Reverse Protonation of Buried Ion-pairs in Staphylococcal Nuclease Mutants

Jiahua Deng[†] and Qiang Cui[*,‡]

†*Department of Chemistry, Boston University, 590 Commonwealth Avenue Boston, MA 02215*

‡*Departments of Chemistry, Physics and Biomedical Engineering, Boston University, 590 Commonwealth Avenue Boston, MA 02215*

E-mail: qiangcui@bu.edu, Tel:(+1)-617-353-6189

Abstract

Although buried titratable residues in protein cavities are often of major functional importance, it is generally challenging to understand their properties such as the ionization state and factors of stabilization based on experimental studies alone. A specific set of examples involve buried Glu-Lys pairs in a series of variants of *Staphylococcal* Nuclease, for which recent structural and thermodynamic studies appeared to suggest that both the stability and the ionization state of the buried Glu-Lys pair are sensitive to its orientation (i.e., Glu23-Lys36 vs. Lys23-Glu36). To further clarify the situation, especially ionization states of the buried Glu-Lys pairs, we have conducted extensive molecular dynamics simulations and free energy computations. Microsecond molecular dynamics simulations show that the hydration level of the cavity depends on the orientation of the buried ion-pair therein as well as its ionization state; free energy simulations recapitulate the relative stability of Glu23-Lys36 (EK) vs. Lys23-Glu36 (KE) mutants measured experimentally, although the difference is similar in magnitude regardless of the ionization state of the Glu-Lys pair. A complementary set of free energy simulations strongly suggests that, in contrast to the original suggestion in the experimental analysis, the Glu and Lys residues prefer to adopt their charge-neutral rather than the ionized states. This result is consistent with the low dielectric constant computed for water in the cavity, which makes it difficult for the protein cavity to stabilize the a pair of charged Glu-Lys residues, even with water penetration. The current study highlights the role of free energy simulations in understanding the ionization state of buried titratable residues and the relevant energetic contributions, forming the basis for rational design of buried charge-networks in proteins.

1 Introduction

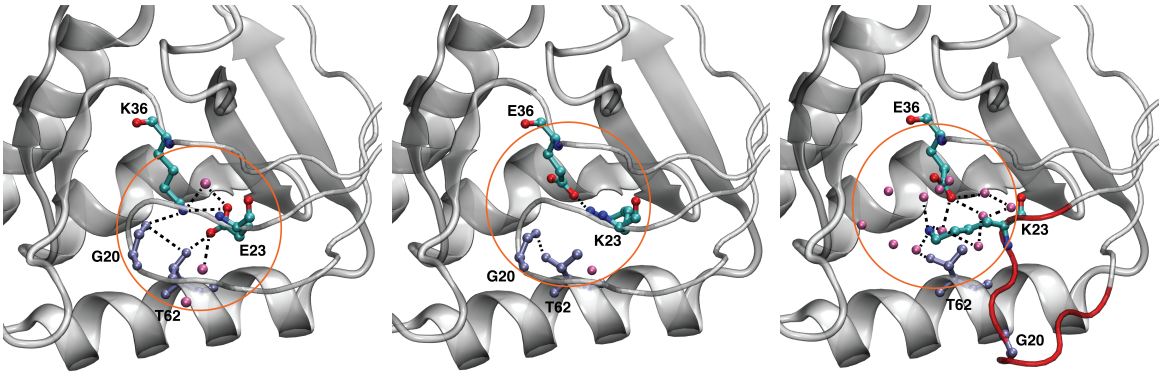
Despite their highly polar nature, titratable residues are commonly observed in enzyme active sites[1] and the interior of ion transporters[2] because they provide crucial stabilization of charged species in otherwise relatively non-polar environments. The design of new enzymes[3] also requires strategically positioning titratable residues in a confined cavity to either stabilize a high-energy transition state and/or to explicitly participate in general-acid-general-base catalysis.[4–6] Therefore, understanding the properties of buried titratable residues and physical factors that stabilize their presence in protein interior is of great fundamental and practical importance.

Burying polar and charged residues in the interior of proteins is generally expected to be energetically costly since the amount of stabilization provided by the pre-organized protein backbone is limited relative to bulk solution.[7,8] Although it is increasingly recognized that water molecules can penetrate into protein cavities to further stabilize polar and charged groups,[9–14] the degree of stabilization provided by water penetration, however, often remains unclear. First, it is generally difficult to unambiguously determine the number of water molecules in protein cavities using experimental approaches. For example, water penetration is likely coupled to protein dynamics, which is significantly damped in the crystalline environment, especially at cryogenic temperatures under which high-resolution crystal structures are often obtained.[15,16] Moreover, mobility of internal water molecules makes them difficult to resolve with X-ray diffraction,[17,18] and determining the number of internal water based on spectroscopy alone[19,20] is also not straightforward. Second, the degree of stabilization of buried charge/dipole provided by internal water molecules depends on not only latter's locations but also their fluctuations,[21,22] which are not readily available from experiments.

An important question regarding buried titratable residues concerns their ionization states,[23,24] which are intimately related to the degree of stabilization provided by the protein microenvironment. For an isolated titratable residue, it is possible to drive the change of titration state by altering pH, which may lead to substantial structural changes at both local

and global scales.[25–27] For a cluster of strongly coupled titratable residues, such as a pair of basic and acidic residues (e.g., Glu-Lys), the response to pH is suppressed[28] and determination of titration states requires detailed spectroscopic analysis such as NMR and IR using isotopically labeled amino acids.[29–31]

A case in point is a set of mutants of *Staphylococcal* nuclease (SNase) studied by Garcia-Moreno and co-workers,[32,33] who employed Δ+PHS (PDB entry: 3BDC) and Δ+PHS* (PDB entry: 3SK6) as the mutation backgrounds as they are highly stable variants of SNase; the more stable Δ+PHS* features several additional mutations (D21N, T33V, T41V, and S58A) relative to Δ+PHS. In the background of the Δ+PHS mutant, Robinson et al.[33] mutated two non-polar residues in the hydrophobic core, V23 and L36, to a series of polar (Q) or titratable (E, K) residues. As expected, these mutants exhibited substantially lower folding stabilities compared to the Δ+PHS background; at neutral pH (pH=7), the magnitude of folding free energy decreased from ∼-12 kcal/mol for the Δ+PHS background to the range of ∼-1 to -6 kcal/mol for the mutants. Of particular interest are two mutants that nominally involve a buried ion-pair: V23E/L36K and V23K/L36E, which are referred to as the "EK" and "KE" variants, respectively, in the following discussion. While EK remains reasonably stable with a folding free energy of -2.6±0.4 kcal/mol, the KE variant is only marginally stable with a folding free energy of -0.6±0.5 kcal/mol; indeed, heteronuclear single quantum coherence (HSQC) spectra indicated a significant population of unfolded state for the KE variant in solution. Introducing the Lys23/Glu36 mutations into the more stable Δ+PHS* background led to a folding free energy of -3.1±0.2 kcal/mol, and HSQC spectra indicated a largely folded native structure;[33] nevertheless, introducing the Lys23/Glu36 mutations still cost more than 11 kcal/mol, similar to that observed in the Δ+PHS background and larger in magnitude than the cost of ∼9 kcal/mol for introducing the Glu23/Lys36 mutations. These differences between the EK and KE variants indicated that the degree of stabilization of a buried dipole depends on orientation, highlighting the pre-organized nature of protein structure as emphasized by Warshel and co-worker.[34]



(a) (b) (c)

Figure 1: Crystal structures that illustrate the microenvironments of (a) V23E/L36K in the Δ+PHS background (3NHH,[32] referred to as EK), (b) V23K/L36E in the Δ+PHS background (6AMF,[33] referred to as KE), and (c) V23K/L36E in the Δ+PHS* background (3SK6[33]). The spherical cavity region is shown as an orange circle. The center of the cavity is defined as the center of geometry for the OE1, OE2 atoms in Glu and the NZ atom in Lys with a radius of 6 Å. Oxygen and nitrogen atoms in Lys and Glu are shown in red and blue, respectively, while carbon atoms are in cyan. Crystallographic water molecules within the cavity are shown as mauve spheres. Two residues, Gly20 and Thr62, which can form hydrogen bonds with either Lys/Glu or crystallographic water are also shown in ice blue. Hydrogen bonds are represented as dashed lines. β-strands near the cavity are in tube representation for better visualization. Residues 16-24 of the KE variant in the Δ+PHS* background are observed in an open conformation (colored in red), which leads to a substantially higher level of hydration and further separation of Lys23 and Glu36.

Crystal structures of these mutants[32,33] further highlighted additional nuances in the EK and KE variants. While the Glu23-Lys36 pair was observed to be surrounded by a number of water molecules in the cavity[32] (Fig. 1a), the Lys23-Glu36 pair in the KE variant of Δ+PHS was apparently dry[33] (Fig. 1b; one water molecule is resolved in the cavity region, although it is too far to form any direct interaction with the Lys23-Glu36 pair). The crystal structure for the KE variant in the Δ+PHS* background captured an "open" conformation of the nearby $\beta 1 - \beta 2$ strands, leading to the presence of several water molecules surrounding the further separated Lys23-Glu36 pair (Fig. 1c). In the EK mutant, the sidechains of Glu23 and Lys36 are able to form hydrogen bonding interactions with nearby residues (main chain of Gly20 with amine in Lys36 and sidechain of Thr62 with carboxylate of Glu23), while no such interactions are visible in the crystal structure for the KE mutant.

