UC Riverside
UC Riverside Previously Published Works

Title

Substrate-independent activation pathways of the CRISPR-Cas9 HNH nuclease.

Permalink

https://escholarship.org/uc/item/785050nr

Journal

Biophysical Journal, 122(24)

Authors

Wang, Jimin

Maschietto, Federica

Qiu, Tianyin

et al.

Publication Date

2023-12-19

DOI

10.1016/j.bpj.2023.11.005

Peer reviewed

Substrate-independent activation pathways of the CRISPR-Cas9 HNH nuclease

Jimin Wang,[1,*] Federica Maschietto,[2] Tianyin Qiu,[2] Pablo R. Arantes,[3] Erin Skeens,[4] Giulia Palermo,[3,4,*] George P. Lisi,[5,*] and Victor S. Batista[2,*]
[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut; [2]Department of Chemistry, Yale University, New Haven, Connecticut; [3]Department of Bioengineering, University of California, Riverside, Riverside, California; [4]Department of Chemistry, University of California, Riverside, Riverside, California; and [5]Department of Molecular and Cell Biology and Biochemistry, Brown University, Providence, Rhode Island

ABSTRACT A hallmark of tightly regulated high-fidelity enzymes is that they become activated only after encountering cognate substrates, often by an induced-fit mechanism rather than conformational selection. Upon analysis of molecular dynamics trajectories, we recently discovered that the Cas9 HNH domain exists in three conformations: 1) Y836 (which is two residues away from the catalytic D839 and H840 residues) is hydrogen bonded to the D829 backbone amide, 2) Y836 is hydrogen bonded to the backbone amide of D861 (which is one residue away from the third catalytic residue N863), and 3) Y836 is not hydrogen bonded to either residue. Each of the three conformers differs from the active state of HNH. The conversion between the inactive and active states involves a local unfolding-refolding process that displaces the Cα and side chain of the catalytic N863 residue by ~5 Å and ~10 Å, respectively. In this study, we report the two largest principal components of coordinate variance of the HNH domain throughout molecular dynamics trajectories to establish the interconversion pathways of these conformations. We show that conformation 2 is an obligate step between conformations 1 and 3, which are not directly interconvertible without conformation 2. The loss of hydrogen bonding of the Y836 side chain in conformation 3 likely plays an essential role in activation during local unfolding-refolding of an α-helix containing the catalytic N863. Three single Lys-to-Ala mutants appear to eliminate this substrate-independent activation pathway of the wild-type HNH nuclease, thereby enhancing the fidelity of HNH cleavage.

SIGNIFICANCE CRISPR-Cas9 has two endonucleases, RuvC (homology domain of UV-sensitive gene product C activity for resolving Holliday junction) and HNH (His-Asn-His endonuclease), which catalyze the double-strand cleavage of foreign DNA duplexes as a form of innate immunity. RuvC is responsible for cleavage of the single-stranded "nontarget" strand of the DNA duplex, whereas the HNH domain is reoriented and repositioned through twisting and swiveling motions for cleavage of the target strand in heteroduplex with guide RNA. Subsequent to these critical domain motions, the wild-type HNH nuclease is then activated through both substrate-dependent and substrate-independent mechanisms. This study applies molecular dynamics simulations to explore a molecular basis for (substrate-independent) activation of the enzyme with potential to inform enzyme redesign for enhanced substrate specificity.

INTRODUCTION

The HNH nuclease of Cas9 is a self-contained catalytic domain for cleavage of target DNA (tDNA) that is base-paired with guide RNA (gRNA) in a heteroduplex (1,2). In contrast, the Cas9 RuvC nuclease is a divided catalytic domain for cleavage of nontarget DNA (ntDNA) that is single-stranded (ss) after forming an R-loop as gRNA slowly displaces this strand of invading foreign DNA duplexes. Two pieces of the RuvC sequence are connected to the HNH domain through two short linkers, which play important roles in selection of the ssDNA substrate for activation of RuvC. The catalytic sites of both nucleases are completely inaccessible to either ssDNA or heteroduplex substrates in the apo form of Cas9 to ensure that other unrelated ssDNA and heteroduplexes of nonsubstrates are not accidentally cut by the enzyme. Many nonsubstrate DNA molecules are present inside cells, including replication intermediates during DNA replication and recombination (3), particularly during lagging-strand DNA synthesis, such as RNA primers synthesized in Okazaki fragments.

Heteroduplexes are also part(s) of transcriptional intermediates when replicational and transcriptional machineries collide (4–6). Highly regulated HNH and RuvC endonuclease activities are therefore required so as not to interfere with the essential cellular processes of replication and transcription while surveilling for invading DNA duplexes. As a part of innate immunity, Cas9 has evolved some promiscuity in its substrate specificity so that it can defend against the invasion of the original and any evolved pathogens containing mutations. However, as a gene-editing tool, the enzyme must be reengineered for extremely high substrate specificity (i.e., to cut only one designed sequence). Understanding the molecular basis for substrate specificity of Cas9 is key to such reengineering.

Cas9 recognizes and binds the protospacer adjacent motif (PAM) basepairs without unpairing them (1,2). Upon this binding, the DNA duplex bends ∼90° so that its downstream ntDNA strand can be unwound to initiate tDNA basepairing with gRNA (7). A twisting rotation of 102° of the HNH domain acts as a helicase that continuously unwinds the downstream DNA duplex for strand displacement by gRNA (7). This twisting motion changes the topology of linkers that bind tDNA relative to the two catalytic sites, together with formation of an R-loop in a rate-determining step (8,9). Correct basepairing between tDNA and gRNA provides free energy to cooperatively assist in the twisting rotation of the HNH domain. When the complete tDNA/gRNA basepairs are formed, the displaced strand of ntDNA substrate is then buried inside the RuvC active site between the RuvC and HNH domains. The RuvC active site is fully formed and is then activated via repositioned linkers relative to the bound ntDNA strand to initiate cleavage. After the ntDNA cleavage, the HNH domain is displaced toward the minor groove of the nearby tDNA/gDNA duplex through attractive, long-range electrostatic interactions. HNH then undergoes a 43° swiveling motion, which irreversibly releases the cleaved ntDNA and disassembles the RuvC active site (7). This motion is responsible for repositioning the HNH catalytic site near the scissile phosphate group of the tDNA strand of the heteroduplex for tDNA cleavage.

