electrostatic free energies between proteins and between charged groups on a protein, for representing potential surfaces, for calculating free energies of solvation of charged groups as well as the pK_a values of ionizable residues, and for other applications.[24,26] This approach is semimacroscopic: the distribution of charges on the macromolecule is treated explicitly on the basis of crystallographic data, while the solvent is treated as a continuum, with a density of charge to represent dissolved ions.[24] The protein is represented as a cavity of low dielectric constant—either uniform or as a function of position[47]—embedded in a high dielectric medium.

Numerical solutions of the Poisson–Boltzmann equation predict the electrostatic potentials present at protein surfaces, and can clarify the role of electrostatics in protein–protein and protein–nucleic acid interactions.[24] For example, trypsin and bovine pancreatic trypsin inhibitor form a strong complex, even though both proteins have a large positive net charge. The Poisson–Boltzmann equation revealed a local domain of negative potential at the binding site of trypsin that probably contributes to the affinity of the two proteins.[24]

Solutions to the Poisson–Boltzmann equation are computationally intensive, and therefore cannot be easily coupled to dynamic simulations. There are also uncertainties in the parameters of the model (e.g. atomic radii that define the boundary between the protein and solvent, dielectric constants of the protein and solvent at the interface, and the "intrinsic" pK_a values used to determine proton equilibria).

Calculations using the Poisson–Boltzmann equation indicate that many electrostatic interactions that are likely to be encountered in proteins and complexes of proteins and ligands are sufficiently weak that they are currently undetectable by experiment. As an example, we calculated the

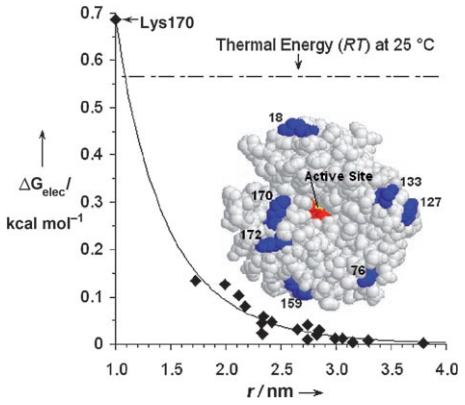


Figure 3. The electrostatic free energy ΔG_{elec} of interactions between the positive charge on each Lys ε-NH_3^+ group of human carbonic anhydrase II (HCAII), and a positive "test" charge situated at the active site of the enzyme, as a function of distance r between the charges. The electrostatic free energy is calculated using numerical solutions of the linearized Poisson–Boltzmann equation. The solvent surrounding the protein is modeled with a dielectric constant of $\varepsilon = 80$; the interior of the protein is modeled with $\varepsilon = 4$. The solid line is a fit of a screened Coulombic potential [Eq. (28) in the text]) with both ε and κ varied to fit the data. The inset shows the structure of HCAII: the active site is colored red with the Zn^{2+} ion in yellow; the 7 of the 21 total Lys residues that are visible in this view are colored blue; the number indicates their position in the sequence of the protein.

electrostatic free energy of interaction between each Lys ε-NH_3^+ residue of carbonic anhydrase and a positive "test" charge situated at the active site of the enzyme (Figure 3). All but one interaction are below the thermal energy (RT).

4. Protein Charge Ladders—A New Approach to Probing Networks of Interactions among Charged Groups on Proteins

Protein charge ladders are formed by chemical modifications of charged groups which alter the charge state of that group, (e.g. -NH_3^+→NHCOCH$_3$; -CO_2^-→-CO_2CH_3). The reactions are, we believe, generally relatively nonspecific in their position—most or all of the Lys residues in a protein may, in principle, react with similar probability. As a result, the reaction mixture contains protein derivatives with different numbers and sites of modifications. The mixture of the protein derivatives can be separated into distinct bands by capillary electrophoresis (CE). CE readily measures and separates molecules on the basis of electrophoretic mobility (μ [$m^2 V^{-1} s^{-1}$])—a property that depends on the ratio of two fundamental biophysical characteristics of proteins: net charge and hydrodynamic drag.[127] Each peak consists of a mixture of regio-isomers of approximately the same net charge.

The combination of protein charge ladders and CE makes it possible to measure or estimate the electrostatic properties and interactions of proteins. The following sections review the synthesis, characterization, and the use of charge ladders and CE to examine the following properties of proteins: 1) the net charge, 2) the extent of charge regulation upon acetylation of a Lys residue, 3) the electrostatic contributions to the free energy of binding of ligands, 4) the electrostatic contributions to the free energy of folding and stability, and 5) the role of the charge of a protein and of the electrostatic interactions in bioprocessing (in particular in ultrafiltration). Two other subjects that we review that do not fall in to the realm of electrostatics, but can nevertheless be addressed using charged ladders and CE, are: 1) determination of the hydrodynamic drag of proteins, and 2) the effects of surface charge on the pattern of protein ionization in electrospray mass spectrometry.

5. Model Proteins in Studies with Charge Ladders

The development of a new technique often requires a system that is straightforward to handle. In protein biophysics, the models used have included barnase[128] for folding, acylphosphatase for studies of in vitro aggregation and amyloid formation,[129] and triosphosphate isomerase and dihydrofolate reductase[124] for catalysis. We have used carbonic anhydrase (CA, EC 4.2.1.1) as the model protein in our studies. Both the bovine CA (BCA) with 18 Lys ε-NH_3^+ groups and the human CA (HCA) with 21 Lys ε-NH_3^+ groups are commercially available and their structures are determined. Both proteins have naturally acetylated N-terminal α-amino groups. CA is stable (T_m(HCA) = 60 °C; T_m(BCA) = 65 °C)[130] and therefore easy to handle. More importantly, the charge ladders of CA are well-resolved by CE (see Figures 1b and 4), and acetylation of CA does not disrupt its tertiary structure.[131] CA does not interact with the walls of the capillary when the electrophoresis buffer has a pH value of greater than 7.5. In addition, a variety of well-characterized inhibitors to CA (almost all aryl sulfonamides) are commercially available or easy to synthesize.[132] We have also used insulin,[133] lysozyme,[107,134] and α-lactalbumin[56] in studies with charge ladders. Many proteins other than these few form well-resolved charge ladders,[54] but we have not carried out detailed analysis of their electrostatic properties by using this technique.

6. Synthesis and Characterization of Protein Charge Ladders

6.1. Preparation of Protein Charge Ladders

The reactions used to form charge ladders convert residues that are charged at the pH value of interest (e.g the Lys ε-NH_2 group, which is ε-NH_3^+ at physiological pH values) into neutral species; certain modifications—for example, acylation of amines with succinic or glutaric anhydride—reverse the charge (Table 2).[19,54] When the reactions are not carried to completion, they result in statistical mixtures of derivatives of the protein; these derivatives differ in the extent of modification, and thus in the number of charged groups. We refer to the set of derivatives with the same number of modifications as a "rung" of the ladder. The

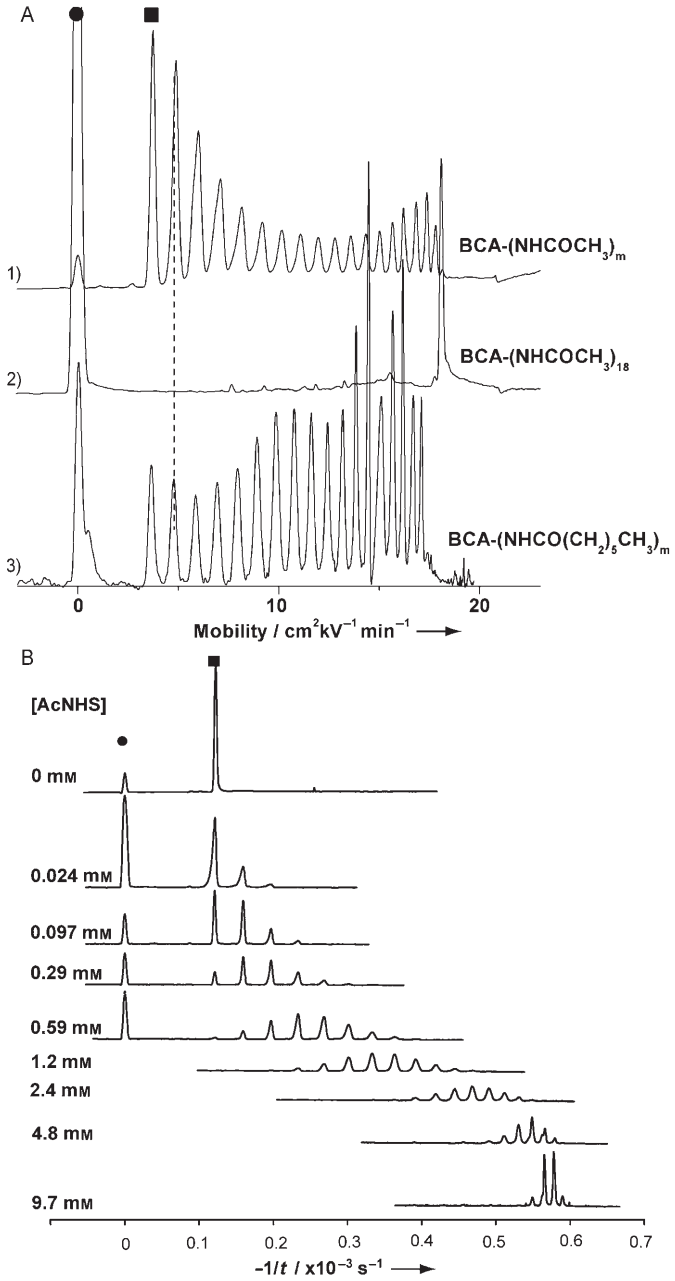
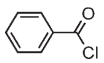
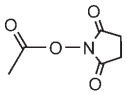
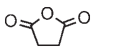
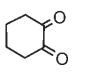
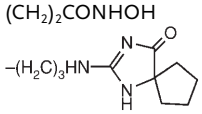


Figure 4. A) Electropherograms of charge ladders of bovine carbonic anhydrase II. 1) Charge ladder of BCAII generated with acetic anhydride (BCA-$(NHCOCH_3)_m$). 2) The reaction can be driven to completion to generate peracetylated protein (BCA-$(NHCOCH_3)_{18}$). 3) Charge ladder generated with hexanoic anhydride (BCA-$(NHCO(CH_2)_4CH_3)_m$). ● indicates native protein, and ■ indicates an electrically neutral marker (mesityl oxide or *p*-methoxybenzyl alcohol) used to monitor electroosmotic flow. A dashed vertical line indicates that the mobilities of the first rung of the acetyl and hexanoyl ladder are indistinguishable. Late rungs of the hexanoyl ladder show the influence of (probably) increasing drag. All separations are done in 25 mM Tris/192 mM Gly buffer on a capillary 47 cm long. B) High-resolution electropherograms of BCAII charge ladders generated with the *N*-hydroxysuccinimide ester of acetic acid. The acetylation reactions were conducted in borate buffer of pH 8.56 and analyzed in D_2O-based Tris-Gly buffer on a capillary 117 cm long. The scale of $-1/t$ is proportional to mobility μ [Eq. (30)]. (Adapted from Ref. [137].)

Table 2: Examples of possible modifications of the Lys, Asp, Glu, and Arg residues of a protein that result in a change in the charge of the residue.

Residue	Reagent	Product	ΔZ
$-(CH_2)_4NH_3^+$	$(CH_3CO)_2O$	$-(CH_2)_4NHCOCH_3$	−1
$-(CH_2)_4NH_3^+$	$(CH_3(CH_2)_4CO)_2O$	$-(CH_2)_4NHCO(CH_2)_4CH_3$	−1
$-(CH_2)_4NH_3^+$		$-(CH_2)_4NHCOC_6H_5$	−1
$-(CH_2)_4NH_3^+$		$-(CH_2)_4NHCOCH_3$	−1
$-(CH_2)_4NH_3^+$		$-(CH_2)_4NHCO(CH_2)_2COO^-$	−2[a]
$-(CH_2)_4NH_3^+$	$N\equiv C-O^-$	$-(CH_2)_4NHCONH_2$	−1
$-CH_2COO^-$ or $-(CH_2)_2COO^-$	CH_2N_2	$-CH_2COOCH_3$ or $-(CH_2)_2COOCH_3$	+1
$-CH_2COO^-$ or $-(CH_2)_2COO^-$	EDAC[b]/ $NH_2OH\cdot HCl$	$-CH_2CONHOH$ or $-(CH_2)_2CONHOH$	+1
			−1[c]

[a] In some circumstances, the reaction can form the cyclic, neutral intermediate. [b] 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide. [c] This product results when the reaction is carried out at pH > 12. Reactions carried out at pH 7–9 results in *N*,*N*-(1,2-dihydroxycyclohex-1,2-ylene)arginine (no change in charge), and pH 10–11 in multiple uncharacterized products.[142]

literature on protein modifications[135,136] lists many additional reactions that might be used to form charge ladders.

Acylation of ε-NH_3^+ residues of Lys by anhydrides or *N*-hydroxysuccinimide (NHS) esters of carboxylic acids annihilates a positive charge (ε-$NH_3^+ \rightarrow$ε-NHCOR) and makes the protein more negatively charged with each modification. Acylation of N-terminal α-NH_3^+ groups also occurs, but the difference in the pK_a values of α-NH_3^+ and ε-NH_3^+ groups allows for preferential acylation of one group over another at certain pH values, if necessary.[133]

The procedure for forming charge ladders by acylation of lysines usually involves the addition of 5–25 equivalents of an anhydride or an NHS ester (relative to the protein) to the solution of a protein in 10% v/v 0.1N NaOH (pH ≈ 10).[54,133] An excess of the reagent is required because the acylation reaction proceeds in competition with aqueous hydrolysis. For proteins with large numbers of lysine groups, we typically generate protein charge ladders showing complete sets of rungs by combining two batches of modified protein, one prepared using 10 equivalents of acylating agent, and one using 25 equivalents. This dual procedure favors the formation of the early and late rungs of the charge ladder; their combination gives a mixture containing all rungs. The reactions can also be carried out in buffers,[137] but the buffer must have low nucleophilicity so that reaction with the buffer does not compete with acylation. Carbamylation

reactions using potassium cyanate can also be used to neutralize lysine residues by converting them into homocitrulline.[138]

Reactions of carboxylic acid residues (Glu-COOH, Asp-COOH) with diazo compounds[136,139] or with nucleophiles such as hydroxylamine[111] or glycine methyl ester[140] result in the annihilation of a complete or partial negative charge, and make the overall charge of a protein more positive. Another possibility is the reaction between arginine and α-dicarbonyl compounds.[141] The products of this reaction, however, vary depending on the pH value of the reaction, and have not been thoroughly characterized.[142]

Another class of acylating reagents that can be used to form charge ladders comprises reagents that themselves generate charge, such as succinic anhydride, 1,2,4-benzenetricarboxylic anhydride, and 1,2,4,5-benzenetetracarboxylic dianhydride.[54,143] Upon reaction and subsequent hydrolysis, these reagents introduce one, two, and three carboxylic acid groups, respectively, in place of a positively charged lysine residue. We use these reagents primarily to improve the resolution of the rungs of charge ladders of proteins of high molecular weight.

The reactions of Lys ε-NH_3^+ residues with anhydrides or NHS esters of large, more hydrophobic groups, such as hexanoyl or benzoyl (Figure 4), result in what we call "hydrophobic charge ladders." Here, we use the change in the charge of the Lys residue upon acylation primarily to count the number of hydrophobic groups added to the protein, and the main objective of acylation is to modify the hydrophobicity of the protein rather than the charge. We have also extended the idea of a ladder to hydrodynamic drag.[144] Hydrodynamic ladders are generated by reactions of lysine residues with NHS-activated poly(ethylene glycol) (PEG) chains (see Section 8.3).

We have driven some of these reactions to more than 90% completion in an aqueous phase by using an excess of the reagents and carefully controlling the reaction conditions:[145] these procedures generate a single species (the perfunctionalized proteins), rather than a mixture. The ability to generate a single, perfunctionalized version of a protein makes it possible to use analytical techniques other than capillary electrophoresis to characterize it.[131,146] Figure 1b and 4 show electropherograms of different ladders of BCA, made with acetic anhydride, hexanoic anhydride, and hydroxylamine, as well as a demonstration of a peracetylated BCA.

6.2. Characterization and Analysis of Charge Ladders by Capillary Electrophoresis

We analyze protein charge ladders by capillary electrophoresis (CE) in free solution. CE separates molecules on the basis of competition between electrokinetic and hydrodynamic forces.[127] In comparison to other techniques for separating biomolecules based on charge—such as isoelectric focusing (IEF), poly(acrylamide) gel electrophoresis (PAGE), or ion exchange chromatography—CE offers high resolving power, minute consumption of material, rapid analysis, and ease of operation.

The basics of CE and its application to charge ladders have been reviewed extensively,[57,127,147] and thus we describe it here only briefly. The electrophoretic mobility of an analyte (μ [$m^2 V^{-1} s^{-1}$]), which is defined as the steady-state velocity per unit applied field, results from the balance of two opposing forces: an electrokinetic force that accelerates the charged molecule and a hydrodynamic drag force that slows it. The mobility μ can be expressed by Equation (19), where Z

$$\mu = \frac{eZ}{f_{\mathrm{eff}}} \tag{19}$$

(with no units) is the net charge of the molecule, e [C] is the charge of an electron, and f_{eff} [$N s m^{-1}$] is the effective hydrodynamic drag coefficient. The value of f_{eff} is a complex function of the shape and size (molecular weight) of the protein and the properties of the electrophoretic buffer (viscosity and ionic strength).

The experimentally observed mobility μ_{obs} of a molecule is the sum of two terms: the electrophoretic mobility μ of the molecule itself and the electroosmotic velocity (the velocity of the buffer) per unit field strength μ_{os}. Electroosmotic flow can be measured by adding an electrically neutral marker to the solution of the protein injected into the capillary.

The electrophoretic mobility of an analyte is thus calculated from the measured values of the migration times of the analyte and a neutral marker from Equation 20, where

$$\mu = \mu_{\mathrm{obs}} - \mu_{\mathrm{os}} = \frac{L_t L_d}{V}\left(\frac{1}{t} - \frac{1}{t_m}\right) \tag{20}$$

V [V] is the applied voltage, L_t [m] is the total length of the capillary, L_d [m] is length of the capillary from the inlet to the detector, t_{nm} [s] is the time that the neutral molecule takes to reach the detector, and t [s] is the time that the analyte takes to reach the detector. The mobility of the neutral marker is thus set to zero.[148]

CE separates the collection of derivatives of a protein charge ladder in free solution into the individual peaks or "rungs" of the ladder, with each rung containing the same number of modifications, and approximately the same net charge. Since small reagents such as acetic anhydride add <1% of mass or volume to the protein, they do not significantly alter the hydrodynamic drag (Figure 4), and the rungs of the charge ladder are separated by CE primarily on the basis of charge. Values of mobilities of the nth rung of the ladder, μ_n, for all the rungs of the charge ladder are measured in a single experiment; typical times for separation range from 3 to 30 minutes.