Double mutant cycles involving both titratable (K, E) and polar (Q) mutations were used to estimate the interaction free energy ($\Delta\Delta G_{int}$) between Glu and Lys in the EK and KE variants. The $\Delta\Delta G_{int}$ value estimated for the KE variant was $\sim$-2 kcal/mol in both the Δ+PHS and Δ+PHS* backgrounds; this value was close to the $\Delta\Delta G_{int}$ measured for the KQ variant, suggesting that Lys23 and Glu36 are likely charge-neutral in the KE variant. By contrast, the $\Delta\Delta G_{int}$ for the EK variant was -3.4 kcal/mol more stable than its neutral analogue in the QK variant, which appears to suggest that Glu23 and Lys36 form a charged ion-pair or salt-bridge, an assignment that is consistent with the presence of multiple polar interactions with surrounding water molecules and amino acids in the corresponding crystal structure[32] (Fig. 1a).

As discussed in the original study of Garcia-Moreno and co-workers,[33] the observation of a completely dry environment surrounding Lys23-Glu36 in the crystal structure of the KE variant in the Δ+PHS background was rather unexpected. To what degree this is due to the crystalization condition warrants further analysis, considering the challenges associated with the determination of water penetration as discussed above. The suggestion that Lys23/Glu36 residues adopt neutral states in the KE variant while Glu23/Lys36 are both

in the charged states in EK also calls for further consideration; the implication to protein design is extraordinary if the presence of a few water molecules (two were resolved in the crystal structure[32]) plus minor adjustments in the protein microenvironment may qualitatively shift the ionization states of the two residues in a buried ion-pair. Along this line, we note that the crystal structure for the KE variant of the Δ+PHS* background captured several water molecules near the further separated Lys23-Glu36 pair (Fig. 1c), although the measured $\Delta\Delta G_{int}$ value of -2.2 kcal/mol is comparable to that in the Δ+PHS background, hinting at a neutral Lys-Glu pair[33] (however, see discussion below).

Motivated by these considerations, we have carried out extensive molecular dynamics simulations of the EK and KE variants of SNase in the Δ+PHS background. In addition to an evaluation of hydration level of the protein cavity that bears the Glu-Lys pair, we also conduct alchemical free energy simulations to estimate differences in folding stability among the EK and KE variants as well as relative to the Δ+PHS background. The results clearly indicate that, despite potentially higher levels of hydration in the protein cavity as compared to the corresponding crystal structures, the Glu-Lys residues are most likely in the charge neutral states in both EK and KE variants. This assignment is consistent with the computed local dielectric constant of the partially hydrated cavity, which is too low to stabilize the ionized form of the Glu-Lys pair. The magnitude of the effect is sufficiently large that an explicit treatment of electronic polarization is not expected to alter the favored charge state. These results have major implications to the introduction of titratable residue pairs into protein interior for protein engineering applications, and highlight the value of free energy simulations in guiding such efforts.

2 Computational Methods

2.1 Simulation set-up

Crystal structures for the highly stable Δ+PHS variant of SNase, its EK and KE mutants (with PDB ID: 3BDC,[35] 3NHH[32] and 6AMF,[33] respectively) are used as starting structures; no ligand or Ca^{2+} is included. Missing residues 1-6 and 142-149 are modeled with CHARMM-GUI.[36] For the EK and KE variants, simulations are carried out with the introduced Glu-Lys residues either both ionized or charge neutral; the simulations are labeled as EK/KE and $EK_{neutral}$/$KE_{neutral}$, respectively.

The systems are assembled using the CHARMM-GUI solution builder.[36] For each system, the initial structure is solvated in a rectangular TIP3Pm[37,38] water box with a 10.0 Å of edge distance under periodic boundary conditions. 150 mM NaCl ions are randomly placed to neutralize the system and mimic the physiological condition. The initial simulation box size is around $76 \times 76 \times 76$ Å^3.

Molecular dynamics simulations are performed with the OpenMM package[39] with GPU-acceleration using the CHARMM36m[40,41] force field; as discussed in Sect.3.3, since the free energy difference between different charge states of the ion-pair is very large, we limit ourselves to additive force field simulations here and will report in-depth analyses using polarizable force fields[42,43] separately. Particle-mesh Ewald (PME)[44] with an Ewald error tolerance of 0.0005 is used to calculate electrostatic interactions. Van der Waals (vdW) interactions are treated with a non-bonded cutoff of 12 Å and a switch distance of 10 Å. Each system is first minimized for 5000 steps with the L-BFGS algorithm. Then, equilibration run in the NVT ensemble is carried out with restraints on backbone and side chain heavy atoms using a force constant of 400 kJ/mol/nm^2 and 40 kJ/mol/nm^2, respectively. Production runs are carried out in the NPT ensemble for 1000 ns at 303.15 K. All bonds involving hydrogen atoms are constrained using SHAKE,[45] allowing an integration time step of 2 fs using the Langevin integrator with a collision frequency of 1 ps^{-1}. Pressure is controlled to be 1 bar

with the MonteCarloBarostat and an update frequency of every 100 steps.

2.2 Charging free energy simulations

To compute *relative* folding stabilities, we use the commonly employed thermodynamic cycles,[46] which are explained in Sect.3.3 in greater detail. For this work, two major assumptions are made to make the computations feasible and efficient. First, we assume that the unfolded state has minimal structural features such that the local environment of the mutation site can be captured with a simple model of the corresponding residue in bulk solution; specifically for the problem at hand, it is reasonable to assume that residues 23 and 36 no longer interact with each other and feature similar local environments in the unfolded state for both EK and KE variants. Second, since the sidechain pairs involved in the analyses, Glu-Lys and Val-Leu, are comparable in size, we assume that the free energy changes are dominated by the electrostatic component of intermolecular interactions (see Table S4 and discussions below). With these assumptions, the relative folding free energy evaluations are reduced to (de)charging free energy simulations, which are straightforward to conduct by directly modifying the partial charges of the relevant sidechain atoms (see below).

There might be concern about the potential caveats using charging free energies for evaluating the difference in folding stability. The statistical errors related to the small difference between large electrostatic interaction energies might be significant for (highly) charged systems. In this particular study, however, the charging free energy approach is well-suited for several reasons. First, in all cases, the net charge associated with the perturbed sidechains is zero; as a result, the free energy derivative contains a large but fairly constant value that corresponds to the direct interaction between Glu and Lys due to the generally stable structure of the ion-pair in the protein cavity. The interaction between the ion-pair and the surrounding environment, including the fluctuating water molecules, is not exceedingly large. These are the reasons that the statistical errors associated with the computed free energy quantities are modest (see Table 2). In fact, the free energy derivatives from reg-

ular dual-topology alchemical free energy simulations would be comparable in magnitude, especially for the relative stability of EK and Δ+PHS, to those in our charging free energy simulations, except for the fairly constant interaction between Glu and Lys. Finally, a particularly attractive aspect of the charging free energy simulation approach is that GPU computation is straightforward. This allows the sampling of each λ window for hundreds of nanoseconds using OpenMM without the need of any code modification; as discussed below, such extensive sampling is crucial due to slow fluctuations of hydration level in the protein cavity, especially for intermediate λ windows (*vide infra*).

In the charging free energy simulations, the partial charges of the sidechain atoms in residues 23 and 36 are scaled by λ; 11 equally spaced λ windows (0.0, 0.1, ..., 0.9, 1.0) are used for each system, and each λ window is sampled for 240 ns. The charging free energy is computed using thermodynamic integration,[47]

$$\Delta G_Q = \int_0^1 \frac{\partial G_Q}{\partial \lambda} d\lambda = \int_0^1 \left\langle \frac{\partial U^{elec}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda, \tag{1}$$

in which

$$U^{elec}(\lambda) = \sum_{i \in mut} \sum_{j \in mut, j \neq i} \frac{\lambda q_i \lambda q_j}{r_{ij}} + \sum_{i \in mut} \sum_{j \in env} \frac{\lambda q_i q_j}{r_{ij}} + U^{elec}_{env}, \tag{2}$$

and therefore $\langle \partial U^{elec}(\lambda)/\partial\lambda \rangle_\lambda = \langle 2\lambda U^{elec}_{intra} + U^{elec}_{inter} \rangle_\lambda$. Here, U^{elec}_{intra} represents the electrostatic interaction energy within the mutation sites (residues 23 and 36) and U^{elec}_{inter} represents electrostatic interaction energy between the mutation sites and the protein/solvent environment.

During post-processing, snapshots from different λ windows (saved with a frequency of 5000 steps) and the unscaled (fully-charged) psf file are used to calculate the electrostatic interactions ($U^{elec}_{intra}, U^{elec}_{inter}$) using the INTE module in CHARMM;[48] although the trajectories are sampled using PME, the post-processing is done using extended electrostatics[49] with a cutoff of 14 Å. To estimate statistical uncertainty, we calculate the statistical inefficiency, s_λ, for $\partial U^{elec}(\lambda)/\partial\lambda$ in each λ window using the pymbar package;[50] s_λ determines the lower bound of an interval, beyond which two data points are statistically independent. The

standard deviation in $\left\langle \partial U^{elec}(\lambda)/\partial\lambda \right\rangle_\lambda$ can then be written as,

$$\sigma(\left\langle \partial U^{elec}(\lambda)/\partial\lambda \right\rangle_\lambda) = \sqrt{\frac{s_\lambda \sigma^2(\partial U^{elec}(\lambda)/\partial\lambda)}{n_\lambda}}, \tag{3}$$

where $\sigma^2(\partial U^{elec}(\lambda)/\partial\lambda) = \frac{1}{n_\lambda - 1}\sum_{\tau=1}^{\tau=n_\lambda}(\partial U^{elec}(\lambda)/\partial\lambda - \left\langle \partial U^{elec}(\lambda)/\partial\lambda \right\rangle_\lambda)^2$ is the sample variance and n_λ is the number of snapshots used in the calculation. Eventually, the standard deviations in each window are propagated to obtain the statistical uncertainty for ΔG_Q.

