

Electrostatic effects in proteins are governed by pH-redistribution of the conformational ensemble

Christos M. Kougentakis^{1,2}, Ananya Majumdar³, and Bertrand García-Moreno, E.^{1*}

¹T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD.

²Graduate Program in Molecular Biophysics, Johns Hopkins University, Baltimore, MD.

³Biomolecular NMR Center, Johns Hopkins University, Baltimore, MD.

*Corresponding author (Email: bertrand@jhu.edu)

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 groups¹. These groups are tolerated because they are usually buried in the neutral state². However, when they become charged they can drive structural transitions to open states in which the charge can be stabilized, mostly through interactions with water³. 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 transduction⁴, the identification of pH-sensing proteins in nature⁷, and for the engineering of pH-sensitive dynamics and function in *de novo* designed proteins⁸.

Buried ionizable residues in proteins are critical for many biochemical processes, especially energy transduction⁴, catalysis¹, and pH-sensing⁷. 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 state². In proteins such as ATP synthase⁴, cytochrome C oxidase⁵, and bacteriorhodopsin⁶, 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 normal⁵. 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 positions² presented here addresses this question directly.

The pK_a values of the internal Lys residues in SNase (Fig. 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. S1)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 solvent¹¹. 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.

Advances in data acquisition¹² and analysis¹³ 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 1H , ^{13}C , and ^{15}N nuclei were measured for each variant at different pH values, or between the background and variants (Fig. 1). In two cases the ionization of the internal Lys triggered global unfolding (Fig. S2). In the other 23 variants the CSPs and/or loss of resonance intensities that were observed were consistent with local or subglobal changes in structure¹⁴ or dynamics¹⁵. In some instances, these changes were concomitant with an increase in random coil resonances (Fig. S3-26, Table S1-3). Structural analysis¹⁶ 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. 2A). In other cases crystal structures confirm that the nature of the pH-dependent structural changes observed by NMR is partial unfolding (Fig. 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.

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. 2D)^{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, 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 1 and 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 for Lys in water the stability of the closed states of each protein decrease in a pH-dependent manner (Fig. 2D). 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.

At pH values below pK_a^{APP} the open state is the dominant state in solution (Fig. 2D), enabling structural characterization of the open state (with NMR spectroscopy) and measurement of its stability (with chemical denaturation experiments²). 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). 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, 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, red in Fig. 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. 3, Fig. 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) and the interface made between α -2, α -3 and the β -barrel (yellow in Fig. 3), 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 (^1H - ^{15}N HSQC spectra in Fig. S4-7), 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 experiments¹⁷. Ionization of the buried Lys residue leads to global unfolding in only two variants (red in Fig. 3); in one case (N100K), fluorescence and circular dichroism spectroscopy provide evidence of some residual structure². 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 response.

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 (Fig. S27)¹⁷, suggesting that the partially unfolded states in Figure 2 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 states are only populated to a significant extent by the truly local nature of these pH-driven structural changes. 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 experiments² (Fig. S28). Recent studies have shown that sparsely populated, partially unfolded states play important roles in tuning activity and allostery in well folded enzymes¹⁸. 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 conditionally disordered states in protein function has been underappreciated, given the difficulties in capturing these sparsely populated states with traditional structural biology methods. Although advances in structural methods have begun to access heterogeneity in the conformational ensemble¹⁹, this present study offers the most complete characterization of the native state ensemble of a protein.

Our data with SNase show that in the physiological range of pH²⁰ a single ionizable residue with an anomalous pK_a can affect the stability and conformation of a protein in a significant way^{2,21}. This suggests a mechanism whereby biological activity can be tuned through pH-dependent structural reorganization²², allosteric regulation²³, and order-to-disorder transitions²⁴. 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. photoillumination²⁵, redox reactions⁵) that introduce charges in hydrophobic environments. With the large amount of structures available in the Protein Data Bank that can be mined with informatics approaches⁷, as well as the development of chemical probes that can selectively modify ionizable residues with anomalous pK_a values²⁶, 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^{27,28}. 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. 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 signal²⁰, and that pH dysregulation is observed in pathological states such as cancer²⁹. 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 transduction⁴ and regulatory processes²⁰ but also for improved prediction of pathological mutations that can make proteins sensitive to pH dysregulation³⁰. 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 function⁸.

