Leveraging QM/MM and Molecular Dynamics Simulations to Decipher the Reaction Mechanism of the Cas9 HNH Domain to Investigate off-Target Effects

Yazdan Maghsoud,[1,†] Vindi M. Jayasinghe-Arachchige,[2,†] Pratibha Kumari,[2] G. Andrés Cisneros,*[,1,3] and Jin Liu*[,2]

[1]*Department of Chemistry and Biochemistry, University of Texas at Dallas, Richardson, TX 75080, USA*

[2]*Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, TX 76107, USA*

*E-mail: Jin.Liu@unthsc.edu

[3]*Department of Physics, University of Texas at Dallas, Richardson, TX 75080, USA*

*E-mail: andres@utdallas.edu

[†]*These authors contributed equally to this manuscript*

ABSTRACT

The clustered regularly interspaced short palindromic repeats (CRISPR) technology is an RNA-guided targeted genome-editing tool using Cas family proteins. Two magnesium-dependent nuclease domains of this enzyme, termed HNH and RuvC, are responsible for cleaving the target DNA (t-DNA) and non-target DNA (nt-DNA) strands, respectively. The HNH domain is believed to determine the DNA cleavage activity of both endonuclease domains and is sensitive to complementary RNA-DNA base pairing. However, the underlying molecular mechanisms of CRISPR-Cas9, by which it rebukes or accepts mismatches, are poorly understood. Thus, investigation of the structure and dynamics of the catalytic state of Cas9 with either matched or mismatched t-DNA can provide insights for improving its specificity by reducing off-target cleavages. Here, we focus on a recently discovered catalytic-active form of the *Streptococcus pyogenes* Cas9 (SpCas9) and employ classical molecular dynamics (MD) and coupled quantum mechanics/molecular mechanics (QM/MM) simulations to study two possible mechanisms of t-DNA cleavage reaction catalyzed by the HNH domain. Moreover, by designing a mismatched t-DNA structure called MM5 (C to G in the fifth position from the PAM region), the impact of single-guide RNA (sgRNA) and t-DNA complementarity on the catalysis process was investigated. Based on these simulations, our calculated binding affinities, minimum energy paths, and analysis of catalytically important residues provide atomic-level details of the differences between matched and mismatched cleavage reactions. In addition, several residues exhibit significant differences in their catalytic role for the two studied systems, including K253, K263, R820, K896, and K913.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR) technology is an RNA-guided targeted genome engineering platform that utilizes Cas proteins for its function.[1, 2] CRISPR-Cas systems were first discovered in *E-coli* in 1987,[3] which provide adaptive immunity to prokaryotic and archaeal microorganisms against invaders from phages or plasmids.[4-9] Since the first reported use in mammalian cells in 2013,[10] extensive research efforts have been advancing this technology.[11-17] CRISPR-Cas systems have been categorized into two major classes,[18] where class 1 combines Cas proteins for RNA-guided targeting.[19] In contrast, only a single protein is required for RNA-guided DNA recognition and cleavage in class 2.[20] Six distinct Cas protein types are grouped into these two classes, i.e., I, III, IV under class 1 and II, V, VI in class 2.[21, 22] Among them, Cas9, the class 2 type II protein accompanying the CRISPR system (CRISPR-Cas9), has been purposed as a powerful tool with the introduction of a single-guide RNA (sgRNA) that fuses the CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) for programmable DNA binding and cleavage.[20] Furthermore, the type II-A *Streptococcus pyogenes* Cas9 (SpyCas9 or SpCas9) is the widely studied variant for gene editing applications in various living cells and organisms to improve the Cas9 toolbox for its DNA cleavage specificity.[23-28]

Site-specific DNA recognition and cleavage require the assembly of SpCas9 with a sgRNA, resulting in a binary complex. A 20-nt sgRNA sequence segment in this binary complex should complement the one-strand target DNA (t-DNA) of the incoming target/foreign double-strand DNA for the tertiary complex formation and subsequent cleavage activity. Additionally, a short sequence of nucleotides on the non-target DNA strand (nt-DNA) called protospacer adjacent motif (PAM) facilitates the identification of the desired DNA sequence across the genome for programmable editing.[29-31] PAM recognition instigates the pairing of the sgRNA with t-DNA by forming an RNA:DNA hybrid, whereas the other strand (nt-DNA) is displaced. Afterward, SpCas9 uses its two-magnesium-ion-dependent endonuclease domains, HNH and RuvC, to cleave the t-DNA and nt-DNA, respectively. However, the catalytic mechanisms of DNA cleavage, a preliminary step for genome editing, are not entirely understood. Moreover, the risk of off-target DNA cleavage is one of the primary concerns hindering this editing tool's use for therapeutic applications.[32-34] In the past, numerous studies demonstrated that the DNA binding process and the RNA-DNA complementary play a significant role in the catalytic activity of SpCas9's endonuclease domains.[35, 36] A detailed molecular understanding of the catalytic mechanism of DNA cleavage in CRISPR-Cas9 with mismatched DNA is imperative for developing specific SpCas9 variants with improved targeting specificity.

Multiple SpCas9 crystal structures,[37-42] cryo-electron microscopy (cryo-EM) structures,[43-45] and molecular dynamics (MD) simulation models[46-48] in different binding forms have been reported over the past few years. Despite these efforts, the cleavage-competent conformation of the catalytic HNH nuclease domain of SpCas9 remained largely elusive and debatable due to the high flexibility of the HNH domain during the different stages of SpCas9 mediated cleavage process. Nevertheless, the HNH domain of SpCas9 was observed to share structural similarities with other nucleases, i.e., periplasmic nuclease Vvn, Endonuclease Colicin E9, Staphylococcal nuclease, and T4 Endonuclease VII.[49-51] These structures contain a conserved histidine residue and an

aspartate/glutamate in their catalytic site, arranged in an orientation consistent with the characteristic catalytic mechanism of one-metal-ion-dependent nucleic acid-cleaving enzymes.[52, 53] Different cleavage mechanisms have been proposed for these systems depending on the catalytic base, nucleophile, and metal ion in the catalytic site.[54-56] In previous efforts to understand the catalytic mechanism of t-DNA cleavage reaction in the HNH domain, QM/MM studies revealed a strategy to model a potential cleavage conformation for the HNH active site using the cryo-EM structure[54, 55] available at the time of their studies (PDB ID: 5Y36)[57]. Detailed mechanisms either using metal-bound water (**Figure 1A**) or a second coordination shell water (**Figure 1B**) were proposed from these QM/MM studies.