Capillary electrophoresis can resolve changes in μ of about 1×10^{-9} $m^2 V^{-1} s^{-1}$. The ability of CE to resolve the rungs of a particular charge ladder depends on the difference in the mobility caused by a change in a single unit of charge. We found empirically that proteins with molecular weights below 50 kDa form resolvable charge ladders in which the rungs differ by about 1 unit of charge.[54] In larger proteins (> 50 kDa), the change in one unit of charge does not change the μ value sufficiently to give resolvable rungs of the ladders. In this case, reagents that change the charge by more than one

unit per modification can be used to resolve the rungs of the ladder.[54]

For small proteins such as insulin (MW = 3 kDa), regioisomers with different acetylation patterns can sometimes be resolved and assigned.[133] The assignment of the acetylation patterns in insulin can be done by exploiting the differences in the pK_a values of the lysine group and the two N-terminal α-NH_2 groups, and by carrying out acetylation reactions in buffers of different pH values. Such assignments would be hard to carry out in a larger protein containing multiple lysine residues of similar reactivities.

Interactions between proteins and the walls of the capillary can complicate the analysis of charge ladders by CE. We usually analyze negatively charged proteins (that is, proteins analyzed at values of pH > p*I*; p*I* = isoelectric point of the protein) on bare silica capillaries. Such capillaries are negatively charged because of the ionization of silanol to siloxide groups ($SiOH \rightleftharpoons SiO^- + H^+$, $pK_a \approx 2.2$).[127] To analyze basic proteins (p*I* > pH of the electrophoresis buffer), it is necessary to reverse the charge of the capillaries to minimize the adsorption of positively charged proteins onto the negatively charged capillary wall. A variety of chemical methods can be used to modify and screen the negative charge on the wall of the capillary through covalent chemical modification[149] or physical adsorption of polymers.[134,150] We choose the straightforward and convenient method of noncovalent adsorption of a cationic polymer such as polybrene (Poly(*N*,*N*,*N*′,*N*′-tetramethyl-*N*-trimethylenehexamethylenediammonium dibromide)) to reverse the charge of the capillary to analyze basic proteins.[134,147] A technical difficulty in analysis arises when the charge ladder of a protein contains both net positively and negatively charged species: the rungs having charge opposite to that of the capillary tend to adsorb to its walls.

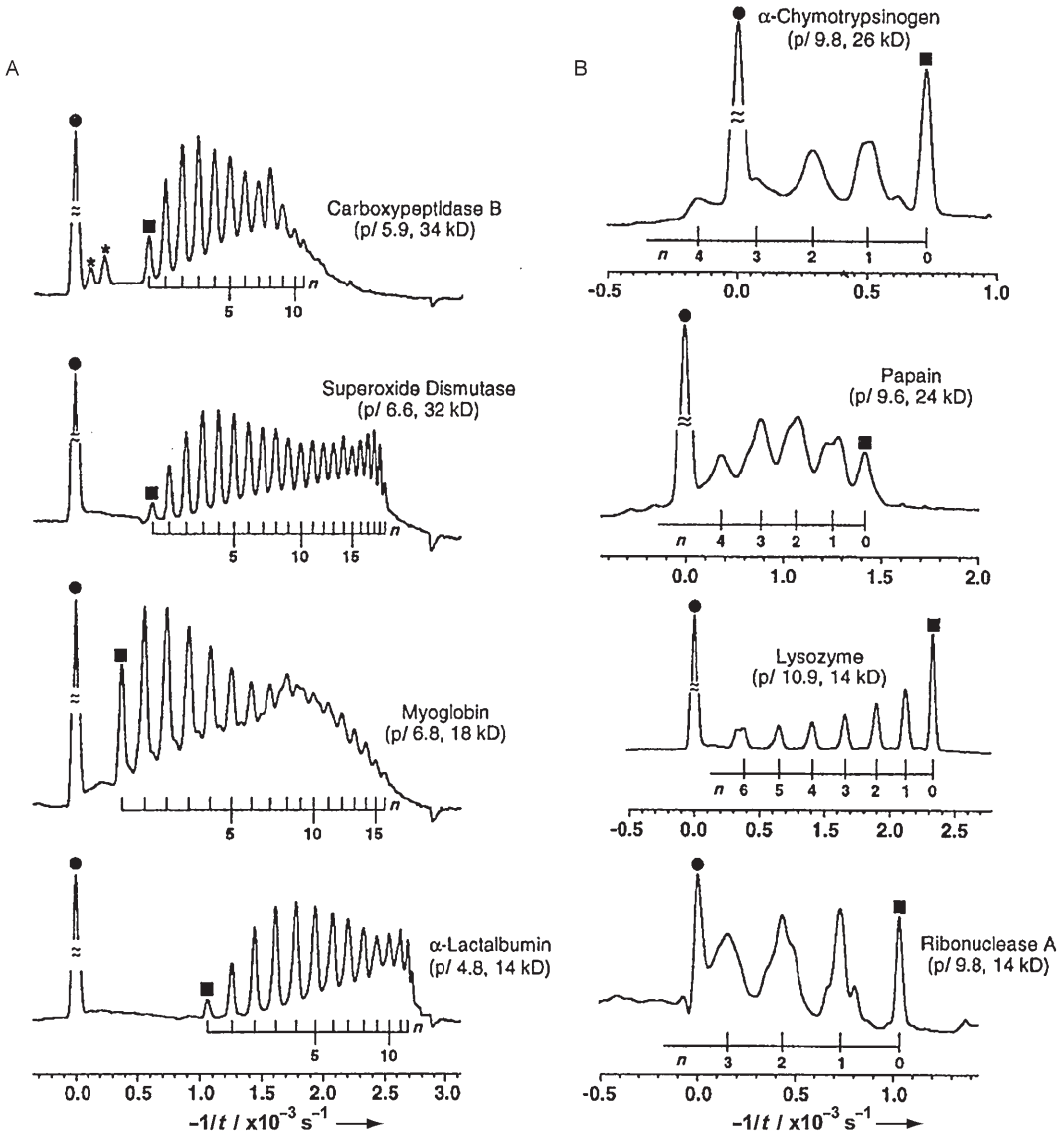


Figure 5. Examples of charge ladders formed by acetylation of Lys ε-NH_3^+ residues. ●: neutral marker, ■: native protein, and *: impurities in the starting sample. The number of acetylated Lys groups (*n*) is indicated below each electropherogram. A) Charge ladders of proteins having values of p*I* < 8.4, analyzed on an uncoated negatively charged capillary. B) Charge ladders of proteins having values of p*I* > 8.4, analyzed on a polybrene-coated (positively charged) capillaries. (Reproduced with permission from Ref. [54].)

Figure 5 shows a survey of different proteins—both basic and acidic, of various molecular weights, and containing various numbers of lysine residues—for their ability to form resolvable charge ladders. While many proteins form useful charge ladders, some (e.g. myoglobin) show the rungs of the ladder superimposed on a broad background; we do not understand the shape of those electropherograms. We believe we can form charge ladders of most proteins, unless acylation of lysine residues spontaneously denatures the protein or causes the derivatives to aggregate.

Increased resolution of the rungs of the charge ladder can be achieved by carefully choosing the conditions of the separation. We have worked out the conditions for baseline separation of the rungs of a charge ladder of carbonic anhydrase by using a long (117 cm) capillary and D_2O-based Tris-Gly buffer.[137] We believe that the D_2O-based buffer increases the resolution of the separation by increasing the viscosity of the buffer in comparison to the H_2O-based buffer. Figure 4B shows high-resolution electropherograms of early and late rungs.

7. Models for the Analysis of Data from Capillary Electrophoresis

Electrokinetic models, such as those used in colloidal science, can aid in the analysis of data from charge ladders, and can help to characterize the physical properties of proteins (that is, values of net charge and effective hydrodynamic size). As we show, a priori prediction of electrophoretic mobilities of proteins is challenging because it requires the combination of two models: an electrostatic one, to predict the distribution of charges and values of the electrostatic potential in the vicinity of the protein, and an electrokinetic one, which uses the values of electrostatic potential, information about the size and shape of the protein and properties (viscosity, concentration of different ions) of the solution to predict values of electrophoretic mobility.

The values of mobility of the rungs of a charge ladder with the smallest magnitude of net charge typically correlate linearly with the number of modifications and with the charge, if we assume the change in charge ΔZ between the rungs to be constant. One of the more useful and straightforward models for the analysis of the mobility data from a capillary electrophoresis experiment is that of Hückel,[151] with subsequent modification by Henry.[152] This model relates the mobility (μ [$m^2V^{-1}s^{-1}$]) of a particle to its surface potential (ψ_s [V]) through Equation (21), where ε_0 [$CV^{-1}m^{-1}$] is the permittivity of a vacuum, ε is the dielectric constant of the solution medium, η [Pa s] is the viscosity of the solution, κ [m^{-1}] is the inverse Debye length, and R [m] is the radius of the particle of interest. The function $f(\kappa R)$[153] added by Henry ranges in value from 1.0 to 1.5 and extends the applicability of the model of Hückel to particles of arbitrary size with respect to the Debye length.

$$\mu = \frac{2\varepsilon\varepsilon_0\psi_s}{3\eta} f(\kappa R) \tag{21}$$

The Debye model relates the surface potential ψ_s to the charge Z [Eq. (22)] and comes from the solution to the linearized Poisson–Boltzmann equation for a spherical object assuming a low average absolute surface potential ($|\psi_s| < 25$ mV; this assumption is known as the Debye–Hückel approximation[154]) and uniform charge distribution. Combining Equations (21) and (22) results in a simple expression that relates mobility directly to net charge [Eq. (23)]. The mobilities extracted from a capillary electrophoresis experiment for the rungs of a protein charge ladder can be analyzed, using Equation (23) to extract the values of charge of each rung, provided that the parameters η, κ, and R are known.

$$\psi_s = \frac{eZ}{4\pi\varepsilon_0\varepsilon R(1+\kappa R)} \tag{22}$$

$$\mu = \frac{eZ}{6\pi\eta R(1+\kappa R)} f(\kappa R) \tag{23}$$

Alternatively, Henry's model can be used in conjunction with mobilities from protein charge ladders to extract values of the hydrodynamic radius of a protein.[58] The procedure requires estimating or assuming the value of ΔZ, and involves a linear least-squares fit of the values of mobility of the first several rungs of a charge ladder to Equation (23). This procedure provides a value of the effective hydrodynamic radius R of the protein, which can be interpreted as the radius of a sphere that would have the same hydrodynamic drag as the protein.

The assumptions that underlie these colloid models do not describe proteins accurately. The Debye model assumes that the charge on the surface of a protein is uniformly distributed; since charged groups on proteins are clearly localized, this assumption is an approximation. Proteins are also only approximately spherical; theoretical predictions of the effective hydrodynamic radius R of a protein based on its real shape is an involved task that requires atomic-level structural data, and consideration of the effects of layers of hydration by the solvent.[155] Both the electrokinetic models of Henry and Hückel and the electrostatic model of Debye are valid only when the absolute average surface potential of the protein is less than about 25 mV.[151] As the magnitude of the charge on the protein increases beyond this value, the use of the Debye–Hückel approximation is no longer valid. Also, the ionic atmosphere surrounding the protein can no longer be assumed to be at its equilibrium value: the applied field can distort (that is, polarize) the ion atmosphere, and the finite mobility of these ions results in a distortion of this atmosphere as the protein moves (ion relaxation). The models of Hückel and Henry are, however, straightforward to use and can provide quick estimates of the net charge of a protein Z°_{CE} (defined in Section 2.1) under a particular set of conditions, provided there is some estimate of R.

There have been multiple extensions of the model of Henry and Hückel to address these approximations.[156–158] Yoon and Kim[156] relaxed the assumption that the protein is a sphere, and instead treated the protein as an ellipsoid with uniform surface potential; the electrostatics are still treated with the Debye–Hückel approximation, and ion relaxation

and polarization are ignored. If this model is parameterized with experimental protein diffusivities, and with some estimate of the shape of the ellipsoid from crystallographic data, it accurately predicts experimental mobilities for charge ladders[159] if the net absolute protein charge is low and the average absolute surface potential is less than approximately 25 mV.

O'Brien and White[157] treated the protein as a sphere with charge uniformly distributed on the surface; this model uses the nonlinear Poisson–Boltzmann equation and includes the effects of polarization and relaxation of ions. Since this model treats the electrokinetic problem for a uniformly charged sphere and its ion atmosphere exactly, and without further approximation, it is referred to in the colloids literature as the "Standard model" of electrokinetics.[160]

To assess the limits of Henry's model for describing the mobility of proteins in charge ladders, we compared experimental values of electrophoretic mobility of the charge ladder of HCAII with predictions from the models of O'Brien and White and the model of Henry using the electrostatic potential calculated with the Debye–Hückel theory [Eq. (22)].[58] Calculations with the Standard model were carried out using two values of ΔZ: a value of -1.0 to represent the "ideal" case and a value of -0.9 to approximate the effects of charge regulation (see Section 8.1). Figure 6 shows that both models give values that agree with the experimental mobilities for the rungs of the charge ladder with the smallest absolute values of net charge. The net charge of the protein becomes increasingly more negative with increasing number n of acetylated lysine ε-NH_3^+ groups, and the values of mobility no longer correlate linearly with n. The onset of nonlinearity at $n = 8$ in the experimental data coincides with deviation of the nonlinear Standard model from the linear model of Henry. Thus, both the nonlinear electrostatics and ion polarization and relaxation contribute to the observed nonlinearity. The fact that the experimental data fall between the two theoretical curves is consistent with other estimates of ΔZ for HCAII (Section 8.1).

The boundary element (BE) model of Allison et al.[161] requires the fewest approximations in modeling the electrokinetics of proteins. This approach uses the crystal structure of the protein to construct its hydrodynamic shear surface (that is, the particle–solution interface where the local viscosity changes from a high to the bulk value[161]) and to calculate its electrostatic potential. The BE model is thus an extension of the Standard model to particles with shapes and distribution of charge that more accurately describe the protein than does the Standard model.

In collaboration with Allison, we compared values of mobility calculated by the BE model with experimental mobilities of the charge ladders produced from five different proteins (bovine α-lactalbumin, hen egg-white lysozyme, bovine superoxide dismutase, human carbonic anhydrase II, and hen ovalbumin).[159] All calculated mobilities were in good-to-excellent agreement with experiments. The BE model also allowed the prediction of the change in mobility resulting from an incremental change in the charge of one unit. Comparison of these predicted changes in mobility with measured changes arising from modification of the charged

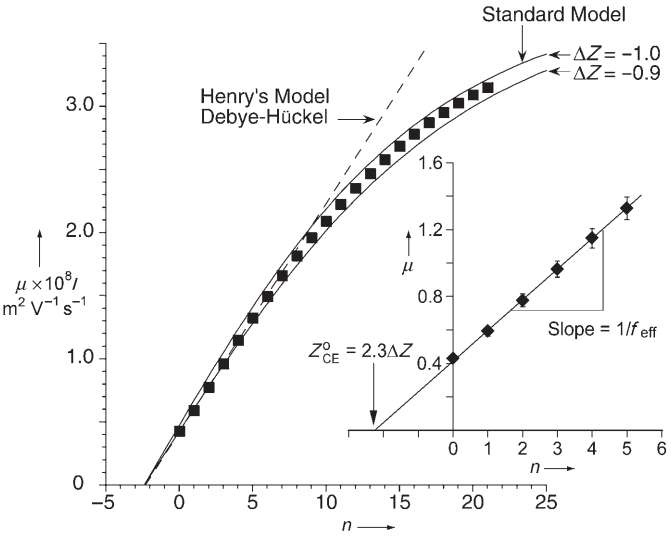


Figure 6. Values of the electrophoretic mobility μ of the rungs of the acetyl charge ladder of HCAII, plotted as a function of the number n of $NHCOCH_3$ groups on the protein. -----: values of μ predicted by Henry's equation using the electrostatic potential calculated from the Debye–Huckel equation; ——: values of μ predicted by the standard model of O'Brien and White, determined using values of $\Delta Z = -1.0$ and $\Delta Z = -0.9$. The inset shows the linear section of the graph of μ versus n. The solid line is a linear least-squares fit of these data: the intercept of the x-axis gives the number of amine groups that would have to be modified to make the electrophoretic mobility of the protein, and by inference the net charge of the protein, zero. The intercept indicated by the arrow in the inset shows that 2.3 amine groups would have to be added to the protein in this case. The net charge of the unmodified protein determined by CE is $Z^{\circ}_{CE} = 2.3\Delta Z$; here ΔZ is the change in net charge as a result of the acetylation of a Lys ε-NH_3^+ group. The value of ΔZ must be determined or estimated by some other means to assign a quantitative value to Z°_{CE}. The slope of the line gives the reciprocal of the effective hydrodynamic drag coefficient $1/f_{eff}$ of the protein.

groups provided estimates of ΔZ that were in reasonable agreement with values calculated using models of charge regulation (Section 8.1).[159] The main limitations to the BE calculations are that they are not yet available in an easily distributed software package, and are expensive computationally. This approach also requires knowledge of the crystal structure of the protein and an independent measure of the translational diffusion coefficient of the protein.

We found that the main challenge to the accurate prediction of electrophoretic mobilities of proteins is in the modeling of the electrostatics—the distribution of charged groups on the protein—and not in modelling the electrokinetics. Comparison of measured electrophoretic mobilities with predictions from electrokinetic models may be a useful approach to the development and testing of new models of electrostatics and proton equilibria in proteins.

8. Charge Ladders for Measuring the Net Charge and Hydrodynamic Drag of a Protein

Charge ladders provide a self-calibrating tool useful for the estimation of certain basic physical parameters of a

protein such as charge and hydrodynamic radius, both of which can be related to the electrophoretic mobility. Plots of the values of electrophoretic mobilities of the rungs of a protein charge ladder versus the number n of modifications typically shows a linear correlation, at least for small values of n (inset of Figure 6). If we know the change in charge ΔZ per modification of an ε-NH_3^+ group, and replot the mobility of the rungs versus the change in charge of each rung $n\Delta Z$, we are then able to determine the net charge of the unmodified protein Z°_{CE} by extrapolation of the best-fit line to the intersection with the x-axis. Equation (24) expresses the mobility of rung n as a function of Z°_{CE} and ΔZ.

$$\mu_n = \frac{eZ^n_{CE}}{f_{eff}} = \frac{e(Z^{\circ}_{CE} + n\Delta Z)}{f_{eff}} \tag{24}$$

This analysis (inset of Figure 6) allows the determination of the net charge of a protein in the absence of any models of electrophoresis and underlying assumptions. It requires, however, a value of ΔZ—the change in charge upon acetylation—to calibrate the x-axis.