As mentioned above, we assume that the interactions between residues 23 and 36 are negligible in the unfolded state. The charging free energy simulations for the unfolded states, therefore, are conducted in a water droplet that contains only one sidechain analog (23 or 36) with acetylated N-terminus and methylamidated C-terminus as blocking groups; the use of a water droplet avoids technical complexities associated with charging free energy simulations using Ewald summation.[51] Each blocked sidechain analog is solvated in a 25 Å radius water droplet with TIP3Pm water. A weak restraint is applied to keep the solute in the center of the droplet without any special solvent boundary potential. Each system is minimized for 50 steps with steepest descent and then 5000 steps with adapted basis Newton-Rhapson. Non-bonded interactions are computed with extended electrostatics[49] and a cutoff of 14 Å, beyond which the electrostatic interactions are computed with group-based multipoles. The post-processing protocols for the unfolded states are the same as above for the folded state. A Born correction term[52] (R=25 Å) for the droplet is included for charged Glu and Lys simulations.

2.3 Structural analysis

For structural analysis, all frames are aligned against the corresponding crystal structure to remove overall translations and rotations of the protein. MDAnalysis[53] is used to calculate the RMSD, RMSF, and the number of water molecules in the cavity. The center of the cavity is defined as the center of geometry for the OE1, OE2 atoms in Glu and the NZ

atom in Lys; for the Δ+PHS system, the cavity center is defined as the center of geometry for the CG1, CG2 atoms in Val, CD1 and CD2 atoms in Leu. The radius of the cavity is set to be 6 Å from radial distribution function of water molecules with respect to the cavity center. CHARMM[48] is used for the dipole moment analysis and the calculation of electrostatic interactions. Secondary structure for residues is assigned based on backbone amide and carbonyl positions using the DSSP algorithm implemented in MDtraj.[54] VMD[55] is used to visualize the trajectories.

2.4 Estimate of local dielectric constant

To better understand the ability of the protein microenvironment to stabilize charge, we compute the local dielectric constant using the Kirkwood-Fröhlich model.[56,57] Since the radius of the cavity, r_1, is much smaller than the effective radius of the protein, a simplified equation is used:[10]

$$G = \frac{\left\langle \Delta M_p^2 \right\rangle}{k_B T r_1^3} = \frac{(\epsilon_1 - 1)(1 + \epsilon_2)}{\epsilon_1 + 2\epsilon_2}, \tag{4}$$

where G is the Kirkwood G factor, $\left\langle \Delta M_p^2 \right\rangle$ is the fluctuation of the collective dipole moment of water molecules in the cavity computed from simulation trajectories, k_B is the Boltzmann constant, T is the temperature. The cavity radius, r_1, is taken to be 6 Å; ϵ_2 is the dielectric constant for the surrounding protein and taken to be either 10 or 20 considering the proximity of the cavity to the bulk solvent. The value of ϵ_1 is then computed based on Eq. 4.

3 Results and Discussion

3.1 Local structural stability and cavity hydration depend on ion-pair orientation and charge state

Overall, the structure remains stable during the simulation for both the Δ+PHS background and the two ion-pair variants, regardless of the charge state of the Glu-Lys pair introduced into the EK/KE mutants. The RMSD relative to the starting crystal structure hovers around 1 Å for the Δ+PHS background and $EK_{neutral}$/$KE_{neutral}$, in which the Glu/Lys residues are charge-neutral (Fig. 2a); for the cases in which Glu/Lys residues are charged, the RMSD values are slightly higher but usually below 2 Å. The RMSF profiles (Fig. 2c) are also similar among the five simulated systems, with the only exception being notably higher fluctuations for residues around Gly20 and His115 in the KE variant with both Lys/Glu being charged. The latter is also reflected in the secondary structure comparison shown in Fig. 2b,d, which indicates that residues 16-25 mostly consist of extended strands, bend, and hydrogen-bonded turns in the EK variant, while the KE simulation exhibits more irregular structures such as loop and isolated β-bridge; the hydrogen bonding interactions between main chains of Ile18, Asp19 and Thr22 are well-maintained in EK but much more dynamical in the KE simulations (Fig. S5a-b). In addition, residues 105-115 also undergo bend to loop transition between EK and KE simulations, although this region is not in the immediate vicinity of the Glu-Lys pair. By comparison, the secondary structures remain highly stable in simulations for the Δ+PHS background and $EK_{neutral}$/$KE_{neutral}$ (see Figs. S1-S2).

We note that NMR studies[33] indicated that the KE variant in the Δ+PHS background features a considerable population of unfolded conformation. This is apparently difficult to sample in the MD simulations at the μs time scale, as reflected by the low RMSD values for the KE variant, especially when both Glu/Lys residues are charge-neutral. The KE variant in the Δ+PHS* background was observed to largely adopt the native fold in solution, and the crystal structure captured an open conformation of the $\beta 1 - \beta 2$ strands, which exposed

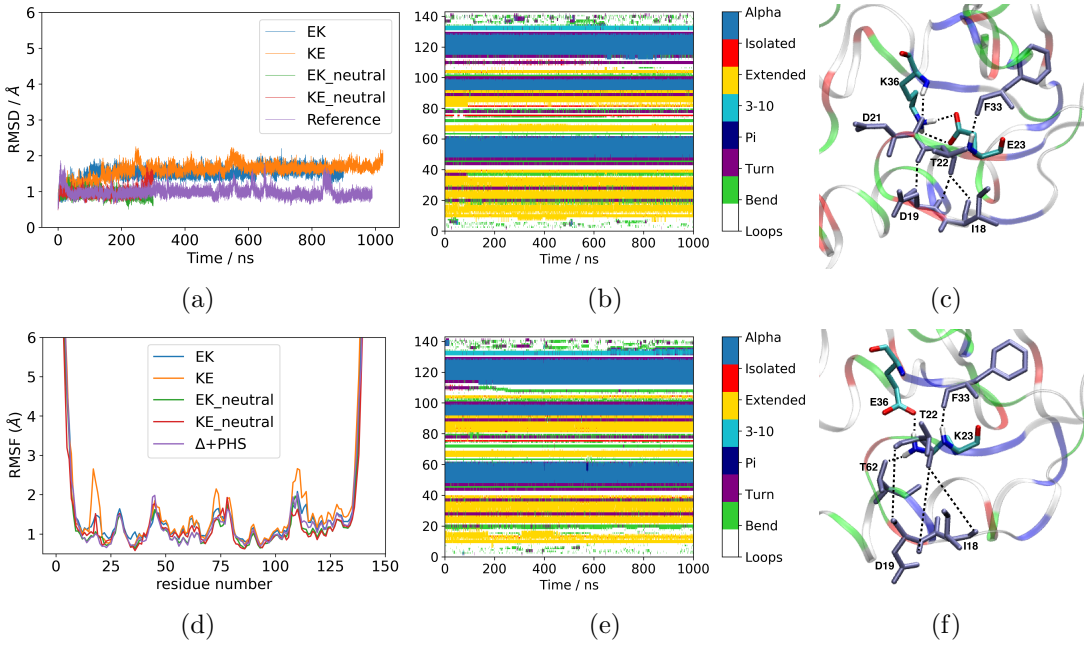


Figure 2: Structural properties of SNase during MD simulations. (a) Root Mean Squared Deviation (RMSD) relative to the corresponding crystal structures as a function of simulation time; (b) secondary structure evolution during the MD simulation for the EK variant; (c) a snapshot of EK at 600 ns that illustrates the stable hydrogen-bonding interactions near the Glu-Lys pair; water molecules are excluded for clarity; (d) Root Mean Squared Fluctuation (RMSF) relative to the average structure; (e) secondary structure evolution during the MD simulation for the KE variant, which exhibits structural instability in residues 15-20 and 115-120 (see text); (f) A snapshot of KE at 600 ns that illustrates the widened $\beta 1 - \beta 2$ turn due to broken main-chain interactions, possibly driven by the formation of a hydrogen-bonding interaction between Lys23 and Thr62. For secondary structure evolution of the Δ+PHS background and results for $EK_{neutral}$/$KE_{neutral}$, see **Supporting Information**.

the ion-pair to solvent. In our KE simulations, while the hydrogen bonding interactions between the $\beta1-\beta2$ strands are less well-maintained and therefore leading to a wider turn (compare snapshots in Fig. 2c, f), the open conformation of the $\beta1-\beta2$ strands is not observed. In the $KE_{neutral}$ simulations, backbone hydrogen bonds that hold the $\beta1-\beta2$ strands are substantially more stable (Fig. S6b).

At the sidechain level, both residues 23 and 36 exhibit a considerable level of flexibility in all simulations as illustrated by the χ angles (Fig. S3), with the χ_1 angle of residue 36 being most stable regardless of the amino acid identity (Glu or Lys) or the charge state. The distance between Glu and Lys also undergoes mild fluctuations in all simulations. In EK variant simulations, the most populated distances are $\sim$3 Å and close to the value in the crystal structure regardless of the charge state of the Glu-Lys pair (Fig. S4a). By contrast, in the KE simulations, the dominant distance is $\sim$2.8 Å, which is much shorter than the value of 3.8 Å in the crystal structure; in the $KE_{neutral}$ simulations, the distance is substantially longer and fluctuates between 4 and 6 Å (Fig. S4b). When both Glu/Lys are charged, the hydrogen-bonding interactions between them and nearby water or protein groups observed in the crystal structure (e.g., with Gly20 and Thr62 as shown in Fig. 1a) are not very stable (Fig. S5c-f) and readily replaced by interactions with penetrated water molecules (see below). In KE, the side chain hydroxyl group of Thr62 rotates toward the side chain of Lys23, and hydrogen bonds occasionally form between HZ1 of Lys23 and OG1 of Thr62 or between NZ of Lys23 and HG1 of Thr62. In the $EK_{neutral}$ and $KE_{neutral}$ simulations, hydrogen-bonding interactions between the Glu-Lys pair with nearby protein groups also undergo fluctuations (Fig. S6), albeit with higher populations that feature short distances due likely to the lower level of hydration in the cavity.