Acknowledgements:

Drs. Juliette Lecomte, Vincent Hilser, and Doug Barrick are thanked for discussions. Drs. Dominique Frueh, Scott Nichols and Bradley Harden assisted in the implementation of non-uniformly sampled pulse sequences. All data were acquired at the JHU Biomolecular NMR Center. This work was funded by NSF grant MCB1517378 and NIH grant GM061597 to B.G.M.E.

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. & E, B. G.-M. 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., E, B. G.-M. & 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 F₀-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. Chimentì, 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. in *Methods in Enzymology* vol. 338 3–34 (Elsevier, 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. 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).
18. 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).
19. Frank, J. New Opportunities Created by Single-Particle Cryo-EM: The Mapping of Conformational Space. *Biochemistry* **57**, 888–888 (2018).
20. Casey, J. R., Grinstein, S. & Orłowski, J. Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **11**, 50–61 (2010).
21. Zheng, Z. & Sosnick, T. R. Protein Vivisection Reveals Elusive Intermediates in Folding. *J. Mol. Biol.* **397**, 777–788 (2010).
22. 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).
23. Bongard, J. *et al.* Identification of Noncatalytic Lysine Residues from Allosteric Circuits via Covalent Probes. *ACS Chem. Biol.* **13**, 1307–1312 (2018).

24. Salmon, L. *et al.* The Mechanism of HdeA Unfolding and Chaperone Activation. *J. Mol. Biol.* **430**, 33–40 (2018).
25. Xie, A. *et al.* Formation of a New Buried Charge Drives a Large-Amplitude Protein Quake in Photoreceptor Activation. *Biochemistry* **40**, 1510–1517 (2001).
26. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).
27. 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).
28. 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).
29. 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).
30. 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).
31. Castañeda, C. A. *et al.* Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease. *Proteins Struct. Funct. Bioinforma.* **77**, 570–588 (2009).
32. 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).
33. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
34. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
35. Lee, W. & Markley, J. L. PINE-SPARKY.2 for automated NMR-based protein structure research. *Bioinforma. Oxf. Engl.* **34**, 1586–1588 (2018).
36. Harms, M. J., Schlessman, J. L., Sue, G. R. & E, B. G.-M. Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci.* **108**, 18954–18959 (2011).
37. 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).
38. 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).