8.1. Estimating the Change in Charge ΔZ upon Acetylation

The value of ΔZ depends both on the initial extent of protonation of the charged group being modified, that is, on the pK_a value of the group and on the pH value of the solution, and on the response of the protein to the modification. The Lys ε-NH_3^+ groups (nominal p$K_a = 10.4$) are 99% protonated at pH 8.4 while the Glu and Asp -COO^- groups (p$K_a \approx 4$) are more than 99% deprotonated. The elimination of charge on the Lys ε-NH_3^+ groups or the Glu/Asp -COO^- groups would then be expected to result in $|\Delta Z| = 1$ at, or close to, this pH value. The discussion in Section 2 of the protein as a network of ionizable groups rather than as a collection of independent charges emphasizes, however, that $|\Delta Z| < 1$. Other ionizable groups on the protein adjust to a change of electrostatic potential on the surface by shifting their charges. This shift reduces the magnitude of the change in net charge. Whether this shift is considered to reflect changes in pK_a values or in local concentration of protons at the protein–solvent interface is immaterial: these two mechanisms are equivalent both mathematically, and in their effect.[106] It is also possible that the adjustment take place through binding of the buffer ions. This type of charge adjustment is, however, more likely to be specific to the structure of the protein and composition of the buffer, rather than is that due to proton equilibria, and thus difficult to generalize or model.

8.1.1. The Linderstrøm-Lang Model

A simple model to use in analyzing this cooperative interaction among ionizable groups is that of Linderstrøm-Lang,[108] which was developed originally to interpret the curves of pH titrations of solutions of proteins. Menon and Zydney were the first to apply this model to the analysis of charge ladders.[105] The two key assumptions of the Linderstrøm-Lang model (which we abbreviate as the "LL model") are that 1) the protein is spherical and has no detailed structure and 2) the distribution of electrostatic potential on the surface of this spherical protein is uniform. The model assumes that the concentration of the protons near the surface $[H^+_{surface}]$ follows the Boltzmann distribution, described by Equation (25), where $[H^+]$ is the concentration of protons in the bulk solution.

$$[H^+_{surface}] = [H^+]\exp\left(\frac{-e\psi_s}{k_B T}\right) \tag{25}$$

The ionization state υ_I of each ionizable amino acid is determined by Equation (26), where $pH_{surface}$ is defined as $-\log[H^+_{surface}]$ and p$K_{a,i}$ is the ionization constant of the residue i in the absence of an electrostatic potential. The net charge Z°_{LL} of the protein is determined by summing all charges [Eq. (27)].

$$\upsilon_i = \frac{1}{1 + 10^{(pH_{surface} - pK_{a,i})}} \tag{26}$$

$$Z^{\circ}_{LL} = \sum_{\text{basic groups}} \upsilon_i - \sum_{\text{basic groups}} (1 - \upsilon_i) \tag{27}$$

The net charge Z°_{LL} of a protein calculated using the LL model can be related to its surface potential ψ_s by the Debye equation [Eq. (22)], and calculations are repeated iteratively using Equations (25)–(27) and (22) until the solution converges to a final value of Z°_{LL}. This calculation can be applied to the unmodified protein, and to the consecutive rungs of the protein charge ladder derived from it, in order to estimate the net charges Z°_{LL} and Z^n_{LL}, and thus the difference in charge upon a modification $\Delta Z = {}^1_{LL}\text{-}Z^{\circ}_{LL}$.

Menon and Zydney[105] applied the LL model to the acetyl charge ladders of bovine carbonic anhydrase to calculate the net charge of the rungs. They compared the results of the calculations to the values of net charge estimated from the measured values of electrophoretic mobility using Henry's model of electrophoresis. They concluded that cooperativity had a significant effect on the value of ΔZ, changing it from the ideal value of -1 to -0.86 in buffer of pH 8.4, $I = 10$ mM. We[106] compared values of net charge measured from the linear regression analysis with the LL calculations and with the charge calculated from experimental mobilities using the Hückel model of electrophoresis, and concluded that the value of ΔZ for the early rungs of the ladder of BCA was -0.93 at pH 8.4, $I = 10$ mM.[106] The difference in the value of ΔZ between our calculation and those of Menon and Zydney results from our consideration of details of the structure of BCA—post-translational acetylation of the N-terminus, coordination of three histidine residues and water to the Zn^{II} cofactor, and ionization of the water ($ZnOH_2^{2+} \rightleftharpoons ZnOH^+ + H^+$, p$K_a \approx 7$).[162] With these inclusions, the calculations agree that $\Delta Z \approx 0.9$, and perhaps more importantly, that the combination of the LL and Henry models provide a useful and reliable method to estimate ΔZ.

We have also studied the dependence of ΔZ for the acetyl charge ladder of BCA on the ionic strength and pH value of

the solution (Figure 7). The ability of each type of residue to participate in charge regulation depends on the pH value of the solution: as the pK_a value of a residue and the pH value of

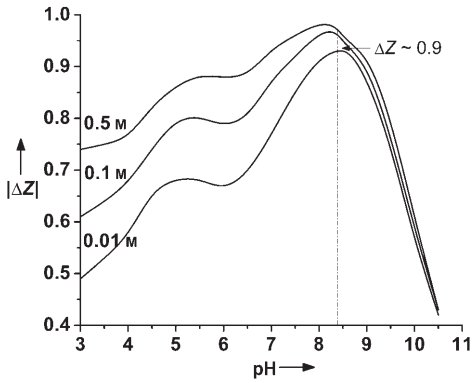


Figure 7. Plot of Linderstrøm-Lang calculated change in charge $|\Delta Z|$ upon acetylation, for a low number of modifications, for BCAII at three different ionic strengths. The vertical line indicates the pH value of the Tris-Gly buffer used in the CE. Increasing ionic strength diminishes the effect of charge regulation, and brings $|\Delta Z|$ close to unity at pH 8. (Reproduced with permission from Ref. [106].)

the solution approach one another, large adjustments in the ionization state of that residue can occur upon a change in the local electrostatic potential at the protein–solution interface. At physiologically relevant pH values of 7–8, among the major functional groups, the histidine residues (p$K_a \approx 6.5$) are most sensitive to electrostatic interactions and most prone to adjust their extent of ionization.[163] Other amino acids (Lys, Arg, Tyr, Glu, Asp) can also adjust to the change in the electrostatic potential, albeit to a lesser extent. BCAII, for example, contains the following numbers of ionizable groups: 18 Lys, 30 Asp/Glu, 9 Arg, 8 Tyr, 11 His, 1 C-terminal α-CO_2^-, 1 $ZnOH^+$. The range of pH values at which the deviation of the value of $|\Delta Z|$ from the nominal value of 1.0 is minimal is pH 8.0–8.5 (Figure 7). At high pH values, the lysine ε-NH_3^+ groups deprotonate, and at low pH values, the aspartate and glutamate CO_2^- groups easily protonate, thus reducing the magnitude of ΔZ. Increasing the ionic strength diminishes the effect of charge regulation through more efficient shielding of charges. Although the LL model makes several important simplifying assumptions, and neglects the explicit distance dependence of the electrostatic force between charged residues, it clarifies the cooperativity among ionizable groups and factors that influence it.

8.1.2. *Poisson–Boltzmann Monte Carlo (PBMC) Simulations*

We also used CE and charge ladders to measure the pK_a values of the N-terminal α-NH_3^+ group of two forms of lysozyme:[107] the native protein (with no $NHCOCH_3$ groups) and the modified protein with all six of its Lys ε-NH_3^+ groups acetylated (Figure 8A). If the ionization states of the Lys ε-NH_3^+ groups influence the ionization of the α-NH_3^+ group, then the pK_a value of NH_3^+ group in the native lysozyme will differ from that in the modified protein. We then compared the experimentally determined pK_a values with those determined by the LL model, and with those from Monte Carlo

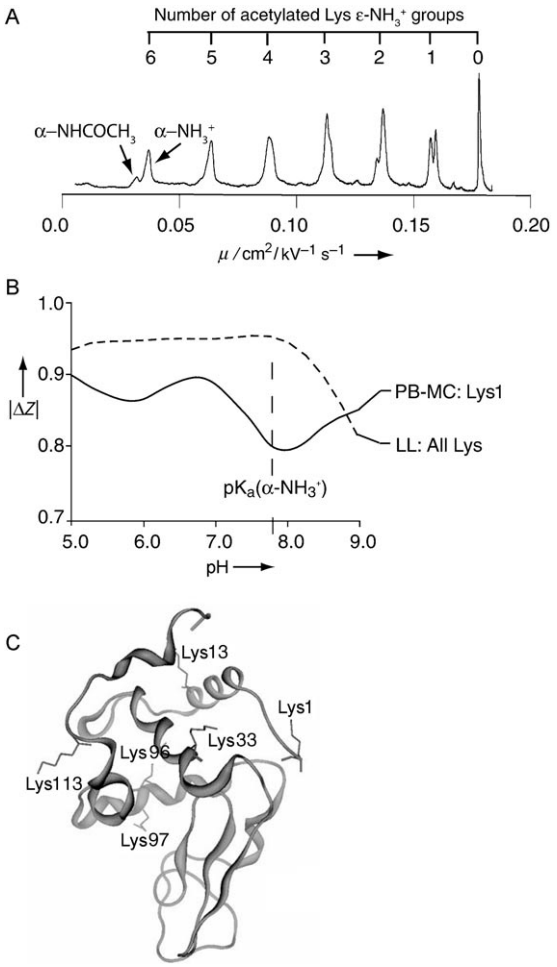


Figure 8. A) High-resolution electropherogram of the charge ladder of lysozyme showing the splitting of the rungs because of the presence of species with and without acetylated N-terminal α-NH_3^+ groups. B) The change in net charge ΔZ of lysozyme predicted by the PB-MC model as a result of the specific acetylation of Lys1 as a function of pH, and average ΔZ value calculated with the LL model. The LL model does not differentiate between different Lys ε-NH_3^+ groups. C) A ribbon diagram showing the position of the different lysine residues.

sampling of protonation states based on electrostatic potentials from numerical solutions of the linearized Poisson–Boltzmann equation (PB-MC model). Lysozyme is a good model for this kind of detailed study, because the pK_a values of all of its ionizable groups have been determined experimentally[100] and calculated by using continuum methods.[164]

The experimental measurements of the pK_a values of the α-NH_3^+ group in native and modified lysozyme showed that the pK_a values of the α-NH_3^+ group differed by 0.3 units at low ionic strength ($I = 33$ mM; pK_a(α-NH_3^+, native) $= 7.5 \pm 0.1$, pK_a(α-NH_3^+, peracetylated) $= 7.8 \pm 0.1$). The values of these two pK_a values at high ionic strength ($I = 108$ mM) were the same (p$K_a = 7.9$). This observation indicates that high salt concentrations can screen the cooperative interactions among charged groups in this small protein sufficiently that they become undetectable.

The LL model predicts that $\Delta Z = -0.94$ upon acetylation of a Lys group at any position at pH 8.4 ($I = 33$ mM), and is

approximately constant between pH 5.5 and 8 (Figure 8B). By using the PB-MC model, we were able to study in detail the response of lysozyme to the acetylation of a specific lysine residue.[107] We found that values of ΔZ upon acetylation of Lys13, Lys33, Lys97, and Lys116 agreed well with the value of ΔZ predicted by the LL model; the values of ΔZ upon acetylation of Lys1 and Lys96 differed significantly from the results from the LL model: The ΔZ value decreases to about 0.8 at pH 8 in the case of Lys1, and decreases to approximately 0.7 at pH 6 for Lys96. The reason behind these differences is, we believe, the proximity of other ionizable residues to the lysines of interest. For example, the α-NH_3^+ group of lysozyme and the ε-NH_3^+ group of Lys1 are only 0.7 nm apart in the crystal. The α-NH_2 group would therefore be expected to change its extent of protonation significantly under the influence of a change in charge on the Lys1 residue. A similar argument can be applied to His15, which is located only 0.8 nm away from Lys96.

In general, we found that significant cooperativity in ionization states only exists between acetylated lysine residues and other titratable groups on lysozyme separated by less than about 1.5 nm. This specific cooperativity can contribute up to approximately 0.2 units to the reduction in magnitude of ΔZ per pair of closely spaced residues. Interactions of lysine residues with a number of other distant titratable groups can contribute about 0.1 units to the reduction in magnitude of ΔZ.

The results of the detailed PB-MC model can be expressed by a simple empirical relationship, originally developed by Lee et al.[165] [Eqs. (28) and (29)]. This relationship describes the cooperativity in proton binding between titratable groups in terms of a shift of the pK_a value by using a screened Coulombic potential (SCP) that depends explicitly on the distance r_{ij} between the two groups on the protein. The cooperativity between groups is modeled by using the Debye–Hückel model with the dielectric constant ε_{eff} as a fitting parameter. This relationship should be useful for the quick analysis of cooperative interactions between specific groups i and j to identify those large enough to influence the values ΔZ.

$$\Delta\psi_{ij} = \frac{e}{4\pi\varepsilon_0\varepsilon_{eff}}\frac{\Delta Z_j}{r_{ij}}\exp(-\kappa r_{ij}) \quad (28)$$

$$\Delta pK_{a,ij} = \frac{e\Delta\psi_{ij}}{2.303 k_B T} \quad (29)$$

8.1.3. Determination of ΔZ though Independent Measure of the Hydrodynamic Radius of Proteins (The Linear Regression–Taylor Dispersion Method, LR-TD)

We propose another method for calibrating the value of ΔZ for the analysis of charge ladders.[166] This method involves independent measurement of the diffusivities of proteins by using the analysis of the dispersion of the analyte sample in pressure-driven flows through thin capillaries.[167,168] The concentration profile of the analyte, which is shaped by the axial convection and radial diffusion, is related to the diffusivity of the protein through the expression of Taylor,[169] and then to the hydrodynamic radius R through the Stokes–Einstein equation. Conveniently, the concentration profile of the analyte can be measured in a standard capillary electrophoresis setup, equipped with pressure flow capability.[166,168]

With the experimentally determined values of R, one can determine the effective hydrodynamic drag f_{eff} of proteins by Henry's law of electrophoresis, and thus ΔZ [Eq. (24)]. Table 3 shows the results for the measurement of diffusivity D_0 for several proteins by using analysis of dispersion, calculation of R from the D_0 value, and subsequent estimation of the ΔZ value through linear regression (abbreviated as the LR-TD method). The Table also compares the ΔZ values from the LR-TD approach with those calculated by the model of Linderstrøm-Lang and by the screened Coulombic potential approach. In most cases, the values of ΔZ agree well among the three methods for most proteins and conditions, although there is no consistency in over- or under-estimation of ΔZ by one method or another. It is possible that the large discrepancy in the case of α-lactalbumin is due to the partial denaturation of the protein.[170]

Table 3: Comparison of ΔZ found by linear regression/Taylor dispersion method (LR-TD) with ΔZ calculated by Linderstrøm-Lang (LL) and screened Coulombic potential (SCP) models.

Protein	D_0[a] [$\times 10^6$ cm s^{-1}]	R [Å]	I [mM]	ΔZ, LL-TD	ΔZ, LL[b]	ΔZ, SCP
BCAII	0.939 ± 0.009	26.1 ± 0.2	8	-0.87 ± 0.01	-0.91	-0.93
BCAII			33	-0.97 ± 0.01	-0.93	-0.95
BCAII			133	-0.97 ± 0.01	-0.95	-0.97
α-lactalbumin	1.10 ± 0.01	22.3 ± 0.2	8	-0.81 ± 0.02	-0.93	
α-lactalbumin			28.1	-0.96 ± 0.02	-0.95	
lysozyme	1.19 ± 0.02	20.6 ± 0.2	8	-0.86 ± 0.02	-0.91	-0.86
lysozyme			33	-0.93 ± 0.02	-0.94	-0.92
lysozyme			108	-0.98 ± 0.02	-0.96	-0.95
myoglobin	1.21 ± 0.02	20.2 ± 0.3	8	-0.82 ± 0.02	-0.91	
ovalbumin[c]	0.730 ± 0.002	33.5 ± 0.1	8	$+0.91 \pm 0.01$	$+0.91$	

[a] No dependence of D_0 and thus R on ionic strength was found. [b] Values of R, used for LL calculations are the experimentally determined values obtained by the Taylor dispersion method. [c] Positive charge ladder formed by amidation of carboxylic acid residues.

8.1.4. Conclusions concerning the Change in Charge ΔZ upon Acetylation

Charge regulation, or cooperativity in proton equilibria, must be included in the analysis of charge ladders. Annihilation of charge on one group of a protein perturbs the extent of protonation of other groups, in the direction that tends to decrease the net change in charge $|\Delta Z|$. Calculations of accurate values of ΔZ will increase the accuracy of the parameters extracted from charge ladders.[107] PB-MC mod-

eling or the empirical relation [Eqs. (28), (29)] requires the crystal structure to estimate the value of ΔZ. If only the sequence of a protein is known, the LL model can quickly provide an estimate of ΔZ. The LR-TD method provides an experimental way of determining both R and ΔZ values at any value of pH or ionic strength without requiring that the crystal structure or sequence of the protein be known. For most proteins, however, we believe that using a value of $\Delta Z = 0.9 \pm 0.05$ at pH 8.4 and $I = 10$ mM will minimize the error in calculations based on charge ladders.

8.2. Determination of Net Charge and Hydrodynamic Drag

When we have a value of ΔZ, we can determine the net charge of the protein under the conditions of electrophoresis by linear extrapolation (Figure 6). For HCAII in Tris-Gly buffer (25 mM Tris, −192 mM Gly, pH 8.4, $I = 10$ mM, 25 °C) with the ΔZ value estimated as −0.90, $Z^{\circ}_{CE} = -2.1$. This value differs from the value of charge calculated from the sequence of the protein Z_{seq} by almost a full unit of charge, with $Z_{seq} = -1.2$. This difference results from the fact that values of Z_{seq} are calculated using standard values of pK_a that neglect the unique electrostatic environment—including the effects of charge regulation—of the particular protein.

The slope of the linear regression line through the values of mobility of the charge ladder depending on the number of rungs n provides the value of the effective hydrodynamic drag f_{eff} of the protein [Eq. (24)].[58] The effective hydrodynamic drag can then be converted into the hydrodynamic radius R of the protein by using Henry's model of electrophoresis (Section 7). Such analysis, of course, would be of interest only if ΔZ is estimated by a method other than LR-TD or simply assumed. The value of R for HCAII was estimated to be 2.5 nm when a value of $\Delta Z = 0.90$ was used.[58] Thus, this approach allows the charge and size of a protein to be determined in aqueous solutions that mimic biological environments in a single experiment.

