ELECTROSTATIC EFFECTS IN PROTEINS ARE GOVERNED BY REDISTRIBUTION OF THE CONFORMATIONAL ENSEMBLE

by

Christos Michael Kougentakis

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

August 2020

Abstract

The ability to regulate pH dynamics is one of the hallmarks of all living cells. pH affects biological function predominantly through its effect on proteins, through the binding or release of H^+ to ionizable groups. To understand the structural basis of the pH sensitivity of proteins, it is necessary to understand the molecular basis of H^+ binding events. The H^+-binding affinities of ionizable groups (i.e. pK_a values) are known to be sensitive to the properties of their microenvironments. The inability of sophisticated computational methods for structure-based calculations of pK_a values to reproduce the experimentally measured pK_a values suggest that H+ binding to proteins is a more complex process than currently acknowledged. To examine in detail the molecular determinants of pK_a values, I studied variants of staphylococcal nuclease with buried Lys residues. These buried Lys residues were chosen because they have highly anomalous pK_a values (depressed, which favors the neutral state) that are difficult to reproduce with structure-based pK_a calculations. In dehydrated environments, electrostatic effects such as Coulomb interactions and dehydration energies are stronger relative to those on the well hydrated protein surface and are highly sensitive to the polarity and polarizability of their environments.

The studies described in this dissertation examined the hypothesis that the magnitude of electrostatic effects in proteins are not as severely influenced by the properties of the microenvironments of ionizable moieties as they are by the propensity of the conformational ensemble to redistribute in response to changes in pH. In chapter 2, the physical properties of the buried Lys residues were probed using ^{15}N NMR spectroscopy. In most cases, pH-dependent structural reorganization was evident in the pH titrations tracked with the $^{15}N\zeta$ resonance of the buried Lys residues. In chapter 3, the nature of this reorganization was characterized in 25 variants of SNase with internal Lys residues. These studies used backbone-based NMR spectroscopy

experiments. The results of this work demonstrate that the apparent pK_a values of the Lys residues measured by linkage thermodynamics report on pH-driven redistribution of the conformational ensemble. The redistribution favors conformational states in which the charged Lys side chains are exposed to solvent. By correlating thermodynamic data (i.e. stabilithy) with detailed structural information about the partially unfolded states that SNase can access, it was possible to map the conformational ensemble of this protein with unprecedented resolution. In chapter 4, the structural and thermodynamic consequences of buried ion pairs in SNase was examined. NMR spectroscopy and X-ray crystallography showed various degrees of structural reorganization that lead to the hydration of at least one of the buried groups. These results demonstrate that not only does structural reorganization govern the properties of individual buried groups, structural reorganization can also determine the properties of buried ion pairs. In chapter 5 I provide irrefutable evidence of strong apparent interactions between buried and surface ionizable groups. NMR spectroscopy and equilibrium thermodynamic experiments demonstrated that the interactions between internal and surface ionizable groups are not Coulomb interactions. Instead, the apparent interactions are mediated by conformational reorganization driven by the ionization of the buried residues.

Most of what is known about protein electrostatics reflects insights from decades of structure-based calculations with continuum electrostatic models, and more recently with constant pH molecular dynamics calculations. The data presented in this dissertation demonstrate that conformational reorganization can be the dominant factor that determines the pK_a values of ionizable groups in proteins. Structure-based prediction of conformational reorganization remains extremely challenging; my results suggest an explanation for the short-comings of structure-based calculations to reproduce experimental pK_a values and other electrostatic properties of proteins.

Advisor:	Bertrand García-Moreno, E., Ph.D.
Reader:	Juliette Lecomte, Ph.D.
Committee:	Doug Barrick, Ph.D.
	Vincent Hilser, Ph.D.
	Albert Lau, Ph.D.

Dedication

This thesis is dedicated to the memory of Anna Kougentakis (née Tzagournis, 1955-2015), for instilling in me the value of education, for encouraging me to pursue my interests, and for making all of this possible.

Acknowledgements

First and foremost, I'd like to thank my PI, Bertrand García-Moreno, for providing me with the right balance of guidance and intellectual freedom necessary to pursue ambitious projects, the end results which are described in this thesis. I'll extend that thanks to the mentors who trained me in molecular biology and biochemistry prior to graduate school; Elizabeth Miller, Amie McClellan, and Amy Keating, as well as all the post-docs and grad students in their labs who gave me the foundations necessary to succeed in graduate school. Thanks to my informal mentors at Hopkins- especially Ananya Majumdar, Jamie Schlessman, and Katie Tripp, for giving me the technical background necessary for my graduate work, as well as my thesis committee (Juliette Lecomte, Vince Hilser, Doug Barrick, and Albert Lau) for critical feedback and inspiration.

I'd also like to thank my peers, both in the García-Moreno/Lecomte labs (Jaime M., Dillon, Anne, Aaron, Miranda H., Jaime S.), as well as my PMB cohort (Miranda R., Emily, Matt, Dagan, Gadd, Mel), for providing both intellectual and personal support throughout my graduate experience. Finally, thanks to my family for support and understanding during my time in Baltimore, especially my sister Alex and Dad; all the non-Hopkins people who made my life great (especially Rebeca and Amber!) and all the great places in Baltimore that kept me sane and entertained for my six years in The Greatest City in America.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many proteins can respond to environmental signals and transduce them to perform biological work. A simple chemical signal that is ubiquitous in the cell and is capable of affecting the structure, function, and stability of proteins is proton (H^+) concentration (pH)[1]. Changes in intracellular and extracellular pH (pHi and pHe, respectively) regulate a diverse range of biological processes, including growth[2], stress[3], and cellular organization[4]. In fact, the electrochemical potential (i.e., the proton motive force) established when a proton gradient develops across a bilayer constitutes the fundamental mechanism of energy storage and transduction in living cells[5].

Changes in pH affect biological function primarily through their effects on proteins. All proteins have ionizable groups that can bind or release H^+, thereby changing their charged state. The H^+ binding affinities of these ionizable groups (pK_a) are sensitive to their surrounding microenvironment and therefore to the conformation of the protein. The factors that govern the pK_a values of these residues can be expressed as the sum of contributions from the self-energy experienced by the charge ($\Delta G_{ii,}$), which in turn is governed by the polarity and polarizability of the protein matrix, and by the Coulomb energy (ΔG_{ij}), which reflects the sum of interactions with other charge moieties of the protein. In principle, these two energy terms are sufficient to calculate meaningful electrostatic energies from protein structures, including calculation of pK_a values of ionizable residues. Considering that electrostatic energy governs the majority of chemical processes in the cell, this approach constitutes the most direct way to correlate structure with energy in proteins. Many computational methods for structure-based energy calculations have been developed in the past century to explain many functional properties of proteins (e.g., structure, interactions, stability, solubility, ligand binding properties, allostery and catalysis). Many of these methods use dielectric constants to describe the response of the protein/water system to a change in ionization state[6]. However, even state-of-the art continuum methods fail to

reproduce the experimentally determined electrostatic parameters such as pK_a values of functional residues underscoring the significant limitations of continuum methods based on static protein structures.

The major shortcoming of current structure-based methods for calculation of pK_a values is with the treatment of the free energy of reorganization of the protein (ΔG_R, governed by the free energy gap between the alternative states that a protein can sample in which the microenvironment of an ionizable group is altered)[7,8]. The reorganization energy is of special interest because the ability of proteins to undergo conformational changes in response to proton binding/release is the basis for most forms of biological energy transduction. Accurate treatment of conformational reorganization in structure-based calculations of electrostatic energy has proved to be challenging. Methods for modeling pH-driven structural changes in proteins based on MD simulations that treat ionization reactions explicitly are available[9] but they are not yet reliable. To reproduce conformational transitions triggered by changes in pH they would have to predict the alternative conformations of a protein and estimate accurately the free energy differences between them. This remains a tall order in even the best of cases.

The studies described in this dissertation examine the role of conformational reorganization as the dominant determinant of pK_a values and electrostatic effects in proteins. The data described in Chapters 2 - 5 constitute irrefutable evidence that H^+-binding and release processes in proteins are governed by the energetics of conformational reorganization: the ability of the conformational ensemble to redistribute in response to changes in pH governs the dielectric response, and therefore electrostatic effects, of proteins. My results have important implications for understanding fundamental aspects of the roles of proteins in energy transduction, in pH-sensing, and in allostery and cooperativity. They have implications for our understanding of molecular mechanisms of

pathological conditions (i.e., cancer, Alzheimers, ischemia) where normal pH homeostasis is perturbed[10]. My results also have important implications for structure-based electrostatic calculations as they identify fundamental flaws in existing computational methods. These results challenge the utility of continuum electrostatic models for mechanistic studies of functional processes governed by electrostatic effects. They also demonstrate that accurate electrostatic calculations will require improved capacity to identify alternative conformational states of proteins and accurate calculation of the free energy gaps between them.

1.1 Regulated pH dynamics in cells

1.1.1 pH affects biological processes through its effects on proteins

There is no living cell on Earth that does not require tight regulation of pH for survival[5]; in this sense, it could be said that the ability of biological systems to respond to changes in pH is a central characteristic to life as we know it on Earth. Although many biological macromolecules have protonatable sites, pH affects biological systems predominantly through its effect on proteins: protonation or deprotonation of ionizable residues in proteins lead to changes in charge that can in turn change structure, dynamics, and function[11]. Other post-translational modifications can add or remove charge to protein side chains (i.e., phosphorylation, acetylation), but none with the properties characteristic of protonation reactions. Protonation-deprotonation reactions are governed by properties intrinsic to the protonatable sites, by equilibrium thermodynamics and pH. They do not require the participation of other catalytic agents. The state of protonation can change in timescales faster than the rate of expression of the enzymes needed for posttranslational modifications[12]. This allows cellular pH changes to rapidly influence the functional properties of multiple proteins and their corresponding regulatory pathways. Not surprisingly, all living cells

have evolved the capacity to regulate cellular pH dynamics; this is necessary to maintain life (Fig. 1.1).

The pH inside a resting cell (pHi) is approximately pH 7.2 in human cells. The extracellular pH (pHe) is slightly higher (approximately 7.4)[5]. Cellular pH is regulated through multiple mechanisms: chemical reactions can consume or produce H^+, and cellular buffers (e.g., proteins are buffers) and proton pumps and transporters that can move H^+ across bilayers can in turn regulate pH[5]. It was previously assumed that pHi is effectively invariant; however, it is now understood that cellular pH is highly dynamic and that fluctuations around the expected value (7.0-7.8 for pHi) are critical for growth[2] and differentiation[13]. Dramatic cytosolic pH decreases (pH < 6) have been observed in single-celled eukaryotes in response to heat[14] and nutrient[15] stress conditions. Changes in pHi can have significant consequences on the activity of a myriad of proteins, including those involved in metabolism (i.e., phosphofructokinase[16,17]), transport (i.e., hemoglobin[18,19,20]), and cytoskeletal organization (i.e., actin dynamics[21,22]).

Dysregulation of cellular pH is a hallmark of many pathological conditions, including cancer[10] and neurodegenerative[23] diseases. In cancer cells of solid tumors, the pH gradient across the plasma member is reversed (from pHi ~ 7.2 and pHe ~ 7.4 in normal cells to pHi ≥ 7.4 and pHe ≤ 7.0 in cancer cells). This reversal in the pH gradient happens initially as a consequence of metabolic adaptions for production of ATP in the absence of oxygen in solid tumors (Warburg effect)[24], and subsequently by the overexpression of H^+ pumps that further exacerbate the flip in the pH gradient[25] (this suggests that gene expression itself is regulated by pH by a mechanism that is poorly understood). In addition to the significant changes in biochemical adaptation and cytoskeleton organization caused by changes in pHi the evidence suggests that the acidified pHe

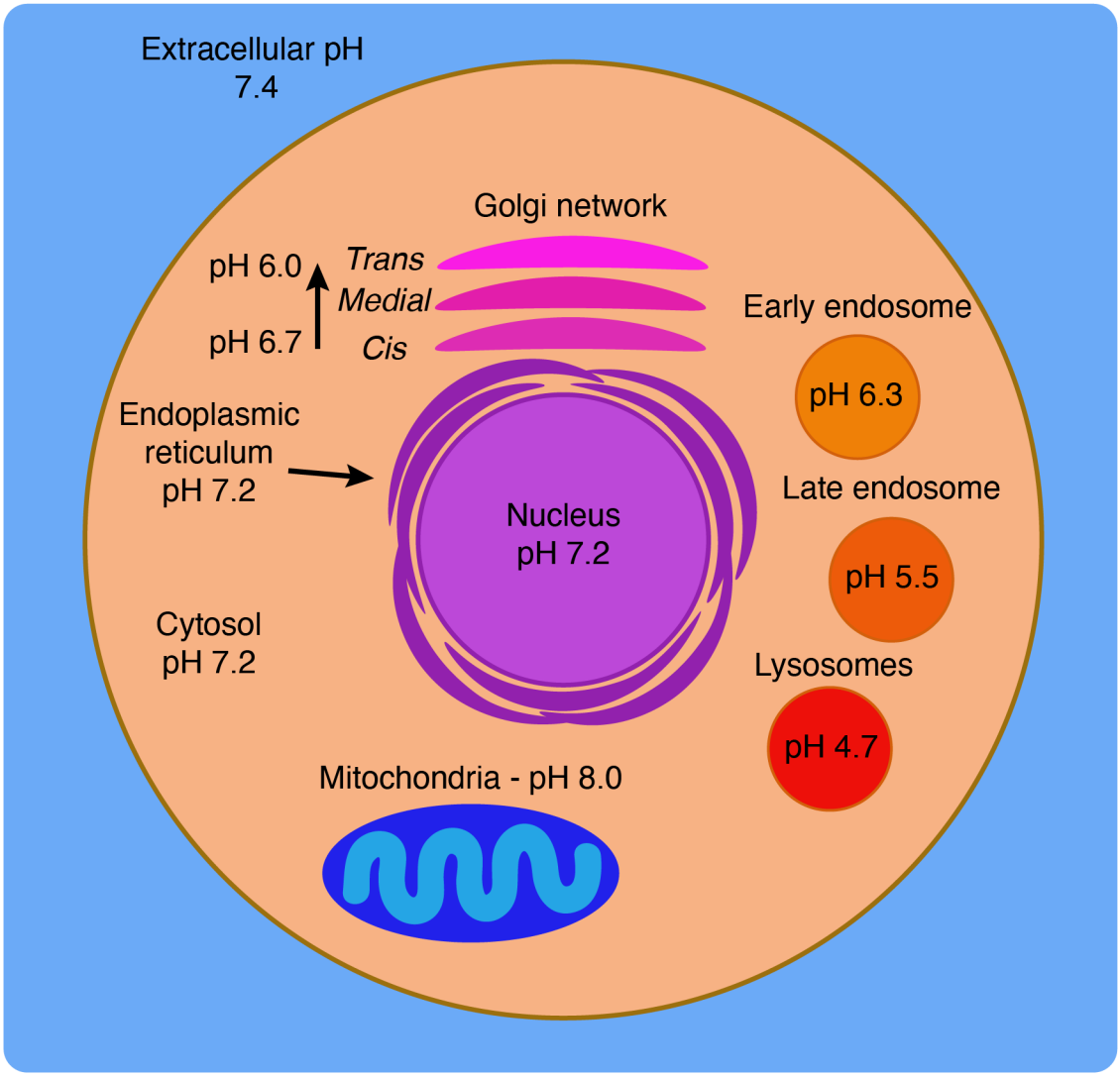

Figure 1.1. pH in cellular compartments. Diagram of a cell identifying the steady state pH values in specific organelles. These values are approximate, and can fluctuate significantly depending on cell cycle stage, environmental conditions, and pathological states. Adapted from reference 7.

in tumor cells promotes migration and invasive behaviors that diminish apoptosis and contribute to metastasis[10].

Inside a normal cell there are pH gradients between the cytosol and many organelles. The pH of membrane-bound organelles varies depending on functional roles; organelle pH dynamics play essential roles in regulating processes such as metabolism and transport. The pH gradient between the cytosol and the more basic mitochondrial interior is necessary for aerobic metabolism[26,27]. Similarly, the pH gradient between compartments of the secretory pathway is important for regulating the conformation and activity of proteins involved in the quality control of cellular trafficking[28]. The decreased pH of lysosomes is necessary for the activity of acid-activated hydrolases that are necessary for catabolic processes in the cell[29]. Many viruses have evolved to exploit natural pH gradients in cells, and to interpret differences in pH between extracellular and intracellular environments as zip codes to detect when they are inside cells in environments amenable for uncoating and replication[30]. These are just some of the ways in which pH changes within a cell and pH differences between membrane bound compartments are involved in regulation of biological function.

1.1.2 pH gradients as energy storage

Some forms of energy transduction in biological systems involve the transformation of the energy stored in proton gradients into energy stored in chemical bonds. ATP-synthase is the quintessential energy transduction machine driven by a H^+-gradient. This is a large membrane protein with two main subunits (F_o and F_1) that can harness the electrochemical potential arising from pH differences across the mitochondrial membrane and use it to synthesize ATP[31]. Specifically, the coupling between the pH gradient and the formation of chemical bonds involves

conformational changes of the protein that are coupled to the protonation-deprotonation of a buried Glu residue in the membrane bound F_o subunit. These protonation-deprotonation reactions lead to conformational changes in the water-soluble F_1 domain where the synthesis of ATP is catalyzed[32,33].

In certain prokaryotic cells, transmembrane proteins such as bacteriorhodopsin[34], can transduce the energy of a photon to create a pH gradient. The critical event in this energy transduction process involve conformational changes that are coupled to proton binding and release that cumulatively lead to the translocation of an H^+ across the bilayer and against the electrochemical gradient. The last enzyme of the electron transport chain in mitochondria, cytochrome c oxidase, operates with a similar mechanism, using proton-coupled electron transfer to pump protons against the pH-gradient with the ultimate goal of driving the synthesis of ATP[35]. In all of these cases the key molecular event involves the creation of charge in an environment that is inherently incompatible with charge, leading to conformational rearrangement that gives rise to biological function.

1.2 Physical chemistry of H^+ binding/release in proteins

1.2.1 Electrostatics provide the most direct means to correlate structure with energy in proteins.

To understand how pH affects proteins it is necessary to understand the molecular determinants of the pK_a values of their ionizable amino acids. Of the 20 amino acids, Arg, His, Lys, Glu, Asp, Tyr, Cys can be ionized in the physiological pH range. The proton binding affinity of the ionizable groups is given by the Boltzmann-weighted probability of being in the proton bound versus unbound state. In principle, these energies can be calculated from first principles from the gas phase[36]; however, this is computationally expensive and difficult as accurate

treatment of solvation energies remains a challenge[37]. In general, the pK_a values of these residues in proteins are calculated relative to the normal values measured in aqueous solutions. The calculation of pK_a values is thus simplified to the calculation of the effects of the microenvironment of the protein on the intrinsic pK_a values measured in solution[7].

The energetic contributions to pK_a shifts observed in the folded state can be described by the contributions of three energetic terms, ΔG_{ij} (Coulomb energy), ΔG_{ii} (self-energy), and ΔG_R (reorganization energy)[7,8] (Fig. 1.2). Structure-based energy calculations have traditionally focused on the first two terms, for which a physical relationship between the electrostatic effects of self and Coulomb interactions and pK_a shifts exists.

$$pK_a^{App} = pK_a^{Int} + \frac{\Delta G_{Elec}}{1.36} \qquad 1.1$$

(assuming a temperature of 25° C)

In this formalism, pK_a^{App} represents the observed log base 10 H^+-binding affinity, pK_a^{Int} is the log base 10 affinity for H^+ in a model compound in water, and ΔG_{Elec} is the sum of electrostatic energies, represented as:

$$\Delta G_{Elec} = \Delta G_{ii} + \Delta G_{ij} \qquad 1.2$$

It is important to note that contributions of the reorganization energy to the pK_a^{App} of ionizable residues cannot be accounted for directly in this formalism, as described below. Rather, it is subsumed in the ΔG_{ij} and ΔG_{ii} terms. There is actually nothing simple about Eq. 1.1. Calculation of ΔG_{Elec} requires knowledge of the charged state of the ionizable groups, which in turn requires knowledge of pK_a^{App}. Recursive relationships and full statistical mechanical methods can be used to solve this problem.

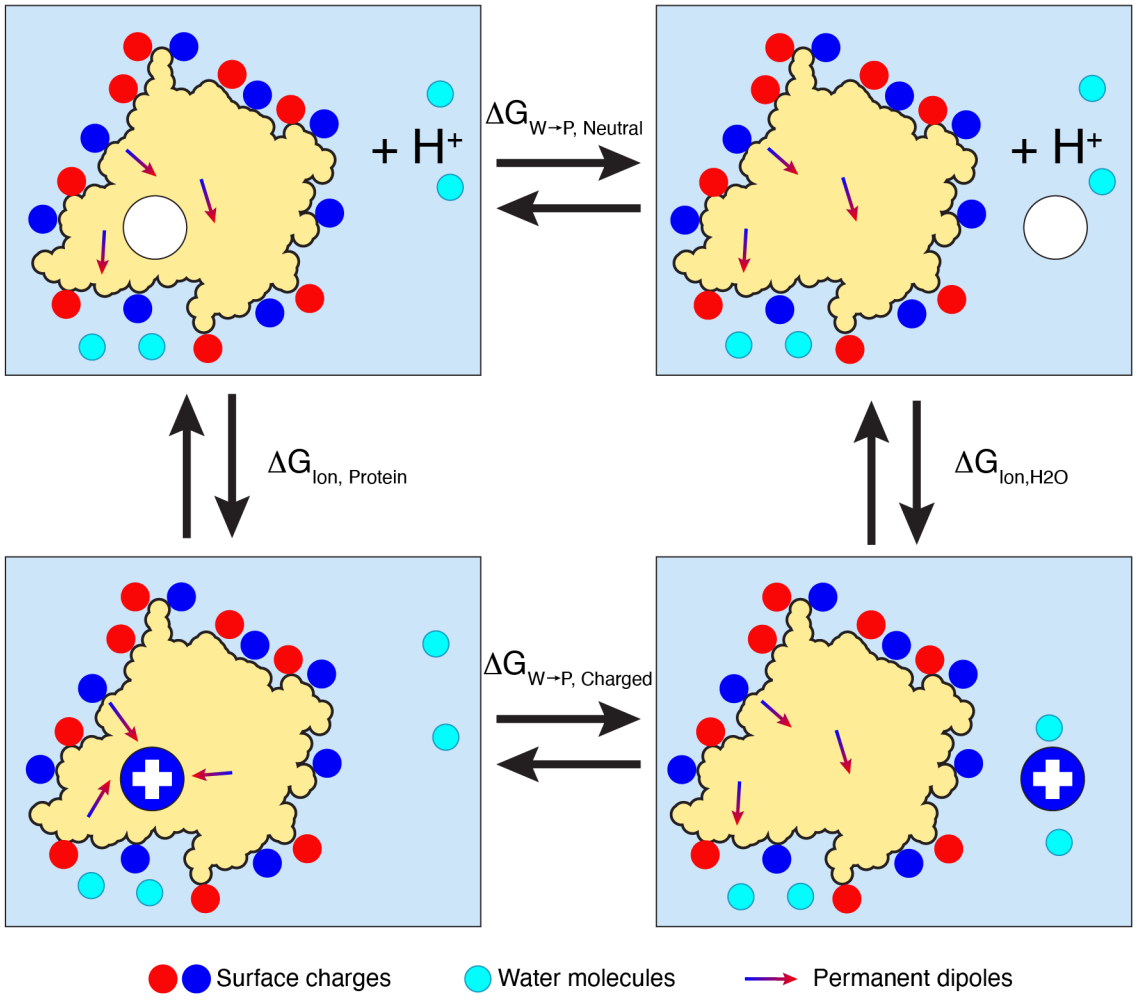

Figure 1.2. Energetics of transfer of a charge from bulk water to the interior of a protein. The $\Delta G_{W->P}$ terms describe the transfer of a neutral or charged residue from bulk solvent to the interior. The ΔG_{Ion} terms describe the energy needed to bind a proton in water and the protein interior.

1.2.2 Coulomb interactions between charges

Coulomb interactions between two charges are distance dependent and can be attractive or repulsive depending on the sign of the interacting charges:

$$\Delta G_{ij} = \frac{z_i z_j}{r_{ij}\varepsilon_{ij}} \qquad 1.3$$

where $Z_{i,j}$ represents the charge state of each group, r_{ij} the distance between them, and ε_{ij} the effective dielectric constant, the interpretation of which is dependent on the electrostatic model used[7]. In most structure-based methods, Coulomb interactions play essential roles in determining the magnitude of electrostatic energies in proteins; in early models where ionizable moieties were treated as point charges, they represented the sole means by which these energies were calculated[38].

Favorable Coulomb interactions shift the pK_a values of ionizable residues in the direction that promotes the charged state. However, when two residues are titrating in similar pH ranges, Coulomb interactions can also be manifested as changes in the cooperativity of proton binding, with favorable interactions promoting positive cooperativity and unfavorable ones promoting negative cooperativity[39,40]. Coulomb interactions have been invoked to explain a myriad of biological processes; short and medium-range Coulomb interactions can play important roles in tuning pK_a values of ionizable residues in functional sites[41], while long-range Coulomb interactions have been invoked to explain effects such as enzyme catalysis exceeding diffusion-limited rates[42]. However, the importance of Coulomb interactions is increasingly controversial. Experimental evidence suggests that in aqueous conditions and physiological salt concentrations (100-200 mM monovalent salts) even short-range Coulomb interactions on the protein surface are weak[43,44,45], much weaker than anticipated from theoretical considerations. Structure-based calculations with continuum methods that treat the protein with the experimentally determined dielectric constants of 2 to 4 exaggerate the magnitude of Coulomb effects. Experimental data

continue to accrue demonstrating that Coulomb interactions in proteins are as weak as they would be in water. This is not to say that cumulatively, in a protein with many charges, the net effect of Coulomb interactions is always weak. In general, the deconvolution of ΔG_{ij} effects from other energetic contributions (ΔG_{ii} and ΔG_R) can be challenging, as described in Chapter 5 of this dissertation.

1.2.3 Self-energies are sensitive to the polarity and polarizability of an environment

A less appreciated but highly significant determinant of electrostatic energies in proteins is the self-energy. Quantitative descriptions of the self-energy were developed by Max Born to describe the electrostatic energies involved in ion solvation, described as the energy necessary to move infinitesimal partial charges into a sphere the size of the cavity occupied by the ion in a solvent[46]. For this reason, this term in the context of protein electrostatics is also referred to as the Born energy, and differences of self-energies (i.e., in water and in a vacuum) are hydration or solvation energies.

While Coulomb interactions refer to the interaction of a charge with the electric field of another charge (ij), the self-energy describes the energy stored in an ion, expressed in the form of the field around it (ii). In its simplest representation, the Born energy of transferring a charge from water to the protein interior can be represented as:

$$\Delta G_{Born\ (ii)} = \frac{z_i^2}{2r_{ion}\varepsilon_{prot}} - \frac{z_i^2}{2r_{ion}\varepsilon_{H2O}} \qquad 1.4$$

where Z_i and r_{ion} represent the charge and radius of the ion, respectively, and ε_{prot} and ε_{H2O} the dielectric constants of protein and water, respectively. The self-energy of a charge is highly sensitive to the properties of its milieu; in nonpolar, low dielectric environments, these energies are very large but in water it is quite low. For this reason, the self-energy acts as an important

organizing force in biology, responsible for the formation of chemical gradients between membranes and the general architecture of many biomacromolecules that places charges always in contact with water, precluding structures in which they are buried in dry environments. It is important to note that the formalism described above is highly simplified and assumes an infinite radius of the protein, which ignores the influence of interfacial hydration or the reaction field of bulk solvent. Multiple approaches have been developed to account for the depth dependence of burial, as described in Chapter 1.3.1. More rigorous microscopic treatments of the self-energy have been used to describe the solvation of charges in terms of permanent and induced dipoles, as well as internal water molecules and the reaction field from bulk solvent[7].

1.2.4 Molecular determinants of pK_a values in proteins

The pK_a values of ionizable residues are highly sensitive to their microenvironments (Fig. 1.3)[41]. Any molecular factor that stabilizes the charged state will shift pK_a values in the direction that favors the charged state. For example, polar environments and those with complementary charges can promote the charged state, while dry, hydrophobic and other environments incompatible with charge will favor the neutral state. When a protein is unfolded, the ionizable residues are fully hydrated and are generally devoid of intramolecular, stabilizing interactions; the pK_a values of the ionizable residues should be close to their normal values, subject only to weak Coulomb effects[47,48]. The process of folding introduces these residues into microenvironments with specific chemical properties. The interactions between the ionizable moieties with permanent and induced dipoles, water molecules, other charges (Coulomb interactions), and the reaction field of bulk water shift the pK_a values away from their normal values in water [7]. This process can be

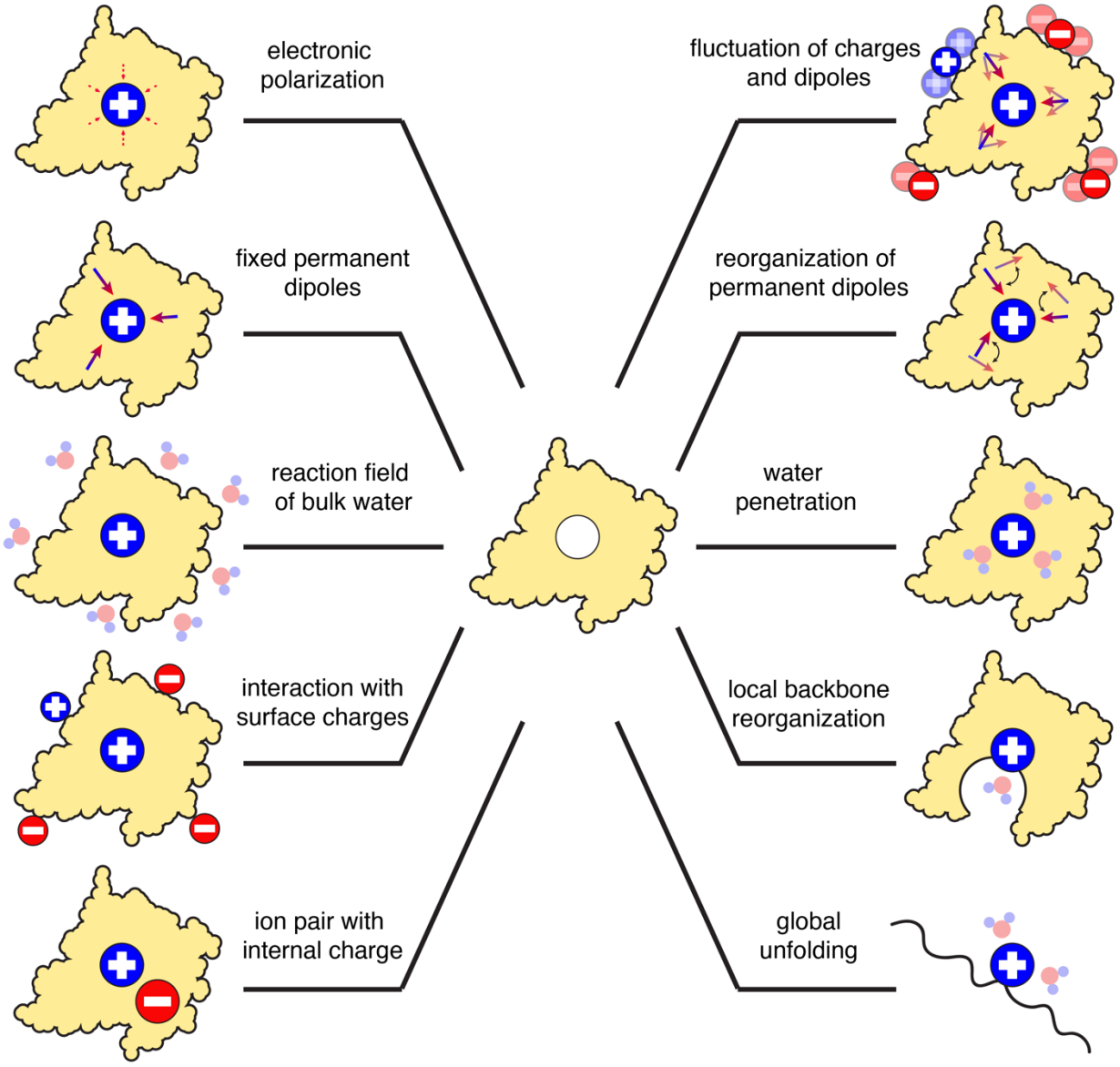

Figure 1.3. Factors that can affect the pK_a values of buried ionizable residues. Static (left) and dynamic (right) contributions to ΔG_{ii}, ΔG_{ij}, and ΔG_R in modulating electrostatic energies in the protein interior.

described in a thermodynamic cycle, which accounts for the transfer of an ion from bulk solvent (i.e., the unfolded state) to a site in the folded protein, where interactions with water are lost (Fig. 1.2).

It is important to emphasize differences between the creation of charge on the protein surface, where charges have been shown experimentally to behave as if they were in bulk water [43,49,44,45], and the creation of a charge buried in the protein interior, devoid of bulk water. It has been argued that capturing the energetics of ion burial is necessary for proper calibration of electrostatic models, as the protein interior contains a heterogenous environment where the contributions of the polarity and polarizability of the protein must be accurately calculated[7,50]. As mentioned above, the pK_a values of the ionizable side chain moieties of the 7 ionizable amino acids, and several phosphoamino acids, have been determined in water[51], providing a reference for interpreting the magnitude of the pK_a shifts upon folding. The shifted pK_a values of ionizable residues have valuable information about ΔG_{ii} and ΔG_{ij}.

1.2.5 Origins of the protein dielectric constant

The magnitude of both Coulomb and self-energies in proteins depend on the dielectric properties of both water and proteins. In continuum electrostatic models the protein is treated as a material with a single dielectric constant. The dielectric constant acts as a single-valued descriptor of the polarity and polarizability of a homogenous material[46]. Dielectric measurements on dry protein powders have provided dielectric constants of 2 to 4[52]. The low dielectric response of protein powders reflects the rigidity of the protein backbone, which is constrained by a myriad of hydrogen bonding and van der Waals interactions. In this highly constrained structure, charges or dipoles to cannot reorganize in response to an electric field; therefore, protein dielectric constants

of 2 to 4 reflect electronic polarization with contributions from structural fluctuations. The protein dielectric constant measured with parallel plate capacitor experiments is highly dependent on the hydration state of the powders; increasing hydration leads to increased dielectric constant, with a discontinuity in the dielectric properties of proteins at approximately 50 water molecules per protein[53]. This discontinuity is observed because hydrated proteins are able to undergo dynamical and conformational rearrangement of dipoles and charges that dehydrated powders cannot. This observation highlights the important contributions of the reorganization energy to electrostatic effects in proteins.

1.2.6 Structural reorganization coupled to H^+-binding or release.

The fundamental property of the pH-sensitivity of proteins described in this dissertation is the linkage between proton binding or release and conformational change. The thermodynamic principles governing pH-driven structural reorganization in proteins are well understood, having been described originally in a completely general form as a linkage relationship by Wyman[54], and for the specific case of proton binding by Charles Tanford in 1961[55]. Conformational reorganization in response to changes in pH requires a minimum of two different conformational states with different proton binding affinities. In other words, at least one ionizable amino acid must have a different pK_a value in the two (or more) states. The linkage between pH changes and structural reorganization are described in the thermodynamic cycle in Fig. 1.4. The equilibrium constants in the four sides of this cycle are the pK_a values of the ionizable residue in each of the conformational states, and the equilibrium constants between conformational states when the ionizable residue are charged or neutral. The force behind the conformational change is the pH-dependence of the equilibrium between states; in cases where the difference in pK_a values is

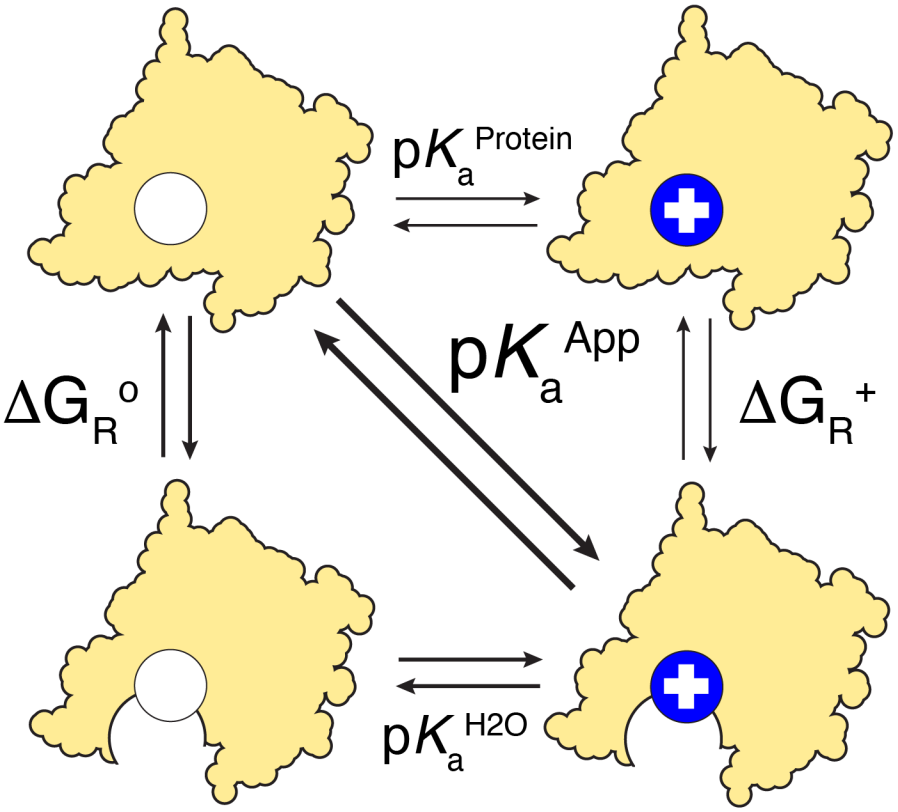

Figure 1.4. Role of pH-driven conformational reorganization in apparent pK_a values. The protein can sample the fully folded (closed) conformation, where the ionizable atom is fully buried and has a depressed pK_a value (pK_a ^{Protein}). It can also sample a partially unfolded (open) conformation where the ionizable atom is in water and titrates with a normal pK_a (pK_a^{H2O}). The reactions on the vertical sides of the cycle describe the energetics of reorganization of the protein when the ionizable group is charged (right) or neutral (left). The apparent pK_a of the buried residue is governed by the energy gap between the open and the closed conformations and independent of the effects of the local microenvironments of the ionizable residue in the two conformations.

significantly greater than the difference in the free energy between conformational states, then the free energy between states governs the apparent pK_a value (pK_a^{App}) of the conformational transition[56].

The ability of proteins to sample alternative states has significant implications for the interpretation of electrostatic energies in proteins, most of which have been missed by the field. For example, if ionization of a residue that is buried and dehydrated drives the protein to a conformation where it is charged and hydrated, then the pK_a of that residue will be governed by the energy gap between states and not the detailed features of the microenvironment around the group when it is buried. In effect, all of the terms that are outlined in Fig. 1.3, which are the terms that are computed in structure-based energy calculations, are irrelevant when conformational reorganization is involved. Several computational models have been developed recently based on the theory originally developed by Tanford to model pH-driven conformational changes, either using Monte-Carlo based methods[57], molecular dynamics simulations[9,56], or ensemble-based partial unfolding models[58]. What has been missing until now is detailed experimental description of the character of the conformational reorganization coupled to change in the charge state of a protein.

1.3 pK_a calculations in proteins

Given the essential role of electrostatics in almost all biochemical processes, there has been significant interest in the development of algorithms that can calculate electrostatic energies from atomic-resolution structures of proteins[6]. The ability to correlate structure with energies using electrostatics provides a means to obtain a quantitative description of the forces that drive biological function (e.g. energy transduction and catalysis)[7]. Additionally, predicting the pH-

dependent properties of proteins is of interest both for understanding how pH regulates biological systems[11] and to guide the design of proteins that can interface with biological systems by functioning as pH-switches[59]. Accurate calculation of electrostatic energies in proteins requires the ability to reproduce experimentally determined pK_a values and reduction potentials; however, despite almost a century of research on protein electrostatics, accurate predictions of pK_a values remains a daunting challenge for reasons explained by the work described in this dissertation[60].

1.3.1 Static continuum methods for the calculation of electrostatic energies in proteins

The first attempts to simulate ionization equilibria in proteins occurred well before the advent of the first protein structures. The groundwork for these studies was laid by Peter Debye and Erich Hückel in 1923, using a generalized form of the Poisson-Boltzmann equation developed a decade earlier. The Debye-Hückel theory of monovalent ions in solutions provided a means to calculate the electrostatic potential in a dielectric cavity surrounded by bulk solvent and ions; by simple extension this could be applied to study ionization equilibria and salt effects in proteins. As early as 1924, Kaj Linderstrøm-Lang recognized that the net charge of a protein would influence the titrations of individual ionizable groups on the protein surface. He developed the first model of protein electrostatics to try to predict acid/base effects in proteins. These first models, which assumed a "smeared" distribution of charge on the protein surface, were clearly unrealistic but the physical chemistry implicit in the model was impeccable. He was decades ahead of his time. Further advances were made by Charles Tanford and John Kirkwood, who in 1957 developed the first continuum model that treated charges as discrete charges[38]. The method assumed that the ionizable residues could be treated as point charges in a dielectric cavity surrounded by water. The spherical approximations used in these models were not expected to significantly affect the ability

of these models to calculate energies from structures, as the energies for each residue were more sensitive to their distances between other charges or within the protein interior than to the actual shape of the cavity. These early models ignored contributions from the self- energy as it was mathematically incompatible with the model-dependent solution of the linearized Poisson Boltzmann equation achieved by Kirkwood. Remarkably, these models were developed a year before the first structure of a protein was solved. In the absence of structure and of computers, they could only be applied to obtain general insight using idealized distribution of charges in fixed geometries.

Despite these advances, it was quickly recognized that that these methods could not accurately calculate electrostatic energies in proteins. By the 1970's, an increase in computing power and the availability of crystal structures allowed for structure-based energy calculations of protein crystal structures[61,62]. Since the pK_a values of the ionizable residues are dependent on both pH and the pK_a values of neighboring residues (and vice versa), iterative methods were developed for pK_a calculations. Early on it was recognized that the calculations were highly sensitive to the depth of burial (a bulk parameter used for all charges in the structure). Attempts to reproduce experimental pK_a values and proton binding curves of proteins with these early continuum models led to the conclusion that ionizable residues in proteins behaved as if they were in bulk water. The next generation of calculations was focused on implementing all manner of ad hoc approximations with the goal of diminishing the magnitude of Coulomb effects. Slowly the field came to recognize that charge-charge interactions alone cannot account for the pK_a shifts observed in folded proteins. At that time it was impossible to incorporate the self-energies of charges. Furthermore, the calculations could not account for conformational flexibility or deviations from the crystal structure. One of the ways in which improvement between calculated and measured

thermodynamic data was achieved was with use of a static solvent accessibility parameter to account for burial (SA-TK)[63]. These semi-empirical methods improved the accuracy of the calculations but the absence of a self-energy term precluded their application to a variety of problems in structure-function relationship related to energy transduction.

The availability of fast computers has allowed for numerical solution of the Poisson-Boltzmann equation, in principle an improvement over the analytical model-dependent solutions[64]. These methods allowed for more realistic treatments of the protein surface, going beyond the simple spherical assumptions and applying more realistic boundary conditions[65,66]. One of the most successful examples of these methods is the finite-difference Poisson-Boltzmann (FDPB) methods[42]. In addition to a more realistic treatment of the protein surface and boundaries between the protein surface and solvent, these models allowed for the incorporation of the Debye-Huckel model to obtain more accurate prediction of salt effects. Importantly, the implementation of the FDPB allowed realistic treatment of solvation parameters to try to effectively capture self-energy terms[67,68].

1.3.2 Limitations of the dielectric "constant": Microscopic models

Despite the early promise of numerical methods applied to continuum models for the calculation of electrostatic energies, it soon became clear that these models could not accurately reproduce the pK_a shifts of ionizable residues in proteins either. The ad hoc solution to the problem involved the use of artificially high dielectric constants[69]. A dielectric constant is a measurable parameter with well-defined physical meaning (the permittivity of a material); in these calculations the dielectric constant effectively becomes a weighting parameter that has lost all physical meaning[7,70]. To address these issues, parallel efforts have tried to incorporate more realistic

treatments of the protein interior. Microscopic approaches such as the protein-dipoles Langevin-dipoles[71,72] (PDLD) and free-energy perturbation (FEP) methods[73,74], which explicitly model structural features governing the solvation of charges (Fig. 1.2 and 1.3), provided more physically relevant descriptions of the heterogenous protein interior that were overlooked in continuum models, and were useful in evaluating electrostatic models in proteins[7]. Significant disadvantages of these methods include the computational cost, lack of convergence, and the difficult treatment of boundary conditions in the accurate calculation of long-range electrostatic effects. To address these issues, semi-microscopic methods that incorporate features of continuum methods to alleviate computational costs and boundary issues have been developed[72]. Even so, limitations of these methods have been demonstrated in their inability to capture large scale conformational changes coupled to change in charge state. Physically unrealistic approaches such as overcharging the ionized residue to force reorganization have been used to address these issues[8].

1.3.3 Incorporation of structural flexibility in structure-based pK_a calculations

As the importance of conformational flexibility as a determinant for electrostatic energies became more apparent, methods were developed to overcome limitations of static continuum models. One of the earliest of these methods is the Monte-Carlo continuum electrostatics (MCCE) method[57]. As the name suggests, MCCE makes use of FDPB continuum methods to calculate electrostatic energies and it accounts for conformational flexibility by incorporating a molecular mechanics forcefield with Monte-Carlo sampling. The backbone of the protein is kept fixed, as backbone relaxation would be too computationally expensive, but the side chains are allowed to sample multiple positions. A Boltzmann-weighted scheme that uses calculated electrostatic and non-electrostatic interaction energies is used to estimate pK_a values. The MCCE method improved

pK_a calculations without resorting to the use of artificially high dielectric constants[75]. Despite these improvements, these algorithms fail to accurately predict the pK_a values of buried residues, which turn out to be the ones that are of functional relevance[76].

Another approach to pK_a calculations with explicit treatment of structural reorganization involves use of molecular dynamics (MD). In general, modeling ionization equilibria using MD is challenging, as it is not possible to create or break bonds during a simulation without causing a fatal discontinuity in the energy function. However, a number of constant-pH MD approximations (CpHMD) have been developed to overcome these obstacles[77]. Two main approaches, discrete Monte-Carlo (MC) and continuous λ-dynamics based methods, have been developed to allow for sampling of alternate protonation states throughout the simulation. In the MC-based approaches, the protonation state of a residue undergoes an MC step to a different protonation state throughout set intervals of an MD simulation. The energy is evaluated at each of these steps, at which point the MC protonation/deprotonation step is accepted or rejected based on the calculated protonation energies. A significant drawback of these methods is the time constraints due to the intensive nature of the calculation of protonation energies at each MC time point, either with continuum methods (Generalized Born[78] or PB[79,80]) or thermodynamic integration simulations[81], both of which are computationally expensive. Issues with discontinuity in the MD simulations during the MC steps can also arise. The λ-dynamics approach differs in that the protonation state is treated through a continuous parameter, λ, which represents a superposition of the protonated and deprotonated states where $\lambda = 1$ and $\lambda = 0$ correspond to specific protonation states[82]. In these methods, the λ parameter is sensitive to the potentials of the titrating residues and pH of the system, allowing it to be tuned in the simulation through pH changes. Because the λ parameter is

continuous, barrier functions are necessary to prevent unrealistic mixed populations of charge at the end points of titrations[83].

CpHMD methods have continued to improve; several approaches have been modified to include various amounts of microscopic detail. Additionally, methods such as pH-replica exchange MD (pH-REMD) allow for extensive sampling of the conformational space, especially of backbone relaxation in pH-driven processes[84]. These advancements represent a significant step forward in the computational modeling of pH-effects in proteins; however, a lack of well-defined benchmarks for these methods has prevented a rigorous analysis of their accuracy. Many of these algorithms have been used to attempt to reproduce the experimentally determined pK_a values of basic and acidic residues buried in the core of the model protein staphylococcal nuclease[9,85,86]. Although some of these methods show significant improvement in their predictive abilities relative to crude continuum methods, in general the nature of the MD-predicted structural changes coupled to ionization of these buried residues are not comparable to what is observed experimentally. The results outlined in this dissertation describe the physical properties of proteins that the computational methods have to reproduce, and they constitute the benchmarks that they will have to be tested against.

1.4 Experimental studies of the dielectric response of proteins to the ionization of buried groups

1.4.1 pK_a values of buried ionizable residues in SNase

Buried ionizable residues are essential for fundamental biochemical processes[87]. Structure-based analysis of buried groups remains unsatisfying; structure-based pK_a calculations cannot reproduce measured values or rationalize the dielectric relaxation processes that define them. Most

calculations fail to reproduce the experimentally determined pK_a values of buried residues because they cannot reproduce the balance of self-energies and Coulomb energies. The problem is that the protein relaxation processes reflected in self-energies and Coulomb energies are very different and they cannot possibly be captured with a single-valued dielectric constant. Furthermore, the heterogenous nature of the protein interior cannot be treated explicitly in continuum methods[7]. In microscopic methods based on MD simulations the key problem is that these methods cannot predict accurately the alternative conformations of proteins, nor can they predict the free energy differences between them.

Previous studies in our laboratory have examined experimentally the dielectric response to ionization of buried residues. These studies have made use of a set of 125 variants of the model protein staphylococcal nuclease (SNase), where 25 internal positions in the protein were substituted for either Lys[88], Glu[89], Asp[90], Arg[91], or His[92]. The pK_a values of these residues were measured by comparing the pH-dependence of the thermodynamic stabilities of the background protein from those of variants with buried ionizable residues. The difference in stabilities ($\Delta\Delta G^o_{H2O}$) were fit to a linkage equation based on the formalism described in Fig. 1.4, where the two states being considered were the folded and unfolded states. Most of them (excluding those of Arg residues) are shifted to favor the neutral state, as expected for ionizable moieties buried in the hydrophobic protein interior (Fig. 1.5). Crystal structures of dozens of these variants have been solved, demonstrating that the protein can tolerate burial of ionizable residues without detectable backbone reorganization[93] at pH values where these residues should be neutral. With some notable exceptions, ionization of buried Lys, Asp and Glu did not trigger structural changes detectable in pH-titrations monitored by circular dichroism or fluorescence spectroscopy[89,88]. This suggests that

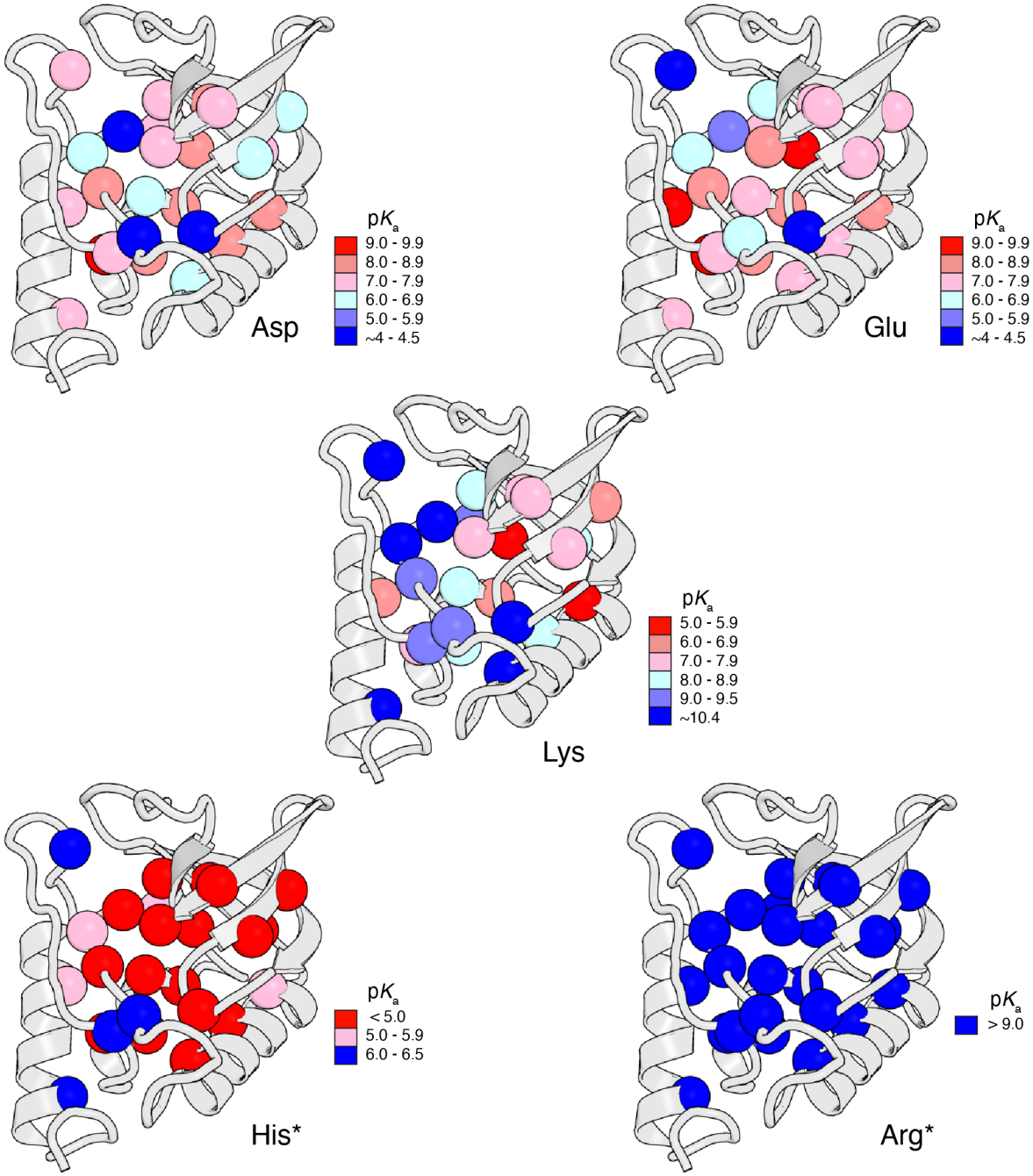

Figure 1.5. Ionizable residues in the hydrophobic interior of staphylococcal nuclease. The pK_a values of Asp, Glu, and Lys in 25 different positions were measured by analysis of the pH-dependence of the thermodynamic stability of the protein[88-92]. The pK_a values of Arg and His at these internal positions were estimated based on the slopes of the thermodynamic stabilities measured at 5 and 7 for His, and pH 7 and 10 for Arg.

any reorganization coupled to ionization of buried residues is local and does not involve significant disruption of secondary structure.