As expected, the hydration level of the cavity that bears the Glu-Lys pair is sensitive to its polarity. In the Δ+PHS background and charge-neutral Glu/Lys simulations, the hydration level remains low and generally similar to the corresponding crystal structures. For example, the $EK_{neutral}$ and $KE_{neutral}$ simulations feature typically $\leq$3 and $\leq$1 water,

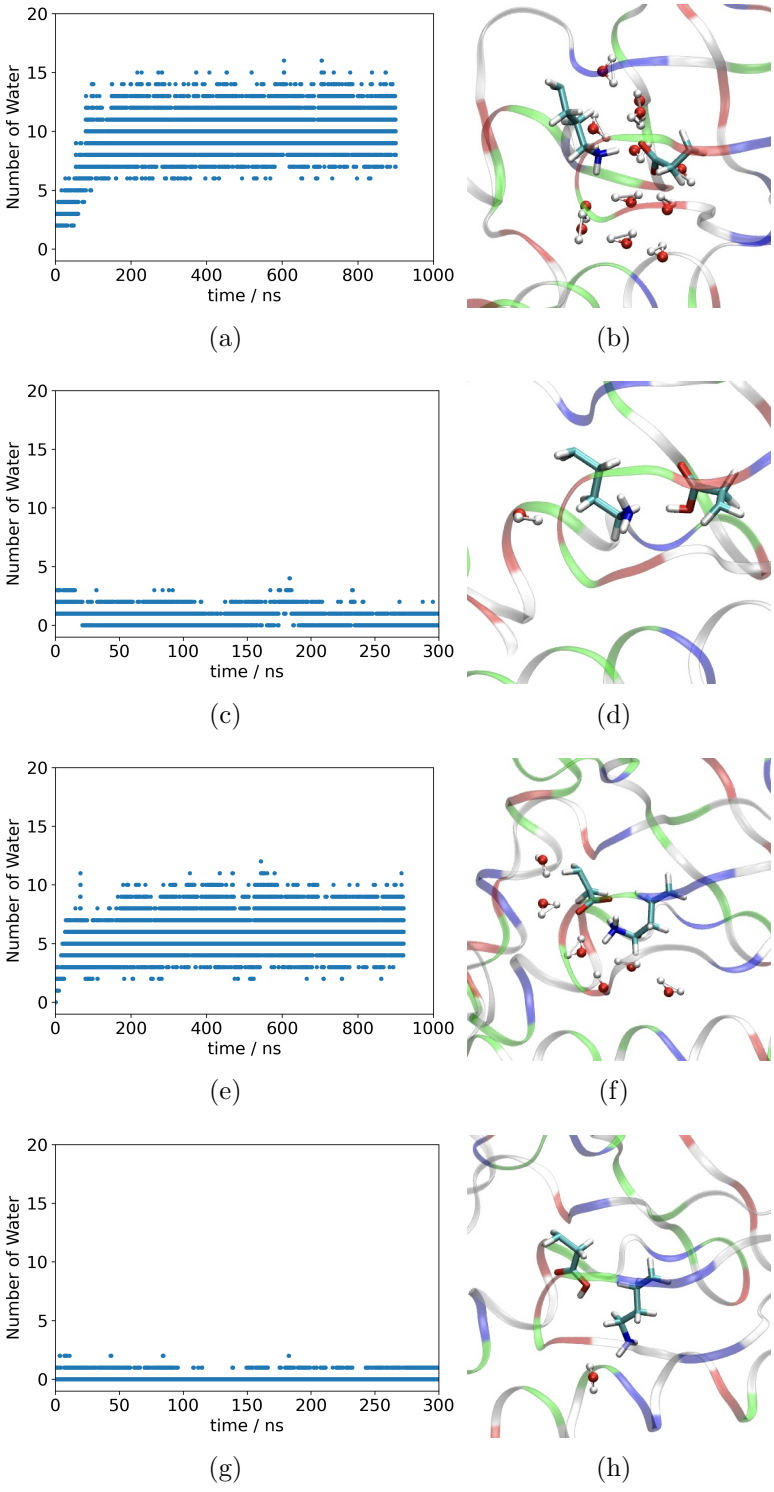


Figure 3: Hydration level of the protein cavity that bears the Glu-Lys pair during the MD simulation. Panels a, c, e, g show the results for EK, $EK_{neutral}$, KE, $KE_{neutral}$, respectively. The right panels show corresponding snapshots by the end of simulations (last frame) to illustrate the structure and hydration environment of the Glu-Lys pair.

respectively; in the corresponding crystal structures, two water molecules were resolved for the EK variant[32] and none for the KE variant[33] (in the Δ+PHS background). When both Glu/Lys residues are charged, the hydration level is substantially higher for the EK and KE variants; the number of water molecules in the cavity reaches ∼15 and ∼10, respectively. As shown in Fig. 3a,e, significant water penetration takes place immediately and reaches a steady level after ∼100 ns for EK and faster (<50 ns) for the KE variant. The fact that a larger number of water molecules may occupy the cavity in the EK variant reflects the subtle structural differences from the KE variant; in particular, the different packings of Glu-Lys sidechains lead to distinct regions available for water molecules, which are able to form a hydrogen-bonding network on both sides of the Glu23 carboxylate in EK, but only a chain of water along one side of the Glu-Lys pair in KE.

3.2 The low-dielectric protein cavity and the linear response approximation

Considering the significant number of water molecules that penetrate into the cavity in mutants that bear a pair of charged Glu/Lys residues (Figs. 3a,e), it is of interest to characterize their ability to stabilize buried charges. To this end, we have estimated the local dielectric constant (ϵ_1) for the cavity based on the Kirkwood G-factor associated with the cavity water molecules (see Sect. 2.4); two different values (10, 20) are used for ϵ_2 to test its impact on the computed ϵ_1. As shown in Table 1, the computed ϵ_1 results are not sensitive to ϵ_2 and the values are low ($\sim$6) for both EK and KE variants; the values are expectedly even lower for the other systems, which have only 1-3 water molecules in the cavity. The generally low dielectric constant reflects the confined nature of the water molecules trapped in the cavity; for example, both experimental[58] and computational[59] studies of confined water in different geometries also observed significantly reduced dielectric constants, with values comparable to those in Table 1 for confining radii of 1 nm. These results suggest that the penetrated water molecules are expected to provide a limited degree of stabilization of buried charges, as analyzed explicitly in the next subsection.

Table 1: Computed Kirkwood G factor and local dielectric constant (ϵ_1) for cavity water.

System	G factor	ϵ_1^a	ϵ_1^b
EK	5.0	7.6	6.7
KE	3.9	5.7	5.3
$EK_{neutral}$	0.6	1.7	1.7
$KE_{neutral}$	0.5	1.5	1.5
Δ+PHS	0.4	1.4	1.4

a. Calculations of ϵ_1 are conducted following Eq. 4 with $\epsilon_2 = 10$. b. $\epsilon_2 = 20$.

Before diving into the results of charging free energy simulations, we note that the stabilization of charges in the protein cavity can be captured rather well with a linear response model,[60] regardless of the charge states of the Glu/Lys residues. This is illustrated by both the λ dependence of the free energy derivatives (Figs. 4a,c), which have linear correlation coefficients 0.95 for EK and 0.96 for KE, and the probability distributions of the "energy

gap"($\partial U^{elec}(\lambda)/\partial\lambda$, see Figs. 4b,d), which largely follow Gaussian statistics[60,61] for all λ windows; the width of the distribution is observed to increase significantly for λ larger than 0.5 (also see Fig. S9b), when the level of hydration in the cavity increases notably. The lack of any significant multi-modal distributions[62] is consistent with the lack of major structural variations of residues that occupy and surround the cavity (Fig. 2). The slight deviation from Gaussian distribution observed for some λ windows most likely reflects the slow convergence of the hydration level, which takes more than 100 ns to equilibrate, especially for intermediate λ windows (see Fig. S8).