8.3. PEG Ladders: Ladders of Hydrodynamic Drag

Hydrodynamic ladders are formed by grafting chains of poly(ethyleneglycol) (PEG), rather than acetyl groups, to the surface of a protein by acylation; each reaction changes both the charge and the hydrodynamic drag of the protein.[144] This type of experiment uses the charges measured by the ladder to count the number of PEG groups attached to the protein. We determined the electrophoretic mobilities of the rungs of PEG ladders of lysozyme by using CE, and calculated the hydrodynamic radius R of each rung using Henry's model of electrophoresis.[58,152] Figure 9 addresses the effects of chain density by showing that the increase in the hydrodynamic radius of lysozyme modified with short PEG (2 kD) chains is smaller than the increase in the radius of lysozyme modified with the longer (5 kD) PEG chain.

It is not directly apparent whether the R values of each protein-PEG construct would be accurately predicted by Henry's model, as the model describes the behavior of smooth, uniformly charged spheres. To test the accuracy of Henry's model, we related the values of R to the partition coefficient of these proteins into neutral gels by using Ogston's model,[171] and compared those partition coefficients with experimentally measured ones.[144] We found excellent agreement between the predicted and measured partition coefficients for all rungs of the PEG ladder. This agreement shows that the analysis of hydrodynamic ladders can be done by using Henry's model of electrophoresis. We believe that the hydrodynamic ladders will be useful in studies of the effects of hydrodynamic size on the transport properties of proteins.

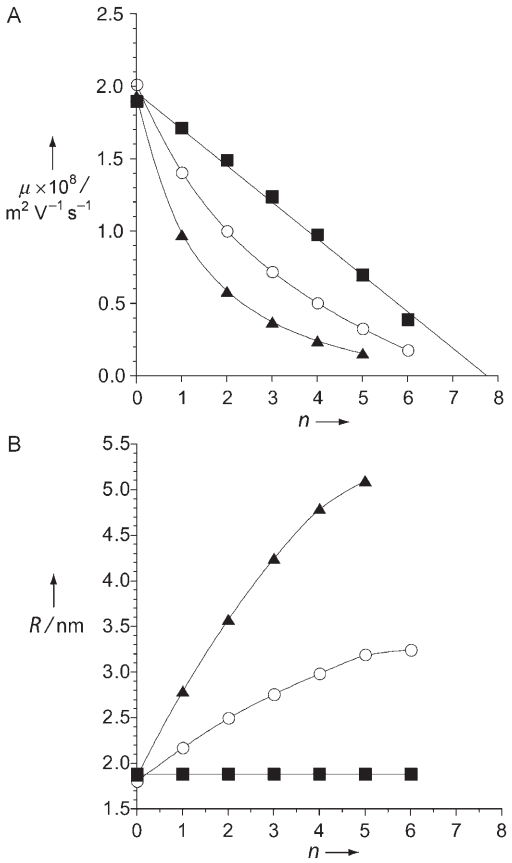


Figure 9. A) Electrophoretic mobility μ of protein charge and radius ladders of lysozyme produced by the partial modification of Lys ε-NH_3^+ groups with polyethylene glycol chains functionalized with *N*-hydroxysuccinimide ester (HO-PEG-OCH_2COO-NHS): molecular weight 2 kDa (○) and 5 kDa (▲). Each acylation results in a change in mobility because of the neutralization of an ε-NH_3^+ group and the increase in the hydrodynamic drag of the protein. The mobilities are compared with values of the charge ladders produced by modification with acetic anhydride (■). B) The hydrodynamic radius R of the rungs of the charge and radius ladders of lysozyme calculated using Henry's model. The values were calculated by assuming differences in mobility of each rung between the PEG and acetyl-modified protein in (A) are due entirely to effects of the PEG on the effective hydrodynamic radius. The hydrodynamic radii of free 2- and 5-kDa PEG chains are 1.4 and 2.2 nm, respectively.

9. Influence of the Net Charge of Proteins (Z_{CE}) on Binding of Ligands: Combining Charge Ladders with Affinity Capillary Electrophoresis

Molecular recognition events involving proteins—for example, ligand binding or folding—in general result in changes in the net charge and/or hydrodynamic drag of proteins. CE detects these changes as a shift in the electrophoretic mobility, and can therefore be used to monitor them.[172] Since protein charge ladders allow the net charge of proteins to be isolated as an independent variable, biophysical studies using charge ladders and CE can also provide measurements of the role of net charge and electrostatics in molecular recognition events involving proteins.[4]

9.1. Affinity Capillary Electrophoresis

The measurement of the binding of ligands to proteins by capillary electrophoresis (affinity CE or ACE) is well described in the literature,[172,173] and we give only a brief description of the technique here. In an ACE experiment, a sample of the receptor is injected onto a capillary and the electrophoretic mobility is measured as a function of the concentration of the ligand in the electrophoresis buffer. The binding of a charged, low-molecular-weight ligand to a protein causes a shift in the electrophoretic mobility of the protein. The shift is due to the fact that the protein and the protein–ligand complex differ (to a first approximation) by the charge of the ligand, but have similar hydrodynamic drag. Measuring the mobility of the protein at different concentrations of the ligand present in the electrophoresis buffer makes it possible to measure binding constants and thus free energies of binding. Affinities of neutral ligands can be measured by using competitive assays against a charged ligand with known binding constant.[174] By using this approach it is straightforward to measure the binding of ligands to multiple proteins (isozymes, homologous proteins, members of a charge ladder) at the same time. For example, ACE makes it possible to measure the mean binding constant for all proteins in each rung of a charge ladder in one set of experiments.[4,5]

9.2. Dependence of the Free Energy of Binding of Charged Ligands on the Net Charge of Carbonic Anhydrase

The binding of a charged ligand to the proteins that make up the rungs of a charge ladder shifts their mobilities.[4,5] Figure 10A shows representative electropherograms of the charge ladder of HCAII measured with different concentrations of the benzene sulfonamide inhibitor **1** in the electrophoresis buffer. As the concentration of this ligand increases, the fraction of the protein–ligand complex also increases. As this ligand carries a net positive charge, the effect of binding is a shift in the positions of the peaks in the electropherograms to the left (that is, to decreasing values of mobility as the protein is negatively charged under the conditions of the experiments). The mobilities of the rungs of the charge ladder measured at 10 μM of the ligand **1** represent the saturated or fully bound state of the protein.

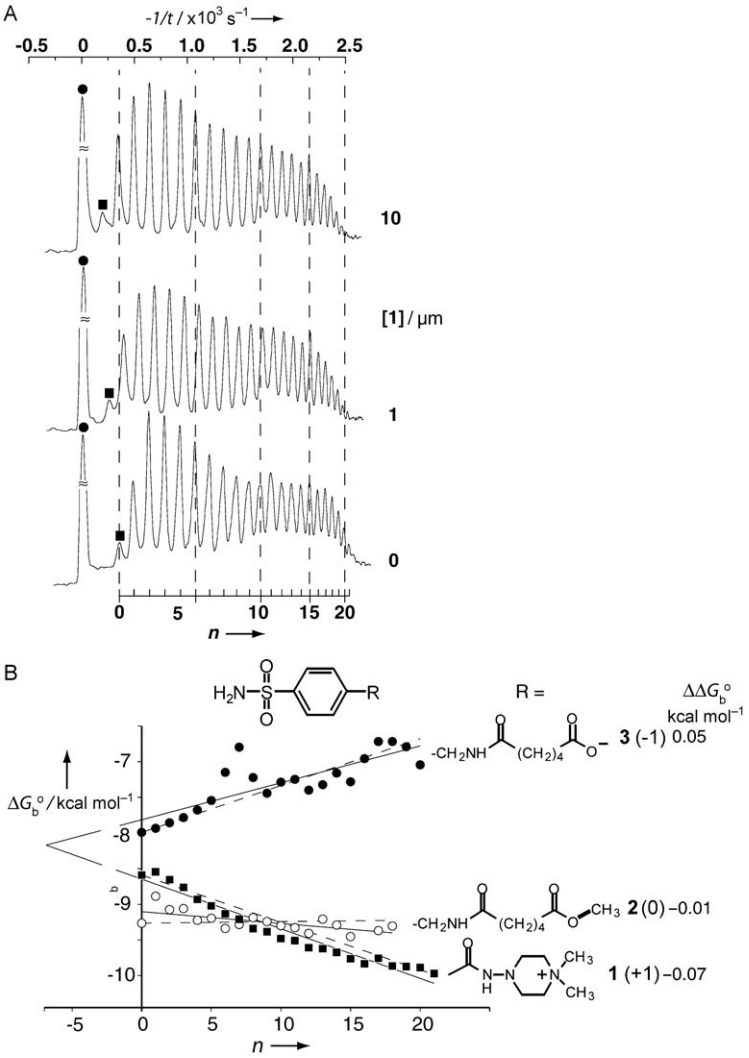


Figure 10. A) Electropherograms demonstrating the changes in the electrophoretic mobility of the rungs of the charge ladder of human carbonic anhydrase II as a result of the binding of ligand **1**. ●: electrically neutral markers of electroosmotic flow, ■: the unmodified protein. The number of acetylated Lys ε-NH_3^+ groups (n) is indicated below the electropherogram. The ACE experiment was performed with inhibitors **1**–**3** in 25 mM Tris/192 mM Gly (pH 8.4). B) Dependence of the standard-state free energy of binding ΔG_b° on the number n of $NHCOCH_3$ groups of the charge ladder of HCAII, and on the charge on the ligands. ——: least-squares fit to the data; -----: results of the continuum electrostatics calculations. The slopes ($\Delta\Delta G_b^\circ/n$) from the linear regression analyses of ΔG_b° versus n yielded the magnitudes of the influence of the charges of the Lys ε-NH_3^+ groups on the HCAII–ligand interactions. The lines from the continuum electrostatics calculations are just the average of the calculated contributions of the 21 different Lys ε- NH_3^+ groups to the free energies of binding of the different sulfonamides. (Reproduced with permission from Ref. [5].)

We carried out experiments, analogous to those in Figure 10A, with three different ligands—each a benzene sulfonamide substituted in the *para* position with a charged or neutral pendant group—and analyzed the data to determine the binding affinities of the rungs of the charge ladder for

those ligands. HCAII becomes more negatively charged as the number of acetylations increases. Figure 10B shows values of the free energies of binding (ΔG_b°) of these ligands as a function of the number of modifications on HCAII. These data allow comparison of the binding affinities of unmodified HCAII for each of these different ligands; they also provide a direct measure of the effects of long-range electrostatic interactions on the binding affinities. Values of ΔG_b° for the neutral ligand are approximately independent of the net charge on the protein. These results imply that the modifications to the amino acids that produced the charge ladders had little effect on the structure of the active site of this enzyme.

For the charged ligands, we observed an approximately linear relationship between the free energy of binding and the net charge of the protein: increasing the net negative charge of the protein resulted in more favorable binding of the positively charged inhibitor **1**, and less favorable binding of the negatively charged inhibitor **3**. This linear relationship implies that the effect of increasing net charge on binding affinity is, in this case, approximately additive.

9.2.1. *Modified Carbonic Anhydrase with Selectivity for Charged Ligands that Increases with Net Charge*

The acetylation of Lys ε-amino groups results in a change in the free energy of interaction between the protein and charged ligands that reflects the collective effects of long-range electrostatic interactions. These interactions, which are less than 0.1 kcal mol^{-1} per unit increment of charge of the protein (that is, 10–20% of RT), are individually too weak to be measured by most biophysical techniques. Increasing the net charge on HCAII by eliminating the positive charge on 21 Lys ε-NH_3^+ groups resulted in an affinity for the anionic inhibitor **3** that was more than 100 times (corresponding to approximately 3 kcal mol^{-1}) that of cationic inhibitor **1**. It is surprising that the most highly charged species follow the linear trends in affinity observed for proteins with lower net charge.

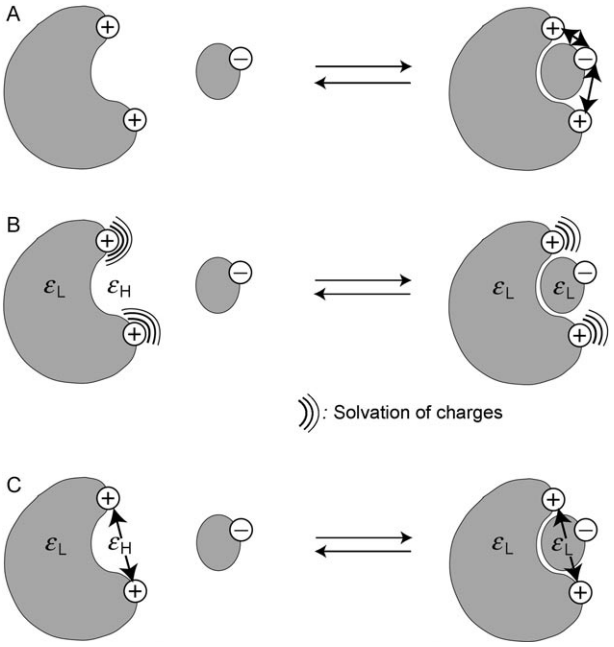
9.2.2. *Modified Carbonic Anhydrase with Large Values of Net Charge without Loss of Stability*

Figure 10B indicates that acetylating all 21 Lys ε-NH_3^+ groups does not change the active site: the affinity of HCAII for the electrically neutral inhibitor is essentially insensitive to increasing numbers of acetylations. Carbonic anhydrase is a particularly stable protein (with a free energy of denaturation at 25 °C of ca. 20 kcal mol^{-1}),[175] and one must be concerned that this conclusion might apply only to such proteins. This conclusion may, however, be applicable to at least some other proteins. For example, Pace and co-workers[22] demonstrated that as many as five charge-reversal mutations (that is, the conversion of five Asp or Glu residues into Lys residues) had only a modest influence on the stability of ribonuclease Sa, and reduced the free energy of denaturation at 25 °C by approximately 0.5 kcal mol^{-1}. Quantitative measurements of the dependence of the thermal stability of α-lactalbumin on the net charge measured by using CE and charge ladders[56] also suggest that net charge is not an important determinant of stability in this protein (Section 10).

9.3. *Comparison of Experimental Values with Predictions by Continuum Electrostatic Modeling Studies*

Continuum electrostatics calculations, which are based on numerical solutions of the linearized Poisson–Boltzmann equation,[24] provided a detailed analysis of the contributions of individual Lys ε-NH_3^+ groups to the electrostatic free energy of binding that is impossible to detect experimentally.[5] We calculated values of ΔG_{elec}°, the contribution of electrostatic free energy to the total free energy of binding, as the difference in the electrostatic free energy of the protein in the presence of a sulfonamide ligand bound to the zinc ion at the active site, and in the absence of ligand (where a hydroxide ion is bound to the zinc ion).[5] We defined the contribution of each Lys ε-NH_3^+ group to the electrostatic free energy of binding as $\Delta\Delta G_i^\circ$, the difference in ΔG_{elec}° when Lys ε-NH_3^+ group i carried a charge of $+1$, and when it had been converted into an $NHCOCH_3$ group with a charge of 0. These calculations made the simplifying assumption that $\Delta Z = -1$. The values of $\Delta\Delta G_i^\circ$ [kcal mol^{-1}] from calculations, averaged over all 21 Lys ε-NH_3^+ groups, were -0.07 for inhibitor **1**, 0.00 for inhibitor **2**, and $+0.05$ for inhibitor **3**. The dashed lines in Figure 10B are plotted with these values as slopes.

These calculations suggested that values of $\Delta\Delta G_i^\circ$ could be divided into three parts (illustrated in Scheme 5): 1) the direct



Scheme 5. Schematic diagram illustrating three contributions to $\Delta\Delta G_{elec}^\circ$ (see text); A) direct effect, B) solvation effect, C) indirect effect. The gray interiors of the protein and ligand are regions of low dielectric ($\varepsilon_L = 2$–4), the regions surrounding the protein and ligand are water with high dielectric ($\varepsilon_H \approx 80$); ↔: electrostatic interaction between charges. (Adapted from Ref. [5].)

effect, $\Delta\Delta G^{\circ}_{dir}$, is the intermolecular Coulombic interactions between charged groups on the protein and charged groups on the ligand or hydroxide ion bound to the zinc ion; 2) the effect of solvation, $\Delta\Delta G^{\circ}_{solv}$, is the change in the free energy of solvation of charged groups on the protein and ligand when the hydroxide ion is replaced by the ligand in the active site; 3) the indirect effect, $\Delta\Delta G^{\circ}_{indir}$, is the change in the intramolecular Coulombic interactions between charged groups on the protein and the ligand when the hydroxide ion, and a significant volume of water in the active site, is replaced by the ligand. Values of $\Delta\Delta G^{\circ}_{solv}$ and $\Delta\Delta G^{\circ}_{indir}$ reflect the changes in screening by the solvent and the ionic atmosphere that result from the change in the shape of the interface between regions of low dielectric constant (the protein and bound ligand, if present), and of high dielectric solvent (water) in the bound and the unbound state.

Values of $\Delta\Delta G^{\circ}_{solv}$ were small (less than 4×10^{-4} kcal mol^{-1} for all lysine groups, except Lys170) because the region of the protein that is desolvated on ligand binding is far from all of the lysines except Lys170 (which had a $\Delta\Delta G^{\circ}_{solv}$ value of 0.02 kcal mol^{-1}). Figure 11 compares the direct and indirect

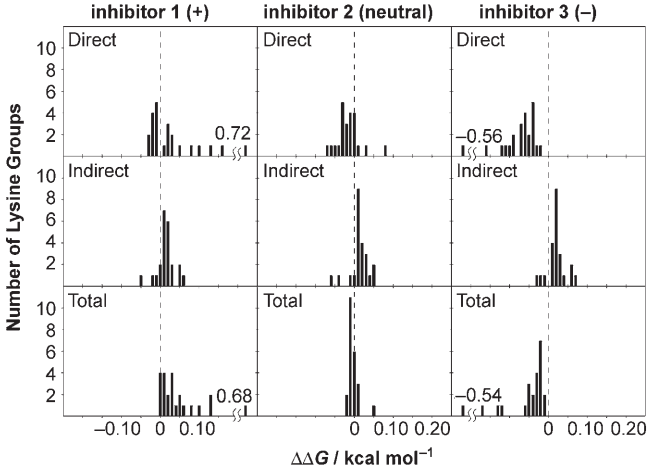


Figure 11. Histograms of $\Delta\Delta G^{\circ}_{dir}$, $\Delta\Delta G^{\circ}_{indir}$, and the total $\Delta\Delta G^{\circ}_{elec}$ (including $\Delta\Delta G^{\circ}_{solv}$) as calculated for each of the three inhibitors from Figure 10B. Each bar indicates the number of Lys ε-NH_3^+ groups on HCAII that contribute an amount of free energy, given by the *x*-axis, to $\Delta\Delta G^{\circ}_{elec}$. The values of $\Delta\Delta G^{\circ}_{dir}$ and the total $\Delta\Delta G^{\circ}_{elec}$ calculated for Lys170 in the presence of inhibitors **1** and **3** are much larger in magnitude than the values for the other Lys residues. The four values for Lys170 are printed explicitly on the histograms. (Reproduced with permission from Ref. [5].)

contributions of different Lys ε-NH_3^+ groups to $\Delta\Delta G^{\circ}_{elec}$. Most previous analyses have focused only on the direct effect; this study shows that both direct and indirect interactions can be of similar magnitude.