Activation and inactivation of each endonuclease for the double-stranded break are highly regulated. When the basepairs between tDNA and gRNA are incomplete, the full twisting motion of HNH is not achieved, and the RuvC active site is not fully assembled (10,11). As a result, the DNA duplex is then rejected without cleavage. Therefore, the presence of the R-loop is essential for the irreversible process of releasing the cleaved ntDNA coupled with the swiveling of HNH toward the bound heteroduplex for tDNA cleavage. Likewise, unrelated heteroduplexes may be accidentally captured by the HNH domain but are often incapable of R-loop formation. Without the R-loop, the scissile position of an accidentally captured DNA strand in the heteroduplex cannot be defined relative to a PAM sequence.

Therefore, HNH would likely remain in an inactive state, which is a subject of this study. Here, we examine one of two possible activation pathways for HNH activity according to analysis of the isolated HNH domain, i.e., a heteroduplex substrate-independent pathway. If HNH has already twisted to expose its catalytic site via substrate-independent activation pathways, it could promiscuously cut any DNA strand near its active site even without the required R-loop formation. Analysis of molecular dynamics (MD) simulations, a well-established method for studies of CRISPR-Cas9 (12), describes this pathway at the molecular level in the wild-type (WT) enzyme and three specificity-enhancing Lys-to-Ala mutants (13). By identifying substrate-independent activation pathways and eliminating them via site-directed mutagenesis, one can achieve a higher degree of substrate selectivity for enzyme engineering.

MATERIALS AND METHODS

Molecular dynamics simulations of the HNH domain were based on our recently published sets of MD trajectories, which included the WT enzyme and three Lys-to-Ala mutants (14–16). In this study, we focus on results of the PCA analysis for new sets of HNH MD simulations, which are almost identical to the published results. This study focuses on underlying conformational landscapes and properties in those MD trajectories.

MD simulations

MD simulations were carried out starting with the high-resolution x-ray structure of the ternary Cas9:RNA:DNA complex (PDB ID: 4UN3) determined at a resolution of 2.59 Å (17). Three single-point mutations, K810A, K848A, and K855A, were introduced into the complex. For the analysis of all titratable residues, we utilized PROPKA3 method (18). The Amber ff12SB force field was used, which included ff99bsc0 corrections for DNA and ff99bsc0+χOL3 corrections for RNA (19–21). The Allnér force field was used for Mg^{2+} ions, and the TIP3P model was employed for water molecules (22,23). These force field parameters and simulation protocols have previously been used in MD studies of CRISPR-Cas9 (24,25). A time step of 2 fs was used for the simulations, and the SHAKE algorithm was employed to constrain all bond lengths involving H atoms. Temperature control was achieved through Langevin dynamics at a temperature of 300 K with a collision frequency (γ) set to 1/ps. Pressure control was accomplished by coupling the system to a Berendsen barostat at a reference pressure of 1 atm, with a relaxation time of 2 ps (26). To relax water molecules and counterions, the systems underwent energy minimization while keeping the protein, RNA, DNA, and Mg^{2+} ions fixed in position, starting with harmonic restraints of 300 kcal/mol·$Å^2$. Subsequently, the systems were heated in two distinct steps in the canonical ensemble (NVT): first from 0 to 50 K and then from 50 to 100K, each using simulations of 5 ps, with reduced restraints of 100 kcal/mol·$Å^2$. The temperature was then increased to approximately 200 K during MD runs lasting around 100 ps in the isothermal-isobaric ensemble (NPT), with further reduced restraints of 25 kcal/mol·$Å^2$. After the release of all restraints, the temperature of the systems was raised to 300 K in a single NPT simulation lasting 500 ps. After an equilibration period of approximately 1.1 ns, NPT runs lasting around 10 ns were conducted to stabilize the system density at around 1.01 g/cm^{-3}.

Finally, MD simulations of approximately 1 μs were performed in an NVT ensemble using the GPU version of AMBER 18 (27), which was utilized for all simulation steps in this study, for each of three single-point mutants and WT enzyme with three replicates each lasting 400 ns after discarding the first 200 ns of the first replicate.
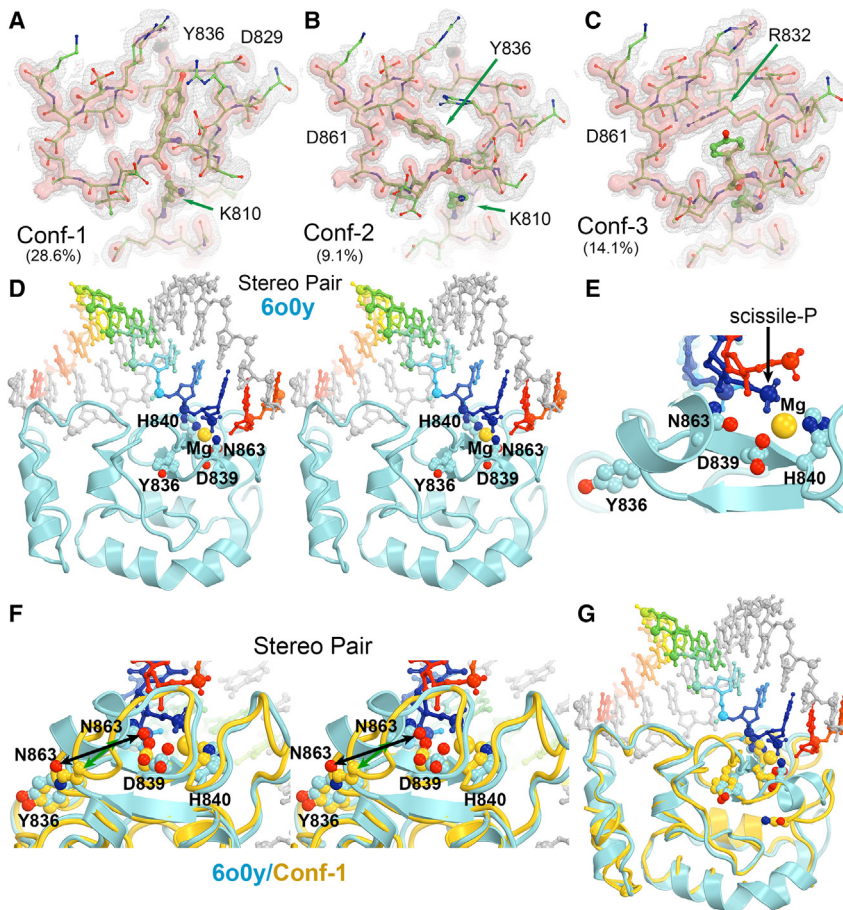
FIGURE 1 Three conformations of WT HNH discovered in MD trajectories and their relationship with the known active state of the cryo-EM structure (PDB ID: 6O0Y). (*A–C*) Three conformations as defined by MD-derived electron density (EM) maps. Arrows indicate labeled residues. (*D*) Stereo-diagram of the HNH domain complexed with the cleavage product DNA duplex taken from the 6o0y structure. The cleavage strand is shown in rainbow colors from 5′ to 3′ for both segments, and the opposite strand is shown in silver. Mg^{2+} ion is shown as a large gold sphere. Three catalytic residues (N863, D839, and H840) and the Y836 residue are shown in ball-and-stick models. (*E*) A close-up of the catalytic site. (*F*) Stereodiagram of superposition of the active structure (PDB ID: 6O0Y, *cyan*) with conformer-1 (*gold*). Large displacements of N863 are indicated by double-headed arrows for Cα (*green*, 5.4 Å) and Oδ1 (*black*, 9.5 Å) atoms. (*G*) Zoom-out view of the entire HNH domain with the same superposition of (*F*). Adapted from Wang et al., 2022a/b with permission.

