Biochemistry

Structural Basis for Reduced Dynamics of Three Engineered HNH Endonuclease Lys-to-Ala Mutants for the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Associated 9 (CRISPR/Cas9) Enzyme

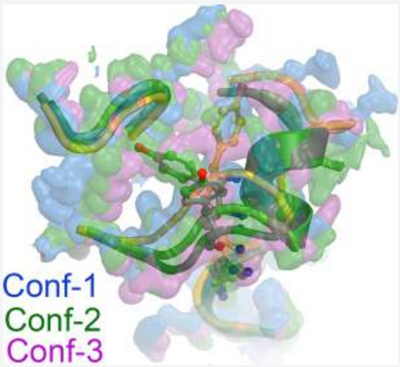
Jimin Wang,* Erin Skeens, Pablo R. Arantes, Federica Maschietto, Brandon Allen, Gregory W. Kyro, George P. Lisi,* Giulia Palermo,* and Victor S. Batista*

Cite This: *Biochemistry* 2022, 61, 785−794

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Many bacteria possess type-II immunity against invading phages or plasmids known as the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated 9 (Cas9) system to detect and degrade the foreign DNA sequences. The Cas9 protein has two endonucleases responsible for double-strand breaks (the HNH domain for cleaving the target strand of DNA duplexes and RuvC domain for the nontarget strand, respectively) and a single-guide RNA-binding domain where the RNA and target DNA strands are base-paired. Three engineered single Lys-to-Ala HNH mutants (K810A, K848A, and K855A) exhibit an enhanced substrate specificity for cleavage of the target DNA strand. We report in this study that in the wild-type (wt) enzyme, D835, Y836, and D837 within the Y836-containing loop (comprising E827-D837) adjacent to the catalytic site have uncharacterizable broadened $^{1}H^{15}N$ nuclear magnetic resonance (NMR) features, whereas remaining residues in the loop have different extents of broadened NMR spectra. We find that this loop in the wt enzyme exhibits three distinct conformations over the duration of the molecular dynamics simulations, whereas the three Lys-to-Ala mutants retain only one conformation. The versatility of multiple alternate conformations of this loop in the wt enzyme could help to recruit noncognate DNA substrates into the HNH active site for cleavage, thereby reducing its substrate specificity relative to the three mutants. Our study provides further experimental and computational evidence that Lys-to-Ala substitutions reduce dynamics of proteins and thus increase their stability.



INTRODUCTION

Discovery of bacterial innate immune systems against invading viruses and plasmids using the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) system has revolutionized genome editing technologies.[1,2] As a precision technology, a high degree of substrate specificity for target DNA sequences being cut is an important factor, even though innate immunity does not require such strict specificity.[3] Cas9 is one of the best studied cleavage systems, containing two endonucleases responsible for double-strand breaks and a single-guide (sg) RNA-binding domain for substrate selectivity.[4] The bound sgRNA recognizes the target DNA duplex by forming base pairs with the target DNA strand after displacing its nontarget DNA strand. This process is dynamic and very complex, starting with base pairing of the DNA seed region after unwinding of the duplex, followed by the repeat−antirepeat duplex formation, involving a series of conformational changes. Only after the full complex is formed are the two endonucleases positioned properly and activated to cut both DNA strands simultaneously. These endonucleases are the HNH domain, cleaving the target DNA strand that is base-paired with sgRNA, and RuvC for cleaving the nontarget DNA strand.

Cas9 has been engineered to improve substrate specificity through the design of better sgRNA sequences for accuracy and efficiency of DNA cleavage. Three single Lys-to-Ala mutations in the Cas9 HNH domain have also achieved this goal: K810A, K848A, and K855A.[5] The structural rationale for initially designing these mutations was to reduce potential electrostatic interactions with the hypothetically modeled nontarget DNA strand.[5] However, that hypothesis was not supported by follow-up structure determination of the Cas9-nucleic acid complexes; thus it remains unclear how the removal of putative electrostatic interactions with nucleic acids could have improved substrate specificity.[6]

Received: March 4, 2022
Revised: April 1, 2022
Published: April 14, 2022

https://doi.org/10.1021/acs.biochem.2c00127

Lys-to-Ala mutations, particularly surface lysine residues, have often been introduced by crystallographers for improved stability of meso-stable proteins for crystallization.[7] This is because large-scale sequence comparisons between thermostable and meso-stable proteins have shown that a reduction of nonessential surface lysine residues is a key feature responsible for enhanced stability of many thermostable proteins (i.e., Lys to Arg substitutions).[8−11] In this study, we provide a structural basis for reduced dynamics associated with three single Lys-to-Ala substitutions in the HNH domain of Cas9 that is beyond the local entropic effects of expected for these substitutions. Recent biophysical investigations have shown that the three HNH mutants, K810A, K848A, and K855A, disrupt allosteric signaling transfer of DNA binding information from the substrate recognition lobe to the cleavage sites.[12−14] This allosteric communication correlates with the specificity enhancement of the three single mutations, with the K855A mutant achieving the highest specificity and also strongly perturbing the allosteric pathway recently reported for HNH.[12−14] The nature of this correlation and the mechanism connecting the allosteric phenomenon to the improved substrate specificity remain intriguing. Establishing the relationship between allostery and the catalytic function of Cas9 is indeed an active area of research,[13] because knowledge of the allosteric relationship can help improving the catalytic efficiency.[15,16]

Following our recent studies,[12−14] we report here $^1H^{15}N$ NMR spectroscopy results for residues within a Y836-containing loop of the wild-type (wt) enzyme and the three mutants and compare them with molecular dynamics (MD) simulations for both the wt and mutant enzymes. We find that this loop, comprising E827-D837, has three distinct conformations in the wt enzyme, whereas the three mutants exhibit a single dominant conformation, consistent with well-known observations that Lys-to-Ala mutations of nonessential surface lysine residues can account for the reduce dynamics of the mutants and thus enhance stability. Given the fact that this loop is immediately adjacent to the HNH catalytic site, we discuss a possible correlation of the existence and reduction of multiple conformations of this loop with substrate selectivity. While the size of the Cas9 HNH domain is ideal for nuclear magnetic resonance (NMR) studies, it limits our ability to directly address the catalytic mechanism of this enzyme, which would require the binding of nucleic acid substrates as well as interactions with many other protein domains within the enzyme that are essential for the high-affinity binding of nucleic acids. Nonetheless, our NMR-based studies of the Cas9 HNH domain should be complementary to many other biophysical and biochemical studies of the intact Cas9 enzyme, including cryo-electron microscopy.[17−21]

■ MATERIALS, EXPERIMENTAL SECTION, AND COMPUTATIONAL METHODS

NMR Spectroscopy. Two-dimensional $^1H^{15}N$ NMR samples of the wt HNH domain (residues 775−908) of Cas9 from *Streptococcus pyogenes* and enhanced-specificity HNH mutants (K810A, K848A, K855A) were expressed in M9 minimal media containing MEM vitamins, $MgSO_4$, and $CaCl_2$ and supplemented with $^{15}NH_4Cl$ (Cambridge Isotopes Laboratories) and $^{12}C_6H_{12}O_6$ as the sole nitrogen and carbon sources, respectively. Cells were grown to an OD_{600} of 0.8−0.9, induced with 0.5 mM IPTG, and incubated for 16−18 h at 20 °C. Cells were then harvested by centrifugation and purified as described previously.[14] Briefly, cells were resuspended in 20 mM HEPES, 500 mM KCl, and 5 mM imidazole at pH 8.0 and lysed by ultrasonication. After centrifugation, the supernatant was purified via an Ni-NTA column, and HNH was eluted with 20 mM HEPES, 250 mM KCl, and 220 mM imidazole at pH 7.4. The N-terminal His-tag was cleaved by TEV protease and removed with a subsequent Ni-NTA column. NMR samples were dialyzed into a final buffer containing 20 mM HEPES, 80 mM KCl, 1 mM DTT, 1 mM EDTA, and 7.5% (v/v) D_2O at pH 7.4.