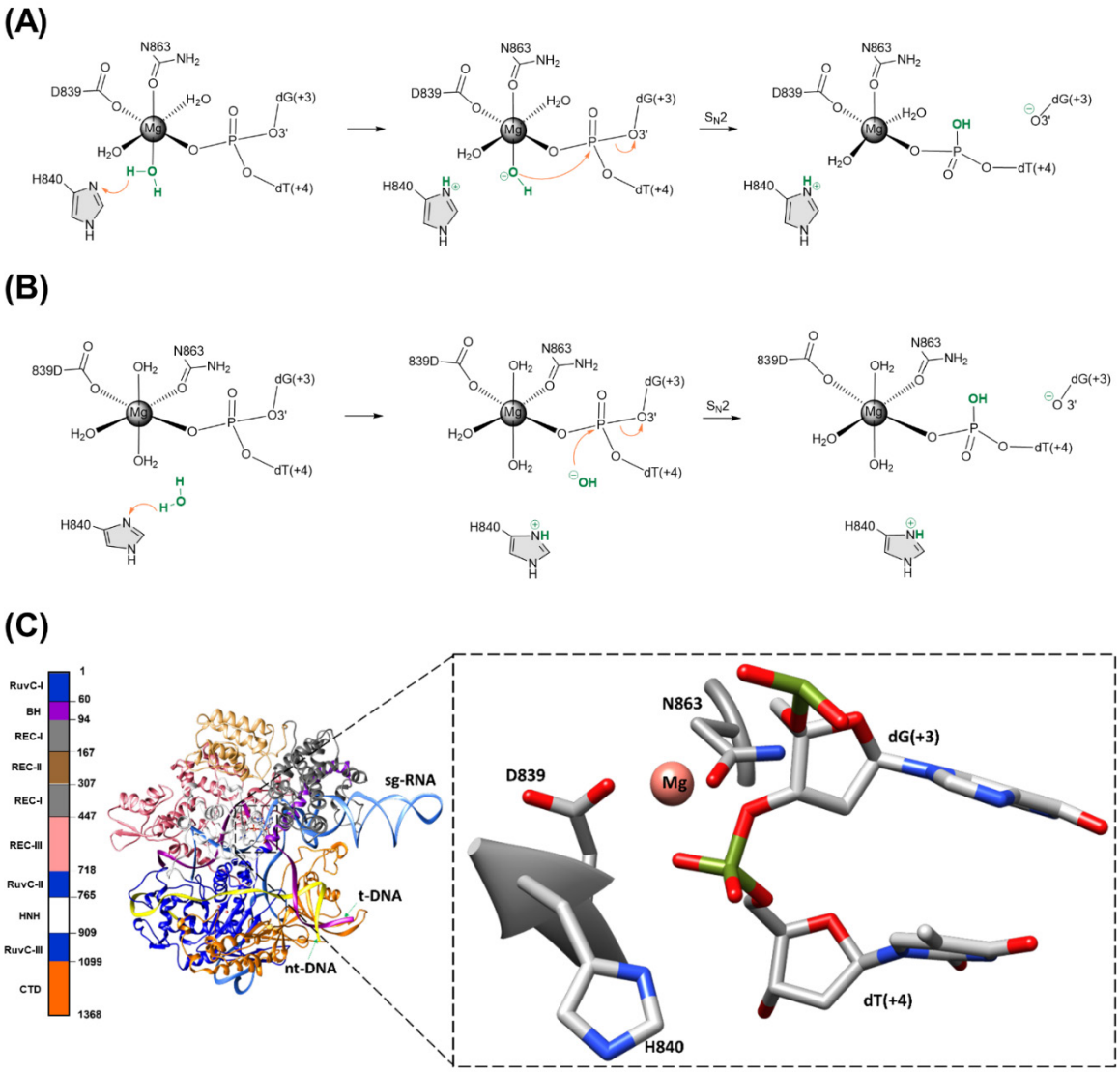


Figure 1. Schematic representation of the reaction mechanisms for the DNA cleavage at the HNH domain of SpCas9 by **(A)** first-shell water coordinated to Mg^{2+} or hydrolysis by **(B)** second-shell water around Mg^{2+}. **(C)** The initial model of SpCas9 (PDB ID: 6O0Y) and the close-up of the HNH's active site. Three coordinated water molecules to the magnesium ion are not shown.

In previous studies, the cryo-EM structure used (PDB ID: 5Y36) was solved at 5.2 Å resolution, where the position of the histidine base is located ~8 Å away from the scissile phosphate, and H840 was mutated to alanine

in the structure to inactivate the enzyme. Furthermore, N863, a catalytic residue known to hold a divalent cation in the HNH active site is ~10 Å away from the catalytic site (see **Figure S1**). Additionally, D861 shows coordination to the Mg^{2+} ion in the active site suggesting it to be a catalytic residue. However, the experiments by Zuo et al.[58] have demonstrated that D861 is not critical for HNH domain-catalyzed t-DNA cleavage, unlike what would be expected from the reported SpCas9 structures like 5Y36.

A recent cryo-EM study by Zhu et al.[59] resolved structures (at 3.4 Å resolution) of precatalytic, postcatalytic, and product states of the active SpCas9•sgRNA•DNA complex in the presence of Mg^{2+} ions. This study provides a unique platform for further investigating the DNA cleavage mechanism in the catalytically active conformation. However, the proposed catalytically competent structure (PDB ID: 6O0Y)[59] is missing several residues and the magnesium ions, requiring further modifications (detailed explanation is in the Methods section). Combined with our previous study[60] and incorporating structural features from this cryo-EM structure, we generated a precatalytic/active state model that resembles the catalytically competent complex resolved (see **Figure 1C** and **Figure S2**).

Recently, through molecular dynamics simulations of this catalytically competent active state model, we observed that base pair mismatches in the DNA at the proximal and distal end of the PAM significantly alter the cross-correlations between the catalytic residues of endonuclease domains and the arginine-rich BH helix depending on DNA mismatch positions.[61] Specifically, we noticed that the introduction of the proximal mismatch (at the fifth position from the PAM) of the t-DNA causes conformational shifts that substantially reduce the population of the conformations around the catalytic-active state, which may lead to a decrease in the rate constant observed in the kinetic experiments.[61]

Here, we present further insights into the catalytic mechanism of the HNH domain based on classical molecular dynamics (MD) and hybrid quantum mechanics/molecular mechanics (QM/MM) calculations. We used an active state model of SpCas9 with matched t-DNA and mismatched t-DNA (which we refer to as MM5 throughout the manuscript) to investigate how RNA:DNA complementarity affects the molecular-level reaction mechanism of t-DNA cleavage. Several potential residues have been identified, which can guide the structural engineering of CRISPR-Cas9 to reduce off-target effects.

COMPUTATIONAL METHODS

I. Molecular dynamics (MD) simulations

Structural Model: We used a stepwise approach to build the initial model due to some missing residues in the recently discovered active-state cryo-EM structure of SpCas9 (PDB ID: 6O0Y, **Figure S2A**). The missing regions of the SpCas9 protein are residues: 175−310 (REC2), 713−717, and 1002−1075 (RuvC-III), unresolved nucleotides of the nt-DNA, and the absence of metal ions in the nuclease domains. A structure generated from our previous simulation study,[60] which achieved the HNH precatalytic/active state, was utilized as the starting point. The mentioned structure was based on the most complete X-ray structure of SpCas9 in complex with sgRNA and DNA (PDB ID: 5F9R).[62] In our previous study,[60] Mg^{2+} ion was added to the HNH catalytic center, and nt-DNA

(present in 5F9R) was removed to achieve the HNH precatalytic/active state at a shorter time scale. The missing nt-DNA was included in the current study by a superposition with crystal structure 5F9R, and the unresolved portion was added manually. To retain the conformation of the RuvC catalytic center comparable to that of the cryo-EM structure (6O0Y), the coordinates of H983 and residues 3−12 were replaced by the corresponding regions from the cryo-EM structure. In this structure, the positions of two Mg^{2+} ions of the RuvC domain were derived from the X-ray crystal structure of CRISPR-Cas9 solved in a complex with Mn^{2+} ions (PDB ID: 4CMQ).[63] The final model, which is used as the starting point for the MD simulations, is shown in **Figure S2B**. Furthermore, to investigate the impact of sgRNA and t-DNA complementarity on the catalysis process, a mismatched system called MM5 was created by mutating the fifth position nucleotide downstream of the PAM on t-DNA (C to G). The corresponding nt-DNA nucleotide (G to C) was also mutated to maintain the complementarity between the t-DNA and nt-DNA.

MD setup: The LEaP module of AMBER18[64] was used to add the hydrogen atoms, neutralize the system with the corresponding number of required counterions, and solvate the structure in a rectangular box filled with TIP3P[65] water extending at least 12 Å from the complex surface. The ff14SB,[66] OL156,[67] and OL3[68] force fields were used to describe the molecular characteristics of the protein, DNA, and sgRNA, respectively. The MD simulations were done via AMBER18's pmemd.cuda.[69] Each system was minimized for 10,000 cycles by employing the steepest descent algorithm for the first 1000 cycles and the conjugated gradient algorithm for the remaining cycles with restraints on the solute's heavy atoms. In the next step, each system was heated to 310 K using Langevin dynamics[70-72] with a collision frequency of 2 ps^{-1} followed by equilibration for 1000 ps in an NPT ensemble, keeping lowered restraints on the heavy atoms of solute. Lastly, the production calculations were performed on an unrestrained system in the NPT ensemble. All bonds involving hydrogen atoms were treated using SHAKE,[73] and long-range Coulomb interactions[74] were handled with the smooth particle mesh Ewald method[75] using a 10 Å cutoff for non-bonded interactions. Individual simulations were run in duplicate, each for at least 200 ns with an integration time-step of 2 fs, and trajectories were saved at every 2 ps.