Analytical method

PCA is used for dimension reduction of high-dimensional data as implemented in the pyEMMA package (28), which is a pipeline including a collection of unsupervised analysis methods to interpret molecular mechanics data. The features were selected to be the Cartesian coordinates of non-H atoms. The first two principal vectors with the largest principal components were selected to project each snapshot onto a two-dimensional Euclidean space with their corresponding projected values. The projection matrix, P, was computed using the first n eigenvectors (v_i) corresponding to the n largest eigenvalues of the covariance matrix C of the centered data ($X_{centered} = X − \mu$). The data were then projected onto the lower-dimensional subspace. The reduced data matrix ($X_{projected}$) was calculated by projecting the centered data ($X_{centered}$) onto the subspace spanned by the first n eigenvectors using the projection matrix P ($X_{projected} = X_{centered} \times P$). These projected data were used to build contours for visualization. This analysis was independently carried out for each of three single-point mutant MD trajectories and for the WT enzyme trajectories. All three replicates of individual systems were concatenated for analysis.

RESULTS AND DISCUSSION

Three distinct inactive conformational states of the wild-type HNH domain

The Y836-containing loop of HNH often exhibited very weak crystallographic electron density features and could

not be fully explained using a single-atomic model (29,30). When the crystallographic structure of the isolated HNH domain was used for MD simulations, it was found that this starting model indeed represents the dominant form in MD trajectories, but there exist at least two more distinct minor conformational states (Fig. 1) (15). These three conformations, conformer-1, conformer-2, and conformer-3 were estimated to have fractions of 29%, 9%, and 14% under the following three stringent definitions, respectively: 1) frames with the Y836Oη-D829N distance <3.30 Å, 2) frames with the Y836Oη-D861N distance <3.30 Å, and 3) frames with both the Y836Oη-D829N distance and the Y836Oη-D861N distance >3.50 Å. The Y836Cα displacement is 2.8, 1.5, and 4.0 Å from conformer-1 to -2, and from conformer-2 to -3, and from conformer-1 to -3, respectively. The corresponding displacements of the Y836Oη atom are 8.6, 7.0, and 11.2 Å. Throughout MD trajectories, there is no direct transition event between conformer-1 and -3. All transition events are between conformer-1 and -2 and between conformer-2 and -3 (Fig. 2 *A*).

These three conformations belonged to inactive states in which the catalytic residue N863 was displaced from the known active state (Fig. 1) (15). In the active site, N863
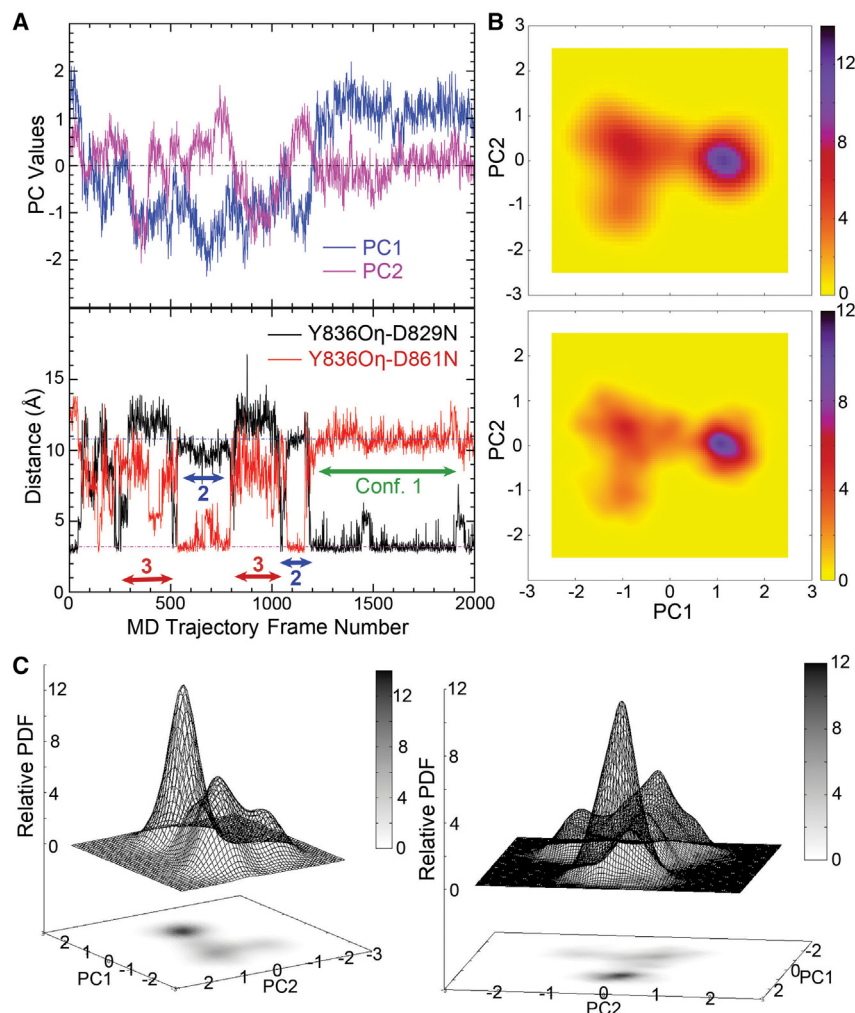
FIGURE 2 The definition of three or more states in Cas9 HNH MD trajectories. (A) Comparison of PC values (*top panel*, PC1, *blue*, PC2, *magenta*) with distance analysis (*lower panel*, Y836Oη-D829N, *black*, Y836Oη-D861N, *red*) as a function of MD frame number. There is only an approximate correlation between the two panels. (B) Relative probability density function plots as a function of PC1 and PC2 in two resolutions of pixel sizes (*top panel*, 0.1 Å × 0.1 Å with Gaussian blurring factor of 0.25 Å; *lower panel*, 0.05 Å × 0.05 Å with Gaussian blurring factor of 0.15 Å). (C) Corresponding 3D plots for the two resolutions of pixel sizes (*left panel*, low resolution; *right*, high resolution).