1.4.2 NMR spectroscopy studies of pH-driven reorganization coupled to ionization of buried residues

To investigate the nature of the dielectric response with higher resolution, NMR spectroscopy was used to characterize the structural responses to ionization of buried Lys residues. Unassigned backbone ^{1}H-^{15}N HSQC spectra collected at pH values above and below the pK_a values of the Lys residues identified 8 cases where ionization produced strong spectroscopic evidence of structural changes[93]. The variants with K66, D66, and E66 were further characterized using chemical shift perturbations, qualitative broadening and hydrogen deuterium exchange studies, providing strong evidence that ionization of residues at position 66 in SNase leads to increased fluctuations of the C-terminus of helix α-1[94]. To examine the nature and timescales of the pH-dependent conformational changes in the Lys variants of SNase, ^{15}N relaxation-based NMR-spectroscopy[95] was used to characterize the pH-dependent dynamics of variants with K25 and K125, both with similar pK_a values (6.3 and 6.2, respectively). In the variant with K125, relaxation dispersion and resonance broadening provided evidence of pH-independent dynamics on μs-ms timescales. In the variant with K25, a hierarchy of pH-dependent dynamics was detected using relaxation dispersion and ZZ-exchange spectroscopy, the timescales of the conformational exchange between folded and subglobally unfolded states in some cases extending to the 100s of ms. Although these initial NMR spectroscopy studies were focused on structural reorganization coupled to the ionization of buried residues, it should be noted that structural changes can occur in response to the protonation of surface residues as well. NMR chemical shift titrations[39], ^{15}N

relaxation and quantitative hydrogen-deuterium exchange studies[96] have provided evidence of subtle structural and dynamical changes in response to the protonation of surface Asp, Glu, and His residues in SNase and posited on theoretical grounds[58].

1.4.3 X-ray crystallography studies of pH-driven reorganization coupled to ionization of buried residues

The initial NMR spectroscopy studies were limited in their ability to characterize the nature of the structural changes coupled to ionization of internal Lys residues. X-ray crystallography studies were used to provide complementary atomic resolution structural information necessary to describe the nature of the pH-dependent structural changes coupled to the ionization of buried residues. The first example of backbone reorganization coupled to charge burial was in the crystal structure of the variant with Arg-109; as Arg-109 is ionized below pH 10, reorganization of a loop allowed for hydration of the buried residue[91]. Further crystallographic studies of variants with Glu-23[97] and Lys-66[98] provided direct evidence of pH-dependent reorganization (Fig. 1.6). In crystal structures of the single variants at pH values where the buried residues were neutral, the backbone of the protein shows no evidence of reorganization to accommodate the buried residues. At pH values where the residues become ionized, crystal structures show clear evidence of partial unfolding to expose the buried residues to bulk water. Consistent with this notion, the apparent pK_a value of Lys-66 was sensitive to Gly substitutions that promoted partial unfolding of the helix the buried Lys residue resides in. These results provided evidence that backbone reorganization was an important factor underlying the apparent pK_a values determined by linkage thermodynamics. It is important to note that the static structures captured by crystallization are not necessarily representative of the full range of conformations available in solution; indeed, NMR

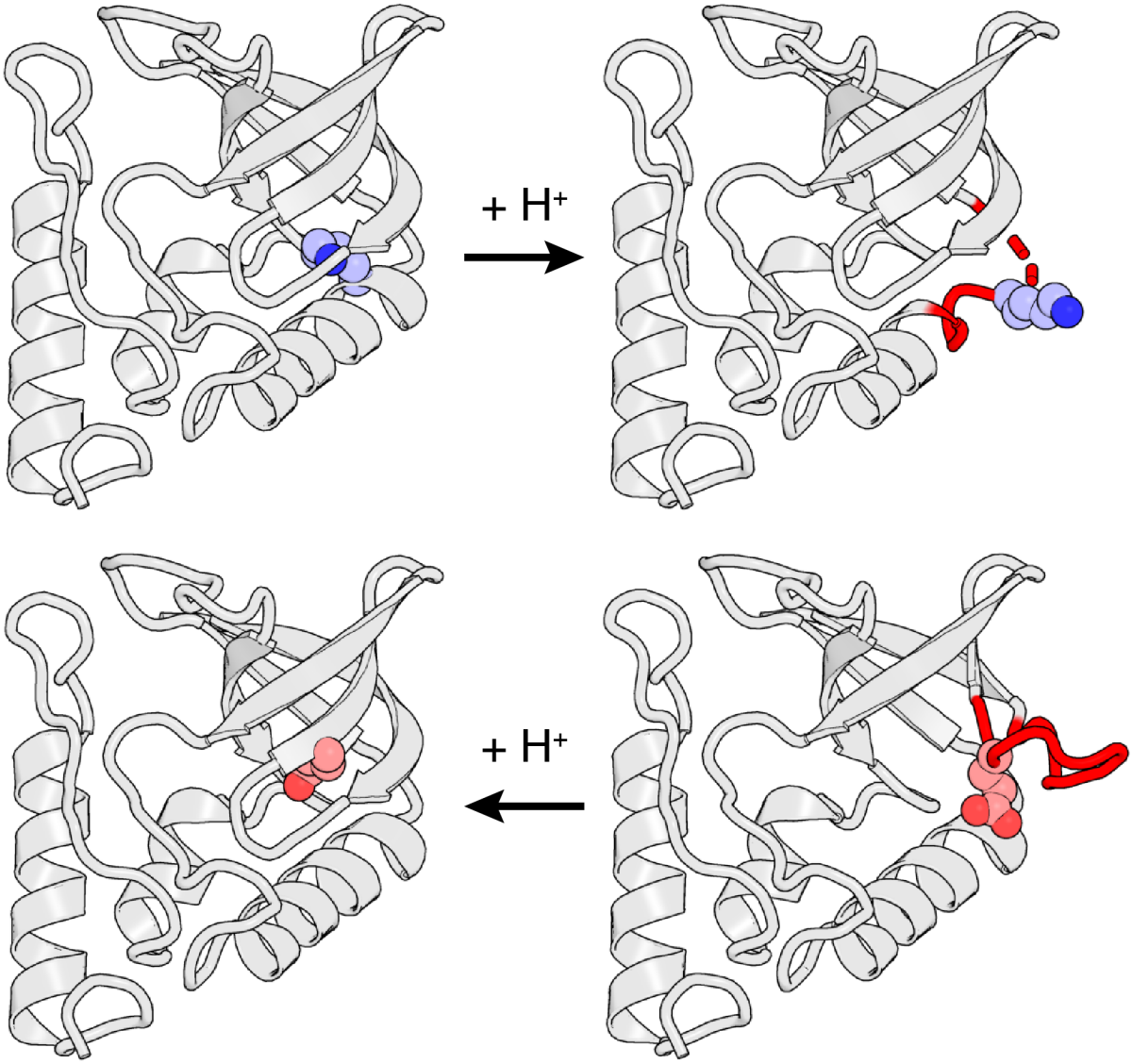

Figure 1.6. Crystal structures showing structural reorganization coupled to the ionization of buried groups. On the left, crystal structures of SNase variants with Lys-66 (top, PDB ID: 3HZX) and Glu-23 (bottom, PDB ID: 3QOL) showing that the residues are fully buried when they are neutral. Structures showing that upon ionization (right), of Lys-66 (top, PDB ID: 5CV5) and Glu-23 (bottom, PDB ID: 3TME) the protein becomes partially unfolded. In this partially unfolded state, the charged moieties are hydrated.

spectroscopy characterization of these variants in solution has shown evidence of resonance broadening[97-98], consistent with conformational dynamics on μs-ms timescales.

1.4.4 Coulomb interactions inside a protein

Buried ionizable residues are necessary for a wide range of biochemical processes. Often they are found in networks with other ionizable or polar residues[99,100,101]. The mechanism that allows proteins to tolerate ion pairs is not well understood. Theoretical studies on the energetics of ion pair burial have been inconclusive. In simple continuum models the highly favorable Coulomb interactions between two buried groups cannot compensate for the unfavorable dehydration of two charges[102]. To investigate the balance between Coulomb and desolvation energies in the protein interior, the structural and thermodynamic consequences of buried ion pairs was investigated in a variant of SNase with an artificial buried ion pair (E23/K36)[103]. These studies demonstrated that the success of ion pair burial in a protein is highly dependent on the microenvironment of the ionizable moieties, and that the ability of the protein to reorganize permanent dipoles and internal water molecules to stabilize the charged pair. Consistent with computational predictions, the protein could not reorganize to accommodate the reversed ion pair (Lys:Glu versus Glu:Lys), as the arrangement of dipoles that stabilized one ion pair destabilized the reverse[104]; the variant with the reversed pair (K23/E36) was marginally stable and sub-globally unfolded.

1.4.5 Characterization of Coulomb interactions between buried and surface charges

Previous studies on thioredoxin variants demonstrated that the pK_a value of a buried Lys was sensitive to the distribution of charges on the protein surface[105]. These results were of interest

as they suggested a means by which to tune the functional properties of active site residues through alterations to the protein surface. These studies were interpreted in terms of Coulomb interactions between the positively charges buried Lys residue and surface Lys or Glu residues. To investigate the role of Coulomb interactions between charges in the protein interior as a determinant of the anomalous pK_a values of buried residues, interactions between buried Lys residues and surface charges were investigated previously in our laboratory[106]. These studies proved inconclusive, as in some cases very strong interactions were observed between buried Lys and surface residues, but in other cases these interactions appeared to be weak. Additionally, it was unclear if coupling was due to direct Coulomb interactions or effects from pH-driven structural reorganization.

1.5 Organization and contributions of this dissertation

The work described in this dissertation examines determinants of the anomalous pK_a values of buried residues in proteins. The goal was to dissect contributions from dehydration, Coulomb interactions, and reorganization systematically using equilibrium thermodynamics and multiple atomic-resolution structural methods. These studies provide a comprehensive view into the determinants of electrostatic energies in proteins, by probing the properties of buried Lys residues (e.g., dynamics, state of hydration), the protein's backbone response to their ionization, and the influence of short and long range electrostatic on their pK_a values. The results described in this dissertation demonstrate conclusively that the apparent pK_a values of buried residues are governed by backbone reorganization coupled to change in charge of the protein. The most important contribution made by this work is the demonstration that the affinity and cooperativity of H^+-binding in protein, that electrostatic effects in general, are governed by the pH-driven redistribution of structural states in the conformational ensemble.

In chapter 2, I describe the use of ^{15}N-detected NMR spectroscopy experiments to study the dynamics, hydration, and protonation state of buried Lys residues in SNase. Studies of five internal Lys variants of SNase suggested that hydration and microenvironment alone were unlikely to account for the magnitude and differences of the pK_a shifts of Lys residues buried in different parts of the protein. In most cases the ionization of buried Lys residues drove backbone reorganization, and the differences between the properties of Lys residues buried in different sites reflect the intrinsic propensity of different regions of the protein to reorganize. These conclusions were only possible due to the ability of ^{15}N-detected NMR experiments to directly observe the ^{15}N resonance of the ionizable Nζ atom of buried Lys, which is generally very difficult to observe with ^{1}H detect experiments. Since the publication of this work[107], the utility of ^{15}N-detection of Lys residues has been extended by other groups to solid-state NMR, to study the role of Lys deprotonation in amyloid:prion liquid to hydrogel transitions[108].

In chapter 3, I describe a systematic, comprehensive NMR spectroscopy survey of the structural changes coupled to the ionization of Lys residues buried in 25 internal locations in SNase. Evidence of backbone structural reorganization was observed in every single protein with an internal Lys residue. This was detected by analysis of chemical shift perturbation between pH values where the Lys residues were charged or neutral, or between the background protein and the variant with charged Lys. In many cases the structural analysis of chemical shift data was consistent with partial unfolding. Correlation of structural detail from NMR spectroscopy with thermodynamic stability measured with equilibrium thermodynamic methods provided a means to map the conformational landscape of SNase, with direct measurements of the free energy gaps between the plethora of partially unfolded states that normally make up vanishingly small

populations of the native state ensemble of this protein. These results demonstrate that the ionization of a single group is sufficient to drive the protein into partially unfolded states.

In chapter 4, I describe studies of backbone reorganization in response to the burial of ion pairs in SNase. Lys:Glu and Glu:Lys pairs were introduced in the i,i+4 positions of an α-helix, where they would be poised to interact. In contrast to previous studies of ion pair burial in SNase, both variants were globally folded, and the residues were charged in both ion pair arrangements. Crystal structures and NMR spectroscopy demonstrated that both ion pair variants underwent some amount of backbone or sidechain reorganization. The region of the backbone that reorganizes is the same as was observed in the other ion pair variants; however, the location and magnitude of the structural response to the burial of ion pairs varied dramatically between the four variants. These data demonstrate, again, that even in cases expected to be dominated by favorable Coulomb interactions, the behavior observed in solution is governed by the propensity of the protein to reorganize into a state that minimizes the energetic penalty associated with dehydration of charges.

In chapter 5, I describe studies of coupling between buried Lys residues and surface residues. Chemical shift perturbations of surface acidic groups were measured between the background protein and 22 internal Lys variants at pH values where Lys was ionized and the protein reorganized. Two variants were chosen for further study with NMR-detected pH-titrations of Asp, Glu, and His residues. Strong couplings were observed between surface acidic groups and both of the buried Lys residues. Mutational charge neutralization studies and thermodynamic analysis of the ensuing pK_a shifts were performed to determine that the origin of the coupling between buried and surface residues is not due to Coulomb interactions but to coupling through conformational reorganization. These data provide direct experimental evidence that the reorganization energy can act as the dominant determinant of not only the pK_a values of buried

residues, but also surface residues which they are coupled to through structural reorganization. The results of these studies have important implications for understanding both allosteric and cooperative ligand binding effects in terms of pH redistribution of the conformational ensemble.

1.6 References

1. Garcia-Moreno, B. Adaptations of proteins to cellular and subcellular pH. *J. Biol.* **8**, 98 (2009).

2. Putney, L. K. & Barber, D. L. Na-H exchange-dependent increase in intracellular pH times G2/M Entry and transition. *J. Biol. Chem.* **278**, 44645–44649 (2003).

3. Munder, M. C. *et al.* A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* **5**, e09347 (2016).

4. Squirrell, J. M., Lane, M. & Bavister, B. D. Altering intracellular pH disrupts development and cellular organization in preimplantation hamster embryos. *Biol. Reprod.* **64**, 1845–1854 (2001).

5. Casey, J. R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **11**, 50–61 (2010).

6. Gunner, M. R. & Baker, N. A. Chapter One - Continuum electrostatics approaches to calculating pK_as and E_ms in Proteins. *Method. Enzymol.* **578,** 1–20 (2016).

7. Schutz, C. N. & Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **44**, 400–417 (2001).

8. Kato, M. & Warshel, A. Using a charging coordinate in studies of ionization induced partial unfolding. *J. Phys. Chem. B* **110**, 11566–11570 (2006).

9. Shi, C., Wallace, J. A. & Shen, J. K. Thermodynamic coupling of protonation and conformational equilibria in proteins: Theory and simulation. *Biophys. J.* **102**, 1590–1597 (2012).

10. White, K. A., Grillo-Hill, B. K. & Barber, D. L. Cancer cell behaviors mediated by dysregulated pH dynamics at a glance. *J. Cell Sci.* **130**, 663–669 (2017).

11. Srivastava, J., Barber, D. L. & Jacobson, M. P. Intracellular pH sensors: Design principles and functional significance. *Physiology* **22**, 30–39 (2007).

12. Schönichen, A., Webb, B. A., Jacobson, M. P. & Barber, D. L. Considering protonation as a posttranslational modification regulating protein structure and function. *Annu. Rev. Biophys.* **42**, 289–314 (2013).

13. Ulmschneider, B. *et al.* Increased intracellular pH is necessary for adult epithelial and embryonic stem cell differentiation. *J. Cell Biol.* **215**, 345–355 (2016).

14. Triandafillou, C. G., Katanski, C. D., Dinner, A. R. & Drummond, D. A. Transient intracellular acidification regulates the core transcriptional heat shock response. *biorXiv,* doi:10.1101/414706 (2018).

15. Isom, D. G. *et al.* Coordinated regulation of intracellular pH by two glucose-sensing pathways in yeast. *J. Biol. Chem.* **293**, 2318–2329 (2018).

16. Trivedi, B. & Danforth, W. H. Effect of pH on the kinetics of frog muscle phosphofructokinase. *J. Biol. Chem.* **241**, 4110–4114 (1966).

17. Ui, M. A role of phosphofructokinase in pH-dependent regulation of glycolysis. *Biochim. Biophys. Acta BBA - Gen. Subj.* **124**, 310–322 (1966).

18. Rossing, R. G. & Cain, S. M. A nomogram relating pO_2, pH, temperature, and hemoglobin saturation in the dog. *J. Appl. Physiol.* **21**, 195–201 (1966).

19. Imai, K. & Yonetani, T. pH dependence of the Adair constants of human hemoglobin. Nonuniform contribution of successive oxygen bindings to the alkaline Bohr effect. *J. Biol. Chem.* **250**, 2227–2231 (1975).

20. Chu, A. H., Turner, B. W. & Ackers, G. K. Effects of protons on the oxygenation-linked subunit assembly in human hemoglobin. *Biochemistry* **23**, 604–617 (1984).

21. Frantz, C. *et al.* Cofilin is a pH sensor for actin free barbed end formation: role of phosphoinositide binding. *J. Cell Biol.* **183**, 865–879 (2008).

22. Srivastava, J. *et al.* Structural model and functional significance of pH-dependent talin–actin binding for focal adhesion remodeling. *Proc. Natl. Acad. Sci.* **105**, 14436–14441 (2008).

23. Hu, Y.-B., Dammer, E. B., Ren, R.-J. & Wang, G. The endosomal-lysosomal system: from acidification and cargo sorting to neurodegeneration. *Transl. Neurodegener.* **4**, 18 (2015).

24. Webb, B. A., Chimenti, M., Jacobson, M. P. & Barber, D. L. Dysregulated pH: a perfect storm for cancer progression. *Nat. Rev. Cancer* **11**, 671–677 (2011).

25. Sennoune, S. R. *et al.* Vacuolar H+-ATPase in human breast cancer cells with distinct metastatic potential: distribution and functional activity. *Am. J. Physiol.-Cell Physiol.* **286**, C1443–C1452 (2004).

26. Padan, E. & Rottenberg, H. Respiratory control and the proton electrochemical gradient in mitochondria. *Eur. J. Biochem.* **40**, 431–437 (1973).

27. Balut, C. *et al.* Measurement of cytosolic and mitochondrial pH in living cells during reversible metabolic inhibition. *Kidney Int.* **73**, 226–232 (2008).

28. Vavassori, S. *et al.* A pH-regulated quality control cycle for surveillance of secretory protein assembly. *Mol. Cell* **50**, 783–792 (2013).

29. Inpanathan, S. & Botelho, R. J. The lysosome signaling platform: Adapting with the times. *Front. Cell Dev. Biol.* **7**, (2019).

30. Lescar, J. *et al.* The fusion glycoprotein shell of Semliki forest virus: An icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**, 137–148 (2001).

31. von Ballmoos, C., Wiedenmann, A. & Dimroth, P. Essentials for ATP synthesis by F_1F_o ATP synthases. *Annu. Rev. Biochem.* **78**, 649–672 (2009).

32. Pogoryelov, D. *et al.* Microscopic rotary mechanism of ion translocation in the Fo complex of ATP synthases. *Nat. Chem. Biol.* **6**, 891–899 (2010).

33. Lau, W. C. Y. & Rubinstein, J. L. Subnanometre-resolution structure of the intact *Thermus thermophilus* H^+-driven ATP synthase. *Nature* **481**, 214–218 (2012).

34. Lanyi, J. K. Bacteriorhodopsin. *Annu. Rev. Physiol.* **66**, 665–688 (2004).

35. Kaila, V. R. I., Verkhovsky, M. I., Hummer, G. & Wikström, M. Glutamic acid 242 is a valve in the proton pump of cytochrome c oxidase. *Proc. Natl. Acad. Sci.* **105**, 6255–6259 (2008).

36. Lim, C., Bashford, D. & Karplus, M. Absolute pK_a calculations with continuum dielectric methods. *J. Phys. Chem.* **95**, 5610–5620 (1991).

37. Klamt, A., Eckert, F., Diedenhofen, M. & Beck, M. E. First principles calculations of aqueous pK_a values for organic and inorganic acids using COSMO−RS reveal an inconsistency in the slope of the pK_a scale. *J. Phys. Chem. A* **107**, 9380–9386 (2003).

38. Tanford, C. & Kirkwood, J. G. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* **79**, 5333–5339 (1957).

39. Castañeda, C. A. *et al.* Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. *Proteins* **77**, 570–588 (2009).

40. McIntosh, L. P. *et al.* Dissecting electrostatic interactions in *Bacillus circulans* xylanase through NMR-monitored pH titrations. *J. Biomol. NMR* **51**, 5 (2011).

41. Harris, T. K. & Turner, G. J. Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98 (2002).

42. Klapper, I., Hagstrom, R., Fine, R., Sharp, K. & Honig, B. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins* **1**, 47–59 (1986).

43. Bycroft, M. & Fersht, A. R. Surface electrostatic interactions contribute little to stability of barnase. *J. Mol. Biol.* **220**, 779–788 (1991).

44. Lee, K. K., Fitch, C. A. & García-Moreno E., B. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci.* **11**, 1004–1016 (2002).

45. Tomlinson, J. H., Ullah, S., Hansen, P. E. & Williamson, M. P. Characterization of salt bridges to lysines in the protein G B1 domain. *J. Am. Chem. Soc.* **131**, 4674–4684 (2009).

46. Gabler, R. *Electrical interactions in molecular biophysics: An introduction*. (Elsevier, 2012).

47. Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry* **38**, 4896–4903 (1999).

48. Whitten, S. T. & García-Moreno E., B. pH dependence of stability of staphylococcal nuclease: Evidence of substantial electrostatic interactions in the denatured state. *Biochemistry* **39**, 14292–14304 (2000).

49. Loewenthal, R., Sancho, J., Reinikainen, T. & Fersht, A. R. Long-range surface charge-charge interactions in proteins: Comparison of experimental results with calculations from a theoretical method. *J. Mol. Biol.* **232**, 574–583 (1993).

50. García-Moreno, B. E. *et al.* Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **64**, 211–224 (1997).

51. Platzer, G., Okon, M. & McIntosh, L. P. pH-dependent random coil ^{1}H, ^{13}C, and ^{15}N chemical shifts of the ionizable amino acids: a guide for protein pK_a measurements. *J. Biomol. NMR* **60**, 109–129 (2014).

52. Bone, S. & Pethig, R. Dielectric studies of the binding of water to lysozyme. *J. Mol. Biol.* **157**, 571–575 (1982).

53. Bone, S. & Pethig, R. Dielectric studies of protein hydration and hydration-induced flexibility. *J. Mol. Biol.* **181**, 323–326 (1985).

54. Wyman, J. & Gill, S. J. *Binding and linkage: Functional chemistry of biological macromolecules*. (University Science Books, 1990).

55. Tanford, C. Ionization-linked changes in protein conformation. I. Theory. *J. Am. Chem. Soc.* **83**, 1628–1634 (1961).

56. Di Russo, N. V., Martí, M. A. & Roitberg, A. E. Underlying thermodynamics of pH-dependent allostery. *J. Phys. Chem. B* **118**, 12818–12826 (2014).

57. Alexov, E. G. & Gunner, M. R. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* **72**, 2075–2093 (1997).

58. Whitten, S. T., García-Moreno E., B. & Hilser, V. J. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci.* **102**, 4282–4287 (2005).

59. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).

60. Nielsen, J. E., Gunner, M. R. & García-Moreno E., B. The pK_a Cooperative: A collaborative effort to advance structure-based calculations of pK_a values and electrostatic effects in proteins. *Proteins* **79**, 3249–3259 (2011).

61. Orttung, W. H. Interpretation of the titration curve of oxyhemoglobin. Detailed consideration of Coulomb interactions at low ionic strength. *J. Am. Chem. Soc.* **91**, 162–167 (1969).

62. Tanford, C. & Roxby, R. Interpretation of protein titration curves. Application to lysozyme. *Biochemistry* **11**, 2192–2198 (1972).

63. Shire, S. J., Hanania, G. I. H. & Gurd, F. R. N. Electrostatic effects in myoglobin. Hydrogen ion equilibriums in sperm whale ferrimyoglobin. *Biochemistry* **13**, 2967–2974 (1974).

64. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).

65. Warwicker, J. & Watson, H. C. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.* **157**, 671–679 (1982).

66. Gilson, M. K., Rashin, A., Fine, R. & Honig, B. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol.* **184**, 503–516 (1985).

67. Gilson, M. K. & Honig, B. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins* **4**, 7–18 (1988).

68. Gilson, M. K., Sharp, K. A. & Honig, B. H. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* **9**, 327–335 (1988).

69. Antosiewicz, J., McCammon, J. A. & Gilson, M. K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **238**, 415–436 (1994).

70. Warshel, A., Russell, S. T. & Churg, A. K. Macroscopic models for studies of electrostatic interactions in proteins: limitations and applicability. *Proc. Natl. Acad. Sci.* **81**, 4785–4789 (1984).

71. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).

72. Sham, Y. Y., Chu, Z. T. & Warshel, A. Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and microscopic approaches. *J. Phys. Chem. B* **101**, 4458–4472 (1997).

73. Warshel, A., Sussman, F. & King, G. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* **25**, 8368–8372 (1986).

74. Lee, F. S. & Warshel, A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.* **97**, 3100–3107 (1992).

75. Song, Y., Mao, J. & Gunner, M. R. MCCE2: Improving protein pK_a calculations with extensive side chain rotamer sampling. *J. Comput. Chem.* **30**, 2231–2247 (2009).

76. Gunner, M. R., Zhu, X. & Klein, M. C. MCCE analysis of the pK_as of introduced buried acids and bases in staphylococcal nuclease. *Proteins* **79**, 3306–3319 (2011).

77. Baptista, A. M., Martel, P. J. & Petersen, S. B. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins* **27**, 523–544 (1997).

78. Mongan, J., Case, D. A. & McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **25**, 2038–2048 (2004).

79. Baptista, A. M., Teixeira, V. H. & Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **117**, 4184–4200 (2002).

80. Dlugosz, M. & Antosiewicz, J. M. Constant-pH molecular dynamics simulations: a test case of succinic acid. *Chem. Phys.* **302**, 161–170 (2004).

81. Bürgi, R., Kollman, P. A. & Gunsteren, W. F. van. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins* **47**, 469–480 (2002).

82. Lee, M. S., Salsbury, F. R. & Brooks, C. L. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* **56**, 738–752 (2004).

83. Abrams, J. B., Rosso, L. & Tuckerman, M. E. Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *J. Chem. Phys.* **125**, 074115 (2006).

84. Swails, J. M., York, D. M. & Roitberg, A. E. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation. *J. Chem. Theory Comput.* **10**, 1341–1352 (2014).

85. Goh, G. B., Laricheva, E. N. & Brooks, C. L. Uncovering pH-dependent transient states of proteins with buried ionizable residues. *J. Am. Chem. Soc.* **136**, 8496–8499 (2014).

86. Liu, J., Swails, J., Zhang, J. Z. H., He, X. & Roitberg, A. E. A coupled ionization-conformational equilibrium is required to understand the properties of ionizable residues in the hydrophobic interior of staphylococcal nuclease. *J. Am. Chem. Soc.* **140**, 1639–1648 (2018).

87. Holliday, G. L., Mitchell, J. B. O. & Thornton, J. M. Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.* **390**, 560–577 (2009).

88. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno E., B. Large shifts in pK_a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci.* **108**, 5260–5265 (2011).

89. Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno E., B. Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci.* **107**, 16096–16100 (2010).

90. Cannon, B. R. Thermodynamic consequences of substitutions of internal positions in proteins with polar and ionizable residues. (The Johns Hopkins University, 2008).

91. Harms, M. J., Schlessman, J. L., Sue, G. R. & García-Moreno E., B. Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci.* **108**, 18954–18959 (2011).

92. Sorenson, J. L. Structural and thermodynamic consequences of internal polar and ionizable residues in staphylococcal nuclease. Thesis, The Johns Hopkins University (2016).

93. Chimenti, M. S. *et al.* Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein. *Structure* **20**, 1071–1085 (2012).

94. Chimenti, M. S., Castañeda, C. A., Majumdar, A. & García-Moreno E., B. Structural origins of high apparent dielectric constants experienced by ionizable groups in the hydrophobic core of a protein. *J. Mol. Biol.* **405**, 361–377 (2011).

95. Richman, D. E., Majumdar, A. & García-Moreno E., B. Conformational reorganization coupled to the ionization of internal Lys residues in proteins. *Biochemistry* **54**, 5888–5897 (2015).

96. Richman, D. E., Majumdar, A. & García-Moreno E., B. pH dependence of conformational fluctuations of the protein backbone. *Proteins* **82**, 3132–3143 (2014).

97. Robinson, A. C., Majumdar, A., Schlessman, J. L. & García-Moreno E, B. Charges in hydrophobic environments: A strategy for identifying alternative states in proteins. *Biochemistry* **56**, 212–218 (2017).

98. Peck, M. T. *et al.* Local backbone flexibility as a determinant of the apparent pK_a values of buried ionizable groups in proteins. *Biochemistry* **56**, 5338–5346 (2017).

99. Barlow, D. J. & Thornton, J. M. Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885 (1983).

100. Kim, J., Mao, J. & Gunner, M. R. Are acidic and basic groups in buried proteins predicted to be ionized? *J. Mol. Biol.* **348**, 1283–1298 (2005).

101. Bush, J. & Makhatadze, G. I. Statistical analysis of protein structures suggests that buried ionizable residues in proteins are hydrogen bonded or form salt bridges. *Proteins* **79**, 2027–2032 (2011).

102. Hendsch, Z. S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226 (1994).

103. Robinson, A. C., Castañeda, C. A., Schlessman, J. L. & García-Moreno E, B. Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl. Acad. Sci.* **111**, 11685–11690 (2014).

104. Robinson, A. C., Schlessman, J. L. & García-Moreno E, B. Dielectric properties of a protein probed by reversal of a buried ion pair. *J. Phys. Chem. B* **122**, 2516–2524 (2018).

105. Pey, A. L., Rodriguez-Larrea, D., Gavira, J. A., García-Moreno E, B. & Sanchez-Ruiz, J. M. Modulation of buried ionizable groups in proteins with engineered surface charge. *J. Am. Chem. Soc.* **132**, 1218–1219 (2010).

106. Khangulov, V. S. Molecular determinants of pKa values of internal ionizable groups in proteins: Interactions of internal ionizable groups with surface charges. Thesis, The Johns Hopkins University (2011).

107. Kougentakis, C. M. *et al.* Anomalous properties of Lys residues buried in the hydrophobic interior of a protein revealed with ^{15}N-detect NMR spectroscopy. *J. Phys. Chem. Lett.* **9**, 383–387 (2018).

108. Kostylev, M. A. *et al.* Liquid and hydrogel phases of PrPC linked to conformation shifts and triggered by Alzheimer's amyloid-β oligomers. *Mol. Cell* **72**, 426-443.e12 (2018).

Chapter 2

Anomalous properties of Lys residues buried in the hydrophobic interior of a protein revealed with ^{15}N-detect NMR spectroscopy

Published as Kougentakis, C.M., Grasso, E.M., Robinson, A.C., Caro, J.A., Schlessman, J.L., Majumdar, A., and García-Moreno, E.B. Anomalous properties of Lys residues buried in the hydrophobic interior of a protein revealed with ^{15}N-detect NMR spectroscopy. *J. Phys. Chem. Lett.* 9. 383-387 (2018).

2.1 Abstract

Ionizable residues buried in hydrophobic environments in proteins are essential for many fundamental biochemical processes. These residues titrate with anomalous pK_a values that are challenging to reproduce with structure-based calculations, as their ionization can trigger conformational reorganization on timescales difficult to simulate. A major factor impeding accurate pK_a calculations is that these residues are often difficult to study experimentally. Here, we use ^{15}N NMR spectroscopy to obtain insight into the protonation state, state of hydration, and conformational dynamics of the Nζ amino group of buried Lys residues. The experiments were applied to five variants of staphylococcal nuclease with internal Lys residues that titrate with pK_a values ranging from 6.2 to 8.1. Direct detection of buried Lys residues with these NMR spectroscopy methods will allow for unprecedented studies of the microscopic determinants of their anomalous pK_a values and the structural transitions their ionization drive.

2.2 Introduction

Ionizable residues buried in the hydrophobic core of proteins are rare but they are essential for energy transduction[1], catalysis[2], macromolecular assembly[3] and other important biochemical processes[4-5]. These buried residues are necessary for proton transport and e^- transfer, and for promoting conformational reorganization in response to binding of protons or ligands. Their functional importance is driven by their unusual properties, such as anomalous pK_a values and attenuated hydrogen exchange rates. The pK_a values of buried residues can be shifted relative to their normal pK_a values in water in the direction that favors the neutral state (*i.e.,* elevated for Asp/Glu and depressed for Lys/His)[4-5]. Owing to shifts in pK_a the ionizable moieties are usually neutral when they are buried, reflecting the incompatibility of charges with hydrophobic environments[4-11]. Despite their importance, the physical properties of buried groups and the determinants of their anomalous pK_a values are not well understood. The pK_a values of buried residues cannot be predicted accurately with structure-based calculations. Constant pH MD simulations show promise as they treat structural relaxation coupled to H^+ binding or release explicitly[6-11]; however, accurate reproduction of the experimental pK_a values measured with a set of Lys, Glu or Asp at 25 internal positions in staphylococcal nuclease (SNase), and of the details of conformational reorganization coupled to the ionization of buried groups, remains challenging[4-11]. NMR spectroscopy is uniquely suited to probe the microscopic determinants of the anomalous pK_a values of buried ionizable residues, as it is sensitive to protonation, conformational dynamics, and solvent exchange effects at atomic resolution.

Previous NMR spectroscopy studies have shown that the $^{15}N\zeta$ chemical shifts of deprotonated Lys residues are distinct from those of protonated Lys residues[12]. Lys residues are generally difficult to detect by ^1H detect experiments because the rapid exchange of the $H\zeta$ protons

with those from water[12-15]. Water penetration and backbone relaxation can cause significant self-decoupling between the $^{15}N\zeta$-$^{1}H\zeta$ of internal Lys residues, requiring the use of low temperatures and specialized pulse sequences to detect them[12]. Owing to the unfavorable relaxation properties of relatively rigid buried Lys residues, the experiments used to study surface Lys residues[14-15] are not sufficiently sensitive to detect pH dependent changes of the internal Lys Nζ chemical shifts[12,16]. To overcome the limitations inherent to ^{1}H detection[17-20], ^{15}N 1D experiments were used to identify the Nζ resonance of internal Lys residues and to monitor their titration with pH. This initial study was performed with five variants of staphylococcal nuclease (SNase) with internal Lys residues.

2.3 Results

2.3.1 Crystal structures

The apparent pK_a values of the internal Lys residues in the T62K, V74K, V99K, L25K, and L125K variants, measured by linkage thermodynamics, are 8.1, 7.4, 6.5, 6.3, and 6.2, respectively[5]. Crystal structures have been solved of all five variants, with each internal Lys sampling different microenvironments (Fig. 2.1, Table 1). Introduction of the internal Lys residues does not affect the fold of the protein, as the variants are superimposable with the background protein. Lys-62, Lys-74, and Lys-99 are all within 3.5 Å of polar atoms, with the latter two being within hydrogen bonding distance of interfacial or internal water molecules. Lys-25 and Lys-125 are both in deeply buried, hydrophobic environments[21].

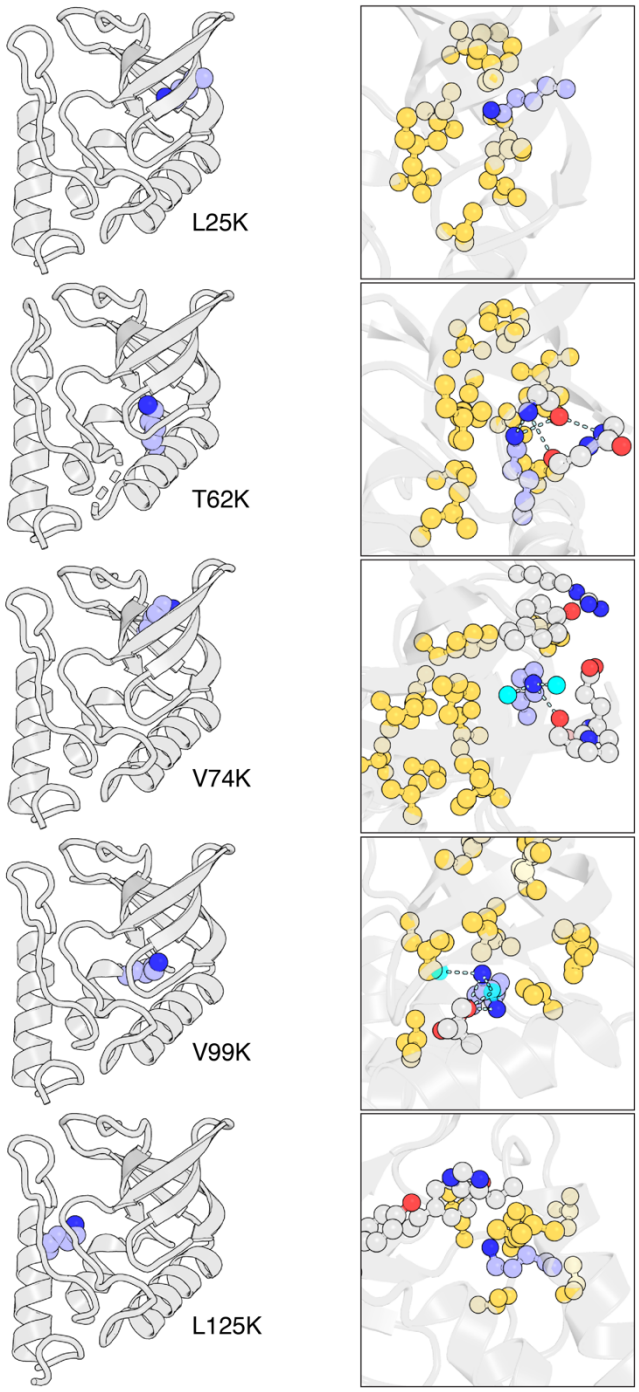

Fig. 2.1. Crystal structures of internal Lys variants of SNase. Backbone conformation shown on left. Close up on microenvironment of buried Lys shown on right. Internal Lys highlighted in light blue, water molecules in cyan, hydrophobic residues in yellow, and polar or ionizable surface residues in gray.

Lys residue (Apparent pK_a[5])	Backbone behavior	Polar contacts with Nζ in structure (within 3.5 Å)	Lys Nζ exchange when neutral	Lys Nζ pH dependent exchange behavior
62 (8.1)	Large CSPs	2 (Thr-22 backbone NH and C=O)	Fast	Intermediate
74 (7.4)	Unperturbed	3 (2 waters, Pro-11 backbone C=O)	Fast	Fast
99 (6.5)	Broadening	2 (1 water, Thr-62 side chain OH)	Fast	Slow
25 (6.3)	Broadening	0	Slow	Slow
125 (6.2)	Broadening	0	Intermediate	Slow

Table 2.1. Summary of microenvironment and NMR exchange behaviors of internal Lys residues.

2.3.2 ^{15}N-NMR spectroscopy

Backbone ^{1}H-^{15}N HSQC experiments on these variants showed previously that the response to changes in pH are very different for these 5 proteins (Table 1), ranging from minor chemical shift perturbations to broadening and increased random coil resonances indicative of partial unfolding, presumably to allow the internal Lys access to solvent[21]. The Nζ resonance of the internal Lys residues in the T62K, V74K, and V99K variants were detectable by ^{15}N 1D pulse acquire experiments and assigned using the H2(C)N Nζ-Hε correlation experiment (Fig. 2.2, S2.1). The resonance belonging to the internal Lys residue was very weak in the ^{15}N 1D experiment of the L25K variant, and completely absent in the case of the L125K variant. Lys-125 was also not detectable using the H2(C)N experiment. In these two cases, ^{1}H-^{15}N-heteronuclear cross polarization[22-27] (HCP) was used to increase the sensitivity of the ^{15}N 1D experiment (Fig. S2.2). In ^{1}H coupled spectra at pH values where the internal Lys residue was expected to be neutral, the Lys-25 Nζ resonance displayed a 1:2:1 triplet splitting with a J_{NH}~65 Hz, consistent with a deprotonated Lys and significantly attenuated solvent exchange[12] (Fig. 2.3A). In the case of the variant with Lys-125, the resulting resonance in the HCP experiment was broad even at pH values significantly above the apparent pK_a value of 6.2, although splittings caused by Nζ-Hζ couplings were observed at low temperatures. In cases where splittings were observed in ^{1}H coupled spectra, the Nζ:Hζ resonance could be clearly observed in Nζ selective ^{1}H-^{15}N HSQC experiments[13] (Fig. 2.3B). The ^{1}H chemical shifts of the neutral NH_2 species are highly upfield shifted relative to the chemical shift of protonated Lys NH_3^+ resonances (6-8 ppm) and consistent with those observed in sp_3-type alkyl-NH_2 compounds (1-2 ppm)[12], though the ^{1}H chemical shift of Lys-25 displayed an even greater upfield shift than those of Lys-125 or the previously studied Lys-66 in the V66K variant of SNase[12]. No splittings were observed in the other variants, consistent with water

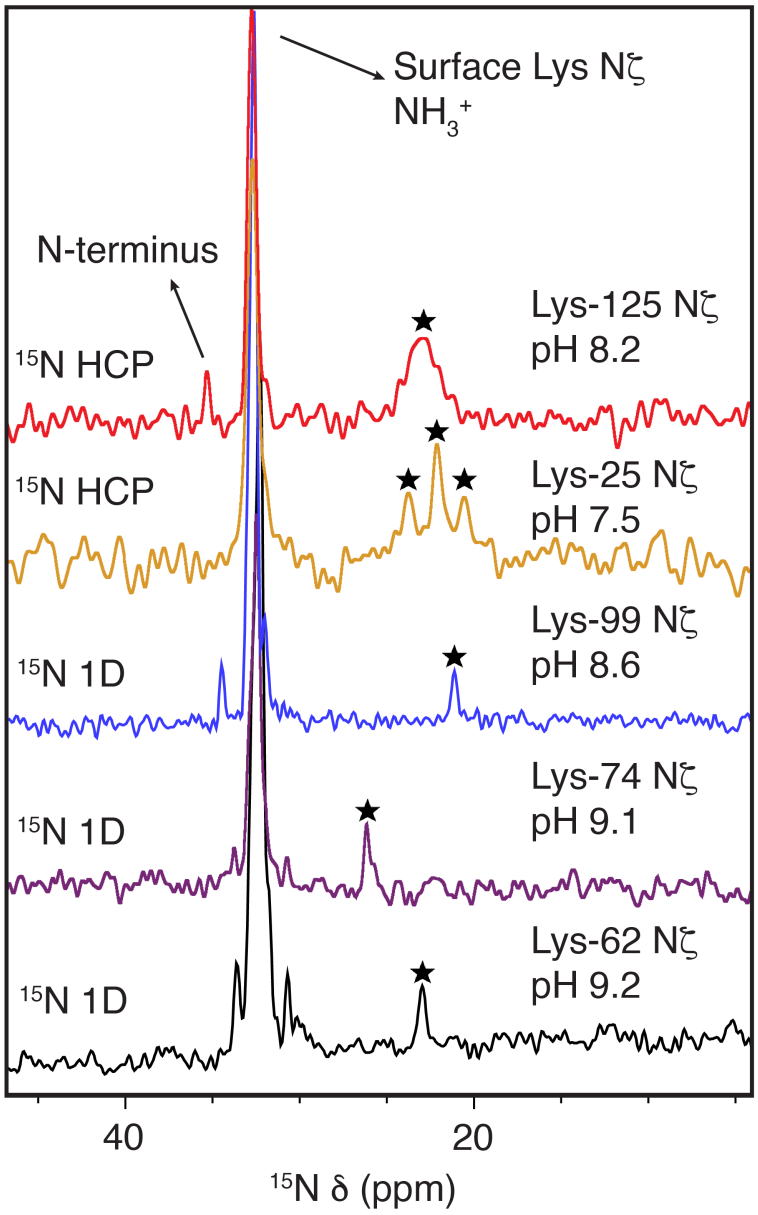

Fig. 2.2. ^{15}N 1D spectra of five internal Lys variants of SNase. Resonances from internal Lys residues are highlighted with stars. All spectra were collected 1 – 2 pH units above the pK_a values of the buried Lys residues.

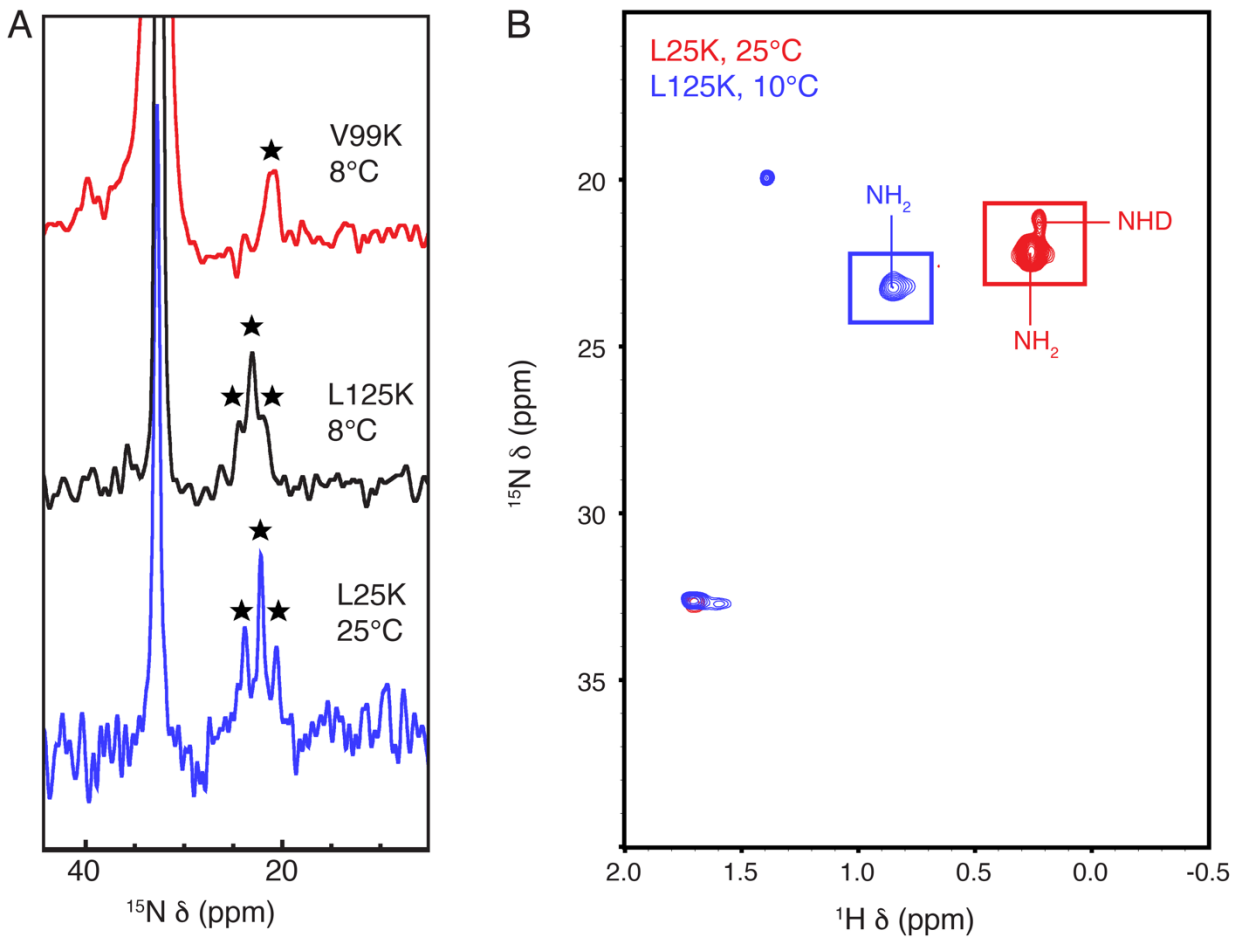

Fig. 2.3. Solvent exchange behavior of variants with Lys-99, Lys-25 and Lys-125. (A) ^{15}N HCP 1D experiments for the variant with V99K at pH 7.0, for L25K at pH 7.3, and for L125K at pH 7.6. Stars identify resonances and splitting patterns. (B) ^{1}H-^{15}N Nζ selective HSQC of variants with Lys-25 and Lys-125.

penetration or transient backbone relaxation exposing the Lys to bulk solvent, causing significant self-decoupling.

2.3.3 pH titrations monitored with ^{15}N 1D experiments

The conformational response of the protein to ionization of the internal Lys residues was probed with pH titrations monitoring their Nζ chemical shifts (Fig. 2.4-6). The Nζ resonance of Lys-74 remained sharp throughout the pH titration in both ^{15}N 1D and H2(C)N experiments (Fig. 2.4). In contrast, the Nζ resonance of Lys-62 was not detectable by the H2(C)N experiment below pH 9.3. In this case, the ^{15}N 1D experiment was necessary to detect the pH dependence of the Nζ chemical shift, the resonance of which became progressively broader as the apparent pK_a value was approached. Lys-99, Lys-25 and Lys-125 exhibited no shift of the Nζ resonance in ^{15}N 1D, ^{1}H-^{15}N HSQC, and ^{15}N HCP 1D experiments, respectively; a pH dependent loss in the intensity of the resonance was observed instead.