Structural analysis: RMSD, RMSF, correlation matrices, and clustering analyses were computed using AMBER's CPPTRAJ program.[76] To perform the clustering analysis, 100,000 trajectories in the 50 to 150 ns range—maintaining the catalytically competent HNH domain—from two replicates of the Matched and MM5 were used for a multi-dimensional analysis via the *k*-means algorithm[77] implemented in AMBER's CPPTRAJ. Each dimension of this analysis on the active site corresponds to a distance between the Mg^{2+} ion and its coordinated residues D839, H840, N863, and dT(+4). Ten clusters, each of which contained three representatives, were initially obtained to find the closest representatives to the centroids of each cluster in the Matched and the MM5 systems. In the next step, four clusters for the Matched and one for the MM5 with the highest population abundance and the best orientations of the active site's residues involved in the cleavage reaction were selected for further QM/MM optimizations.

II. MM/GBSA calculations

The molecular mechanics/generalized Born surface area (MM/GBSA)[78-80] method was employed using the "single-trajectory" protocol[81] to calculate the binding enthalpies for the Matched and MM5 systems via two different

approaches. In the first approach, DNA and the sgRNA+SpCas9 were considered the ligand and receptor, respectively. In the second one, the HNH's active site is regarded as the ligand (residues: 838–841, 863, 1493–1495, and 1541), while the rest of the system is considered as the receptor (residues: 1–837, 842–862, 864–1492, 1496–1540, and 1542–1543). The last 10,000 frames of MD for both replicates of each structure were used for the binding enthalpy calculations. The MM/GBSA calculations were performed via the *MMPBSA.py* internal module of AmberTools.[82] In addition to the computational efficiency of MM/GBSA, several studies have shown that this method results in comparable or even more accurate data in ranking ligand affinities compared to the molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA).[83-86] The offset and surface tension default values were used to correct the nonpolar contribution to the solvation free energy, and the salt concentration in the GB equation was set to 150 mM. Previous studies have shown that MM/GB(PB)SA can satisfactorily compare the relative ligand binding affinities, particularly when dealing with similar ligands.[83-90] Since the only difference between the Matched and MM5 is a G to C and G to C mutations in the fifth position of the ligand (t-DNA & nt-DNA, respectively), the entropic effect is not expected to be highly determinant.

III. QM/MM calculations

All QM/MM calculations were carried out with the LICHEM code[91, 92] combining the Gaussian16[93] and TINKER[94] programs. The ωB97X-D/6–31G(d,p)[95, 96] level of theory and the AMBER ff14SB force field were employed for the QM region and the MM environment, respectively. The QM/MM long-range electrostatic correction (QM/MM-LREC) method[97] was used with a 27 Å cutoff for the QM subsystem coupled with the particle mesh Ewald[74] (PME) method for the MM contribution. The QM subsystem for both systems includes Mg^{2+}, coordinated water molecules, V838, D839, H840, I841, N863, dG(+3), and dT(+4). Residues dC(+5) or its mutation dG(+5) were also added to the QM subsystem in the Matched and MM5 systems, respectively. In addition, the nucleophilic water in the second shell around Mg^{2+} was also included in the QM subsystem of the Matched$^{2nd\ shell}$ system. The remaining residues and all solvent molecules are described by the AMBER ff14SB potential. The pseudobond approach[98] was also applied to treat the covalent boundaries for the nucleic acid, i.e., dG(+3) and dC(+5)/dG(+5), and protein residues (V838, I841, and N863) of the QM subsystem. In all cases, the optimizations were carried out using the iterative QM/MM optimization protocol implemented in LICHEM,[91, 92] where all atoms in the MM subsystem within a radius of 27 Å from the center of the active site (Mg^{2+}) were optimized, and the rest were kept frozen.

After optimizing all the selected representatives of the Matched and MM5, the one with the lowest QM/MM optimization energy in each structure was considered the most stable reactant and was used to design the initial structure of the product. The simulated products were then used for further QM/MM calculations at the same level of theory. Based on the optimized reactant and product structures of each system (Matched and the MM5), the potential energy surface of the reaction path was tried to be obtained and compared using the quadratic string model (QSM) combined with a restrained MM procedure as implemented in LICHEM.[92] The restraint on the MM environment started at 50 kcal mol^{-1} $Å^{-2}$ and gradually decreased to zero. A chain of fourteen beads between the reactant (bead 0) and the product (bead 15), resulting in sixteen beads, was employed for guessing the reaction

path. The ESP charges of the reactant, approximate TS, and the product were also calculated using the Merz-Singh-Kollman[99] scheme from the QM/MM-optimized structures embedded with the electrostatic charges of the MM region at the same level of theory.

The critical points were approximated by using the QM/MM-optimized structures and obtaining the frequencies and thermochemistry using only the electrostatically-embedded system. These structures were then used for vibrational analysis via the Gaussian16[93] at the same levels of theory to investigate the approximate free energies. One negative imaginary frequency was obtained for the approximate TS of the Matched and MM5 corresponding to the motion along the reaction coordinates (see animations in the ESI). The activation Gibbs free energies ($\Delta G^{\ddagger}$) of the approximate TS in solvent were computed at 310 K and pH 7.0 based on transition state theory (TST)[100, 101] as implemented in the Eyringpy code.[102, 103]

Non-covalent interactions (NCI) were analyzed using the promolecular density method[104] implemented in the Multiwfn[105] code, using a cubic grid of 200 au. This analysis gives a qualitative view into the chemical bonding and weak noncovalent interactions between the molecule(s) of interest and the surrounding residues based on the relationship between the electronic density and the reduced density gradient in regions of low electron density. The isovalue of 0.4 au with the color scale of –0.05 au < sign(λ_2)ρ < 0.05 au was used to illustrate the NCI surfaces. The specific RGB colors of the NCI surfaces show the strength and characteristics of the interactions. For example, red surfaces show repulsive interactions, while green and blue surfaces represent weak and strong interactions like Van der Waals and hydrogen bonds.

The QM/MM-optimized structures of the reactant, product, and approximate TS were used for further MD simulations with restraints on the QM region to perform the energy decomposition analysis (EDA). In all cases, in addition to the optimized coordinates, the calculated ESP charges of the QM region (QM atoms and pseudobond atoms) were employed and transferred to the new topology files by AMBER's ParmEd module.[106] Transient non-standard residues dG–O^{-}(+3) and dT…OH(+4) forming during the phosphodiester bond cleavage at the TS, were initially parameterized by the R.E.D. Server,[107-110] while the missing bonded parameters were added by ANTECHAMBER.[111, 112] The non-standard residue dT–OH(+4), which forms after DNA cleavage, was parameterized using the R.E.D. server. In the next step, the LEaP module was employed to generate the coordinate and topology files of the TS and products for the MD simulations. Lastly, 10 ns of MD simulation with 100 kcal mol^{-1} Å^{-2} restraint on the QM atoms was performed at a temperature of 310 K via the NVT ensemble. All bonds involving hydrogen atoms were treated using SHAKE. Long-range Coulomb interactions were handled with the smooth particle mesh Ewald method using a 10 Å cutoff for non-bonded interactions. The CPPTRAJ module was used to analyze the RMSD and RMSF values of the MD simulations to monitor the stability of the TS and the product in the Matched and the MM5 systems throughout the simulation (**Figures S3** and **S4**). All the 2500 frames of these 10 ns of MD on the Matched and the MM5 products were also employed to calculate relative binding enthalpies via the MMGBSA method, as explained in the second approach of the "MM/GBSA calculations" section.

EDA implemented in an in-house Fortran90 program was employed to calculate the non-bonded inter-molecular interaction energies along the cleavage reaction path.[113-115] This analysis was performed on the MD-

simulated trajectories by considering the changes in Coulomb and Van der Waals interaction energies between the QM subsystem and the residues of the MM region for the reaction process. This difference in the non-bonded intermolecular interaction energy $\Delta\Delta E_{Intermol.\ Interact.}$ can be calculated as:

$$\boldsymbol{\Delta\Delta E_{Intermol.\ Interact.} = \Delta E^{TS/Product}_{Intermol.\ Interact.} - \Delta E^{Reactant}_{Intermol.\ Interact.}} \qquad \textbf{Eqn. 1}$$

where $\Delta E^{TS/Product}_{Intermol.\ Interact.}$ represents the difference between the non-bonded intermolecular interactions of the TS or product and $\Delta E^{Reactant}_{Intermol.\ Interact.}$ represents the same values for the reactant. This analysis, which can be applied to QM/MM optimized structures, or based on the MD-generated ensemble, gives a qualitative assessment of the catalytic role of residues surrounding the active site with stabilizing or destabilizing effects on the catalytic reaction.[116-122] The UCSF Chimera,[123] VMD,[124] and GaussView 6.1[125] programs were used for rendering the images.