How do we reconcile the complex distribution of interaction free energies (shown in Figure 11) with the relatively simple dependence of the ΔG°_{b} value on the number of acetylations (shown in Figure 10B)? It may be that these details are obscured when values of $\Delta\Delta G^{\circ}_{elec}$ are averaged over the different protein derivatives that make up each rung of the charge ladder. To test this idea we performed Monte Carlo simulations to examine the influence of different patterns of acetylation on the average binding free energies of the different rungs of the ladder.[5] These simulations showed that a large number of different patterns of acetylation gave results that were in good agreement with the experimental results.

What do these results tell us about the ability of ACE and charge ladders to resolve changes in the electrostatic free energy of binding that result from changes in the net charge of the protein? There are two issues: 1) the ability of ACE to measure differences in values of ΔG°_{b} between adjacent rungs, and 2) the ability of ACE to resolve the presence of species with different values of ΔG°_{b} within a single rung. The results in Figure 10 show that ACE can measure differences in the ΔG°_{b} values between adjacent rungs of the ladder that are as small as 0.1 kcal mol^{-1}. This accuracy, combined with the large number of rungs of the ladders that are simultaneously monitored for binding by CE, provides an accurate measure of the average value for $\Delta\Delta G^{\circ}_{elec}$ over a large number of derivatives. (It would be very difficult to estimate energetic terms of this magnitude using site-directed mutagenesis.)

These results cannot rule out the existence of outliers from the measured average values of $\Delta\Delta G^{\circ}_{elec}$ in the pool of derivatives within a single rung. Calculated contributions to the free energy of binding of different lysine groups ranged from 0.01 to 0.15 kcal mol^{-1}, except for one residue in HCAII, Lys170, which contributed between 0.54 and 0.72 kcal mol^{-1}. If we assume that the derivatives acetylated at Lys170 represent the only outliers, then these derivatives will make up only approximately 5% of the population in a rung. The resolution of CE (defined as the ability to resolve species with different values of electrophoretic mobility into separate peaks) also limits this kind of experiment. The difference in mobility between adjacent rungs (ca. 2×10^{-9} m^2 V^{-1} s^{-1}; Figure 10A) represents the maximum shift in mobility from the binding of a ligand that presents a single unit of charge (+ or −) μ_{PL}-μ_P. From Figure 10A, we estimate a peak width to correspond to the difference in mobility of about 0.8×10^{-9} m^2 V^{-1} s^{-1}: a reasonable estimate of the resolution of CE.

We use these estimates to determine the ability of ACE to resolve differences in the values of K_d among different species within the same rung of the ladder. The maximum difference in mobilities will occur at concentrations of ligand near the value of K_d. We assume a rung of the ladder is composed of two populations: the first population includes those derivatives where Lys170 is not acetylated and has a value of $\Delta G^{\circ}_{b} = -8$ kcal mol^{-1} at 25 °C (which corresponds to a value of $K_d = 1.36$ μM); the second population includes those derivatives acetylated at Lys170 and has a value of $\Delta G^{\circ}_{b} = -8.5$ kcal mol^{-1} at 25 °C (which corresponds to a value of $K_d = 0.58$ μM). With the concentration of ligand of 1.36 μM, the first population will be half-saturated, while the fraction of second population with bound ligand will be 0.7. From our previous estimate of μ_{PL}-$\mu_P = 2\times10^{-9}$ m^2 V^{-1} s^{-1}, we estimate that these two populations will differ in electrophoretic mobility by about 0.4×10^{-9} m^2 V^{-1} s^{-1}, a value smaller than the width of a single peak of the charge ladder. Therefore, we do not expect the presence of different populations that differ in values of

ΔG_b° on the order of 0.5 kcal mol^{-1} to result in the presence of detectable shoulders or splitting of peaks during an ACE experiment. This estimate was supported by simulations of the ACE experiment.[5] The relatively uniform behavior observed with CE and charge ladders may in fact be masking more complex behavior, which is homogenized through the combined effects of averaging over a population of different protein derivatives and limited resolution of CE.

9.4. Conclusions

The modification of as many as 21 primary amino groups on HCAII does not alter its binding to neutral inhibitors; we infer that the binding pocket of this enzyme is not altered by these modifications. Large numbers of weak (< 0.1 kcal mol^{-1} per unit of charge), long-range electrostatic interactions contribute significantly to the free energy of binding of charged ligands to proteins and can influence the selectivity of binding. These weak interactions can be measured accurately by the combination of ACE and charge ladders because of the large number of derivatives that are studied simultaneously.

Continuum electrostatic theory accurately predicts the average contributions of long-range electrostatics to the free energy of binding of charged ligands to HCAII. There is a distribution in the contribution of different charged groups to the free energy of binding; ACE usually does not resolve these differences within individual rungs of a ladder. These calculations also show that changes in the shape of the boundary between regions of low and high dielectric constant that accompany the binding of a ligand to the protein contribute an amount to the electrostatic free energy of binding that is similar in magnitude to the Coulombic interactions between charged groups on the protein and ligand. Detailed models that include explicitly the molecular shape of the protein and protein–ligand complex are required to capture this effect.

The combination of charge ladders, ACE, and continuum electrostatic theory is a useful method for studying the contributions of charged groups to the free energy of ligand binding. ACE and protein charge ladders measure the average effects of long-range electrostatic interactions, and can serve as a check of models based on continuum electrostatic theory that, in turn, provide the details of the contributions individual charged groups make to the free energy of binding.

10. Study of Electrostatic Contributions to Protein Folding and stability by Using Charge Ladders

10.1. The Role of Net Charge in Protein Stability

Our use of protein charge ladders to study the role of net charge on protein stability to thermal denaturation was motivated by two questions: 1) Do charged groups on proteins influence stability in a way that is similar to their influence on affinity for charged ligands?[81] 2) What is the role of charged groups in the pH-dependent stability of proteins?[21]

Many proteins denature under acidic and alkaline conditions.[2] There are two explanations for this instability at extremes of pH values. 1) At values of pH far from the isoelectric point, proteins will develop a large net excess of positive charge (at low pH) or negative charge (at high pH). This large net charge will destabilize the compact, native state, relative to that of the more extended denatured state, by increasing the repulsion between like-charged groups. 2) Values of pK_a change when the protein denatures; differences in pK_a values would link charge regulation directly to thermal denaturation, and would result in differences in the net charge of proteins in the native and denatured states. Previously, these two possibilities could not be disentangled.

The combination of CE and charge ladders provides both the thermodynamics (that is, the free energy of denaturation $\Delta G_{D\text{-}N}^\circ$, D: denatured, N: native) and structural changes (that is, the net charge Z_{CE} and effective hydrodynamic radius R of proteins in both the native and denatured states) associated with protein denaturation, and hence it is well suited to study protein stability. Measurements of the temperature at which different rungs of the ladder denature provide values of $\Delta G_{D\text{-}N}^\circ$ for those rungs and demonstrate the effects of net charge on stability. The Z_{CE} and R values of the native and denatured species reflect the changes in the number of bound protons and in the hydrodynamic size that accompany denaturation.

10.2. Using CE to Measure the Free Energy of Protein Denaturation

The procedure for measuring the free energy of protein denaturation by using CE is similar to ACE. Rather than using different concentrations of ligand, we measure values of electrophoretic mobility at different temperatures or concentrations of chemical denaturants (e.g. urea) in the electrophoresis buffer. Changes in mobility are proportional to the equilibrium distribution of proteins between native and denatured states, and thus can be used to calculate the equilibrium constant $K_{D\text{-}N}$ for denaturation.[176] The details of the procedure and the relationships between the melting temperature and free energy of denaturation $\Delta G_{D\text{-}N}^\circ$ are described elsewhere.[56,147]

10.3. Monitoring the Thermal Denaturation of a Charge Ladder of α-Lactalbumin by CE

α-Lactalbumin (α-LA, MW = 14,200 Da, pI = 4.8) denatures under mildly acidic and alkaline conditions.[177] We were interested in the role that proton binding and long-range electrostatic interactions play in the thermal stability of this protein. Figure 12 shows electropherograms of the charge ladder of α-LA produced by the partial acetylation of Lys ε-NH_3^+ groups, measured at different temperatures and corrected for changes in the viscosity resulting from increasing temperature.[56] Shifts in the position of peaks thus reflect only

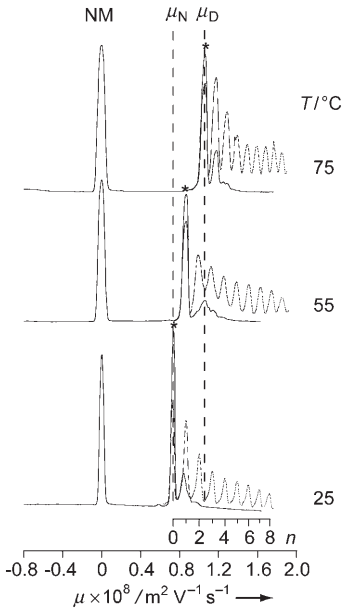


Figure 12. Electropherograms of the charge ladder of α-LA (-----) superimposed on those of the unmodified protein (——) at different values of temperature; NM: neutral marker. The separations were done in a buffer composed of 25 mM Tris/192 mM Gly (pH 8.4), 20 mM NaCl, and 35 μM $CaCl_2$. The peak marked with (*) corresponds to unmodified protein; the number of acetylated Lys ε-NH_3^+ groups *n* is indicated below the corresponding peak. The vertical dashed lines indicate the electrophoretic mobilities of unmodified α-LA ($n=0$) in the native (μ_N) and denatured (μ_D) states. The rungs of the charge ladder, as well as the peak that corresponds to the unmodified protein, broaden near the melting temperature of α-LA (ca. 56 °C). This broadening may reflect the finite rate of interconversion of the native and denatured states or heterogeneity in the free energies of unfolding of different derivatives of α-LA that have the same number of acetylated Lys ε-NH_3^+ groups. (Reproduced with permission from Ref. [56].)

changes in the charge and hydrodynamic drag of the protein with increasing temperature. The impurity to the right of the main peak in the unmodified sample has approximately the same mobility as the first rung of the charge ladder; it is possible that this peak corresponds to α-LA with a deamidated asparagine or glutamine side chain, with a change in charge similar to that produced by the acetylation of a Lys ε-NH_3^+ group.[178]

α-LA forms a compact denatured state, known as a molten globule, with a melting temperature of 56 °C under the conditions of these separations (pH 8.4 Tris-Gly buffer containing 25 mM NaCl and 35 μM $CaCl_2$).[102,179] Individual rungs of the charge ladder were resolved at all temperatures as the protein was denatured. We assume that all the proteins in all the rungs have completed the transition to the compact denatured state at 75 °C. An interesting feature of these data is that the mobility of the proteins increases upon denaturation. We expected the configuration of the protein in the molten globule to be more extended than the native state and, therefore, the value of f_{eff} to increase as the protein denatured. Since the mobility is the ratio of Z_{CE} to f_{eff}, the observed increase in mobility must be due to an increase in the magnitude of the net charge of the protein: that is, changes in Z_{CE} outweigh effects of changes in f_{eff} on the mobility of α-LA as it denatured.

To quantify these effects we analyzed the data in Figure 12 by using the methods described in Section 8.2 to determine *R* and Z_{CE} of α-LA as it denatured thermally. Figure 13 A shows

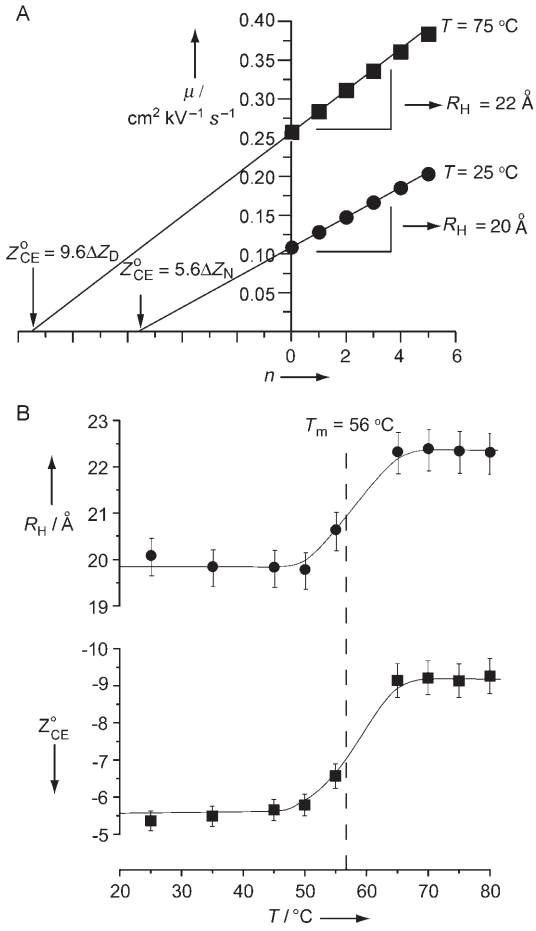


Figure 13. A) Determination of the charge Z°_{CE} and hydrodynamic radius *R* of proteins in the native and denatured state. Values of the electrophoretic mobility of the rungs of the charge ladder of α-LA from Figure 12 are plotted as a function of *n*. ●: native protein at 25 °C, ■: compact denatured state at 75 °C. The net charges $Z^\circ_{CE,N} = -5.4$ (native protein) and $Z^\circ_{CE,D} = -9.2$ (denatured protein), were calculated assuming $\Delta Z_N = \Delta Z_D = -0.96$. The hydrodynamic radius *R* is determined from the slope of the line using Henry's model. B) Values of the hydrodynamic radius *R* and net charge Z°_{CE} of α-LA determined at different temperatures from analysis of the data in (A). The dashed line indicates the melting temperature of this protein (56 °C). (Reproduced with permission from Ref. [56].)

plots of mobility versus *n* for the charge ladder of α-LA measured in the native state (25 °C) and thermally denatured states (75 °C). Linear least-squares analysis of these data yielded values of Z_{CE} and *R* for the unmodified protein at different temperatures (Figure 13 B). We measured an increase in *R* of approximately 0.2 nm or 11 %, relative to the native state, as α-LA transformed from the native to the compact denatured state. A large change in the magnitude of the net charge accompanied this relatively small change in hydrodynamic size: the value of Z°_{CE} nearly doubled from

-5.4 ($Z^{\circ}_{CE,N} = 5.6\Delta Z$, $\Delta Z_N = -0.96$) in the native state to -9.2 ($Z^{\circ}_{CE,D} = 9.6\Delta Z$, $\Delta Z_D = -0.96$) in the denatured state. We calculated the value of ΔZ for the native state by using the LL model (Section 8.1.1). Since charge regulation is expected to be less significant in the denatured state than in the native state, ΔZ_D was assumed to be the same as ΔZ_N. Figure 13B shows that values of Z°_{CE} and R of the unmodified protein could be measured throughout the transition from the native to the denatured states, and the melting temperature of a protein can be estimated from that set of data.

Similar denaturation experiments carried out with the neutral denaturant urea indicated an increase in R of approximately 0.5 nm on denaturation: this value is approximately twice that measured for thermal denaturation. The change in net charge from $5.6\Delta Z$ in the native state to $9.0\Delta Z$ in the denatured state measured with urea denaturation was similar to that measured with thermal denaturation. These results suggest the significant change in the Z°_{CE} value that accompanies denaturation is independent of the denaturant in the case of α-LA.

10.4. The Role of Proton Equilibria in the Thermal Stability of α-Lactalbumin

The measured change in the Z°_{CE} value that accompanied thermal denaturation of α-LA indicated that several titratable residues had shifted their extents of protonation in the denatured ensemble from their values in the native state. This observation implies that proton equilibria were linked energetically to denaturation, and values of $\Delta G^{\circ}_{D\text{-}N}$ should depend on the pH value.[21] This dependence was quantified by Wyman and Gill[96] as a thermodynamic identity [Eq. (30)].

$$\frac{\partial \Delta G^{\circ}_{D\text{-}N}}{\partial \mathrm{pH}} = 2.303\, RT(Q_n - Q_D) \qquad (30)$$

This expression describes the contribution of differences in proton equilibria between the native and denatured states to the pH-dependence of the free energy of unfolding, where Q_N is the number of protons associated with the protein in the native state, and Q_D is the number in the denatured state.

We used CE to measure values of $\Delta G^{\circ}_{D\text{-}N}$ for unmodified α-LA at different values of pH, and determined a value of $\partial \Delta G^{\circ}_{D\text{-}N}/\partial \mathrm{pH} = 5.3$ kcal mol^{-1} per unit of pH.[56,147] From Equation (30), this value corresponds to a $\Delta Q_{D\text{-}N} = -3.9$—a value similar to $\Delta Z_{CE,D\text{-}N} = -3.8$ measured by using CE and charge ladders. We conclude that at pH 8.4, values of $\Delta Z_{CE,D\text{-}N}$ reflect primarily a difference in the extent of protonation between native and denatured states of α-LA arising from differences in pK_a values, dielectric constants, and electrostatic potentials of the protein in the native and denatured states.

10.5. The Role of Net Charge in the Thermal Stability of α-Lactalbumin

To quantify the effects of long-range electrostatics on the thermal stability, we determined the value of $\Delta G^{\circ}_{D\text{-}N}$ for the rungs of the charge ladder of α-LA by determining the fraction of denatured protein for each rung as a function of temperature.[56] Figure 14 compares values of $\Delta G^{\circ}_{D\text{-}N}$ for the

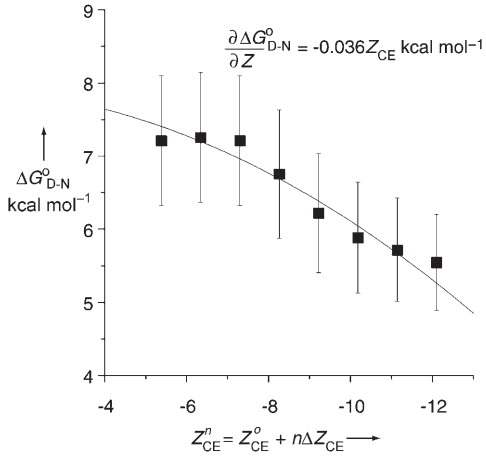


Figure 14. Dependence of free energy of unfolding $\Delta G^{\circ}_{D\text{-}N}$ on the net charge Z^{n}_{CE} of the rungs of charge ladders of α-LA. (Reproduced with permission from Ref. [56].)

members of ladders of α-LA as a function of the net charge measured by CE (using $\Delta Z = -0.96$, calculated by the LL model). The solid line is a quadratic regression fit of the data. We chose this functional form for the dependence of $\Delta G^{\circ}_{D\text{-}N}$ on the net charge because it is consistent with a simple model, in which the electrostatic free energy of protein denaturation is proportional to the free energy of charging a sphere, with the charge distributed evenly on the surface. This charging free energy is proportional to the square of net charge.[180] The fact that this quadratic equation fits the data implies that this model is consistent with the results for α-LA.