The pH dependence of the lineshape behavior observed in these experiments provided insight into the protonation dynamics of the internal Lys residues. Chemical exchange, originating from either solvent exchange or conformational exchange between two distinct species, can influence the lineshape of NMR resonances in several ways, depending on the timescales of inter-conversion, populations of states, and differences in chemical shift (referred to as Δω, where ω is the Larmor frequency)[28,29]. The titration of Lys-74 offers an example of fast exchange (k_{ex}>>Δω) in which the chemical shift is the population-weighted average of both the protonated and deprotonated forms of the residue. Backbone amide chemical shift perturbations (CSPs) above and below the pK_a value of Lys-74 were minor and on the scale observed in the background

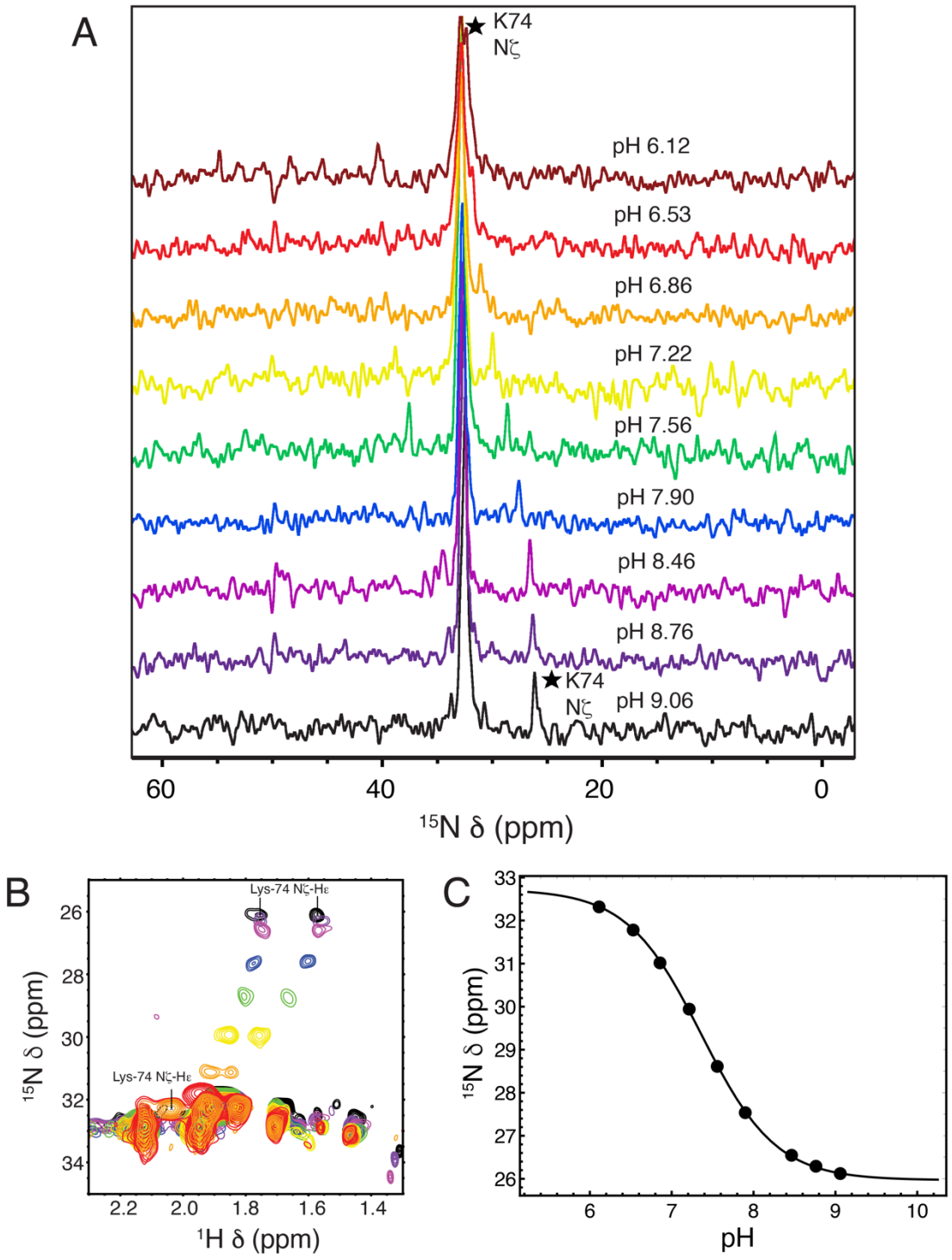

Fig. 2.4. pH titration of K74. **(A)** ^{15}N 1D titration of V74K. **(B)** H2(C)N titration confirms assignment of K74 Nζ resonance. **(C)** Fit of K74 ^{15}Nζ chemical shift to a Henderson-Hasselbalch-Hill equation. The pK_a value was determined to be 7.4, and Hill coefficient 0.96.

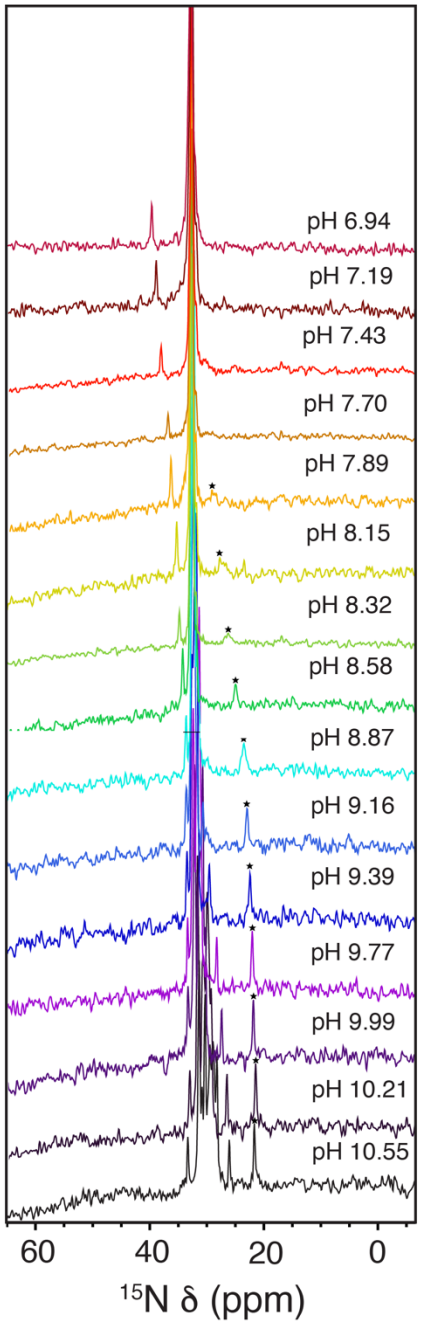

Fig. 2.5. pH titration of K62. ^{15}N 1D pH titration of T62K. Resonances from K62 Nζ are denoted with stars.

protein (Fig. 2.7), suggesting that ionization of Lys-74 does not result in significant structural reorganization of the protein. The pK_a value of Lys-74 determined by analysis of the pH-dependence of the ^{15}Nζ chemical shift[30] was shown to be in excellent agreement with that obtained through linkage thermodynamics[5] (Fig. 2.4C).

In contrast, the Nζ resonance of Lys-62 exhibited pH dependent line broadening in both the ^{15}N 1D and H2(C)N experiments (Fig. 2.5). Broadening is indicative of intermediate exchange (k_{ex}~Δω). Consistent with this observation, large backbone amide CSPs were observed in regions adjacent to the amine of Lys-62 (Fig. 2.7)[16]. The broadening of the Lys-62 ^{15}Nζ resonance is consistent with an internal Lys that is highly dynamic, while the large backbone CSPs reflects reorganization to hydrate the charged Lys residue, or changes in the electronic environment due to the presence of a partially buried NH_3^+ group.

The pH dependent decrease in resonance intensity observed in variants with Lys-25, Lys-99, Lys-125 is caused by slow exchange between two populations (k_{ex}<<Δω, Fig. 2.6). In backbone amide ^{1}H-^{15}N HSQC experiments, the resonances of the V99K and L25K[31] variants underwent pH dependent broadening at or below the apparent pK_a values. The broadening of folded resonances is concomitant with an increase in disordered resonances, with exchange rates between them in the hundreds of ms, as measured by ZZ-exchange spectroscopy[31]. L125K exhibited more complex behavior, as both the ^{15}Nζ resonance and the backbone amide resonances[31] on the helix hosting Lys-125 undergo pH independent broadening, though previous relaxation dispersion[31] and circular dichroism spectroscopy experiments[5] have shown that a pH dependent transition occurs below the apparent pK_a of Lys-125.

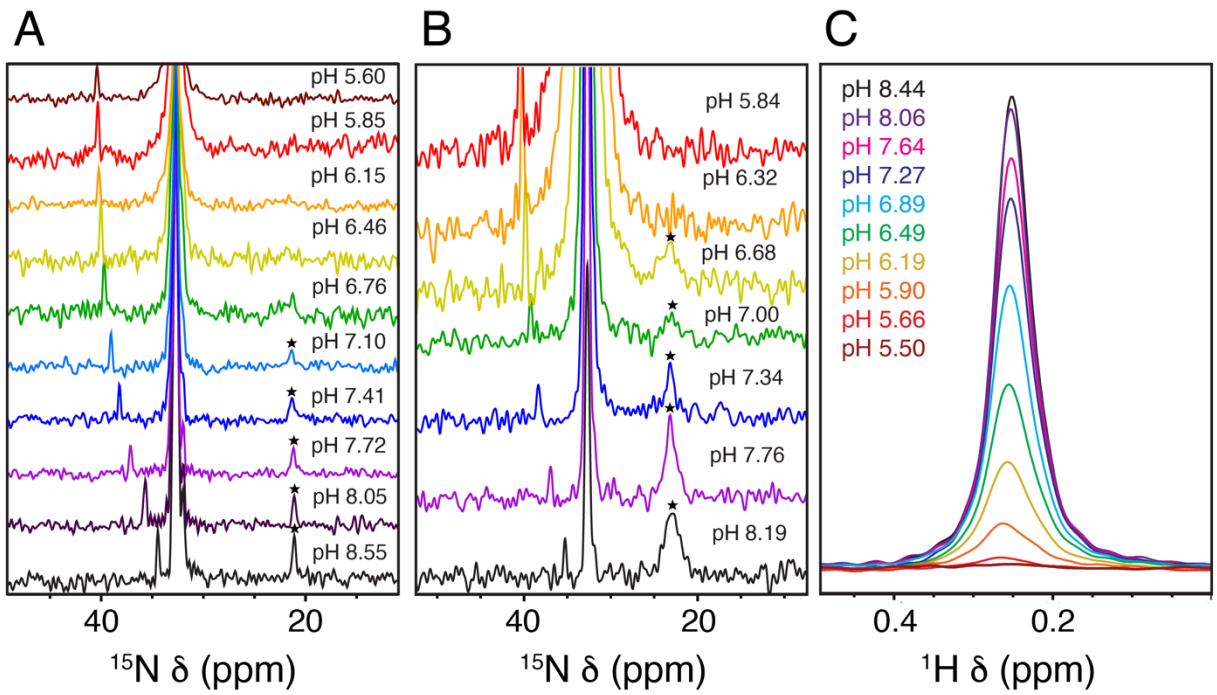

Fig. 2.6. Slow exchange pH titration of K25, K99, and K125. **(A)** ^{15}N 1D titration of V99K. **(B)** ^{15}N-^{1}H 1D HCP titration (^{15}N detected) of L125K. **(C)** Nζ/Hζ selective ^{1}H-^{15}N 1D HSQC titration (^{1}H detected) of L25K. In all three cases, the protonated resonances are not observable over the surface Lys Nζ/Hζ resonances.

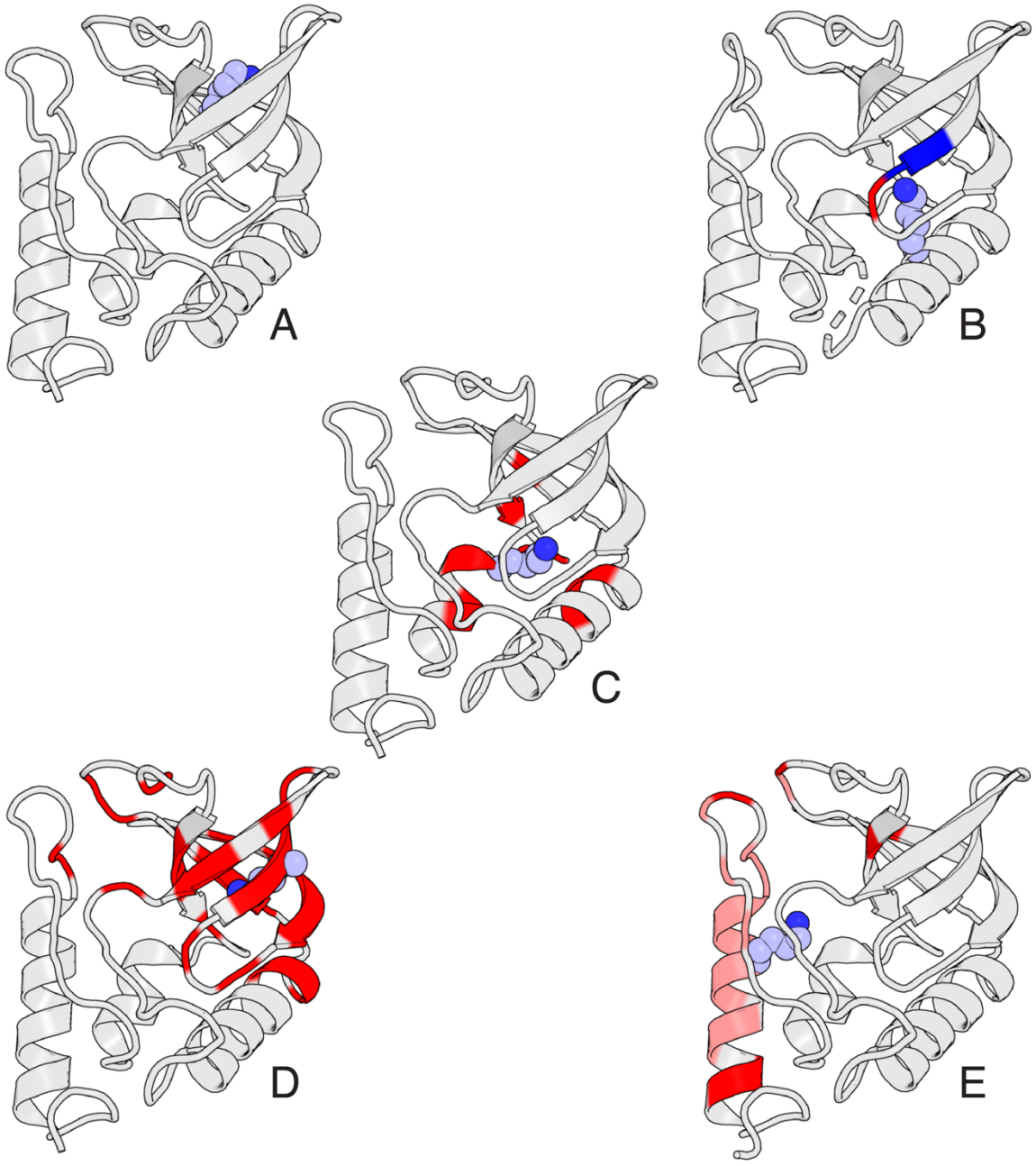

Fig. 2.7. pH-dependent backbone reorganization detected by ^{1}H-^{15}N HSQC. Variants with **(A)** K74, **(B)** K62, **(C)** K99, **(D)** K25 and **(E)** K125. Residues experiencing fast exchange behavior with weighted ^{1}H-^{15}N chemical shift perturbations greater than 0.5 ppm (α = 0.24) shown in blue, residues showing pH-dependent resonance broadening below the pK_a value of the internal Lys in red, and residues showing pH-independent resonance broadening in light red.

The slow exchange observed in these experiments reflects the shift in populations from a fully folded state where the Lys is buried, to a partially folded state where the internal Lys residue is solvent exposed and indistinguishable from the surface Lys residues in ^{15}N 1D experiments. In these proteins the ^{15}N detect experiments provided direct evidence of coupling between protonation of internal Lys residues and conformational reorganization. The similarity in the pH dependent exchange behavior and pK_a values of these three variants is noteworthy given the very different microenvironments sampled by the internal Lys side chains at positions 25, 99 and 125, as evidenced by the crystal structures and splitting patterns of the ^{15}Nζ resonance in ^{1}H-coupled spectra (Table 1). These three internal Lys residues demonstrate that when internal Lys residues are in positions where the protein cannot reorganize easily, increased hydration (e.g., Lys-99) or dynamics (e.g., L125K) may not be enough to compensate for the dehydration penalty associated with burial of an ionizable residue. In these cases, backbone reorganization is necessary to allow internal residues to access solvent.

2.4 Discussion

This study demonstrates that ^{15}N detect experiments can provide detailed insight into the properties of buried Lys residues and the pH dependent conformational transitions they drive in proteins. These experiments will enable unprecedented functional studies in proteins where buried Lys residues are essential[32]. Recent advances in probe technology have led to dramatic increases in signal-to-noise in ^{15}N detect experiments[17], so these experiments should be broadly applicable to larger proteins. This is underscored by the fact that in cases where both ^{15}N 1D and H2(C)N experiments were successful, similar acquisition times were needed for both experiments despite the ^{1}H detect experiments being done on a ^{1}H optimized cryogenic probe, while the ^{15}N 1D

experiments were performed on a much older, room temperature broad band probe. In cases where solvent or conformational exchange leads to pH independent broadening in ^{15}N 1D or H2(C)N experiments, sensitivity enhancement using heteronuclear cross polarization was absolutely essential for the detection of Lys residues in intermediate exchange. The range of behaviors observed in the experiments with SNase already suggest the serious challenge facing structure-based energy calculations of buried ionizable groups in proteins. Methods for pK_a calculations will have to be able to distinguish between cases where ionization of an internal residue causes minor structural fluctuations (V74K and T62K) and cases where ionization drives large-scale backbone rearrangements (V99K, L25K, L125K). Accurate prediction of the anomalous pK_a values of internal residues will not only require identification of the alternative conformational states accessible, some of which are sampled in timescales of hundreds of milliseconds[31], but also accurate calculation of the free energy gap between states.

2.5 Methods

2.5.1 Sample preparation

Proteins were expressed using pET24a+ plasmids with the SNase Δ+PHS variant sequence; variants were prepared previously[5]. For NMR studies, *E. coli* BL21 DE3 cells transformed with the plasmids were grown in 5 ml of LB for 4 hours at 37 ºC before being transferred to 250 ml of M9 minimal media supplemented with ^{15}N labelled ammonium chloride, and when appropriate, ^{13}C labelled glucose (purchased from Cambridge Isotopes). Cells were grown for a further four hours before being left at 4 ºC overnight. The following morning, the 250 ml culture was added to a further 1 L of M9. Cells were allowed to grow at 37 ºC till reaching an OD_{600} of 0.7-0.9, and then induced with 1 mM IPTG for four hours. Unlabeled

proteins for crystallography were prepared by adding a single colony of BL21 cells to 1.5 L of LB, which were then left to grow for 12-14 hours at 37 ºC before induction with IPTG; afterwards, the purification procedure was identical for both unlabeled and isotopically labeled proteins. Cells were spun down at 2,500 g for 10 minutes, lysed by resuspension (with agitation at 4 ºC for 20-30 minutes) with extraction buffer 1 (EB1, 25 mM Tris, 2.5 mM EDTA, 6 M Urea, pH 8.0), and then spun down again at 6,000 g for 15 minutes to isolate inclusion bodies. The supernatant was discarded, and the pellet resuspended (with agitation at 4 ºC for 40-50 minutes) in extraction buffer 2 (EB2, extraction buffer 1 with 400 mM NaCl). The high salt of EB2 solubilizes inclusion bodies; a further centrifugation at 6,000 g for 15 minutes isolated soluble SNase in the supernatant, which was then 100% ethanol (EtOH) to a final concentration to 50% EtOH and left to precipitate for 2.5-5 hours at -20 ºC. The mixture was then spun down at 6,000 g for 20 minutes to eliminate contaminating nucleic acids and protein, before further diluting the supernatant to a final volume of 75% EtOH. The EB2/75% EtOH mixture was left at -20 ºC overnight to precipitate SNase. The mixture was then centrifuged to separate the pellet of partially purified SNase from the EtOH (6,000 g for 20 minutes), resuspended in 30 ml of EB1, and then passed through SP Sepharose Fast Flow cation exchange resin. Column was washed with 100 ml of EB1, and protein was eluted with 40 ml of EB2. The 40 ml of EB2, which contains the purified SNase protein, was then diluted with EtOH to a final volume of 25% EB2: 75% EtOH and left to precipitate for 3-5 hours at -20 ºC. A final centrifugation at 6,000 g for 20 minutes was used to separate the purified SNase protein from the EtOH, which was then resuspended in 5-10 ml of EB2 and left to dialyze against 4 L of 1 M KCl overnight (4 ºC). The following morning, the protein was dialyzed out of 1 M KCl in steps of 200-250 mM (30-90

minutes each step) until a final dialysis in 4 L of water overnight. The following day, the protein was centrifuged to remove aggregates and flash frozen in liquid nitrogen for storage at -80 ºC.

Proteins were buffer exchanged in appropriate buffers for experimental work using Amicon Ultra 4 mL centrifugal filters (10 kDa cut off). Protein concentrations were determined as previously described. For titrations, sample concentrations were 1-1.5 mM, in 100 mM KCl and 5 mM Tris (T62K) or 5 mM HEPES (V74K) or no buffer (V99K, L25K and L125K), in a 90/10 H_2O:D_2O solution. For side chain assignments, sample concentrations were between 0.9-1.5 mM, in 100 mM KCl and 25 mM CHES (T62K), 25 mM MES (V74K), or 25 mM HEPES (V99K and L25K), in 90/10 H_2O:D_2O. Backbone resonances for L25K[31], L125K[31], and T62K[16] were assigned previously.

2.5.2 NMR Spectroscopy

All ^{1}H detect experiments were performed on a Bruker Avance NMR spectrometer operating at a 600 MHz ^{1}H frequency, equipped with a TXI cryoprobe. All ^{15}N 1D pulse acquire experiments were performed on a 500 MHz Varian Inova spectrometer with a room temperature Broad Band Observe (BBO) probe. All ^{15}N 1D Heteronuclear Cross Polarization (HCP) experiments were performed on a 400 MHz Bruker Avance III spectrometer with a room temperature Broad Band and Fluorine Observe (BBFO) probe. All H2(C)N and Nζ-selective HSQC experiments were performed with 20-30 ms acquisition times in the indirect dimension, and 1 s relaxation delays (1.5 s for L25K 1D Nζ HSQC titration). 64 (V74K titration), 128 (L25K and V99K assignment) and 256 (T62K assignment) scans per FID were required for the H2(C)N experiment. Nζ HSQC titrations on L25K were performed with 8 (2D) and 800 (1D) scans per FID. ^{15}N 1D pulse acquire experiments were performed with 60° pulses to increase the

number of scans per unit time, and 2.5 s relaxation delays. 4000 scans were collected for each pH point for the V74K titration (~1.5 hours). At high pH values (>9), 6000-8000 scans were collected for each pH point for the T62K titration, but this was increased to 16,000-24,000 per pH point at pH values near and below the apparent pK_a value of Lys-62. 10,000 scans were acquired per point for V99K at high pH, which was increased to 15,000-20,000 near and below the apparent pKa value of Lys99. Details of ^{15}N 1D HCP experiment are provided in Figure S14. Assignments were performed using standard triple resonance experiments (CBCA(CO)NH, HNCACB, (H)CC(CO)NH and H(CCCO)NH TOCSY). 1D data was analyzed using Bruker Topspin 3.5. 2D and 3D data were processed using the NMRpipe suite[33] and analyzed using Sparky 3.115[34].

2.5.3 Crystallography

Crystal structures of variants with L25K, T62K, and L125K substitutions were reported previously1. Crystals of variants with V74K and V99K were grown at 4 °C using the hanging drop vapor-diffusion method, using a protein concentration >9 mg/ml. Conditions are provided in Table S3. 4 μl of protein solution (in H2O) mixed with 4 μl of reservoir solution were used. Crystals appeared within 5 weeks and were suspended in mother liquor in a cryoloop (Hampton Research), before being flash frozen in liquid nitrogen. Data collection source and statistics are shown in Table S3, with all data 2 being indexed, integrated, scaled, and merged using the manufacturer's software. Subsequent processing was performed with the CCP4 suite[35], and initial phasing by molecular replacement done with PHASER[36] using the coordinates of the background protein (Δ+PHS variant of SNase, PDB ID: 3BDC) as a search model. Iterative model building and refinement were performed using COOT[37] and Refmac5[38], respectively.

2.5.4 Analysis

All fitting was performed with Mathematica 10.1. The pH dependence of ^{15}N chemical shifts was fitted to a modified Henderson-Hasselbalch equation[39], where the protonated and deprotonated chemical shifts (δ_+ and δ_o), as well as the Hill coefficient (n) and pK_a value were fitted parameters:

$$\delta(pH) = \delta_+ - (\delta_+ - \delta_o)\frac{10^{n(pH-pKa)}}{1+10^{n(pH-pKa)}}$$

Chemical shift perturbation analysis of backbone amides was performed using the following equation[40]:

$$\Delta\delta_{NH} = \sqrt{\frac{1}{2}\left(\Delta\delta_H^2 + \frac{\Delta\delta_N^2}{25}\right)}$$

2.6 References

1) von Ballmoos, C.; Wiedenmann, A.; Dimroth, P. Essentials for ATP Synthesis by F_1F_o ATP Synthases. *Annu. Rev. Biochem.* **2009**, 78, 649-672.

2) Ho, M.-C.; Ménétret, J.-F.; Tsuruta, H.; Allen, K.N. The Origin of the Electrostatic Perturbation in Acetoacetate Decarboxylase. *Nature*. **2009**, 459, 393-397.

3) Malevanets, A.; Chong, P.A.; Hansen, D.F.; Rizk, P.; Sun, Y.; Lin, H.; Muhandiram, R.; Chakrabartty, A.; Kay, L.E.; Forman-Kay, J.D.; Wodak, S.J. Interplay of Buried Histidine Protonation and Protein Stability in Prion Misfolding. *Sci. Rep.* **2017**, 7, 882.

4) Isom, D.G.; Castañeda, C.A.; Cannon, B.R.; Velu, P.D.; García-Moreno E, B. Charges in the Hydrophobic Interior of Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, 107, 16096-16100.

5) Isom, D.G.; Castañeda, C.A.; Cannon, B.R.; García-Moreno E, B. Large Shifts in pK_a Values of Lysine Residues Buried Inside a Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, 108, 5260-5265.

6) Di Russo, N.V.; Estrin, D.A.; Mart, M.A.; Roitberg, A.E. pH-Dependent Conformational Changes in Proteins and their Effect on Experimental pK_as: the Case of Nitrophorin 4. *PLOS Comput. Biol.* **2012**, 8, e1002761.

7) Damjanovic, A.; Brooks, B.R.; García-Moreno E, B. Conformational Relaxation and Water Penetration Coupled to Ionization of Internal Groups in Proteins. *J. Phys. Chem. A.* **2011**, 115, 4042-4053.

8) Goh, G.B.; Laricheva, E.N.; Brooks III, C.L. Uncovering pH-Dependent Transient States of Proteins with Buried Ionizable Residues. *J. Am. Chem. Soc.* **2014**, 136, 8496-8499.

9) Wu, X.; Lee, J.; Brooks, B.R. Origin of pK_a shifts of Internal Lysine Residues in SNase Studied Via Equal-Molar VMMS Simulations in Explicit Water. *J. Phys. Chem. B.* **2017**, 121, 3318-3330.

10) Zheng, Y.; Cui, Q. Microscopic Mechanisms that Govern the Titration Response and pK_a Values of Buried Residues in Staphylococcal Nuclease Mutants. *Proteins: Struct., Funct., Bioinf.* **2017**, 85, 268-281.

11) Shi, C.; Wallace, J.A.; Shen, J.K. Thermodynamic Coupling of Protonation and Conformational Equilibria in Proteins: Theory and Simulation. *Biophys. J.* **2012**, 102, 1590-1597.

12) Takayama, Y.; Castañeda, C.A.; Chimenti, M.; García-Moreno E, B.; Iwahara, J. Direct Evidence for Deprotonation of a Lysine Side Chain Buried in the Hydrophobic Core of a Protein. *J. Am. Chem. Soc.* **2008**, 130, 6714-6715.

13) Iwahara, J.; Jung, Y.S.; Clore, G.M. Heteronuclear NMR Spectroscopy for Lysine NH_3 Groups in Proteins: Unique Effect of Water Exchange on ^{15}N Transverse Relaxation. *J. Am. Chem. Soc.* **2007**, 129, 2971-2980.

14) Poon, D.K.Y.; Schubert, M.; Au, J.; Okon, M.; Withers, S.G.; McIntosh, L.P. Unambiguous Determination of the Ionization State of a Glycoside Hydrolase Active Site Lysine by ^{1}H-^{15}N Heteronuclear Correlation Spectroscopy. *J. Am. Chem. Soc.* **2006**, 128, 15388-15389.

15) André, I.; Linse, S.; Mulder, F.A.A. Residue-Specific pK_a Determination of Lysine and Arginine Side Chains by Indirect ^{15}N and ^{13}C NMR Spectroscopy: Application to *apo* Calmodulin. *J. Am. Chem. Soc.* **2007**, 129, 15805-15813.

16) Khangulov, V. Molecular Determinants of pK_a Values of Internal Ionizable Groups in Proteins: Interactions of Internal Ionizable Groups with Surface Charges. **2011**, Ph.D. Thesis, Johns Hopkins University, Baltimore, MD.

17) Takeuchi, K.; Heffron, G.; Sun, Z.-Y.J.; Frueh, D.P.,; Wagner, G. Nitrogen-Detected CAN and CON Experiments as Alternative Experiments for Main Chain NMR Resonance Assignments. *J. Biomol. NMR*. **2010**, 47, 271-282.

18) Lin, I-J.; Chen, Y.; Fee, J.A.; Song, J.; Westler, W.M.; Markley, J.L. Rieske Protein from *Thermus thermophilus*: ^{15}N NMR Titration Study Demonstrates the Role of Iron-Ligated

Histidines in the pH Dependence of the Reduction Potential. *J. Am. Chem. Soc.* **2006**, 128, 10672-10673.

19) Zhu, L.; Kemple, M.D.; Yuan, P.; Prendergast, F.G. N-Terminus and Lysine Side Chain pK_a Values of Melittin in Aqueous Solutions and Micellar Dispersions Measured by ^{15}N NMR. *Biochemistry*. **1995**, 34, 13196-13202.

20) Preimesberger, M.R.; Majumdar, A.; Lecomte, J.T.J. Dynamics of Lysine as a Heme Axial Ligand: NMR Analysis of the *Chlamydomonas reinhardtii* Hemoglobin THB1. *Biochemistry*. **2017**, 56, 551-569.

21) Chimenti, M.S.; Khangulov, V.S.; Robinson, A.C.; Heroux, A.; Majumdar, A.; Schlessman, J.L.; García-Moreno E, B. Structural Reorganization Triggered by Charging of Lys Residues in the Hydrophobic Interior of a Protein. *Structure*. **2011**, 20, 1071-1085.

22) Bertrand, R.D; Moniz, W.B.; Garroway, A.N.; Chingas, G.C. Sensitivity Enhancement in ^{15}N NMR Spectroscopy of Liquids Using the J Cross-Polarization Technique. *J. Mag. Res.* **1978**, 32, 465-467.

23) Zuiderweg, E.R.P. Analysis of Multiple-Pulse-Based Heteronuclear J Cross Polarization in Liquids. *J. Mag. Res.* **1990**, 89, 533-542.

24) Majumdar, A.; Wang, H.; Morshauser, R.C.; Zuiderweg, E.R.P. Sensitivity Improvement in 2D and 3D HCCH Spectroscopy Using Heteronuclear Cross-Polarization. *J. Biomol. NMR*. **1993**, 3, 387-397.

25) Kellogg, G.W.; Schweitzer, B.I. Two- and Three-Dimensional ^{31}P-Driven NMR Procedures for Complete Assignment of Backbone Resonances in Oligodeoxyribonucleotides. *J. Biomol. NMR*. **1993**, 3, 577-595.

26) Majumdar, A.; Zuiderweg, E.R.P. Efficiencies of Double- and Triple-Resonance J Cross Polarization in Multidimensional NMR. *J. Mag. Res*. **1995**, 113, 19-31.

27) Yuwen, T.; Skrynnikov, N.R. J. CP-HISQC: A Better Version of HSQC Experiment for Intrinsically Disordered Proteins Under Physiological Conditions. *J. Biomol. NMR*. **2014**, 58, 175-192.

28) Kleckner, I.R.; Foster, M.P. An Introduction to NMR-Based Approaches for Measuring Protein Dynamics. *Biochim. Biophys. Acta*. **2011**, 1814, 942-968.

29) Millet, O.; Loria, J.P.; Kroenke, C.D.; Pons, M.; Palmer III, A.G. The Static Magnetic Field Dependence of Chemical Exchange Linebroadening Defines the NMR Chemical Shift Time Scale. *J. Am. Chem. Soc*. **2000**, 122, 2867-2877.

30) Markley, J.L. Observation of Histidine Residues in Proteins by Nuclear Magnetic Resonance Spectroscopy. *Acc. Chem. Res*. **1975**, 8, 70-80.

31) Richman, D.E.; Majumdar, A.; García-Moreno E, B. Conformational Reorganization Coupled to the Ionization of Internal Lys Residues in Proteins. *Biochemistry*. **2015**, 54, 5888-5897.

32) Hacker, S.M.; Backus, K.M.; Lazear, M.R.; Forli, S.; Correia, B.E.; Cravatt, B.F. Global Profiling of Lysine Reactivity and Ligandability in the Human Proteome. *Nat. Chem*. **2017**, doi:10.1038/nchem.282

33) Delaglio, F.; Grzesiek, S.; Vuister, G.W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR*. **1995**, 6, 277-293.

34) Goddard, T.; Kneller, D. Sparky 3. University of California San Francisco, **2008**.

35) Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; *et. al*. Overview of the CCP4 Suite and Current Developments. *Acta Crys. D*. **2011**, 67, 235-242.

36) McCoy, A.J.; Grosse-Kunstleve, R.W.; Storoni, L.C.; Read, R.J. Likelihood-Enhanced Fast Translation Functions. *Acta Crys. D*. **2005**, 61, 458-464.

37) Emsley, P. and Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crys. D*. **2004**, 60, 2126-2132.

38) Bailey, S. The CCCP4 Suite: Programs for Protein Crystallography. *Acta Crys. D*. **1994**, 50, 760-763.

39) Markley, J.L. Observation of histidine residues in proteins by nuclear magnetic resonance spectroscopy. *Acc. Chem. Res*. **1975**, 8, 70-80.

40) Grzesiek, S.; Bax, A.; Clore, G.M.; Gronenborn, A.M.; Hu, J.-S.; Kaufman, J.; Palmer, I.; Stahl, S.J.; Wingfield, P.T. The Solution Structure of HIV-1 Nef Reveals an Unexpected Fold and Permits Delineation of the Binding Surface for the SH3 Domain of Hck Tyrosine Protein Kinase. *Nat. Struc. Mol. Biol*. **1996**, 3, 340-345.

2.7 Supporting Information

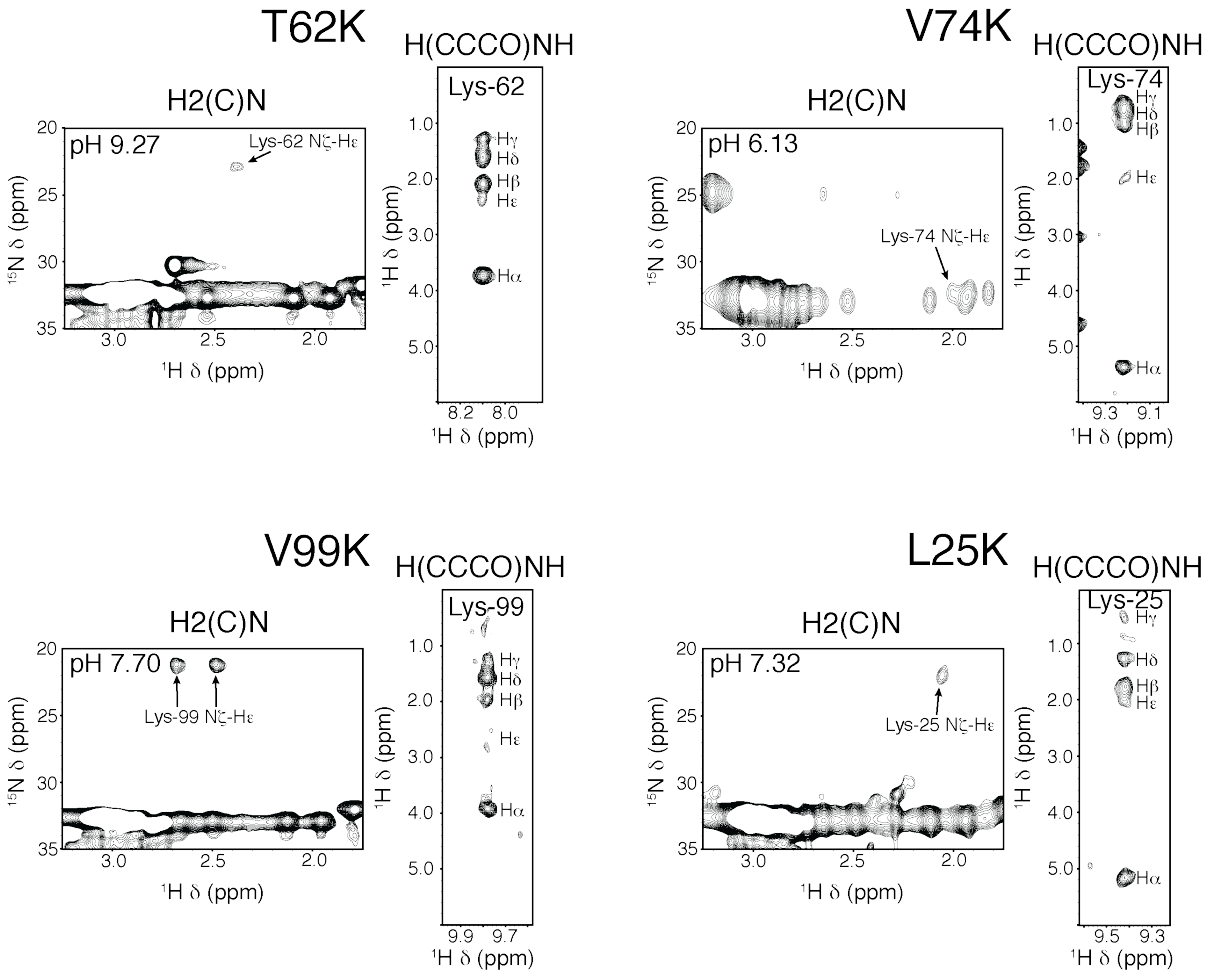

Fig. S2.1. Assignment of Nζ resonances observed in ^{15}N 1D experiments. Nζ-Hε correlations were obtained using the H2(C)N experiment, which were assigned to the internal Lys residues using the H(CCCO)NH-TOCSY experiment. 2D planes from 3D TOCSY experiment corresponding to the internal Lys of interest are shown.

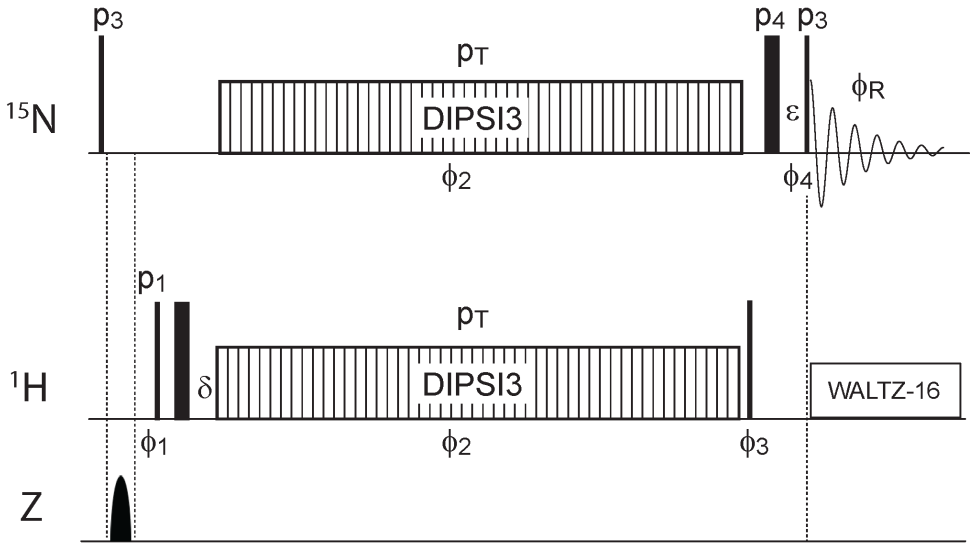

Fig. S2.2 Heteronuclear ^{1}H-^{15}N Cross-Polarization (HCP) pulse sequence for ^{15}N detected 1D spectra. Thin and thick lines represent high power 90° and 180° pulses, respectively. All pulse phases are x, unless otherwise specified. RF carrier positions: ^{1}H: H_2O (4.77 ppm at 25 °C), ^{15}N: 30 ppm. High power pulses p1 (^{1}H) = 14-15 μs and p3 (^{15}N) = 16 μs. Delays $\delta = 2\times p1/\pi$, ε=p1. Cross-polarization parameters: pulse train = DIPSI-3, pT = 60 μs ($\nu 1 \sim 4.2$ kHz), duration 39 ms. Pulse phases: $\varphi 1$ = y, $\varphi 2$ = x, $\varphi 3$ = -y, $\varphi 4$ = x, x, -x, -x; φR = x, -x. Gradient: 1 ms, 50% of maximum. Recycle delay: 2 s. Optional WALTZ ^{1}H decoupling during acquisition.

Chapter 3

Protein electrostatics governed by pH-redistribution of the ensemble

Version currently available on the bioRxiv preprint server as Kougentakis, C.M., Majumdar, A., and García-Moreno, E.B. Electrostatic effects in proteins are governed by pH-redistribution of the conformational ensemble. *biorXiv*, doi:10.1101/2020.02.02.931253, (2020).

3.1 Abstract

The imperative for charges to be hydrated is one of the most important organizing principles in biology, responsible for the general architecture of biological macromolecules and for energy storage in the form of electrochemical gradients. Paradoxically, many functional sites in proteins have buried ionizable groups[1]. These groups are tolerated because they are usually buried in the neutral state[2]. However, when they become charged they can drive structural transitions to open states in which the charge can be stabilized, mostly through interactions with water[3]. This coupling between the ionization of a buried group and conformational reorganization is precisely the mechanism used by proteins to perform energy transduction[4,5,6]. By applying this principle to a family of 25 variants of staphylococcal nuclease with internal Lys residues, it was possible to characterize in detail the range of localized partial unfolding events that even a highly stable protein that unfolds cooperatively can undergo in response to H^+-binding. Conformational states that constitute vanishingly small populations of the equilibrium native state ensemble of this protein were identified by correlation of structural and thermodynamic data, providing a map of the conformational landscape of this protein with unprecedented detail. The data demonstrate that the apparent pK_a values of buried ionizable residues are not determined by the properties of their microenvironment but by the intrinsic propensity of the protein to populate open states in which internal charged residues can be hydrated. The role of buried residues in functional sites in proteins relies on their ability to tune the conformational ensemble for redistribution in response to small changes in pH. These results provide the physical framework necessary for understanding the role of pH-driven conformational changes in driving biological energy transduction[4], the identification of pH-sensing proteins in nature[7], and for the engineering of pH-sensitive dynamics and function in *de novo* designed proteins[8].

3.2 Introduction

Buried ionizable residues in proteins are critical for many biochemical processes, especially energy transduction[4], catalysis[1], and pH-sensing[7]. Owing to the incompatibility of charges with hydrophobic environments, the pK_a values of these residues can be highly anomalous, shifted in the direction that promotes the neutral state[2]. In proteins such as ATP synthase[4], cytochrome c oxidase[5], and bacteriorhodopsin[6], the key molecular events of the energy transduction step involve reorganization of the protein in response to the transient presence of a charge in a hydrophobic environment. This reorganization exposes the buried group to polar environments where their pK_a values are more normal relative to their values in water[5]. It is well established that H^+-coupled conformational changes require that proteins sample multiple conformations with different proton binding affinities (i.e., ionizable residues must have different pK_a values in the different conformational states)[9]. What is not understood is the nature of the conformational reorganization that can be coupled to the change in ionization state of a buried group. The study of 25 variants of staphylococcal nuclease (SNase) with Lys residues at internal positions[2] presented here addresses this question directly.

The pK_a values of the internal Lys residues in SNase (Fig. 3.1A)[2] were measured previously by linkage analysis of the pH-dependence of thermodynamic stability. In nineteen cases, the internal Lys pK_a is depressed relative to its pK_a in water, some by more than 5 pH units. Crystal structures show that the overall protein fold is unaffected when the buried Lys residues are neutral (Fig. S3.1)[10]. No correlation between either the structural features of the microenvironment of the buried Lys residues or the magnitude of the pK_a shifts is apparent. In a small number of cases, NMR spectroscopy and X-ray crystallography have provided evidence of structural reorganization in response to ionization of buried Lys residues, presumably to expose the group to bulk solvent[11].

The low throughput of these methods has previously prevented a systematic study of the types of conformational changes coupled to H^+- binding to the buried Lys residues.

To address these limitations, recently developed NMR methods were used to characterize the backbone response to ionization of internal Lys residues in all 25 variants of SNase. The apparent pK_a values measured by linkage thermodynamics were found to report on the pH-driven redistribution of the conformational ensemble coupled to the ionization of buried Lys. By systematically examining the alternatively folded or partially unfolded states sampled in response to ionization of Lys at different internal positions throughout the protein structure, it was possible to construct a conformational landscape of the dominant states in the conformational landscape of SNase. These results demonstrate how pH-driven redistribution of the ensemble can act as a determinant of electrostatic effects in proteins, providing the physical insight necessary to inform the next generation of structure-based pK_a calculation methods.

3.3 Results

3.3.1 Chemical shift (CS) perturbations and structural analysis of CS

Advances in data acquisition[12] and analysis[13] protocols have allowed for a dramatic increase in the throughput of NMR spectroscopy experiments, providing a means to describe with atomic detail the location and amplitude of structural reorganization that is coupled to the ionization of buried Lys residues in all 25 SNase variants. Chemical shift perturbations (CSP) of the backbone ^{1}H, ^{13}C, and ^{15}N nuclei were measured for each variant at different pH values, or between the background and variants (Fig. 3.1B, Fig. S3.2-3). In two cases the ionization of the internal Lys triggered global unfolding (Fig. S3.4). In the other 23 variants the CSPs and/or loss

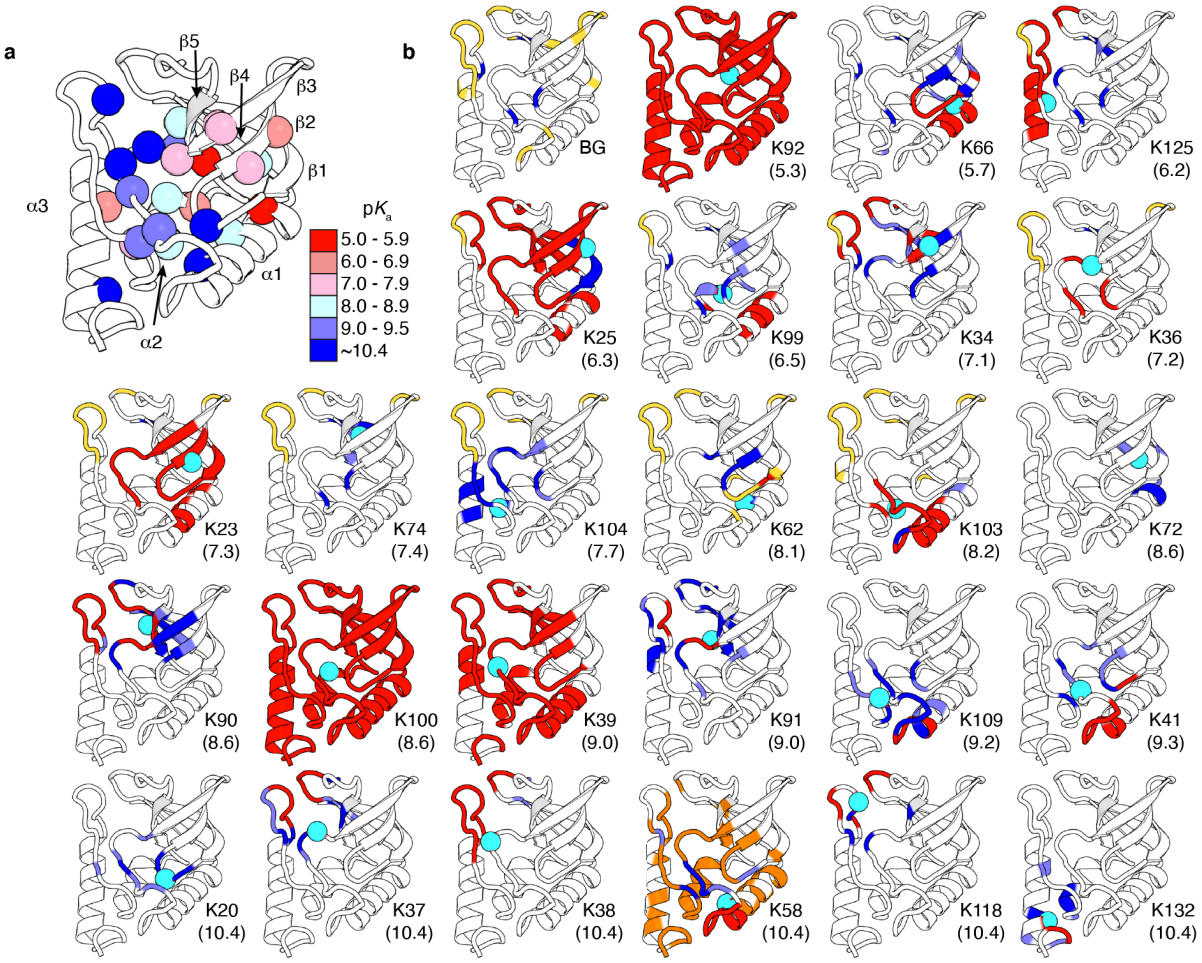

Fig. 3.1. Structural reorganization coupled to the ionization of buried Lys residues. (a) Crystal structure of staphylococcal nuclease (PDB ID: 3BDC). Spheres identify sites of substitutions, color-coded according to the pK_a value of Lys residue at that position. **(b)** Chemical shift perturbations (CSPs) of backbone $^{1}H_N$, $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, and ^{15}N shifts. Background protein (BG) pH-dependent CSPs between pH 9.4 and 4.6 are shown for reference. For variants with internal Lys residues with pK_a values below 8.5, CSPs compare 1.2-1.5 pH units above and below the pK_a of the internal Lys. In order to avoid assignment difficulties at high pH due to base catalyzed exchange, CSP analysis for variants with pK_a values above 8.5 were calculated by comparing the variant with internal Lys at pH 7.2 with the background protein at pH 7.4. The pK_a values of the buried Lys are listed in parenthesis for each

variant. Red denotes loss of resonance intensity, light and dark blue denote regions with moderate and large CSPs, respectively. Residues that were not observable owing to base catalyzed exchange at high pH are shown in yellow. Residues in slow exchange with two assignable conformations are shown in orange. A cyan sphere identifies the site of substitution with Lys.