RESULTS AND DISCUSSION

Matched and mismatched systems maintain stable conformations for HNH catalytic state. All-atom MD simulations in an aqueous solution were performed to obtain the initial conformation of the DNA and sgRNA-bound SpCas9 with catalytically active HNH domain for Matched and MM5 systems (in two replicates). Throughout the simulations, the distance between the nitrogen atom of H840 and the scissile phosphate (OP1–dT(+4)) was maintained between 5.61 Å and 5.65 Å for Matched and MM5, respectively. Hence, we considered this range of the MD simulation for further analysis, representing a suitable coordination geometry of the DNA substrate and the active site residues with Mg^{2+} ion in the HNH domain. The time-dependent root-mean-square deviation (RMSD) plots for the alpha carbon (Cα) atoms of the SpCas9 protein for the Matched and MM5 systems are shown in **Figure S5A**. The RMSD values converged within 50 ns for Matched and MM5, indicating that systems have reached a stable state. However, the RMSD of the backbone of the SpCas9 is slightly lower for the MM5 system than that of Matched, suggesting that the SpCas9 protein explores alternative dynamics and conformation in the presence of mismatched RNA:DNA pair. Moreover, we found that the introduction of PAM proximal mismatched DNA has a distinct effect on the flexibility in the different regions of the SpCas9•sgRNA•DNA complex, as depicted in **Figure S5B**. It can also be observed in this figure that various regions of SpCas9, i.e., REC-I, REC-III, HNH, RuvC, and CTD, have higher flexibility in MM5 than in the Matched system.

The PAM proximal mismatch (MM5) instigates conformational changes and domain motion alterations in the CRISPR-Cas9 system. The overall conformation of the SpCas9•sgRNA•DNA tertiary complex remains stable with a mismatch at the fifth position from the PAM in the MM5 system (see **Figure 2**). However, this mismatch in the DNA substrate induces several local structural changes in the SpCas9 and the nucleotides attached to it. For instance, as shown in **Figure 2B**, the RNA:DNA interactions of the mismatched and adjacent nucleotides are affected. Furthermore, the PAM distal end of the nt-DNA displays higher flexibility and loses interactions with the 3'-end of the t-DNA (**Figure 2C** & **D**, and **Figure S5B**). These differences partially explain the calculated binding affinity reduction of around 22% for the MM5 compared to the Matched when considering the complexation of DNA with the SpCas9-sgRNA binary complex (see **Figure S6**).

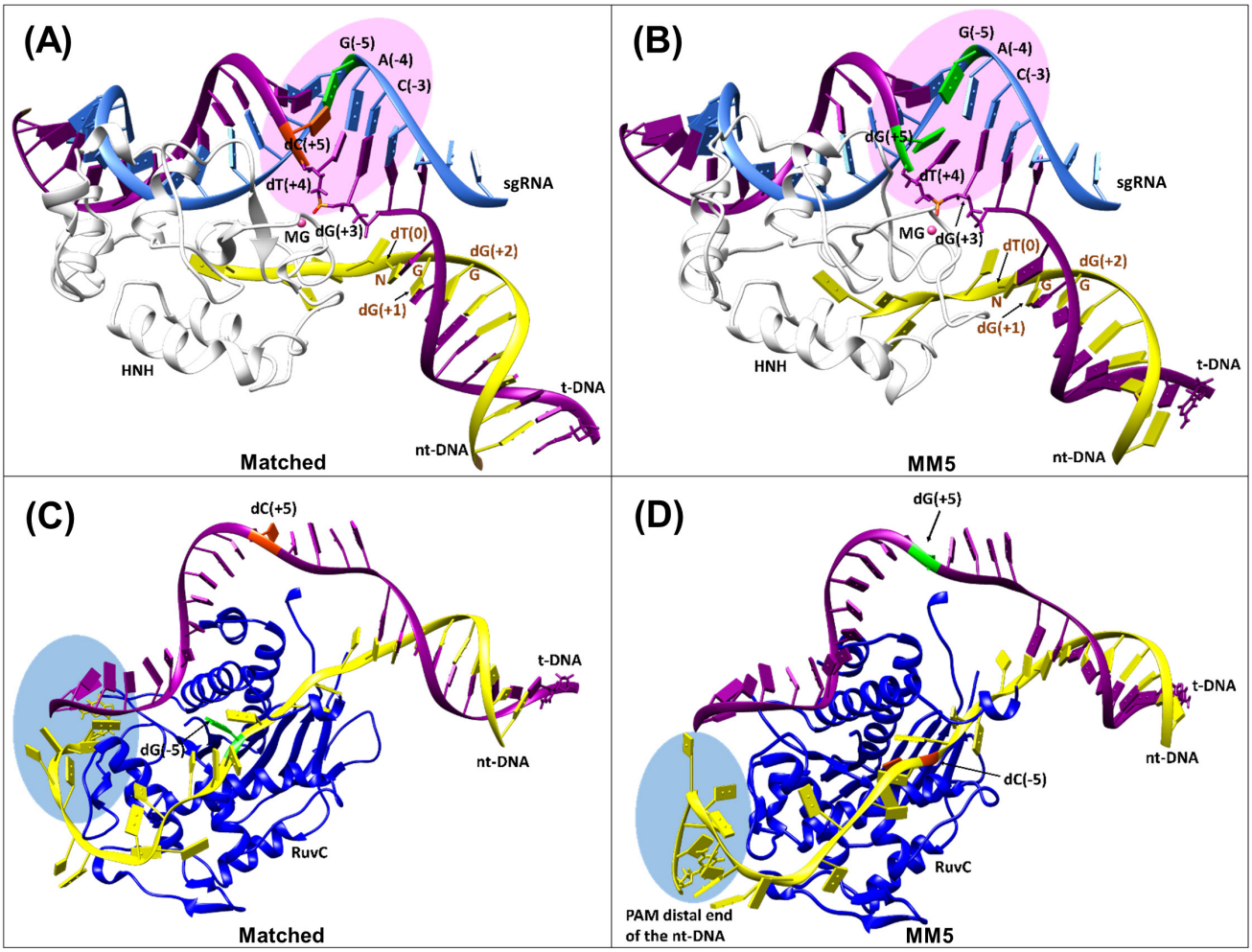


Figure 2. DNA, sgRNA, and protein interactions for **(A)** the Matched-SpCas9 and **(B)** MM5-SpCas9 focusing HNH catalytic site and PAM (NGG) region. **(C** & **D)** show different views of the Matched and MM5 zooming out the PAM distal end and RuvC region interactions. The t-DNA, nt-DNA, and sgRNA are colored magenta, yellow, and light blue, respectively. Two nuclease domains of SpCas9, HNH, and RuvC are shown in white and deep blue.

It has been observed that the association of DNA to the binary complex of Cas9 and gRNA is rate-limiting during the first catalytic turnover of Cas9, while DNA cleavage from a pre-formed ternary complex of SpCas9.sgRNA.DNA is rapid.[126] Besides, mutations designed to lower Cas9 off-target activity often found to be result in a decreased affinity for its target sequence (DNA) and reduced mutagenesis rates resulting the low cleavage efficiency issues.[127] Thus, DNA binding plays a critical part in the cleavage mechanism of SpCas9. Additionally, we observed large amplitude motions of the protein domains directly involved with the nucleic acids in MM5: the recognition region (REC-I) interacting with the stem of sgRNA and the C-terminal domain that binds the DNA.