lies at the C-terminal end of an α-helix and coordinated with the single catalytic divalent metal ion (31). In all three inactive conformations, this helix is completely unfolded and adopts a similar coiled-coil conformation with an N863 Cα displacement of 5.4 Å and an Oδ1 displacement of 9.5 Å relative to the active state. Therefore, the activation pathway involves a refolding process to form the α-helix from the coiled-coil conformation. The inactive coiled-coil conformation is stabilized by multiple hydrogen bonds, including one in conformer-2 between Y836Oη and D861N, which is only one residue away from N863. Disruption of this hydrogen bond is therefore an essential step for the activation process, which led to a hypothesis that conformer-3 (in which the Y836 side chain has no hydrogen bond) could be an intermediate for HNH activation. In all three inactive states, the conformation of the N863-containing sequence remained a coiled-coil. Likewise, this stretch of sequence in three single-point specificity-enhancing Lys-to-Ala mutants (i.e., K810A, K848A, and K855A) also remained in the inactive coiled-coil conformation. It is noted that K848 interacts with the linker-1 loop that con-

nects HNH to RuvC and thus becomes part of the RuvC active site when it binds the single-stranded ntDNA substrate (32).

Previous identification of the three conformations used conventional crystallographic model refinement procedures through a series of guessed models built to explain MD-derived electron density maps (15). The procedure was tedious and required a large amount of crystallographic refinement for the full resolution of these conformations. The analysis described in this study can bridge the conventional model refinement and PCA for identification of individual conformational states and resolution of distinct substructures at equilibrium.

Visualization of principal component distributions in MD trajectories

A question also arises as to whether we may have missed other short-lived, but essential, conformational states in MD trajectories through local geometric analysis. For a given local hydrogen bond criterion between Y836Oη and
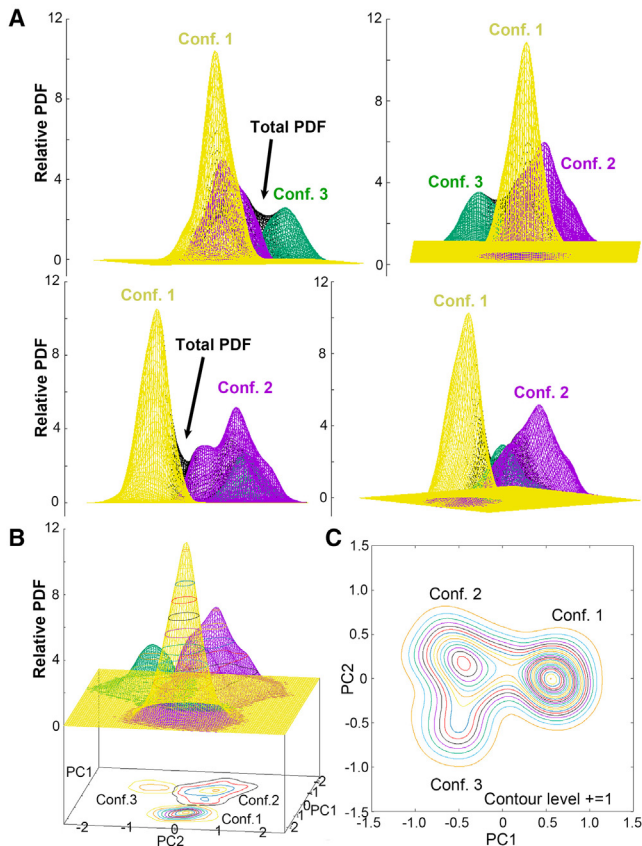
FIGURE 3 Clustering of MD trajectories using PCA methods. (A) Minimal three clusters in four views of the entire MD trajectories: (i) conformer-1 (yellow): PC1 > 0.4, (ii) conformer-2 (magenta, which can be further divided into multiple substates if there are necessary frames): PC1 ≤ 0.4 and PC2 > −0.5), and (iii) conformer-3 (green): PC1 ≤ 0.4 and PC2 ≥ −0.5. (B) Decomposition of three main peaks in three-dimensional plot and its projection. (C) Two-dimensional plot of the conformers contoured at +1σ increments.