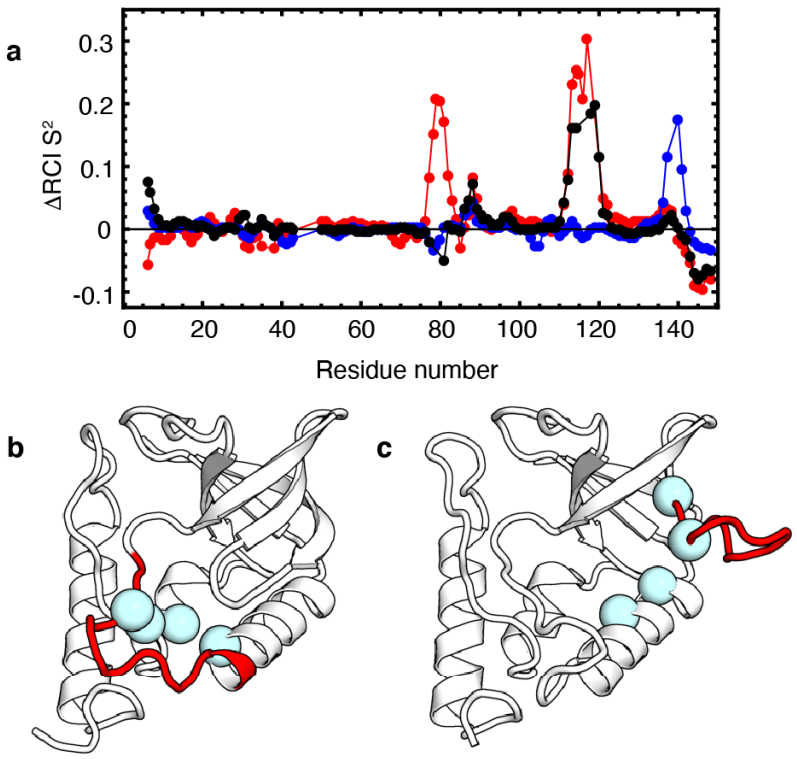

Fig. 3.2. Structural characterization of pH-dependent conformational changes necessary to hydrate buried Lys residues. (a) Difference in predicted order parameters (predicted using secondary shifts and the Random Coil Index) between the background protein at pH 7.4 and variants with K91 (red), K118 (black), and K132 (blue) at pH 7.2 (pH values where Lys is ionized). Positive values indicate regions that are predicted to be less ordered in the variant relative to the background protein. Based on error estimates of the RCI analysis, values below 0.08 or from terminal residues should not be considered significant. **(b-c)** Crystal structures of variants with charged **(b)** R109 (PDB ID: 3D4W) and **(c)** E23 (PDB ID: 3TME). Regions showing evidence of structural reorganization relative to the background protein are highlighted in red. In Lys-containing variants where chemical shift changes or broadening are in regions showing partial unfolding as in the structures, the positions of the Lys substitutions are shown with spheres.

of resonance intensities that were observed were consistent with local or subglobal changes in structure[14] or dynamics[15]. In some instances, these changes were concomitant with an increase in random coil resonances (Fig. S3.4-6, Table S3.1-2). Structural analysis[16] of the chemical shifts of variants at pH values where Lys is charged suggests that in many cases these changes reflect a loss of structure relative to the background protein or conformation at high pH (Fig. 3.2A, S3.7). In other cases crystal structures confirm that the nature of the pH-dependent structural changes observed by NMR is partial unfolding (Fig. 3.2B-D)[11]. These data demonstrate unequivocally that ionization of the buried Lys residues is coupled to structural reorganization, shifting the population distribution from a predominantly fully folded (closed) state into an ensemble of locally, partially or fully unfolded (open) states in which the charged Lys residues are hydrated.

3.3.2 Linkage thermodynamics

The thermodynamically determined pK_a values of these buried Lys residues are apparent pK_a values (pK_a^{App}) that do not report on the pK_a of the Lys in the buried state, but on the pH-dependent conformational equilibrium between the closed and the open states (Fig. 3.3)[2,11]. In the open states the pK_a values should be comparable to the normal pK_a of Lys in water whereas in the closed states the pK_a must be lower than the measured pK_a^{App}. This observation has important implications for structure-based electrostatic calculations of pK_a values: the apparent pK_a values of buried groups are not determined by the electrostatic properties of the microenvironments around the buried ionizable moieties[17], but by the propensity of the protein to populate the open states.

To understand the physical origin of these pH-driven structural transitions, it is necessary to correlate the structural data from Figure 3.1 and 3.2 with equilibrium stability measurements. Owing to the shifted pK_a values of the buried Lys residues, at pH below the normal pK_a of 10.4

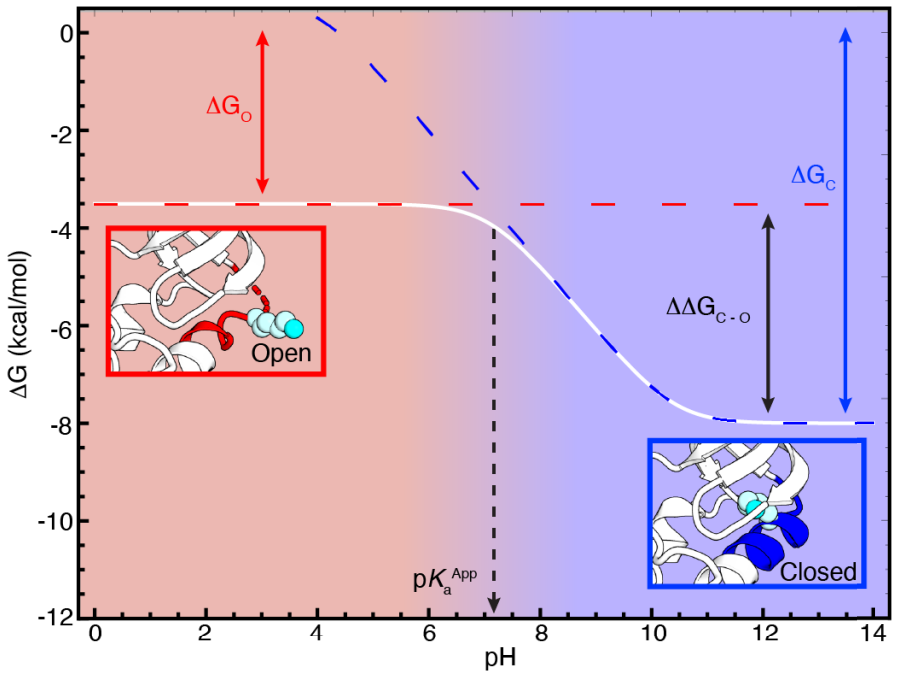

Fig. 3.3. Thermodynamic characterization of pH-dependent conformational changes necessary to hydrate buried Lys residues. Simulation showing the pH-dependent thermodynamic stability of a protein with buried Lys. The solid white line is simulated as the observed free energy difference between folded and unfolded states. ΔG_C is the free energy difference between the closed and unfolded states, which is pH-dependent below pH 10.4 (blue dashed line). ΔG_O is the free energy difference between the open and unfolded states (red dashed line). The energy gap between closed and open states is $\Delta\Delta G_{C-O}$, which is proportional to ΔpK_a. Closed and open states are represented by crystal structures of variants with K66 at pH 9 (closed, PDB ID: 3HZX) and pH 7 (open, PDB ID: 5CV5).

for Lys in water the stability of the closed states of each protein decrease in a pH-dependent manner (Fig. 3.3). Although the open states are less stable than the closed states at high pH, their stability is pH-independent because in these states the Lys residues are exposed to water and their pK_a values are normal. As the pH decreases from values where Lys is neutral in water to where it is charged, the free energy gap between the closed and the open states decreases; pK_a^{App} corresponds to the pH where this gap is zero.

3.3.3 Mapping the conformational landscape of SNase

At pH values below pK_a^{App} the open state is the dominant state in solution (Fig. 3.3), enabling structural characterization of the open state (with NMR spectroscopy) and measurement of its stability (with chemical denaturation experiments[2]). Because many variants showed similar patterns of CSPs or open state stabilities, it was possible to cluster variants based on similarity in the location or magnitude of their structural response. This allowed correlation between specific structural states populated by changes in pH with measured free energies (Fig. 3.4). The landscape shown two dimensionally in Fig. 3 shows that the region of SNase least destabilized by the ionization of buried Lys residues is the loop region between α-2 and α-3 (black in Fig. 3.4, black in Fig. 2A). Variants in which both this loop and the loop between β-4 and β-5 and in which the edge of β-3, β-4, and β-5 are perturbed (green in Fig. 3.4, red in Fig. 3.2A) are significantly less stable than the variants where only the loop between α-2 and α-3 appears to unfold. In both sets of variants ionization of Lys residues at three different positions triggered structural responses of similar magnitudes with similar open state stabilities. In contrast, the variants in which the turn between β-1 and β-2 or the loop between β-3 and α-1 are disrupted (orange and magenta in Fig.

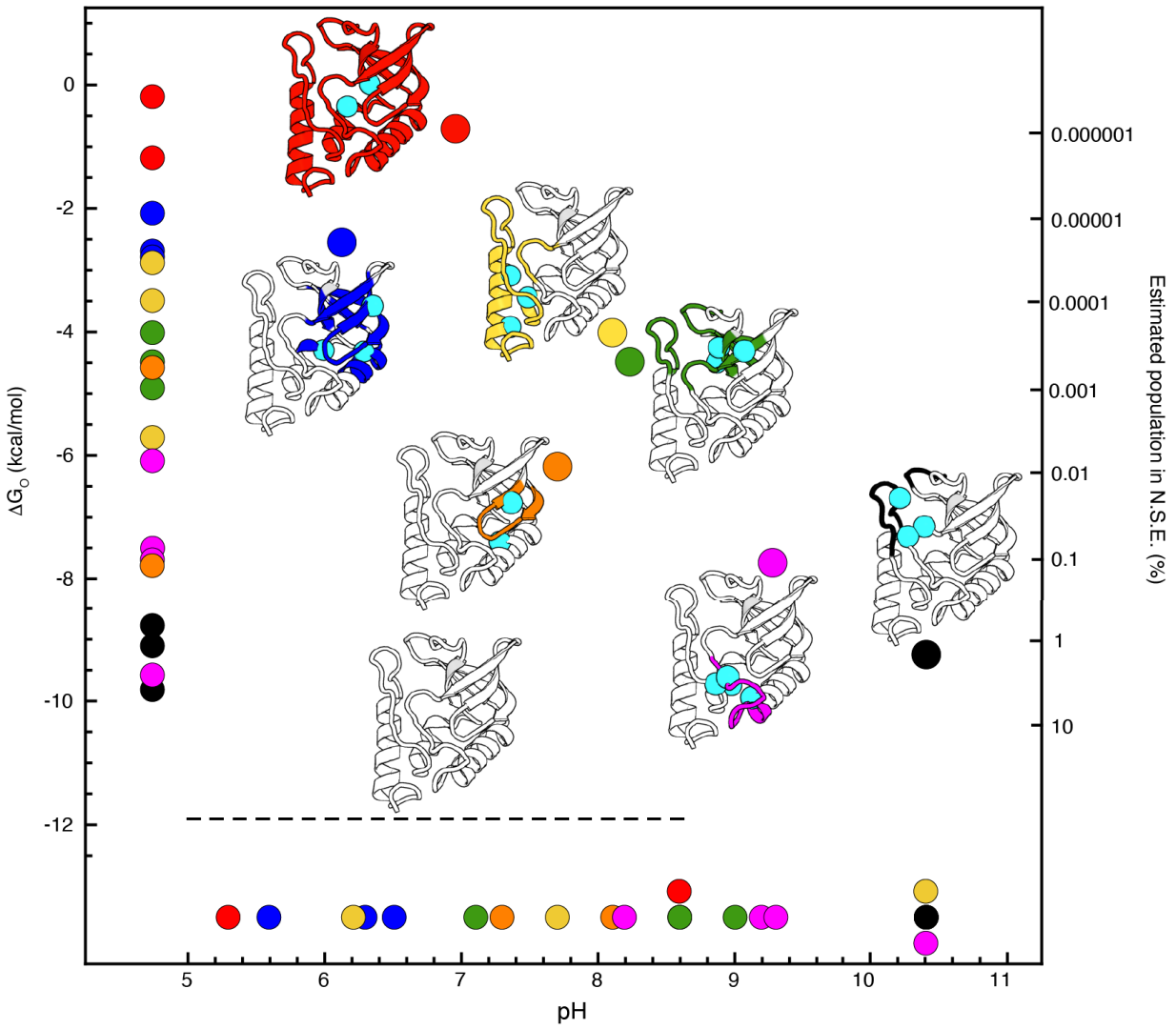

Fig. 3.4. Conformational landscape of staphylococcal nuclease modulated by changes in pH. Correlation of the structural perturbations induced by the ionization of an internal Lys—detected with NMR spectroscopy—with thermodynamic stability measured by chemical denaturation[2] for each protein under conditions of pH where the internal Lys is charged (small spheres, left axis). The pK_a values of the buried Lys residues are shown with small spheres on the bottom axis. The average open state stabilities and pK_a values are shown with large spheres. Structures in red represent variants with K92 and K100; blue with K25, K66 and K99; yellow with K104, K125 and K132; green with K34, K90 and K91; orange with K23 and K62; magenta with

K41, K58, K103, and K109; and black with K37, K38, and K118. Cyan spheres identify the sites of substitution with Lys. The ground state in this diagram (bottom structure, black dotted line) corresponds to the thermodynamic stability of the background protein. All open state stabilities were made in a pH regime where stability of the background protein is pH-independent (11.8 kcal/mol, pH 5-8.5). The difference in stabilities between background protein and open states at identical solution conditions was used to estimate the approximate population of each state in the native state ensemble of this protein (right axis). Each point on the vertical axis represents a single measurement from a set of measurements of the pH-dependence of the thermodynamic stabilities of the proteins; the standard error from fits and the pH-dependence of these thermodynamic stabilities have been reported and were between 0.1 - 0.2 kcal/mol for individual points[2] (Table S3.3). Errors for pK_a measurements were determined from fits of thermodynamic stability to a linkage formalism and range from 0.1-0.3 pH units[2].

3.4, Fig. 3.2B, C) exhibit greater variation in the character of the structural perturbations, and the associated stabilities of their open states.

The least stable open states of SNase are those in which the two main subdomains of the protein, the OB-fold (blue in Fig. 3.4) and the interface made between α-2, α-3 and the β-barrel (yellow in Fig. 3.4), are significantly disrupted. Presumably the ionization of Lys residues buried in these regions disrupts important packing interactions. In the most destabilized variant in the yellow cluster (L125K) an increase in random coil resonances suggests partial unfolding of α-3. The variants in the blue cluster all show evidence of significant partial unfolding (^{1}H-^{15}N HSQC spectra in Fig S3.2), with the β-barrel and α-2 affected to varying degrees; the C-terminus of α-1 is disrupted significantly in all of them. In all of these variants extensive disruption of the hydrophobic core of the OB-fold would be necessary to hydrate the buried Lys residue. Consistent with these observations, the blue regions have among the highest protection factors in hydrogen/deuterium exchange experiments[18]. Ionization of the buried Lys residue leads to global unfolding in only two variants (red in Fig. 3.4); in one case (N100K), fluorescence and circular dichroism spectroscopy provide evidence of some residual structure[2]. In five variants the structural responses could not be categorized either due to extensive resonance intensity loss in the native state or because the response is subtle and strictly local.

3.4 Discussion

The conformational substates that were identified in SNase by ionization of buried Lys residues are consistent with those identified previously with native state hydrogen deuterium exchange[18], suggesting that the partially unfolded states in Figure 3.1 are relevant as low lying substates of the native state ensemble of this protein. The energy differences between the

background protein and the open states allows for estimates of the population of these species in the native state ensemble of this protein. In most cases the substates could only exist at vanishingly small populations. These partial unfolding reactions leave the rest of the cooperative folding core of the protein unaffected, as observed by the high m-values observed in chemical denaturation experiments[2] (Table S3.3). Recent studies have shown that sparsely populated, partially unfolded states play important roles in tuning activity and allostery in well folded enzymes[19]. The results of this study are direct evidence that even well folded, stable proteins can access a wide range of partially unfolded states in response to relatively small changes in environmental conditions. The role of these partially disordered states in protein function has been underappreciated, given the difficulties in capturing these sparsely populated states with traditional structural biology methods. Both structural[20,21,22,23,24] and thermodynamic approaches[25,26] have previously been used to characterize sparsely populated states in the conformational ensemble. Despite advances, structural characterization of sparsely populated and partially disordered states remains challenging, and thermodynamic methods either lack atomic resolution structural detail or rely on perturbations that are not local. This present study resolves these limitations by demonstrating that the truly local nature of proton-binding can be used to populate high energy states as the dominant species in solution, which can be characterized using atomic resolution structural methods.

Our data with SNase show that in the physiological range of pH a single ionizable residue with an anomalous pK_a can affect the stability and conformation of a protein in a significant way[2,27,28,]. This suggests a mechanism whereby biological activity can be tuned through pH-dependent structural reorganization[29], allosteric regulation[30], and order-to-disorder transitions[31]. The role of buried residues is then to poise the conformational ensemble to respond to changes in pH, as well as other stimuli (e.g. photoillumination[32], redox reactions[5]) that introduce charges in

hydrophobic environments. With the large amount of structures available in the Protein Data Bank that can be mined with informatics approaches[7], as well as the development of chemical probes that can selectively modify ionizable residues with anomalous pK_a values[33,34], it should be possible to identify key residues and regions in proteins necessary for pH-sensing and regulation.

State-of-the-art electrostatics calculations, especially calculation of pK_a values, attempt to reproduce conformational reorganization coupled to the ionization of residues[35,36,37,38,39,40]. The data with SNase suggest two problems will have to be solved to increase the accuracy of these calculations. First, the calculations have to predict the structures of the open states, which will be challenging giving the highly local and disordered character of the reorganization (Fig. 1-2). Second, the Gibbs free energy differences between conformations will have to be calculated with high accuracy and precision, which will be difficult given the structural nuances described by Figure 3.

The conformation and stability of proteins has evolved to be sensitive to the chemical and physical properties of their environment[41]. This is why proteins are endowed with the remarkable ability to respond to environmental changes, to interpret these changes as biological signals, and to amplify and transduce them into essential biological reactions. Understanding the mechanisms whereby pH regulates these properties of proteins is of special interest owing to increased recognition of pH as an essential biological signal[42], and that pH dysregulation is observed in pathological states such as cancer[43]. This detailed structural and thermodynamic demonstration that a single ionizable residue can redistribute the conformational ensemble of a protein in response to small changes in pH in the physiological range has important implications for how proteins behave inside cells. Deeper understanding of the physical principles at play is necessary for improved mechanistic understanding not only of biological energy transduction[4] and regulatory

processes[42] but also for improved prediction of pathological mutations that can make proteins sensitive to pH dysregulation[44]. Buried ionizable groups will also be useful to introduce pH-sensitive dynamics in engineered proteins by allowing them to transition between distinct conformational states, a hallmark of biological function[8].

3.5 Methods

3.5.1 Sample preparation

Proteins was prepared and stored as previously described (Chapter 2.5.1). For NMR experiments, proteins were buffer exchanged into 25 mM buffer and 100 mM KCl in 90% H_2O/10% D_2O. Reported pH values are the pH of buffer stocks protein was buffer exchanged into. Potassium acetate was used between pH 4-5.5, MES between pH 5.8-6.6, HEPES between pH 7-8, TAPS between pH 8-9, and CHES between pH 9-10. No buffer effects have been previously observed in SNase, and spectra collected in this study overlay with spectra from previous studies collected in different buffers or water[10]. Sample concentrations varied between 0.7-2.5 mM, usually from 1.0-1.5 mM. Variants with pK_a values between 5-8.5 were studied at two pH values corresponding to 1.2-1.5 pH units above and below the apparent pK_a value of the internal Lys residue (pH values for samples given in Fig. S3-14). All other samples were made at pH 7.2 and compared to a sample of the background protein at pH 7.4. Deviation of sample pH values after data collection from original buffer pH was observed to be less than 0.1 pH units.

3.5.2 NMR spectroscopy

All NMR experiments were performed on Bruker Avance or Avance II 600 MHz spectrometers, equipped with cryogenic TCI probes. All data was collected at 25 °C. For backbone

resonance assignments, 2D ^{1}H-^{15}N HSQC and 3D HNCACB, CBCACONH, HNCO, HBHACONH, CCC(CO)NH, and H(CCCO)NH spectra were collected. All spectra were collected with acquisition times of 6-8 ms in ^{13}C (10-12 ms for HNCO), 8-10 ms in ^{1}H and 23-24 ms in ^{15}N in the indirect dimensions. Spectral widths in the indirect dimension for ^{13}C were 62, 16, and 70 ppm for HNCACB/CBCACONH, HNCO, and CC(CO)NH spectra, respectively. Spectral widths in the indirect dimension for ^{1}H were 8 and 11 ppm for HBHACONH and H(CCCO)NH experiments, respectively. The ^{15}N spectral width was set to 25.5 ppm for all experiments. Carriers were adjusted based on spectra, but were generally 116-119 in ^{15}N, 42 ppm for aliphatic ^{13}C spectra and 177-178 ppm for the HNCO experiment. ^{1}H carrier was consistently on the water frequency (4.76 ppm when accounting for salt concentration). 8 scans were collected per FID for each experiment except the CCCONH and H(CCCO)NH experiments, for which 16 scans per FID were required for adequate signal to noise, and the HNCO, for which only 4 scans per FID were necessary. All 3D experiments were collected using non-uniform sampling procedures with schedules generated using the Poisson Gap sampler developed by the Wagner lab[12]. Experiments were sampled at 15%, except for the CCCONH and (H)CCCONH, which were sampled at 17.5%. Spectra were reconstructed using NESTA-NMR[45]. Spectra were processed using NMRPipe[46] and analyzed using NMRFAM-Sparky[47].

The ^{1}H-^{15}N HSQC, HNCACB, CBCACONH, and HNCO were manually peak picked. The automated assignment protocol of PINE-SPARKY was used to rapidly assign manually picked peaks[13,48]. In general, accuracy of this protocol was 90-100%, except in cases where broadening broke inter-residue connectivities, in which accuracy droppee to 50-60%. Spectra were manually assigned to correct misassigned resonances from PINE, and to assign the ^{1}Hα resonances using the HBHACONH and H(CCCO)NH spectra. Side chain patterns in the CC(CO)NH and

H(CCCO)NH spectra were used to confirm backbone assignments from CBCACONH and HNCACB spectra, though no side chains past the Cβ carbon were assigned for this study. All spectra underwent at least one more round of manual assignment to prevent misassignment of resonances.

For the 12 variants for which data was collected at two pH values, the CSPs were calculated by subtracting the low pH data from the high pH data set. The CSPs were evaluated based on the CSPs observed between pH 9.4 and 4.6 in the background protein. This range of pH covers the entire range of pH values used in this study, and accounts for the titration of multiple residues (Fig. S3). The Cβ resonances of Asp-21, His-121, and His-8 were excluded from the analysis of Cβ CSPs as their large pH dependent CSPs reflect a change of charge state of an adjacent moiety, and significantly skew the standard deviation. Residues in the variants that showed at least two atoms having CSPs more than two standard deviations from the mean of the background protein CSPs were marked as having moderate CSPs (light blue), while residues with at least two atoms having CSPs more than three standard deviations from the mean of the background protein were marked as having large CSPs (blue). This cutoff was chosen because the only residues that met the criteria for moderate to large CSPs in the background protein were titrating residues (His-8, Asp-21, His-121) and one residue that is hydrogen bonded to the side chain of Asp-21 (Thr-41). Residues were identified as having intensity loss if the resonances for 4 out of 6 of the backbone nuclei are broadened beyond identification below the apparent pK_a value of the buried Lys. This accounts for the fact that very broad peaks (indicative of intermediate exchange) in the ^{1}H-^{15}N HSQC were not assignable using triple resonance experiments. If this was only observed at pH values above the apparent pK_a value of the buried Lys residue, this was assumed to be due to base catalyzed exchange at the high pH values necessary to deprotonate the buried Lys residue. Residues for

which less than 3 chemical shifts could be measured for different reasons (e.g., multiple resonance overlap) were excluded from this cutoff. Mapping structural changes using weighted chemical shifts did not change the distribution of CSPs shown in Fig. 3.1. For detailed descriptions of NMR experiments and assignment procedure, see general methods.

3.5.3 Structural studies of alternative states

Crystal structure of background protein and variants with neutral Lys were solved in previous studies[10]. Crystal structures of alternative states (A109R, V23E, and V66K) were solved in previous studies[49,50,11]. The V23E and V66K variants required secondary mutations to crystallize (D21N, T33V, T41I, S58A for V23E, K64G and E67G for V66K).

Structural information was obtained through analysis of the 6 backbone chemical shifts measured in this study. TALOS-N was used to for torsion angle and secondary structure prediction[51]. Predicted order parameters (RCI S^2) were generated using the RCI algorithm[16], to assess changes in local dynamics.

3.5.4 Simulation of thermodynamic stabilities

The thermodynamic stability of a protein with a buried Lys residue was simulated as[2]:

$$\Delta G = \Delta G_C + RT\left(\ln\left(\frac{1 + e^{-2.3(pH-pKa,O)}}{1 + e^{-2.3(pH-pKa,C)}}\right)\right)$$

Because structural perturbations caused by ionization of internal Lys residues do not cause significant changes in Trp fluorescence or m-values[2], the closed and open states observed in this study cannot be distinguished in a chemical denaturation experiment. The ΔG should reproduce the thermodynamic stability of a protein with a single buried residue with a shifted pK_a value, as observed in a chemical denaturation experiment. For the purpose of the simulation (solid gray

curve in Figure 3.3), the initial stability (set as the maximum stability of the closed state, ΔG_C) was set to 8 kcal/mol. The terms $pK_{a,O}$ and $pK_{a,C}$ refer to the fully unfolded state pK_a value of the Lys residue (set to 10.2; a slightly lower value was used than the pK_a value of Lys in water (10.4) for aesthetic purposes in Figure 3.3) and the apparent pK_a value of its titration when the protein is folded (set to 7.1), respectively. For the simulation of the closed state (dashed blue line in Figure 3.3), the maximum stability was set to 8 kcal/mol, and the terms $pK_{a,C}$ and $pK_{a,O}$ refer to the pKa values in closed (set to 4 in this simulation, much lower than pK_a^{App}) and open (set to 10.2) states. Since the pK_a value of the Lys residue in the open state is similar to that of the Lys in water, its stability is not pH dependent and was set to reflect the loss in the stability of the closed state necessary to shift the pK_a value of the Lys from the unfolded state to its pK_a^{App}:

$$\Delta G_O = \Delta G_C + 1.36(pK_{a,O} - pK_a^{App})$$

3.5.5 Measurement and analysis of thermodynamic parameters

The values for ΔG_O were obtained from a previous study on the pH-dependence of the folding stability of these variants[2], measurements and errors of which are available in Extended Data Table 3. These measurements were made using chemical denaturation (GdnHCl). In general, the errors for stability measurements was between 0.1-0.2 kcal/mol. For the ΔG_O term, $\Delta G°_{H2O}$ values were chosen that were 0.5-1 pH unit below the apparent pK_a value of the Lys residue, unless the variant showed a second region of pH dependence in its thermodynamic stability after the Lys is protonated (indicative of the a pK_a shift of a surface residue[2]). In these cases, a $\Delta G°_{H2O}$ value 0-0.5 pH units below the apparent pK_a value were chosen. To estimate the population of partially folded state in the native state ensemble, the energy difference between the open state and the

background at equivalent solution conditions was used to calculate a relative population of open versus native protein, using the following relation:

$$\Delta\Delta G^\circ = \Delta G_O{}^\circ - \Delta G_{BG}{}^\circ = -RT\left(\ln\left(\frac{[O]}{[U]}\right) - \ln\left(\frac{[BG]}{[U]}\right)\right) = -RT\ln\frac{\frac{[O]}{[U]}}{\frac{[BG]}{[U]}} = -RT\ln\frac{[O]}{[BG]}$$

It is important to note that this analysis is solely meant to provide an estimate of populations, as the substitution itself could lead to favorable or unfavorable interactions that influence the energy differences between states. These are not likely to significantly shift the order of magnitude of the populations, as the energy landscape described by charge burial agrees well with that described by hydrogen deuterium exchange studies. It is also important to note that in many cases, different Lys residues that disrupt similar structural elements of the protein are not likely to share similar environments in their open states, as in many cases they are on opposing elements of secondary structure. This is true for variants in the black, green, and blue clusters in Figure 3, as well as to a lesser extent the yellow cluster (i.e., variants with K34 and K90, K66 and K25, K66 and K99, or K104 and K125); nevertheless, these cases have similar open state stabilities (within 1 kcal/mol of each other), again suggesting that the populations obtained from the thermodynamic analysis are indeed relevant to the native state ensemble of these proteins.

3.6 References

1. Holliday, G. L., Mitchell, J. B. O. & Thornton, J. M. Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.* **390**, 560–577 (2009).

2. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno E., B. Large shifts in pK_a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci.* **108**, 5260–5265 (2011).

3. Whitten, S. T., García-Moreno E., B. & Hilser, V. J. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci.* **102**, 4282–4287 (2005).

4. Bai, C. & Warshel, A. Revisiting the protomotive vectorial motion of F_o-ATPase. *Proc. Natl. Acad. Sci.* **116**, 19484–19489 (2019).

5. Pisliakov, A. V., Sharma, P. K., Chu, Z. T., Haranczyk, M. & Warshel, A. Electrostatic basis for the unidirectionality of the primary proton transfer in cytochrome c oxidase. *Proc. Natl. Acad. Sci.* **105**, 7726–7731 (2008).

6. Subramaniam, S. & Henderson, R. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature* **406**, 653 (2000).

7. Isom, D. G. *et al.* Protons as second messenger regulators of G protein signaling. *Mol. Cell* **51**, 531–538 (2013).

8. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).

9. Tanford, C. Ionization-linked changes in protein conformation. I. Theory. *J. Am. Chem. Soc.* **83**, 1628–1634 (1961).

10. Chimenti, M. S. *et al.* Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein. *Structure* **20**, 1071–1085 (2012).

11. Peck, M. T. *et al.* Local backbone flexibility as a determinant of the apparent pK_a Values of buried ionizable groups in proteins. *Biochemistry* **56**, 5338–5346 (2017).

12. Hyberts, S. G., Takeuchi, K. & Wagner, G. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J. Am. Chem. Soc.* **132**, 2145–2147 (2010).

13. Lee, W., Westler, W. M., Bahrami, A., Eghbalnia, H. R. & Markley, J. L. PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics* **25**, 2085–2087 (2009).

14. Wishart, D. S. & Case, D. A. Use of chemical shifts in macromolecular structure determination. *Method. Enzymol.* **338**, 3–34 (2002).

15. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1814**, 942–968 (2011).

16. Berjanskii, M. V. & Wishart, D. S. A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971 (2005).

17. First, J. T., Novelli, E. T. & Webb, L. J. Beyond pK_a: Experiments and simulations of nitrile vibrational probes in staphylococcal nuclease show the importance of local interactions. *J. Phys. Chem. B* **124,** 3387-3399 (2020).

18. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **21**, 996–1005 (2012).

19. Saavedra, H. G., Wrabl, J. O., Anderson, J. A., Li, J. & Hilser, V. J. Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324–328 (2018).

20. Keedy, D. A. *et al.* Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **4**, e07574 (2015).

21. Kerns, S. J. *et al.* The energy landscape of adenylate kinase during catalysis. *Nat. Struct. Mol. Biol.* **22**, 124–131 (2015).

22. Horowitz, S. *et al.* Visualizing chaperone-assisted protein folding. *Nat. Struct. Mol. Biol.* **23**, 691–697 (2016).

23. Frank, J. New opportunities created by single-particle Cryo-EM: The mapping of conformational space. *Biochemistry* **57**, 888–888 (2018).

24. Alderson, T. R. & Kay, L. E. Unveiling invisible protein states with NMR spectroscopy. *Curr. Opin. Struct. Biol.* **60**, 39–49 (2020).

25. Mello, C. C. & Barrick, D. An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci.* **101**, 14102–14107 (2004).

26. Fossat, M. J. *et al.* High-resolution mapping of a repeat protein folding free energy landscape. *Biophys. J.* **111**, 2368–2376 (2016).

27. Zheng, Z. & Sosnick, T. R. Protein vivisection reveals elusive intermediates in folding. *J. Mol. Biol.* **397**, 777–788 (2010).

28. Davis, C. M. & Dyer, R. B. The role of electrostatic interactions in folding of β-proteins. *J. Am. Chem. Soc.* **138**, 1456–1464 (2016).

29. Gayen, A., Leninger, M. & Traaseth, N. J. Protonation of a glutamate residue modulates the dynamics of the drug transporter EmrE. *Nat. Chem. Biol.* **12**, 141–145 (2016).

30. Bongard, J. *et al.* Identification of noncatalytic lysine residues from allosteric circuits via covalent probes. *ACS Chem. Biol.* **13**, 1307–1312 (2018).

31. Salmon, L. *et al.* The mechanism of HdeA unfolding and chaperone activation. *J. Mol. Biol.* **430**, 33–40 (2018).

32. Xie, A. *et al.* Formation of a new buried charge drives a large-amplitude protein quake in photoreceptor activation. *Biochemistry* **40**, 1510–1517 (2001).

33. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).

34. Matos, M. J. *et al.* Chemo- and regioselective lysine modification on native proteins. *J. Am. Chem. Soc.* **140**, 4004–4017 (2018).

35. Georgescu, R. E., Alexov, E. G. & Gunner, M. R. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophys. J.* **83**, 1731–1748 (2002).

36. Kato, M. & Warshel, A. Using a charging coordinate in studies of ionization induced partial unfolding. *J. Phys. Chem. B* **110**, 11566–11570 (2006).

37. Shi, C., Wallace, J. A. & Shen, J. K. Thermodynamic coupling of protonation and conformational equilibria in proteins: Theory and simulation. *Biophys. J.* **102**, 1590–1597 (2012).

38. Goh, G. B., Laricheva, E. N. & Brooks, C. L. Uncovering pH-dependent transient states of proteins with buried ionizable residues. *J. Am. Chem. Soc.* **136**, 8496–8499 (2014).

39. Liu, J., Swails, J., Zhang, J. Z. H., He, X. & Roitberg, A. E. A coupled ionization-conformational equilibrium Is required to understand the properties of ionizable residues in the hydrophobic interior of staphylococcal nuclease. *J. Am. Chem. Soc.* **140**, 1639–1648 (2018).

40. Fossat, M. J. & Pappu, R. V. q-canonical Monte Carlo sampling for modeling the linkage between charge regulation and conformational equilibria of peptides. *J. Phys. Chem. B* **123**, 6952–6967 (2019).

41. Guinn, E. J., Jagannathan, B. & Marqusee, S. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat. Commun.* **6**, 1–8 (2015).

42. Casey, J. R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **11**, 50–61 (2010).

43. White, K. A., Grillo-Hill, B. K. & Barber, D. L. Cancer cell behaviors mediated by dysregulated pH dynamics at a glance. *J. Cell Sci.* **130**, 663–669 (2017).

44. White, K. A. *et al.* Cancer-associated arginine-to-histidine mutations confer a gain in pH sensing to mutant proteins. *Sci Signal* **10**, eaam9931 (2017).

45. Sun, S., Gill, M., Li, Y., Huang, M. & Byrd, R. A. Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms. *J. Biomol. NMR* **62**, 105–117 (2015).

46. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

47. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

48. Lee, W. & Markley, J. L. PINE-SPARKY.2 for automated NMR-based protein structure research. *Bioinforma. Oxf. Engl.* **34**, 1586–1588 (2018).

49. Harms, M. J., Schlessman, J. L., Sue, G. R. & García-Moreno E., B. Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci.* **108**, 18954–18959 (2011).

50. Robinson, A. C., Majumdar, A., Schlessman, J. L. & García-Moreno E, B. Charges in hydrophobic environments: A strategy for identifying alternative states in proteins. *Biochemistry* **56**, 212–218 (2017).

51. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).

52. Richman, D. E., Majumdar, A. & García-Moreno E., B. pH dependence of conformational fluctuations of the protein backbone. *Proteins* **82**, 3132–3143 (2014).

53. Carra, J. H. & Privalov, P. L. Energetics of denaturation and m-values of staphylococcal nuclease mutants. *Biochemistry* **34**, 2034–2041 (1995).

54. Shortle, D. Staphylococcal nuclease: A showcase of m-value effects. *Adv. Protein Chem.* **46**, 217–247 (1995).

3.7 Supporting Information

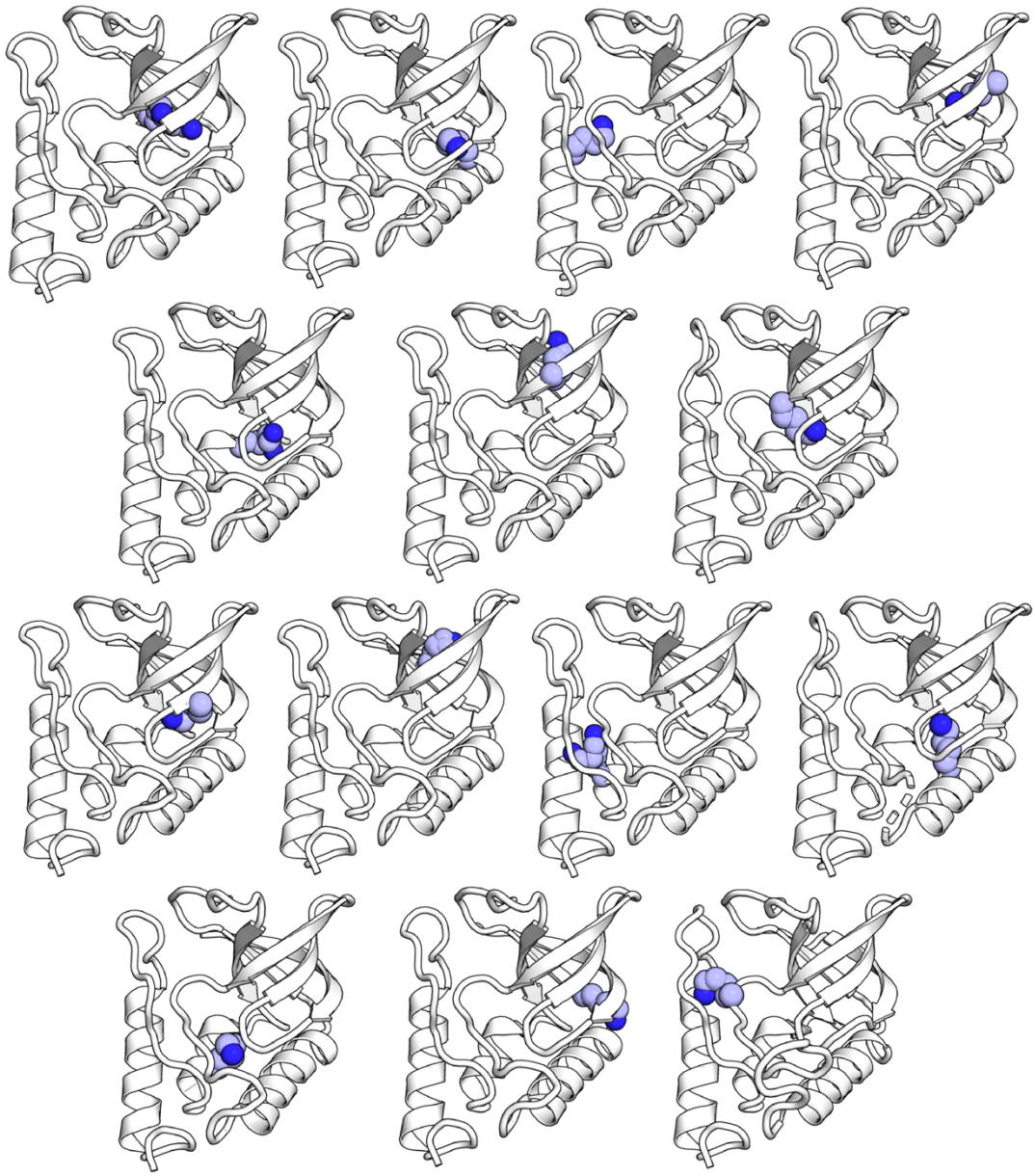

Fig. S3.1. Crystal structures of 14 variants of SNase with buried Lys residues. From top left to bottom right (PDB ID in parentheses): I92K (1TT2), V66K (3HZX), L125K (3C1E), L25K (3ERQ), V99K (4HMI), F34K (3ITP), L36K (3EJI), V23K (3QOJ), V74K (3RUZ), V104K (3C1F), T62K (3DMU), L103K (3E5S), I72K (2RBM), and L38K (2RKS). All crystal structures were previously published[10].

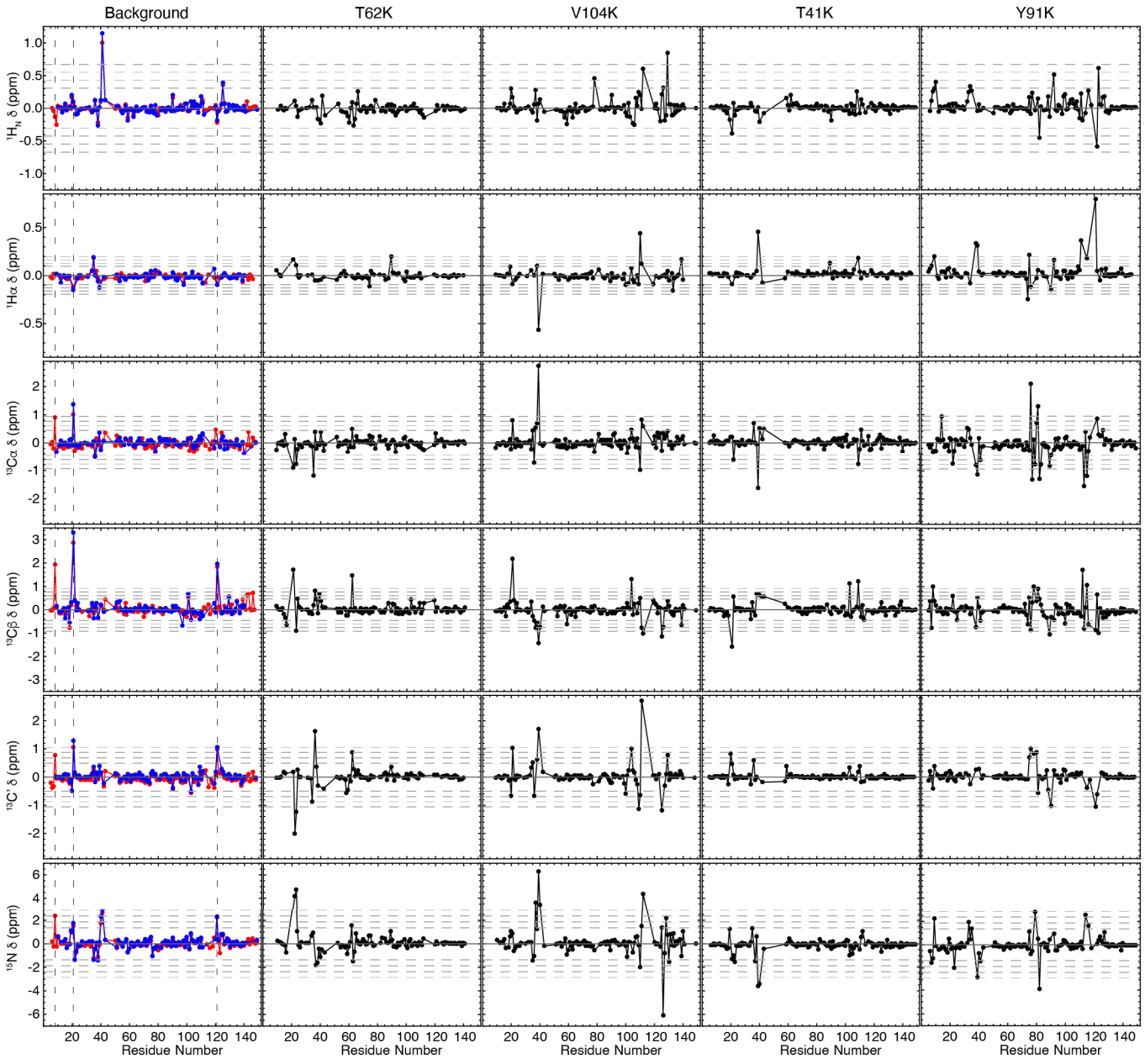

Fig. S3.2. Representative backbone chemical shift perturbations. Background CSPs are shown between pH 9.4 - 4.6 (blue) and pH 7.4 - 4.6 (red). CSPs shown for variants with K62 and K104 are between pH 9.6 – 6.6 and pH 9.2 – 6.4, respectively. Vertical dashed lines denote residues whose ionization state changes between pH 7.4 and 4.6 (H8, D21, H121). CSPs shown for variants with K41 and K91 are between the background protein (pH 7.4) and variants (pH

7.2). Horizontal dashed lines denote 2-5 standard deviations from mean of background protein CSPs (between pH 9.4 – 4.6). Full CSPs are included in supplementary information.

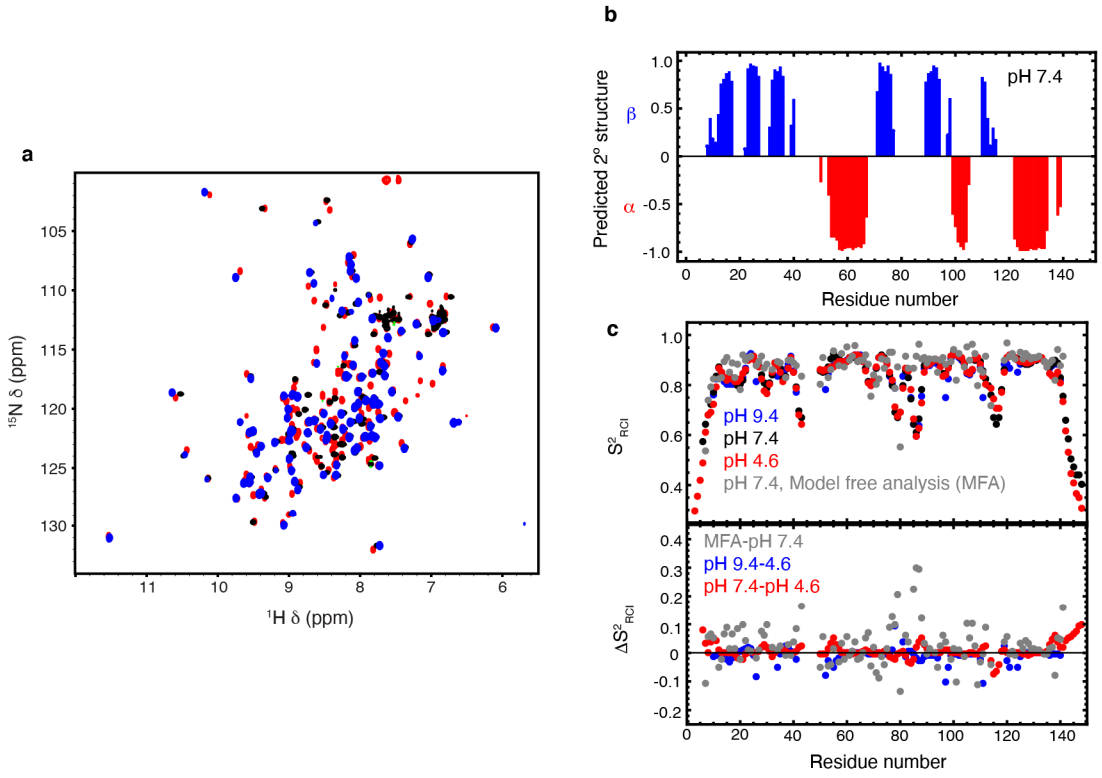

Fig. S3.3. NMR characterization of SNase variant Δ+PHS (background protein). (a) ^{1}H-^{15}N HSQC spectra of background protein at pH 9.4 (blue), 7.4 (black), and 4.6 (red). **(b)** TALOS+ secondary structure prediction using backbone chemical shifts. **(c) Top:** RCI predicted order parameters (S^2), from backbone chemical shifts. S^2 values from ^{15}N relaxation model free analysis[52] shown as comparison. **Bottom:** Comparison of RCI S^2 values; no evidence of pH-dependent local disorder is observed in the background protein between pH 4.6 to 9.4. In general, the predicted order parameters are within 0.1 of the values determined by model free analysis, with the exception of residues in the loop region comprising residues 78-88, where the RCI analysis overestimates the flexibility of this region of the protein.

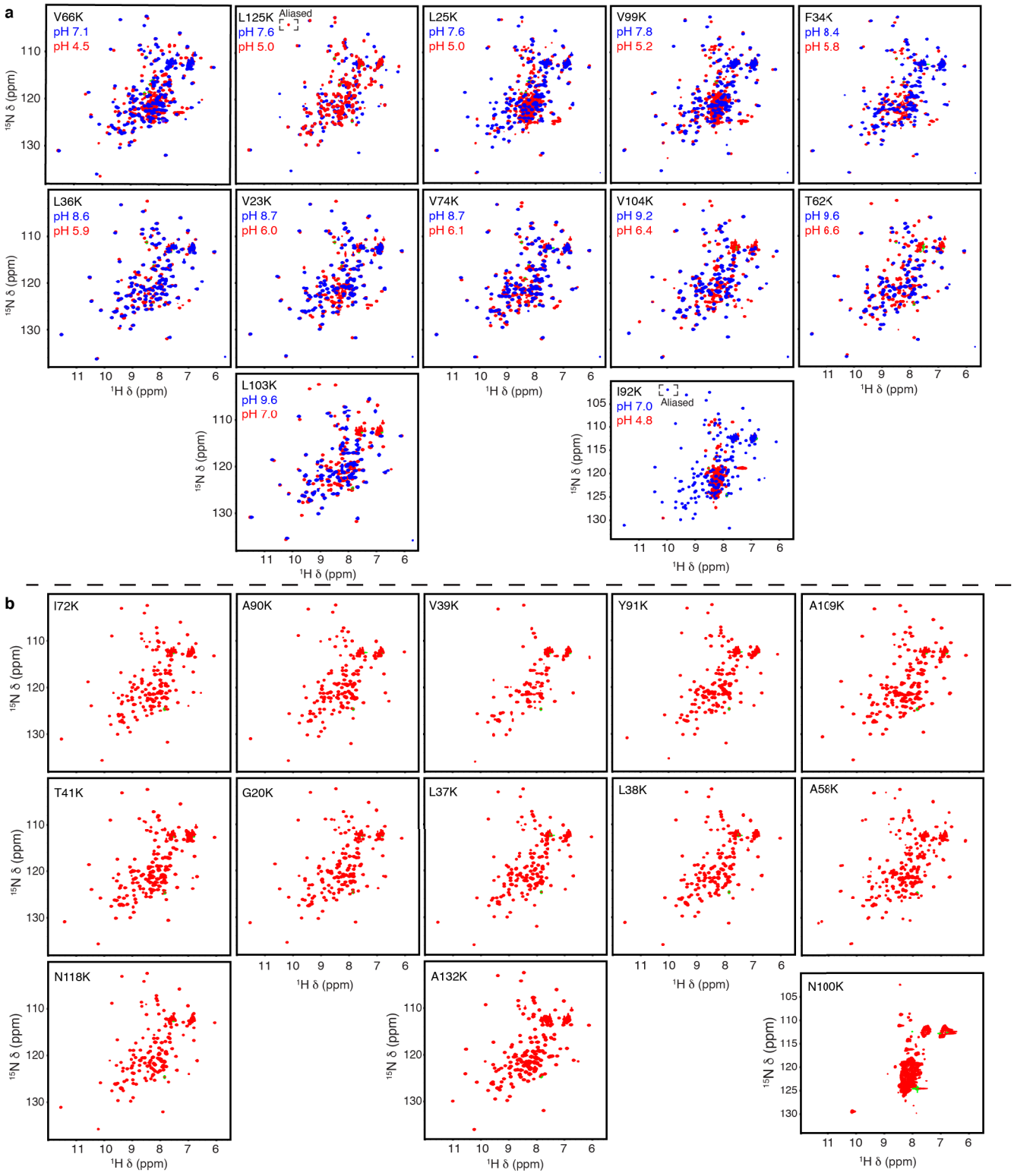

Fig. S3.4. ^{1}H-^{15}N HSQC spectra of internal Lys variants of SNase. (a) pH-dependent spectra of 12 SNase variants with Lys pK_a values below 8.5. **(b)** Spectra of SNase variants with Lys pK_a

values above 8.5, at pH 7.2. Variants are organized by increasing pK_a^{App}, except for those undergoing global unfolding (I92K and N100K), which are shown as final panels in **(a)** and **(b)**.