A dynamic cross-correlation analysis has been performed to characterize the large-scale motions of the SpCas9 protein domains for the Matched and MM5, respectively (see **Figure S7**). Several deviations of the correlated motions of SpCas9 domains are observed in the MM5 upon the incorporated mismatch. The REC-II (167–307) and a part of the REC-III (450–500) domain's movements along the direction of the HNH and RuvC-III

domains (765–1099) in the Matched change into the opposite direction in the MM5. On the other hand, the REC-I (94–167 and 307–447) region's anti-correlated motion in the Matched exhibits somewhat correlated motion in the MM5 with these two nuclease domains. The HNH and RuvC-III regions show a positively correlated motion with a part of the CTD domain (1200–1368) in Matched, while it is changed to a negatively correlated motion in MM5. Conversely, two regions of the REC-III (300–400 and 600–700) domain display an increased paired motion with the same CTD region in the MM5, indicating a relative opening of the protein in the MM5, which could affect the nucleotides and protein binding. Thus, the mismatch affects the overall motion of the SpCas9.

The mismatch weakens the cleavage point at the HNH catalytic site conformations. Considering the most conducive orientations in the active site for the cleavage reaction, four clusters (10 representatives) from the Matched system and one cluster (3 representatives) from the MM5 system were obtained (see **Tables S1** & **S2** and **Figures S8** & **S9**). In the case of Matched system, when one of the coordinated waters to the Mg^{2+} (termed first-shell water) considered being the nucleophile, the orientations of the active site are relatively suitable in three clusters, including around 60% of the 100,000 simulated snapshots. Therefore, seven representatives of these three clusters, in which the catalytic water is also hydrogen-bonded to H840 (Matched-1 to Matched-7 in **Figure S8**), were used for further QM/MM calculations. In comparison, when non-coordinated water around the Mg^{+2} (termed second-shell water) is in a reasonable distance and orientation toward the H840 and the phosphate group, three representatives of the fourth cluster with a population abundance of 16.7% were considered for further QM/MM studies (Matched-8 to Matched-10 in **Figure S8**).

Contrary to the observed trend for the Matched system, about 13% of the clustered structures for the MM5 show a rotation of H840 that hinders its catalytic competence as the generalized base to activate the nucleophile. In addition, among the remaining 87%, only 16% (cluster 1) maintained catalytically conducive orientations, while even among the three representatives of this cluster, just one structure displays a reasonable O3'–P…O_w angle (see **Table S2** and **Figure S9**). Furthermore, based on the detailed results in **Table S2**, the first-shell water was the only potential nucleophile in the MM5 structure. All the other representatives with the second-shell water are either too far from H840 and phosphorus or the O3'–P…O_w angle in the active site is unsuitable for an S_N2-like reaction. Considering the clustering results, more than 72% of the simulated trajectories of the Matched favor the HNH active site conformation, leading to the catalytic cleavage of the t-DNA between the third and fourth nucleotides from the PAM region. In comparison, only in 5% of the MM5 simulated trajectories, the orientations of the residues of the HNH active site can lead to the cleavage reaction. This indicates a reduction of the precise and efficient cleavage of the t-DNA by mismatch containing MM5 compared to its native matched form.

Conformation of the reactants for the Matched and MM5 systems. Based on the clustering analysis results, ten representatives of the Matched shown in **Figure S8** were selected for further hybrid QM/MM studies. Since representatives with either the first- or the second-shell water were chosen from the clustering analysis; thus, two sets of structures were considered separately to be optimized. A summary of the clustering analysis for the selected representatives and the calculated relative optimization energies are listed in **Table S3**. As shown in the table, Matched-4 and Matched-8 are the most stable structures of the first- and second-shell water reactants,

which termed Matched$^{1st\ shell}$ and Matched$^{2nd\ shell}$ for the rest of the paper. In the case of MM5, since only one representative (MM5-1) had reasonable orientations in the active site, this structure was optimized and used for designing the product (termed MM5 for the rest of the paper). The active sites for the optimized structures of the Matched$^{1st\ shell}$, Matched$^{2nd\ shell}$, and MM5 active sites are shown in **Figure 3**. None of the structures from our MD simulations for either the Matched or MM5 systems show the involvement of K848 for the active site conformation of the HNH nuclease domain (**Figures S10**–**S12**). This is in contrast to previously proposed computational models based on an inactive crystal structure.[54, 55]

Based on the position of the nucleophilic water in the selected representative structures, two pathways are considered for the DNA cleavage mechanism at the HNH domain via an S_N2-like reaction. In the first pathway shown in **Figure 1A**, predominantly seen for the Matched$^{1st\ shell}$ and the MM5, the first-shell water plays the role of the nucleophile. As a result, proton transfer occurs from the water to H840, and the resulting OH^- attacks the phosphorus with concomitant cleavage of the P–O3' bond of the dG(+3). In contrast, as shown in **Figure 1B**, the second-shell water between the phosphate bridge and the H840 undergoes the proton transfer and performs the cleavage reaction in the second pathway.

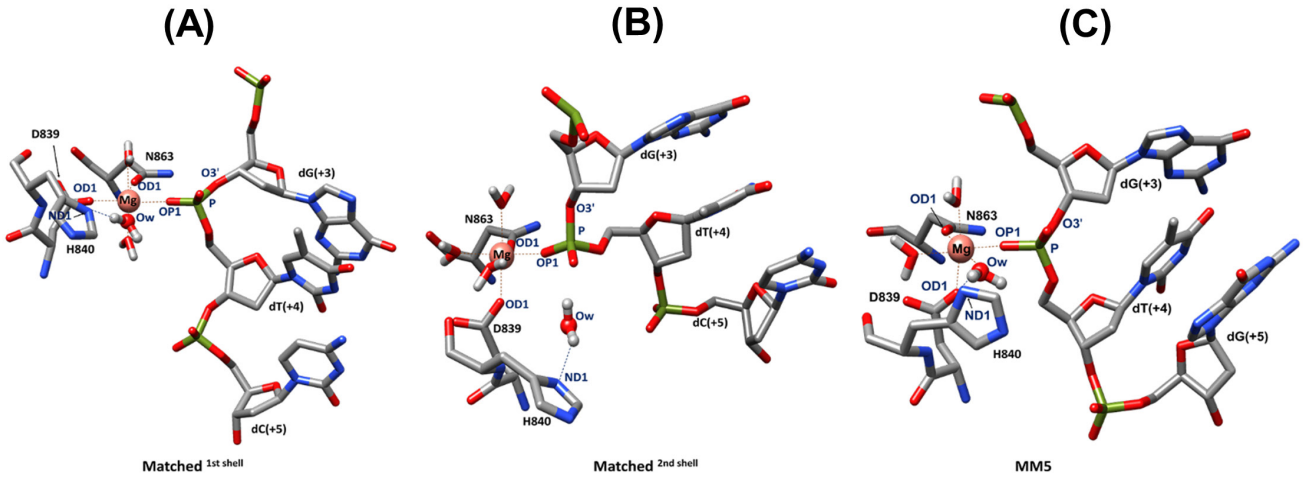


Figure 3. Optimized active site geometries for **(A)** Matched$^{1st\ shell}$, **(B)** Matched$^{2nd\ shell}$, and **(C)** MM5.

The t-DNA hydrolysis by the Matched system. As mentioned above, two reaction mechanisms for the hydrolysis of t-DNA by the HNH domain of endonuclease SpCas9 in the Matched system have been considered: (1) metal-bound water/first-shell water-mediated pathway or (2) second-shell water-catalyzed pathway.

(1) ***Metal-bound water/first-shell water-mediated pathway (M1 pathway).*** As shown in the reactant ($\mathbf{R_{M1}}$) of this pathway in **Figure 4A**, one phosphoryl oxygen (OP1) atom of dT(+4) is bound to the Mg^{2+} ion (Mg…OP1 = 2.05 Å, **Table S4**), while the other phosphoryl oxygen (OP2) interacts with Q844 through a hydrogen bond. This metal-substrate (t-DNA) coordination activates the scissile P–O3' bond of dG(+3) compared to the P–O5' bond of dT(+4) (P–O3' = 1.66 Å and P–O5' = 1.60 c). The positive charge of the magnesium ion (1.96 *e*, **Table S4**) plays a vital role in activating the P–O3' bond. Additionally, this coordination mode helps polarize the scissile phosphodiester bond's P atom (1.42 *e*). In $\mathbf{R_{M1}}$, the base residue H840 is hydrogen bonded to an Mg-bound water

molecule, H_WO_WH (Mg–O_W = 2.07 Å and H_W–Nδ = 1.90 Å). The catalytic site residues (D839 and N863) and two additional water molecules complete the octahedral coordination geometry around the Mg^{2+} ion. In the TS ($\mathbf{TS_{M1}}$), the H_W proton of the catalytic water (H_WO_WH) transfers to H840, and the resulting nucleophile O_WH^- attacks the electrophilic P atom (1.25 *e*) of the dT(+4) concomitantly, elongating the P–O3' bond. The TS structure shows key reacting distances that suggest a concerted mechanism for this step (H_W–Nδ = 1.38 Å, O_W–P = 2.47 Å and P–O3' = 2.45 Å in **Figure 4A** and **Table S4**).