The slope of the curve in Figure 14 provides the differential dependence of $\Delta G^{\circ}_{D\text{-}N}$ of α-LA on the net charge $\partial \Delta G^{\circ}_{D\text{-}N}/\partial Z_{CE} = -0.036\, Z_{CE}$ kcal mol^{-1} per unit increase in net charge. The effect of acetylation on the thermal stability of α-LA therefore increases as more amino groups are acetylated (that is, as the protein becomes more negatively charged). This differential provides a direct measure of the average effects of long-range electrostatic interactions to the free energy of folding of α-LA. In the case of α-LA, the difference in $\Delta G^{\circ}_{D\text{-}N}$ between the native and the peracetylated protein is not large (ca. 2 kcal mol^{-1}): this particular protein can accommodate changes in several charged groups without significant loss of stability.

10.6. Conclusions

This section has illustrated the quantitation of contributions of both proton binding and long-range electrostatic interactions to the stability of a representative protein, α-LA. The primary influence of the pH value on the stability of α-LA was a shift in the proton equilibria between the native and denatured states (namely, a shift in the pK_a values of several groups that result in a change in the number of bound protons between the native and denatured states). Increasing the net

charge on the protein by chemical modification has comparably little effect on stability. As in the case of ligand binding, the average contribution that a single lysine amino group makes to the stability of a protein is small, although not zero. In contrast to ligand binding, the influence of charged groups on stability is not simply additive: the larger the net charge already on a protein, the greater the influence of an additional charge.

11. The Net Charge of Proteins in the Gas Phase

Electrospray ionization mass spectrometry (ESI-MS) is an important analytical tool for characterizing proteins and protein–ligand complexes.[181] The factors that influence the formation and relative abundance of ions in ESI-MS are not completely understood. We wanted to determine whether the net charge of a protein in solution determines its net charge in the gas phase. Charge ladders of BCAII (Figure 15) provide an excellent system with which to solve this, since each rung of the ladder differs in charge, but not in any other important parameter, such as sequence or state (native versus denatured).[182]

The separation of the charge ladder by CE followed by analysis by ESI-MS confirmed the previous assignment of the composition of the rungs of the charge ladder: each rung is composed of derivatives that have the same number of modified charged groups. The fact that only a few charge states are observed in the gas phase is also consistent with the retention of the compact structure of the native protein (denatured proteins typically demonstrate a different distribution of charge states in the gas phase[183]). More significant in terms of the physical chemistry of ESI is the observation that both the magnitude and distribution of charged states of the ions generated in the gas phase do not correlate with the net charge of the protein derivatives in solution (that is, the number of Lys ε-NH_3^+ groups they contain).

An interesting correlation found through the examination by ESI-MS of charge ladders of BCAII, lysozyme, and bovine pancreatic trypsin inhibitor is that the magnitude of charge states increases with the surface area of the proteins in such a way that the surface density of charge is retained over a fairly narrow range of 0.9–1.5 units of charge per 1000 Å^2. This observation suggests that proton binding to the proteins in the gas phase is determined largely by electrostatic interactions among charged groups, and not by the intrinsic affinity of basic functional groups of the protein for protons as measured, for example, by values of proton affinity of amino acids in the vapor phase.

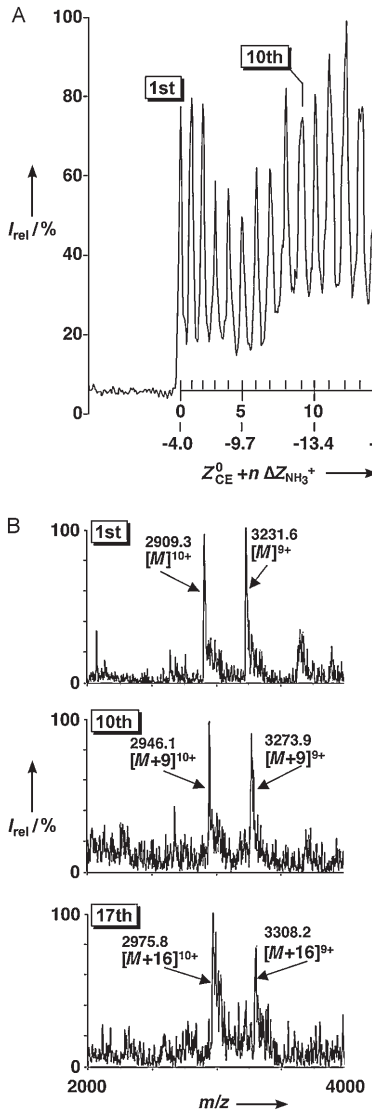


Figure 15. A) CE-ESI-MS electropherogram of an acetyl charge ladder of BCAII. The number of acetylated Lys ε-NH_3^+ groups n, as well as the estimated charge $Z^0_{CE}+n\Delta Z$ of the proteins that make up each of the rungs of the ladder, are indicated below the electropherogram. B) ESI mass spectra obtained from the accumulation of scans for the 1st, 10th, and 17th peaks to emerge after separation by CE. Each peak is labeled with its value of m/z and identified using the notation $[M+n]^{\pm z}$. The quantity in brackets represents the net mass of the ion, where M is the mass of the native protein; the superscript $\pm z$ represents the charge of the ion in the gas phase. (Reproduced from Ref. [182].)

12. Using Charge Ladders to Study the Role of Electrostatics in Bioprocessing

Ultrafiltration is a common technique for the purification of proteins. The selectivity of an ultrafiltration membrane for a particular protein is measured by its value of sieving coefficient S, the ratio of concentration of the protein in the filtrate and the retentate. The values of the sieving coefficient depend both on the size of a protein and on the electrostatic interactions between the protein and the membrane.[82] Thus, for a particular combination of protein and membrane, the sieving coefficient is a function of both the ionic strength and the pH value of the solution. Changes in the pH value can alter the density of charged groups on both the protein and the membrane.[83] Changes in ionic strength can alter the electrostatic free energy of interaction between the charged protein and charged membrane.[184] If there is significant charge regulation between charged groups on either the protein or the membrane, then the density of charged groups

can also change with ionic strength. It has therefore been difficult to determine directly the effects of net charge of the protein in the ultrafiltration process.

Menon and Zydney[84,85] used charge ladders of BCAII and myoglobin to measure the contributions of net charge to the ultrafiltration selectivity of these proteins, and to test models of this process. At the pH value of their experiments (pH 7), both unmodified proteins were negatively charged. The values of S decreased with increasing magnitude of net (negative) charge for charge ladders of both proteins through a negatively charged membrane (Figure 16). Selectivity also

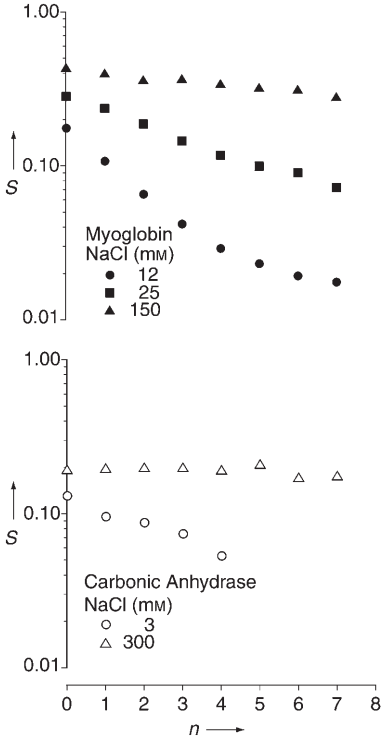


Figure 16. Effect of charge of myoglobin and BCAII on S through negatively charged polyethersulfone (Biomax) ultrafiltration membranes at different ionic strengths. (Reproduced with permission from Ref. [85].)

depended on the ionic strength of the solution. In the presence of 150 mM NaCl, the value of S for the different rungs of the charge ladder of myoglobin went from about 0.45 for the unmodified protein to 0.30 for the rung composed of proteins with eight acetylated amino groups; in the presence of 12 mM NaCl, the value of S went from approximately 0.2 to 0.02 for the same proteins. At a concentration of 300 mM NaCl, there was no measurable difference in the values of S for the rungs of the charge ladder of BCAII.

Menon and Zydney[85] rationalized these changes in S as resulting from an increase in the exclusion of the protein from the membrane as a result of the repulsion between the negatively charged walls of the pores of the membrane and the negatively charged proteins. Increases in ionic strength effectively screened these interactions, decreased the magnitude of the electrostatic free energy of interaction between the protein and membrane, and thereby reduced the dependence of S on the net charge of the protein. The authors calculated the effects of charge on S by using an equilibrium model of partitioning of proteins into charged, cylindrical pores.[185] This model predicted values of S accurately for the first four rungs of the charge ladder of myoglobin at an ionic strength of 12 mM. The later rungs of the charge ladder, which had the largest magnitude of net charge, showed significant deviations between the model and experiments.

Menon and Zydney[85] pointed out that CE and charge ladders should also be useful in measuring electrostatic interactions involving membrane systems in much the same way that solutions of dextran with different distributions in molecular weight have been useful in measuring the effects of steric interactions.[186] The combination of charge ladders and CE can also be extended to other bioprocessing techniques. Examples would include ion-exchange chromatography (where proteins are separated on the basis of their affinity for charged surfaces[187]) and aqueous two-phase extraction (where proteins are separated on the basis of their relative solubility in the two immiscible solutions of polymers.[188]).

The development of models that predict the selectivity of these different bioprocessing strategies on the basis of charge is an on-going, fundamental research effort in chemical engineering.[82–86,88,105,184,189] The value of such models is that they could lead to strategies for the engineering of proteins by manipulating their composition to improve processability, similar to the way in which proteins are currently engineered for improved stability and activity. Charge ladders, and their extensions to other properties of proteins such as hydrodynamic size and surface hydrophobicity, will help to improve our understanding of the fundamental interactions that serve as the basis for bioprocessing techniques.

13. Summary and Outlook

13.1. Summary

This Review describes the procedures used to form, characterize, and manipulate protein charge ladders, and to use them to explore problems in the physical chemistry of proteins. The combination of protein charge ladders and CE provides internally self-consistent data (in the form of electrophoretic mobilities) for a large number of derivatives of a protein—which differ incrementally in number of charged groups—in a single, simple experiment. These data permit the extraction of net charge[111,190] and hydrodynamic radius[58] of the parent, native protein.

The chemical modifications used to generate charge ladders are performed on folded proteins, and thus involve charged residues on their exterior. There is currently no way to target groups in specific locations, so charge ladders are not appropriate for studying the roles of specific residues, such as those buried in the interior of the protein or those in its active site; questions concerning these roles are best addressed by site-directed mutagenesis. Charge ladders are, however, very useful for examining collective effects of charged residues, and for studying phenomena that originate on the surface of a

protein. Electrostatic interactions in proteins are often relatively weak ($< RT$, Figure 3), but numerous, and are thus difficult to quantify individually. Charge ladders and CE allow simultaneous analysis of a large number of charge variants, and thus provide a way to begin to quantify those interactions.

We have examined the effect of net charge on ligand binding[4] (with carbonic anhydrase, Figure 11), on thermal stability[56] (with α-lactalbumin, Figure 14), and on ion formation in the vapor phase[182] (with carbonic anhydrase, Figure 15). Zydney and co-workers have used charge ladders to study the effects of the net charge of proteins on their transport through charged ultrafiltration membranes[84,85] (Figure 16). Other phenomena that are likely to be dependent on the surface charges—aggregation, crystallization, protein–protein association, catalytic activity—have not been yet examined, but their examination should be facilitated by experiments using protein charge ladders.

We have also combined experimental data from CE and charge ladders with continuum electrostatic theory to validate the ability of the theory to predict the effects of net charge on the binding of charged and neutral ligands;[5] the theory provided details of electrostatic interactions that could not be revealed experimentally (Figure 11). This combination of theory and experiment circumvents some of the limitations of both charge ladders and site-directed mutagenesis alone. It can provide information on the role of individual charges in specific locations and can analyze the electrostatic interactions in proteins with many permutations of charges.

The ability to extract the values of net charge of the native protein, and of its derivatives, depends on the ability to estimate the value of ΔZ—the difference in charge between successive rungs of a ladder, or between proteins differing by one modification. The value of ΔZ is influenced not only by the charge of the group that is modified, but also by the cooperative interactions among other ionizable groups on the protein. Although we cannot measure ΔZ directly, we can estimate it theoretically by using three different models:

1) The Linderstrøm-Lang model. This model requires knowing the composition of amino acids of the protein and the pK_a values of ionizable residues, and an estimate of the effective hydrodynamic radius of the protein.[105,106]
2) Poisson–Boltzmann calculations with Monte Carlo simulations of proton equilibria. This technique requires knowledge of the three-dimensional structure of the protein and significant numerical computation.[107]
3) Screened Coulombic potential model. This simple model also requires knowing the three-dimensional structure of the protein (especially the distances between ionizable groups), but does not require significant computation: it can be solved on a spreadsheet.[107]

Values of ΔZ can also be estimated experimentally by combining measured electrophoretic mobilities of charge ladders with an independent measurement of hydrodynamic radius; this measurement can be accomplished conveniently on a CE instrument by using analysis of Taylor's dispersion.[168] At pH values of about 8.5 and low (ca. 10 mM) ionic strengths, the value of ΔZ per acetylation of a lysine lies between −0.8 and −0.9, based on the analysis of several proteins. At high ionic strengths (ca. 100 mM), the value of ΔZ approaches −1.

The concept of "ladders" can also be extended to properties of proteins other than charge. We have described derivatives of proteins with incremental changes in hydrodynamic size (Figure 9) and proposed how the surface hydrophobicity of proteins can be altered by successive acylation of amino groups (Figure 4). In these experiments, CE serves primarily to count the number of modifications of the protein, not to estimate the charge.

13.2. Outlook

Protein charge ladders, site-directed mutagenesis, and theory allow us to begin to ask sophisticated, focused questions about roles of charged groups in macromolecules. Here we give several representative questions, and sketch experimental evidence relevant to their eventual answer; in no case do we have a complete answer, and in no case can these questions be easily answered with one experimental technique.

13.2.1. Is it the Value of Net Charge of a Protein that is Important in the Intracellular and Extracellular Environments, or Is it the Distribution of Charges on a Protein That Is Important?

To start to address this question, we analyzed the structure of charges and net charge of 20 isoforms of carbonic anhydrase (CA).[131] This analysis showed that the net charge Z_{seq} calculated from the sequence of amino acids varied from +1.8 to −2.3, with 90% of the isoforms falling between $Z = +1.5$ and −1.5, even though the total number of charges (positive and negative) ranged from 36 to 75. These results imply that there is evolutionary pressure to maintain the net charge of CA between +1 and −2, even if the total number of charged groups is not conserved. It would be interesting to extend the analysis to other families of proteins.

It is possible that retaining a small net charge on proteins both reduces the osmotic pressure[191] and, at the same time, prevents phase separation inside the cell[192] and minimizes nonspecific protein–protein interactions in an environment crowded with ions, nucleic acids, and proteins.[193] Sear has started to explore this subject by calculating the net charge Z_{seq} of all proteins (ca. 4000) in the genome of *E. Coli*.[191] The distribution of net charge is approximately Gaussian, centered around −3, with a standard deviation of 8. Few proteins (< 1%) had magnitudes of charge greater than 30. Such analysis, done on the proteome of an organism, supports the idea that net charge, as a parameter, is important for the function of a cell.

The distribution of charge in proteins is certainly involved in formation of protein–protein complexes,[8] binding of highly charged nucleic acids,[24] and may be important in more subtle phenomena, such as electrostatic steering.[10,78] Thus, the conservation of both net charge and of charge distribution are likely to occur in some circumstance. We

therefore cannot yet answer the question of the relative importance of net charge and charge distribution.

13.2.2. Are Charged Groups on Proteins Involved in "Negative Design"?

In the complex environment of a cell, specific interactions required for life must be balanced against other interactions detrimental to life;[194,195] the latter may result in crystallization and homogeneous or heterogeneous aggregation. Charged groups on the surfaces of proteins may be involved in the "negative design", where the presence of charge does not necessarily improve the function of an enzyme, but instead reduces undesirable, noncatalytic interactions (such as the protein–protein interactions involved in crystallization or aggregation).

Richardson and Richardson described a specific example of the role of charges in negative design in β-sheet proteins.[195] Naturally occurring β-sheet proteins are usually soluble, but fragments or mutants of these proteins can aggregate and, in some cases, form amyloid fibers. An analysis of 75 proteins having β-sheet structures led the authors to conclude that the presence of a single charged group on one edge of the β sheet was one of the most common natural features associated with inhibition of aggregation; the authors also argued that out of all available charged residues, Lys would be the best choice because of its flexibility. These arguments, supported by experiments done by Wang and Hecht,[196] showed that this design principle was effective in converting proteins, designed de novo to form β sheets and to assemble into amyloid-like fibers, into proteins that remained as soluble monomers in solution.

13.2.3. Do Cells Utilize Electrostatics To Control the Activity and Stability of Proteins?

Cells actively regulate the concentration of ions and protons in compartmentalized volumes: for example, endosomes and lysosomes are quite acidic (pH 5–6), while the cytosolic pH value is close to neutral (ca. 7.2).[197] During apoptosis, cells undergo acidification of their intracellular pH value by 0.3–0.4 units,[198] sometimes with a transient alkalinization. The cytosolic pH values of tumor and multidrug resistant cells are different from those of normal cells.[199] Cells utilize these pH differences, and thus differences in protonation states of proteins, to activate certain enzymes and pathways only where and when necessary.

One example of such control is the transport of proteins to the lysosomes. Proteins to be transported to the lysosome are tagged with a marker, mannose-6-phosphate, in the Golgi complex (at neutral pH). Vesicles containing a receptor for mannose-6-phosphate bind the marked proteins. These carriers then fuse with pre-lysosomal vesicles that are acidic; lowering of pH value causes the marked protein to dissociate from the receptor and continue its way to the lysosome.[41]

Another question that rises immediately is: "*How do lysosomal enzymes refrain from uncontrolled, and likely destructive, activity en route to the lysosome?*" Heikinheimo et al.[200] proposed the following mechanism for the activation of lysosomal α-mannosidase at a low pH value: at cytosolic pH values, multiple salt bridges stiffen the loops near the active site of the enzyme, thus reducing the catalytic activity; at the acidic pH value of the lysosome, protonation of glutamic and aspartic acids breaks the network of salt bridges and allows for the necessary plasticity of the enzyme required for catalysis. There are many other examples of pH-activated events (activation of hemagglutinin at low pH values in influenza infection may be another well-known one[201]), and all cannot be listed here. The point is that the forces that are modulated by a change in the pH value are in part electrostatic.