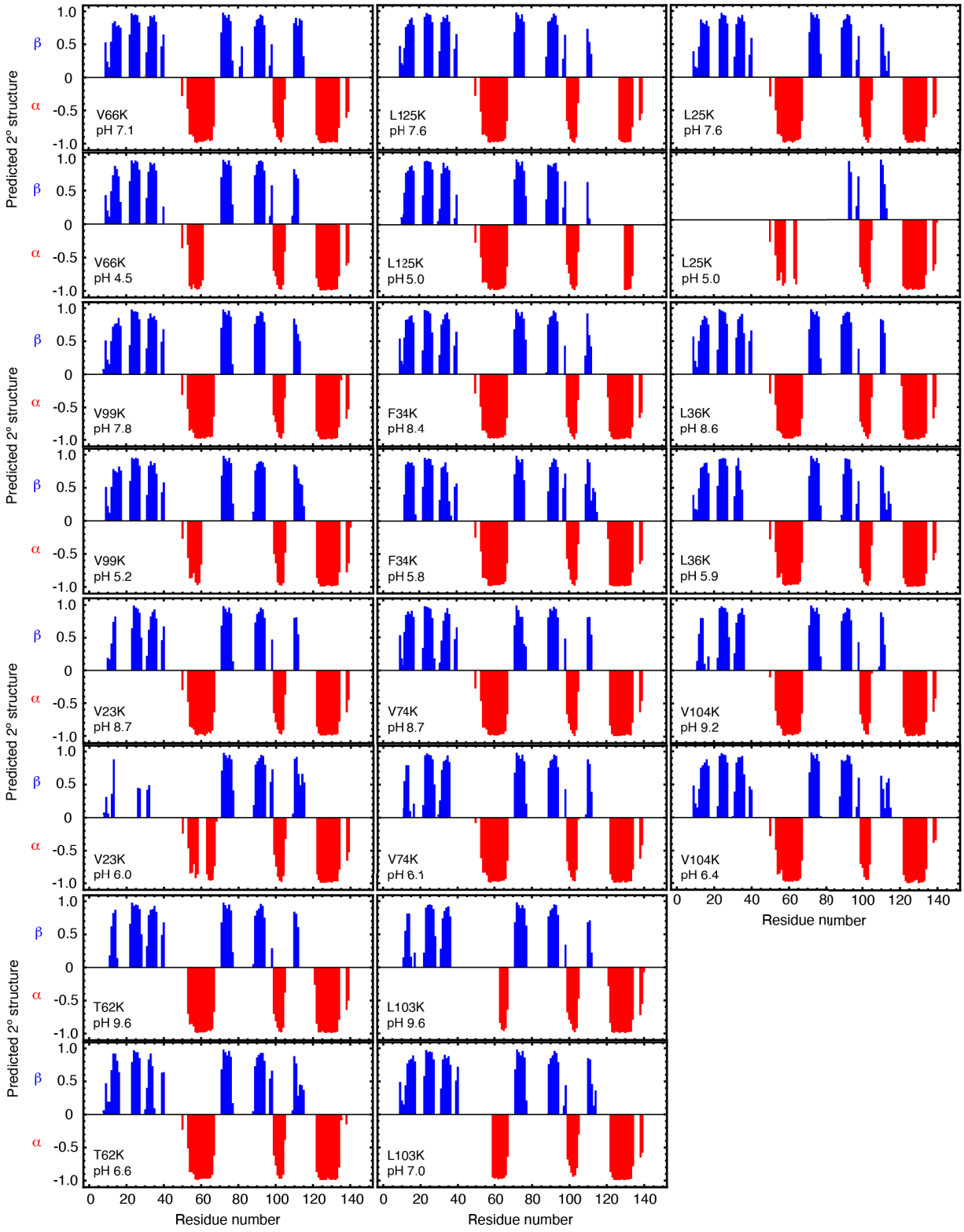

Fig. S3.5. TALOS+ secondary structure predictions of internal Lys variants of SNase with pK_a values below 8.5. Predictions made above and below the pK_a^{App} of each buried Lys residue.

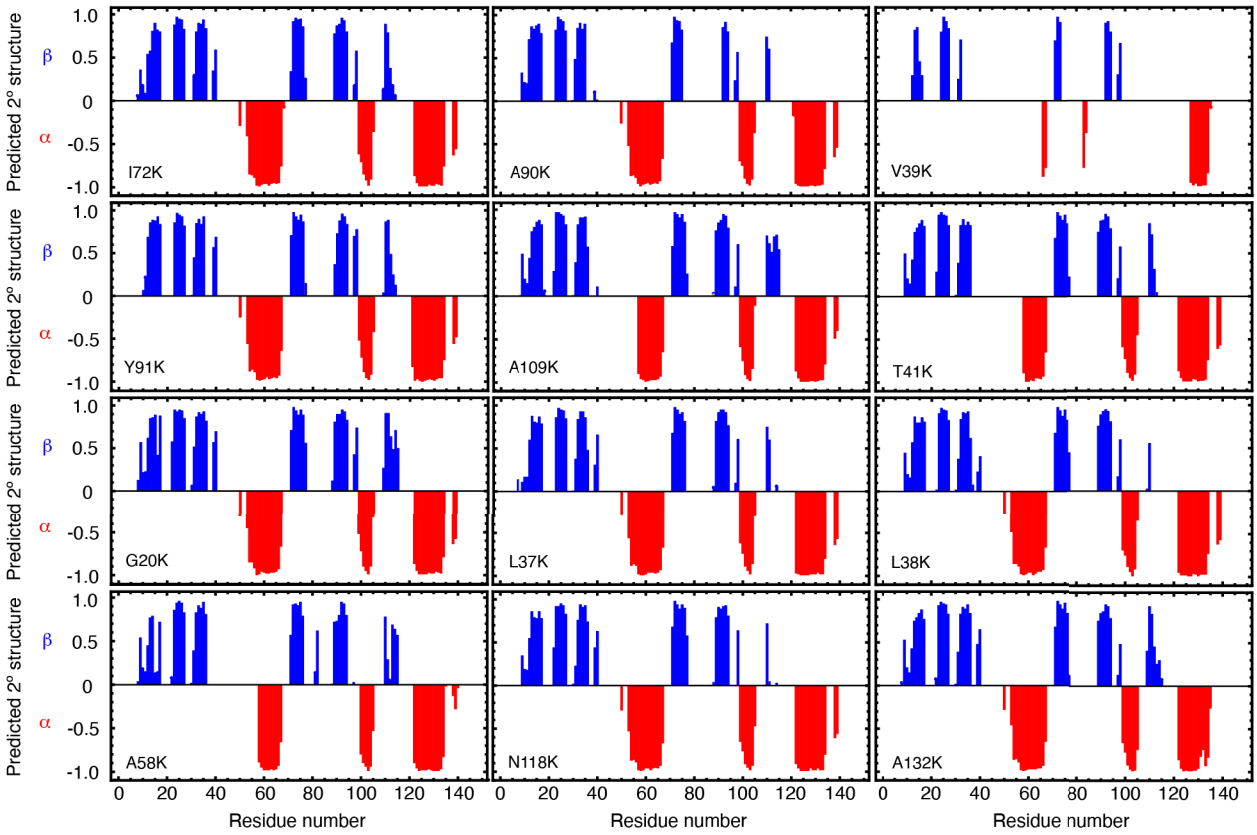

Fig. S3.6. TALOS+ secondary structure predictions of internal Lys variants of SNase with pK_a values above 8.5. All chemical shift data were collected at pH 7.2.

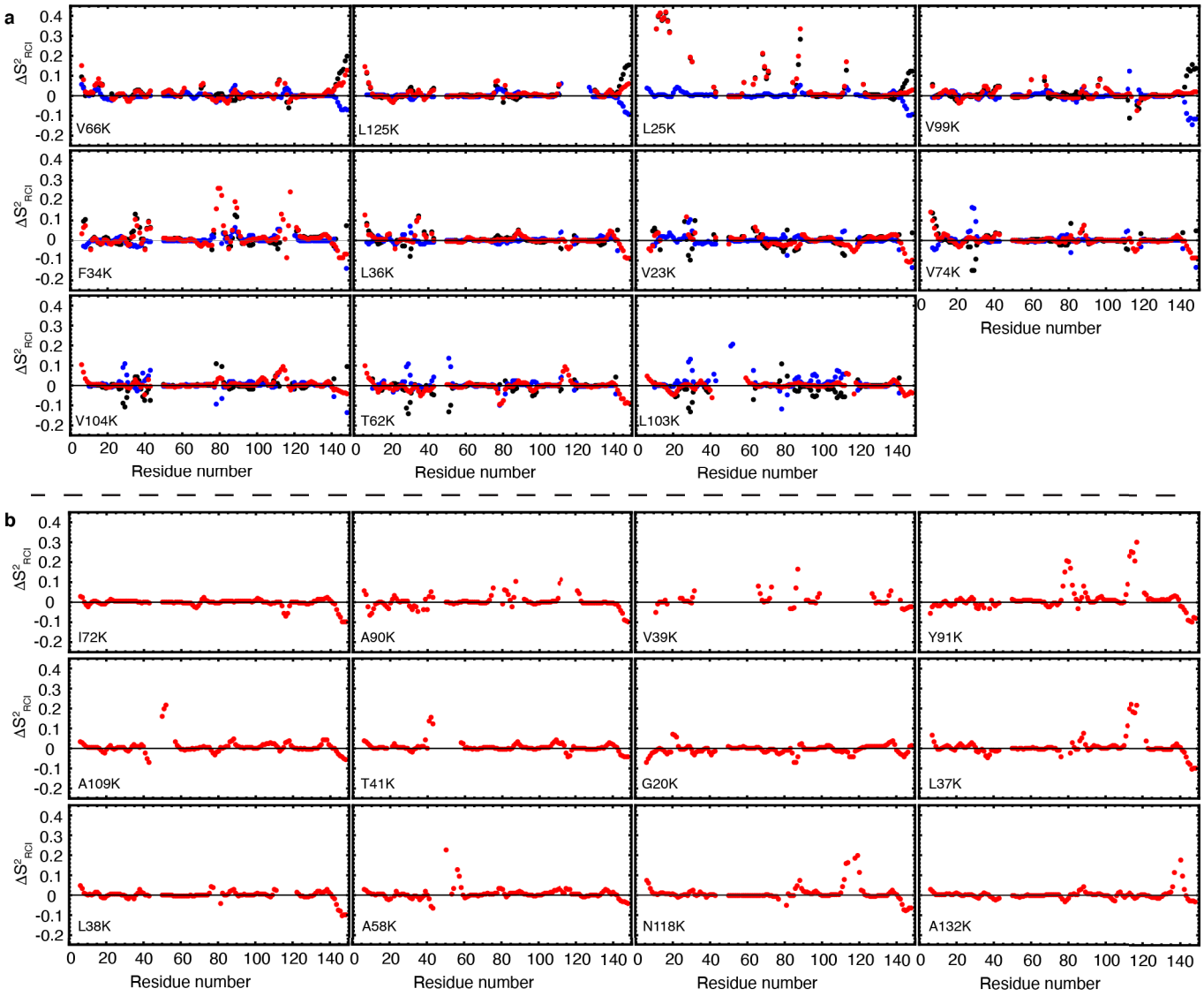

Fig. S3.7. RCI analysis of internal Lys variants of SNase. (a) pH dependent differences in predicted RCI S^2 values between high (closed) and low (open) pH in black, between high pH (closed state) and background protein in blue, and low pH (open state) and background protein in red. Only variants with Lys pK_a values below 8.5 are shown. One set of pH values (pH 7.4) was used for background protein RCI S^2 values since differences between the RCI S^2 values between pH 9.4 and 4.6 are small. **(b)** Comparison between background protein (pH 7.4) and variants (pH 7.2) with Lys pK_a values above 8.5.

	$^{1}H_{N}$	$^{1}H\alpha$	$^{13}C\alpha$	$^{13}C\beta$	^{13}CO	^{15}N
Average	0.063	0.028	0.11	0.15	0.13	0.38
Std. Dev.	0.122	0.033	0.16	0.15	0.19	0.50
1 std. dev. from mean	0.185	0.061	0.28	0.30	0.31	0.87
2 std. dev. from mean	0.306	0.094	0.44	0.46	0.50	1.37
3 std. dev. from mean	0.428	0.128	0.61	0.61	0.69	1.87
4 std. dev. from mean	0.550	0.161	0.77	0.76	0.87	2.36
5 std. dev. from mean	0.672	0.194	0.93	0.91	1.06	2.86

Table S3.1. Average and standard deviation of pH-dependent chemical shift perturbations in background protein, between pH 9.4 and 4.6. Calculated from absolute value of chemical shift perturbations.

Lys substitution (pK_a value)	Residues with large chemical shift changes.	Residues with very large chemical shift changes	Residues with loss of native state resonance intensity.
K92 (5.3)			1-149
K66 (5.6)	14,25,94,106	8*,17,22,23,41,72,121*	15,18-21,62-69
K125 (6.2)	89	8*,21*,38,41,74,75	78-80,112-129
K25 (6.3)		11-18	7-10,19-29,31-41,60-62,66,67,72-83,89-92,(115,116),117-121
K99 (6.5)	22,34,74,94,97,100	8*,21*,41*,121*	61-66,98,99,102, (115,116)
K34 (7.1)	26,34,37,75-77, 90-91	8*,21*,22,33,38,74,121*	10,35,78-81,114, (115,116),117-120
K36 (7.2)		8*	20,21,(28,29),36,37,40,41,(113-118)
K23 (7.3)		8*	14-25,(28,29),33-41, 59-62,66,(80,113-118)
K74 (7.4)	9,74,75,92	8*,10,21*,41*	(28,29,80,113-118)
K104 (7.7)	20,35,36,38	21*,37,39,40,104,110-112,125,126,129	(7,8,28,29,79,80,113-118)
K62 (8.1)		21*-23,36,62	(17),18,(19,20,28,29,50,79,80,96,113-118)
K103 (8.2)	66	52	(7,8,28,29),40-51,53-62,(79,80,96),109,(113-118,123)
K72 (8.6)	14,70,73	66,67	72
K90 (8.6)	7,23,73,88,121	8.9,22,33-35,38,39, 74,75,80,92	36,37,41,76-79,89-91,113-120
K100 (8.6)			1-149
K39 (9.0)			7-10,18-23,33-65,74-83,88-91,100-105,107-126,138-141
K91 (9.0)	7,22,41,75,116	9,38,39,74,76,78,80-82,89,90,92,113,115, 122-123	36,37,118-120
K109 (9.2)	20,39,58,59,110-111	21,40,41,43-52,57,108	53-56,109
K41 (9.3)	21,22,36,40	20,39,109	19,41,50-57
K20 (10.4)	36,39,41,43,111	19,21,40,	20
K37 (10.4)	39,74,76,114-116	8,38,81,82,89,90,113, 121	37,77-80,117-120
K38 (10.4)	76,82		38,78-80,112-121
K58 (10.4)	19,43,113	40-42,109	***17,20-23,35-39***, 51-58, ***60-62,64-68,72,75, 78,88,91,92, 95,99-108, 110-112,114,118121, 124,129,131,133, 137-141***
K118 (10.4)		38,89,113,119	78-80,115,116, 118,120
K132 (10.4)	106,111,125,134, 137	104,105,131	132,138-139

Table S3.2. Summary of pH-dependent chemical shift behaviors of 25 variants of SNase with buried Lys. Residues denoted with stars show large pH-dependent chemical shift changes in the background protein. Residues in parentheses display native state intensity loss at high pH

in the background protein due to base-catalyzed exchange. Residues in bold italic have observable native state resonances, but are observable as two distinct resonances in slow exchange (variant with K58).

Lys substitution	ΔG_O	$\Delta\Delta G_{C-O}$	*m*-value
K92 (5.3)	0.2 (4.83, 0.1)	6.9 (0.3)	7.5 (0.2)
K100 (8.6)	1.2 (7.93, 0.1)	2.4 (0.3)	4.4 (0.1)
K25 (6.3)	2.7 (5.45, 0.1)	5.6 (0.3)	5.4 (0.1)
K66 (5.7)	2.8 (4.88)	6.5 (0.3)	6.5 (0.3)
K99 (6.5)	2.1 (5.99, 0.1)	5.3 (0.3)	6.1 (0.1)
K104 (7.7)	3.5 (7.00, 0.1)	3.7 (0.3)	6.2 (0.1)
K125 (6.2)	2.9 (5.45, 0.1)	5.7 (0.3)	5.0 (0.1)
K132 (10.4)	5.7 (7.00, 0.2)	0	5.4 (0.1)
K34 (7.1)	4.0 (6.48, 0.1)	4.5 (0.3)	5.8 (0.1)
K90 (8.6)	4.5 (7.93, 0.1)	2.4 (0.3)	5.7 (0.1)
K91 (9.0)	4.9 (8.49, 0.2)	1.9 (0.3)	5.0 (0.2)
K23 (7.3)	4.6 (7.00, 0.1)	4.2 (0.3)	6.0 (0.1)
K62 (8.1)	7.8 (7.53, 0.1)	3.1 (0.3)	5.4 (0.1)
K41 (9.3)	9.6 (8.46, 0.2)	1.5 (0.3)	5.2 (0.1)
K58 (10.4)	7.7 (7.03, 0.1)	0	5.7 (0.1)
K103 (8.2)	6.1 (7.01, 0.1)	3.0 (0.3)	5.8 (0.1)
K109 (9.2)	7.5 (8.45, 0.1)	1.6 (0.3)	5.3 (0.1)
K37 (10.4)	8.8 (7.00, 0.1)	0	4.7 (0.1)
K38 (10.4)	9.1 (7.01, 0.2)	0	5.0 (0.1)
K118 (10.4)	9.8 (7.00, 0.6)	0	5.0 (0.3)
K20 (10.4)	9.2 (6.98, 0.1)	0	5.4 (0.1)
K36 (7.2)	4.4 (6.50, 0.1)	4.4 (0.3)	6.3 (0.1)
K39 (9.0)	4.7 (8.43, 0.1)	1.9 (0.3)	5.7 (0.1)
K72 (8.6)	5.5 (7.02, 0.2)	2.4 (0.3)	5.6 (0.1)
K74 (7.4)	4.4 (6.48, 0.1)	4.1 (0.3)	5.5 (0.1)

Table S3.3. Summary of thermodynamic parameters in energy landscape of Lys variants. The pK_a values of the Lys residues are given in parentheses. All free energies are in units of kcal/mol. The pH value at which the measurements were made (for ΔG_O) and the error associated with each measurement are given in parentheses. The value for ΔG_{C-O} was calculated based on the shift of the pK_a^{App} from its normal pK_a value in water (such that $\Delta G_{C-O} = 1.36 * \Delta pK_a$, as described in Isom *et. al.*, PNAS 2011). It is important to note that the ΔG_{C-O} reflects the energy

difference between the closed and open states of the variant, and not between the background and the open state of the variant. The latter calculation must include a term to account for the pH-independent cost of inserting a large neutral Lys residue into the core of the protein (described in Isom *et. al.*, PNAS 2011). The denaturation *m*-values, which in SNase are highly sensitive to folding cooperativity and the presence of intermediates[53,54], were determined at the same pH as the ΔG_O values. The thermodynamic stability of the background protein is 11.7-11.9 kcal/mol between pH 5 - 8.5; the *m*-value at pH 7.0 is 4.8 kcal/mol*M.

Chapter 4

The properties of buried ion pairs are governed by the propensity of proteins to reorganize

Version currently available on the bioRxiv preprint server as Kougentakis, C.M., Skettirr, L., Majumdar, A., Schlessman, J.L., and García-Moreno, E.B. The properties of buried ion pairs are governed by the propensity of proteins to reorganize. *biorXiv*, doi:10.1101/2020.02.03.932012, (2020).

4.1 Abstract

Charges are incompatible with the hydrophobic interior of proteins, yet proteins use buried charges, often in pairs or networks, to drive energy transduction processes, catalysis, pH-sensing, and ion transport. The structural adaptations necessary to accommodate interacting charges in the protein interior are not well understood, as the Coulomb interaction between two buried charges cannot offset the highly unfavorable penalty of dehydrating two charges. This was investigated experimentally with two variants of staphylococcal nuclease (SNase) with Glu:Lys or Lys:Glu pairs introduce at internal i, i+4 positions on an α-helix. Contrary to expectations from previous theoretical and experimental studies, the proteins tolerated the charged ion pairs in both orientations. Crystal structures and NMR spectroscopy studies showed that in both variants, side chains or backbone are reorganized to the expose at least one of the two buried groups to water. Comparison of these ion pairs with a highly stable buried ion pair in SNase shows that the location and the amplitude of structural reorganization can vary dramatically between ion pairs buried in the same general region of the protein. The propensity of the protein to populate alternative conformation states in which internal charges can contact water appears to be the factor that governs the magnitude of electrostatic effects in hydrophobic environments. The net effect of structural reorganization is to weaken the Coulomb interactions between charge pairs; however, the reorganized protein no longer has to pay the energetic penalty for burying charges. These results provide the framework necessary to understand the interplay between the dehydration of charges, Coulomb interactions and protein reorganization that tunes the functional properties of proteins.

4.2 Introduction

Buried ionizable residues in proteins are rare but essential for biochemical energy transduction. When buried alone, the pK_a values of these residues are shifted in the direction that favors the neutral state, relative to their values in water, consistent with the unfavorable transfer of an ionizable moiety from water to the hydrophobic interior of a protein[1]. In many natural systems, such as in enzyme active sites[2,3], proton pumps[4,5], ion channels[6], and pH-sensing motifs in signaling proteins[7,8], ionizable residues can be buried in pairs or in networks. Despite being able to form ion pairs[9,10], continuum electrostatics theory predicts that a favorable Coulomb interaction is not strong enough to compensate for the unfavorable dehydration of two charges[11,12]. Detailed structural and thermodynamic characterization of the energetics of buried ion pairs in natural systems is rarely feasible since many of these proteins are difficult to work with experimentally. The lack of experimental data has hampered the development of computational methods capable of accurately capturing the interplay between favorable Coulomb interactions and unfavorable dehydration energies governing the energetics of ion pair burial. A detailed understanding of charge-charge interactions in hydrophobic environments is necessary to gain mechanistic insight into biochemical energy transduction processes, and to design novel proteins (e.g., enzymes, H^+ pumps) with these important functional properties.

Recent studies have shown that it is possible to introduce two ionizable residues in the hydrophobic core of the model protein staphylococcal nuclease (SNase) as a charged pair[12], without significantly affecting the structure of the background protein. When buried alone, the pK_a values of the introduced Glu-23 and Lys-36 are highly anomalous and favor the neutral state, relative to their values in water. When introduced together, subtle reorganization of backbone and side chain dipoles along with water penetration creates a highly polar environment that stabilizes

the charged E23/K36 pair, with a coupling energy of 5 kcal/mol. Previous theoretical studies predicted that the arrangement of dipoles that would stabilize the arrangement of one ion pair would destabilize the reverse orientation[13]. Consistent with this prediction, reversing the introduced ion pair in SNase (V23K/L36E) leads to a protein of marginal stability, as the protein appears incapable of reorganizing dipoles or introducing internal water molecules to stabilize the reverse pair[14].

In this study, we have characterized the structural and thermodynamic properties of two variants of SNase with internal Lys:Glu or Glu:Lys pairs introduced on the i,i+4 positions of an α-helix, where they would be poised to form interactions in both orientations[15161718192021]. The pK_a values of the buried residues in the single variants are shifted to favor the neutral state, relative to their values in water. The pK_a values of the introduced Glu/Lys residues in both T62K/V66E (KE) and T62E/V66K (EK) variants are normalized relative to those measured in the single variants (i.e. more similar to their values in water)[1,22]. These proteins exhibit more complex conformational responses than observed in the 23:36 ion pair variants, which allows the protein to tolerate the burial of charged EK and KE pairs regardless of orientation. Comparison of the structural response of these proteins with those of previously studied ion pair variants of SNase demonstrates that charged groups are incompatible with the protein interior even in the presence of stabilizing interactions, and that correctly predicting alternative conformations accessible in the conformational ensemble is necessary to understand how proteins are able to tolerate the burial of charged pairs and networks.

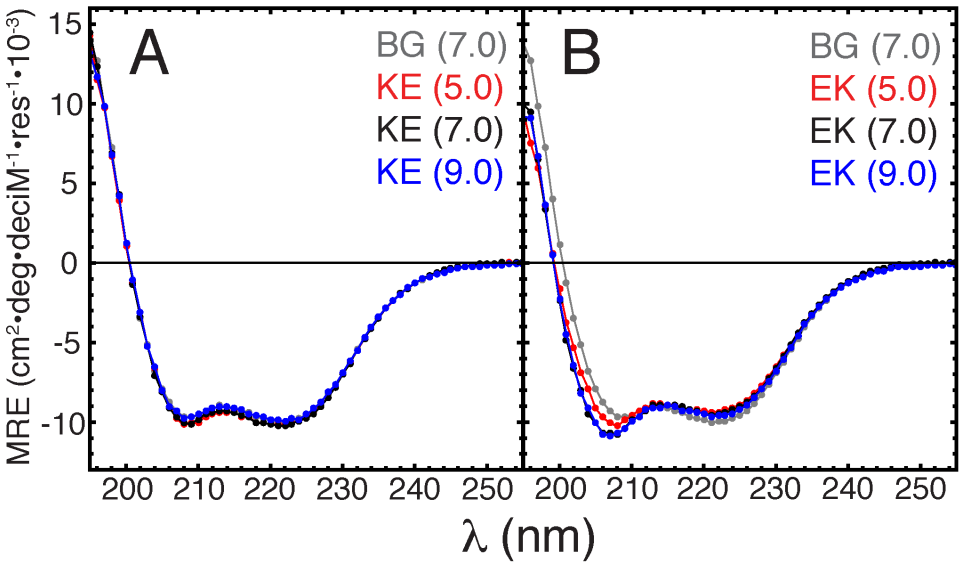

Fig. 4.1. Circular dichroism spectra of background protein, **(A)** KE and **(B)** EK variants as a function of pH. The pH value of each sample is shown in parenthesis.

4.3 Results

4.3.1 Circular dichroism (CD) spectroscopy

Circular dichroism spectra of the EK and KE variants collected at pH values between 5-9 (Figure 1) were indicative of well-folded proteins. The spectra of the KE variant are pH independent and were superimposable with those of the background protein. Conversely the EK variant displays small but reproducible shifts at 208 and 222 nm relative to the background protein across the range of pH values studied, consistent with minor structural reorganization. Similar shifts in the CD spectra are observed in both the T62E and V66K single variants at pH values where their respective introduced residues are ionized (Figure S1).

4.3.2 Thermodynamic stability

The thermodynamic stabilities of the two ion pair variants were measured as a function of pH using chemical denaturation with guanidinium hydrochloride monitored with Trp fluorescence (Fig. 2, Table 1 and S1-3). Introduction of an ionizable residue with a perturbed pK_a leads to a characteristic pH dependence to its thermodynamic stability[1]. Thermodynamic linkage analysis of the difference in the pH-dependence of the thermodynamic stability between the background protein and the variant protein with buried ionizable residues can be used to determine apparent pK_a values of these introduced residues. Previous studies have shown that buried Lys residues at positions 62 and 66 titrate with pK_a values of 8.1 and 5.7, respectively, far from the pK_a value of 10.4 of Lys in water[22]. Buried Glu residues at the same positions have pK_a values of 7.7 and 8.5, respectively, far from the pK_a value of 4.5 of Glu in water[1].

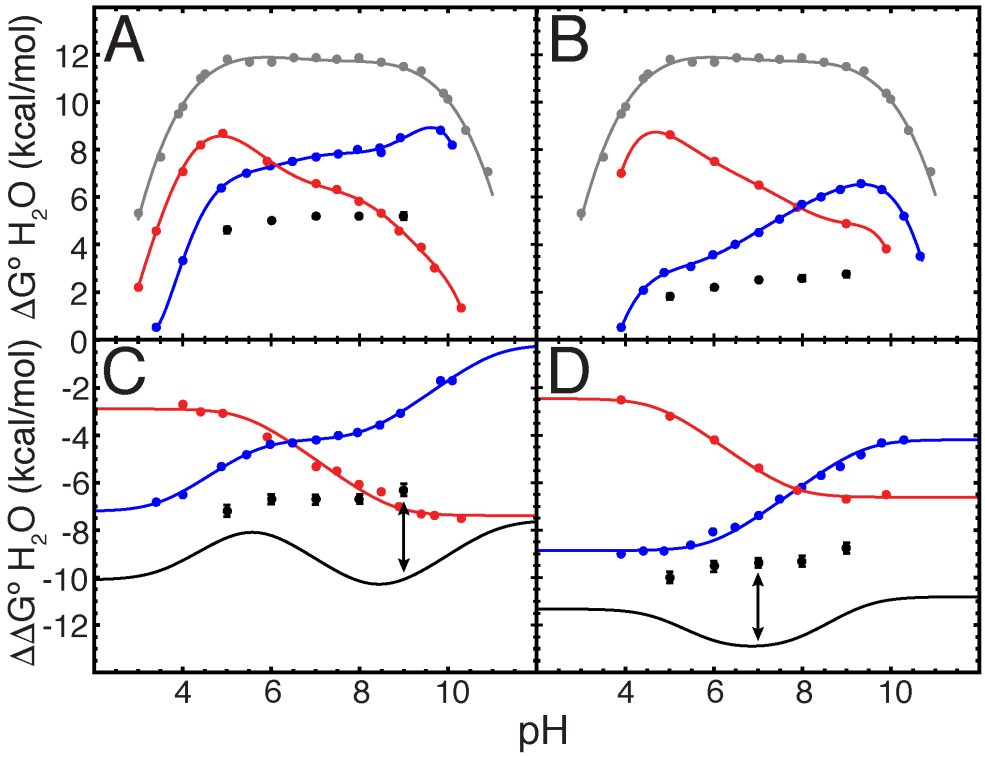

Fig. 4.2. pH dependence of the thermodynamic stabilities of the KE **(A and C)** and EK **(B and D)** variants. Data for single Lys variants are shown in blue, single Glu variants in red, double variants in black, and background protein in grey. Solid lines in A and B are meant solely to guide the eye. Solid colored lines in C and D show fits to a thermodynamic linkage equation used to determine the pK_a values of single variants, while black line represents an additive model where there is no coupling energy between the two introduced residues. Error bars for double variant data represent standard errors from triplicate measurements.

The $\Delta G°_{H2O}$ of the KE variant ranged from 4.6 to 5.2 kcal/mol at pH values between 5 to 9, with a stability of 5.2 kcal/mol at pH 7. The $\Delta G°_{H2O}$ of the EK variant ranged from 1.8 to 2.7 kcal/mol at pH values between 5 to 9, with a stability of 2.5 kcal/mol at pH 7. The thermodynamic stabilities of the KE and EK variants relative to the background protein ($\Delta\Delta G°_{H2O}$) are relatively constant between pH 5-9. These data are consistent with the pK_a value of both groups being either normalized or further shifted in opposite directions (opposing pK_a shifts of two residues would cancel out the pH-dependence of the protein's thermodynamic stability[12]).

The double variants were significantly more stable than expected from the introduction of two destabilizing substitutions. The coupling energy ($\Delta\Delta G_{int}$) was determined by comparing the relative stabilities of these variants compared to background to an additive model of the single variants, which assumes no interaction between the two residues[23]:

$$\Delta\Delta G_{int} = \Delta\Delta G^{\circ}_{H2o,KE/EK} - (\Delta\Delta G^{\circ}_{H2o,Single\ Lys} + \Delta\Delta G^{\circ}_{H2o,Single\ Glu})$$

At the pH values where the additive model is predicted to be most destabilizing (~pH 9 for KE and ~pH 7 for EK), the KE and EK variants are 3.9 and 3.4 kcal/mol more stable than predicted, respectively. It is important to note that these coupling energies do not reflect a direct interaction between the two residues, as the assumption that the protein structure is not perturbed by introduction of the two ionizable side chains was shown to not be correct (see below).

4.3.3 Crystal structures

Crystal structures of the KE and EK variants were solved to 1.7 and 2.2 Å, respectively, allowing for a structural comparison with the background protein (Figure 3A-C). The microenvironment of Thr-62 and Val-66 is completely hydrophobic in the background protein. In the T62K variant, Lys-62 is completely buried and excluded from solvent (Figure 3D), while the

ionizable moieties of the buried residues in T62E, V66E, and V66K are within hydrogen bonding distance (< 3.5 Å) of buried or interfacial water molecules (Figure 3E-G).

The crystal structure of KE was solved in complex with Ca^{2+} and thymidine 3',5'-diphosphate (pdTp) at pH 9 (Figure 3A, H). The protein backbone is superimposable with that of the background protein, with a backbone Cα RMSD of 0.19 Å. Lys-62 and Glu-66 are buried and within hydrogen bonding distance to each other, though Glu-66 is observed in two conformations. In the more deeply buried conformation (55% occupancy) the Oε1 and Oε2 atoms of Glu-66 are 2.6 and 3.4 Å away from the Nζ atom of Lys-62, and a single internal water molecule is found 2.7 Å away Oε2. The minor conformation of Glu-66 (45% occupancy) is more solvent exposed with Oε2 overlapping the position of the H-bonded water in the major conformation; Oε1 is within 2.8 Å from the Nζ atom of Lys-62, while Oε2 is 2.7 and 3.1 Å from two interfacial water molecules. Lys-62 makes no other polar contacts besides those to Glu-66.

The crystal structure of EK was solved at pH 7 in the absence of inhibitor was also highly similar to the background protein, with a overall backbone Cα RMSD of 0.47 Å, however significant structural changes were observed (Figure 3B, I). This variant crystallized in a P6(3) space group, which has not been observed in any of the 250+ previously solved structures of SNase. The loop region between residues 113-120 adopts a unique conformation due to different intermolecular contacts in this crystal form. Electron density for residues 15-24 in the β-1 and β-2 strands is very weak or missing and could not be reliably modeled. This structural change effectively exposes Glu-62 and Lys-66 to bulk solvent. The ionizable moieties of Glu-62 and Lys-66 are shifted slightly relative to their positions in the single variants to be in hydrogen bonding distance from each other, with an Oε to Nζ distance of 3.2 Å.

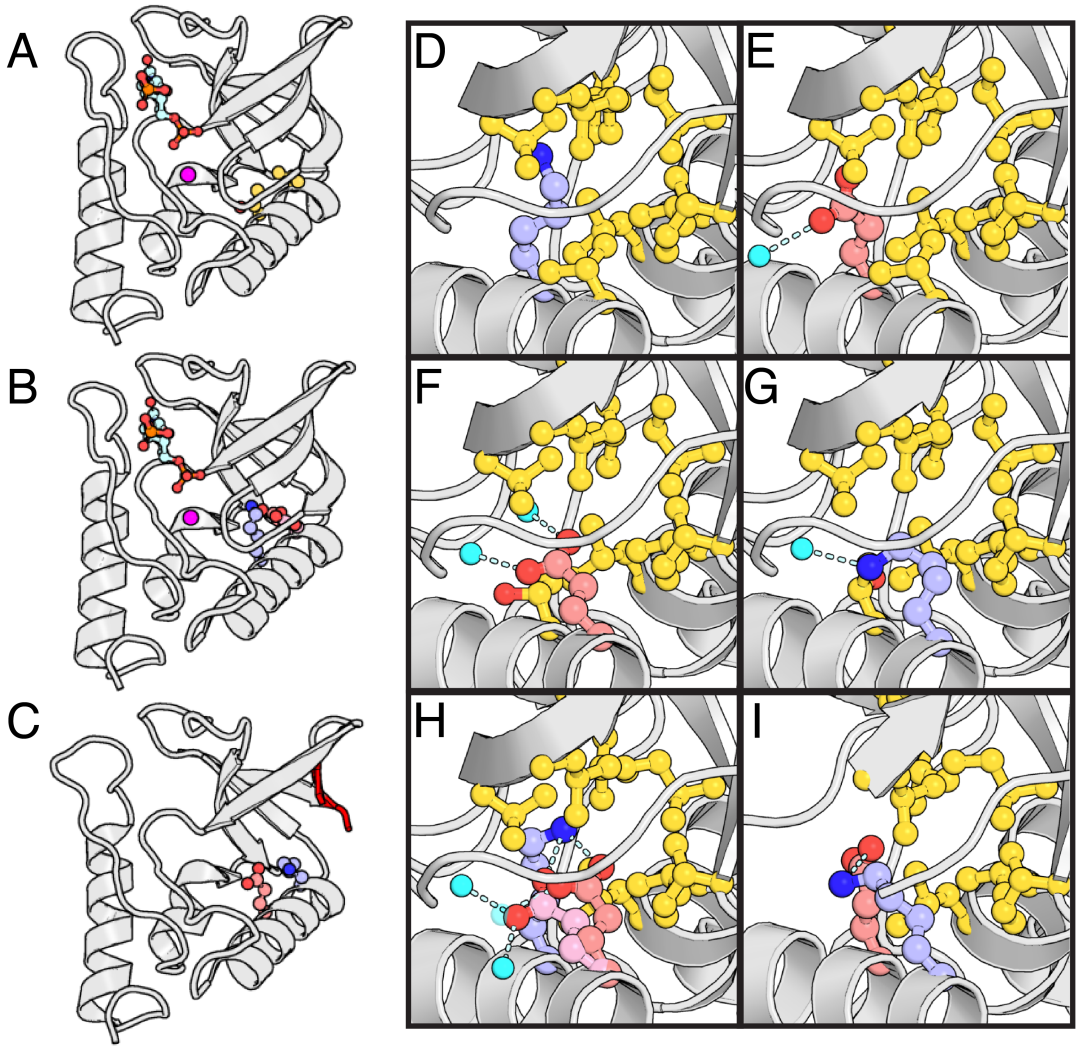

Fig. 4.3. Crystal structures. **(A)** Crystal structure of background protein at pH 8, in complex with calcium (purple) and inhibitor pdTp. (**B**) Crystal structure of KE at pH 9, in complex with calcium and inhibitor pdTp. **(C)** Crystal structure of EK at pH 7. **(D-I)** Microenvironment of 62:66 pairs in variants with K62 at pH 8 (**D**), E66 at pH 6 (**E**), E62 at pH 9 (**F**), K66 at pH 9 (**G**), K62/E66 at pH 9 (**H**), and E62/K66 at pH 7 (**I**). Hydrophobic residues are highlighted in yellow, and water molecules in cyan. Transparent water molecule in KE is partial occupancy. Lower occupancy Glu-66 side chain in KE is shown in light pink.

4.3.4 NMR spectroscopy: Backbone characterization

The ^{1}H-^{15}N HSQC spectra of both KE and EK variants contain well-dispersed, sharp resonances, indicative of folded proteins, though a small but significant degree of random coil resonances were observed in the EK spectra (Figure 4A, B). Each variant had a single resonance missing that was assignable in the background protein (residues 23 and 21 in EK and KE, respectively). Chemical shift perturbations (CSPs) were observed for both the backbone N and H_N resonances with the largest changes occurring around the mutation site and the β-1 and β-2 strands, the latter region being where resonance loss was observed (Figure S2-3). Similar trends were observed for the Cα CSPs, which are highly sensitive to structural perturbations in proteins[24], in both variants near the β-1 and β-2 strands (Figure 4C, D, S4). The CSPs around the β-1 and β-2 strands in all nuclei were larger in the EK variant. ^{1}H-^{15}N NOE measurements demonstrate that the β-1/β-2 turn was more dynamic on the ps-ns timescales in EK than in KE or the background protein, consistent with partial unfolding (Figure 4E). The chemical shifts of the residues in this turn in EK were consistent with partial unfolding, based on Random Coil Index[25] predictions (Figure 4F).

4.3.5 NMR spectroscopy: Direct detection of buried ion pairs

Side chain Cβ/γ-CO correlation spectra (^{13}C-^{13}C CBCGCO) of the single Glu variants and ion pair variants (Figure 5A, B, S5) were collected to investigate the charge state and hydrogen bonding patterns of the introduced side chain carboxylate moieties. The chemical shifts of the neutral Glu resonances in the single variants were distinct from those of the surface residues. The chemical shifts of E62 and E66 in the double variants were more similar to those of surface residues than the buried, neutral Glu residues, and very similar to the previously studied ionized

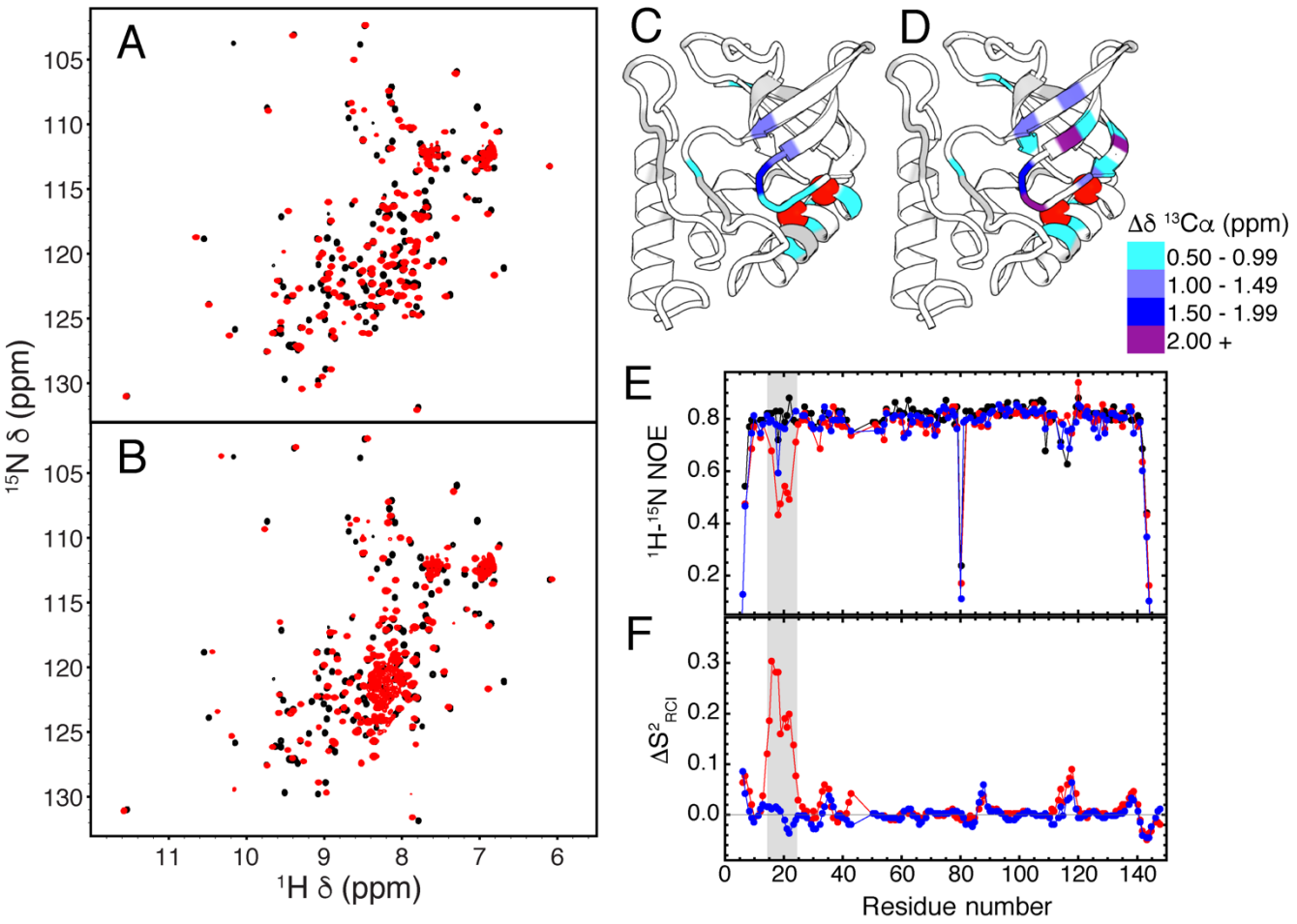

Fig. 4.4. Backbone characterization by NMR spectroscopy. ^{1}H-^{15}N HSQC spectra of background protein (pH 6.6), **(A)** KE (pH 7.0), and **(B)** EK (pH 6.6) variants. The pH value of each sample is denoted in parenthesis. ^{13}Cα chemical shift perturbations (CSPs) of **(C)** KE, pH 7.0 and **(D)** EK, pH 6.6, relative to the background protein at pH 6.6. The mean CSP was 0.21 for KE and 0.28 for EK, and the standard deviation was 0.48 for KE and 0.60 for EK. Substitution sites are shown as red spheres, but are not included in analysis due to the sensitivity of the Cα chemical shift to the residue identity. (E) ^{1}H-^{15}N heteronuclear NOE analysis of background protein (pH 7.4, black), KE (pH 6.1, blue), and EK (pH 7.4, red). The pH values were chosen based on spectral resolution. (F) Comparison of Random Coil Index S^2 values (predicted from chemical shifts), shown as background (pH 6.6) minus variant (ΔS^2). ΔS^2 values for KE variant (pH 7.0) are shown in blue,

and those for EK (pH 6.6) in red. More positive values indicate regions that are more dynamic in variant than the background protein. Values below 0.8 should not be considered significant.

E23 in the V23E/L36K variant of SNase[12]. Two resonances in the CBCGCO spectra of KE were assignable to Glu-66, indicative of slow exchange behavior caused by conformational exchange on ms or longer timescales[26] consistent with the alternate conformations observed in the crystal structure. In the double variants the ^{13}C chemical shift of the side chain carboxylate of both E62 and the major form of E66 did not show evidence of a deuterium isotope shift upon transfer from H_2O to D_2O. In general, protonated Glu residues show evidence of ~0.23 ppm isotope shift upon transfer to D_2O[27]. The lack of deuterium isotope shift in E62 or E66 demonstrates that these Glu residues are deprotonated between pH 7-8. The resonance of the minor form of E66 was too broad for isotope shift studies. Both the major and minor form of E66 and E62 lost resonance intensity below pH 6, presumably due to exchange with protonated states.

The protonation state of the introduced Lys residues was determined by detection of their side chain ^{15}N and ^{13}C nuclei. At pH values where Lys-62 and Lys-66 are deprotonated in the single variants, no Nζ resonances consistent with a deprotonated Lys (20-26 ppm[28,29]) residue were observed in the double variants in ^{15}N 1D experiments (Figure 5C), suggesting that either the Lys residues were protonated and indistinguishable from the surface resonances, or were broadened beyond detection due to μs-ms dynamics. To distinguish the between those possibilities, the side chain carbon resonances were assigned, as the $^{13}C\delta$ chemical shift is also a sensitive reporter of charge state[30]. All side chain ^{13}C resonances of Lys-62 were broadened beyond detection past the β position, suggestive of dynamic behavior. Similar behavior is observed previously in the T62K single variant when Lys-62 ionizes[29] (Figure S7). In contrast, the Cδ and Cε resonances of Lys-66 were observed in CC(CO)NH and ^{1}H-^{13}C-^{13}C NOESY experiments. The Cδ of Lys-66 was shifted upfield by 4.5 ppm in the EK variant relative to its value in the V66K variant at pH values where

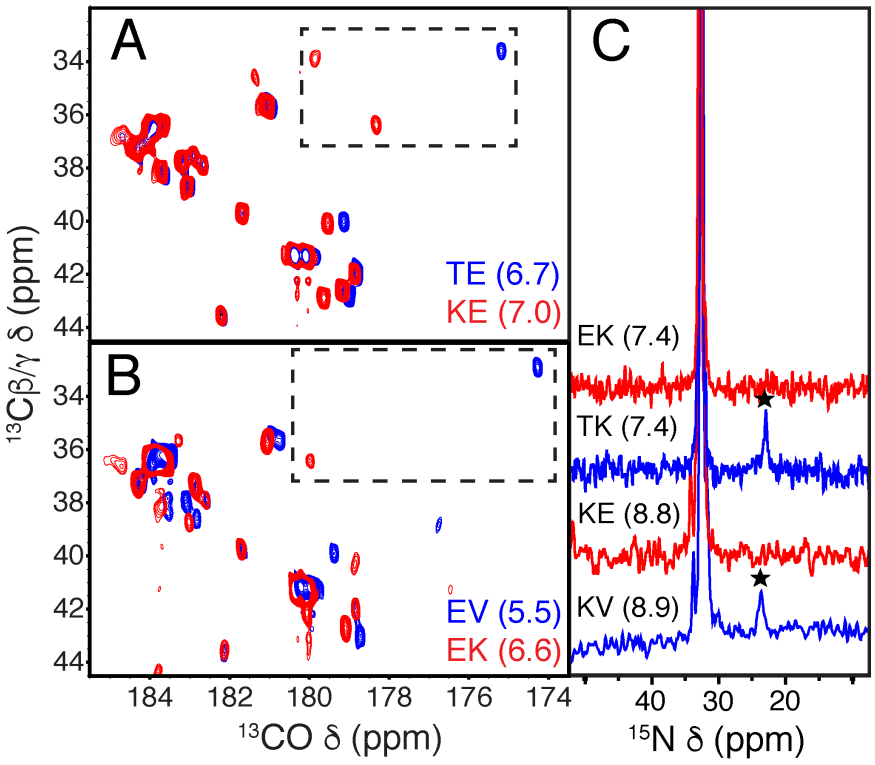

Fig. 4.5. NMR detection of buried ion pairs. ^{13}C-detect CBCGCO spectra of **(A)** KE and **(B)** EK single and double variants. Resonances from E66/E62 are highlighted in dashed boxes. The pH for EK and single variant T62E (EV) were not matched, as spectra were chosen based on pH at which backbone assignments for each variant were made; subsequent pH-titrations showed that changes in the position of E62 in single variant (EV) between pH 5.5 and 6.6 were minimal (Figure S6). **(C)** ^{15}N 1D spectra of T62K (KV) and KE (pH 8.9, bottom) and V66K (TK) and EK (pH 7.4, top). Resonances belonging to deprotonated Lys residues are denoted with stars, while the protonated Lys and the N-terminal NH_3^+ groups resonate between 30-40 ppm.

Lys-66 should be neutral. The magnitude of the shift is consistent with the ionization of the Lys side chain[27,30]. NOEs from deeply buried methyl groups to Lys-66 were also observed in both the V66K and EK variants, consistent with Lys-66 being partially buried in the hydrophobic protein interior (Figure S8).

4.4 Discussion

Classic continuum electrostatic theory describes the energetics of ion pair burial through the self-energies experienced by dehydrating each charge (ΔG_{ii}) and the Coulomb interaction experienced by the two charges (ΔG_{ij}). These energies are sensitive to the dielectric constant of the surrounding medium; although the measured dielectric constant of dry protein powders is between 2-4[31], continuum calculations require artificially high dielectric constants to accurately reproduce electrostatic effects in proteins[32]. The origin of the high dielectric constants has been controversial, having been attributed to factors including water penetration[33,34,35], side chain heterogeneity[36,37], and backbone reorganization[38,39,28]. Previous structural and thermodynamic characterization two ion pairs in SNase[12,14], along with the two additional pairs studied here has demonstrated that backbone reorganization is necessary to accommodate these groups in their charged state. The novel insight this study provides is the structural detail into how proteins reorganize to accommodate buried charges; the range of conformational responses observed in this study illustrates the challenges that structure-based energy calculations must overcome to properly capture the complexities of the local structural responses in response to charge burial.

4.4.1 Dielectric breakdown of the hydrophobic core

The EK and KE variants in this study were observed to undergo reorganization by CD, NMR spectroscopy and X-ray crystallography. The crystal structure and NMR spectroscopy data are completely consistent with partial unfolding of the β-1/β-2 region of the EK variant, which exposes the buried residues to bulk solvent. In the KE variant, the crystal structure showed both residues are internal, with NMR evidence suggesting only minor reorganization of the β-1/β-2 region. Instead, conformational heterogeneity of the Glu-66 side chains was observed in KE structure and supported by intermediate and slow exchange behavior of side chain by NMR. Although it has been previously demonstrated that Lys residues are capable of rapidly making and breaking hydrogen bonds on ps-ns timescales[40,41,42], the KE pair appears to exhibit exchange at much slower timescales (greater than μs). The slow exchange may reflect a large energy barrier between conformations due to the presence of dehydrated intermediate states. Finally NMR spectroscopy and analysis of the pH-dependence of the thermodynamic stability of these variants demonstrate that the Lys and Glu residues are charged between pH 5-9 suggesting the observed structural change is sufficient to shift the pK_a values of the groups away from their values in the single variants and back towards their values in bulk solvent.

The degree of reorganization observed in the EK and KE variants show that the dielectric environment governing the interaction between the charged residues is not governed by the properties of the protein interior, but rather the protein-water interface. In this sense, the responses of the EK and KE variants both represent dielectric breakdown as the protein interior must reorganize to the extent that its properties no longer dictate the interaction of the charged residues[43]. Instead, the localized structural responses in both variants are necessary to effectively hydrate the buried residues, allowing the residues to behave as if they were in water. Consistent with this view,

in the crystal structure of the KE variant the side chain of Glu-66 was observed to adopt conformations where it would be accessible to interfacial water molecules; although no water was observed in the lower resolution EK variant structure, partial unfolding would presumably expose the residues to interfacial or bulk water.