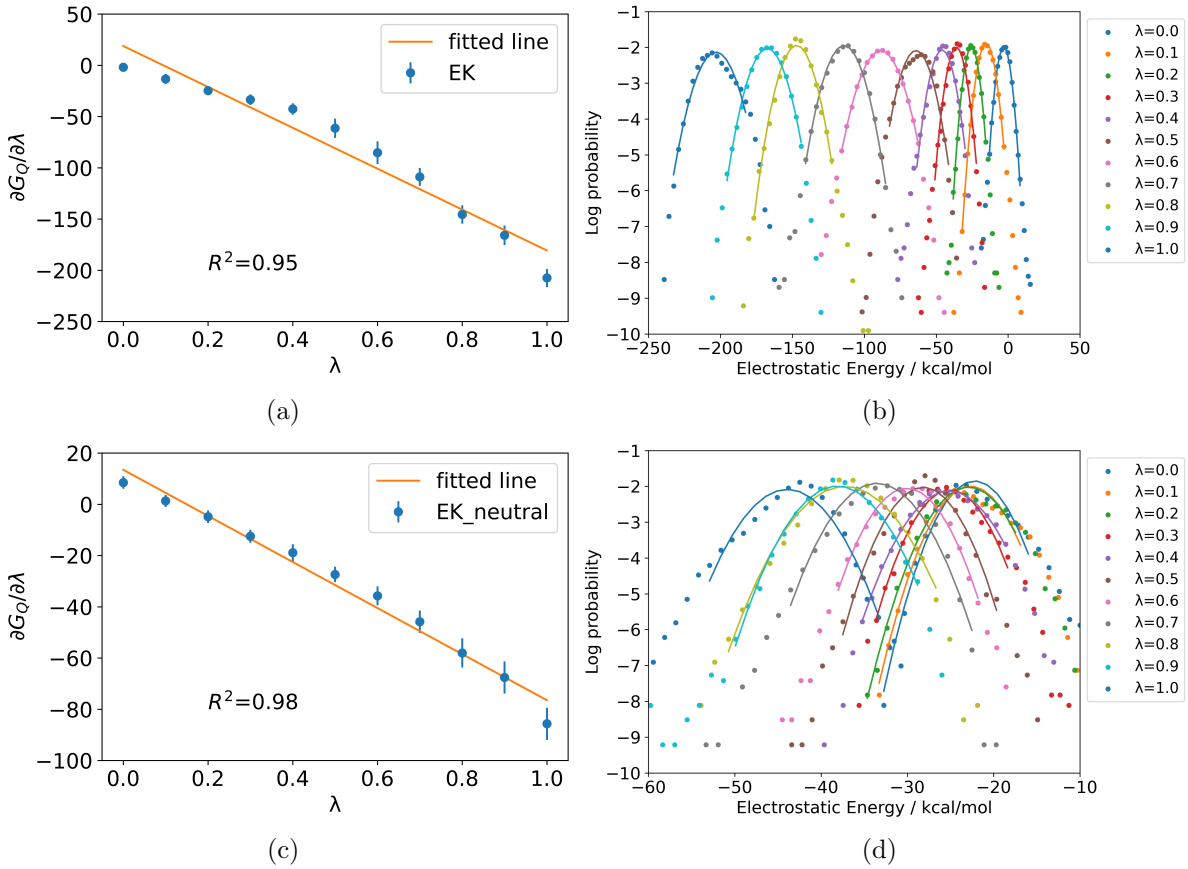


Figure 4: Results of charging free energy simulations for the (top) EK and (bottom) $EK_{neutral}$ simulations (corresponding plots for the KE variants are included in the **Supporting Information**). Left panels: λ dependence of the free energy derivative (blue dots) and the corresponding linear fit (orange lines); right panels: probability distributions of the "energy gap" ($\partial U^{elec}(\lambda)/\partial\lambda$, dots) and Gaussian fits (lines) to the distributions (note the logarithm scale for the distribution).

3.3 Free energy simulations strongly suggest reverse protonation for the buried ion-pair in both mutants

We start by evaluating the relative folding stability of the EK and KE variants following the thermodynamic cycle shown in Fig. 5; a similar set of thermodynamic cycle is applicable to the $EK_{neutral}$ and $KE_{neutral}$ pair. As discussed in Sect.2.2, by assuming that the unfolded state does not exhibit any significant structural feature (or more precisely, no major difference between the EK and KE variants in terms of the local environments of residues 23 and 36 in the unfolded state) and that van der Waals interactions do not contribute significantly to the free energy difference, the set of thermodynamic cycles suggests that the folding free energy difference can be cast into the difference in the decharging free energies ($\Delta G_{dQ} = -\Delta G_Q$) for the Glu-Lys pair in the two mutants:

$$\Delta G_f^{KE} - \Delta G_f^{EK} = \Delta\Delta G_f^{EK/KE} = \Delta G_{dQ}^{F(EK)} - \Delta G_{dQ}^{F(KE)}. \tag{5}$$

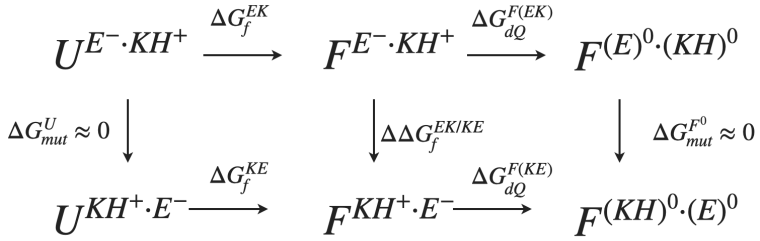


Figure 5: The thermodynamical cycle used to compute the relative stability $\Delta\Delta G_f^{EK/KE}$ of the EK and KE variants in which both Glu and Lys are charged; a similar cycle applies to the $EK_{neutral}$/$KE_{neutral}$ pair. The label U represents the unfolded state, while F represents the folded state. The superscript 0 indicates that the partial charges of residues 23 and 36 are set to zero. $\Delta G_{dQ}^{F(EK)}$ and $\Delta G_{dQ}^{F(KE)}$ are the de-charging free energies of residues 23 and 36 in the EK and KE variants, respectively ($\Delta G_{dQ} = -\Delta G_Q$). With the key assumptions made in this study (see Sect. 2.2), the stability of the unfolded state is identical for the two variants (i.e., $\Delta G_{mut}^{U} \approx 0$); similarly, stability of the folded state in which residues 23 and 36 are decharged is also identical for the two variants (i.e., $\Delta G_{mut}^{F^0} = 0$), which is further supported by Figs. S10-S11 and explicitly calculated van der Waals interactions between the EK/KE sidechains and the surrounding environment (Table S4).

Table 2: Relative thermodynamical stability (in kcal/mol) from alchemical free energy simulations and experiments.[a]

$\Delta\Delta G_f$	CHARMM36m	Exp.[b]
EK → KE	2.9±1.6	2.0±0.6
$EK_{neutral}$ → $KE_{neutral}$	1.1±0.6	
Δ+PHS → $EK_{neutral}$	15.9±0.6[c]	9.3±0.4
Δ+PHS → EK	55.8±1.6[c]	
Δ+PHS → $KE_{neutral}$	17.0±0.1[c]	11.3±0.4
Δ+PHS → KE	58.7±1.0[c]	

a. Individual free energy components following thermodynamic cycles shown in Figs. 5 and 6 are summarized in **Supporting Information**. b. Experimental values are the thermodynamical stabilities measured by chemical denaturation in Ref. 33. c. The values are further reduced by ∼4.4 kcal/mol when the difference between EK (or KE) and VL in terms of their van der Waals interactions with the surrounding environment (see Table S4) is included.

When both Glu/Lys residues are charged, the calculated $\Delta\Delta G_f^{EK/KE}$ is 2.9±1.6 kcal/mol; when the Glu/Lys residues are charge-neutral, the computed value is 1.1±0.6 kcal/mol. Considering the statistical uncertainty in the calculated and experimental values (Table 2), both of these results are in fair agreement with the experimental measurement of 2.0±0.6 kcal/mol. Therefore, the protein microenvironment preferentially favors the Glu23-Lys36 pair, regardless of charge state of the residues. Although the interaction between the Glu-Lys pair and the protein environment is expected to be substantially stronger when both residues are charged, the higher hydration level drawn by the charged residues apparently screens the interactions significantly. Therefore, the persistent preferential stabilization of the Glu23-Lys36 pair over the reverse ion-pair, regardless of charge states of both residues, clearly reflects the pre-organized nature of the surrounding protein structure, especially the polar backbone atoms.[34]

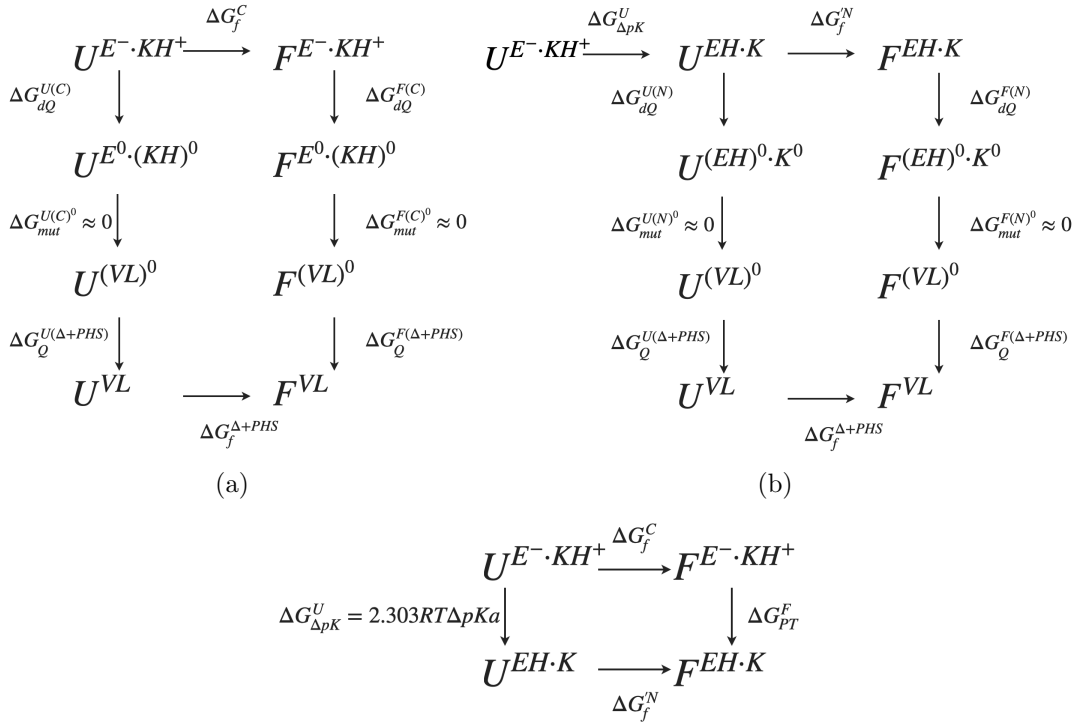


Figure 6: The thermodynamical cycles used to compare the stabilities of EK and the Δ+PHS background. (a) The pair of Glu/Lys residues are in the charged (C) states. VL represents the original residues 23, 36 in Δ+PHS. $\Delta G_{mut}^{(U/F)^0}$ is the free energy of mutating de-charged Glu-Lys to de-charged Val-Leu in the unfolded/folded states, respectively; they are set to zero considering the similar sizes of the sidechain pairs (also see Figs. S10-S11 and footnote c of Table 2). $\Delta G_Q^{U/F}$ represents the charging free energy of side chains in residues 23, 36 in the unfolded (U) and folded (F) state, respectively, and $\Delta G_{dQ}^{U/F}$ represents the decharging free energy of these sidechains. (b) The corresponding thermodynamical cycle for $\text{EK}_{neutral}$ in which the Glu/Lys residues are in their charge neutral (N) states; note that for comparing to experimental folding stabilities, pK_a correction for the unfolded state needs to be taken into consideration (see text). (c) The thermodynamical cycle used to explicitly compare the relative stability of the Glu-Lys pair when both residues are in their charged or neutral states; the difference between $\Delta G_f^{\prime N}$ and ΔG_f^C is computed based on the thermodynamic cycles shown in (a-b), and $\Delta G_{\Delta pK}^U$ is computed based on the experimental pK_a values of Glu and Lys in bulk solution under neutral pH.