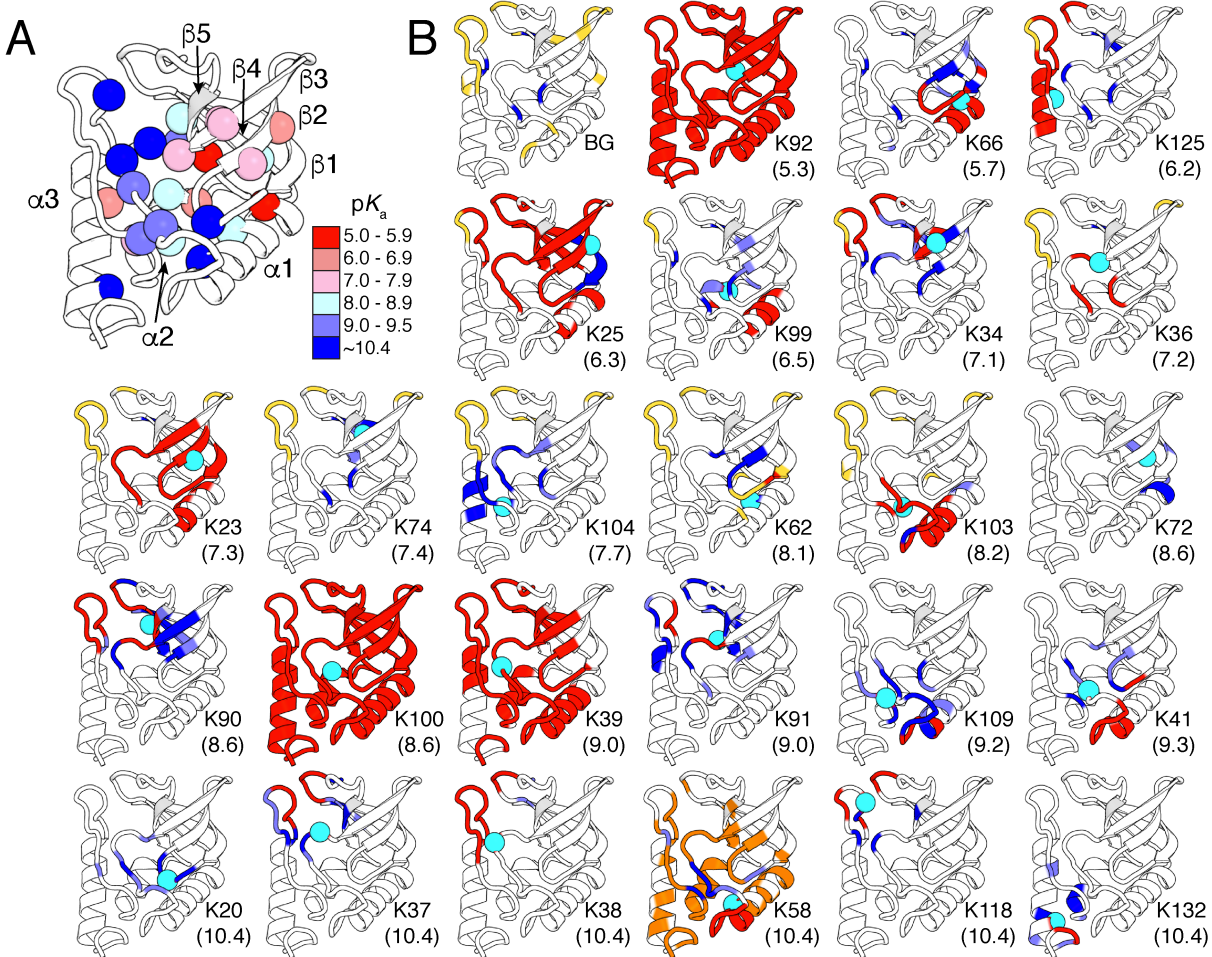


Figure 1. Structural reorganization coupled to the ionization of buried Lys residues. (A) Crystal structure of staphylococcal nuclease (PDB ID: 3BDC). Spheres identify sites of substitutions, color-coded according to the pK_a value of Lys residue at that position. (B) Chemical shift perturbations (CSPs) of backbone 1H_N , $^1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$, 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.

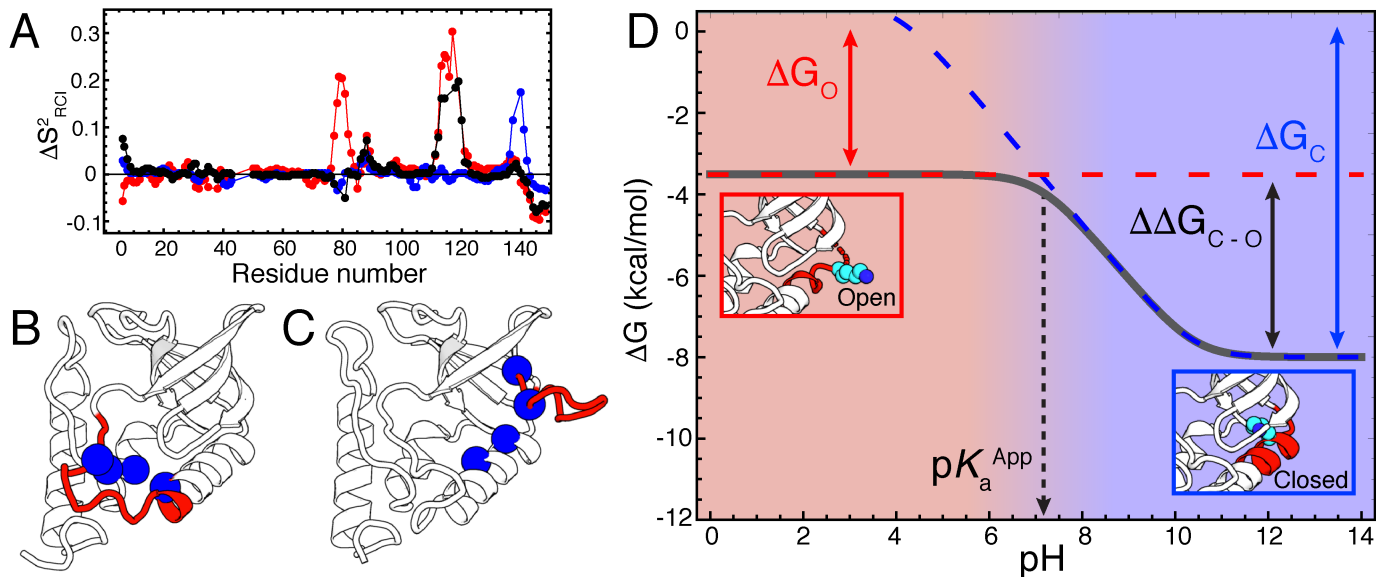


Figure 2. Structural and thermodynamic characterization of pH-dependent conformational changes necessary to hydrate buried Lys residues. (A) Difference in predicted order parameters (predicted using secondary shifts and the Random Coil Index) between the background protein at pH 7.4 and variants with 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 structured in the variant relative to the background protein. Based on error estimates of the RCI analysis, values below 0.07-0.1 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 in blue spheres. (D) Simulation showing the pH-dependent thermodynamic stability of a protein with buried Lys. The solid gray 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).