As shown in **Figure 4B**, the energy barrier and the approximate activation free energy for this process are 14.3 and 16.0 kcal mol^{-1}, respectively, consistent with the estimated activation barriers from experimental data and previously reported simulations. Several experimental and computational studies have been reported on the catalytic mechanism of SpCas9.[20, 54, 55, 61, 128-133] Sue and coworkers employed various kinetic techniques and successfully characterized each major step of the CRISPR/Cas9 mechanism.[126] They showed that the DNA cleavage (chemistry step) from a pre-formed ternary complex (SpCas9•sgRNA•DNA) to form DNA products is fast ($K_{chem} \geq 700\ s^{-1}$). The estimated free energy barrier (ΔG‡) based on Eyring's TST for the cleavage reaction is ~14.1 kcal mol^{-1}. Taylor and coworkers[134] and Singh et al.[132] measured a k_{cat} of 4.3 s^{-1} (corresponding to ΔG‡ of ~16–17 kcal mol^{-1}) for the HNH catalyzed hydrolysis step.

In recent work based on the same catalytically-active structure of SpCas9 employed in the present study, Palermo and coworkers[56] conducted calculations to investigate the reaction mechanism associated with the chemical step involved in the catalytic activity of the HNH domain, with calculated free energy barriers of ~16–18 kcal mol^{-1}. Previous computational studies based on the inactive crystal structure also reported energy barriers of 21.0 and 17.8 kcal mol^{-1} for the first-[54] and second-shell[55] water mechanisms, respectively, which align with the experimental values. It is worth mentioning that this previous study found that achieving catalytic activation in the inactive structure necessitates a significant conformational change. Specifically, the movement of K848 or another positively charged group from a considerable distance towards the scissile phosphate is required. This conformational change results in a shift in the position of the Mg^{2+} ion and significantly lowers the activation barrier for the catalytic reaction.

Upon completion of the cleavage reaction, the P–O3' phosphodiester bond is cleaved to generate the product ($\mathbf{P_{M1}}$), resulting in the separation of the t-DNA into two segments. As shown in **Figure 4A**, the octahedral geometry around the Mg^{2+} ion changes to trigonal bipyramidal during the cleavage reaction ($\mathbf{R_{M1}}$ to $\mathbf{P_{M1}}$), and its coordination number changes from six to five. As shown in **Figure 4B**, the formation of $\mathbf{P_{M1}}$ is exergonic by 15.4(13.7) kcal mol^{-1} from $\mathbf{R_{M1}}$. Sue and coworkers also showed that the process of the DNA product release is the slowest step during the multiple-turnovers ($t_{1/2}$ ~ 43–91 h), which makes it a single-turnover nuclease.[126] Other studies also have shown that Cas9 has a long cutting half-life and catalytic lifetime, which is less efficient than other nucleases like restriction enzymes.[128, 134, 135] As a result, the rate of DNA cleavage is constrained by the time required Cas9 to detach from its DNA-substrate and revisit the population of target sites within a cell.

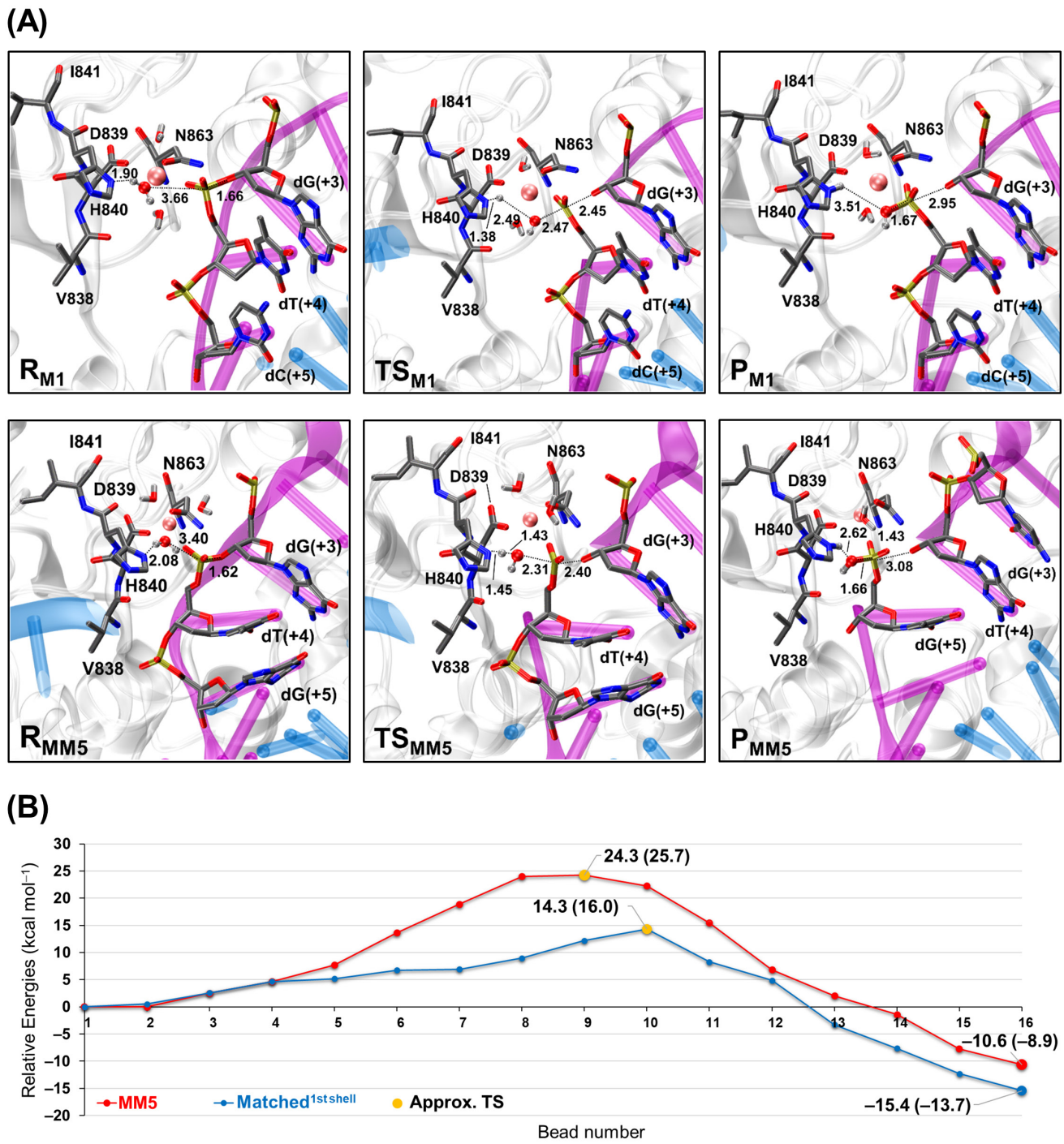


Figure 4. (A) Optimized geometries of reactant (R), transition state (TS), and the product (P) of the cleavage reaction at the HNH domain of Matched$^{1st\ shell}$ (top) and MM5 (middle). The nucleophilic water and the Mg^{2+} are shown in ball-and-sticks, while all the other atoms are in sticks. The chain of the protein, t-DNA, and sgRNA are shown in white, purple, and blue ribbons, respectively. **(B)** Optimized minimum energy path for the cleavage reaction at the HNH domain of the Matched$^{1st\ shell}$ and MM5 systems. Values for critical points correspond to potential (Gibbs free) energies.