Since the computed $\Delta G_f^{EK/KE}$ values do not provide any distinction between the charge states of the Glu/Lys residues, we next compare the stability of the EK mutant relative to the Δ+PHS background. For this purpose, we employ the thermodynamic cycles shown in Figs. 6a-b, which apply to the cases in which both Glu/Lys residues are in the charged and

neutral states, respectively. When they are both charged, for example, we have,

$$\Delta G_f^{\Delta+PHS} - \Delta G_f^C = \Delta G_{dQ}^{F(C)} + \Delta G_{mut}^{F(C)^0} + \Delta G_Q^{F(\Delta+PHS)} - \Delta G_{dQ}^{U(C)} - \Delta G_{mut}^{U(C)^0} - \Delta G_Q^{U(\Delta+PHS)} \tag{6}$$

$$\approx \Delta G_{dQ}^{F(C)} + \Delta G_Q^{F(\Delta+PHS)} - \Delta G_{dQ}^{U(C)} - \Delta G_Q^{U(\Delta+PHS)}, \tag{7}$$

in which we approximate $\Delta G_{mut}^{F(C)^0}$ and $\Delta G_{mut}^{U(C)^0}$ as zero; this is expected to be a reasonable approximation as the E/K and V/L sidechains are of similar sizes and therefore expected to have similar van der Waals interactions with the local environment (see below). We note that for the $\text{EK}_{neutral}$ system, the thermodynamic cycle shown in Fig. 6b involves a protonated Glu and deprotonated Lys in the unfolded state ($U^{EH\cdot K}$), thus the computed free energy difference ($\Delta G_f^{\Delta+PHS} - \Delta G_f^{'N}$) should be corrected by the free energy difference relative to the expected protonation states of these residues at neutral pH ($U^{E^-\cdot KH^+}$), which is $\sim$ 9 kcal/mol.

Following these schemes, the computed relative stability between the EK variant and the Δ+PHS background is a staggering 55.8$\pm$0.6 kcal/mol when both Glu/Lys residues are charged, while it is 15.9$\pm$0.6 kcal/mol when they are charge neutral; these values are further reduced by $\sim$4.4 kcal/mol when the difference between Glu-Lys and Val-Leu in terms of their van der Waals interactions with the surrounding environment (see Table S4) are included. The value for $\text{EK}_{neutral}$ compares favorably with the experimental value of 9.3$\pm$0.4 kcal/mol (Table 2), considering the approximations made in our thermodynamic cycles and the non-polarizable nature of the force field.

An alternative way of looking at these free energy results is that the difference between ΔG_f^C and $\Delta G_f^{'N}$ can be used to compute the relative stability of the Glu-Lys pair when the

residues adopt different charge states. In particular, following Fig. 6c,

$$\Delta G^{F}_{PT} = \Delta G^{U}_{\Delta pK} - \Delta G^{C}_{f} + \Delta G^{\prime N}_{f} = 2.303RT\Delta pK_a(Glu, Lys) + (\Delta G^{\Delta+PHS}_{f} - \Delta G^{C}_{f}) - (\Delta G^{\Delta+PHS}_{f} - \Delta G^{\prime N}_{f}). \tag{8}$$

Using the $\Delta G^{\Delta+PHS}_{f} - \Delta G^{C}_{f}$ and $\Delta G^{\Delta+PHS}_{f} - \Delta G^{\prime N}_{f}$ values just discussed, we obtain ΔG^{F}_{PT}=-39.9±1.3 kcal/mol, which suggests that the charge-neutral states of Glu/Lys residues in the EK variant are strongly favored than the charged states.

3.4 Discussion

3.4.1 Burial of an ion-pair in protein cavity is energetically costly

The most compelling result that has emerged from the current computational analysis is the significant amount of energetic penalty for burying a pair of charged residues in a protein cavity, even when the the latter is hydrated with tens of water molecules. This result can be understood by realizing that these confined water molecules have limited ability to reorient and therefore feature a relatively low effective dielectric constant of less than 10 (see Table 1). Using a simple Onsager reaction field model with a spherical cavity,[63,64] the solvation free energy for a dipole is given by,

$$\Delta G_{slv} = -\frac{\mu^2}{a^3}\frac{\epsilon - 1}{2\epsilon + 1}, \tag{9}$$

in which μ is the dipole moment of the solute (ion-pair), a the radius of the cavity that encloses the dipole and ϵ the surrounding dielectric constant. With the simplest model that μ and a remain constant in different environments, transferring the dipole from bulk solution ($\epsilon \sim 80$) to a low-dielectric environment ($\epsilon \sim 7$) reduces the solvation free energy by 20%; considering that the solvation free energy of a pair of charged residues in solution is highly favorable, the desolvation free energy penalty is significant. Using realistic molecular models, the charging free energy for a pair of charged Glu/Lys residues in solution using

the CHARMM36 force field is about -130 kcal/mol (which is consistent with the solvation free energies of Glu/Lys sidechains,[65] see Table S2), while the corresponding values in EK and KE variants are around -80 kcal/mol (see Table S1), pointing to even more significant amount of desolvation penalty.

By adopting the reverse protonation states thus burying a pair of charge-neutral residues clearly does not cause as large a desolvation penalty. Using the Glu-Lys pair as an example, charging free energies of the neutral pair in solution and protein cavity differ only by $\sim$ 7 kcal/mol (Table S1). However, as evident in the thermodynamic cycle in Fig. 6b, adopting the reverse protonation states is penalized by the pK_a difference of the residue pair in solution, which is a significant amount ($\sim$ 9 kcal/mol for Glu-Lys); i.e., burying a pair of charge-neutral titratable residues is also energetically costly and therefore requires stabilization of the protein interior by polar interactions.

The current work employed a popular fixed-charge force field (CHARMM36) along with the TIP3Pm water model.[37,38] It has been observed that TIP3Pm based protein force fields tend to favor overly collapsed structures (e. g., with over-estimated helical contents)[66] compared to experimental values from small angle X-ray scattering and fluoresence resonance energy transfer spectroscopy, although such discussions have largely focused on unfolded or intrinsically disordered proteins.[67–70] The radius of gyration of the protein remains substantially higher than that of the crystal structure for all the variants analyzed here (see Fig. S7), thus there is no compelling evidence suggesting that the use of CHARMM36/TIP3Pm leads to over-compact structures and thus underestimated level of cavity hydration in the current systems. Moreover, the magnitude of the desolvation penalty estimated for the charged Glu-Lys pair is so large ($\sim$50 kcal/mol) that including electronic polarization explicitly in the force field is not expected to reverse the trend; indeed, energy decomposition analyses using different quantum mechanical schemes consistently suggest that electronic polarization usually contributes up to 20% of intermolecular interactions for charged systems (e.g., a charged ion-pair).[71–73] Nevertheless, in the future, it is of interest to study the current set of systems,

along with several analyzed previously,[7,74] using popular polarizable force fields to establish and compare the quantitative contributions of different electronic polarization models[42,43,75] to the stabilization of buried charges and dipoles in protein cavities.

The magnitude (∼50 kcal/mol) of destabilization by the pair of charged Glu/Lys residues as compared to the Val-Leu pair observed in our analysis is substantially larger than the values reported in continuum electrostatic analyses.[7,74] The work of Tidor and co-workers[7] indicated that a majority of salt-bridges in the analyzed proteins, especially those buried in protein interior, were destabilizing (as compared to their hydrophobic isosteres) due to the large (up to ∼24 kcal/mol) desolvation penalty, which was defined to exclude any specific interactions involving the ion-pair residues and thus not directly comparable to the charging free energy difference of the ion-pair in different environments as computed in this work. The analysis of Nussinov and co-workers,[74] by contrast, observed that a large fraction of the analyzed salt-bridges in high-resolution crystal structures, were stabilizing, including a notable number of buried ones. Many of the buried stabilizing salt-bridges involve Arg, which features significant van der Waals interactions with the surrounding as well. The larger number of stabilizing salt-bridges observed in Ref. 74 was possibly due to the careful selection of ion-pairs that had geometries well-suited for favorable interactions.

Since the parameters used in these pioneering studies were very different from current atomistic force fields, and no structural relaxation (especially water penetration) was included in the continuum electrostatic analysis, it is difficult to quantitatively compare our work with these previous results. On the one hand, in light of the current findings, it is worth conducting free energy simulations to revisit some of the analyzed systems, especially concerning the ionization state of some of the most destabilizing salt-bridges. On the other hand, it is worth noting that all salt-bridges analyzed in those previous studies were in the context of naturally evolved proteins, which likely feature local environments that interact more favorably with the ion-pairs than the hydrophobic cavity of SNase (see next subsection); this is in line with analysis of interaction energies in proteins based on molecular dynamics

simulations, which found that many buried charged groups contribute favorably to protein enthalpic stability.[76] For example, for the buried Lys 42a-Asp 46a pair in uteroglobin, the overall contribution was found to be destabilizing[7] due to the significant (∼24 kcal/mol) desolvation penalty, the salt-bridge was nevertheless stabilized by helix-capping interactions that involve backbone carbonyl and C_αNH groups.