NMR experiments were performed on a Bruker Avance NEO 600 MHz spectrometer at 25 °C. NMR data were processed using NMRPipe[22] and analyzed in NMRFAM-SPARKY.[23] $^1H-^{15}N$ HSQC spectra were collected with the ^1H and ^{15}N carriers set to the water resonance and 120 ppm, respectively. Combined chemical shifts were determined by $\Delta\delta = \sqrt{[(\delta_{HN}^2/2) + (\delta_N^2/25)]/2}$. R_{ex} values were determined from the analysis of reported NMR relaxation data.[12]

MD Simulations and Analysis of the Resulting Maps. MD simulations started with the 4UN3 X-ray structure of the wt CRISPR-Cas9 from which three single mutations of K810A, K848A, or K855A were made.[5,24] All systems were solvated within a periodic box for ∼340,000 total atoms. A new AMBER ff99SBnmr2 force field was used, which improves the consistency of the backbone conformational ensemble with NMR experiments as shown in our previous NMR-MD studies.[25] Parameters for nucleic acids included the ff99bsc0 + χOL3 corrections.[12,14,26] The TIP3P model was used for water molecules.[27] Simulations were performed in an ensemble with temperature held at 310 K using the Bussi thermostat.[27] The pressure was held at 1 bar with the Parrinello−Rahman barostat.[28] An ∼1.2 μs-long trajectory was collected in three replicas for the wt CRISPR-Cas9 system and for each of the K810A, K848A, and K855A variants with a step increment of 2 fs as done previously.[29] Analysis of the first replica was performed after discarding the first ∼200 ns of MD, to enable appropriate equilibration and a reliable comparison. Overall results were confirmed in the second and third replicas. All simulations were performed using Gromacs (v. 2020).[30]

The HNH domain (2039 protein atoms) was extracted from the MD simulations. The electrostatic potential (ESP) and electron density (ED) maps were calculated for the domain from coordinates of individual MD frames as described elsewhere[31] and briefly summarized here, using the CCP4 package (sfall and fft).[32] Maps were added and averaged after certain alignment using the CCP4 suite (mapsig with lsqkab and pdbset).[32] The coordinates were placed in the center of a box of a fixed size for map calculations. The equilibrated structures of the MD simulations were independently interpreted from MD-ESP maps and partially refined using both Phenix and CCP4's refmac5.[32−35] The alignment of the resulting equilibrated structures was made using the CCP4 package (lsqkab) for making graphic representations using Pymol.[32,36]

■ RESULTS

$^1H^{15}N$ NMR Characterization of a Y836-Containing Loop. $^1H^{15}N$ NMR spectral overlays of backbone amide chemical shifts of the HNH Y836-containing loop reveal large chemical shift differences between the wt and three mutant enzymes (Figure 1, Supporting Figure S1). Furthermore, many of the other residues with large chemical shift perturbations are

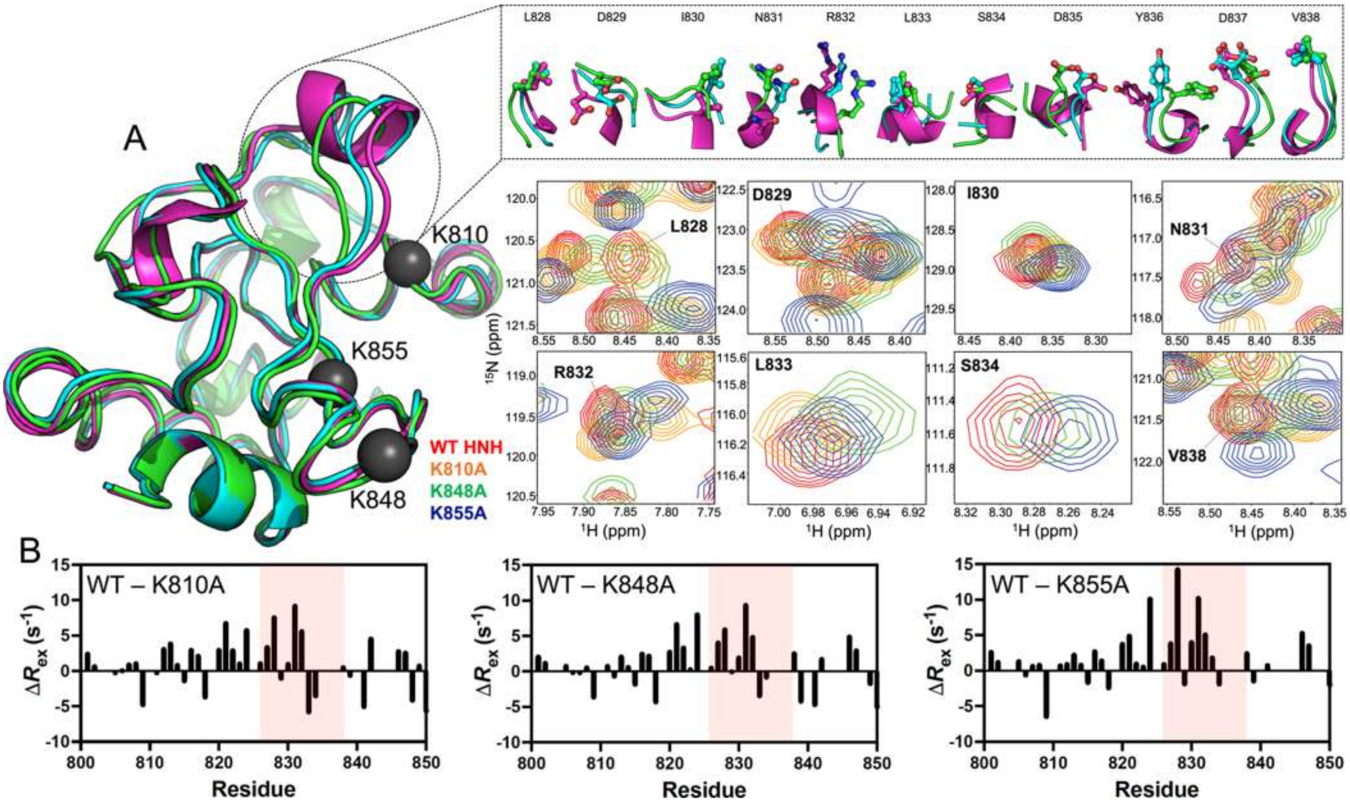


Figure 1. Conformational heterogeneity of the Y836-containing loop. (A) Three MD-derived conformers (green, cyan, and purple) of the wt HNH domain are overlaid with major structural differences localized to the loop containing residues 828−838. The three Lys-to-Ala mutation sites are highlighted by gray spheres. The various conformations of the amino acid side chains of the wt HNH domain via MD simulations are shown at top. ^{1}H^{15}N NMR spectral overlays of the wt HNH (red) domain and each of the three Lys-to-Ala mutations of K810A (orange), K848A (green), and K855A (blue) highlights widespread conformational differences in this region. Residues D835, Y836, and D837 are broadened beyond detection in spectra of wt HNH. (B) Dynamics of the region surrounding the 828−838 loop reported as ΔR_{ex} (wt-minus-mutant). Positive values indicate that wt HNH is more flexible in this region than three mutants. The loop region highlighted by NMR chemical shifts in (A) is shown in a red box. See Figure S1 for additional data.

interacting with this loop. These data suggest that there are structural changes in the Y836-containing loop and surrounding regions resulting from the specificity-enhancing mutations. The backbone resonances for D835, Y836, and D837 were so broadened in the wt enzyme that they could not be observed, suggesting that the Y836-containing loop is highly flexible. Residues for their flanking sequences (L833, S834, and V839) also have broadened reasonances, but they remained measurable (Figure 1A). Consistent with the observed line broadening, this loop exhibits positive residual R_{ex} values when comparing wt HNH to mutants via spin relaxation experiments (ΔR_{ex}, Figure 1B). This indicates that the wt enzyme is substantially more flexible in this region than the mutants. In addition to D835 and D837, there are two other carboxylates nearby, D839 and D861, which are stabilized by a Mg^{2+} ion in MD simulations immediately next to H840, which is one of the HNH catalytic residues. Our observations of structural perturbations and altered dynamic properties of the Y836-containing loop in the wt enzyme and specificity-enhancing mutants may be correlated with the different efficiencies of catalysis for DNA cleavage.