4.4.2 The conformational ensemble modulates electrostatic interactions in the protein interior

It is important to note that the ion pairs investigated here and in previous studies all perturb the same structural region of the protein, the β-1/β-2 strands and the intervening hairpin turn, though the magnitude of reorganization varies dramatically between the four variants. Crystal structures and NMR spectroscopy have shown that the β-1/β-2 strands can undergo partial unfolding in response to the ionization of a single Glu residue[39], although more subtle responses localized to just the hairpin turn between the strands have been observed[29]. In variants where E23/K36 or K62/E66 pairs are introduced the backbone response is subtle, with similar chemical shift perturbations in both variants localized to the β-1/β-2 turn (Figure 6, S4). Subtle backbone reorganization around this β-turn in these double variants may be underestimated in crystal structures. Both structures are bound to inhibitor and calcium, the latter requiring residues on the β1/2-turn for binding. Although the backbone behavior of variants with E23/K36 and K62/E66 are similar, the side chains in the protein interior have very different behaviors. The E23/K36 pair is stabilized by a "polar cage" of internal water molecules and protein dipoles. In the K62/E66 pair, both side chains show evidence of conformational heterogeneity; minor backbone reorganization alone cannot effectively stabilize the internal pair, the side chains must sample multiple conformations to remain effectively hydrated.

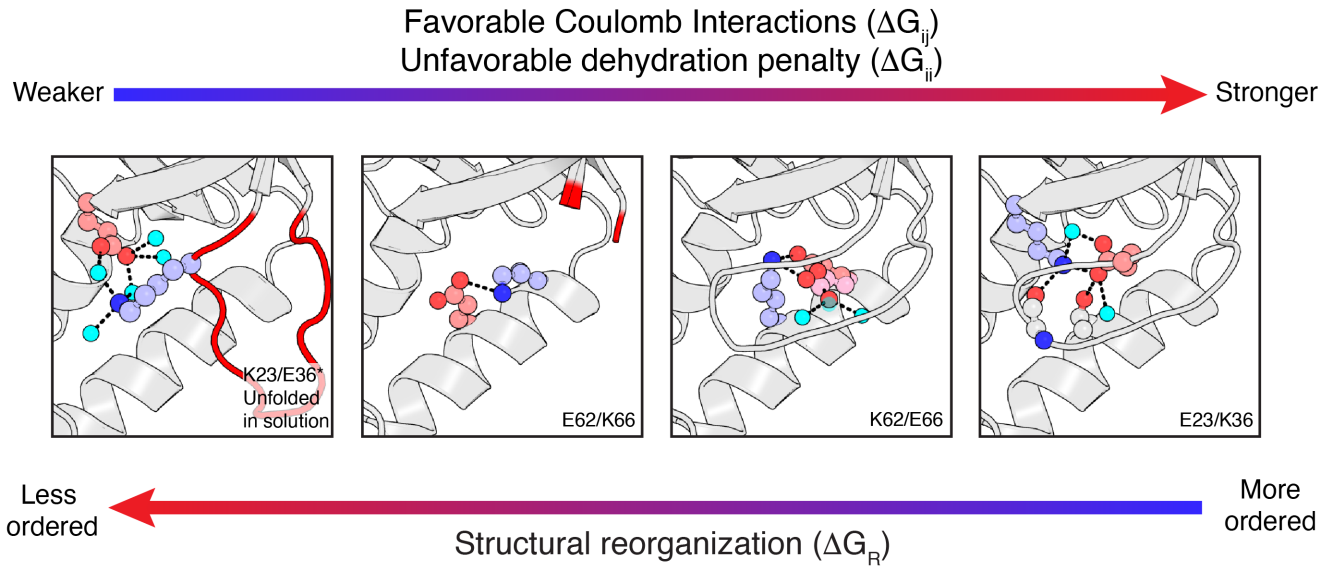

Fig. 4.6. Energetics of ion pair burial in proteins. Crystal structures of four ion pair variants of SNase are shown. Structures of variants with E23/K36 and K23/E36* were previously solved (PDB ID: 3NHH[12] and 3SK6[14], respectively); the structure shown for K23/E36* was solved in a more stable background protein (Δ+PHS*). The terms governing the energetics of ion pair burial (Coulomb, self-energy/dehydration, and reorganization) and their relative contributions in tolerating the four ion pairs in SNase are shown.

These subtle responses are in stark contrast to the backbone response to the ionization of E62/K66, which leads to unfolding of a 10 residue stretch of the β-1/β-2 strands, and K23/E36, which triggers subglobal unfolding. The K23/E36 pair must be introduced into a more stable variant of SNase, in order to be folded in solution. Even then, the protein responds similarly to the variant with E62/K66, where a portion of β-1/β-2 is unfolded but the rest of the protein is unperturbed[14]. In this state, K23 and E36 do not form a direct, "contact" ion pair[42] in the crystal structure, but appear to instead form a solvent-separated ion pair[44].

Although each variant studied has a different mechanism for tolerating burial of the charged pairs, in every case some degree of backbone reorganization was necessary, either as a dielectric response – making the protein interior a better solvent, or as dielectric breakdown – reorganization to expose the buried residues to solvent. What is remarkable is that these different behaviors are modulated by the same structural element in the protein, which can adopt multiple conformations to stabilize the ion pairs. The inherent conformational heterogeneity of this region of the protein is likely to be important for the success of these ion pair studies in SNase. Unsurprisingly, the energetics of ion pair burial appear to be more complex than of burial of single ionizable groups. In a survey of 25 internal Lys substitution in SNase, the cost of ionizing the buried side chain was correlated to the cost of ionizing the buried Lys and the extent of structural reorganization (manuscript in preparation). In contrast, the most ordered protein in the ion pair studies, the E23/K36 variant, is as stable as the E62/K66 variant (ΔG of both proteins = 2.5 kcal/mol at pH 7), despite the latter protein undergoing significant partial backbone unfolding. On the other hand, the K62/E66 variant is the most stable of the four variants studied (ΔG = 5.2 kcal/mol at pH 7) yet shows evidence of significant side chain and minor backbone reorganization.

4.4.3 Implications

The results of this study demonstrate that conformational heterogeneity can and does modulate charge-charge interactions in the protein interior. Previous statistical and structure-based calculation studies have suggested that buried ion pairs in nature can be stabilized through Coulomb interactions[45,46]. At least in SNase, Coulomb interactions alone are insufficient to stabilize the ionization of buried residues; the protein interior must either reorganize to increase its polarity or the buried residues must be exposed to bulk solvent. Both processes involve some degree of structural reorganization. The results of this study contradict a recent computational study that suggests ionized acidic residues in the protein interior can be stabilized by pairing with hydronium ions[47]. Although methods that explicitly account for reorganization have already been used to study the backbone response of the E23/K36 variant, it represents perhaps the simplest case given the relatively minor magnitude of the response [48,49]. The highly local, yet varied nature of the backbone and sidechain conformational responses of the variants in this study will be difficult to capture using current structure-based methods for studying pH-dependent conformational changes in proteins[50-54].

In nature, dynamical processes involving ion pairs in interfacial and hydrophobic environments are essential for a wide range of biochemical processes including both H^+ transport and e^- transfer and for modulating protein dynamics and pH-sensitivity in proteins. Ion pairs in protein-protein interfaces have been shown to be important for driving allosteric responses [5556]. Enzyme active sites also utilize buried ion pairs to modulate the reactivity of nucleophiles, tune the pK_a values of general acids/bases, and stabilize transition states. Although dynamics have been invoked to explain the catalytic power of some enzymes[57-60], the physical basis for how these motions are related to catalysis remains controversial[61]. Nevertheless, experimental work has

suggested that evolution of effective enzymes requires a balance of stability and flexibility to effectively tune function[62] for specific environmental niches[63]. The results of this study demonstrate how the interplay between conformational heterogeneity, flexibility[64,62], and stability[63] can be used as a mechanism for tuning electrostatic energies in functional sites in proteins.

4.4.5 Conclusions

The role of conformational reorganization in modulating the dielectric properties of proteins is of great interest in understanding biological energy transduction, catalysis, and pH-sensing. The results of this study indicate the challenges facing structure based energy calculations and design algorithms; although it may be possible to design proteins with enough thermostability to tolerate the introduction of functional sites, the arrangement of an active site in a crystal structure or the determination of the lowest energy conformation in structure prediction algorithm will not accurately capture the range of conformations observed in solution. Rather, methodologies that seek to explicitly model the conformational heterogeneity of proteins[65] will likely prove to be more successful in the design of novel proteins with the ability to drive energy transduction, respond to changes in pH, and catalyze novel chemical reactions. This will be difficult, as the highly local nature of the structural responses observed in this study, which range from side chain disorder to backbone unfolding, will be difficult to accurately capture for even the most state-of-the-art structure-based energy calculation methods.

4.5 Methods

4.5.1 Sample preparation

Plasmids containing the SNase variants with K62/E66 and E62/K66 were prepared by Quikchange mutagenesis, as previously described[1], using primers from Integrated DNA Technologies. Details of protein purification procedure are described in Chapter 2.5.1.

4.5.2 Optical spectroscopy

Protein samples consisted of 100 mM KCl, 10 mM buffer (CD) or 25 mM buffer (fluorescence), and 125 μg/ml (CD) or 50 μg/ml (fluorescence) protein. Potassium acetate was used as a buffer for pH 5 measurements, MES for pH 6, HEPES for pH 7 (substituted with Tris for CD), TAPS for pH 8, and CHES for pH 9. All CD experiments were performed on an Aviv model 420 circular dichroism spectrometer, as previously described[22], except each point (collected every 1 nm between 195-300 nm) was averaged for 10 seconds. All guanidinium melts were performed on an Aviv model 107 Automated Titration Fluorimeter as previously described[22], using ultra-pure guanidinium hydrochloride purchased from Fischer. All data were collected at 25 °C.

4.5.3 NMR Spectroscopy

Assignment and ^{13}C detected experiments were performed on Bruker Avance or Avance II spectrometers operating at a 600 MHz ^{1}H frequency, equipped with TCI cryogenic probes. NOESY experiments were performed on a Varian 800 MHz Inova spectrometer with a room temperature probe. ^{15}N detected data were collected on Varian 500 MHz Inova spectrometers, on a room temperature broadband observe probe. Protein samples consisted of 0.9-1.2 mM protein in 25 mM buffer (HEPES pH 7.0 for KE, MES pH 6.6 for EK), 100 mM KCl, and 10% D_2O for assignments.

For ^{1}H-^{15}N heteronuclear NOE experiments, data was collected at pH 6.1 in 25 mM MES for the KE variant, and pH 7.4 in 25 mM HEPES for the EK variant (100 mM KCl, 10% D_2O). The pH 7.4 sample of EK was also used for NOESY experiments. Assignments were obtained using standard triple resonance experiments (HNCACB, CBCA(CO)NH, HNCO, HBHA(CO)NH, H(CCCO)NH, CC(CO)NH). A CCH-TOCSY was necessary to assign the minor Glu-66 resonance in KE. NOESY experiments were performed using CH_3 selective ^{1}H-^{13}C-^{13}C NOESY-HMQC and ^{1}H-^{13}C-^{1}H NOESY-HSQC pulse sequences. ^{1}H-^{15}N NOE experiments were performed with a 4 s saturation period or recycle delay. NOE was calculated as the ratio of resonance intensities with versus without NOE saturation. ^{13}C detect CBCGCO experiments were performed using a previously published pulse sequence[39]. 64-128 scans were necessary to detect the resonances of E62 and E66, with acquisition times of 20 ms in the indirect dimension. For deuterium isotope shift experiments, a sample of each variant was made in 25 mM HEPES and 100 mM KCl (pH 7.3 and 7.8 for KE and EK, respectively). Samples were split into two tubes of identical volume (400 ml), lyophilized, and then resuspended in equivalent volumes of either 90% H_2O/10% D_2O or 99% D_2O. The pH at which the spectra were collected were chosen such that any shift in activity due to shift in pD (approximated as ±0.5 pH units) would not lead to a significant chemical shift change, which could obscure a potential isotope shift. ^{15}N 1D and heteronuclear cross polarization experiments were performed using the same parameters as previously described[29]. 1D data was visualized using Topspin 3.7. 2D and 3D data were processed with the NMRpipe[67] suite and analyzed using Sparky 3.115[68]. All data were collected at 25 °C.

4.5.4 Crystallography

Crystals of T62E, V66E, and KE (all in the Δ+PHS background) were grown using the hanging-drop vapor diffusion method at 4 °C. Protein concentrations of 8-12 mg/ml were used for each protein. Crystals of EK were grown at room temperature (20 °C) with a concentration of 23 mg/ml. 4 μl of protein solution was mixed with 4 μl of mother liquor. Crystals appeared within a month, and all were flash frozen prior to data collection except for the EK crystals, which were pushed through paratone oil prior to freezing and data collection. Details of data collection and statistics are provided in Table S4-5. All data was indexed, integrated, scaled and merged using the manufacturer's software. Processing was performed using the CCP4 suite, with initial phasing by molecular replacement done using PHASER, with the background protein's crystal structure used for the starting model (PDB ID: 3BDC). Model building and refinement was performed using COOT and Refmac-5, respectively.

4.6 References

1. Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno E., B. Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci.* **107**, 16096–16100 (2010).

2. Harris, T. K. & Turner, G. J. Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98 (2002).

3. Schwans, J. P., Sunden, F., Gonzalez, A., Tsai, Y. & Herschlag, D. Uncovering the determinants of a highly perturbed tyrosine pK_a in the active site of ketosteroid isomerase. *Biochemistry* **52**, 7840–7855 (2013).

4. Lanyi, J. K. Proton transfers in the bacteriorhodopsin photocycle. *Biochim. Biophys. Acta BBA - Bioenerg.* **1757**, 1012–1018 (2006).

5. Ballmoos, C. von, Wiedenmann, A. & Dimroth, P. Essentials for ATP synthesis by F_1F_o ATP synthases. *Annu. Rev. Biochem.* **78**, 649–672 (2009).

6. Hong, H., Szabo, G. & Tamm, L. K. Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. *Nat. Chem. Biol.* **2**, 627 (2006).

7. Isom, D. G. *et al.* Protons as second messenger regulators of G protein signaling. *Mol. Cell* **51**, 531–538 (2013).

8. Isom, D. G., Sridharan, V. & Dohlman, H. G. Regulation of Ras paralog thermostability by networks of buried ionizable groups. *Biochemistry* **55**, 534–542 (2016).

9. Barlow, D. J. & Thornton, J. M. Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885 (1983).

10. Rashin, A. A. & Honig, B. On the environment of ionizable groups in globular proteins. *J. Mol. Biol.* **173**, 515–521 (1984).

11. Hendsch, Z. S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226 (1994).

12. Robinson, A. C., Castañeda, C. A., Schlessman, J. L. & García-Moreno E, B. Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl. Acad. Sci.* **111**, 11685–11690 (2014).

13. Hwang, J.-K. & Warshel, A. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature* **334**, 270 (1988).

14. Robinson, A. C., Schlessman, J. L. & García-Moreno E, B. Dielectric properties of a protein probed by reversal of a buried ion pair. *J. Phys. Chem. B* **122**, 2516–2524 (2018).

15. Marqusee, S. & Baldwin, R. L. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci.* **84**, 8898–8902 (1987).

16. Lyu, P. C., Marky, L. A. & Kallenbach, N. R. The role of ion pairs in .alpha. helix stability: two new designed helical peptides. *J. Am. Chem. Soc.* **111**, 2733–2734 (1989).

17. Scholtz, J. M., Qian, H., Robbins, V. H. & Baldwin, R. L. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* **32**, 9668–9676 (1993).

18. Sivaramakrishnan, S., Spink, B. J., Sim, A. Y. L., Doniach, S. & Spudich, J. A. Dynamic charge interactions create surprising rigidity in the ER/K α-helical protein motif. *Proc. Natl. Acad. Sci.* **105**, 13356–13361 (2008).

19. Baker, E. G. *et al.* Local and macroscopic electrostatic interactions in single α-helices. *Nat. Chem. Biol.* **11**, 221 (2015).

20. Meuzelaar, H., Vreede, J. & Woutersen, S. Influence of Glu/Arg, Asp/Arg, and Glu/Lys salt bridges on α-helical stability and folding kinetics. *Biophys. J.* **110**, 2328–2341 (2016).

21. Fossat, M. J. & Pappu, R. V. q-canonical Monte Carlo sampling for modeling the linkage between charge regulation and conformational equilibria of peptides. *J. Phys. Chem. B* **123**, 6952–6967 (2019).

20. Isom, D. G., Castañeda, C. A., Cannon, B. R., and García-Moreno E., B. Large shifts in pK_a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci.* **108,** 5260–5265 (2011).

23. Schreiber, G. & Fersht, A. R. Energetics of protein-protein interactions: Analysis of the Barnase-Barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486 (1995).

24. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).

25. Berjanskii, M. V. & Wishart, D. S. A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971 (2005).

26. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1814**, 942–968 (2011).

27. Platzer, G., Okon, M. & McIntosh, L. P. pH-dependent random coil ^{1}H, ^{13}C, and ^{15}N chemical shifts of the ionizable amino acids: a guide for protein pK_a measurements. *J. Biomol. NMR* **60**, 109–129 (2014).

28. Takayama, Y., Castañeda, C. A., Chimenti, M., García-Moreno, B. & Iwahara, J. Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. *J. Am. Chem. Soc.* **130**, 6714–6715 (2008).

29. Kougentakis, C. M. *et al.* Anomalous properties of Lys residues buried in the hydrophobic interior of a protein revealed with ^{15}N-detect NMR spectroscopy. *J. Phys. Chem. Lett.* **9**, 383–387 (2018).

30. Hass, M. A. S. & Mulder, F. A. A. Contemporary NMR studies of protein electrostatics. *Annu. Rev. Biophys.* **44**, 53–75 (2015).

31. Harvey, S. C. & Hoekstra, P. Dielectric relaxation spectra of water adsorbed on lysozyme. *J. Phys. Chem.* **76**, 2987–2994 (1972).

32. Schutz, C. N. & Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins Struct. Funct. Bioinforma.* **44**, 400–417 (2001).

33. Damjanović, A., Brooks, B. R. & García-Moreno E., B. Conformational relaxation and water penetration coupled to ionization of internal groups in proteins. *J. Phys. Chem. A* **115**, 4042–4053 (2011).

34. Nguyen, D. M., Leila Reynald, R., Gittis, A. G. & Lattman, E. E. X-ray and thermodynamic studies of staphylococcal nuclease variants I92E and I92K: Insights into polarity of the protein interior. *J. Mol. Biol.* **341**, 565–574 (2004).

35. Chakrabarty, S. & Warshel, A. Capturing the energetics of water insertion in biological systems: The water flooding approach. *Proteins Struct. Funct. Bioinforma.* **81**, 93–106 (2013).

36. Alexov, E. G. & Gunner, M. R. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* **72**, 2075–2093 (1997).

37. Georgescu, R. E., Alexov, E. G. & Gunner, M. R. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophys. J.* **83**, 1731–1748 (2002).

38. Kato, M. & Warshel, A. Using a charging coordinate in studies of ionization induced partial unfolding. *J. Phys. Chem. B* **110**, 11566–11570 (2006).

39. Robinson, A. C., Majumdar, A., Schlessman, J. L. & García-Moreno E, B. Charges in hydrophobic environments: A strategy for identifying alternative states in proteins. *Biochemistry* **56**, 212–218 (2017).

40. Zandarashvili, L., Li, D.-W., Wang, T., Brüschweiler, R. & Iwahara, J. Signature of mobile hydrogen bonding of lysine side chains from long-range ^{15}N–^{13}C Scalar J-couplings and computation. *J. Am. Chem. Soc.* **133**, 9192–9195 (2011).

41. Anderson, K. M. *et al.* Direct observation of the ion-pair dynamics at a protein–DNA interface by NMR spectroscopy. *J. Am. Chem. Soc.* **135**, 3613–3619 (2013).

42. Chen, C. *et al.* Dynamic equilibria of short-range electrostatic interactions at molecular interfaces of protein–DNA complexes. *J. Phys. Chem. Lett.* **6**, 2733–2737 (2015).

43. Karp, D. A., Stahley, M. R. & García-Moreno E., B. Conformational consequences of ionization of Lys, Asp, and Glu buried at position 66 in staphylococcal nuclease. *Biochemistry* **49**, 4138–4146 (2010).

44. Yu, B., Pettitt, B. M. & Iwahara, J. Experimental evidence of solvent-separated ion pairs as metastable states in electrostatic interactions of biological macromolecules. *J. Phys. Chem. Lett.* **10**, 7937–7941 (2019).

45. Kim, J., Mao, J. & Gunner, M. R. Are acidic and basic groups in buried proteins predicted to be ionized? *J. Mol. Biol.* **348**, 1283–1298 (2005).

46. Bush, J. & Makhatadze, G. I. Statistical analysis of protein structures suggests that buried ionizable residues in proteins are hydrogen bonded or form salt bridges. *Proteins Struct. Funct. Bioinforma.* **79**, 2027–2032 (2011).

47. Wu, X. & Brooks, B. R. Hydronium ions accompanying buried acidic residues lead to high apparent dielectric constants in the interior of proteins. *J. Phys. Chem. B* **122**, 6215–6223 (2018).

48. Pathak, A. K. Effect of a buried ion pair in the hydrophobic core of a protein: An insight from constant pH molecular dynamics study. *Biopolymers* **103**, 148–157 (2015).

49. Pathak, A. K. Constant pH molecular dynamics study on the doubly mutated staphylococcal nuclease: capturing the microenvironment. *RSC Adv.* **5**, 94926–94932 (2015).

50. Shi, C., Wallace, J. A. & Shen, J. K. Thermodynamic coupling of protonation and conformational equilibria in proteins: Theory and simulation. *Biophys. J.* **102**, 1590–1597 (2012).

51. Goh, G. B., Laricheva, E. N. & Brooks, C. L. Uncovering pH-dependent transient states of proteins with buried ionizable residues. *J. Am. Chem. Soc.* **136**, 8496–8499 (2014).

52. Zheng, Y. & Cui, Q. Microscopic mechanisms that govern the titration response and pK_a values of buried residues in staphylococcal nuclease mutants. *Proteins* **85**, 268–281 (2017).

53. Liu, J., Swails, J., Zhang, J. Z. H., He, X. & Roitberg, A. E. A Coupled ionization-conformational equilibrium is required to understand the properties of ionizable residues in the hydrophobic interior of staphylococcal nuclease. *J. Am. Chem. Soc.* **140**, 1639–1648 (2018).

54. Damjanovic, A., Miller, B. T., Okur, A. & Brooks, B. R. Reservoir pH replica exchange. *J. Chem. Phys.* **149**, 072321 (2018).

55. Bell, C. E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Mol. Biol.* **7**, 209 (2000).

56. Zhan, H., Sun, Z. & Matthews, K. S. Functional impact of polar and acidic substitutions in the lactose repressor hydrophobic monomer·monomer interface with a buried lysine. *Biochemistry* **48**, 1305–1314 (2009).

57. Bhabha, G. *et al.* A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**, 234–238 (2011).

58. Camilloni, C. *et al.* Cyclophilin A catalyzes proline isomerization by an electrostatic handle mechanism. *Proc. Natl. Acad. Sci.* **111**, 10203–10208 (2014).

59. Hanoian, P., Liu, C. T., Hammes-Schiffer, S. & Benkovic, S. Perspectives on electrostatics and conformational motions in enzyme catalysis. *Acc. Chem. Res.* **48**, 482–489 (2015).

60. Kerns, S. J. *et al.* The energy landscape of adenylate kinase during catalysis. *Nat. Struct. Mol. Biol.* **22**, 124–131 (2015).

61. Warshel, A. & Bora, R. P. Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *J. Chem. Phys.* **144**, 180901 (2016).

62. Bar-Even, A., Milo, R., Noor, E. & Tawfik, D. S. The moderately efficient enzyme: Futile encounters and enzyme floppiness. *Biochemistry* **54**, 4969–4977 (2015).

63. Sugrue, E., Carr, P. D., Scott, C. & Jackson, C. J. Active site desolvation and thermostability trade-offs in the evolution of catalytically diverse triazine hydrolases. *Biochemistry* **55**, 6304–6313 (2016).

64. Elias, M., Wieczorek, G., Rosenne, S. & Tawfik, D. S. The universality of enzymatic rate–temperature dependency. *Trends Biochem. Sci.* **39**, 1–7 (2014).

65. Davey, J. A., Damry, A. M., Goto, N. K. & Chica, R. A. Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* **13**, 1280–1285 (2017).

66. Castañeda, C. A. *et al.* Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. *Proteins* **77**, 570–588 (2009).

67. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

68. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

4.7 Supporting Information

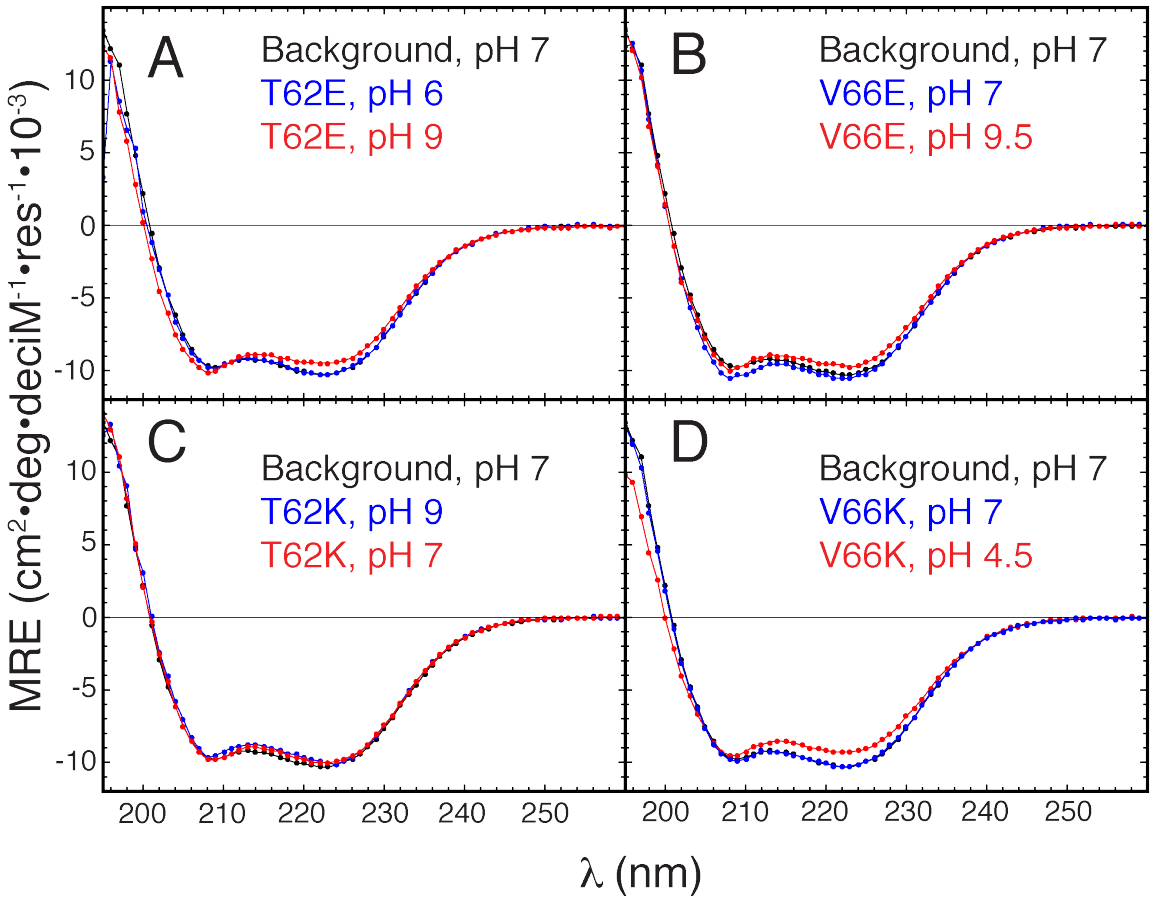

Fig. S4.1. Circular dichroism spectra of each single Lys and Glu variant at positions 62 and 66 relative to background protein. The pH values were chosen to populate the neutral and charged states of the introduced residues, as the pK_a values of the buried residues are **(A)** 7.7 for Glu-62, **(B)** 8.5 for Glu-66, **(C)** 8.1 for Lys-62, and **(D)** 5.7 for Lys-66.

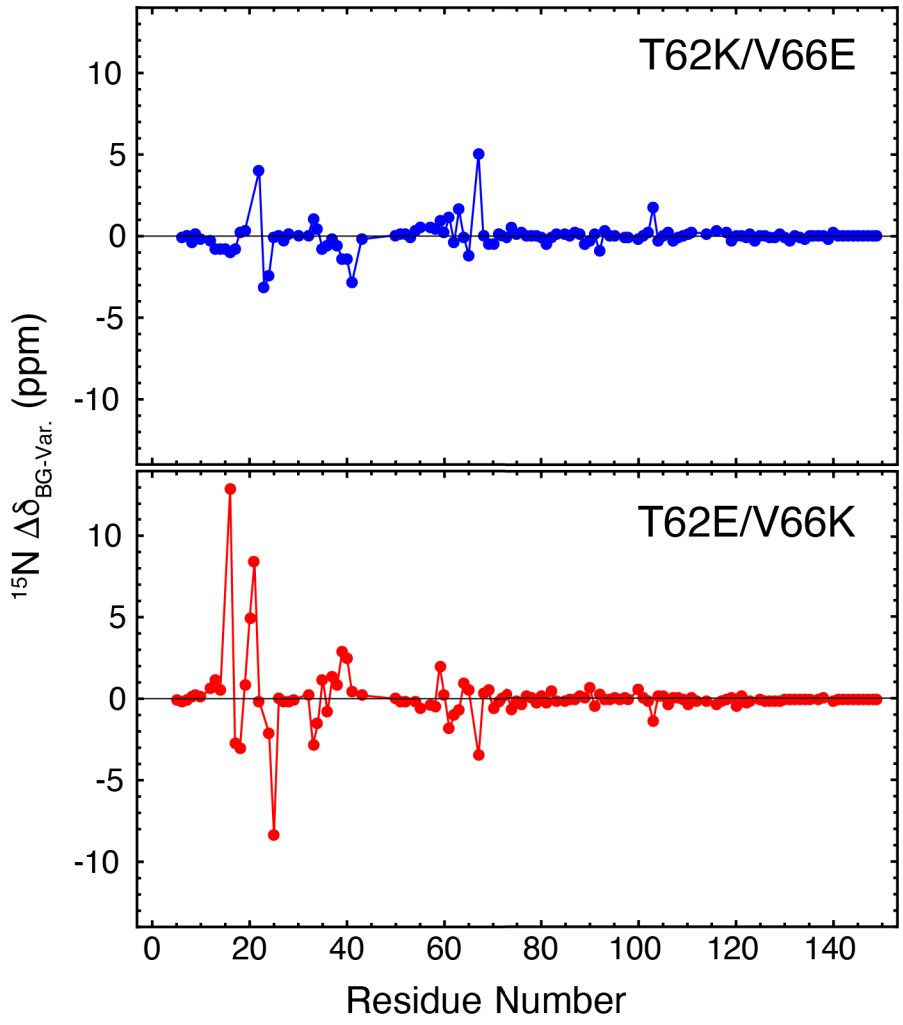

Fig. S4.2. Backbone ^{15}N chemical shift perturbations of background protein versus variants with K62/E66 (top) and E62/K66 (bottom).

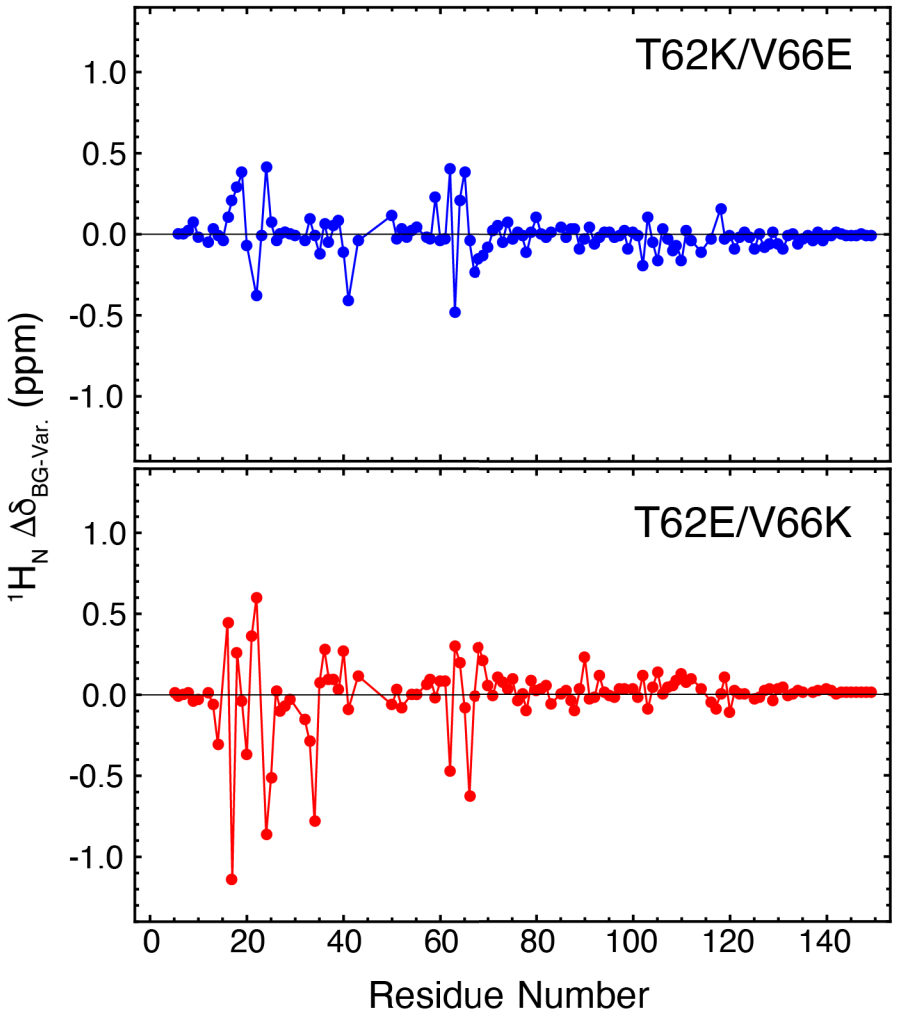

Fig. S4.3. Backbone ^1H_N chemical shift perturbations of background protein versus variants with K62/E66 (top) and E62/K66 (bottom).

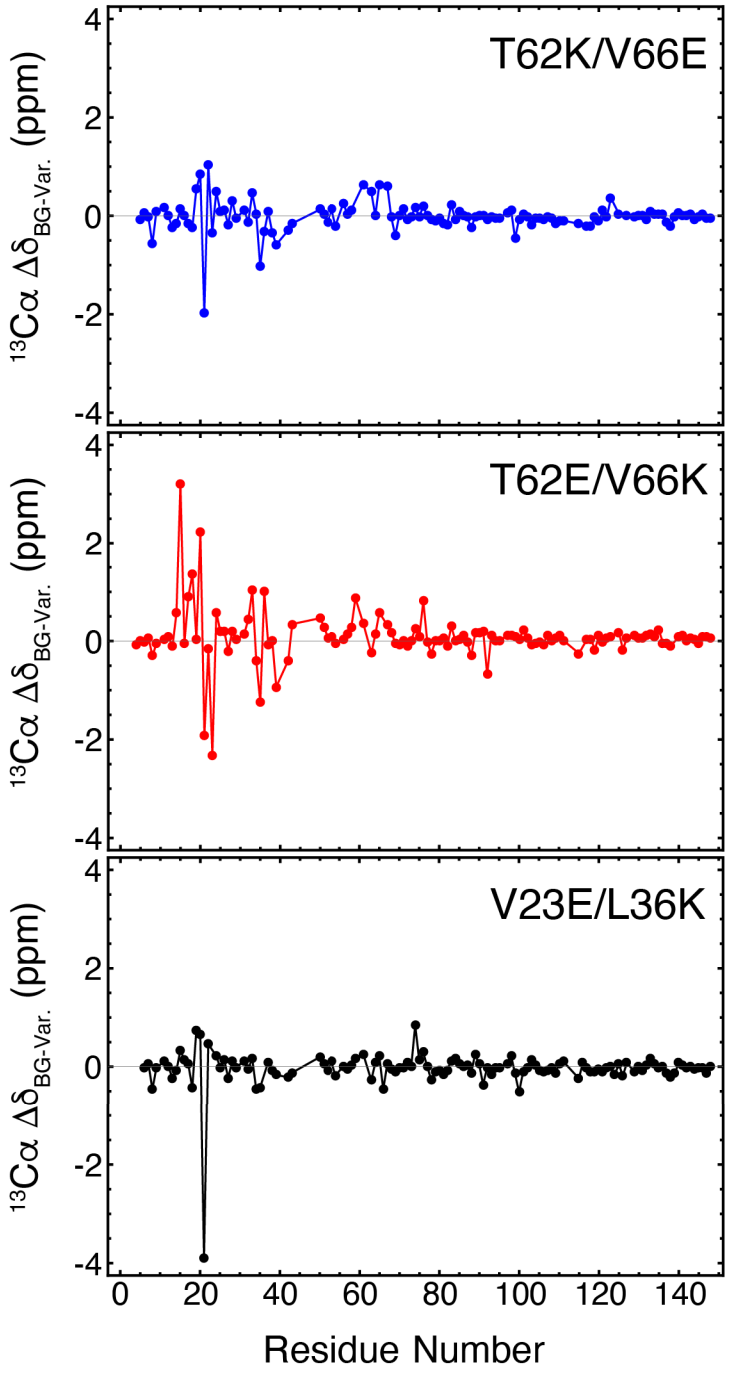

Fig. S4.4. Backbone $^{13}C\alpha$ chemical shift perturbations of background protein versus variants with K62/E66 (top), E62/K66 (middle), and E23/K36 (bottom). The variant with E23/K36 has significant CSPs greater than 0.5 ppm between residues 19-22 but are smaller in magnitude than either of the 62:66 ion pair variants.

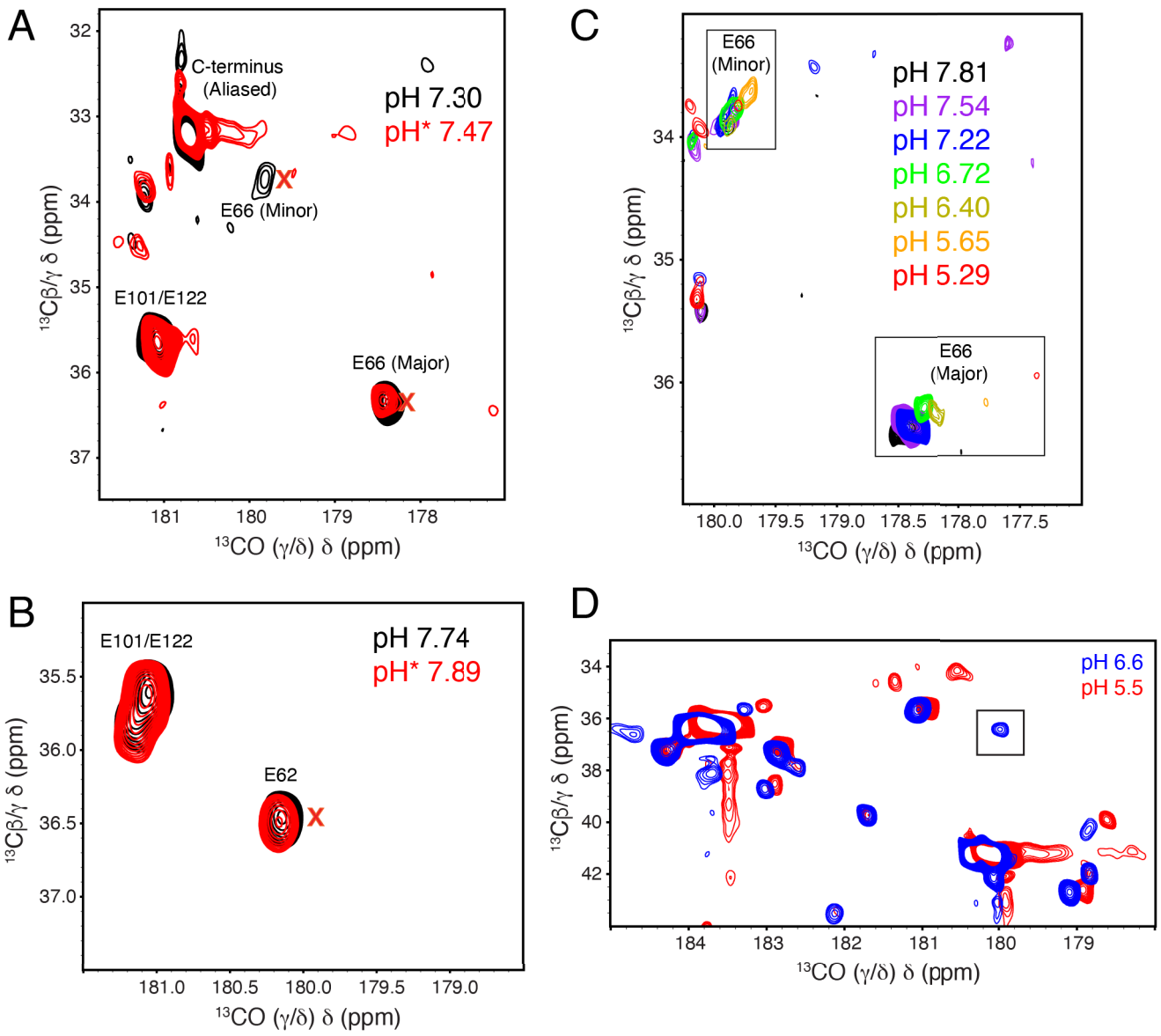

Fig. S4.5. ^{13}C-^{13}C CBCGCO spectra. Deuterium isotope shift experiments on **(A)** KE and **(B)** EK. Sample in 90% H_2O/10% D_2O shown in black, and sample in 99% D_2O shown in red. pH* is the uncorrected pH reading for the D_2O sample. The red X marks the position where the E62/E66 peak in D_2O would be expected to resonate if deuterated and neutral, due to the 0.23 ppm isotope shift observed upon transfer from D_2O to H_2O. Small shifts (less than 0.05 ppm) observed in E62 and E66 variants are not consistent with a neutral Glu, are similar to those

observed for surface Asp and Glu residues and are likely due to small changes in pH versus pD or small variations in salt concentration when resuspending lyophilized protein. **(C)** pH titration of Glu-66 in the KE variant. Major and minor peaks are outlined in boxes, with no detectable intensity for either peak below pH 5.7. **(D)** Resonance of E62 in the EK variant is unobservable below pH 6.0 (resonance at pH 6.6 highlighted in box).

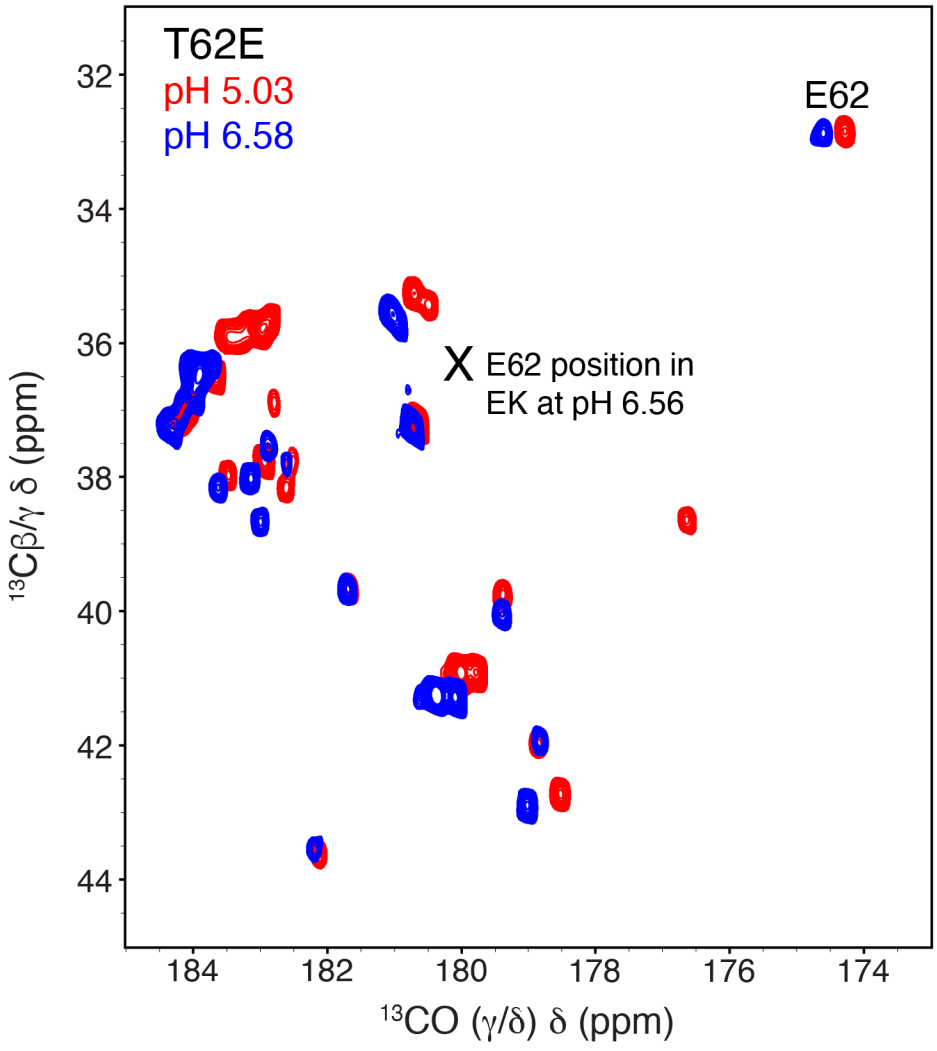

Fig. S4.6. ^{13}C-^{13}C CBCGCO spectra of T62E single variant at pH 5.03 (red) and 6.58 (blue). The chemical shift change of E62 in this range of pH values is minimal (E62 marked at top righthand corner). The position of the E62 resonance in the EK double variant is marked with an X, showing that pH-dependent changes in the position of E62 in the single variant cannot explain the large chemical shift change that E62 undergoes upon interaction with K66 in the EK variant.

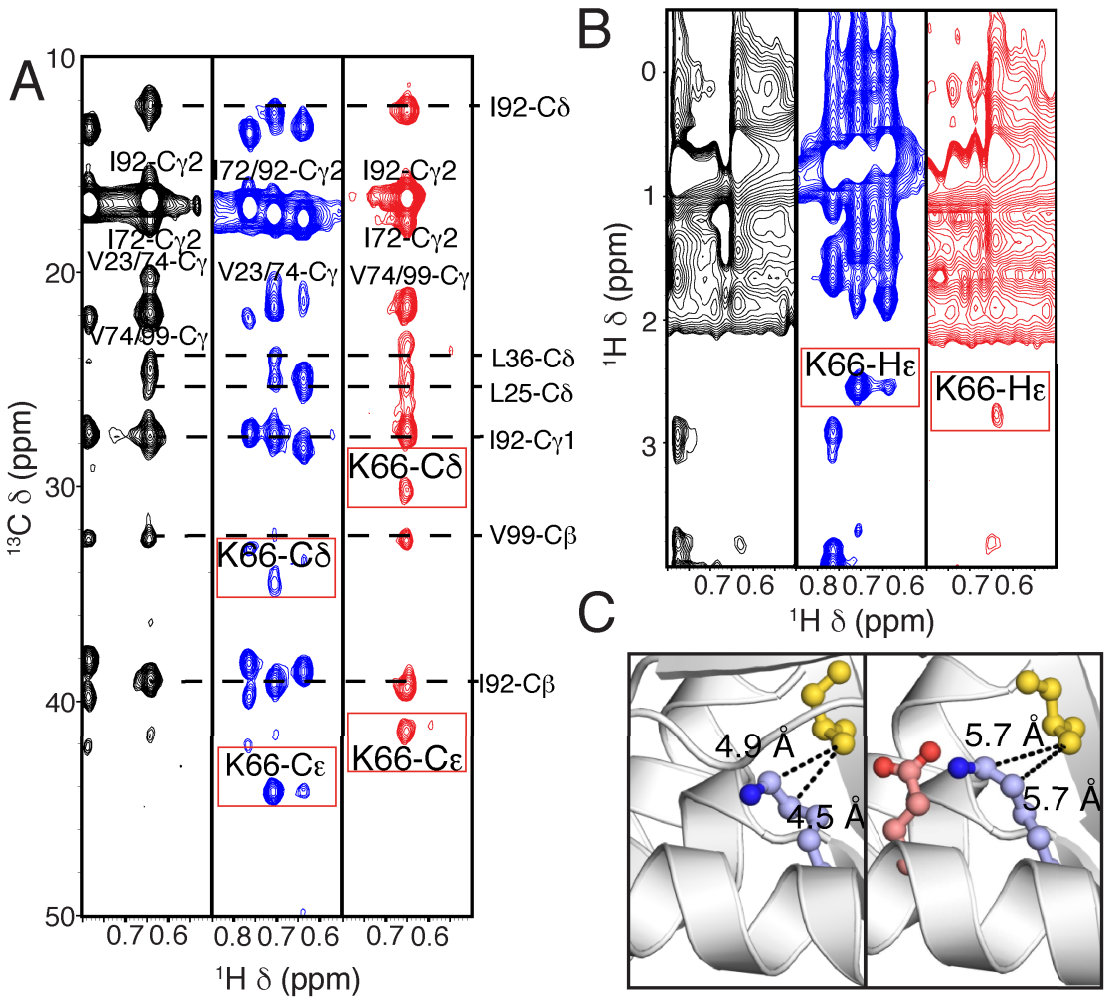

Fig. S4.7. CH_3-selective NOESY spectra of EK variant. Background protein is shown in black, V66K single variant in blue, and EK variant in red. All spectra were collected at pH 7.4, where Lys-66 is neutral in the V66K variant. All NOEs were assigned using CC(CO)NH and H(CCCO)NH experiments. **(A)** ^{1}H-^{13}C-^{13}C NOESY-HMQC slice from Ile-92 Cγ2-Hγ2. NOEs from the side chain of Lys-66 are shown in red boxes. Note the 4.5 ppm difference between Cδ chemical shifts of Lys-66 in V66K and EK, suggestive of a change in charge state. **(B)** ^{1}H-^{13}C-^{1}H NOESY-HSQC slice from Ile-92 Cγ2-Hγ2. NOEs from the side chain of Lys-66 are shown in red boxes. **(C)** Crystal structure close ups showing C-C distances between Ile-92 Cγ2 and Lys-66 Cε and Cδ atoms in V66K (left) and EK (right). These distances imply that the ^{1}Hγ2-^{1}Hε/δ distances

should be close enough (under 6 Å) for NOEs to be observed. 100 ms mixing times were used for each experiment; although the current data cannot account for potential relay effects, the large distance in sequence space between Ile-92 and Lys-66 suggests that even in the case of a relay effect, Lys-66 must be in close proximity to buried residues.