D829N, we cannot be certain that this represents a single conformational state with no other conformational difference elsewhere in the entire structure among the selected MD frames. In fact, this issue occurs during MD simulations of the K855A variant but can be addressed using a global PCA (33), which includes information for all atoms of the entire trajectory. To convert discrete data points of MD trajectories into a continuous distribution function, we applied a Gaussian mixture model (GMM) (34), with simple isotropic Gaussian blurring factors for filtering to compensate for insufficient sampling. When isotropic Gaussian blurring factors of 0.25 or 0.15 Å^2 were applied, we saw only three major peaks in the PC1-PC2 plots, corresponding to the three inactive conformers described above (Fig. 2). Thus, there appear to be no other major substates within the three conformers, whereas minor substates of intermediates do exist when the PCA is combined with two-dimensional distance distribution analysis. PCA of several other MD simulations of the isolated HNH domain, as well as of the intact Cas9, were carried out previously to reveal

similar peak distributions, independent of starting points (16). These peaks in PC1-PC2 plots went unexplained until we discovered the three conformations (15). Additionally, the high-resolution crystal structure of the isolated HNH domain buried the mobile Y836 loop at the interface of a dimer present in the crystallographic asymmetric unit, which likely hampered the discovery of the dynamic nature of this loop (16).

According to PCA, conformer-1 is the most symmetrically distributed single state, whereas conformer-2 has four satellite substates surrounding its main state after peak decomposition (Fig. 3 A). Additional MD simulations could help visualization and characterization of these substates. On the other hand, conformer-3 is not a satellite substate of conformer-2 and is instead a distinct state. Once the dominant conformer-1 is removed from the distribution using GMM, another intermediate between conformer-1 and -2 becomes more clearly visible that was obscured by the dominant conformer-1 in the original distribution (Fig. 3). Thus, this state is clearly visible in one-dimensional distance distributions because it involves switching the hydrogen bond of Y836Oη from D829N to D861N (Fig. 2).

An expansion of distance distribution from one dimension in our original study to two dimensions by treating the Y836Oη-D829N distance and the Y836Oη-D861N distance as independent variables helped us to characterize all intermediate substates among the three conformers (Fig. 4). The connectivity of these substates, three from conformer-1 to -2, one between conformer-2 to -3, and none directly from conformer-1 to -3. These observations suggest that interconversions between conformer-1 and conformer-2 and between conformer-2 and conformer-3 are far more likely than direct interconversion between conformer-1 and conformer-3. When the same analysis was applied to MD trajectories for the three Lys-to-Ala mutants, it was clear that these variants have the single dominant conformer-1 with no detectable conformer-2 or -3 near the Y836-containing loop and its interacting partners (Fig. 5). However, all three mutants have multiple peaks in the PC1-PC2 plots (Fig. 5), suggesting that these mutations reduced the dynamic properties of the Y836 loop and its interacting partners, while also increasing dynamics elsewhere, which are yet to be characterized.

We also carried out PCA using only backbone atoms and obtained very similar results to those using all atoms, although the magnitude of PCA values was reduced proportionally. This implies that backbone atoms alone contain the same information as all atom PCA. To establish how backbone conformations are linked to the three orientations of the Y836 side chains, which form a hydrogen bond with either D829N or D861N or neither, we examined the torsion angle χ1 of the Y836 side chain in MD trajectories (Fig. 6 A). It is worth noting that χ2 rotation does not change the location of its Oη atom and therefore does not alter its hydrogen bond capability (Fig. 6 A). We unexpectedly
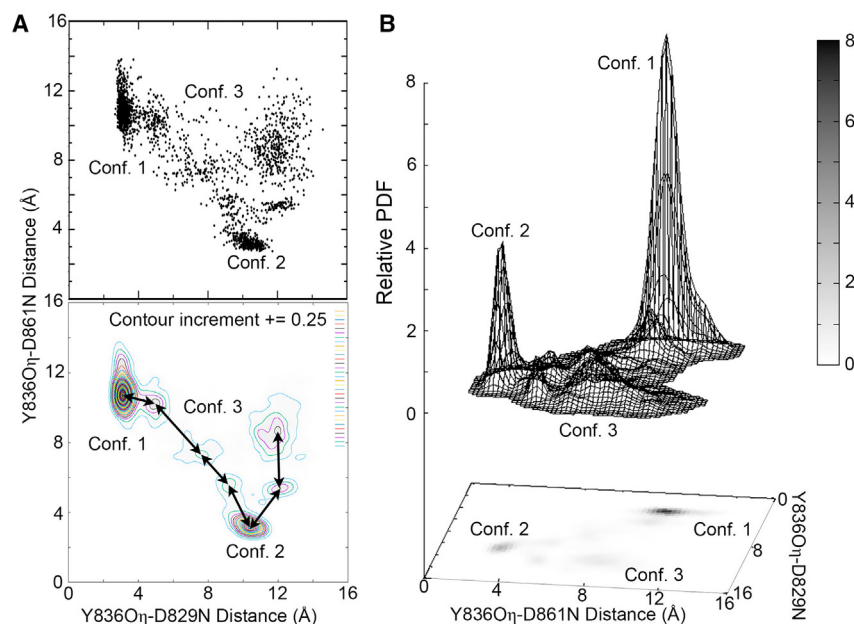
FIGURE 4 Clustering of MD trajectories using distinctive distance features. (*A*) Raw distance plots (*horizontal axis*, Y836Oη-D829N; *vertical axis*, Y836Oη-D861N) in the original discrete data (*upper panel*) and a Gaussian-smoothed plot (*lower panel*, pixel spacing, 0.1 Å, and standard deviation of Gaussian filtering function, 0.1 Å), contoured at 0.25σ increments. A hypothetical pathway is indicated with double-headed arrows. (*B*) Three-dimensional plot of the HNH conformer populations and its projection.

found that χ1 has two peaks in its probability density function but not the expected three peaks for the three conformers. This is due to the contribution of backbone torsion angles Φ and Ψ in defining local side chain conformations and thus the overall conformational state. In the WT enzyme, backbone conformations of both Y836 and D837 residues have multiple states (Fig. 6). When plotted over MD simulation frame number, Ψ values of Y836 exhibit a strong correlation with both the Y836Oη-D829N and Y836Oη-D861N distances (Fig. 6 *B*).