MD simulations were conducted on the full-length wt Cas9 and the K810A, K848A, and K855A variants. Upon global alignment of MD trajectories, root-mean-square-fluctuation (RMSF) was calculated for combined backbone amide N and H atoms of the wt and the three mutants. Differences of RMSF values between the wt enzyme and each of the three mutants were plotted and compared with the differences of the corresponding simulated chemical shifts derived from MD simulations using the program SHIFTX2 (Figures 2 and S2).[37] It is evident that the large change of chemical shifts (i.e., the change of local chemical environments) bears an apparent correlation with the increased RMSF differences (Figure 2). To explain the structural basis of the increased RMSF differences, three distinct conformations of the Y836-containing loop were discovered as follows (Figure 1A).

Eigenvector Centrality Analysis of the Y836-Containing Loop Region. Experimental chemical shift perturbation and change in the composite chemical shift were computed between the wt and each of the three mutants from the experimental NMR ^{1}H and ^{15}N shifts and compared with the eigenvector centrality (Figures 3 and S3). Simulated composite chemical shifts were determined as $\sqrt{[(\delta_{\mathrm{HN}}^{2}/2) + (\delta_{\mathrm{N}}^{2}/25)]/2}$. The difference in eigenvector centrality computed from Kabsch−Sander H-bond energy was also determined between the wt and each mutant.[38] The difference in centrality was calculated as the wt donor−acceptor pair minus the mutant donor−acceptor pair, that is, (centrality$_{\text{[wt/donor]}}$ − centrality$_{\text{[wt/acceptor]}}$) − (centrality$_{\text{[mutant/donor]}}$ − centrality$_{\text{[mutant/acceptor]}}$) (see the Supporting Information, SI). Donor-minus-acceptor centrality was normalized. This metric is complementary to experimental NMR data, as it provides

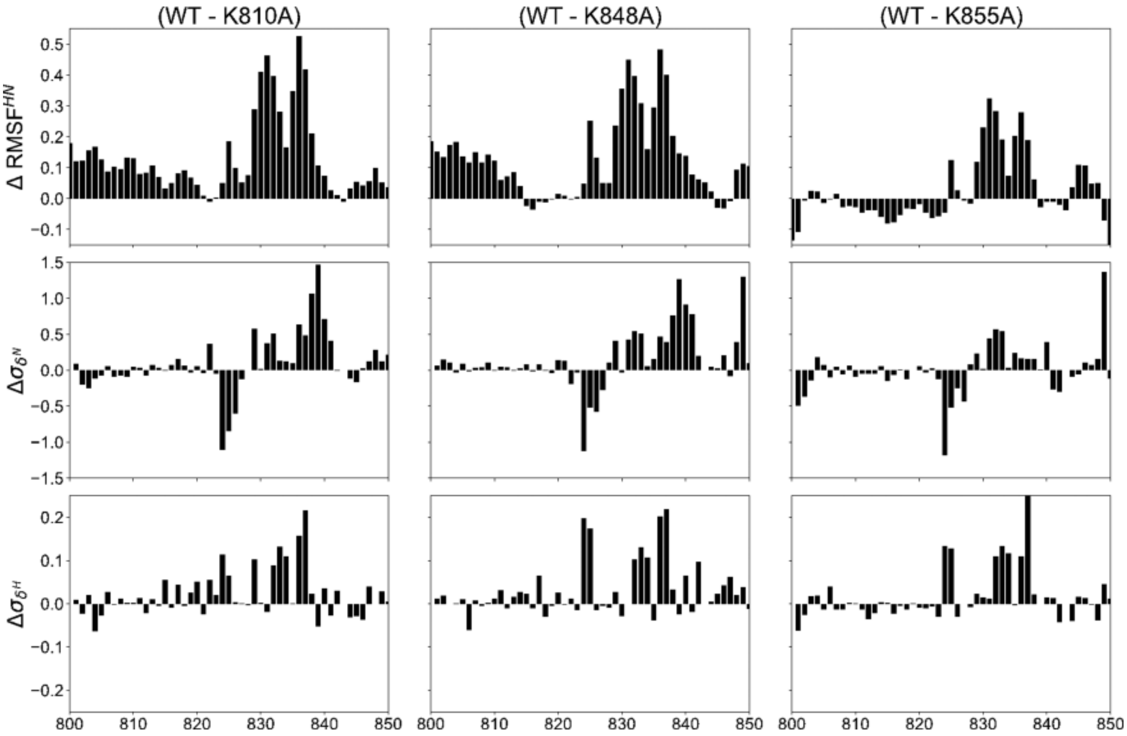


Figure 2. Apparent correlations of differences in RMSF values between the wt and each of the three mutants in the MD simulations with the simulated changes of NMR chemical shifts in the residue range of P800-D850. See Figure S2 for additional analysis.

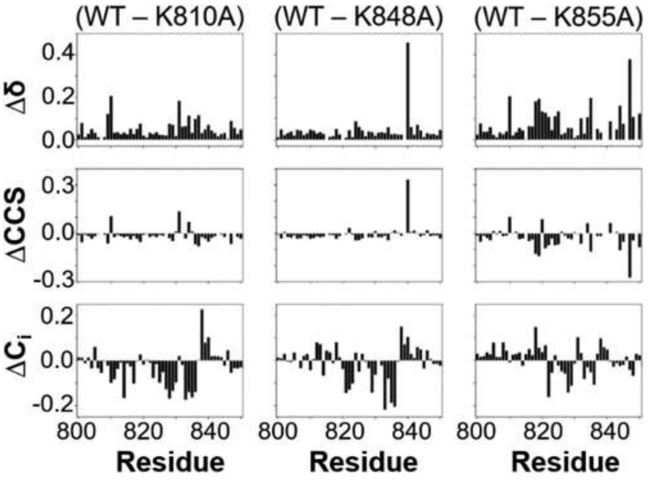


Figure 3. Relationship between the observed changes of the chemical shift and the calculated difference in eigenvector centrality between the wt HNH and mutants K810A, K848A, and K855A. NMR-derived chemical shift perturbation (Δδ), change in simulated composite chemical shift (ΔCCS), and difference in eigenvector centrality derived from Kabsch−Sander hydrogen bond energy (ΔC_i) between WT and mutants K810A, K848A, and K855A. Residues P800−D850 are shown. See Figures S2 and S3 for additional analysis.