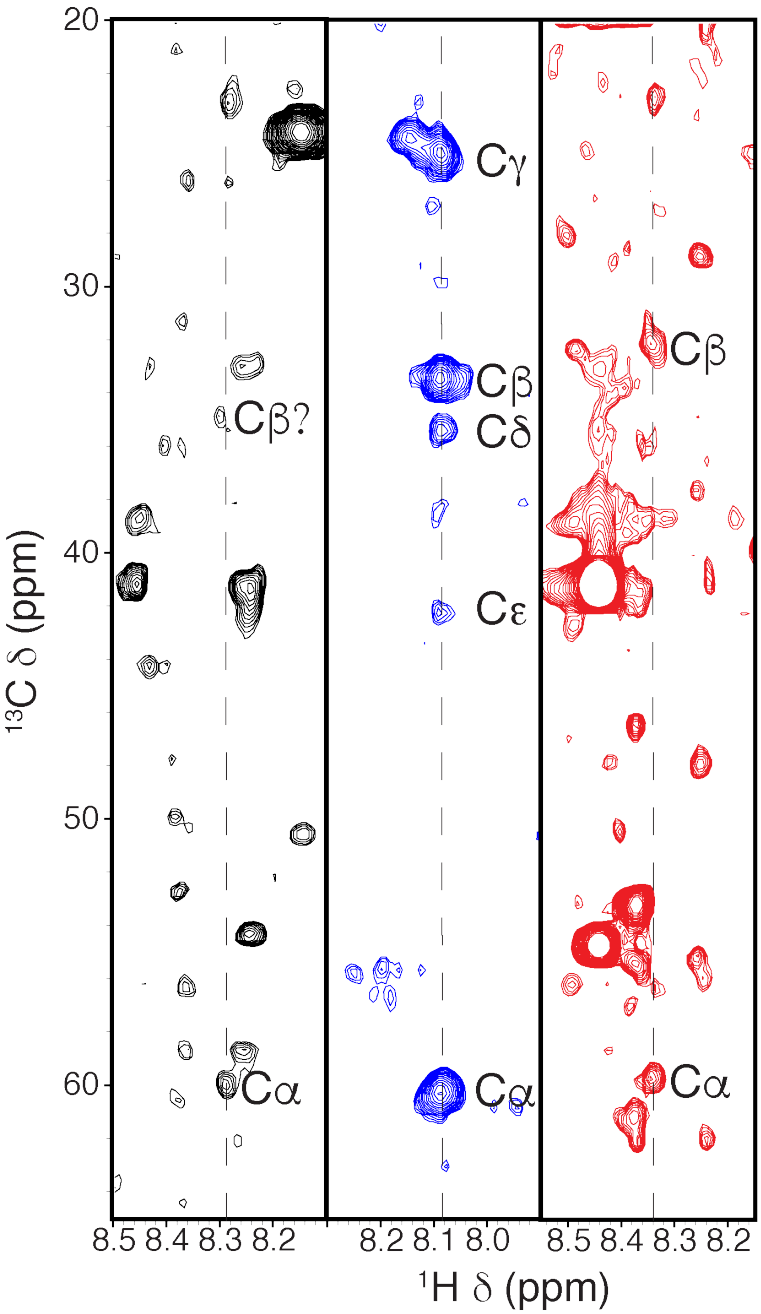

Fig. S4.8. CC(CO)NH spectra of KE variant at pH 7.0 (black), T62K variant at pH 9.6 (blue), and T62K variant at pH 6.6 (red). The pK_a value of Lys-62 in the single variant is 8.1. Resonances corresponding to the side chain of Lys-62 are labeled. Assignments of Cα and Cβ resonances were confirmed using HNCACB and CBCACONH spectra. In the case of KE, this was necessary since the Cβ resonance is barely above noise.

Table S4.1. Thermodynamic stabilities ($\Delta G°_{H2O}$) of EK and KE variants. Errors are standard errors from triplicate measurements (two biological replicates). Units are in kcal/mol.

Variant	$\Delta G°_{H2O, pH\ 5}$	$\Delta G°_{H2O, pH\ 6}$	$\Delta G°_{H2O, pH\ 7}$	$\Delta G°_{H2O, pH\ 8}$	$\Delta G°_{H2O, pH\ 9}$
T62K/V66E	4.61 ± 0.08	5.00 ± 0.02	5.18 ± 0.04	5.21 ± 0.02	5.20 ± 0.08
T62E/V66K	1.81 ± 0.07	2.17 ± 0.05	2.52 ± 0.03	2.59 ± 0.07	2.74 ± 0.07

Table S4.2. Differences in stability between variant and background proteins ($\Delta\Delta G°_{H2O}$)**.** Errors are estimated using errors from table S1 and assuming a standard error of 0.2 for measurements of background protein. Units are in kcal/mol.

Variant	$\Delta\Delta G°_{H2O, pH\ 5}$	$\Delta\Delta G°_{H2O, pH\ 6}$	$\Delta\Delta G°_{H2O, pH\ 7}$	$\Delta\Delta G°_{H2O, pH\ 8}$	$\Delta\Delta G°_{H2O, pH\ 9}$
T62K/V66E	-7.2 ± 0.2	-6.7 ± 0.2	-6.7 ± 0.2	-6.7 ± 0.2	-6.3 ± 0.2
T62E/V66K	-10.0 ± 0.2	-9.5 ± 0.2	-9.4 ± 0.2	-9.3 ± 0.2	-8.8 ± 0.2

Table S4.3. *m*-values from guanidinium hydrochloride melts. Errors are standard errors from triplicate measurements (two biological replicates). Units are in kcal/mol/M.

Variant	$m_{, pH\ 5}$	$m_{, pH\ 6}$	$m_{, pH\ 7}$	$m_{, pH\ 8}$	$m_{, pH\ 9}$
T62K/V66E	5.71 ± 0.10	5.93 ± 0.03	5.99 ± 0.04	6.01 ± 0.02	6.18 ± 0.07
T62E/V66K	5.51 ± 0.15	6.25± 0.02	6.31 ± 0.03	6.30 ± 0.04	6.26 ± 0.09

Table S4.4. Crystallographic statistics for SNase variants T62E and V66E.

Variant	Δ+PHS T62E	Δ+PHS V66E
Crystallization conditions		
Temperature, K	277.0	277.0
pH	9	6
Buffer	25 mM KH_2PO_4/K_2HPO_4	25 mM KH_2PO_4/K_2HPO_4
Precipitant	22% (w/v) MPD	20% (w/v) MPD
Additives	2 eq: pdTp, 3 eq. $CaCl_2$	2 eq: pdTp, 3 eq. $CaCl_2$
Data collection statistics		
Wavelength, Å	1.10	1.54
Space group	P 1 21 1	P 1 21 1
Unit cell dimensions		
a, Å	31.04	31.18
b, Å	60.51	60.35
c, Å	38.18	38.42
α, °	90	90
β, °	93.23	93.52
γ, °	90	90
Temperature, K	110	110
Radiation source	Sealed tube MicroSource	Sealed tube MicroSource
Resolution range, Å	38.12-1.95 (1.98-1.95)	50.0-1.85 (1.92-1.85)
No. of reflections	10364 (456)	12215 (1226)
Completeness, %	100.0 (100.0)	99.9 (100.0)
Rmerge/Rsigma	0.0257 (0.124)	0.0220 (0.118)
Average I/σ (I)		
Redundancy	14.79 (10.8)	10.0 (6.9)
Wilson B, $Å^2$	28.2	26.25
Refinement statistics		
Resolution range, Å	38.12-1.95 (2.00-1.95)	38.35-1.85 (1.894-1.846)
No. of reflections	9827 (707)	11555 (785)
R_{work}	0.1723	0.1722
R_{free}	0.2186	0.2179
RMS distance for ideal geometry		
Bond, Å	0.019	0.017
Angles, °	1.959	1.815
Average B-factors		
Protein (no. of atoms)	21.74 (1050)	22.66 (1097)
Solvent (no. of atoms)	24.72 (82)	23.20 (88)
Ion/Ligand (no. of atoms)	17.10 (26)	15.59 (26)
PDB Code	6EEG	5EGT

Table S4.5. Crystallographic statistics for variants T62K/V66E and T62E/V66K.

Variant	Δ+PHS T62K/V66E	Δ+PHS T62E/V66K
Crystallization conditions		
Temperature, K	277.0	293.0
pH	9	7
Buffer	25 mM KH_2PO_4/K_2HPO_4	100 mM HEPES
Precipitant	40% (w/v) MPD	29% (w/v) Jeffamine ED-2001
Additives	2 eq: pdTp, 3 eq. $CaCl_2$	
Data collection statistics		
Wavelength, Å	1.54	1.54
Space group	P 1 21 1	P 63
Unit cell dimensions		
a, Å	31.13	71.79
b, Å	60.43	71.79
c, Å	38.45	56.80
α, °	90	90
β, °	93.86	90
γ, °	90	120
Temperature, K	110	110
Radiation source	Sealed tube MicroSource	Sealed tube MicroSource
Resolution range, Å	50-1.70 (1.76-1.70)	62.18-2.20 (2.28-2.20)
No. of reflections	15727 (1579)	8555 (860)
Completeness, %	100.0 (100.0)	99.8 (99.8)
Rmerge/Rsigma	0.018 (0.108)	0.023 (0.161)
Average I/σ (I)		
Redundancy	10.0 (6.9)	11.0 (6.9)
Wilson B, $Å^2$	21.48	42.05
Refinement statistics		
Resolution range, Å	38.36-1.70 (1.74-1.70)	62.18-2.20 (2.257-2.20)
No. of reflections	14905 (1104)	8146 (596)
R_{work}	0.1746	0.2538
R_{free}	0.2032	0.3086
RMS distance for ideal geometry		
Bond, Å	0.017	0.023
Angles, °	1.813	2.259
Average B-factors		
Protein (no. of atoms)	14.51 (1093)	41.70 (964)
Solvent (no. of atoms)	19.61 (106)	28.50 (7)
Ion/Ligand (no. of atoms)	8.77 (26)	
PDB Code	5KYI	6EEK

Chapter 5

Electrostatic allostery and cooperativity in proteins

5.1 Abstract

pH dynamics in the cell plays an important role in the regulation of biochemical processes. As a chemical signal, pH acts primarily through H^+ binding at ionizable groups in proteins. Because these binding processes change the charge state of the protein, mechanisms of signal transduction involving H^+ are usually interpreted in terms of Coulomb communication between charges. Here we use NMR spectroscopy and equilibrium thermodynamic experiments to examine how charge effects are actually relayed through the protein matrix in staphylococcal nuclease. Strong coupling was detected between buried Lys residues and surface Glu and Asp residues, of magnitude comparable to short-range Coulomb interactions between surface residues. However, we found no evidence that the coupling between ionizable groups was mediated by direct, pairwise Coulomb interactions. Instead, the coupling is governed by conformational reorganization driven by ionization of one of the groups. These observations challenge the view of proteins as materials with a physically meaningful dielectric constant. They suggest an approach for tuning the affinity and cooperativity of H^+ binding by targeting allosteric mechanisms. These observations offer a novel interpretation of complex pH-dependent behavior of proteins in terms of the pH-redistribution of the conformational ensemble, likely related to the imperative for charges to always be hydrated in biological systems.

5.2 Introduction

Tight regulation of cellular pH is a property shared by all living cells[1]. Not surprisingly, dysregulation of pH homeostasis is a hallmark of many pathological conditions, including cancer and neurodegeneration[2]. Management of pH in the cellular milieu is central to energy transduction[3], to cellular organization[4] and condensate formation[5], and to the regulation of essential biochemical pathways in metabolism[6] and cell signaling[7]. The main way in which intracellular pH dynamics exerts its influence is by affecting fundamental properties of proteins that are governed by their charge state[8]—H^+ are bound or released by proteins when pH changes and their charge changes concomitantly. To understand how cellular pH dictates the fate of cells, and to engineer novel proteins that can effectively interface with biological systems[9], it is necessary to understand the physical determinants of the pK_a values of ionizable residues in proteins. The pK_a values of these residues are sensitive to the presence of other charges in the protein, generally through Coulomb interactions. In most electrostatic models which attempt to calculate the pK_a values of ionizable residues, the protein is treated as a dielectric material whose polarity and polarizability modulate the magnitude of the hydration of charges and of the Coulomb interactions between them. To critically evaluate these models, we have previously examined communication between surface charges[10] and between internal charges[11,12]. Here we examine the character of interactions between internal and surface charges.

Internal ionizable groups[13] are relatively rare but they represent an important motif for pH-sensing[7], and for energy transduction governed by H^+ and inorganic ions[14] and by e^- transfer[15]. Buried ionizable residues can have anomalous pK_a values that are shifted significantly relative to their normal values in water. Owing to the loss of favorable hydration when in the buried state, these shifts are in the direction that favors the neutral state[16,17]; it is not unusual to find buried Glu,

Asp and Lys residues that are neutral under physiological conditions[18,19]. Despite their importance in a wide range of fundamental biochemical processes, accurate prediction of the pK_a values of these essential residues remains extraordinarily challenging[20,21,22]. A set of variants of staphylococcal nuclease with buried Lys residues at 25 internal sites is being used by many groups to determine why computational methods fail to reproduce experimentally determined pK_a values of buried residues[23,24,21,22,25]. The pK_a values of each residue was measured experimentally using linkage thermodynamics analysis of the pH-dependence of the protein stability[17]. The pK_a values of most of the internal Lys residues were depressed relative to their values in water, some by more than 5 pH units. NMR spectroscopy and X-ray crystallography studies have shown that in all cases studied, the Lys side chains are buried when they are neutral and that ionization of the internal residue triggers conformational reorganization into a state where the buried Lys comes in contact with bulk water[26,27]. The pK_a values of the Lys residues report on the pH at which the protein undergoes a pH-dependent conformational transition; they reflect the energy gap between fully folded closed states where the Lys is buried and neutral, and partially or globally unfolded open states where the Lys is exposed to solvent and charged. These results illustrate why structure-based electrostatics calculations have failed to capture functional electrostatic effects and highlight the two challenges facing structure-based pK_a calculation methods: accurate prediction of alternative states of proteins, and accurate calculation of the free energy distance between states.

The properties of individual ionizable residues are governed by differences of hydration of single charges in water and in protein. To examine contributions from Coulomb interactions, here we have examined the communication between buried and surface ionizable groups. Evidence of this communication was detected originally in a significant feature of the pH dependence of stability of many of the SNase variants with internal Lys residues. In 12 of 25 cases, there is

evidence of a secondary transition in the pH-dependence of the protein's thermodynamic stability, suggesting that the pK_a values of surface residues in the protein are influenced by the internal Lys residue[17] (Fig. 1A). These effects were interpreted previously as arising from Coulomb interactions strengthened by the low dielectric protein medium. They were proposed as a mechanism by which alterations of surface charge distribution could influence the functional properties of buried residues[28]. However, the recent evidence of pH-driven reorganization coupled to the ionization on individual internal Lys residues[27] suggests that the pK_a values of surface residues are shifted as ionization of the internal residues shifts the protein into an alternative state where the microenvironments of surface residues are disrupted. To address this possibility, NMR spectroscopy and linkage thermodynamics were used to determine the physical origin of coupling between buried and surface charges. We found no evidence of strong Coulomb interactions between buried and surface residues in this protein. Instead, the apparent, strong medium-to-long-range couplings reflect conformational reorganization of the protein that lead to shifts in the pK_a values of surface groups. These effects influenced both the affinity and cooperativity of the buried Lys residues, providing important insight into the physical basis of H^+-coupled allostery and cooperativity in natural systems.

5.3 Results

5.3.1 Titration of surface Glu/Asp residues in partially unfolded states of SNase

NMR spectroscopy was used to identify cases where surface Asp and Glu residues were perturbed by ionization of internal Lys residues. ^{13}C-^{13}C correlation CBCGCO (Asp Cβ-Cγ, Glu Cγ-Cδ) spectra[29] were collected for 18 variants of SNase with internal Lys residues. Spectra were collected at pH values where the Lys residues were charged and the protein reorganized into states

that are open relative to the fully folded state (Fig. 1B, S1). In most cases, large ^{13}C chemical shift perturbations (CSPs) or resonance intensity loss of the surface Asp or Glu side chains was observed, consistent with perturbation of structural environment or dynamics or both. These data are consistent with crystal structures showing disrupted hydrogen bonding interactions in the open states of SNase[30,29] (Fig. 1C-E). Extensive networks of hydrogen bonds on the surface of SNase (Fig. 1F) are expected to be perturbed by many of the pH-driven structural responses coupled to ionization of buried Lys residues. Two variants, with Lys-34 and Lys-62, were selected for further study based on thermodynamic data and the preliminary NMR survey. The pK_a^{App} values of Lys-62 was 8.1 when measured with linkage thermodynamics[17] and 8.2 when measured by global fitting of 11 backbone NH resonances in ^{1}H-^{15}N HSQC titrations (Hill coefficient of 0.94, Fig. S2). The pK_a^{App} of Lys-34 was 7.1 when determined by linkage thermodynamics[17]; however, twelve $^{1}H_N$ and ^{15}N resonances showed pH dependent changes with pK_a values of 6.6 and 5.8 and Hill coefficients between 0.5-0.8. Many of these titrations contain ambiguities owing to peak overlap or the influence of multiple titration events; three resonances with high quality, complete titrations were chosen for global fitting to obtain a pK_a^{App} of 6.6 and a Hill coefficient of 0.70 (Fig. 4B). As described below, the discrepancy between the NMR and linkage derived values reflects the influence of pK_a shifts of surface residues.

High resolution CBCGCO titrations were collected to follow the pH-dependence of the $^{13}C\gamma/\delta$ chemical shifts of all Asp/Glu residues, respectively (Fig. 2, S3-4). The pK_a values of the surface Asp and Glu residues in the background protein have been measured previously[31]. ^{1}H-^{13}C HSQC experiments were used to follow the $^{1}H\varepsilon$ resonance of His residues (Fig. S5). Accurate determination of the pK_a of Asp and Glu residues is complicated by the fact that the variants are less stable than the background protein and acid unfold near the pK_a values of most of the Asp/Glu

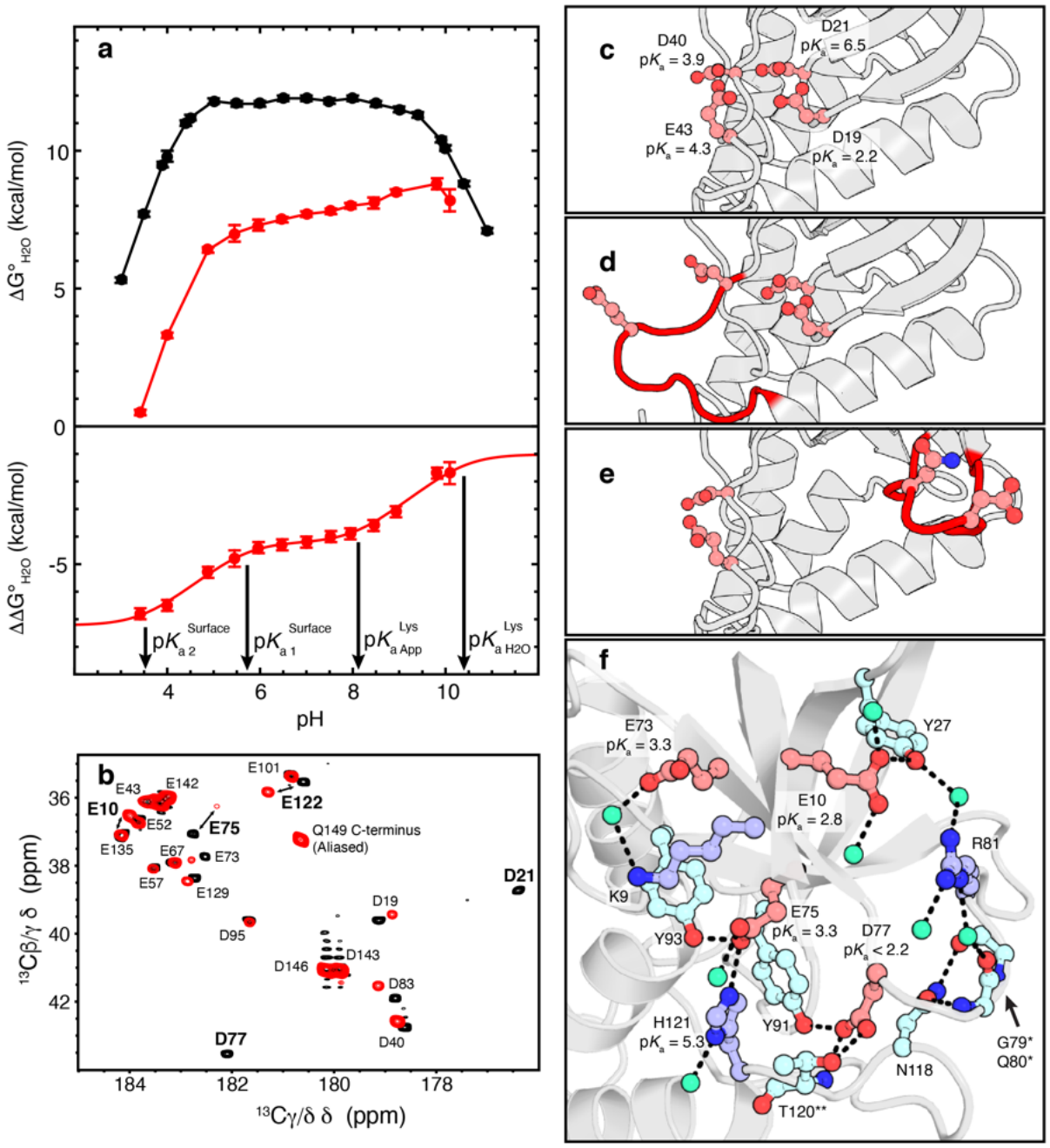

Fig. 1. Evidence of coupling between internal Lys residues and surface ioinizable groups in SNase. (a) (Top) pH-dependence of the thermodynamic stability ($\Delta G°_{H2O}$) of background protein (black) and variant with Lys-62 (red). Solid lines are shown solely to guide the eye. (Bottom) Difference in stability between background and variant proteins ($\Delta\Delta G°_{H2O}$). Solid line is fit to a linkage equation (Eq. 3 in ref 22). The inflection points reflect the pK_a of Lys in water ($pK_a^{Lys}{}_{,H2O}$)

and the apparent pK_a of the buried Lys (pK_a^{App}), and at lower pH values they reflect the perturbation of the pK_a values of surface residues. **(b)** Representative ^{13}C-^{13}C CBCGCO spectra of background protein (black, pH 5.30) and variant with K34 (red, pH 5.35). Residues with significant weighted chemical shift perturbations (greater than 0.4 ppm) are shown in bold. (**c-e**) Crystal structures showing arrangement of the four active site acidic residues in the background protein (**c**, PDB ID: 3BDC) and in partially unfolded states (**d**, variant with Arg-109, PDB ID: 3D4W and **e**, variant with Asn-21/Glu-23, PDB ID: 3TME). **(f)** Crystal structure of background protein highlighting the networks of hydrogen bonds spanning the most N-terminal region of secondary structure (β-1, residues 8-18) to the most C-terminal secondary structure element (α-3, residues 121-135). Water molecules are shown as green spheres.

residues. Therefore, the pK_a values can only be estimated as upper estimates by making assumptions using the protonated chemical shifts of the background protein or the expected range of chemical shift changes for acidic side chains upon protonation. In the variant with Lys-62 the pK_a values of two Asp residues on a β-turn in the active site (Asp-19 and 21), which is disrupted by ionization of Lys-62, were significantly shifted relative to their values in the background protein. The pK_a value of Asp-19, which is depressed to 2.2 in the background protein relative to the normal pK_a of Asp in water, appears to be have a more normal pK_a in the variant with Lys-62. The opposite is true for Asp-21, which has an elevated pK_a value of 6.5 the background protein but is depressed to a normal value near 4 in the variant. The direction of the pK_a shift of Asp-19 (closer to the normal value of Asp in water, favoring the neutral state) is inconsistent with a direct, favorable Coulomb interaction between Lys-62 and Asp-19, which would have depressed it to favor the charged form of Asp-19. However, the observation that the pK_a values of both Asp-19 and Asp-21 are closer to their values in water than in the background protein is consistent with structural reorganization disrupting the hydrogen bonding networks in the active site and exposing the side chains to water. The pK_a values of the other surface Asp, Glu, and His residues are unaffected by ionization of Lys-62; only those in close proximity to the mutation site display small shifts in pK_a (smaller than 0.3 units) or changes in Hill coefficients.

In the variant with Lys-34, three surface residues (Glu-10, Asp-21, Asp-77) had detectable changes in pK_a (≥0.5). A large transition was observed for Asp-77, which resides on the loop between β-4 and β-5, that parallels the titration of Lys-34 (pK_a^{App} value of 6.6, Hill coefficient of 0.71). Asp-77 begins to undergo another titration event due to protonation below pH 5; in the background protein the beginning of this transition is not observed until the onset of acid unfolding below pH 2.5, suggesting that the pK_a value of Asp-77 is shifted significantly by the ionization of

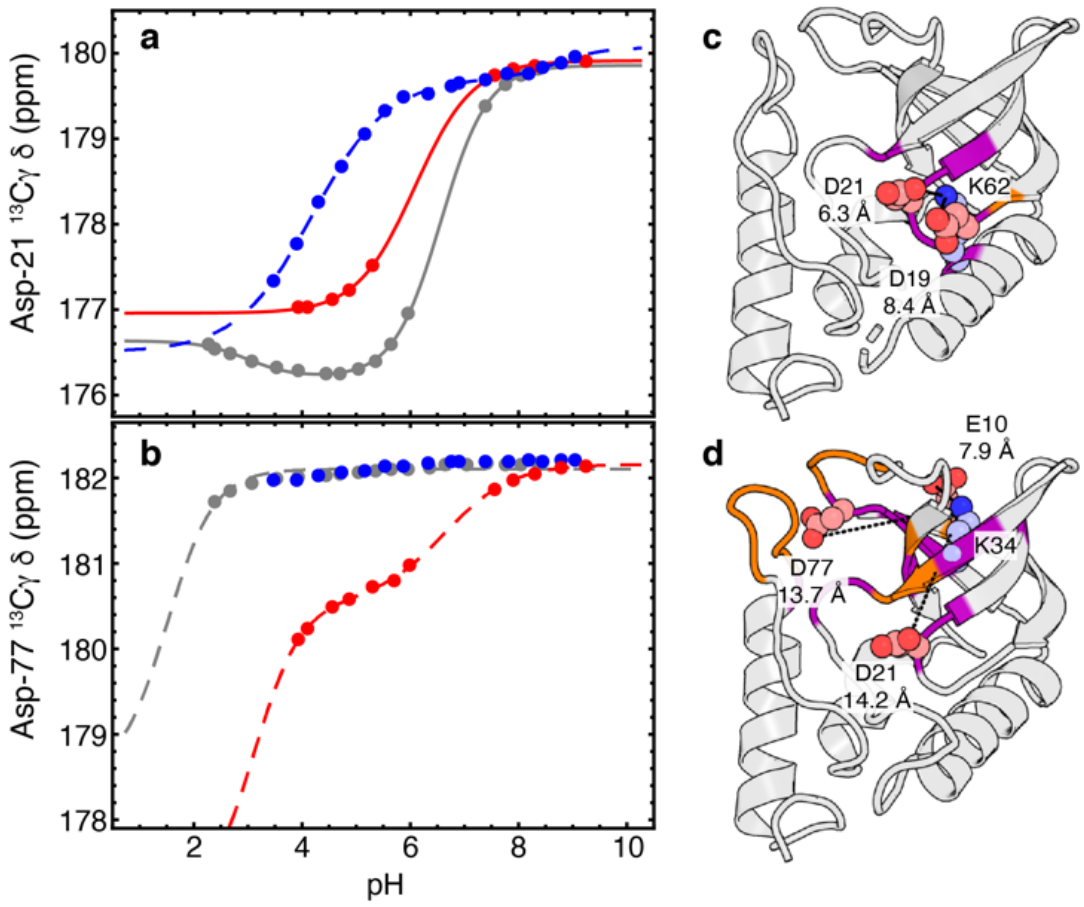

Fig. 2. Site-specific titration curves of surface residues measured with ^{13}C NMR spectroscopy. (a) pH-dependence of ^{13}Cγ chemical shift of Asp-21 in the background protein (gray) and in variants with Lys-34 (red) or Lys-62 (blue). Solid gray and red lines represent fits to a modified Hill equation for single or two-site binding schemes (Eq. 1 or 2 in ref 28); dashed blue line represents the case where the pK_a value is only approximate because the protein acid unfolds before the titration reaches a protonated baseline. **(b)** pH-dependence of ^{13}Cγ chemical shift of D77. Color scheme same as in (a). **(c)** Crystal structures of the variant with Lys-62 **(**PDB ID: 3DMU**)** and **(d)** crystal structure of the variant with Lys-34 **(**PDB ID: 3ITP**).** In **(c) and (d) the** regions showing large backbone CSPs or broadening are shown in purple or orange, respectively.

Distances between Nζ atom of Lys-62 or Lys-34 and closest Oδ/ε atoms of surface Asp or Glu residues whose pK_a values are affected by ionization of buried Lys residues are shown.

Lys-34. An estimated pK_a shift of 1.5 to 2 units corresponds to a shift in the H^+-binding energy of approximately 2-3 kcal/mol, or a change in binding affinity of 30 to 100-fold. What is remarkable is that in the crystal structure the H^+-binding side chain of Asp-77 is almost 14 Å away from Lys-34. Glu-10, which has a highly depressed pK_a value in the background protein, did not display a transition corresponding to the titration of Lys-34, but its protonation transition point occurs at higher pH values relative to the background protein, suggesting an elevated pK_a. Coulomb interactions cannot explain the direction of the shifts in the pK_a of Glu-10 and Asp-77. As in the variant with Lys-62, the pK_a of Asp-21 is more normal (depressed) in the variant with Lys-34, although to a lesser extent; however, the Hill coefficient is significantly lower than in the background protein (0.81 versus 1.00). Although Glu-73, Glu-75, Asp-83 and Glu-122 show evidence of small transitions coupled to the titration of Lys-34 (0.2-0.5 ppm), their protonation transitions occur at similar points as in the background protein, suggesting their pK_a values are not significantly affected by the ionization of Lys-34. This provides important structural detail into the nature of the conformational transition coupled to ionization of Lys-34. The hydrogen bond networks involving Glu-75 and Asp-83 are not significantly disrupted, but the network between them, involving Asp-77, is likely to be perturbed, as is the hydrogen bond network extending from Arg-81 to Tyr-27 and Glu-10. Consistent with this interpretation, RCI analysis of backbone chemical shifts suggests residues 77-82 were unfolded in the open states of the variant with Lys-34 and in other variants that exhibit similar structural perturbations, but the backbone of Glu-75 and Asp-83 appear to be fully folded[27]. Importantly, the pK_a of His-121, which is hydrogen bonded to Glu-75, shifts only slightly in response to ionization of Lys-34 (from 5.3 to 5.5), providing

further evidence that the structural response to ionization of Lys-34 is largely limited to the loop regions between secondary structures. The resonance of His-8 was observed in two conformations in the ^{1}H-^{13}C HSQC in all variants; this was not originally observed in 1D titration experiments[10]. The pK_a value of the major form of His-8 in the variant with Lys-34 was slightly depressed relative to the background protein (6.3 versus 6.4); this side chain is within 4 Å of Glu-10.

5.3.2 Effects of conformational reorganization on apparent pK_a values

To distinguish how Coulomb interactions and structural reorganization contribute towards the normalization of the pK_a of Asp-21 in the variant with Lys-62, the D21→N21 substitution was introduced into variants with Lys-62 or Glu-62. In the variant with Asn-21/Lys-62, the apparent pK_a value of Lys-62 was 7.1 ± 0.2 and 6.7 ± 0.1 by linkage thermodynamics and ^{13}C NMR (indirectly measured from secondary titrations of $^{13}C\beta$ and $^{13}C\gamma$ resonances of Asp-19), respectively. This represents a decrease of 1.0 - 1.5 pH units relative to the pK_a of Lys-62 in the presence of Asp-21 (Fig. 3, S6). This is consistent with a favorable Coulomb interaction worth 1.4 to 2.0 kcal/mol. This represents a strong interaction, but it is significantly weaker than the interaction energy measured by the pK_a shift of Asp-21 in the variant with Lys-62 (1.36 * (6.5 – 4.0) = 3.4 kcal/mol at 298K). To determine whether the interaction is Coulombic in nature, the Asn-21 substitution was also introduced into the variant with Glu-62. The pK_a value of Glu-62 alone is 7.7 ± 0.1, which is high relative to the pK_a of 4.5 for Glu in water[16]. If the interaction between Lys-62 and Asp-21 had a significant favorable Coulomb component, then the absence of an unfavorable Coulomb interaction with Asp-21 in the variant with Glu-62 should normalize the pK_a value of the buried Glu (shift closer to 4.5). Importantly, ionization of Glu-62 perturbs the same region of the protein as Lys-62; however, more extensive line broadening in $^{13}C\alpha$-^{13}C'

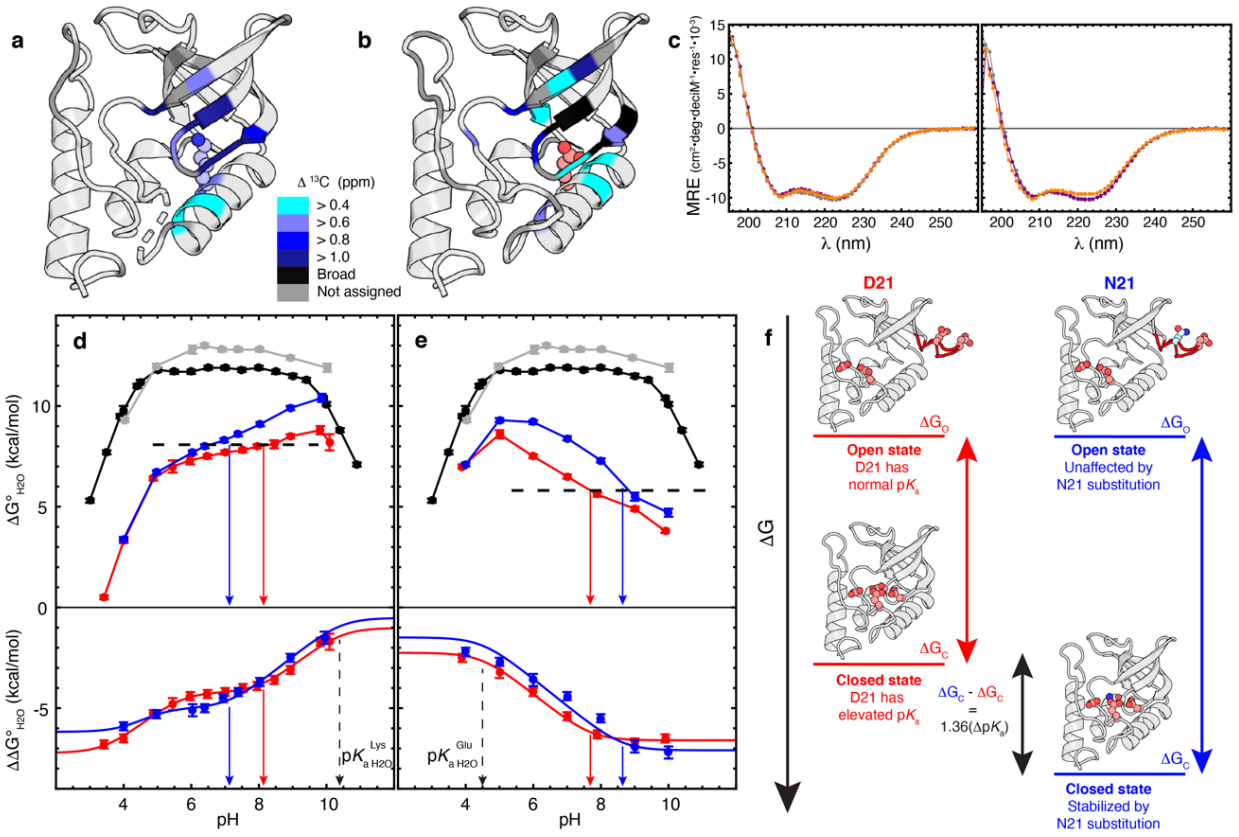

Fig. 3. Apparent coupling between buried and surface ionizable groups does not reflect Coulomb interactions. (a-b) Weighted pH-dependent backbone $^{13}C\alpha/^{13}C'$ CSPs for variants with Lys-62 (**a**, pH 9.6 versus 6.6) and Glu-62 (**b**, pH 9.9 versus 6.6, PDB ID: 6EEG). Note that in the crystal structure Asp-21 is equidistant from Lys-62 and Glu-62. **(c)** CD spectra of variants with Lys-62 (left panel) and Glu-62 (right panel). Spectra of the fully folded, closed state are shown in purple (pH 9.5 for Lys-62, pH 6 for Glu-62) and spectra for the open states are shown in orange (pH 7 for Lys-62, pH 9 for Glu-62). Background spectra are shown in gray for reference at pH 7, though the spectra are invariant between pH 5-9. **(d-e)** pH-dependence of the thermodynamic stability for variants with **(d)** Lys-62 and Asn-21/Lys-62 and **(e)** Glu-62 and Asn-21/Glu-62. Data for the background protein and the variant with Asn-21 are shown in black and gray, respectively. Data for Lys-62 or Glu-62 alone are shown in red, data for the double variant (Lys-62 or Glu-62 with Asn-21) are shown in blue. Solid lines are meant to guide the eye in top panels ($\Delta G°_{H2O}$)

whereas they represent fits of pH-dependence to linkage model in the two bottom panels ($\Delta\Delta G°_{H2O}$). The secondary titration observed in the pH-dependence of Asn-21/Lys-62 is caused by shifts in the pK_a value of Asp-40 and Glu-43 (Fig. S7). The dashed black lines in $\Delta G°_{H2O}$ plots highlight the stability at pK_a value, which represents the stability of the open state that is common to the single and double K62/E62 variants. **(f)** Thermodynamic model illustrating origins of coupling between D21 and K/E62.

correlation spectra and larger pH-dependent changes in the circular dichroism spectra in the variant with Glu-62 suggest the magnitude of the structural response is slightly greater in response to ionizing Glu versus Lys at position 62. In the variant with Asn-21/ Glu-62, the pK_a value of Glu-62 is further elevated to 8.6 ± 0.4, inconsistent with the loss of an unfavorable Coulomb interaction between Glu-62 and Asp-21. Although broadening and complex, multiphasic titration behavior prevented pK_a determination of Glu-62 by NMR in either the single or double variants, the pH-dependence of $^{13}C\alpha/^{13}C'$ resonance broadening and CSPs were consistent with the apparent pK_a value of Glu-62 being approximately 1.0 – 1.5 pH units higher in the presence of Asn-21 (Fig. S5.6).

The fact that the D21→N21 substitution shifts the pK_a values of both Lys-62 and Glu-62—two residues with charge of opposite signs—in the direction that favors the neutral state demonstrates that the coupling between these residues is not governed by a Coulomb interaction. To understand the physical origin of the apparent interaction between positions 62 and 21 it was necessary to examine the global thermodynamic stabilities of the proteins. In the background protein and in the variants Lys-62 or Glu-62, the D21→N21 substitution is stabilizing in a pH-dependent manner, the stability increase being greater above pH 6: at pH-values where Lys-62 or Glu-62 are neutral, the substitution stabilizes the protein. However, the thermodynamic stabilities of variants with Lys-62 and Glu-62 at their respective apparent pK_a values, where they transition to the open, partially unfolded or reorganized states, are almost identical ($\Delta G_O^{K62} = \Delta G_O^{N21/K62} =$ 8.0 kcal/mol, $\Delta G_O^{E62} = \Delta G_O^{N21/E62} =$ 5.6 kcal/mol). This suggests a coupling mechanism between Asp-21 and residue 62 that involves selective stabilization of a specific state in the conformational ensemble (Fig. 3F). In the closed state, the D21→N21 substitution is stabilizing in a pH-dependent manner because the repulsive interactions between Asp-21 and the surrounding active site acidic

residues are removed. In contrast, in the open state, the pK_a values of all of the acidic active site residues are close to their normal values and in this case the D21→N21 substitution does not remove repulsive interactions, therefore the stability of the open state is unaffected. Importantly, the fact that the stability of the open state does not significantly shift upon neutralizing the charge of Asp-21 suggests that any Coulomb interactions between Asp-21 and Lys-62 or Glu-62 are weak. This interpretation explains the nonreciprocal coupling between Asp-21 and Lys-62 observed via pK_a shifts of Asp-21 when Lys-62 is ionized by pH versus those of Lys-62 when Asp-21 is neutralized through substitution; these two cases report on different phenomena. In the first instance, the pK_a value of Asp-21 shifts when Lys-62 is ionized because the network of Coulombic and hydrogen bonding interactions that elevate the pK_a value of Asp-21 are disrupted when Lys-62 is ionized. The second instance reflects the consequences of the selective mutational stabilization of the closed state.

5.3.3 Cooperativity in coupling between conformational reorganization and H^+ binding

Asp-21 and Lys-34 are significantly further apart than Asp-21 and Lys-62. The coupling between 21 and 34 introduces small pK_a shifts but it has a noticeable effect on the cooperativity of the pH-dependent conformational transition. Structural analysis of the Lys-34 variant suggests that ionization of Lys-34 could disrupt the sidechain of Arg-35 that makes structurally important hydrogen bonds near the active site. Disruption of Arg-35 would affect the pK_a values of the active site acidic residues through both Coulomb and structural effects (Fig. 4A). Consistent with this interpretation, the pK_a value of Asp-21 in the Arg35→Gln-35 variant drops from 6.54 to 6.05[32]. The Hill coefficient of the titration coupled to Lys-34 provides evidence of negative cooperativity (n = 0.70); reciprocal effects were observed on Asp-21, which had a decreased Hill coefficient of

0.81 relative to its value of 1.00 in the background protein. The other surface residues that titrate in the same pH range where Lys-34 titrates are His-8 and His-121. These two histidine residues do not show significant changes in Hill coefficients in the variant with Lys-34 relative to the background protein. The Hill coefficient in the variant with Asn-21/ Lys-34 was significantly closer to unity (n = 0.92), confirming that the negative cooperativity observed in the H^+-coupled structural transition coupled to the ionization of Lys-34 is caused by Asp-21. Although the direction of the pK_a shift of Asp-21 in the variant with Lys-34 could be consistent with a long-range Coulomb interaction, the observed negative cooperativity is not, as opposite charges titrating with similar pK_a values should result in positive cooperativity[33]. To dissect the contributions of Asp-21 and Lys-34 to the observed structural transition, the backbone HSQC data can be fit to a two-site model to extract pK_a values of 7.1 and 5.8 (this model does not account for the structural change the protein undergoes, where the two residues would have different pK_a values in different states).

To resolve the inconsistencies between different interpretations, the pH-dependent backbone $^1H_N/^{15}N$ chemical shift data were fit to a simple thermodynamic model[34,35] that accounts for the different pK_a values the buried Lys and surface Asp residue can have in the two states (Fig. 4B). No assumptions were made about electrostatic interactions between the two residues; the charge state of the residues was not explicitly modeled, as the model treats protonation as a simple ligand-binding reaction with known pK_a values. The effects of proton binding to each residue on the conformational transition would be determined by the energetic effects of differing ligand binding affinities on the stabilization of the closed versus open states. In this regard, this model is a quantitative representation of the qualitative diagram shown in Fig. 3F. Accurate extraction of

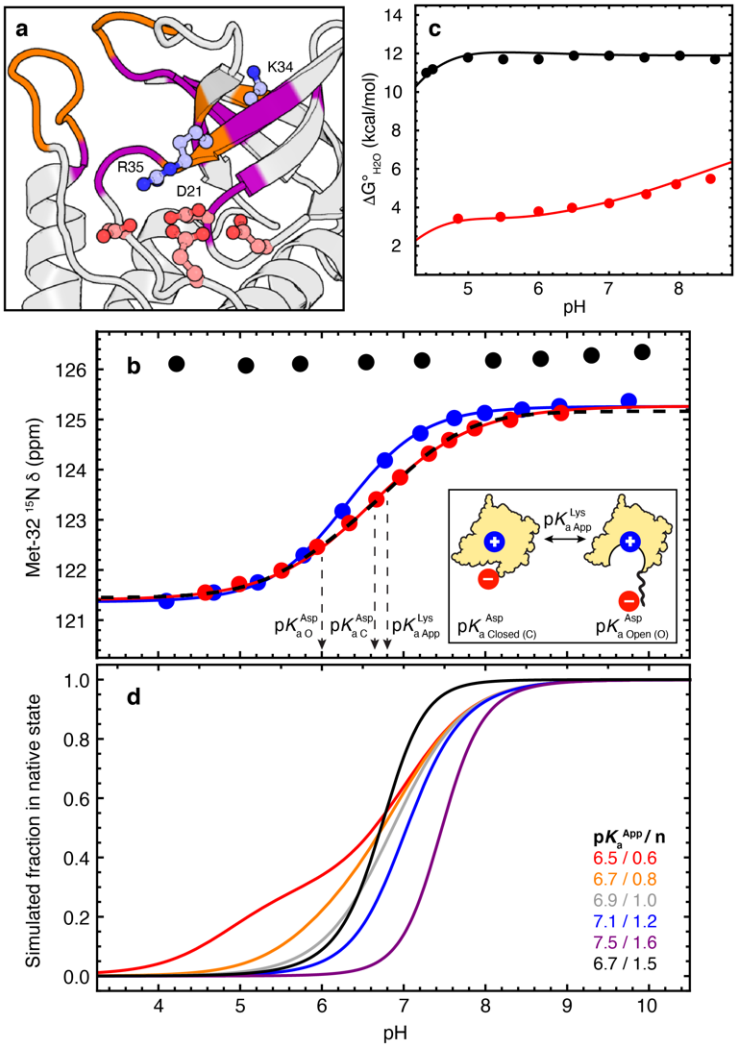

Fig. 4. Cooperative interactions between buried and surface ionizable groups. (a) Crystal structure of variant with Lys-34 showing proximity between Arg-35 and Asp-21, providing a structural basis for understanding how pH-dependent reorganization coupled to ionization of Lys-34 influences the pK_a value of Asp-21 by disrupting interactions mediated by Arg-35. Backbone residues showing large backbone CSPs or broadening are shown in purple or orange, respectively. **(b)** pH-dependence of backbone chemical shifts (^{15}N resonance of M32) in the background protein (black) and in variants with Lys-34 (red) or Asn-21/Lys-34 (blue). This resonance is representative since its chemical shift is pH-insensitive in the background protein, and the shifts in the closed and open states are identical in both single and double variants. This demonstrates that the coupling is

thermodynamic in nature, and not caused by secondary effects from chemical shifts. Solid lines represent fits to Hill equation while the dashed black line represents fit to a thermodynamic model accounting for multiple conformations with different pK_a values. Inset: Visual representation of model used to account for pK_a value shifts caused by structural reorganization. **(c)** pH-dependence of thermodynamic stabilities of the background protein (black) and of the variant with Lys-34 (red). Solid lines represent simulated pH-dependent stabilities based on pK_a values of surface Asp and Glu residues measured by NMR spectroscopy[28]. The simulations for background protein uses a two-state model (folded and unfolded protein only); the pK_a of the folded protein are those measured with NMR spectroscopy and the pK_a of the unfolded state are those of ionizable residues in water (described in Methods). Simulation for the variant with Lys-34 introduces a third state, an open state, and the pK_a values assumed for this state are those measured with NMR spectroscopy. A three-state model can fit the two-state thermodynamic unfolding data because the closed and open states of the variant with Lys-34 are indistinguishable based on identical fluorescence emissions at 326 nm, and highly similar m-values. **(d)** Simulations showing that the pH_{mid} and cooperativity of a pH-driven conformational transition can be tuned by modulating the pK_a value of surface residues and the energy gap between conformational states. The likelihood that positive or negative cooperativity is observed depends on the direction of the pK_a shift of the second residue in response to ionization of the first. The simulations assume the first residue driving the transition has a pK_a^{Open} of 10 and pK_a^{Closed} of 3; the pK_a^{Closed} of the second residue is 6.5 in all simulations. The energy gap between states was set to -4.3 kcal/mol for all simulations, except the black line, which was set to -6 kcal/mol. The pK_a^{Open} of the second residue is varied in the simulations, and set to 5.0 (red), 6.0 (orange), 6.5 (gray, control assuming no pK_a shift between states for second residue), 7.0 (blue), and 8.0 (purple, black).

the pK_a values of Asp-21 in the closed (pK_a = 6.6) and open (pK_a = 6.0) states were obtained, which were within 0.1 pH units of the pK_a values measured by ^{13}C NMR for Asp-21 in the closed state (background protein) and open state (variants with Lys-34). The microscopic pK_a values of Lys-34 in the closed and open states could not be accurately resolved (see Methods for details of fitting procedure and analysis), but an apparent pK_a = 6.8 of Lys-34 could be accurately determined. This value represents the pH at which Lys-34 is 50% protonated and 50% deprotonated and is distinct from the value of 6.6 obtained as the midpoint of the conformational transition, which is influenced by both Asp-21 and Lys-34. Importantly, the predicted pH-dependence of the global stability of the protein using the model described in Fig. 4B and the NMR-derived pK_a values of all Asp, Glu, and His residues in each state agree with the experimental pH-dependence of the thermodynamic stability of this protein (Fig. 4C). The ability of this model to reproduce the conformation-specific pK_a values of Lys-34 in closed and open states of Asp-21 and the apparent pK_a value of Lys-34, determined through independent methods, confirmed that the nature of the coupling between Asp-21 and Lys-34 is through redistribution of the conformational ensemble and not through direct Coulomb interactions. This suggests that the origin of the negative cooperativity comes from an energetic tug-of-war between Asp-21 and Lys-34. Depressing the pK_a value of Lys-34 decreases the global stability of the protein as pH decreases, increasing the probability that the partially unfolded open state is sampled; however, sampling the open state leads to a slight pH-dependent stabilization, as the pK_a value of Asp-21 becomes more normal. The independent actions of the two titrating residues on the global thermodynamic stability and on the population of the conformational ensemble leads to the observed negative cooperativity of the pH-dependent transition coupled to ionization of Lys-34. Simulations of pH-driven conformational transitions demonstrate how two non-interacting ionizable residues can modulate the midpoint and

cooperativity (negative or positive) of proton-coupled conformational transitions (Fig. 4D). These effects are comparable to natural cases where pH-driven conformational transitions are triggered by titration of multiple, independent residues acting in concert.

5.4 Discussion

Communication-at-a-distance in proteins, and how they potentiate for cooperative and allosteric interactions, is a fundamental property of proteins that allows them to act as sensors, amplifiers and transducers of information in their environment. Understanding mechanisms of long-range communication, allostery and cooperativity remains an important goal in protein science. When charged atoms are involved it is tempting to interpret mechanism in terms of direct, Coulomb interactions between atoms. Coulomb effects can in principle be very strong across large distances. This is especially true if the interactions happen in a material that can behave as a material with low dielectric constant. The dielectric constants of proteins measured with parallel plate capacitor experiments with fully dehydrated powders are low, in the order of 2 to 4[36,37]. With this kind of a dielectric response, strong Coulomb interactions across long distances in proteins should be possible. Surface ionizable residues seem not to experience the low dielectric environment inside proteins[10,38,39], presumably because at the protein-water interface conformational fluctuations are sufficient to ensure that surface charges are always primarily in bulk water[35,40]. One way to probe the character of long-range electrostatic communication is to examine the interactions between internal and surface residues, where the interactions would happen through the presumably low-dielectric protein matrix.

The results of this study demonstrate how the proton-binding affinities of two ionizable moieties in a protein can be coupled through structural reorganization. Significant thermodynamic

coupling between proton binding sites worth several kcal/mol (at 298 K) were observed at distances of up to 15 Å. Thermodynamic coupling was manifested as shifts in both pK_a values and cooperativity of H^+ binding. Importantly, no evidence of medium to long range Coulomb interactions was observed. Instead, the couplings are governed by the consequence of shifts in pK_a values on the free energy gaps between different conformation states (closed or fully folded and open or partially unfolded) and therefore through how changes in state of ionization shifts the population of conformational states in the ensemble. These phenomena can be described using simple ligand-binding schemes that account for the presence of multiple states with different H^+-binding affinities. Although formally these processes are not allosteric[41], since the ligand is identical in both cases, the physical principles that describe these effects are relevant for understanding pH-driven allosteric and cooperative effects in natural systems.

The conformational reorganization triggered by the ionization of Lys-34 and Lys-62 appears to impact the properties of surface residues through local unfolding. Local unfolding can disrupt the networks of polar and ionizable surface residues that would normally shift the pK_a values of surface residues relative to their normal values in water. The affinity or cooperativity of H^+-binding reactions can be tuned with secondary residues that themselves do not participate in H^+-binding equilibria but which impact the energy gaps between states. This suggests a generalizable mechanism useful to tune pH-dependent coupled binding-reorganization events, through the disruption of the microenvironments of distal ionizable residues. Evidence of pH-dependent allosteric regulation through the disruption of surface hydrogen bond and Coulomb networks has been observed in natural systems[42]. The detailed NMR and crystallographic illustrate how disrupted hydrogen bonding interactions can affect the affinity or cooperativity of H^+-binding to distant sites. It is important to recognize that the mechanism by which these residues are coupled

is through the effect each individual residue has on the pH-dependence of the global thermodynamic stability, demonstrating how a global parameter can influence local properties in proteins. This point is of great importance, as it demonstrates that functional properties such as cooperativity in H^+-binding or release cannot be attributed to simple structural motifs observed in static structures, such as low-barrier hydrogen bonds[43]. Additionally, no simple linear, mechanical model can explain our results self consistently in terms of direct interactions; it is necessary to invoke a statistical thermodynamic model that invokes pH-driven redistribution of the conformational ensemble[35].