For the three Lys-to-Ala single-point mutants, both K810A and K848A have the single dominant conformation of conformer-1 with no detectable minor conformations near Y836 loop and its interacting partners. However, the K855A variant has a somewhat different distribution next to conformer-1 according to its distance-distance plot
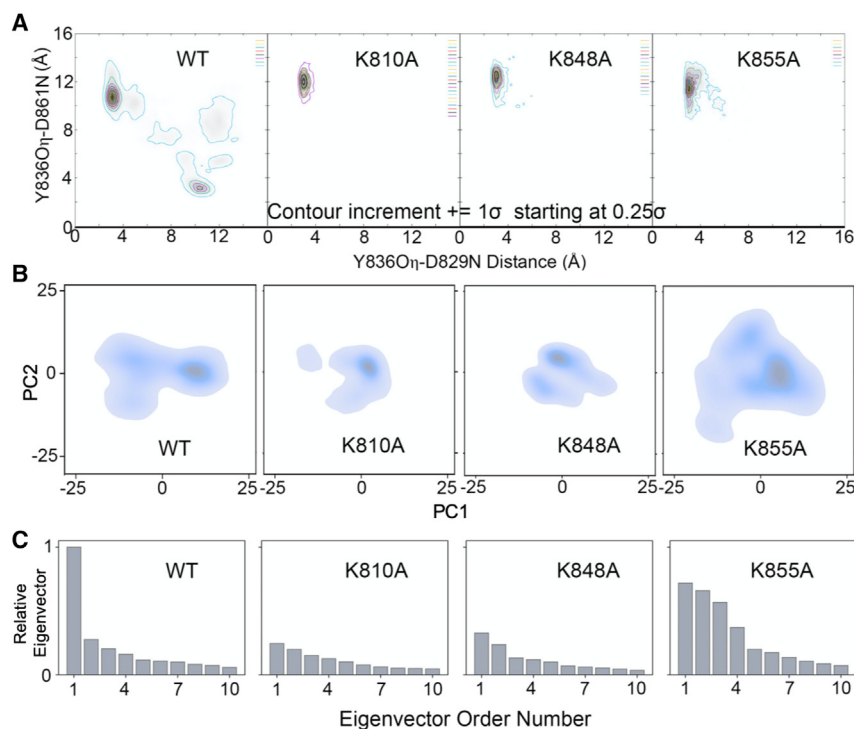

FIGURE 5 Comparison of the WT HNH domain with those of K810A, K848A, K855A variants. (*A*) Bond distance analysis for Y836Oη-D829N and Y836Oη-D861N. (*B*) PC1-PC2 correlations. (*C*) Distribution of normalized eigenvector lengths.
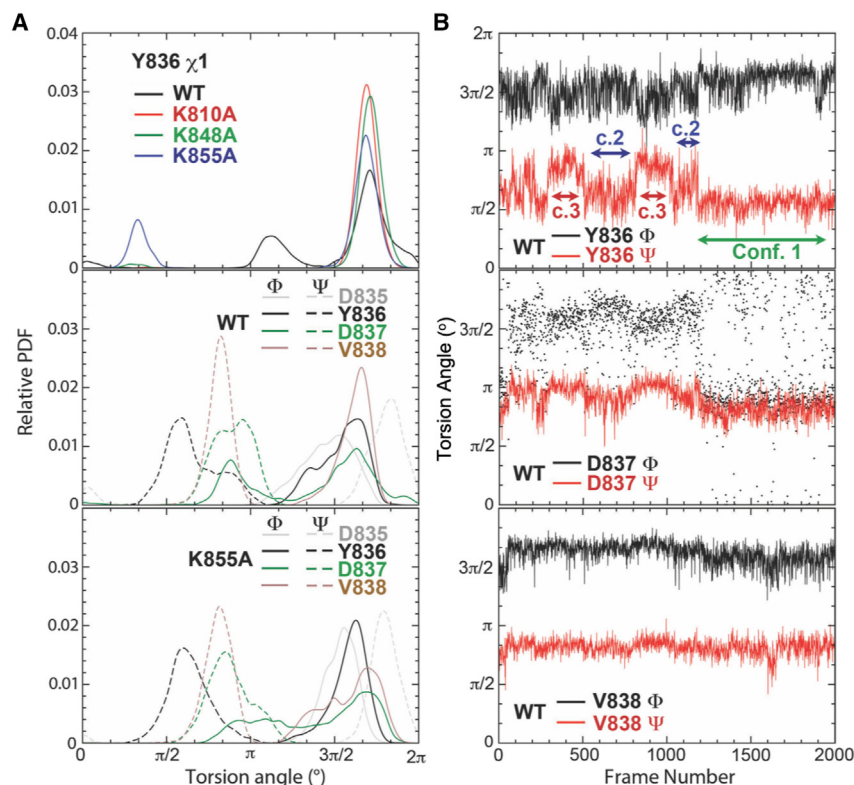
FIGURE 6 Backbone and side-chain torsion angles of Y836 and nearby residues in both WT HNH and Lys-to-Ala variants. (*A*) Relative probability density function of distribution. Top panel: χ1 of the WT Y836 side chain. Middle: Φ (*solid curves*) and Ψ (*dashes*) of D835 (*gray*), Y836 (*black*), D837 (*green*), and V838 (*brown*) of the WT enzyme. Bottom: the same Φ (*solid curves*) and Ψ distributions for the K855A variant. (*B*) MD trajectories of the WT enzyme for Y836 (*top panel*), D837 (*middle*), and V838 (*bottom*), showing that the backbone torsion transition of Y836 is independent of those of D837 and V838. Each frame was sampled at 500 ps with a total of 1 μs for 2000 frames.

(Fig. 5 *B*). Its PC1-PC2 plot also exhibited far more complex patterns than the two other variants (Fig. 5). The Y836 side chain and backbone torsion angles appear to exhibit compensatory co-variations, resulting in a similar single-peak distribution of the Y836-D829N distance (Fig. 6). All other structural differences appear distant from the Y836-containing loop, and PC1-PC2 plots of all three single-point Lys-to-Ala mutants showed some structural differences, likely at the mutation sites. Nevertheless, these mutations reduce the dynamic properties of the Y836-containing loop through allosteric communication, consistent with prior NMR and MD studies (Nierzwicki et al., 2021).