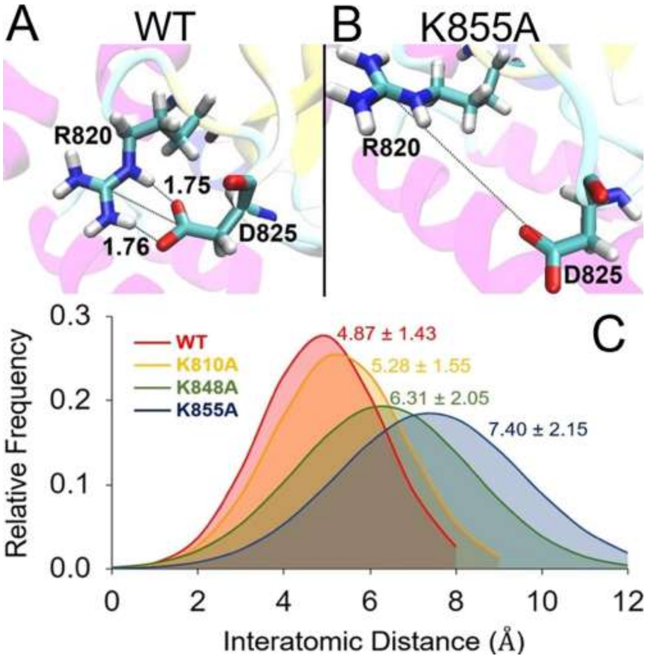


Figure 4. Close-up view of residues R820 and D825 (frame 526 of trajectory) in the HNH domain. (A) wt: two hydrogen bonds are detected (1.75 and 1.76 Å). (B) In the K855A mutant, both H-bonds disappear. (C) Relative distance probability density function as a function of interatomic R820−D825 distance for the wt (red), K810A (orange), K848A (green), and K855A (blue) mutants. See Figure S4 for additional analysis.

insight into which residues experience the greatest electrostatic shift upon mutation. The chemical shift perturbation between wt and each mutant is pronounced in residues 818−842, most prominently in the case of the K855A mutant (Figure 3). These effects can be partially explained by the breaking of intramolecular H-bonds between residues R820 and D825 (Figures 4 and S4).

The interatomic distance between R820 and D825 is shortest for the wt and longest for the K855A mutant, commensurate with the calculated chemical shift perturbation and difference in eigenvector centrality (Figure 3, Figure 4C). In the MD trajectories, we can visualize the kinetics of H-bond formation and breakage of specific pairs such as R820Nη1−

D825Oδ1 and R820Nη2−D825Oδ1, which depend on the flipping frequency of the R820 side chain, or such as R820Nη1−D825Oδ1 and R820Nη1−D825Oδ2, which depend on the flipping frequency of the D825 side chain (Figure S5). An analysis of dwell time distribution of these specific pairs in two specific states (H-bond formed and H-bond broken states, for example) can provide both the kinetics and thermodynamics of state transition,[39,40] which is a subject beyond the scope of this study.

The alterations of this region due to the three mutations can be further understood by looking at the difference in eigenvector centrality computed from the α-carbon displacements along the dynamics. A significant decrease in eigenvector centrality in this region occurring in the three mutants, most pronounced in the K855A mutant (Figure 5), strongly supports the notion that the increased interatomic R820−D825 distance exhibited by the three mutants plays a key role in disrupting the conformation dynamics of the region. The difference in eigenvector centrality between wt and each mutant clearly depicts a centrality decrease from the mutation sites to the region of residues R820−D825, providing evidence that mutation-induced perturbations are responsible for the disappearance of the R820−D825 salt bridge.

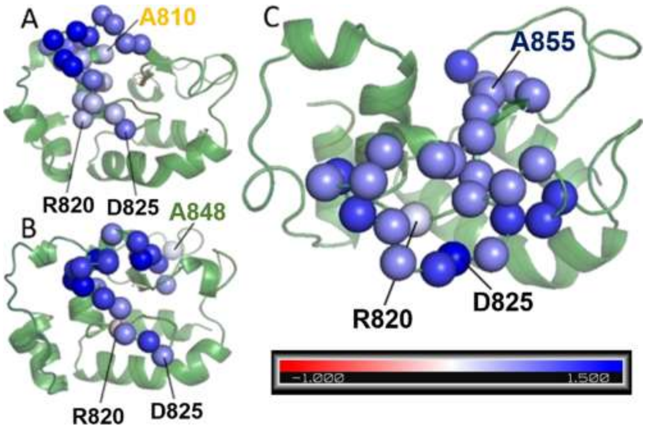


Figure 5. Difference in eigenvector centrality between the wt HNH and each of the K810A (A), K848A (B), and K855A (C) mutants with λ (locality factor) of 20 Å. Residues with differences in eigenvector centralities greater than 2 standard deviations from the mean are represented as spheres and colored according to the color bar (dominated by large values in blue). In addition to given Lys-to-Ala mutations, the locations of R820 and D825 are also indicated.

The Y836-containing loop (comprising E827−D837) adjacent to the catalytic site exhibits a significant decrease in eigenvector centrality upon each of the three mutations, providing a partial explanation for the loss of two conformations in the mutants compared to the wt. This analysis can be further strengthened by looking at the change in MD-derived ED maps between the wt and each of the three Lys-to-Ala mutants, in the regions which show the largest centrality variations.

Y836-Containing Loop Exhibits Three Distinct Conformations. The ED maps derived from all 2000 MD trajectory frames for each of the wt enzyme and the three Lys-to-Ala mutants are shown in Figure 6. The conformational heterogeneity of the loop captured by MD simulations agrees well with the structural and dynamic properties determined experimentally by NMR (Figure 1). Analysis of the resulting maps reveals a striking difference at the level of ED for the Y836-containing loop (residues E827−D837) between the wt enzyme and the three mutants. This loop is only visible at a 2σ contouring level for the wt enzyme, whereas it is still fully visible even at a 6σ contouring level for the three mutants (Figure 6A−D), which is consistent with the heightened flexibility of the loop in the wt enzyme, relative to the mutants, observed by NMR (Figure 1B). This implies that about one third of the population (or less) of the wt enzyme occupies this position and the remaining population occupies an alternate conformer(s) or transits between multiple distinct conformations. When a single-conformation model of this loop was used for standard crystallographic refinement against the MD-derived wt ED map, it revealed that there were two additional conformations of Y836. These three conformations of the Y836 side chain in the wt enzyme have mutually exclusive H-bonding patterns: (i) conformation 1, Y836Oη is H-bonded to the backbone amide D829N, (ii) conformation 2, Y836Oη is H-bonded to the backbone amide D861N, and (iii) conformation 3, the backbone amide D861N is H-bonded to R832, which replaces the Y836 side chain and forces Y836 to move into the third position.