Experimental characterization of mechanisms of communication that are mediated by pH-triggered conformational change will be challenging. In the SNase variants with Lys-34 and Lys-62 there is no evidence of pH-dependent reorganization in low-resolution biophysical methods such as circular dichroism or fluorescence spectroscopy[17]. Rarely is crystallography useful; only three or four open states of SNase have been observed in the almost 300 solved crystal structures of variants of this protein. Crystallization conditions are likely to suppress partially unfolded states in favor of the fully folded form that crystallizes readily. Importantly, the open states that have been observed in crystal structures have also been observed in solution with NMR spectroscopy. The detection of localized unfolding and site specific pK_a measurements performed in this study required the use of NMR methods that are not yet amenable to the study of more complex proteins. However, the general principle that communication between ionizable sites happens through the redistribution of the ensemble is likely to be relevant in many cases (e.g. catalysis[44,45,46], pH-sensing[47], and pH-dependent formation of cellular substructures and condensates[5,48,49]) where subtle structural reorganization and H^+-binding and release of multiple residues are coupled.

In general, pH effects continue to be rationalized in terms of direct Coulomb interactions between titratable residues[50]. The results of this study suggest that Coulomb effects are not sufficient to explain the origin of many electrostatic effects in proteins. The traditional models of proteins as dielectric cavities with low dielectric constants embedded in a continuum of water of high dielectric constant misrepresents the dielectric response of proteins in solution[38,10,39]. The interactions characterized in this study found no evidence of Coulomb interactions between buried and surface residues, even those that were relatively close in space (< 6.5 Å). This contradicts previous studies that did not account for the possibility that the coupling between buried and surface charges involving conformational reorganization rather than Coulomb interactions[28]. This point has significant implications for structure-based energy calculations that seek to model pH-effects and to predict the pK_a values in proteins. Continuum electrostatic models that assume a static structure cannot possibly account for the behavior of ionizable residues observed in this study. These algorithms can only interpret coupling between ionizable residues in terms of Coulomb interactions that are, incidentally, usually greatly overestimated. More sophisticated algorithms have recently been developed which attempt to account for reorganization explicitly[22]. In particular, a recent pH-replica-exchange molecular dynamics studies has observed evidence of negative cooperativity between a buried Lys and a partially buried His in SNase[51]; however, it was unclear if these effects arose from Coulomb interactions or conformational effects, and the calculated apparent pK_a values are inconsistent with the experimental pH-dependence of the thermodynamic stability of the variant. Given the difficulty in the accurate calculation of free energies, and the computational cost in modeling multiple titrating residues with shifted pK_a values, the experimental data provided in this study will be necessary to benchmark and inform

the next generation of structure-based energy calculations that can account for functionally important allosteric and cooperative effects.

In general, structure-based methods for calculation of pH-dependent processes in proteins are likely to be of limited utility to examine the type of mechanism for electrostatic allostery and cooperativity described by these studies and of relevance in many cellular processes. This would require methods that can predict the myriad states that populate the conformational ensemble, accurate calculation of the free energy gaps between states and of their sensitivity to pH. This remains an extraordinarily challenging thing to accomplish even with model cases, such as SNase.

There is ample evidence that globular proteins, even those such as SNase with well-defined cores, do not behave as materials with low dielectric constant where strong Coulomb interactions across large distances are possible. Instead, the protein behaves as a highly deformable material in which charges rarely experience an environment with low dielectric constant because low-lying alternative states are within access (i.e., energy gap comparable to the energy reflected in shifts in pK_a values) in which the charges fully satisfy their hydration. This has been now observed with SNase from the perspective of surface residues[52], of individual buried residues[27], of buried ion pairs[53], and now from the perspective of interactions between internal and surface residues. In all cases, the character of electrostatic effects in the protein is governed by the energetics of redistribution of the conformational ensemble driven by changes in pH. It could be argued that what has been observed with SNase describes an artificial case that did not evolve to stabilize buried ionizable residues, and that the behavior of natural proteins where internal ionizable residues are found will be different. We do not exclude the possibility that other proteins where buried ionizable residues are found naturally, have evolved the rigidity or stability to preclude the type of reorganization of the ensemble we have observed. It is noteworthy, however, to point out

that in many proteins where electrostatic effects are essential for function (e.g., bacteriorhodopsin[54], ATP synthase[14], transporters[55], most enzymes[56]), the mechanism of energy transduction depends entirely on the type of coupling between ionization of residues and conformational transitions that were observed with SNase. The insights provided by studies with SNase will help us understand the mechanisms by which pH-regulates protein function, and to develop strategies to engineer proteins that use changes pH to regulate function.

5.5 Methods

5.5.1 Sample preparation

Proteins were prepared as described in Chapter 2.5.1. Mutagenesis to generate the SNase variants D+PHS D21N/T62K, D21N/F34K, and D21N/T62E was performed as described in chapter 4.5.1, with primers purchased from Integrated DNA Technologies.

5.5.2 Optical spectroscopy

All fluorescence and circular dichroism experiments were performed as described in Chapter 4.5.2.

5.5.3 NMR spectroscopy

Titrations were performed in 100 mM KCl, 90% H_2O:10% D_2O, as previously described[28]. Assignment of backbone resonances was performed as described in Chapter 3.5.2. Detection and titration of surface Asp/Glu residues was achieved using the ^{13}C-detected CBCGCO experiment[26]. For initial survey studies, 16-32 scans were collected per FID, with acquisition times in the indirect dimension of 20-25 ms. 16-48 scans were collected per FID for each titration point, with

acquisition times of 40 ms in the indirect dimension. Assignments were transferred from background protein spectra and confirmed with CCC(CO)NH spectra. Detection and titration of His residues was achieved using a standard aromatic ^{1}H-^{13}C HSQC. 8 scans were collected per FID, with acquisition times of 10 ms used in the indirect dimension. Assignments for His residues were transferred from background protein[12]. Backbone ^{1}H-^{15}N HSQC titrations of variants with T62K, D21N/T62K, F34K and D21N/F34K were performed with 4 scans per FID and 80 ms acquisition times in the indirect dimension; the titrations for F34K and D21N/F34K were duplicated at lower resolution (40 ms AQ). Backbone ^{13}C-^{13}C (HA)CACO spectra were collected with 16 scans per FID and acquisition times of 40 ms in the indirect dimension.

5.5.4 Analysis

All titration data were fit to a modified Hill equation, as described in Chapter 2.5.4, or a modified version accounting for two-site protonation[28]. In the case of the backbone titration data for F34K, the data was fit to the equation:

$$CS_{Obs} = CS_{Open}F_{Open} + CS_{Closed}F_{Closed}$$

where the fractional populations in the closed (F_{Closed}) and open (F_{Open}) states determined through the relationship adapted from reference 31:

$$Q_{pH} = 1 + K_{C-O}\frac{\left(1 + K_{a,Open}^{Lys}[H^+]\right)\left(1 + K_{a,Open}^{Asp}[H^+]\right)}{\left(1 + K_{a,Closed}^{Lys}[H^+]\right)\left(1 + K_{a,Closed}^{Asp}[H^+]\right)}$$

$$F_{Open} = \frac{1}{Q_{pH}}$$

$$F_{Closed} = \frac{K_{C-O}\dfrac{\left(1 + K_{a,Open}^{Lys}[H^+]\right)\left(1 + K_{a,Open}^{Asp}[H^+]\right)}{\left(1 + K_{a,Closed}^{Lys}[H^+]\right)\left(1 + K_{a,Closed}^{Asp}[H^+]\right)}}{Q_{pH}}$$

The energy gap between closed and open states is described by $K_{C\text{-}O}$, and the H^+-binding affinities of Lys and Asp in the closed and open state are treated explicitly. However, in practice the parameters describing $K_{a,Closed}^{Lys}$ and CS_{Open} are not independent; physically realistic values of CS_{Open} were obtained when setting $K_{a,Closed}^{Lys}$ to any value below 5, suggesting the latter parameter only provides an upper limit for the H^+-binding affinity of Lys-34 in the closed state. The parameters $K_{a,Open}^{Lys}$ and K_{C-O} are also coupled and cannot be accurately defined by the titration data. However, the parameters reduce to the apparent pK_a of the Lys-34, as $pK_{a,Open}^{Lys} - \frac{\Delta G_{C-O}}{1.36} = pK_{a,App}^{Lys}$; unlike the two dependent parameters $K_{a,Open}^{Lys}$ and K_{C-O}, the apparent pK_a is well defined by the titration data. This model was also used for the simulations shown in Fig. 5.4d.

5.5.5 Simulation of thermodynamic stabilities

The global thermodynamic stability of the variants with or without K34 was simulated using a modified version of the formalism described above, to account for the unfolded state:

$$Q_{pH} = 1 + K_{C-O}\frac{\left(1 + K_{a,Open}[H^+]\right)^n}{\left(1 + K_{a,Closed}[H^+]\right)^n} + K_U\frac{\left(1 + K_{a,Unfolded}[H^+]\right)^n}{\left(1 + K_{a,Closed}[H^+]\right)^n}$$

$$F_{Open} = \frac{1}{Q_{pH}}$$

$$F_{Closed} = \frac{K_{C-O}\dfrac{\left(1 + K_{a,Open}[H^+]\right)^n}{\left(1 + K_{a,Closed}[H^+]\right)^n}}{Q_{pH}}$$

$$F_{Unfolded} = \frac{K_U\dfrac{\left(1 + K_{a,Unfolded}[H^+]\right)^n}{\left(1 + K_{a,Closed}[H^+]\right)^n}}{Q_{pH}}$$

The pK_a values of His residues (His-8 and His-121) and Asp-21 were modeled explicitly, using pK_a values determined by NMR spectroscopy in this and previous studies[28] (unfolded pK_a set to 6.4 for His and 4.0 for Asp-21). The effects of the other 19 acidic groups were estimated based on the magnitude of the experimentally determined pK_a shifts; three Asp residues were modeled (closed pK_a set to 2.5, unfolded pK_a set to 3.6). Seven Glu residues were modelled as having slightly depressed (closed pK_a set to 3.9, unfolded pK_a set to 4.2), while another three were modeled with more depressed values (closed pK_a set to 3.5, unfolded pK_a set to 4.2). These simplifications were based on the experimentally determined pK_a values of Glu and Asp residues in SNase; using fully microscopic detail and accounting for each of the 19 surface Asp and Glu residues did not alter the pH-dependence of the stability in the pH regime corresponding to the protonation of Lys-34. In the open state, the shift in pKa values of Asp-21, Asp-77, and Glu-10 were modelled explicitly based on the values determined (or in the case of Glu-10 and Asp-77, estimated) using NMR spectroscopy, as were the small shifts observed for His-8 and His-121. The global thermodynamic stability was simulated as:

$$\Delta G = -RT \ln \frac{F_{Closed} + F_{Open}}{F_{Unfolded}}$$

This is appropriate to compare to the two-state unfolding data as the open and closed states have identical fluorescence spectra and highly similar m-values[22]; they are indistinguishable in unfolding experiments.

5.6 References:

1. Casey, J. R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **11**, 50–61 (2010).

2. Webb, B. A., Chimenti, M., Jacobson, M. P. & Barber, D. L. Dysregulated pH: a perfect storm for cancer progression. *Nat. Rev. Cancer* **11**, 671–677 (2011).

3. von Ballmoos, C., Wiedenmann, A. & Dimroth, P. Essentials for ATP synthesis by F_1F_o ATP synthases. *Annu. Rev. Biochem.* **78**, 649–672 (2009).

4. Srivastava, J. *et al.* Structural model and functional significance of pH-dependent talin–actin binding for focal adhesion remodeling. *Proc. Natl. Acad. Sci.* **105**, 14436–14441 (2008).

5. Franzmann, T. M. *et al.* Phase separation of a yeast prion protein promotes cellular fitness. *Science* **359**, (2018).

6. Erecińska, M., Deas, J. & Silver, I. A. The effect of pH on glycolysis and phosphofructokinase activity in cultured cells and synaptosomes. *J. Neurochem.* **65**, 2765–2772 (1995).

7. Isom, D. G. *et al.* Coordinated regulation of intracellular pH by two glucose-sensing pathways in yeast. *J. Biol. Chem.* **293**, 2318–2329 (2018).

8. White, K. A., Grillo-Hill, B. K. & Barber, D. L. Cancer cell behaviors mediated by dysregulated pH dynamics at a glance. *J. Cell Sci.* **130**, 663–669 (2017).

9. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).

10. Lee, K. K., Fitch, C. A. & García-Moreno E., B. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci.* **11**, 1004–1016 (2002).

11. Robinson, A. C., Castañeda, C. A., Schlessman, J. L. & García-Moreno E., B. Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl. Acad. Sci.* **111**, 11685–11690 (2014).

12. Robinson, A. C., Schlessman, J. L. & García-Moreno E., B. Dielectric properties of a protein probed by reversal of a buried ion pair. *J. Phys. Chem. B* **122**, 2516–2524 (2018).

13. Holliday, G. L., Mitchell, J. B. O. & Thornton, J. M. Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.* **390**, 560–577 (2009).

14. Lau, W. C. Y. & Rubinstein, J. L. Subnanometre-resolution structure of the intact *Thermus thermophilus* H^+-driven ATP synthase. *Nature* **481**, 214–218 (2012).

15. Kaila, V. R. I., Verkhovsky, M. I., Hummer, G. & Wikström, M. Glutamic acid 242 is a valve in the proton pump of cytochrome c oxidase. *Proc. Natl. Acad. Sci.* **105**, 6255–6259 (2008).

16. Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno E., B. Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci.* **107**, 16096–16100 (2010).

17. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno E., B. Large shifts in pK_a values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci.* **108**, 5260–5265 (2011).

18. Ho, M.-C., Ménétret, J.-F., Tsuruta, H. & Allen, K. N. The origin of the electrostatic perturbation in acetoacetate decarboxylase. *Nature* **459**, 393–397 (2009).

19. Harris, T. K. & Turner, G. J. Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98 (2002).

20. Shi, C., Wallace, J. A. & Shen, J. K. Thermodynamic coupling of protonation and conformational equilibria in proteins: Theory and simulation. *Biophys. J.* **102**, 1590–1597 (2012).

21. Goh, G. B., Laricheva, E. N. & Brooks, C. L. Uncovering pH-dependent transient states of proteins with buried ionizable residues. *J. Am. Chem. Soc.* **136**, 8496–8499 (2014).

22. Liu, J., Swails, J., Zhang, J. Z. H., He, X. & Roitberg, A. E. A coupled ionization-conformational equilibrium is required to understand the properties of ionizable residues in the hydrophobic interior of staphylococcal nuclease. *J. Am. Chem. Soc.* **140**, 1639–1648 (2018).

23. Nielsen, J. E., Gunner, M. R. & E, B. G.-M. The pK_a Cooperative: A collaborative effort to advance structure-based calculations of pK_a values and electrostatic effects in proteins. *Proteins Struct. Funct. Bioinforma.* **79**, 3249–3259 (2011).

24. Gunner, M. R., Zhu, X. & Klein, M. C. MCCE analysis of the pKas of introduced buried acids and bases in staphylococcal nuclease. *Proteins* **79**, 3306–3319 (2011).

25. Liu, R., Yue, Z., Tsai, C.-C. & Shen, J. Assessing lysine and cysteine reactivities for designing targeted covalent kinase inhibitors. *J. Am. Chem. Soc.* **141**, 6553–6560 (2019).

26. Chimenti, M. S. *et al.* Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein. *Structure* **20**, 1071–1085 (2012).

27. Kougentakis, C. M., Majumdar, A. & García-Moreno, E. B. Electrostatic effects in proteins are governed by pH-redistribution of the conformational ensemble. *bioRxiv*, doi:10.1101/2020.02.02.931253 (2020).

28. Pey, A. L., Rodriguez-Larrea, D., Gavira, J. A., García-Moreno, B. & Sanchez-Ruiz, J. M. Modulation of buried ionizable groups in proteins with engineered surface charge. *J. Am. Chem. Soc.* **132**, 1218–1219 (2010).

29. Robinson, A. C., Majumdar, A., Schlessman, J. L. & García-Moreno E., B. Charges in hydrophobic environments: A strategy for identifying alternative states in proteins. *Biochemistry* **56**, 212–218 (2017).

30. Harms, M. J., Schlessman, J. L., Sue, G. R. & García-Moreno E., B. Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci.* **108**, 18954–18959 (2011).

31. Castañeda, C. A. *et al.* Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. *Proteins* **77**, 570–588 (2009).

32. Doctrow, B. Electrostatic coupling and conformational fluctuations as determinants of pK_a values in proteins. Thesis, Johns Hopkins University (2014).

33. McIntosh, L. P. *et al.* Dissecting electrostatic interactions in Bacillus circulans xylanase through NMR-monitored pH titrations. *J. Biomol. NMR* **51**, 5 (2011).

34. Whitten, S. T., Wooll, J. O., Razeghifard, R., García-Moreno, E. B. & Hilser, V. J. The origin of pH-dependent changes in m-values for the denaturant-induced unfolding of proteins. *J. Mol. Biol.* **309**, 1165–1175 (2001).

35. Whitten, S. T., García-Moreno, E. B. & Hilser, V. J. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci.* **102**, 4282–4287 (2005).

36. Harvey, S. C. & Hoekstra, P. Dielectric relaxation spectra of water adsorbed on lysozyme. *J. Phys. Chem.* **76**, 2987–2994 (1972).

37. Bone, S. & Pethig, R. Dielectric studies of the binding of water to lysozyme. *J. Mol. Biol.* **157**, 571–575 (1982).

38. Bycroft, M. & Fersht, A. R. Surface electrostatic interactions contribute little to stability of barnase. *J. Mol. Biol.* **220**, 779–788 (1991).

39. Tomlinson, J. H., Ullah, S., Hansen, P. E. & Williamson, M. P. Characterization of salt bridges to lysines in the protein G B1 domain. *J. Am. Chem. Soc.* **131**, 4674–4684 (2009).

40. Antosiewicz, J., McCammon, J. A. & Gilson, M. K. The determinants of pK_as in Proteins. *Biochemistry* **35**, 7819–7833 (1996).

41. Fenton, A. W. Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem. Sci.* **33**, 420–425 (2008).

42. Lang, E. J. M., Heyes, L. C., Jameson, G. B. & Parker, E. J. Calculated pK_a variations expose dynamic allosteric communication networks. *J. Am. Chem. Soc.* **138**, 2036–2045 (2016).

43. Dai, S. *et al.* Low-barrier hydrogen bonds in enzyme cooperativity. *Nature* **573**, 609–613 (2019).

44. Hanoian, P., Liu, C. T., Hammes-Schiffer, S. & Benkovic, S. Perspectives on electrostatics and conformational motions in enzyme catalysis. *Acc. Chem. Res.* **48**, 482–489 (2015).

45. Kerns, S. J. *et al.* The energy landscape of adenylate kinase during catalysis. *Nat. Struct. Mol. Biol.* **22**, 124–131 (2015).

46. Warshel, A. & Bora, R. P. Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *J. Chem. Phys.* **144**, 180901 (2016).

47. Vavassori, S. *et al.* A pH-regulated quality control cycle for surveillance of secretory protein assembly. *Mol. Cell* **50**, 783–792 (2013).

48. Iserman, C. *et al.* Condensation of Ded1p promotes a translational switch from housekeeping to stress protein production. *Cell* (2020) doi:10.1016/j.cell.2020.04.009.

49. Triandafillou, C. G., Katanski, C. D., Dinner, A. R. & Drummond, D. A. Transient intracellular acidification regulates the core transcriptional heat shock response. *biorXiv*, doi:10.1101/414706 (2018).

50. Tanford, C. & Kirkwood, J. G. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* **79**, 5333–5339 (1957).

51. Sarkar, A., Gupta, P. L. & Roitberg, A. E. pH-dependent conformational changes due to ionizable residues in a hydrophobic protein interior: The study of L25K and L125K variants of SNase. *J. Phys. Chem. B* **123**, 5742–5754 (2019).

52. Richman, D. E., Majumdar, A. & García-Moreno E., B. pH dependence of conformational fluctuations of the protein backbone. *Proteins* **82**, 3132–3143 (2014).

53. Kougentakis, C. M., Skerritt, L., Majumdar, A., Schlessman, J. L. & García-Moreno E., B. The properties of buried ion pairs are governed by the propensity of proteins to reorganize. *bioRxiv,* doi:10.1101/2020.02.03.932012 (2020).

54. Lanyi, J. K. Bacteriorhodopsin. *Annu. Rev. Physiol.* **66**, 665–688 (2004).

55. Gayen, A., Leninger, M. & Traaseth, N. J. Protonation of a glutamate residue modulates the dynamics of the drug transporter EmrE. *Nat. Chem. Biol.* **12**, 141–145 (2016).

56. Pisliakov, A. V., Cao, J., Kamerlin, S. C. L. & Warshel, A. Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc. Natl. Acad. Sci.* **106**, 17359–17364 (2009).

5.7 Supporting Information

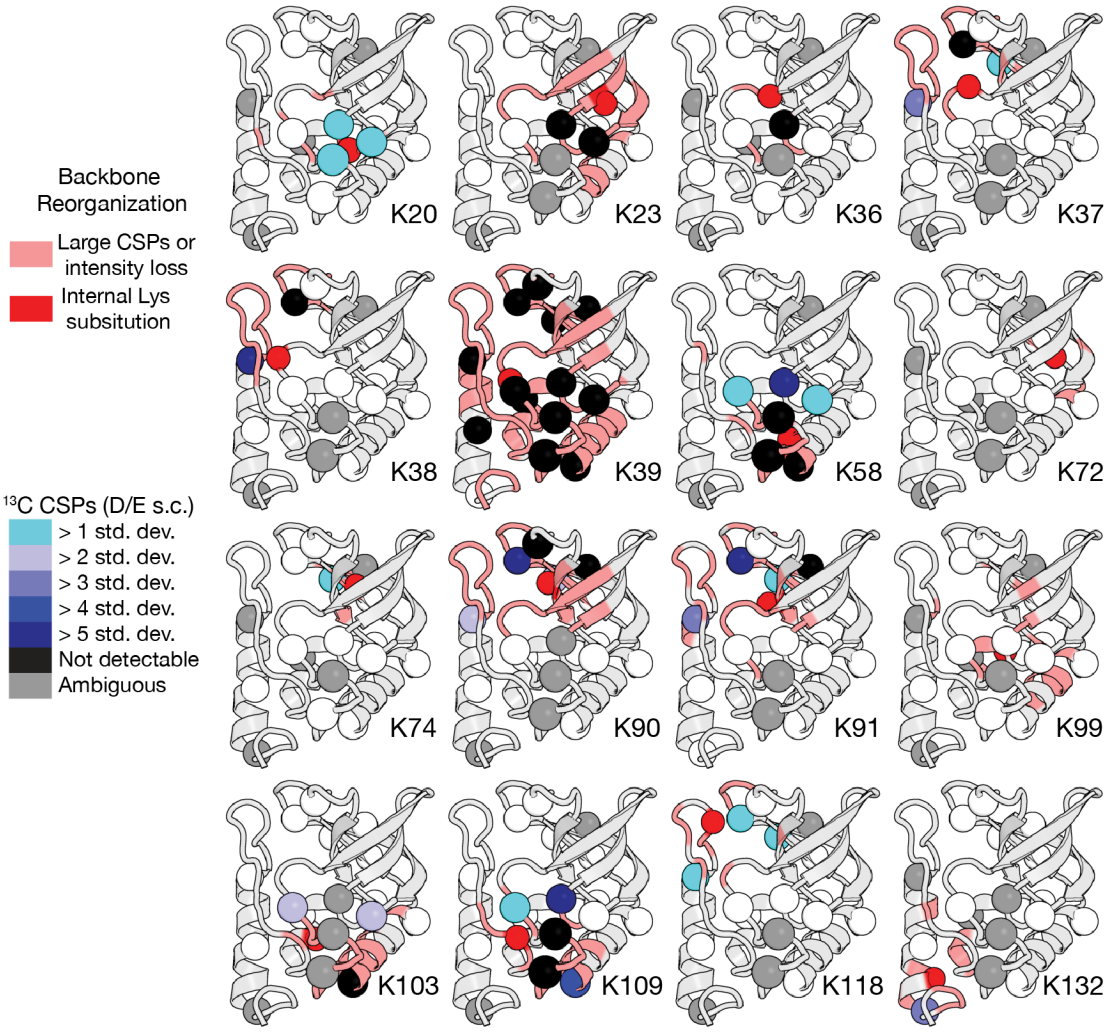

Fig S5.1. Survey of surface acidic residues perturbed by ionization of internal Lys residues. Chemical shift perturbations (weighted CSPs of Cγ/Cδ for Glu and Cβ/Cγ for Asp) were determined by comparing CBCGCO spectra of variants at pH values where the internal Lys is ionized and the protein reorganized (1-2 pH units below pK_a^{App}, except for those variants with normal pKa values, for which data was collected at pH 7.2) with background protein spectra at equivalent pH values.

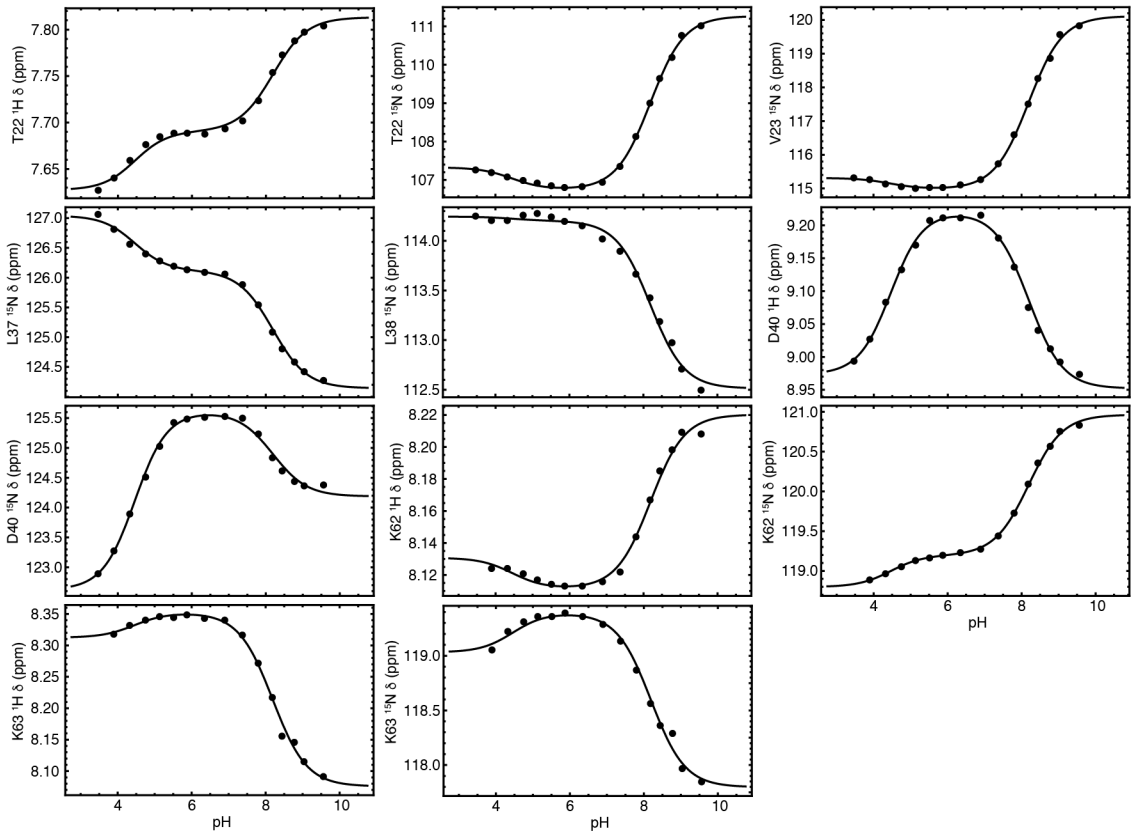

Fig S5.2. Global fitting of backbone titration data for variant with K62.

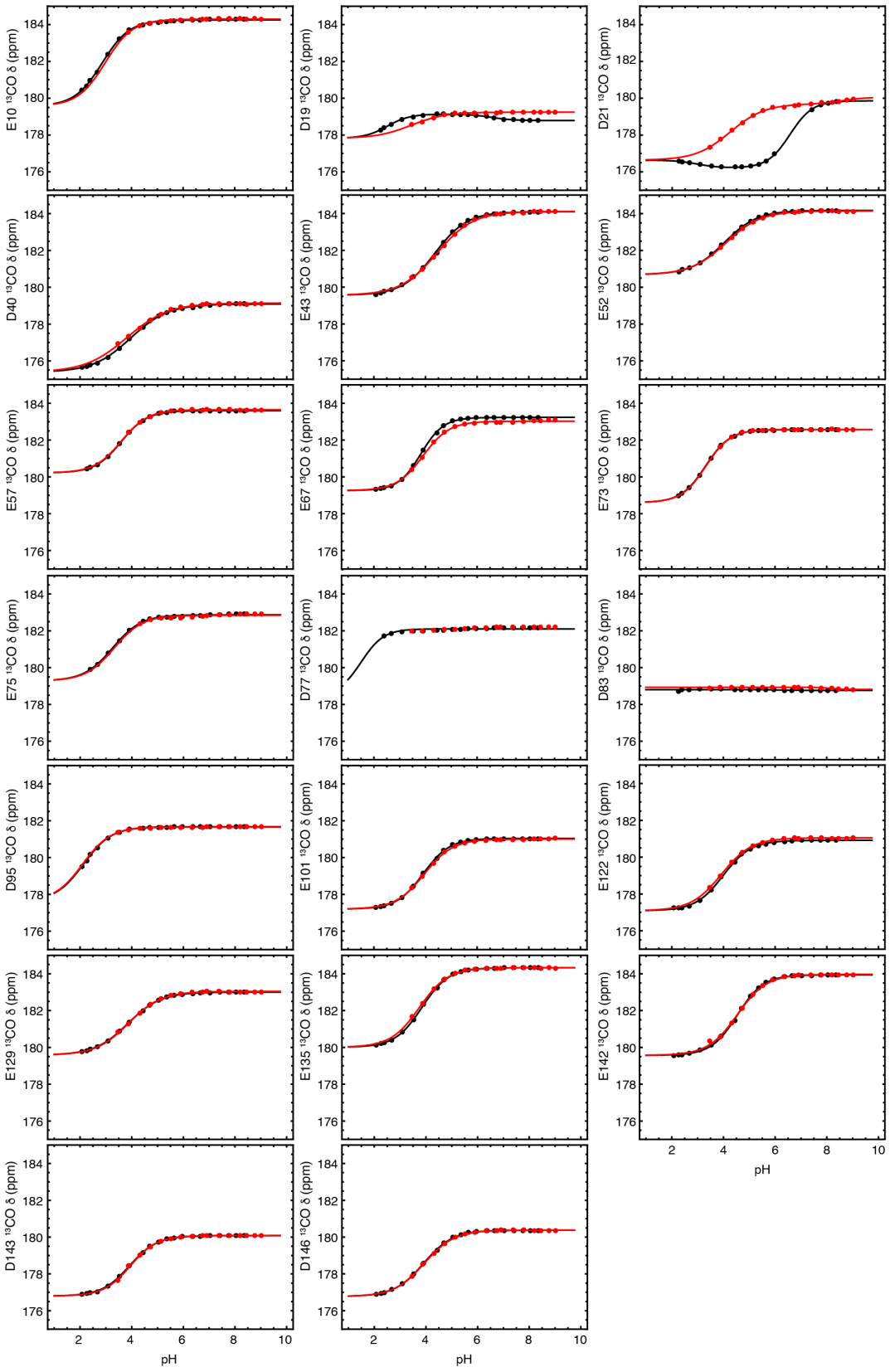

Fig S5.3. CBCGCO titration data for surface Asp and Glu residues in variant with K62.

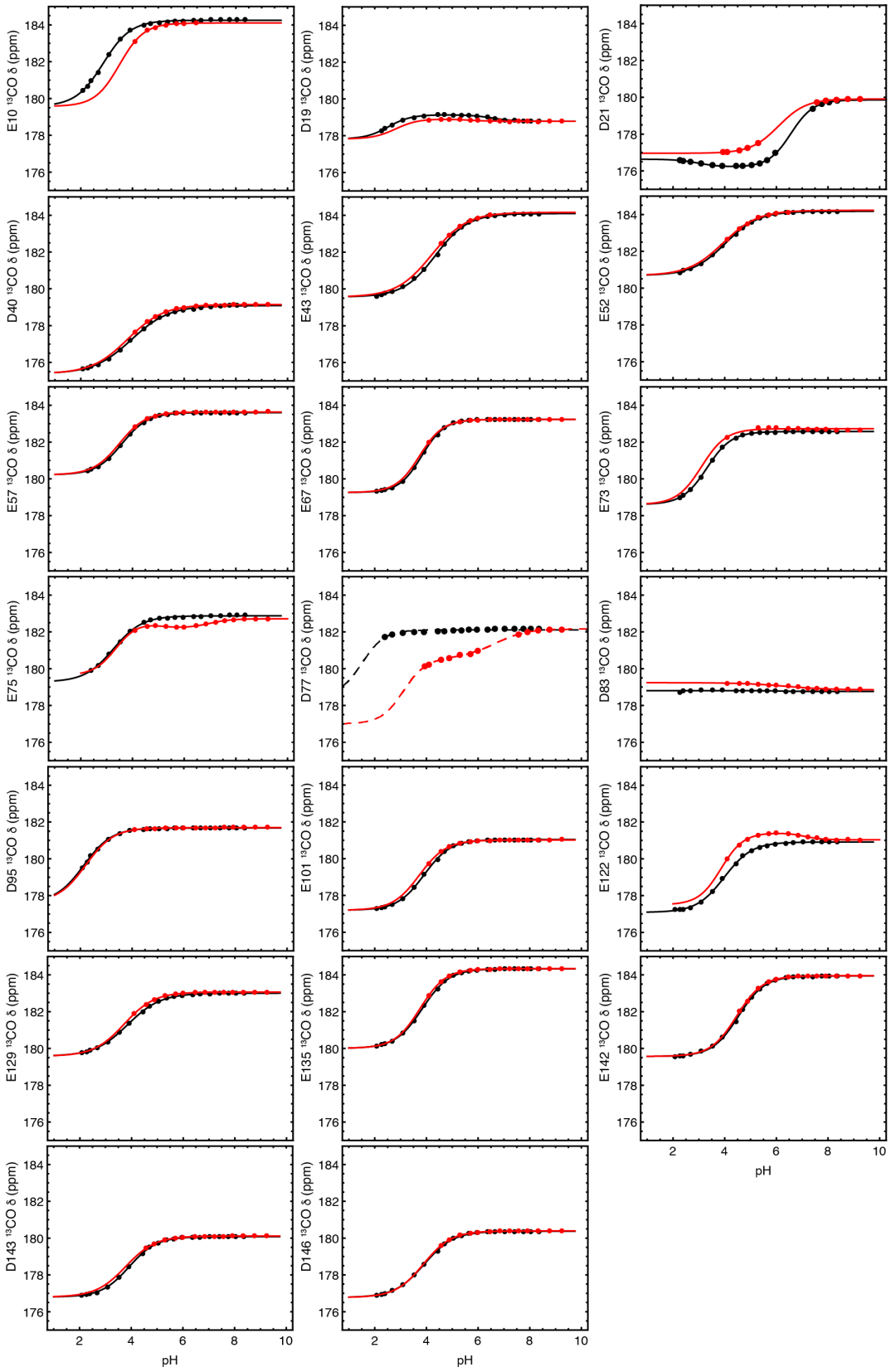

Fig S5.4. CBCGCO titration data for surface Asp and Glu residues in variant with K34.

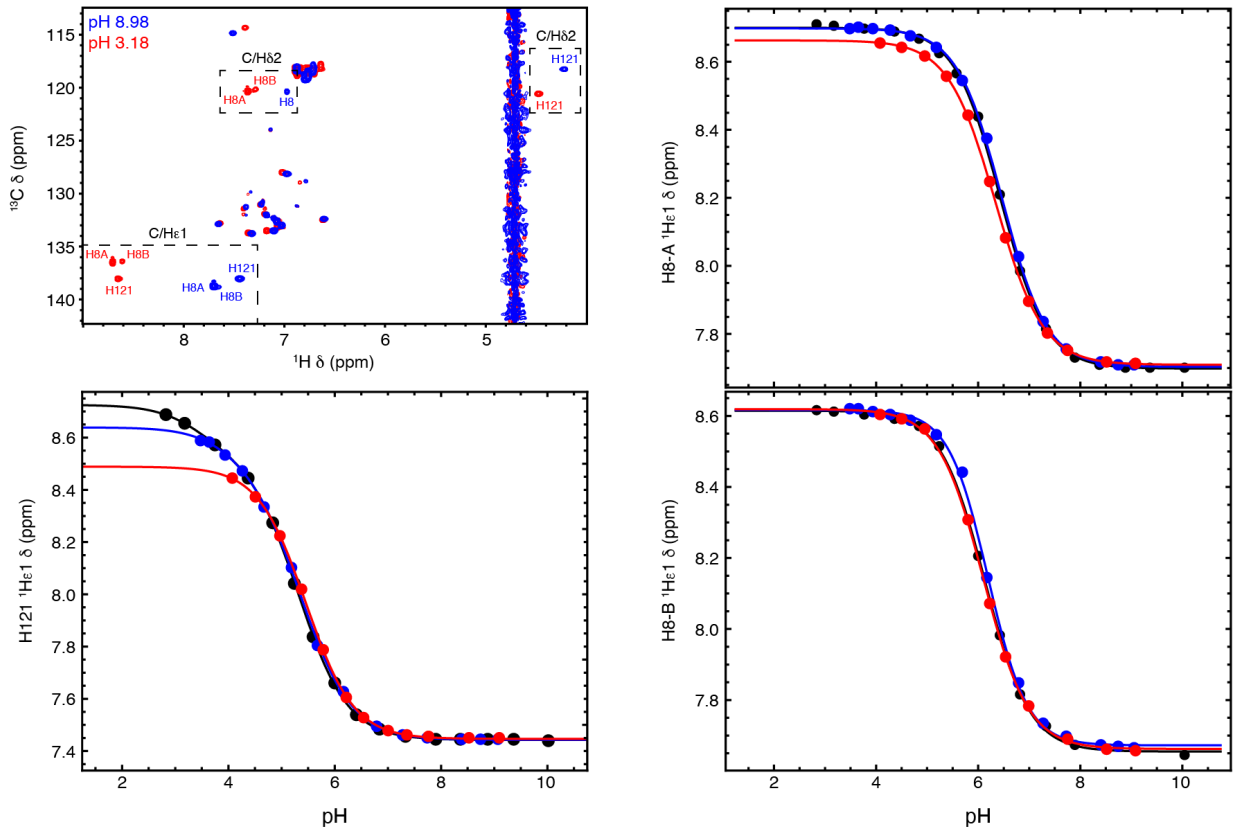

Fig S5.5. Titration data for surface His residues. Representative ^{1}H-^{13}C HSQC spectra of background protein shown in top left. Note the minor form of His-8 (H8B). Fits of pH-dependence of chemical shifts shown in three panels, of background protein (black), and variants with K34 (red) and K62 (blue).

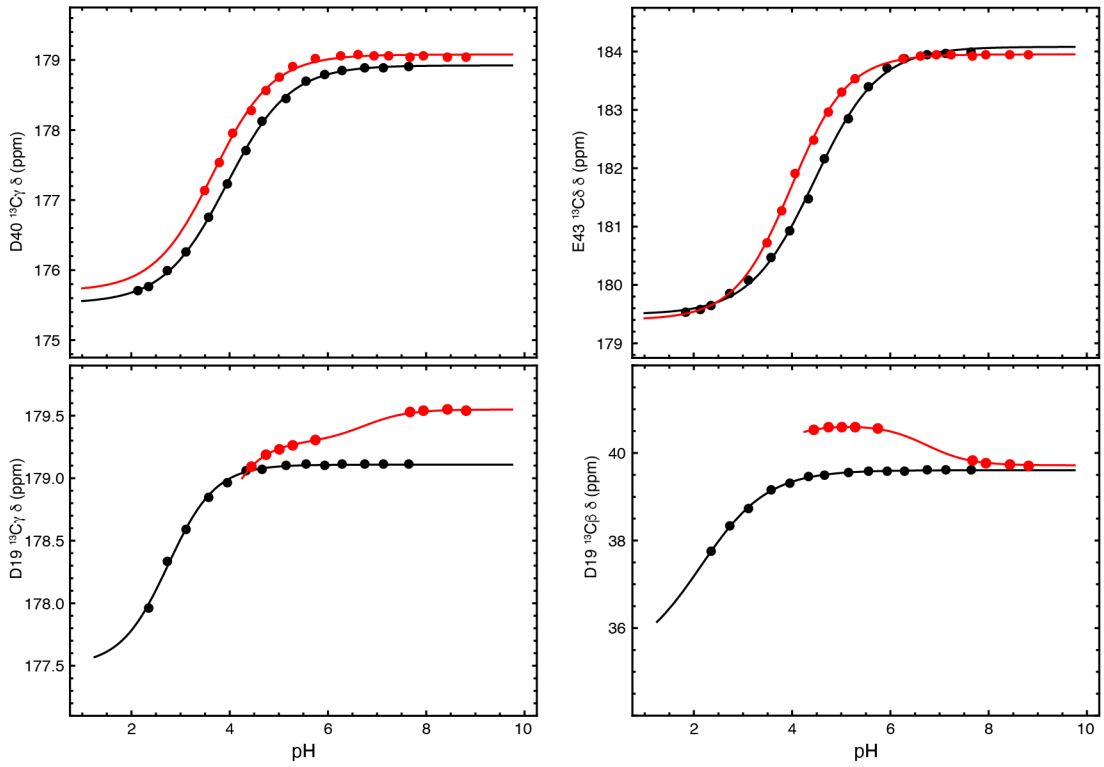

Fig S5.6. CBCGCO titration data for surface Asp and Glu residues in variant with N21/K62. The observed shifts in the transition points for Asp-40 and Glu-43 are consistent with the second region of pH-dependence observed in the thermodynamic data for the variant with N21/K62. Below, a transition coupled to the ionization of K62 is observed in the titration data of Asp-19; this can be fit to obtain a pK_a^{App} of 6.7.

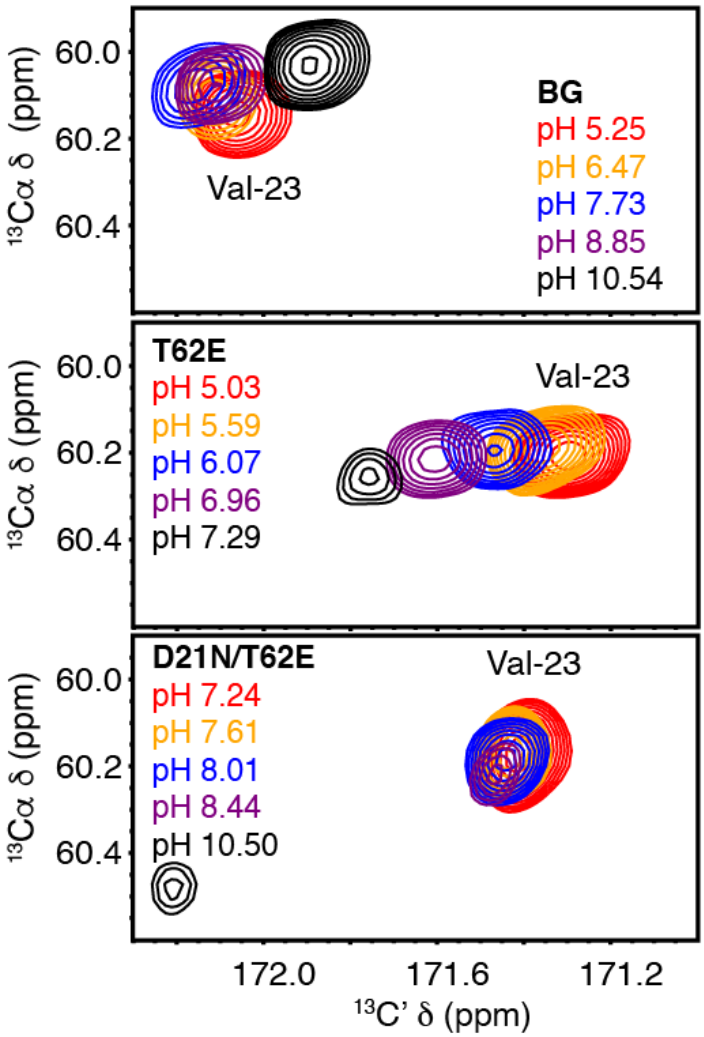

Fig S5.7. pH-dependent (HA)CACO spectra of variants with E62 and N21/E62. Close up of Val-23 backbone ^{13}Cα-^{13}C' resonance is shown, which is highly sensitive to ionization of both Glu-62 and Lys-62. Broadening of the resonance is observed near the apparent pK_a values of Glu-62 in both single and double variants; however, this occurs at approximately 1.0 – 1.5 pH units higher in the double variant relative to the single. The resonance in the background protein (top panel) does not show evidence of broadening over a wide range of pH values.

Chapter 6

Conclusions

Understanding the pH-dependent properties of proteins, and therefore the determinants of the pK_a values of their H^+-binding sites, has been of interest for a century. Although there have been significant advances in computational models over time, it is still difficult to reproduce experimentally determined pK_a values of residues in functional sites in proteins. The results of this thesis provide insight into the shortcomings of these models, and potential approaches to improve structure-based energy calculations.

The most important take away from these studies is the need to accurately predict the alternative states available to the protein. Accurate calculation of electrostatic energies in proteins cannot be achieved from a static structure; rather, the contributions of desolvation and Coulomb interactions in the relevant accessible states to the protein must be correctly reproduced. Many current algorithms continue to underestimate the magnitude of backbone reorganization coupled to the ionization of buried residues and continue to rationalize discrepancies between experimental and calculated pK_a values as being a result of issues with force-fields[1,2] or through physically unrealistic mechanisms (e.g., ion penetration into the hydrophobic core[3]). The studies presented here underscore that it is not enough to obtain a correct pK_a value; if computational methods cannot reproduce the types of structural changes observed in this study, it is unlikely they can contribute any physical insight into biochemical processes.

Correctly identifying alternatively folded or partially unfolded states remains a challenge for even sophisticated methods which take advantage of enhanced sampling techniques, for example through pH-replica exchange[4]. One factor that may contribute to the inability of current force fields to reproduce the structural changes observed in this study is the difficult in modeling local unfolding. Although some molecular dynamics studies have been able to capture a subset of

the conformational changes observed in this study[4,5], capturing changes involving significant unfolding of secondary structure remains challenging.

Another significant issue that structure-based pK_a calculation methods must solve is accurately calculating the energy gap between conformational states. The results from this thesis have shown how the pK_a values of buried residues, and the in some cases the cooperativity of H^+-binding, are governed by the free energy gap between closed and open states. Free energy calculations in proteins remains extremely challenging[6,7]; however, improvements in these methods are required in order to accurately simulate the pH-dependent properties of proteins.

One way in which this work can inform current force-fields to improve pK_a predictions and free energy calculations is through the conclusions drawn about electrostatic interactions in proteins. As has been previously described, many force-fields overestimate the contributions of Coulomb interactions in proteins[8]. Work presented by multiple groups suggests that interactions between surface residues behave as if the charges were well hydrated, in a high dielectric medium[9,10,11]. The work in this thesis expands upon this by showing that even interactions between charges in the protein interior, or between buried and surface charges, are as sensitive or more sensitive to the hydration energies of the individual charges than to the Coulomb interactions between them. Treatments of electrostatic effects in force-fields therefore generally overestimate the contributions from Coulomb interactions and underestimate the desolvation penalties associated with burying charges. Consistent with this view, in computational studies of ionization induced partial unfolding, it was necessary to use physically unrealistic, exaggerated charge states to observe partial unfolding consistent with experimental data[12].

In conclusion, the results presented in this thesis demonstrate that simple treatments of proteins as a low dielectric cavity surrounded by a high dielectric environment are not appropriate

for the accurate calculation of electrostatic energies in proteins. The potential issues, including the inability to simulate partial unfolding and treatment of electrostatics in force-fields, have been identified, along with suggestions for improving these methods. These data will serve as critical benchmarks for the next generation of structure-based pK_a calculation methods, which in turn will allow for more realistic treatments of pH-dependent properties of proteins, as well as the design of *de novo* proteins which can use changes in pH as a biological signal.

References

1. Goh, G. B., Laricheva, E. N. & Brooks, C. L. Uncovering pH-dependent transient states of proteins with buried ionizable residues. *J. Am. Chem. Soc.* **136**, 8496–8499 (2014).

2. Liu, R., Yue, Z., Tsai, C.-C. & Shen, J. Assessing lysine and cysteine reactivities for designing targeted covalent kinase inhibitors. *J. Am. Chem. Soc.* **141**, 6553–6560 (2019).

3. Wu, X. & Brooks, B. R. Hydronium ions accompanying buried acidic residues lead to high apparent dielectric constants in the interior of proteins. *J. Phys. Chem. B* **122**, 6215–6223 (2018).

4. Liu, J., Swails, J., Zhang, J. Z. H., He, X. & Roitberg, A. E. A coupled ionization-conformational equilibrium is required to understand the properties of ionizable residues in the hydrophobic interior of staphylococcal nuclease. *J. Am. Chem. Soc.* **140**, 1639–1648 (2018).

5. Damjanović, A., Brooks, B. R. & García-Moreno E., B. Conformational relaxation and water penetration coupled to ionization of internal groups in proteins. *J. Phys. Chem. A* **115**, 4042–4053 (2011).

6. Hansen, N. & van Gunsteren, W. F. Practical aspects of free-energy calculations: A review. *J. Chem. Theory Comput.* **10**, 2632–2647 (2014).

7. Lin, Y.-L., Aleksandrov, A., Simonson, T. & Roux, B. An overview of electrostatic free energy computations for solutions and proteins. *J. Chem. Theory Comput.* **10**, 2690–2709 (2014).

8. Ahmed, M. C., Papaleo, E. & Lindorff-Larsen, K. How well do force fields capture the strength of salt bridges in proteins? *PeerJ* **6**, e4967 (2018).

9. Bycroft, M. & Fersht, A. R. Surface electrostatic interactions contribute little to stability of barnase. *J. Mol. Biol.* **220**, 779–788 (1991).

10. Antosiewicz, J., McCammon, J. A. & Gilson, M. K. The determinants of pK_as in Proteins. *Biochemistry* **35**, 7819–7833 (1996).

11. Lee, K. K., Fitch, C. A. & García-Moreno E., B. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci.* **11**, 1004–1016 (2002).

12. Kato, M. & Warshel, A. Using a charging coordinate in studies of ionization induced partial unfolding. *J. Phys. Chem. B* **110**, 11566–11570 (2006).

Vita

Christos Michael Kougentakis was born in New York City, New York in 1990 to Greek immigrants Petros and Anna Kougentakis. Christos' interest in biology began at a young age and was cultivated in the New York City public school system, especially at Brooklyn Technical High School where he first began considering a career in scientific research. Christos attended Bennington College starting in 2008, where he continued pursuing his interests in biology, music and political science. During his first two years, he spent summer and winter breaks working in Dr. Elizabeth Miller's laboratory at Columbia University (now MRC Laboratory of Molecular Biology), studying endoplasmic reticulum associated degradation pathways in yeast. His final two years at Bennington were spent working closely with Bennington professor Dr. Amie McClellan, where he continued to study proteostasis and chaperones. Eager to develop a stronger background in fundamental protein science, Christos spent two years working as a technician in Dr. Amy Keating's laboratory at MIT.

In order to obtain a rigorous training in protein science, Christos decided to pursue a Ph.D. in Molecular Biophysics at Johns Hopkins University. During his time in Baltimore, Christos researched fundamental questions in protein electrostatics in Dr. Bertrand García-Moreno's laboratory. To apply his expertise in protein physical chemistry to questions of biological regulation and disease, Christos is currently residing in the Bay Area, where he is working in the laboratory of Dr. Diane Barber at the University of California, San Francisco.