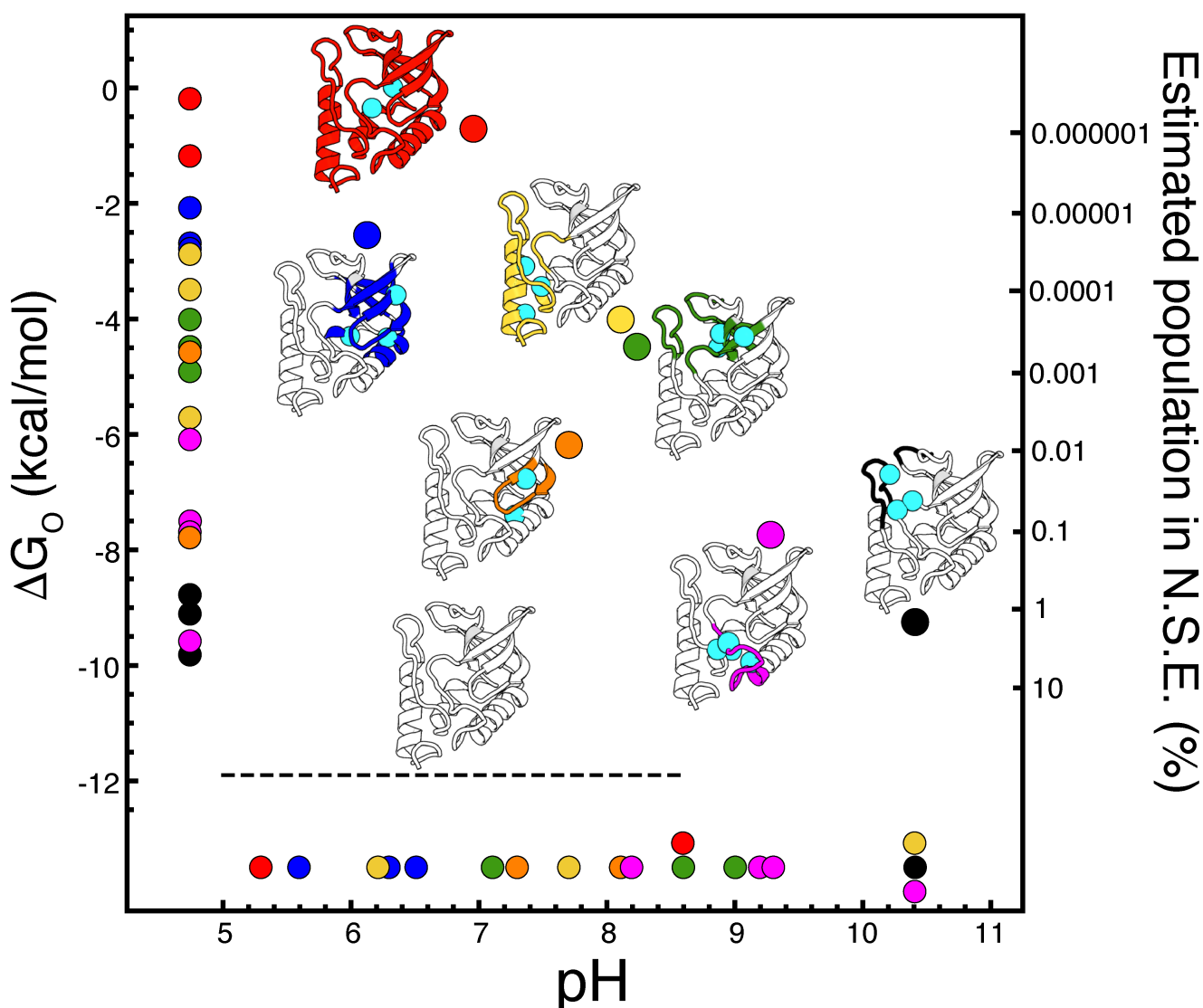


Figure 3. Conformational landscape of staphylococcal nuclease modulated by changes in pH. Correlation of the structural perturbations induced by the ionization of an internal Lys—detected with NMR spectroscopy—with thermodynamic stability measured by chemical denaturation for each protein under conditions 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).