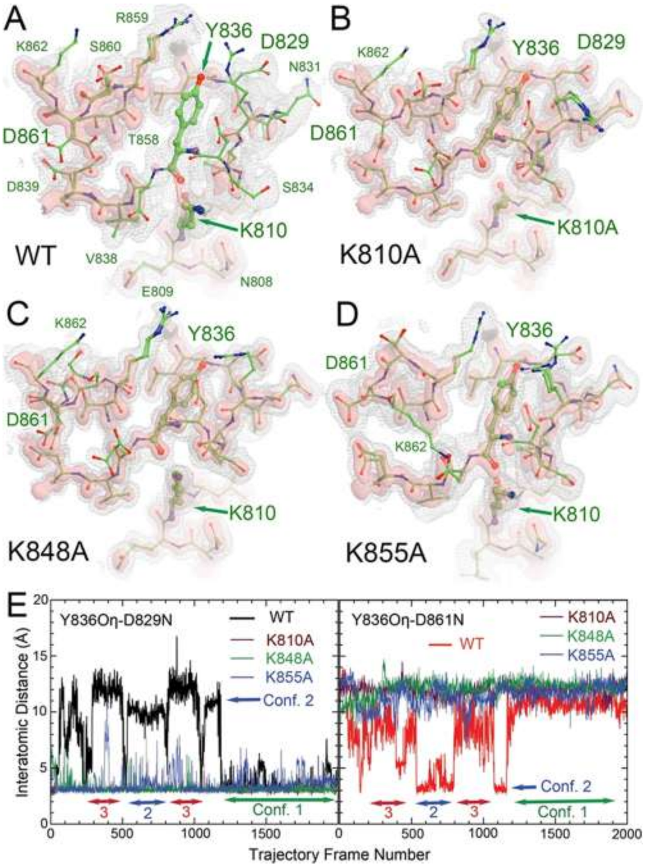


Figure 6. MD-derived ED maps for the wt HNH (A), K810A (B), K848A (C), and K855A (D) mutants at a low (2σ, silver isomesh) and high (6σ, salmon isosurface) contouring level superimposed with the corresponding MD-ESP map-derived equilibrated structures. (E) Interatomic Y836Oη−D829N distance for conformation 1 (left) as a function of MD trajectory frame number for wt (black), the K810A (maroon), K848A (green), and K855A (blue) mutants. Interatomic Y836OY−D861N distance (right) for clustering the second conformation of the wt (red) enzyme. The MD frames for conformations 1, 2, and 3 based on these distances are indicated.

The MD trajectory frames were clustered according to the three H-bonding geometries for Y836 to estimate the fraction of each conformation so that the MD-derived ED map can be calculated for each of the three individual conformations (Figures 6E and 7A–D). Each trajectory can be assigned to only one of the three conformational states (i.e., the three conformational states are mutually exclusive, 51.8%, see below) or to an intermediate transient state that has no defined equilibrated structure (48.2%). The three conformations of the Y836 side chains can be characterized by the Cα–Cβ bond torsion angle in the gauche-minus (300° or −60°) and approximately anticonfigurations (~210°) (Figure 7E,F), plus two Cβ–Cγ configurations (data not shown). The Y836 side chain of the K848A mutant has a distinct population with the Cα–Cβ bond torsion angle near the gauche-plus (~60°) configuration, which is not found in the other three mutant enzymes (Figure 7F), highlighting MD differences between them.

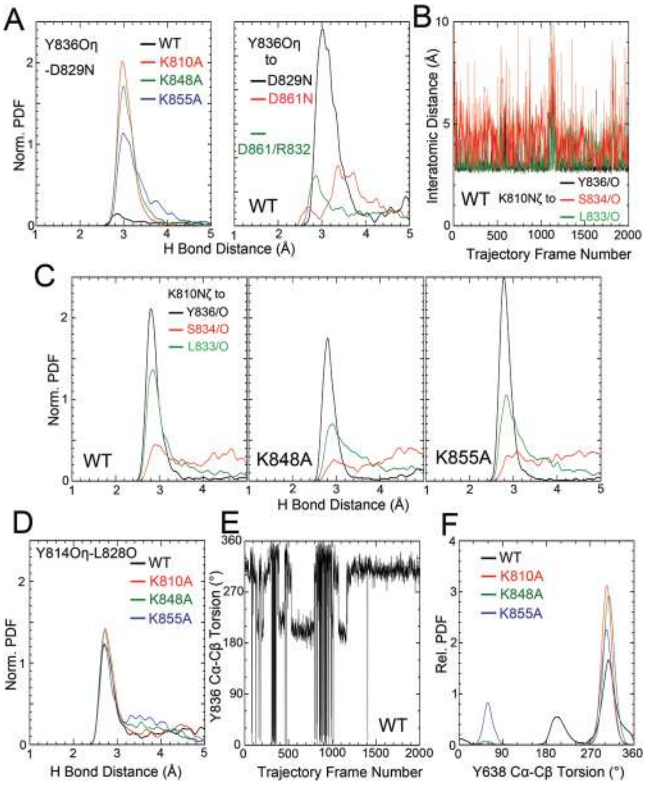


Figure 7. Population and nature of interatomic distances within clustering of each conformation. (A) Relative PDF of the Y836Oη–D829N distance for the wt enzyme (black), K810A (red), K848A (green), and K855A (blue) mutants and a close-up view of PDFs for the three conformations of the WT enzyme as defined by three interatomic distances (Y836Oη–D829N for conformation 1, Y836Oη–D861N for conformation 2, and R832–D861N for conformation 3). (B) Interatomic distance of K810Nζ to Y836O (black), S834O (red), and L833O (green) as a function of MD trajectory frame. (C) Corresponding relative PDF of K810 in the wt enzyme and in the K848A and K855A mutants. (D) Relative PDF of the tertiary interaction of the Y814Oη–L828O distance. (E) Cα–Cβ torsion angle of Y836 in the wt enzyme as a function of MD trajectory frame. (F) Relative PDF of the Cα–Cβ bond dihedral (or torsion) angle of the wt enzyme (black), and K810A (red), K848A (green), and K855A (blue) mutants.

The probability density function (PDF) of the Y836Oη–D829N distance distribution is very sharp for the K810A mutant, but it becomes slightly broader for the K848A and K855A mutants, corresponding to a single conformation (Figure 7A). However, this distribution is very broad for the wt enzyme (Figure 7A). The distribution of the Y836Oη–D861N and R832–D861N distances reveals that two other distinct conformations exist for the wt enzyme (Figure 7A).

Because three conformations of the wt HNH domain are mutually exclusive, that is, the Y836 side chain can adopt only one of the three conformations in any MD frame; the fractions should be correlated with corresponding fractions of specific H-bonds described above (Figure 7). Additionally, other residues are also involved in H-bonding interactions within this loop, including two interloop tertiary contacts and one intraloop backbone H-bond (Figure 7). One interloop tertiary H-bond involves K810Nζ with Y836O, S834O, and/or L833O in multiple H-bonding interactions (Figure 7B,C). The second tertiary O bond is between Y814Oη and L822O (Figure 7D). These tertiary interactions are maintained to some extent and are independent of the three conformations of Y836. Therefore, the local conformational change of this loop has been propagated to its surrounding environment. The distribution of the intraloop backbone H-bond between Y836N and R832O is similar among the wt enzyme and the K810A and K855A mutants (see below), highlighting this loop moving largely as a single rigid body regardless of the orientations of the Y836 and R832 side chains.

Characterization of Three Equilibrated Structures of the Y836 Loop. When all MD trajectory frames with the Y836Oη–D829N interatomic distance <3.30 Å of the first conformation were clustered together for the calculation of the MD-ED map, this cluster represented 28.6% of the population (Figure 8A). The ED level for Y836 and the Y836-containing loop in such clustering of MD frames was restored to the full level of other atoms (Figure 8A), and the equilibrated structure for this conformation can be accurately determined. The cluster of the second conformation having the Y836Oη–D861N interatomic distance <3.30 Å represented 9.1% of the population (Figure 8B). The cluster of the third conformation having the D861N–R832Nη1 or R832Nη2 distances <3.30 Å represented 14.1% of the population (Figure 8C).

The MD-ED maps for the three conformations were generated upon alignment of all main chain atoms, revealing that the largest coordinate differences are located at the side chain and main chain of Y836 (Figure 8). From the first to second conformations, the Y836Oη atom is displaced by 8.6 Å and the Y836Cα atom is by 2.8 Å (Figure 8D,E). From the first to third conformations, the corresponding displacements are by 11.2 and 4.0 Å, respectively. From the second to the third conformations, they are 7.0 and 1.5 Å, respectively. Depending on the Y836 side chain rotamers, the Y836-containing loop largely undergoes subtle rotation motions with the rotation axis passing through the Y815 side chain so that the loops surrounding Y815 have the same rotational motions (Figure 8F).