(2) ***Second-shell water-assisted pathway (M2 pathway).*** As shown in **Figure 1A**, the major difference in this mechanism is that the base residue H840 creates a nucleophile by activating an external water molecule that is not bound to the Mg^{2+} ion. In the optimized reactant (**R_{M2}**), the P–O3' bond is 0.18 *e* less activated than in **R_{M1}**

due to the low Lewis acidity of the Mg^{2+} ion in this configuration (see **Table S4** and **Figure S13**). In addition, the charge on the HO^-_W nucleophile of the second-shell water is 0.35 *e* lower than that of the metal-bound water in the previous pathway, resulting in a relatively weak nucleophile. Moreover, the charge of the P atom is reduced by 0.25 *e* compared to that of $\mathbf{R_{M1}}$. Although the O_W…P–O3' angle in $\mathbf{R_{M2}}$ (~163°) is closer to the desired angle for an S_N2-type reaction than that of $\mathbf{R_{M1}}$ (~151°), there seems to be a competition among the H840 and the free phosphoryl oxygen (OP2) of the t-DNA substrate to abstract a proton from the nucleophile water (H_WO_WH). This is supported by an additional strong hydrogen bond (1.80 Å) between the H atom of the nucleophile water and OP2 in $\mathbf{R_{M2}}$. In addition, this water molecule's orientation is not favorable for the nucleophilic attack on the P atom of the substrate. The optimized product ($\mathbf{P_{M2}}$) is endergonic by 32.6 kcal mol^{-1} from $\mathbf{R_{M2}}$, indicating the unfavorable nature of this mechanism. Thus, our calculations suggest that the second-shell water molecule is a weaker nucleophile than the metal-bound water for this reaction. This is also seen in previous studies related to phosphodiester bond hydrolysis reactions by single metal-containing nucleases.[136, 137]

The t-DNA hydrolysis by MM5 system. As mentioned previously, only one of the extracted representative structures provides a suitable active site configuration for the t-DNA cleavage reaction by the HNH catalytic site of the MM5. This structure possessed a water molecule bound to Mg^{2+} ion and hydrogen bonded to H840 that can be used as the potential nucleophile for the hydrolysis reaction. Thus, we investigated the M1 pathway for MM5 to understand the structural and mechanistic details involved to implement our findings to mitigate the knowledge gap between mismatch sensitivity and specificity of SpCas9.

The optimized reactant of the MM5 ($\mathbf{R_{MM5}}$ in **Figure 4A**) is different from the optimized reactant of the $Matched^{1st\ shell}$ ($\mathbf{R_{M1}}$ in **Figure 4A**) due to the position of a water molecule (WAT2) bound to the Mg^{2+} ion (see **Figure S14**). A reduction of 0.12 *e* charge for the Mg^{2+} ion in $\mathbf{R_{MM5}}$, along with a decrease of 0.59 *e* for the P–O3' bond compared with the Matched system ($\mathbf{R_{M1}}$), are observed. The nucleophile HO^-_W of $\mathbf{R_{MM5}}$ has a reduced charge of 0.34 *e*, and the P atom of the scissile phosphodiester bond shows a reduction of 0.66 *e* compared to $\mathbf{R_{M1}}$. Moreover, the calculated Ow…P–O3' angle in the $\mathbf{R_{MM5}}$ system is ~141°, which is smaller than the expected 180° for an S_N2 attack. Additionally, the oxygen of the nucleophilic water does not face the phosphorus in a catalytically conducive orientation, and the H_W–O_W…P angle is unfavorable (~40°), resulting in a significant rotation required by the water in the reactant to reach a catalytically competent orientation. These differences may help explain (at least in part) the higher activation barrier for the MM5 ($\mathbf{TS_{MM5}}$ = 24.3 kcal mol^{-1}). In the approximate TS, the breaking and forming bond distances (O_W–P = 2.31 Å and P–O3' = 2.40 Å, **Table S4**) display a concerted (S_N2-like) dissociative pathway[138] where a slightly more bond cleavage to the leaving group than bond formation to the nucleophile is observed (P–O3' is 0.09 Å longer than O_W–P).

The Matched system follows a concerted pathway[61, 139, 140] with a similar extent of partial bond formation to the nucleophilic oxygen and partial bond cleavage to the leaving group at the transition state, $\mathbf{TS_{M1}}$ (O_W–P = 2.47 Å and P–O3' = 2.45 Å in **Table S4**). Unlike the Matched system (M1 path), Mg^{2+} loosely binds to the nucleophile water (Mg–O_W = 2.17 Å, in $\mathbf{R_{MM5}}$), indicating that it does not act as a suitable Lewis acid in the MM5. The tightness of the transition state in the mechanisms of phosphoester hydrolysis reactions, described in terms of the O_W–P

(nucleophile) and P–O3' (leaving group) bond distances, decreases from mono- to triesters.[140] This value calculated for the Matched and MM5, decreases from the Matched to MM5 (4.92 Å *vs.* 4.71 Å). The sum of the O_W–P and P−O3' distances illustrates the hydrolytic reaction progression. Comparison of these distances for the Matched system in R_{M1} and TS_{M1} indicates a significant increase (0.40 Å) in tightness from 5.32 Å to 4.92 Å.

Conversely, the increment of tightness is only 0.31 Å in the MM5 case (**R_{MM5}** and **TS_{MM5}**), indicating a relatively low reaction progression, which is also consistent with the calculated higher activation barrier for the MM5 compared to the Matched system. Moreover, in one of our recent papers,[118] kinetic rates of the DNA cleavage reaction for a similar system have been calculated using a kinetic model designed for plasmid DNA cleavages. The relative cleavage rates for the Matched and MM5 DNA were 1.23 ± 0.13 min^{-1} and 0.68 ± 0.09 min^{-1}, respectively. Since these rates involve the entire kinetic process up to the cleavage step, a direct comparison with our values is not possible. Based on the QM/MM energies of the optimized reactant and product in **Figure 4B**, the cleavage reaction catalyzed by the Matched$^{1st\ shell}$ system is exoergic with a reaction energy of –15.4 kcal mol^{-1}, compared with –10.6 and 32.6 kcal mol^{-1} for MM5 and Matched$^{2nd\ shell}$, respectively.

Figures 5A and **5C** show that the nucleophilic water in the Matched$^{1st\ shell}$ and MM5 structures have strong, attractive interactions with Mg^{2+} (in the NCI scale) and show hydrogen bonds with H840, although for MM5, the color of the surfaces between the nucleophilic water and H840 indicate a weaker H_W…Nδ hydrogen bond interaction. It also can be seen that the second hydrogen of the nucleophilic water in the Matched$^{1st\ shell}$ has a strong hydrogen bond with an adjacent water. In contrast, the second hydrogen of the nucleophilic water in MM5 has weak hydrogen bond interaction with the OP1 of dT(+4). The NCI plot for the Matched$^{2nd\ shell}$ in **Figure 5B** shows that the nucleophilic water forms a hydrogen bond with the ζ-hydrogen of K862, and OP1 of dT(+4), while displaying weak interactions with H840. The calculated ESP charges in **Table S4** also show that the nucleophilic water in the Matched$^{1st\ shell}$ is more polarized than the nucleophilic water in the Matched$^{2nd\ shell}$ and MM5, facilitating the proton transfer from the water to H840 in the Matched$^{1st\ shell}$ system.