3.4.2 Factors of stabilization

In terms of the mechanism of polar stabilization, previous studies[7,13,14,25,34,74,77] and current calculations highlight roles of both amino acid mainchain/sidechain groups and water molecules that penetrate into the cavity. While the former have limited ability to reorient due to restrictions imposed by the protein structure,[34] the latter can generally reorient more readily and therefore accommodate different charge distributions. Nevertheless, water penetration into a small and relatively non-polar cavity has significantly energetic and entropic costs,[22] thus the degree of water penetration and amount of stabilization are likely dependent on the structural and chemical features of the buried charge and the surrounding cavity. Indeed, the EK and KE variants in which both Glu/Lys residues are charged exhibit different levels of cavity hydration (with a difference up to five water molecules) and different thermodynamic stabilities (∼1-2 kcal/mol). For $EK_{neutral}$ and $KE_{neutral}$, which feature much less hydration in the cavity and thus the pair of neutral Glu/Lys residues is stabilized mainly by interactions with protein groups, the difference in thermodynamic stability is less surprising.[33,34]

3.4.3 Analysis of ionization state of a buried pair of titratable residues

The simulation analysis highlights that it is difficult to determine the ionization state of a buried ion-pair based solely on structural features. Other than the somewhat expected higher level of water penetration, simulations with a pair of charged Glu/Lys residues do not exhibit major structural differences from those in which Glu/Lys are charge-neutral,

at least with unbiased simulations on the scale of microseconds, which are too slow to observe major structural unfolding events. Even the structural features of the ion-pair (e.g., distance separation between the carboxylate and the amine groups) are generally similar, with an exception of the KE simulation, in which the charged Glu-Lys separation is consistently shorter than that captured by the crystal structure of the variant in the Δ+PHS background.[33] Therefore, quantitative free energy simulations are required to firmly establish the ionization state and to thoroughly understand the origin of stabilization. Along this line, it is encouraging that such calculations are able to capture the modest (∼2 kcal/mol) difference in stability between the EK and KE variants, although it is remarkable that the difference is found to be rather insensitive to the charge state of the Glu-Lys pair.

As discussed in **Introduction**, $\Delta\Delta G_{int}$ measured by double-mutant cycle analyses[33] found that the interaction between Glu and Lys in the KE variant was similar to that in the KQ mutant, suggesting that the Lys/Glu residues in the KE variant are charge-neutral. By contrast, $\Delta\Delta G_{int}$ measured for the EK variant is more favorable by 3.4 kcal/mol than that for the QK mutant; this was interpreted to indicate that the Glu/Lys residues are charged in the EK variant. Our free energy simulations strongly suggest that the Glu/Lys residues are charge neutral, thus the measured difference in $\Delta\Delta G_{int}$ between EK and QK variants is likely due to differences in the microenvironment of the buried Glu-Lys vs. Gln-Lys pairs. An explicit analysis will benefit from a high-resolution structure of the QK variant, which is not yet available.

The structure of the KQ variant (i.e., Lys23-Gln36) has been solved,[33] which features a very dry cavity; therefore, if the QK mutant also features a dry cavity, the presence of additional water molecules in $EK_{neutral}$ might contribute to the difference in $\Delta\Delta G_{int}$, which is a thermodynamic quantity and therefore not always straightforward to interpret at the molecular level. As another example, $\Delta\Delta G_{int}$ measured for Lys-Glu interaction in the Δ+PHS* background was similar in magnitude as that for the KE variant in the Δ+PHS background, although the two KE mutants clearly exhibit rather different structures under

both crystalline and solution conditions.[33] Therefore, it is worthwhile to conduct free energy simulations for $\Delta\Delta G_{int}$ in these systems in future studies, which can provide much needed molecular level interpretation and serve as powerful benchmarks for molecular force fields, such as the roles of electronic polarization.

3.4.4 Further experimental connections

Reverse protonation of ion-pairs have been noted in several enzyme active sites and deemed essential to the catalytic mechanism,[78–80] although the reverse protonation state is often a minor population even at the pH of maximum activity, thus the situation is different from the EK/KE variants of SNase discussed here. The most direct way to establish the ionization states of titratable residues in protein interior is to employ spectroscopic techniques, such as NMR[30,31] and IR.[81] pH-dependent ^{13}C and ^{15}N chemical shifts of nearby residues and vibrational frequencies of the titratable groups are highly sensitive to the individual ionization state of the ion-pair, including the degree of proton delocalization between the residues.[29,82,83] Thus additional spectroscopic analyses for the EK/KE variants of SNase and other systems with buried salt-bridges[7,74] are of great interest regarding the physicochemical nature of the ion-pair. Moreover, the two interacting titratable sites involving cooperative protonations (e.g. ion-pairs) represent interesting cases to compare different microscopic, macroscopic and decoupled site representations and further clarify their connections in NMR titration studies.[31] Along this line, it is worth stressing that, in general, a destabilizing charged ion-pair relative to non-polar isosteres[7] does not necessarily suggest that the ion-pair prefers the charge-neutral ionization state; as evident from panel (c) of Fig. 6, the charge-neutral ionization state is favored (i.e., $\Delta G^F_{PT} < 0$) only when the folding stability difference between the two ionization states ($\Delta G'^N_f - \Delta G^C_f$) offsets the pK_a difference between the two titratable residues in solution ($\Delta G^U_{\Delta pK}$).

Finally, we note that in a recent study,[84] Kaila and co-workers have used computational design to engineer one or multiple ion-pairs into the interior of a set of highly stable helical

bundles. Structural characterizations using crystallography and NMR along with thermodynamic measurements supported the overall stability of these mutants with buried ion-pairs; in fact, one of the mutants exhibited impressive thermodynamic stability comparable to the original protein background. Compared to the situation in SNase, the local environment of the engineered ion-pair(s) is less hydrophobic, with polar residues such as Gln that form more extensive hydrogen-bonding interactions, especially when a pair of ion-pairs are simultaneously inserted. Moreover, burying a pair of ion-pairs next to each other gains additional stabilization due to favorable interactions between the ion-pairs. Nevertheless, bearing the observations from this study in mind, free energy analyses similar to those conducted here are expected to be informative in terms of the precise ionization states of the ion-pairs and nature of stabilization.

4 Conclusions

Buried titratable residues in protein cavities are often of major functional importance, thus understanding their properties and factors that stabilize them is of both fundamental and practical significance. One particularly challenging task is to determine the ionization states of buried charges, which are often difficult to determine based on structural studies alone. In this study, using extensive molecular dynamics simulations and free energy calculations, we investigate the properties of buried ion-pairs in a set of variants of *Staphylococcal* Nuclease, for which recent structural and thermodynamic studies appeared to suggest that both stability and ionization state of the buried Glu-Lys pair are sensitive to its orientation (i.e., Glu23-Lys36 vs. Lys23-Glu36).

Unbiased molecular dynamics simulations at the microsecond time scale show that the hydration level of the cavity depends on the orientation of the buried Glu-Lys pair therein as well as its ionization state; free energy simulations using a non-polarizable force field recapitulate the relative stability of EK vs. KE mutants measured experimentally, although

the difference is similar in magnitude regardless of the ionization states of the Glu/Lys residues. However, a complementary set of free energy simulations strongly suggests that, in contrast to the original suggestion in the experimental analysis, $EK_{neutral}$ is more stable than its ionized form (EK), and the free energy difference is significant that explicitly including electronic polarization is not expected to reverse the trend. This result is consistent with the calculated local dielectric constants of water in the cavity, which are low (< 10) compared to bulk water due to the confined environment. As a result, it remains difficult for the protein to stabilize the buried Glu-Lys pair when both residues are in the ionized form, even with water penetration. The current study highlights the role of free energy simulations in understanding the ionization state of buried titratable residues and the relevant energetic contributions. The study also underscores the importance of adequate sampling for such free energy simulations, due mainly to the kinetics of water penetration, which can be as slow as hundreds of nanoseconds even for a relatively shallow cavity in a small protein such as SNase.

In the future, it is worthwhile conducting additional free energy simulations to compare with experimental double mutant cycles, especially with polarizable force fields to calibrate the importance and accuracy of different polarizable models. It is also of great interest to conduct more detailed spectroscopic analysis of residues in the cavity so as to further understand the properties of buried ion-pairs and nearby water molecules.

Acknowledgement

This work was supported by NSF-CHE1829555. Computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number OCI-1053575, are greatly appreciated; part of the computational work was performed on the Shared Computing Cluster which is administered by Boston University's Research Computing Services (URL: www.bu.edu/tech/support/research/).

Supporting Information Available

Additional analysis of the charging free energy simulations, structural and dynamical properties of the EK/KE variants with different ionization states are included. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W.H. Freeman and Company, 1999.

(2) Wikström, M.; Sharma, V.; Kaila, V. R. I.; Hosler, J. P.; Hummer, G. New Perspectives on Proton Pumping in Cellular Respiration. *Chem. Rev.* **2015**, *115*, 2196–2221.

(3) Huang, P. S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320–327.

(4) Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190–194.

(5) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013**, *503*, 418.

(6) Lassila, J. K.; Baker, D.; Herschlag, D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4937–4942.

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

(8) Warshel, A. Computer simulations of enzyme catalysis: Methods, Progress and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

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

(10) Goyal, P.; Lu, J.; Yang, S.; Gunner, M. R.; Cui, Q. Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in Cytochrome c Oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 18886–18891.

(11) Zheng, Y. Q.; 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.

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

(13) Son, C. Y.; Yethiraj, A.; Cui, Q. Cavity Hydration Dynamics in Cytochrome c Oxidase and Functional Implications. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E8830–E8836.

(14) Kaila, V. R. I.; Wikström, M.; Hummer, G. Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6988–6993.