Nature of the Y836-Containing Loop Movement. The orientation of the Y836 side chain is well defined and can be readily characterized in the MD trajectories. However, the nature of main chain movement of the Y836-containing loop is more difficult to analyze because there is no clear-cut boundary of the Y836-containing loop. Strong backbone interactions often retain any large deformation of the backbone structure at

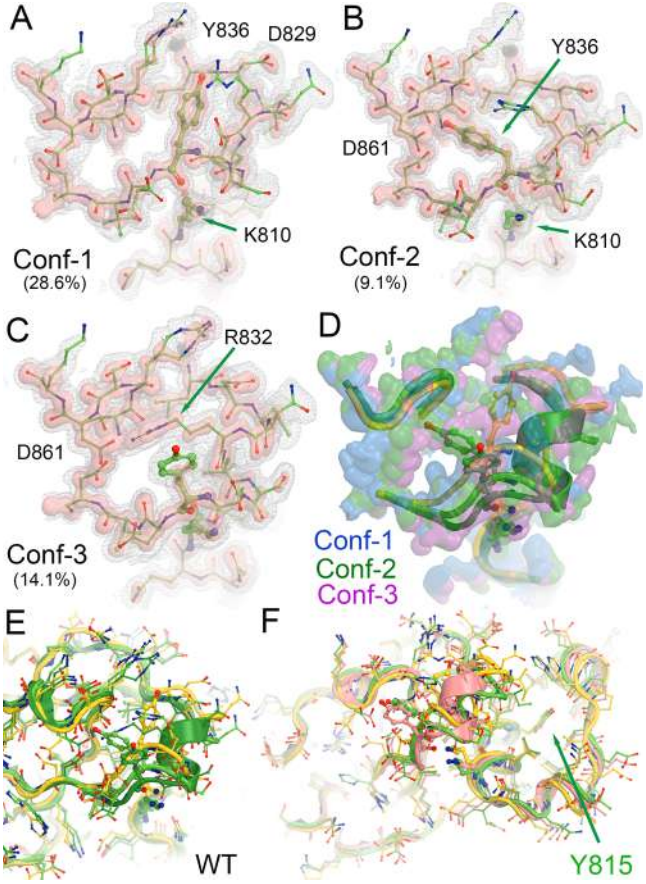


Figure 8. Decomposition of the three conformations of the wt enzyme through clustering analysis. (A) Conformation 1 (28.6%) at a low (2σ, silver isomesh) and high (6σ, salmon isosurface) contouring level superimposed with the corresponding equilibrated structure. (B) Conformation 2 (9.1%). (C) Conformation 3 (14.1%). (D) Superimposition of three conformations as unit population. (E,F) Two views of superposition of equilibrated structures corresponding to the three conformations.

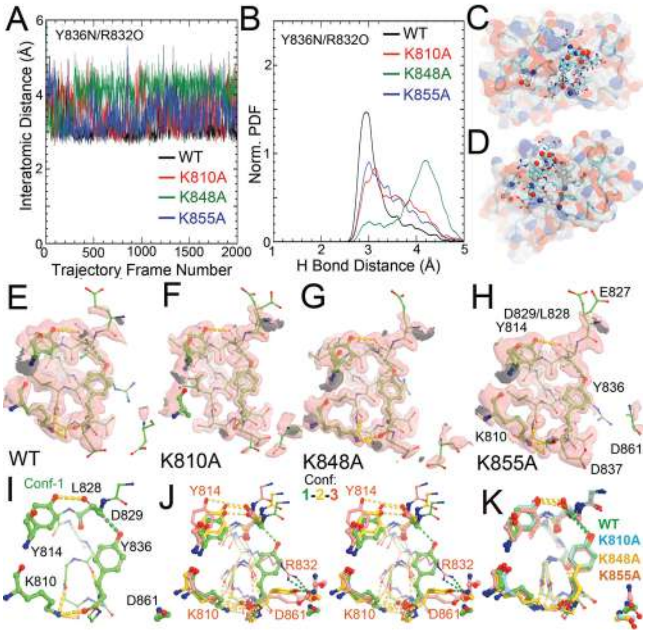


Figure 9. Structures of the Y836-containing loop in the wt enzyme and the three mutants. (A) Interatomic intraloop Y836N−R832O distance as a function of MD trajectory frame for the wt enzyme (black), the K810A (red), K848A (green), and K855A (blue) mutants. (B) Corresponding relative PDFs. (C,D) Two views of the location of the Y836-containing loop within the HNH domain. (E–H) Y836 loop-aligned MD-ED maps contoured at +6σs (isosurface) for the wt enzyme (E), the K810A (F), K848A (G), and K855A (H) mutants. (I) Two interloop H-bonds (Y814Oη/L828O, K810Nζ/Y836 and S834 and L833) and one intra-loop H-bond (Y836N/R832O). (J) Stereodiagram of the Y836 loop in three conformations (1, green, 2, gold, and 3 in salmon). (K) Comparison of the conformation 1 of the wt enzyme (green), and the K810A (blue), K848A (gold), and K855A (salmon) mutants.

any single residue. Instead, the large backbone displacement of one residue is often surrounded by smaller displacements in its connected residues. In the alignment of main chain atoms of the entire domain, the summed population from the three distinct conformations is 51.8%, meaning that the analysis has not included the remaining 48.2% of trajectory frames.

To address the structure of the Y836-containing loop in the entire population, the MD maps were calculated after the alignment of main chain atoms of this loop only (Figure 9). This alignment shows that the loop has a single dominant equilibrated structure in the wt enzyme and in each of the three mutants. The ED distributions of MD-ED maps for the K810A and K848A mutants are sharper than those of the wt enzyme and the K855A mutant, and that of the K855A mutant is sharper than that of the wt enzyme. This observation suggests that the large side chain conformational change of this loop in the wt enzyme is associated with a larger motion of its main chain relative to the three mutants, whereas the loop mainly undergoes a single-body rigid motion. Within the loop, the intraloop backbone H-bond between Y836N and R832O has a larger fluctuation in the trajectories of the wt enzyme than other residues, more so than those of the three mutants (Figure 7A). This H-bond is largely absent from the K848A mutant (Figure 7A). Interloop tertiary H-bonds involving K810 and Y814 remain relatively unchanged among the wt enzyme and the three mutants (Figure 9I,J). In this alignment, the density for the remaining structure is completely smeared out, which is consistent with the notion that this loop undergoes rotational motions relative to the entire domain.

Relationship of the Y836 Dynamics to the Endonuclease Catalytic Site. We observed that each of the K810A, K848A, and K855A mutations has significantly altered dynamics of Y836 and the Y836-containing loop (Figures 1 and S1), which are located next to the catalytically important residues D839 and H840, identified by site-directed mutagenesis.[6] The 6o0y structure represents the catalytically competent state in which the target DNA has a nick located at the catalytic site (Figure 10).[6] D839 and H840 are directly connected to the Y836-containing loop by two residues (and thus they can be considered as an extended part of the Y836-containing loop). Residues H840 and N863 are part of the His-Asn-His-(HNH) motif.[6] The active site should also include both K866 (which may act as a second-shell residue stabilizing the catalytic site) and R864 residues (Figure 10E,F). K855 is the only one of the mutated Lys residues near the active site, while K810 and K848 are more distant. Likely because of its proximity to the catalytic site, the K855A mutant exhibits the highest substrate specificity.[5] Because of the limited resolution of cryo-EM map at ∼3.37 Å for the 6o0y model,[6] many of the important interactions identified in our study could not be fully validated experimentally, including H-