Substrate-independent activation pathways of the HNH domain

Combined analysis of the global PC features of the entire HNH domain with the local structural features of single hydrogen bond geometry centered at the Y836 side chain allows us to define the temporal relationship between the activated state and three inactive conformations of HNH that are present in the WT enzyme but absent in Lys-to-Ala specificity-enhancing variants (Fig. 6). Even though high-resolution, catalytically relevant HNH structures with a tDNA/gRNA duplex remain sparse, the catalytic mechanism likely involves one metal ion (but unlikely two as in the classic two-metal-ion mechanisms) according to kinetic studies as

well as site-directed mutagenesis (8,35–38). Additional MD and quantum mechanics/molecular mechanics studies also support a proposed mechanism in which N863 is catalytically important (25,38,39). We emphasize here that MD simulations and computational analyses are instrumental in explaining biochemical data, reconciling them with structural biology (40).

Comparison of the known active state with each of the three conformers identified for the WT HNH nuclease of Cas9 reveals that conformer-3 is a gateway to the active state (Fig. 7). During this step, the Y836 side chain frees its hydrogen bond with the D861 backbone, permitting a local unfolding and refolding event of the region surrounding the D861-N863 pair (Fig. 7). This activation pathway does not require the binding of a cognate tDNA/gRNA heteroduplex, which makes it more promiscuous in substrate selection and cleavage. For adapted innate immunity, a modest amount of substrate promiscuity for HHN could serve as an advantage in recognizing closely related DNA sequences of subtly mutated pathogens. On the other hand, from the perspective of engineering, it is advantageous to eliminate any promiscuity to improve substrate selectivity. The three Lys-to-Ala single-point mutants have indeed eliminated this substrate-independent activation pathway and also improved substrate selectivity, providing a rationale for their high degree of substrate selectivity that absolutely requires formation of an R-loop before the HNH nuclease can be activated. In addition, removal of
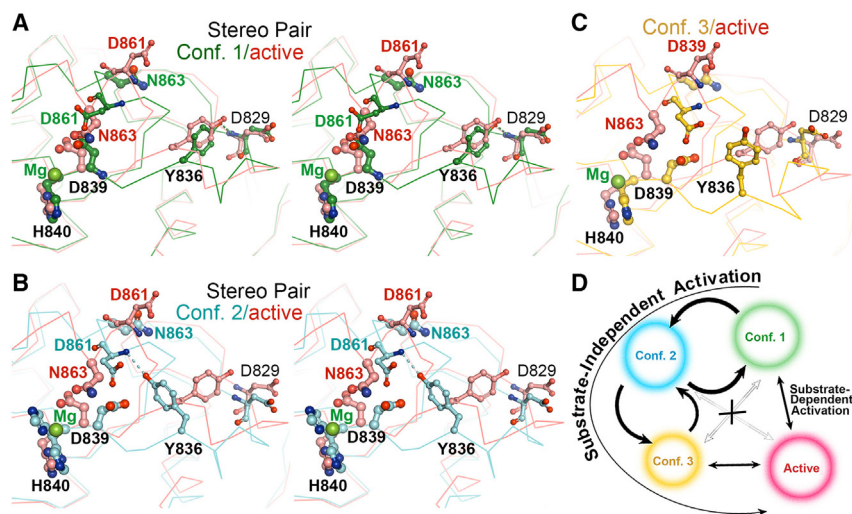
FIGURE 7 Structural basis for substrate-independent (conformation 1 to conformation 2, to conformation 3, and to the active state) and substrate-dependent (conformation 1 to the active state through a mechanism yet to be established) pathways. (*A*) Stereodiagram comparison between conformation 1 (*green*) and the active state (*salmon*). (*B*) Stereodiagram comparison between conformation 2 (*cyan*) and the active state (*salmon*). (*C*) Comparison between conformation 3 (*gold*) and the active state (*salmon*). (*D*) Scheme of the two possible activation pathways. Pathways that are not directly interconvertible are shown in open double-headed arrows.

positively charged Lys side chains could help to prevent any binding of nonspecific DNA without first recognizing its sequence, as well as from prolonged interaction with non-substrate nucleic acids that can lead to enzymatic errors during the substrate selection process.

Broad implications and prospective

Although the focus of this study was to identify three distinct inactive states of HNH and to establish their relationship to the active state, we may have identified a bottleneck in the Cas9 activation process, likely involving a folding-refolding transition of the catalytic loop and substantial conformational dynamics of the Y836-containing loop. In such a case, the overall catalytic rate of the HNH endonuclease could be altered by designed mutations that specifically affect the rate of this folding-refolding process. This hypothesis could be validated experimentally as was previously done for adenylate kinase (41). Furthermore, interesting parallels exist between high binding affinity and a high degree of selectivity of both substrate(s) and inhibitor(s) to an enzyme. These scenarios can describe both induced-fit and conformational selection mechanisms (42). For example, DNA polymerases undergo a substrate-independent opening-closing motion of the active site pocket when binding both cognate and noncognate substrates via a conformational selection mechanism. It is likely that an additional induced-fit process enables the cognate substrate to bind with much higher affinity than noncognate substrates (43,44). As demonstrated recently (15), three individual states of the HNH domain can be characterized by NMR spectroscopy; thus a more complete view of conformational selection in HNH may emerge once kinetic rates and the complete energy landscape of this small, but important, enzymatic subdomain are accurately determined.

CONCLUSIONS

In summary, we describe an integrated approach of global PCA with local geometric analysis (such as backbone and side-chain torsional angles and selected interatomic distances) of MD simulation trajectories. This approach helps to identify and site-specifically localize important structural and dynamic properties required to interpret and understand experimental data. Our approach provides direct visualization of conformational states and minor substates in MD simulations and dynamic properties of large enzymes using a global PCA alongside local geometry analysis. Our integrative analysis of MD simulations also provides a further molecular basis for how three Lys-to-Ala single-point mutations enhance the selectivity of substrate binding and cleavage relative to WT Cas9. This work helps to further understand the mechanism of HNH and principles that can redesign it with improved specificity. Our approach could have broad applications to a variety of enzymes, such as DNA polymerases that exhibit a very high degree of base selectivity.