13.2.4. Can One Make "Better" Proteins—That is, Proteins with Improved Stability, Greater Activity, or New Functionality—by Designing an Appropriate Pattern of Charge?

Changing the patterns of charge on protein surfaces may be a strategy to achieve desired properties of proteins for applications in biotechnology: for example, in organic synthesis, diagnostics and sensing applications, and protein therapy. Leist et al. have recently demonstrated an excellent example of an improved charge-modified therapeutic with erythropoietin (EPO). Carbamylation of lysine residues on EPO did not interfere with its binding to the tissue-protective receptors, but interfered with its binding to the receptor responsible for the hematopoietic activity of EPO.[202] Thus, EPO-stimulated tissue-protective pathways could be activated without the unwanted effects of up-regulated production of red blood cells.

Sarkar et al. demonstrated a clever utilization of the difference in electrostatic interactions between a therapeutic protein and its receptor at two different pH values to control the lifetime of the therapeutic.[203] The authors engineered a cytokine GCSF (granulocyte colony-stimulating factor) to contain a His residue at the binding interface to its receptor (GCSFR). At the extracellular pH value of about 7 when the His is deprotonated, GCSF maintained its tight binding to the receptor. At the endosomal pH value of 5–6, unfavorable electrostatic interactions resulting from the protonation of His reduced the affinity of GCSFR to bind GCSF by several fold. As a result of the weakened interaction of the cytokine with its receptor in the endosome, the cytokine was recycled back to the surface of the cell rather than being sent for degradation in the lysosome. This increased recycling of the engineered GCSF increased the lifetime and, thereby, the activity of the cytokine, relative to the wild-type. The authors believe that modulation of binding affinity between ligands and receptors by carefully engineered "histidine switches" may be a general strategy for the enhancement of endosomal recycling.

13.2.5. What Are the Energetics of Interaction and Forces between Ions and Charged Macromolecules in Solution?

Before we are able to answer any questions about the role of charges on proteins in complex mixtures, we need to understand better the behavior of ions in solution. Interestingly, we do not understand the behavior of even simple ions

in solutions,[50,74,75] especially at biologically relevant concentrations.

In contrast to a vacuum, where the electrostatic free energy of interaction between charged groups is entirely enthalpic, the free energy of interaction between charged groups in biological fluids is often dominated by the contributions of entropy. It is the entropic contribution to the electrostatic free energy that can make the electrostatic interactions between opposite charges favorable. The electrostatic enthalpy of interaction between opposite charges can actually be unfavorable in aqueous solutions.

Ninham and co-workers[74,75] have shown that charge–charge interactions do not always dominate the total free energy of interactions between ions. Dispersion forces among ions and water must be included in accurate thermodynamic descriptions of solutions of electrolytes. A striking, counterintuitive result of the studies by Ninham and co-workers is that the net interaction between some ions of *like* charge is *attractive* when dispersion forces are included.[75] The dispersion force depends on the size and polarizability of the ions, and therefore, may play an important role in describing the effects of specific ions in experimentally measured osmotic coefficients and surface tensions. The authors have also argued that dispersion forces can explain such observations as the dependence of net charge and ionization constants of a protein on the background electrolyte,[51] and the "salting-in" and "salting-out" of proteins by different salts (also known as Hofmeister effect).[204] These specific ion effects, which are still incompletely understood, are omitted in most models of protein solutions. A more complete picture of protein–ion interactions will need to include these effects.

In addition, if we are to understand fully electrostatic interactions in biology, we must measure more than just equilibrium constants and the associated standard state free energies of interaction; we must also know the individual contributions of enthalpy and entropy to this free energy. To see a complete picture, we will need to measure the thermodynamic parameters that describe molecular recognition directly, by using techniques such as calorimetry.

While experiments with protein charge ladders by themselves will not be able to directly answer the questions we have posed, and others of a similar character, they will certainly advance our understanding of electrostatic interactions involving proteins in solutions. Complimentary calculations with continuum electrostatic theory will add to this understanding. Protein charge ladders, and extensions to other ladders that systematically alter the surface chemistry of proteins, will, we believe, also aid in the development of tools for protein separation and characterization.

This work was supported by the National Institute of Health (grant GM 51559 (G.M.W.)), the National Science Foundation (grant CTS-0134429), and the Camille and Henry Dreyfus Foundation (J.D.C.). We thank Katie Gudiksen, Demetri Moustakas, Logan McCarty, and Max Narovlyansky for helpful discussions.

Received: July 20, 2005
Published online: April 18, 2006

[1] M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **2004**, *303*, 186.
[2] A. R. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, *3rd ed.*, Freeman, New York, **1998**.
[3] S. Miller, J. Janin, A. M. Lesk, C. Chothia, *J. Mol. Biol.* **1987**, *196*, 641.
[4] J. M. Gao, M. Mammen, G. M. Whitesides, *Science* **1996**, *272*, 535.
[5] J. A. Caravella, J. D. Carbeck, D. C. Duffy, G. M. Whitesides, B. Tidor, *J. Am. Chem. Soc.* **1999**, *121*, 4340.
[6] D. F. Green, B. Tidor, *J. Mol. Biol.* **2004**, *342*, 435.
[7] C. Park, R. T. Raines, *J. Am. Chem. Soc.* **2001**, *123*, 11472.
[8] S. J. Davis, E. A. Davies, M. G. Tucknott, E. Y. Jones, P. A. van der Merwe, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5490.
[9] F. Dong, M. Vijayakumar, H. X. Zhou, *Biophys. J.* **2003**, *85*, 49; F. B. Sheinerman, B. Honig, *J. Mol. Biol.* **2002**, *318*, 161; L. P. Lee, B. Tidor, *Nat. Struct. Biol.* **2001**, *8*, 73.
[10] G. Schreiber, A. R. Fersht, *Nat. Struct. Biol.* **1996**, *3*, 427.
[11] G. Schreiber, A. R. Fersht, *J. Mol. Biol.* **1995**, *248*, 478.
[12] I. Moarefi, D. Jeruzalmi, J. Turner, M. O'Donnell, J. Kuriyan, *J. Mol. Biol.* **2000**, *296*, 1215; D. Tworowski, A. V. Feldman, M. G. Safro, *J. Mol. Biol.* **2005**, *350*, 866.
[13] I. Muegge, P. X. Qi, A. J. Wand, Z. T. Chu, A. Warshel, *J. Phys. Chem. B* **1997**, *101*, 825; R. G. Alden, W. W. Parson, Z. T. Chu, A. Warshel, *J. Am. Chem. Soc.* **1995**, *117*, 12284; K. K. Rodgers, S. G. Sligar, *J. Am. Chem. Soc.* **1991**, *113*, 9419.
[14] Y. X. Jiang, V. Ruta, J. Y. Chen, A. Lee, R. MacKinnon, *Nature* **2003**, *423*, 42; D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. L. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, R. MacKinnon, *Science* **1998**, *280*, 69; R. S. Eisenberg, *J. Membr. Biol.* **1996**, *150*, 1.
[15] S. Kumar, R. Nussinov, *J. Mol. Biol.* **1999**, *293*, 1241.
[16] H. X. Zhou, F. Dong, *Biophys. J.* **2003**, *84*, 2216; S. Kumar, R. Nussinov, *Proteins Struct. Funct. Genet.* **2000**, *41*, 485; S. Kumar, B. Y. Ma, C. J. Tsai, R. Nussinov, *Proteins Struct. Funct. Genet.* **2000**, *38*, 368; S. Spector, M. H. Wang, S. A. Carp, J. Robblee, Z. S. Hendsch, R. Fairman, B. Tidor, D. P. Raleigh, *Biochemistry* **2000**, *39*, 872.
[17] Z. S. Hendsch, B. Tidor, *Protein Sci.* **1994**, *3*, 211.
[18] D. Stigter, D. O. V. Alonso, K. A. Dill, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4176.
[19] M. Hollecker, T. E. Creighton, *Biochim. Biophys. Acta* **1982**, *701*, 395.
[20] A. H. Elcock, *J. Mol. Biol.* **1999**, *294*, 1051; M. Schaefer, M. Sommer, M. Karplus, *J. Phys. Chem. B* **1997**, *101*, 1663.
[21] A. S. Yang, B. Honig, *J. Mol. Biol.* **1993**, *231*, 459.
[22] K. L. Shaw, G. R. Grimsley, G. I. Yakovlev, A. A. Makarov, C. N. Pace, *Protein Sci.* **2001**, *10*, 1206.
[23] J. P. Schmittschmitt, J. M. Scholtz, *Protein Sci.* **2003**, *12*, 2374; F. Chiti, M. Stefani, N. Taddei, G. Ramponi, C. M. Dobson, *Nature* **2003**, *424*, 805.
[24] B. Honig, A. Nicholls, *Science* **1995**, *268*, 1144.
[25] A. Warshel, A. Papazyan, *Curr. Opin. Struct. Biol.* **1998**, *8*, 211; Y. Y. Sham, Z. T. Chu, A. Warshel, *J. Phys. Chem. B* **1997**, *101*; T. Simonson, *Curr. Opin. Struct. Biol.* **2001**, *11*, 243; N. Sinha, S. J. Smith-Gill, *Curr. Protein Pept. Sci.* **2002**, *3*, 601.
[26] F. B. Sheinerman, R. Norel, B. Honig, *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.
[27] W. Kauzmann, *Adv. Protein Chem.* **1959**, *12*, 1.
[28] K. A. Dill, *Biochemistry* **1990**, *29*, 7133.
[29] K. A. Dill, S. Bromberg, K. Z. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, H. S. Chan, *Protein Sci.* **1995**, *4*, 561.
[30] H. S. Chan, K. A. Dill, *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447.
[31] K. A. Dill, *Protein Sci.* **1999**, *8*, 1166.

[32] C. N. Pace, B. A. Shirley, M. McNutt, K. Gajiwala, *FASEB J.* **1996**, *10*, 75.
[33] P. L. Privalov, S. J. Gill, *Adv. Protein Chem.* **1988**, *41*, 191.
[34] G. Hummer, S. Garde, A. E. Garcia, M. E. Paulaitis, L. R. Pratt, *J. Phys. Chem. B* **1998**, *102*, 10469.
[35] C. Hansch, *Acc. Chem. Res.* **1993**, *26*, 147.
[36] J. Kyte, *Biophys. Chem.* **2003**, *100*, 193.
[37] N. T. Southall, K. A. Dill, A. D. J. Haymet, *J. Phys. Chem. B* **2002**, *106*, 521.
[38] D. A. McQuarrie, *Statistical Mechanics*, Harper & Row, New York, **1975**.
[39] P. G. de Gennes, *Scaling concepts in polymer physics*, Cornell University Press, Ithaca, NY, **1979**.
[40] K. A. Dill, *Biochemistry* **1985**, *24*, 1501.
[41] L. Stryer, *Biochemistry, 4th ed.*, Freeman, New York, **1995**.
[42] D. Voet, J. G. Voet, *Biochemistry, 3rd ed.*, Wiley, New York, **2004**.
[43] B. Schiott, B. B. Iversen, G. K. H. Madsen, F. K. Larsen, T. C. Bruice, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12799; A. Warshel, *J. Biol. Chem.* **1998**, *273*, 27035; W. W. Cleland, P. A. Frey, J. A. Gerlt, *J. Biol. Chem.* **1998**, *273*, 25529; J. G. Chen, M. A. McAllister, J. K. Lee, K. N. Houk, *J. Org. Chem.* **1998**, *63*, 4611; E. L. Ash, J. L. Sudmeier, E. C. DeFabo, W. W. Bachovchin, *Science* **1997**, *278*, 1128; A. Warshel, A. Papazyan, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13665; J. P. Guthrie, *Chem. Biol.* **1996**, *3*, 163; A. Warshel, A. Papazyan, P. A. Kollman, *Science* **1995**, *269*, 102; W. W. Cleland, M. M. Kreevoy, *Science* **1994**, *264*, 1887; P. A. Frey, S. A. Whitt, J. B. Tobin, *Science* **1994**, *264*, 1927.
[44] D. Bashford, M. Karplus, *Biochemistry* **1990**, *29*, 10219.
[45] C. Tanford, *Adv. Prot. Chem.* **1962**, *17*, 69.
[46] C. N. Schutz, A. Warshel, *Proteins* **2001**, *44*, 400.
[47] T. Simonson, C. L. Brooks III, *J. Am. Chem. Soc.* **1996**, *118*, 8452.
[48] K. A. Dill, S. Bromberg, *Molecular Driving Forces*, Garland, New York, **2003**.
[49] B. N. Dominy, H. Minoux, C. L. Brooks, *Proteins Struct. Funct. Genet.* **2004**, *57*, 128.
[50] K. D. Collins, M. W. Washabaugh, *Q. Rev. Biophys.* **1985**, *18*, 323.
[51] M. Bostrom, D. R. M. Williams, B. W. Ninham, *Biophys. J.* **2003**, *85*, 686.
[52] M. Bostrom, D. R. M. Williams, B. W. Ninham, *Eur. Phys. J. E* **2004**, *13*, 239.
[53] F. H. Arnold, G. Georgiou in *Methods in Molecular Biology, Vol. 230*, Humana, Totowa, NJ, **2003**.
[54] I. J. Colton, J. R. Anderson, J. Gao, R. G. Chapman, L. Isaacs, G. M. Whitesides, *J. Am. Chem. Soc.* **1997**, *119*, 12701.
[55] Y. Hagihara, Y. Tan, Y. Goto, *J. Mol. Biol.* **1994**, *237*, 336.
[56] R. S. Negin, J. D. Carbeck, *J. Am. Chem. Soc.* **2002**, *124*, 2911.
[57] J. D. Carbeck, I. J. Colton, J. Gao, G. M. Whitesides, *Acc. Chem. Res.* **1998**, *31*, 343.
[58] J. D. Carbeck, R. S. Negin, *J. Am. Chem. Soc.* **2001**, *123*, 1252.
[59] K. Sharp, B. Honig, *J. Phys. Chem.* **1990**, *94*, 7684; W. Rocchia, E. Alexov, B. Honig, *J. Phys. Chem. B* **2001**, *105*, 6507.
[60] A. R. Fersht, S. E. Jackson, L. Serrano, *Philos. Trans. R. Soc. Lond. A* **1993**, *345*, 141; A. R. Fersht, L. Serrano, *Curr. Opin. Struct. Biol.* **1993**, *3*, 75.
[61] H. Nakamura, *Q. Rev. Biophys.* **1996**, *29*, 1; T. Simonson, *Rep. Prog. Phys.* **2003**, *66*, 737; K. Sharp, B. Honig, *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*.
[62] D. J. Winzor, *Anal. Biochem.* **2004**, *325*, 1.
[63] L. R. Deyoung, A. L. Fink, K. A. Dill, *Acc. Chem. Res.* **1993**, *26*, 614.
[64] G. H. Iyer, S. Dasgupta, J. A. Bell, *J. Cryst. Growth* **2000**, *217*, 429.
[65] A. George, W. W. Wilson, *Acta Crystallogr. Sect. D* **1994**, *50*, 361.
[66] T. Hill, *An Introduction to Statistical Thermodynamics*, Dover, **1986**.
[67] P. M. Tessier, A. M. Lenhoff, S. I. Sandler, *Biophys. J.* **2002**, *82*, 1620; P. M. Tessier, A. M. Lenhoff, *Curr. Opin. Biotechnol.* **2003**, *14*, 512.
[68] P. B. Warren, *J. Phys. Condens. Matter* **2002**, *14*, 7617; A. H. Elcock, J. A. McCammon, *Biophys. J.* **2001**, *80*, 613; W. C. K. Poon, S. U. Egelhaaf, P. A. Beales, A. Salonen, L. Sawyer, *J. Phys. Condens. Matter* **2000**, *12*, L569; R. C. Chang, D. Asthagiri, A. M. Lenhoff, *Proteins Struct. Funct. Genet.* **2000**, *41*, 123; C. J. Coen, H. W. Blanch, J. M. Prausnitz, *AIChE J.* **1995**, *41*, 996.
[69] B. L. Neal, D. Asthagiri, O. D. Velev, A. M. Lenhoff, E. W. Kaler, *J. Cryst. Growth* **1999**, *196*, 377.
[70] B. Guo, S. Kao, H. McDonald, A. Asanov, L. L. Combs, W. W. Wilson, *J. Cryst. Growth* **1999**, *196*, 424.
[71] J. P. K. Doye, A. A. Louis, M. Vendruscolo, *Phys. Biol.* **2004**, *1*, 9.
[72] A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. King, G. B. Benedek, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6116; P. G. Vekilov, A. R. Feeling-Taylor, D. N. Petsev, O. Galkin, R. L. Nagel, R. E. Hirsch, *Biophys. J.* **2002**, *83*, 1147.
[73] J. Z. Wu, D. Bratko, J. M. Prausnitz, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15169.
[74] M. Bostrom, B. W. Ninham, *J. Phys. Chem. B* **2004**, *108*, 12593.
[75] W. Kunz, L. Belloni, O. Bernard, B. W. Ninham, *J. Phys. Chem. B* **2004**, *108*, 2398.
[76] P. M. Tessier, A. M. Lenhoff, *Curr. Opin. Biotechnol.* **2003**, *14*, 512.
[77] A. H. Elcock, D. Sept, J. A. McCammon, *J. Phys. Chem. B* **2001**, *105*, 1504.
[78] C. J. Camacho, Z. Weng, S. Vajda, C. DeLisi, *Biophys. J.* **1999**, *76*, 1166.
[79] A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425; A. Warshel, *Acc. Chem. Res.* **2002**, *35*, 385.
[80] A. Warshel, J. Florian, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5950; A. H. Elcock, M. J. Potter, D. A. Matthews, D. R. Knighton, J. A. McCammon, *J. Mol. Biol.* **1996**, *262*, 370; J. Antosiewicz, S. T. Wlodek, J. A. McCammon, *Biopolymers* **1996**, *39*, 85.
[81] J. M. Sanchez-Ruiz, G. I. Makhatadze, *Trends Biotechnol.* **2001**, *19*, 132.
[82] D. B. Burns, A. L. Zydney, *AIChE J.* **2001**, *47*, 1101.
[83] D. B. Burns, A. L. Zydney, *Biotechnol. Bioeng.* **1999**, *64*, 27.
[84] M. F. Ebersold, A. L. Zydney, *Biotechnol. Bioeng.* **2004**, *85*, 166.
[85] M. K. Menon, A. L. Zydney, *J. Membr. Sci.* **2001**, *181*, 179.
[86] D. Nagrath, B. W. Bequette, S. M. Cramer, A. Messac, *AIChE J.* **2005**, *51*, 511; C. A. Brooks, S. M. Cramer, *AIChE J.* **1992**, *38*, 1969.
[87] Y. Yao, A. M. Lenhoff, *Anal. Chem.* **2004**, *76*, 6743; J. Stahlberg, *J. Chromatogr. A* **1999**, *855*, 3.
[88] C. A. Haynes, F. J. Benitez, H. W. Blanch, J. M. Prausnitz, *AIChE J.* **1993**, *39*, 1539; A. P. Sassi, A. J. Shaw, S. M. Han, H. W. Blanch, J. M. Prausnitz, *Polymer* **1996**, *37*, 2151.
[89] W. L. Jorgensen, D. S. Maxwell, J. TiradoRives, *J. Am. Chem. Soc.* **1996**, *118*, 11225; W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179; D. Sitkoff, K. A. Sharp, B. Honig, *J. Phys. Chem.* **1994**, *98*, 1978; M. E. Davis, J. D. Madura, B. A. Luty, J. A. McCammon, *Comput. Phys. Commun.* **1991**, *62*, 187; B. H. Besler, K. M. Merz, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*, 431; B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187.