(15) Yabukarski, F.; Biel, J. T.; Pinney, M. M.; Doukov, T.; Powers, A. S.; Fraser, J. S.; Herschlag, D. Assessment of enzyme active site positioning and tests of catalytic mechanisms through X-ray-derived conformational ensembles. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 33204–33215.

(16) Fraser, J. S.; van der Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16247–16252.

(17) Ernst, J. A.; Clubb, R. T.; Zhou, H. X.; Gronenborn, A. M.; Clore, G. M. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* **1995**, *267*, 1813–1817.

(18) Yu, B.; Blaber, M.; Gronenborn, A. M.; Clore, G. M.; Caspar, D. L. D. Disordered water within a hydrophobic protein cavity visualized by x-ray crystallography. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 103–108.

(19) Freier, E.; Wolf, S.; Gerwert, K. Proton transfer via a transient linear water-molecule chain in a membrane protein. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11435–11439.

(20) Wolf, S.; Freier, E.; Cui, Q.; Gerwert, K. Infrared spectral marker bands characterizing a transient water wire inside a hydrophobic membrane protein. *J. Chem. Phys.* **2014**, *141*, 22D524.

(21) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006**, *106*, 3210–3235.

(22) Rasaiah, J. C.; Garde, S.; Hummer, G. Water in nonpolar confinement: From nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713–740.

(23) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the prediction of pK(a) values in proteins. *Proteins: Struct., Funct., & Bioinf.* **2011**, *79*, 3260–3275.

(24) Goh, G. B.; Laricheva, E. N.; C. L. Brooks III, Uncovering pH-Dependent Transient

States of Proteins with Buried Ionizable Residues. *J. Am. Chem. Soc.* **2014**, *136*, 8496–8499.

(25) Isom, D. G.; Castaneda, C. A.; Velu, P. D.; B. Garcia-Moreno E., Charges in the hydrophobic interior of proteins. *Proc. Nat. Acad. Sci. USA* **2010**, *107*, 16096–16100.

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

(27) Chimenti, M. S.; Khangulov, V. S.; Robinson, A. C.; Heroux, A.; Majumdar, A.; Schlessman, J. L.; Garcia-Moreno, B. Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein. *Structure* **2012**, *20*, 1071–1085.

(28) Ullmann, G. M. Relations between Protonation Constants and Titration Curves in Polyprotic Acids: A Critical View. *J. Phys. Chem. B* **2003**, *107*, 1263–1271.

(29) Molina, P. A.; Jensen, J. H. A Predictive Model of Strong Hydrogen Bonding in Proteins: The $N^{\delta 1}$-H-$O^{\delta 1}$Hydrogen Bond in Low-pH α-Chymotrypsin and α-Lytic Protease. *J. Phys. Chem. B* **2003**, *107*, 6226–6233.

(30) Chivers, P. T.; Prehoda, K. E.; Volkman, B. F.; Kim, B. M.; Markley, J. L.; Raines, R. T. Microscopic pK(a) values of Escherichia coli thioredoxin. *Biochem.* **1997**, *36*, 14985–14991.

(31) Hass, M. A. S.; Mulder, F. A. A. Contemporary NMR Studies of Protein Electrostatics. *Annu. Rev. Biophys.* **2015**, *44*, 53–75.

(32) Robinson, A. C.; Castaneda, C. A.; Schlessman, L. J.; Garcia-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. USA* **2014**, *111*, 11685–11690.

(33) Robinson, A. C.; Schlessman, J. L.; B. Garcia-Moreno E., Dielectric Properties of a Protein Probed by Reversal of a Buried Ion Pair. *J. Phys. Chem. B* **2018**, *122*, 2516–2524.

(34) Hwang, J. K.; Warshel, A. Why Ion Pair Reversal by Protein Engineering Is Unlikely to Succeed. *Nature* **1998**, *334*, 270–272.

(35) Castañeda, C. A.; Fitch, C. A.; Majumdar, A.; Khangulov, V.; Schlessman, J. L.; García-Moreno, B. E. Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease. *Proteins: Struct. Funct. & Bioinf.* **2009**, *77*, 570–588.

(36) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(38) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

(39) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; C. D. Stern et al., OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

(40) Huang, J.; A. D. MacKerell Jr., CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comp. Chem.* **2013**, *34*, 2135–2145.

(41) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; A. D. MacKerell Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2016**, *14*, 71–73.

(42) Lin, F. Y.; Huang, J.; Pandey, P.; Rupakheti, C.; Li, J.; Roux, B.; A. D. MacKerell Jr., Further Optimization and Validation of the Classical Drude Polarizable Protein Force Field. *J. Chem. Theory Comput.* **2020**, *16*, 3221–3239.

(43) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(44) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(45) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(46) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden thermodynamics of mutant proteins - a molecular dynamics analysis. *Science* **1989**, *244*, 1069–1072.

(47) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.

(48) Brooks, B. R.; III, C. L. B.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(49) Stote, R. H.; States, D. J.; Karplus, M. On the treatment of electrostatic interactions in biololecular simulation. *J. Chim. Phys.* **1991**, *88*, 2419–2433.

(50) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.

(51) Lin, Y. L.; Aleksandrov, A.; Simonson, T.; Roux, B. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2690–2709.

(52) Born, M. Volumen und Hydratationswärme der Ionen. *Zeitschrift für Physik* **1920**, *1*, 45–48.

(53) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(54) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.

(55) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(56) Barrat, J.-L.; Hansen, J.-P. *Basic concepts for simple and complex liquids*; Cambridge University Press: London, 2003.

(57) Simonson, T.; Perahia, D. Internal and interfacial dielectric-properties of cytochrome-c from molecular- dynamics in aqueous-solution. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1082–1086.

(58) Fumagalli, L.; Esfandiar, A.; Fabregas, R.; Hu, S.; Ares, P.; Janardanan, A.; Yang, Q.; Radha, B.; Tangiguchi, T.; Watanabe, K.; Gomila, G.; Novoselov, K. S.; Geim, A. K. Anomalously low dielectric constant of confined water. *Science* **2018**, *360*, 1339.

(59) Mondal, S.; Bagchi, B. Water in Carbon Nanotubes: Pronounced Anisotropy in Dielectric Dispersion and Its Microscopic Origin. *J. Phys. Chem. Lett.* **2019**, *10*, 6287–6292.

(60) Simonson, T. Gaussian fluctuations and linear response in an electron transfer protein. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6544–6549.

(61) Levy, R.; Belhadj, M.; Kitchen, D. B. Gaussian fluctuation formula for electrostatic free-energy changes in solution. *J. Chem. Phys.* **1991**, *95*, 3627–3633.

(62) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.

(63) Kirkwood, J. G. Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *J. Chem. Phys.* **1934**, *2*, 351–361.

(64) Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.

(65) Deng, Y. Q.; Roux, B. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004**, *108*.

(66) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528–552.

(67) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98–105.

(68) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field

for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4758–E4766.

(69) Best, R. B.; Zheng, W. W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non- Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

(70) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

(71) Kitaura, K.; Morokuma, K. New energy decomposition scheme for molecular-interactions within hartree-fock approximation. *Int. J. Quant. Chem.* **1976**, *10*, 325–340.

(72) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.

(73) Mao, Y.; Loipersberger, M.; Horn, P. R.; Das, A.; Demerdash, O.; Levine, D. S.; Veccham, S. P.; Head-Gordon, T.; Head-Gordon, M. From Intermolecular Interaction Energies and Observable Shifts to Component Contributions and Back Again: A Tale of Variational Energy Decomposition Analysis. *Annu. Rev. Phys. Chem.* **2021**, *72*, 1.

(74) Kumar, S.; Nussinov, R. Salt Bridge Stability in Monomeric Proteins. *J. Mol. Biol.* **1999**, *293*, 1241–1255.

(75) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications Annual Review of Biophysics. *Annu. Rev. Biophys.* **2019**, *48*, 371–394.

(76) Dadarlat, V. M.; Post, C. B. Contribution of charged groups to the enthalpic stabilization of the folded states of globular proteins. *J. Phys. Chem. B* **2008**, *112*, 6159–6167.

(77) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta* **2006**, *1764*, 1647–1676.

(78) Knuckley, B.; Bhatia, M.; Thompson, P. R. Protein Arginine Deiminase 4: Evidence for a Reverse Protonation Mechanism. *Biochem.* **2007**, *46*, 6578–6587.

(79) Frankel, B. A.; Kruger, R. G.; Robinson, D. E.; Kelleher, N. L.; McCafferty, D. G. *Staphylococcus aureus* Sortase Transpeptidase SrtA: Insight into the Kinetic Mechanism and Evidence for a Reverse Protonation Catalytic Mechanism. *Biochem.* **2005**, *44*, 11188–11200.

(80) Sims, P. A.; Larsen, T. M.; Poyner, R. R.; Cleland, W. W.; Reed, G. H. Reverse Protonation Is the Key to General Acid-Base Catalysis in Enolase. *Biochem.* **2003**, *42*, 8298–8306.

(81) Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem. Rev.* **2014**, *114*, 126–163.

(82) Phatak, P.; Ghosh, N.; Yu, H.; Cui, Q.; Elstner, M. Amino acids with an intermolecular proton bond as the proton storage site in bacteriorhodopsin. *Proc. Acad. Natl. Sci. USA* **2008**, *105*, 19672–19677.

(83) Goyal, P.; Ghosh, N.; Phatak, P.; Clemens, M.; Gaus, M.; M. Elstner,; Cui, Q. Proton storage site in bacteriorhodopsin: new insights from QM/MM simulations of microscopic pK_a and infrared spectra. *J. Am. Chem. Soc.* **2011**, *133*, 14981–14997.

(84) Baumgart, M.; Röpke, M.; Mühlbauer, M. E.; Asami, S.; Mader, S. L.; Fredriksson, K.;

Groll, M.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Design of buried charged networks in artificial proteins. *Nat. Comm.* **2021**, *12*, 1895.

Graphical TOC Entry