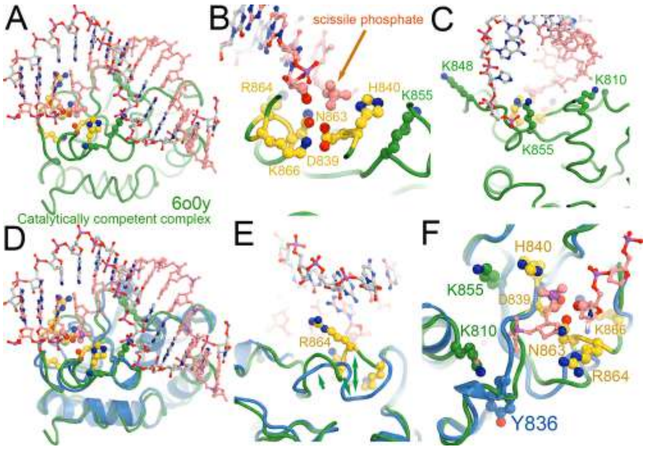


Figure 10. Relationship of the Y836-containing loop with the two catalytically important D839 and H840 residues and with the locations of three mutations. (A−C) Three views of the catalytically competent conformation derived from 6o0y coordinates. The scissile phosphate group (salmon arrow), the catalytically important residues D839 and H840, and three other important residues N863, K866, and R864 are shown in large spheres (gold and salmon), as are the three mutants (green). (D−F) Three views of superposition of conformation 1 of the wt enzyme (sky-blue cartoon) with the 6o0y structures. The R864-containing loop undergoes large conformational changes as indicated by green arrows (E).

bonds of main chain conformations, which would limit our ability to correlate dynamics of this enzyme from our study with experimental data.

Least-squares alignment of the first conformation of the wt enzyme with the catalytically competent 6o0y HNH domain shows that the main chains of H840 and R864 are displaced by 1.2 and 5.8 Å, respectively (Figure 10).[6] In the 6o0y structure,[6] the Y836Oη−D829N distance is ∼4.2 Å; thus it is not H-bonded. In the three mutants, as well as in the first conformation of the wt enzyme, the Y836Oη and D829N maintain a strong H-bond. In the 4un3 crystal structure[24] of the catalytically inactive complex from which our MD simulations started, the catalytic H840 side chain was not built because of weak ED while the R864 side chain was partially buried, involving five H-bonds to its guanidium group, including bidentate H-bonds between D839 and R864. R864 in the 4un3 structure is completely inaccessible for the substrate of the target DNA strand to interact, which is required for the HNH endonuclease activity as observed in the 6o0y complex.[6,24] If breaking the H-bond between Y836Oη and D829N is an essential step for activation of this enzyme, these three mutations would enhance this H-bond and may thus prevent promiscuous activation of the HNH activity by stabilizing their inactive state until the cognate substrate is properly positioned. We expect that results of MD studies could help to define catalytically important residues more accurately in cryo-EM structure determination, many of which may be addressed computationally using enhanced sampling to capture the inactive-to-active conformational transition of the HNH domain.[41]

■ DISCUSSION

Precision gene editing tools differ from innate immunity in substrate specificity. Innate immunity surveillance, rapid detection, and hydrolysis of closely related invading DNA sequences, which may be divergent due to continuous evolution, are more important than cutting a single sequence. As a precision gene editing tool, substrate specificity for cutting only one single sequence is most critical. Thus, understanding the structural basis of existing enzyme variants for improved substrate specificity could help to rationally design better mutants with an even higher degree of substrate specificity. Here, the structural and dynamical basis for improved substrate specificity of the three Lys-to-Ala mutants in CRISPR-Cas9 has been examined. Our investigations reveal that the Y836-containing loop displays remarkable flexibility in the wt Cas9 enzyme, accessing three distinct conformations. However, each of the three Lys-to-Ala mutations reduces dynamics of this loop to only a single conformation. The structural basis for enhanced specificity of the three mutants suggests that they share the same signaling pathway through the Y836 loop, which may explain why multiple combinations of these mutations do not exhibit additive effects.[5] This result agrees with our recent observations that these Lys-to-Ala mutations strongly perturb an allosteric signaling pathway.[12,13]

Traditionally, allosteric phenomena in proteins and nucleic acids are primarily analyzed through network theory, which offers an excellent tool kit to decrypt the signal transmission,[42] and whose power is magnified through the combination of solution NMR.[43] MD trajectories are also primarily analyzed using backbone atoms only, which would not easily uncover the dynamics of the Y836 side chain and the Y836 loop as described above. The present analysis offers a new view of conformational networking in proteins, primarily through the side chain rotamer configurations. For Y836 to change from the first to the second conformations, the R832 side chain would have to transiently move away from the current position because R832 is located midway between the orientations of Y836 and completely blocks this transition. For Y836 to change from the third to the second conformations, the R832 side chain would have to break its H-bond to D861N first. Each of these transitions and corresponding energy barriers are not dependent on the backbone coordinates of residues and thus are often invisible in the backbone (or Cα)-based conformational analysis.

In DNA polymerases, which are the most extensively studied enzymes, the base selectivity of substrates and the catalytic efficiency often bear a positive correlation.[44] The most efficient replicative DNA polymerases often exhibit the highest base selectivity because only the correct substrate can achieve the highest stabilization effect to the transition state of the polymerization reaction.[45] The latter exploits a metal dependent mechanism to chemically process nucleic acids, similar to what observed in the Cas9 enzyme.[15,46] Considering this similarity, an equivalent correlation may also exist for Cas9. In either case, allosteric regulation of dynamical properties could play an essential role in the correlation of substrate specificity and catalytic efficiency of both classes of enzymes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.2c00127.

One computational procedure section for centrality and five supporting figures (PDF)

AUTHOR INFORMATION

Corresponding Authors

Jimin Wang − *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114, United States;* orcid.org/0000-0002-4504-8038; Email: jimin.wang@yale.edu

George P. Lisi − *Department of Molecular and Cell Biology and Biochemistry, Brown University, Providence, Rhode Island 02912, United States;* orcid.org/0000-0001-8878-5655; Email: george_lisi@brown.edu

Giulia Palermo − *Department of Bioengineering and Department of Chemistry, University of California Riverside, Riverside, California 92521-9800, United States;* orcid.org/0000-0003-1404-8737; Email: gpalermo@engr.ucr.edu

Victor S. Batista − *Department of Chemistry, Yale University, New Haven, Connecticut 06511-8499, United States;* orcid.org/0000-0002-3262-1237; Email: victor.batista@yale.edu

Authors

Erin Skeens − *Department of Molecular and Cell Biology and Biochemistry, Brown University, Providence, Rhode Island 02912, United States*

Pablo R. Arantes − *Department of Bioengineering and Department of Chemistry, University of California Riverside, Riverside, California 92521-9800, United States;* orcid.org/0000-0003-1946-2750

Federica Maschietto − *Department of Chemistry, Yale University, New Haven, Connecticut 06511-8499, United States;* orcid.org/0000-0002-5995-2765

Brandon Allen − *Department of Chemistry, Yale University, New Haven, Connecticut 06511-8499, United States;* orcid.org/0000-0002-5512-1892

Gregory W. Kyro − *Department of Chemistry, Yale University, New Haven, Connecticut 06511-8499, United States;* orcid.org/0000-0002-0095-8548

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.2c00127

Author Contributions

J.W., G.P.L., G.P., and V.S.B. designed the experiments. E.S. and G.P.L. carried out NMR studies. P.A. and G.P. carried out MD simulations. A.B., F.M., and G.K. carried out correlation analysis of MD simulations with NMR. J.W. carried out electron-density analysis of MD trajectories and wrote the draft manuscript with input of all coauthors. All authors were involved in the interpretation of results and writing of the final version of manuscript.