[90] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
[91] A. Giletto, C. N. Pace, *Biochemistry* **1999**, *38*, 13379.
[92] B. E. Garcia-Moreno, J. J. Dwyer, A. G. Gittis, E. E. Lattman, D. S. Spencer, W. E. Stites, *Biophys. Chem.* **1997**, *64*, 211.
[93] D. E. Anderson, W. J. Becktel, F. W. Dahlquist, *Biochemistry* **1990**, *29*, 2403.
[94] L. Polgar, *Euro. J. Biochem.* **1973**, *33*, 104.
[95] C. A. Ibarra, B. S. Nieslanik, W. M. Atkins, *Biochemistry* **2001**, *40*, 10614.
[96] J. Wyman, S. J. Gill, *Binding and Linkage: The Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, **1990**.
[97] C. Tanford, R. Roxby, *Biochemistry* **1972**, *11*, 2192.
[98] R. Roxby, C. Tanford, *Biochemistry* **1971**, *10*, 3348.
[99] Y. Tan, M. Oliveberg, A. R. Fersht, *J. Mol. Biol.* **1995**, *254*, 980.
[100] S. Kuramitsu, K. Hamaguchi, *J. Biochem.* **1980**, *87*, 1215.
[101] A. L. Fink, L. J. Calciano, Y. Goto, M. Nishimura, S. A. Swedberg, *Protein Sci.* **1993**, *2*, 1155.
[102] S. Kim, J. Baum, *Protein Sci.* **1998**, *7*, 1930.
[103] Y. Goto, N. Takanashi, A. L. Fink, *Biochemistry* **1990**, *29*, 3480.
[104] Y. Goto, L. J. Calciano, A. L. Fink, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 573.
[105] M. K. Menon, A. L. Zydney, *Anal. Chem.* **2000**, *72*, 5714.
[106] I. Gitlin, M. Mayer, G. M. Whitesides, *J. Phys. Chem. B* **2003**, *107*, 1466.
[107] U. Sharma, R. S. Negin, J. D. Carbeck, *J. Phys. Chem. B* **2003**, *107*, 4653.
[108] K. C. Linderstrom-Lang, *C. R. Trav. Lab. Carlsberg* **1924**, *15*, 1.
[109] M. Lund, B. Jonsson, *Biochemistry* **2005**, *44*, 5722; T. Simonson, J. Carlsson, D. A. Case, *J. Am. Chem. Soc.* **2004**, *126*, 4167; T. Grycuk, *J. Phys. Chem. B* **2002**, *106*, 1434.
[110] J. M. Gao, F. A. Gomez, R. Harter, G. M. Whitesides, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12027.
[111] J. D. Carbeck, I. J. Colton, J. R. Anderson, J. M. Deutch, G. M. Whitesides, *J. Am. Chem. Soc.* **1999**, *121*, 10671.
[112] M. C. Justice, J. C. Justice, *J. Solution Chem.* **1976**, *5*, 543; H. Yokoyama, H. Yamatera, *Bull. Chem. Soc. Jpn.* **1975**, *48*, 1770.
[113] J. E. Prue, *J. Chem. Educ.* **1969**, *46*, 12.
[114] M. Mihailescu, M. K. Gilson, *Biophys. J.* **2004**, *87*, 23.
[115] D. R. Lide, *Handbook of Chemistry and Physics*, *84th ed.*, CRC, Boca Raton, **2003**.
[116] D. F. Evans, H. Wennerstrom, *The colloidal domain: Where physics, chemistry, biology and technology meet*, *1st ed.*, VCH, New York, **1994**.
[117] J. Israelachvili, *Intermolecular and Surface Forces: With Applications to Colloidal and Biological Systems*, *2nd ed.*, Academic Press, New York, **1992**.
[118] A. V. Morozov, T. Kortemme, D. Baker, *J. Phys. Chem. B* **2003**, *107*, 2075.
[119] D. Perl, U. Mueller, U. Heinemann, F. X. Schmid, *Nat. Struct. Biol.* **2000**, *7*, 380.
[120] P. Barth, I. Guillouard, P. Setif, B. Lagoutte, *J. Biol. Chem.* **2000**, *275*, 7030.
[121] R. Loewenthal, J. Sancho, T. Reinikainen, A. R. Fersht, *J. Mol. Biol.* **1993**, *232*, 574.
[122] S. Dao-pin, U. Sauer, H. Nicholson, B. W. Matthews, *Biochemistry* **1991**, *30*, 7142.
[123] S. E. Jackson, A. R. Fersht, *Biochemistry* **1993**, *32*, 13909.
[124] S. J. Benkovic, S. Hammes-Schiffer, *Science* **2003**, *301*, 1196.
[125] J. Warwicker, H. C. Watson, *J. Mol. Biol.* **1982**, *157*, 671.
[126] A. Nicholls, B. Honig, *J. Comput. Chem.* **1991**, *12*, 435.
[127] P. D. Grossman, J. C. Colburn, *Capillary Electrophoresis. Theory and Practice*, Academic Press, San Diego, CA, **1992**.
[128] A. R. Fersht, *FEBS Lett.* **1993**, *325*, 5; A. R. Fersht, *Philos. Trans. R. Soc. Lond. B* **1995**, *348*, 11.
[129] F. Chiti, M. Calamai, N. Taddei, M. Stefani, G. Ramponi, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16419.
[130] D. Matulis, J. K. Kranz, F. R. Salemme, M. J. Todd, *Biochemistry* **2005**, *44*, 5258.
[131] K. Gudiksen, I. Gitlin, J. Yang, A. R. Urbach, D. T. Moustakas, G. M. Whitesides, *J. Am. Chem. Soc.* **2005**, *127*, 4707.
[132] C. T. Supuran, A. Scozzafava, A. Casini, *Med. Res. Rev.* **2003**, *23*, 146.
[133] J. Gao, M. Mrksich, F. Gomez, G. M. Whitesides, *Anal. Chem.* **1995**, *67*, 3093.
[134] E. Cordova, J. Gao, G. M. Whitesides, *Anal. Chem.* **1997**, *69*, 1370.
[135] J. E. Coligan, B. M. Dunn, H. L. Ploegh, D. W. Speicher, P. T. Wingfield, *Current Protocols in Protein Science*, Wiley, New York, **2003**.
[136] A. N. Glazer, *Annu. Rev. Biochem.* **1970**, *39*, 101.
[137] J. R. Anderson, O. Cherniavsky, I. Gitlin, G. S. Engel, L. Yuditsky, G. M. Whitesides, *Anal. Chem.* **2002**, *74*, 1870.
[138] S. Rimon, G. E. Perlmann, *J. Biol. Chem.* **1968**, *243*, 3566; V. V. Loladze, G. I. Makhatadze, *Protein Sci.* **2002**, *11*, 174.
[139] I. Gitlin, J. Yang, G. M. Whitesides, unpublished results.
[140] D. G. Hoare, D. E. Koshland, Jr., *J. Biol. Chem.* **1967**, *242*, 2447.
[141] K. Toi, E. Bynum, E. Norris, H. A. Itano, *J. Biol. Chem.* **1967**, *242*, 1036.
[142] L. Patthy, E. L. Smith, *J. Biol. Chem.* **1975**, *250*, 557; L. Patthy, E. L. Smith, *J. Biol. Chem.* **1975**, *250*, 565.
[143] A. A. Vinogradov, E. V. Kudryashova, V. Y. Grinberg, N. V. Grinberg, T. V. Burova, A. V. Levashov, *Protein Eng.* **2001**, *14*, 683.
[144] U. Sharma, J. D. Carbeck, *Electrophoresis* **2005**, *26*, 2086.
[145] J. Yang, I. Gitlin, V. M. Krishnamurthy, J. A. Vazquez, C. E. Costello, G. M. Whitesides, *J. Am. Chem. Soc.* **2003**, *125*, 12392.
[146] S. Julka, F. E. Regneir, *Anal. Chem.* **2004**, *76*, 5799.
[147] U. Sharma, J. D. Carbeck, *Methods Mol. Biol.* **2004**, *276*, 189.
[148] M. Mammen, I. J. Colton, J. D. Carbeck, R. Bradley, G. M. Whitesides, *Anal. Chem.* **1997**, *69*, 2165.
[149] S. Hjerten, *J. Chromatogr.* **1985**, *347*, 191; J. Horvath, V. Dolnik, *Electrophoresis* **2001**, *22*, 644.
[150] H. Katayama, Y. Ishihama, N. Asakawa, *Anal. Chem.* **1998**, *70*, 5272; T. W. Graul, J. B. Schlenoff, *Anal. Chem.* **1999**, *71*, 4007.
[151] R. J. Hunter, *Foundations of Colloid Science*, *Vol. 1*, Oxford University Press, New York, **1991**.
[152] D. C. Henry, *Proc. R. Soc. London Ser. A* **1931**, *133*, 106.
[153] $f(\kappa R) = (1 + \frac{\kappa^2 R^2}{16} - \frac{5\kappa^3 R^3}{48} - \frac{\kappa^4 R^4}{96} + \frac{\kappa^5 R^5}{96} - \frac{11}{98} e^{\kappa R} \int_{\infty}^{\kappa R} \frac{e^{-r}}{r} dr)$;
in this expression, κ [m^{-1}] is the Debye screening parameter, r is a dimensionless distance (namely, distance multiplied by κ) and R [m] is the radius of the particle.
[154] This assumption comes from the linearization of the Poisson–Boltzmann equation which involves expanding the exponential and retaining only the term linear in potential. Doing so is valid when $e\psi_s/kT$ and $|\psi_s| < 25$ mV at 25 °C.
[155] J. G. De la Torre, M. L. Huertas, B. Carrasco, *Biophys. J.* **2000**, *78*, 719.
[156] B. J. Yoon, S. Kim, *J. Colloid Interface Sci.* **1989**, *128*, 275.
[157] R. W. O'Brien, L. R. White, *J. Chem. Soc. Faraday Trans. 2* **1978**, *74*, 1607.
[158] M. V. Piaggio, M. B. Peirotti, J. A. Deiber, *Electrophoresis* **2005**, *26*, 3232.
[159] S. A. Allison, J. D. Carbeck, C. Y. Chen, F. Burkes, *J. Phys. Chem. B* **2004**, *108*, 4516.
[160] W. B. Russel, D. A. Saville, W. R. Schowalter, *Colloidal Dispersions*, Cambridge University Press, Cambridge, UK, **1999**.

[161] S. A. Allison, V. T. Tran, *Biophys. J.* **1995**, *68*, 2261; S. A. Allison, *Macromolecules* **1996**, *29*, 7391; S. A. Allison, M. Potter, J. A. McCammon, *Biophys. J.* **1997**, *73*, 133.
[162] L. L. Kiefer, S. A. Paterno, C. A. Fierke, *J. Am. Chem. Soc.* **1995**, *117*, 6831.
[163] M. Raghavan, V. R. Bonagura, S. L. Morrison, P. J. Bjorkman, *Biochemistry* **1995**, *34*, 14649; W. L. Martin, A. P. West, L. Gan, P. J. Bjorkman, *Mol. Cell* **2001**, *7*, 867; E. R. Sprague, W. L. Martin, P. J. Bjorkman, *J. Biol. Chem.* **2004**, *279*, 14184.
[164] P. Beroza, D. A. Case, *J. Phys. Chem.* **1996**, *100*, 20156.
[165] K. K. Lee, C. A. Fitch, B. E. Garcia-Moreno, *Protein Sci.* **2002**, *11*, 1004.
[166] U. Sharma, J. D. Carbeck, unpublished results.
[167] M. S. Bello, R. Rezzonico, P. G. Righetti, *Science* **1994**, *266*, 773.
[168] U. Sharma, N. J. Gleason, J. D. Carbeck, *Anal. Chem.* **2005**, *77*, 806.
[169] G. Taylor, *Proc. R. Soc. London B* **1953**, *140*, 219.
[170] E. A. Permyakov, L. A. Morozova, E. A. Burstein, *Biophys. Chem.* **1985**, *21*, 21.
[171] A. G. Ogston, *Trans. Faraday Soc.* **1958**, *54*, 1754.
[172] I. J. Colton, J. D. Carbeck, J. Rao, G. M. Whitesides, *Electrophoresis* **1998**, *19*, 367.
[173] Y.-H. Chu, L. Z. Avila, J. Gao, G. M. Whitesides, *Acc. Chem. Res.* **1995**, *28*, 461; R. M. Guijt-van Duijn, J. Frank, G. W. van Dedem, E. Baltussen, *Electrophoresis* **2000**, *21*, 3905.
[174] L. Z. Avila, Y.-H. Chu, E. C. Blossey, G. M. Whitesides, *J. Med. Chem.* **1993**, *36*, 126.
[175] R. Lavecchia, M. Zugaro, *FEBS Lett.* **1991**, *292*, 162.
[176] V. J. Hilser, E. Freire, *Anal. Biochem.* **1995**, *224*, 465.
[177] F. M. Robbins, F. E. Andreotti, L. G. Holmes, M. J. Kronman, *Biochim. Biophys. Acta* **1967**, *133*, 33.
[178] S. J. Wearne, T. E. Creighton, *Proteins Struct. Funct. Genet.* **1989**, *5*, 8.
[179] C. Redfield, B. A. Schulman, M. A. Milhollen, P. S. Kim, C. M. Dobson, *Nat. Struct. Biol.* **1999**, *6*, 948.
[180] C. Tanford, *Physical Chemistry of Macromolecules*, Wiley, New York, **1961**.
[181] F. Rusconi, R. Guillonneau, D. Praseuth, *Mass Spectrom. Rev.* **2002**, *21*, 5; M. Mann, R. C. Hendrickson, A. Pandey, *Annu. Rev. Biochem.* **2001**, *70*, 437.
[182] J. D. Carbeck, J. C. Severs, J. Gao, Q. Wu, R. D. Smith, G. M. Whitesides, *J. Phys. Chem. B* **1998**, *102*, 10596.
[183] J. A. Loo, R. R. Loo, H. R. Udseth, C. G. Edmonds, R. D. Smith, *Rapid Commun. Mass Spectrom.* **1991**, *5*, 101.
[184] N. S. Pujar, A. L. Zydney, *Ind. Eng. Chem. Res.* **1994**, *33*.
[185] F. G. Smith, W. M. Deen, *J. Colloid Interface Sci.* **1980**, *78*, 444.
[186] L. Zeman, A. L. Zydney, *Microfiltration and Ultrafiltration: Principles and Applications*, Marcel Dekker, New York, **1996**.
[187] F. E. Regneir, *Science* **1987**, *238*, 319.
[188] A. D. Diamond, J. T. Hsu, *Adv. Biochem. Eng./Biotechnol.* **1992**, *47*.
[189] N. S. Pujar, A. L. Zydney, *J. Chromatogr. A* **1998**, *796*, 229.
[190] J. Gao, G. M. Whitesides, *Anal. Chem.* **1997**, *69*, 575.
[191] R. P. Sear, *J. Chem. Phys.* **2003**, *118*, 5157.
[192] R. P. Sear, J. A. Cuesta, *Phys. Rev. Lett.* **2003**, *91*; B. L. Neal, D. Asthagiri, A. M. Lenhoff, *Biophys. J.* **1998**, *75*, 2469.
[193] R. J. Ellis, A. P. Minton, *Nature* **2003**, *425*, 27; D. Hall, A. P. Minton, *Biochim. Biophys. Acta* **2003**, *1649*, 127.
[194] J. S. Richardson, D. C. Richardson, N. B. Tweedy, K. M. Gernert, T. P. Quinn, M. H. Hecht, B. W. Erickson, Y. B. Yan, R. D. McClain, M. E. Donlan, M. C. Surles, *Biophys. J.* **1992**, *63*, 1186; M. J. Nohaile, Z. S. Hendsch, B. Tidor, R. T. Sauer, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3109.
[195] J. S. Richardson, D. C. Richardson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754.
[196] W. Wang, M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2760.
[197] B. Alberts, J. Alexander, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, *4th ed.*, Garland, New York, **2002**.
[198] D. Lagadic-Gossmann, L. Huc, V. Lecureur, *Cell Death Differ.* **2004**, *11*, 953.
[199] S. Simon, D. Roy, M. Schindler, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1128.
[200] P. Heikinheimo, R. Helland, H. K. S. Leiros, I. Leiros, S. Karlsen, G. Evjen, R. Ravelli, G. Schoehn, R. Ruigrok, O. K. Tollersrud, S. McSweeney, E. Hough, *J. Mol. Biol.* **2003**, *327*, 631.
[201] M. Lakadamyali, M. J. Rust, X. W. Zhuang, *Microb. Infect.* **2004**, *6*, 929.
[202] M. Leist, P. Ghezzi, G. Grasso, R. Bianchi, P. Villa, M. Fratelli, C. Savino, M. Bianchi, J. Nielsen, J. Gerwien, P. Kallunki, A. K. Larsen, L. Helboe, S. Christensen, L. O. Pedersen, M. Nielsen, L. Torup, T. Sager, A. Sfacteria, S. Erbayraktar, Z. Erbayraktar, N. Gokmen, O. Yilmaz, C. Cerami-Hand, Q. W. Xie, T. Coleman, A. Cerami, M. Brines, *Science* **2004**, *305*, 239.
[203] C. A. Sarkar, K. Lowenhaupt, T. Horan, T. C. Boone, B. Tidor, D. A. Lauffenburger, *Nat. Biotechnol.* **2002**, *20*, 908.
[204] M. Bostrom, V. S. J. Craig, R. Albion, D. R. M. Williams, B. W. Ninham, *J. Phys. Chem. B* **2003**, *107*, 2875; M. Bostrom, W. Kunz, B. W. Ninham, *Langmuir* **2005**, *21*, 2619; M. C. Gurau, S. M. Lim, E. T. Castellana, F. Albertorio, S. Kataoka, P. S. Cremer, *J. Am. Chem. Soc.* **2004**, *126*, 10522.