Methods

Protein purification and sample preparation

Bacterial stocks containing plasmids with single Lys substitutions introduced into a highly stable variant (deletion of a loop comprising residues 44-49, G50N, V51F, P117G, H124L, S128A, known as the Δ +PHS variant) of staphylococcal nuclease were previously prepared. Protein was prepared and stored as previously described³¹. For NMR experiments, proteins were thawed and buffer exchanged into 25 mM buffer and 100 mM KCl in 90% H₂O/10% D₂O. 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 water¹⁰. 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.

NMR Spectroscopy

All NMR experiments were performed on Bruker Avance or Avance II spectrometers equipped with cryogenic TCI probes. All data was collected at 25 °C. For backbone resonance assignments, 2D ¹H-¹⁵N 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 ¹³C (10-12 ms for HNCO), 8-10 ms in ¹H and 23-24 ms in ¹⁵N in the indirect dimensions. Spectral widths in the indirect dimension for ¹³C were 62, 16, and 70 ppm for HNCACB/CBCACONH, HNCO, and CC(CO)NH spectra, respectively. Spectral widths in the indirect dimension for ¹H were 8 and 11 ppm for HBHACONH and H(CCCO)NH experiments, respectively. The ¹⁵N spectral width was set to 25.5 ppm for all experiments. Carriers were adjusted based on spectra, but were generally 116-119 in ¹⁵N, 42 ppm for aliphatic ¹³C spectra and 177-178 ppm for the HNCO experiment. ¹H 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 lab¹². Experiments were sampled at 15%, except for the CCCONH and (H)CCCONH, which were sampled at 17.5%. Spectra were reconstructed using NESTA-NMR³². Spectra were processed using NMRPipe³³ and analyzed using NMRFAM-Sparky³⁴.

Assignment procedure

The ¹H-¹⁵N HSQC, HNCACB, CBCACONH, and HNCO were manually peak picked. The automated assignment protocol of PINE-SPARKY was used to rapidly assign manually

picked peaks^{13,35}. In general, accuracy of this protocol was 90-100%, except in cases where broadened broke inter-residue connectivities, in which accuracy could drop to 50-60%. Spectra were then manually assigned to correct misassigned resonances from PINE, and to assign the ¹H α 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 C β carbon were assigned for this study. All spectra underwent at least one more round of manual assignment to prevent misassignment of resonances.

Chemical shift perturbation analysis

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 C β resonances of Asp-21, His-121, and His-8 were excluded from the analysis of C β 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 large 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 very large CSPs (blue). This cutoff was chosen because the only residues that met the criteria for large or very 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 more than 3 out of 6 of the backbone nuclei are broadened beyond identification below the apparent pK_a value of the buried Lys. 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 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. 1.

Structural studies of alternative states

Crystal structure of background protein and variants with neutral Lys were solved in previous studies^{31,10}. Crystal structures of alternative states (A109R, V23E, and V66K) were solved in previous studies^{36,37,11}. 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 prediction³⁸. Predicted order parameters (S²_{RCI}) were generated using the RCI algorithm¹⁶, to assess changes in local dynamics.

Simulation of thermodynamic stabilities

The thermodynamic stability of a protein with a buried Lys residue was simulated as²:

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

Because structural perturbations caused by ionization of internal Lys residues do not affect Trp. fluorescence, the closed and open states observed in this study cannot be distinguished in a chemical denaturation experiment, therefore the ΔG represents the simulated thermodynamic stability of the protein as observed in a chemical denaturation experiment. For the purpose of the simulation (solid gray curve in Figure 2D), 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) 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 2D), the maximum stability was set to 8 kcal/mol, and the terms $pK_{a,C}$ and $pK_{a,O}$ refer to the pK_a 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} : $\Delta G_O = \Delta G_C + 1.36(pK_{a,O} - pK_a^{APP})$

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 variants², measurements and errors of which are available in Table S4. In general, the errors for stability measurements was between 0.1-0.2 kcal/mol, For the ΔG_O term, $\Delta G^{\circ}_{H_2O}$ 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 residue²). In these cases, a $\Delta G^{\circ}_{H_2O}$ 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.