Funding

This material is based upon work supported by the National Institute of Health under Grant No. R01GM136815 (awarded to VSB, GP, and GPL) and Grant No. R01GM141329 (awarded to GP). This work was also funded by the National Science Foundation under Grant No. CHE-1905374 (awarded to GP) and under Grant No. MCB-2143760 (awarded to GPL). Computer time for MD has been awarded by XSEDE under Grant No. TG-MCB160059 and by NERSC under Grant No. M3807 (to GP).

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

CRISPR: clustered regularly interspaced short palindromic repeat; Cas: CRISPR-associated system; HNH: histidine-asparagine-histidine motif endonuclease; ESP: electrostatic potential; ED: electron density; MD: molecular dynamics

REFERENCES

(1) Doudna, J. A.; Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, No. 1258096.

(2) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014**, *343*, No. 1247997.

(3) Fu, Y.; Foden, J. A.; Khayter, C.; Maeder, M. L.; Reyon, D.; Joung, J. K.; Sander, J. D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **2013**, *31*, 822−826.

(4) Jiang, F.; Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017**, *46*, 505−529.

(5) Slaymaker, I. M.; Gao, L.; Zetsche, B.; Scott, D. A.; Yan, W. X.; Zhang, F. Rationally engineered Cas9 nucleases with improved specificity. *Science* **2016**, *351*, 84−88.

(6) Zhu, X.; Clarke, R.; Puppala, A. K.; Chittori, S.; Merk, A.; Merrill, B. J.; Simonovic, M.; Subramaniam, S. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **2019**, *26*, 679−685.

(7) Longenecker, K. L.; Garrard, S. M.; Sheffield, P. J.; Derewenda, Z. S. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 679−688.

(8) Yokota, K.; Satou, K.; Ohki, S. Comparative analysis of protein thermo stability: Differences in amino acid content and substitution at the surfaces and in the core regions of thermophilic and mesophilic proteins. *Sci. Technol. Adv. Mater.* **2006**, *7*, 255−262.

(9) Sokalingam, S.; Raghunathan, G.; Soundrarajan, N.; Lee, S. G. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One* **2012**, *7*, No. e40410.

(10) Osire, T.; Yang, T.; Xu, M.; Zhang, X.; Li, X.; Niyomukiza, S.; Rao, Z. Lys-Arg mutation improved the thermostability of Bacillus cereus neutral protease through increased residue interactions. *World J. Microbiol. Biotechnol.* **2019**, *35*, 173.

(11) Czepas, J.; Devedjiev, Y.; Krowarsch, D.; Derewenda, U.; Otlewski, J.; Derewenda, Z. S. The impact of Lys−>Arg surface mutations on the crystallization of the globular domain of RhoGDI. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 275−280.

(12) Nierzwicki, L.; East, K. W.; Morzan, U. N.; Arantes, P. R.; Batista, V. S.; Lisi, G. P.; Palermo, G. Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *Elife* **2021**, *10*, No. e73601.

(13) Nierzwicki, L.; Arantes, P. R.; Saha, A.; Palermo, G. Establishing the allosteric mechanism in CRISPR-Cas9. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, No. e1503.

(14) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142*, 1348−1358.

(15) Casalino, L.; Nierzwicki, L.; Jinek, M.; Palermo, G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal.* **2020**, *10*, 13596−13605.

(16) Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **2018**, *22*, 359−371.

(17) Sun, W.; Yang, J.; Cheng, Z.; Amrani, N.; Liu, C.; Wang, K.; Ibraheim, R.; Edraki, A.; Huang, X.; Wang, M.; Wang, J.; Liu, L.; Sheng, G.; Yang, Y.; Lou, J.; Sontheimer, E. J.; Wang, Y. Structures of Neisseria meningitidis Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. *Mol. Cell* **2019**, *76*, 938.e5−952.e5.

(18) Tang, H.; Yuan, H.; Du, W.; Li, G.; Xue, D.; Huang, Q. Active-Site Models of Streptococcus pyogenes Cas9 in DNA Cleavage State. *Front. Mol. Biosci.* **2021**, *8*, No. 653262.

(19) Huai, C.; Li, G.; Yao, R.; Zhang, Y.; Cao, M.; Kong, L.; Jia, C.; Yuan, H.; Chen, H.; Lu, D.; Huang, Q. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat. Commun.* **2017**, *8*, 1375.

(20) Zuo, Z.; Zolekar, A.; Babu, K.; Lin, V. J.; Hayatshahi, H. S.; Rajan, R.; Wang, Y. C.; Liu, J. Structural and functional insights into the bona fide catalytic state of Streptococcus pyogenes Cas9 HNH nuclease domain. *Elife* **2019**, *8*, No. e46500.

(21) Zuo, Z.; Liu, J. Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Sci. Rep.* **2017**, *7*, 17271.

(22) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277−293.

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

(24) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **2014**, *513*, 569−573.

(25) Yu, L.; Li, D. W.; Bruschweiler, R. Balanced Amino-Acid-Specific Molecular Dynamics Force Field for the Realistic Simulation of Both Folded and Disordered Proteins. *J. Chem. Theory Comput.* **2020**, *16*, 1311−1318.

(26) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(27) 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.

(28) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(29) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139*, 16028−16031.

(30) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(31) Wang, J.; Shi, Y.; Reiss, K.; Allen, B.; Maschietto, F.; Lolis, E.; Konigsberg, W. H.; Lisi, G. P.; Batista, V. S. Insights into Binding of Single-Stranded Viral RNA Template to the Replication-Transcription Complex of SARS-CoV-2 for the Priming Reaction from Molecular Dynamics Simulations. *Biochemistry* **2022**, *61*, 424.

(32) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2011**, *67*, 235−242.

(33) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **1997**, *53*, 240−255.

(34) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2010**, *66*, 213−221.

(35) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.

(36) Delano, W. L. *Pymol*; Schrodinger, Inc.. http://pymol.org/.

(37) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **2011**, *50*, 43−57.

(38) Maschietto, F.; Zavala, E.; Allen, B.; Loria, P. J.; Batista, V. S. MptpA kinetics enhanced by allosteric control of an active conformation. *J. Mol. Biol.* **2022** (accepted for publication).

(39) Min, W.; English, B. P.; Luo, G.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. Fluctuating enzymes: lessons from single-molecule studies. *Acc. Chem. Res.* **2005**, *38*, 923−931.

(40) Xie, X. S. Biochemistry. Enzyme kinetics, past and present. *Science* **2013**, *342*, 1457−1459.

(41) Igaev, M.; Kutzner, C.; Bock, L. V.; Vaiana, A. C.; Grubmuller, H. Automated cryo-EM structure refinement using correlation-driven molecular dynamics. *Elife* **2019**, *8*, No. e43542.

(42) Arantes, P. R.; Patel, A. C.; Palermo, G. Emerging methods and applications to decrypt allostery in proteins and nucleic acids. *J. Mol. Biol.* **2022**, No. 167518.

(43) East, K. W.; Skeens, E.; Cui, J. Y.; Belato, H. B.; Mitchell, B.; Hsu, R.; Batista, V. S.; Palermo, G.; Lisi, G. P. NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys. Rev.* **2020**, *12*, 155−174.

(44) Xia, S.; Konigsberg, W. H. RB69 DNA polymerase structure, kinetics, and fidelity. *Biochemistry* **2014**, *53*, 2752−2767.

(45) Wang, J.; Konigsberg, W. H. Two-metal-ion catalysis: Inhibition of DNA polymerase activity by a third divalent metal ion, *Fronit. Mol. Biosci.* **2022**, in press. DOI: 10.3389/fmolb.2022.824794/abstract.

(46) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59*, 2394−2406.