AUTHOR CONTRIBUTIONS

J.W., V.S.B., G.P.L., and G.P. conceptualized this study. P.R.A. performed molecular dynamics simulations. F.M., T.Q., and J.W. carried out analysis and made illustrations. J.W. wrote the draft manuscript. All authors contributed to writing and editing the final version of the manuscript.

DECLARATION OF INTERESTS

Authors declare no competing interest in publishing this study.

REFERENCES

1. Jiang, F., and J. A. Doudna. 2017. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* 46:505–529.

2. Doudna, J. A., and E. Charpentier. 2014. Genome editing The new frontier of genome engineering with CRISPR-Cas9. *Science.* 346, 1258096.

3. Alberts, B. 2003. DNA replication and recombination. *Nature.* 421:431–435.

4. Lang, K. S., and H. Merrikh. 2018. The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria. *Annu. Rev. Microbiol.* 72:71–88.

5. Liu, B., and B. M. Alberts. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science.* 267:1131–1137.

6. Merrikh, H., C. Machón, …, P. Soultanas. 2011. Co-directional replication-transcription conflicts lead to replication restart. *Nature.* 470:554–557.

7. Wang, J., P. R. Arantes, …, G. Palermo. 2022. Twisting and Swiveling Domain Motions in Cas9 to Recgonize Target DNA Duplexes, Make Double-Strand Break, and Release Cleaved Duplexes. *Front. Mol. Biosci.*, In Press.

8. Raper, A. T., A. A. Stephenson, and Z. Suo. 2018. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.* 140:2971–2984.

9. Gong, S., H. H. Yu, …, D. W. Taylor. 2018. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* 22:359–371.

10. Dagdas, Y. S., J. S. Chen, …, A. Yildiz. 2017. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* 3, eaao0027.

11. Chen, J. S., Y. S. Dagdas, …, J. A. Doudna. 2017. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature.* 550:407–410.

12. Saha, A., P. R. Arantes, and G. Palermo. 2022. Dynamics and mechanisms of CRISPR-Cas9 through the lens of computational methods. *Curr. Opin. Struct. Biol.* 75, 102400.

13. Slaymaker, I. M., L. Gao, …, F. Zhang. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science.* 351:84–88.

14. Nierzwicki, L., K. W. East, …, G. Palermo. 2021. Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *Elife.* 10, e73601.

15. Wang, J., E. Skeens, …, V. S. Batista. 2022. 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. *Biochemistry.* 61:785–794.

16. East, K. W., J. C. Newton, …, G. P. Lisi. 2020. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* 142:1348–1358.

17. Anders, C., O. Niewoehner, …, M. Jinek. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature.* 513:569–573.

18. Olsson, M. H. M., C. R. Sondegaard, …, J. H. Jensen. 2011. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theor. Computation.* 7:525–537.

19. Pérez, A., I. Marchán, …, M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* 92:3817–3829.

20. Banáš, P., D. Hollas, …, M. Otyepka. 2010. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theor. Comput.* 6:3836–3849.

21. Zgarbová, M., M. Otyepka, …, P. Jurečka. 2011. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theor. Comput.* 7:2886–2902.

22. Allnér, O., L. Nilsson, and A. Villa. 2012. Magnesium Ion-Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theor. Comput.* 8:1493–1502.

23. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79:926–935.

24. Ricci, C. G., J. S. Chen, …, G. Palermo. 2019. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* 5:651–662.

25. Palermo, G. 2019. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J. Chem. Inf. Model.* 59:2394–2406.

26. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* 81:3684–3690.

27. Case, D. A., R. M. Betz, …, P. A. Kollman. 2018. AMBER 2018. University of California.

28. Scherer, M. K., B. Trendelkamp-Schroer, …, F. Noé. 2015. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theor. Comput.* 11:5525–5542.

29. Nishimasu, H., F. A. Ran, …, O. Nureki. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell.* 156:935–949.

30. Hirano, S., H. Nishimasu, …, O. Nureki. 2016. Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol. Cell.* 61:886–894.

31. Zhu, X., R. Clarke, …, S. Subramaniam. 2019. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* 26:679–685.

32. Jiang, F., D. W. Taylor, …, J. A. Doudna. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science.* 351:867–871.

33. David, C. C., and D. J. Jacobs. 2014. Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol. Biol.* 1084:193–226.

34. McLachlan, G. J., S. X. Lee, and S. I. Rathnayake. 2019. Finite Mixture Models. *Annu. Rev. Stat. Appl.* 6:355–378.

35. Zuo, Z., A. Zolekar, …, J. Liu. 2019. Structural and functional insights into the bona fide catalytic state of Streptococcus pyogenes Cas9 HNH nuclease domain. *Elife.* 8, e46500.

36. Tang, H., H. Yuan, …, Q. Huang. 2021. Active-Site Models of Streptococcus pyogenes Cas9 in DNA Cleavage State. *Front. Mol. Biosci.* 8, 653262.

37. Gasiunas, G., R. Barrangou, …, V. Siksnys. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA.* 109:E2579–E2586.

38. Nierzwicki, Ł., K. W. East, …, G. Palermo. 2022. Principles of target DNA cleavage and role of Mg2+ in the catalysis of CRISPR-Cas9. *Nat. Catal.* 5:912–922.

39. Casalino, L., Ł. Nierzwicki, …, G. Palermo. 2020. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal.* 10:13596–13605.

40. Brunk, E., N. Ashari, …, U. Rothlisberger. 2011. Pushing the frontiers of first-principles based computer simulations of chemical and biological systems. *Chimia.* 65:667–671.

41. Kerns, S. J., R. V. Agafonov, …, D. Kern. 2015. The energy landscape of adenylate kinase during catalysis. *Nat. Struct. Mol. Biol.* 22:124–131.

42. Agafonov, R. V., C. Wilson, …, D. Kern. 2014. Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. *Nat. Struct. Mol. Biol.* 21:848–853.

43. Millar, D. P. 2022. Conformational Dynamics of DNA Polymerases Revealed at the Single-Molecule Level. *Front. Mol. Biosci.* 9, 826593.

44. Turvey, M. W., K. N. Gabriel, …, P. G. Collins. 2022. Single-molecule Taq DNA polymerase dynamics. *Sci. Adv.* 8:eabl3522.