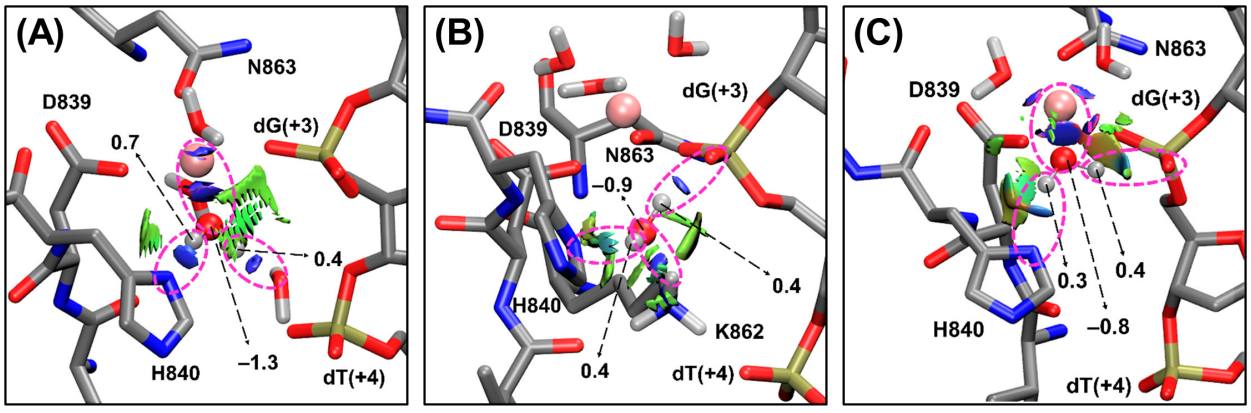


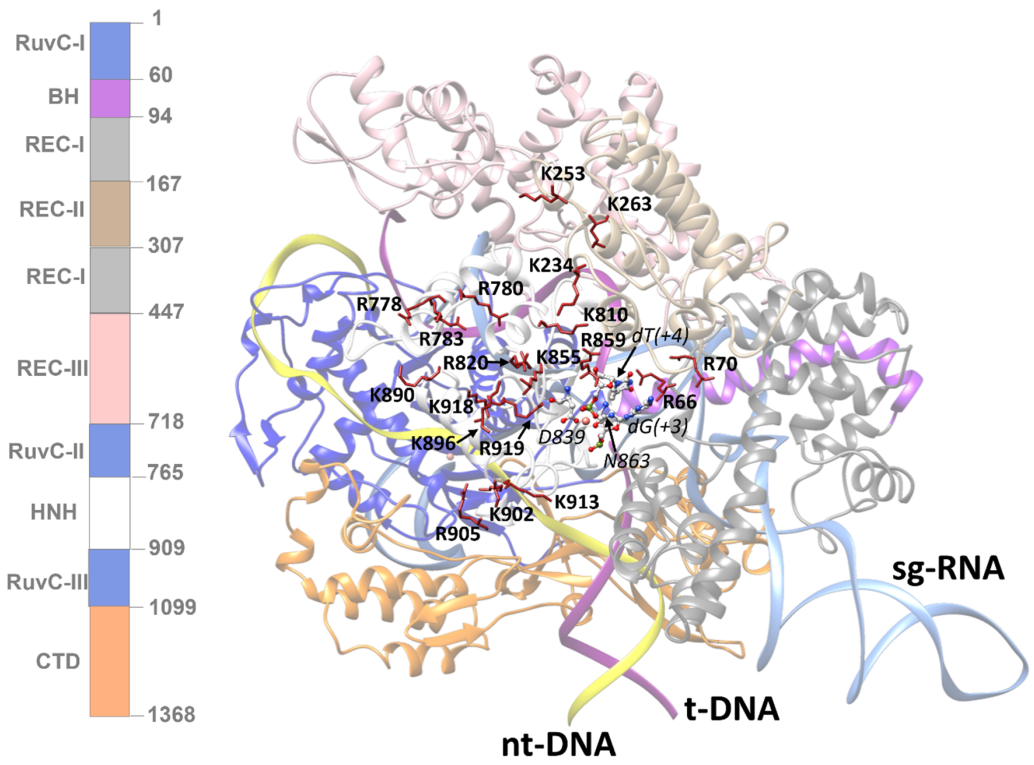
Figure 5. Calculated ESP charges for the nucleophilic water and NCI index surfaces between the nucleophilic water and the surrounding residues in the reactants of **(A)** Matched$^{1st\ shell}$, **(B)** Matched$^{2nd\ shell}$, and **(C)** the MM5. The ESP charges are extracted from the optimized structures of the reactants at ωB97X-D/6–31G(d,p) level of theory with AMBER ff14SB Force Field. The nucleophilic water and the Mg^{2+} are shown in ball-and-sticks, while all the other atoms are in licorice. Hydrogen atoms of the amino acids and the nucleotides are not presented for more clarity except for the ζ-hydrogens of K862 in (B).

Energy Decomposition Analysis (EDA) reveals SpCas9 residues involved in t-DNA mismatch selectivity. EDA was performed on the reactant, TS, and product of the Matched$^{1st\text{-}shell}$ (termed Matched for the rest of the EDA section) and MM5 to study the non-bonded intermolecular interactions (Coulomb and van der Waals) between the SpCas9 and the residues of the active site. EDA is a qualitative tool that can offer useful information on the role of the enzyme's individual residues during the chemical step. Calculated intermolecular interaction energy differences, $\Delta\Delta E_{Intermol.\ Interact.}$ for the protein and nucleic acid residue between the reactant and TS of the Matched and MM5 are –294 and –2 kcal mol^{-1}, respectively (see **Eqn. 1**). These results suggest that the protein environment significantly stabilizes the transition state of the Matched system compared to MM5. The corresponding values between the product and the reactant of the Matched and MM5 are –346 and –57 kcal mol^{-1}, respectively, suggesting that the protein environment also favors the product of the Matched system compared to MM5. Corresponding graphs showing individual values for all residues between the reactant/TS and reactant/product pairs are given in **Figures S15** and **S16**, respectively.

We were also interested in comparing the stability of the reactant and product of the Matched system to that of MM5, which provides valuable insights into residues with significant stabilizing or destabilizing effects on these structures. The calculated $\Delta\Delta E_{Intermol.\ Interact.}$ between the MM5 and Matched reactants is 246 kcal mol^{-1}, while the same value for the products is 175 kcal mol^{-1} (detailed results in **Figures S17** and **S18**). This suggests that the MM5 system is destabilized compared to the Matched system and this destabilizing effect is more significant in the reactant than the product. In fact, SpCas9 helps stabilize the product of the reaction during the catalytic reaction by the MM5, but this stabilization is less than that in the Matched case. Decomposition of the free enthalpy contributions to the binding enthalpies of the Matched and MM5 systems on a per-residue basis was also performed to study the binding affinities between the active site and the rest of the system via the MM/GBSA approach. Our calculations show that the binding affinities in the reactant of the Matched are higher than the MM5. The average values of ΔH_{total} are ~ –161 and –143 kcal mol^{-1} for the reactants of the Matched and the MM5, respectively (detailed values in **Table S5**).

Several residues that show differential effects on the Matched and MM5 systems were identified (**Figure 6** and **Table S6**), some of which have been previously recognized.[133, 141-143] For instance, the high-fidelity SpCas9 variants (SpCas9-HF1 to SpCas9-HF4) identified by Joung and coworkers[144] contain a mutation at residue R661 (R to A), which is also one of the identified residues (extended values in **Table S6**). Slaymaker et al.[141] employed a structure-guided engineering approach on SpCas9 to improve its DNA targeting specificity. Three high-fidelity variants of SpCas9 (K855A), (K810A/K1003A/R1060A, eSpCas9 1.0), and (K848A/K1003A/R1060A, eSpCas9 1.1) were identified after a comprehensive mutational study focusing on PAM distal mismatches. The two top residues (K855 and K810) found by our EDA method using MM5 were also seen in these variants proposed by this group. A recent study by Liu and coworkers[145] proposed two SpCas9 variants (HSC 1.1 and HSC 1.2) with enhanced specificity using a structure-guided engineering method. The K1246 residue found from our EDA method was also seen in the HSC 1.1 variant. R691A (HiFi Cas9),[146] K526E, R661Q (evoCas9),[147] and K890N (sniper Cas9)[148] are some of the other residues mentioned in previous studies, which are also observed in our energy decomposition analysis as listed in **Table S6**.

(A)



(B)

Cas9 domain	Residue	$E^{TS}_{Intermol.Interact.} - E^{Reactant}_{Intermol.Interact.}$ (Kinetics of the reaction)		$E^{Product}_{Intermol.Interact.} - E^{Reactant}_{Intermol.Interact.}$ (Thermodynamics of the reaction)	
		MM5	Matched	MM5	Matched
BH	R66[149]	–9.3	2.3	–9.6	1.5
	R70[149]	–10.4	1.8	–9.6	1.4
REC-II	K234	–5.4	1.1	–5.7	2.8
	K253	–4.4	3.0	–4.4	2.5
	K263	–7.3	2.9	–4.1	1.9
HNH	R778	–3.6	3.6	–3.9	3.5
	R780[141]	–3.9	7.3	–4.1	7.2
	R783[141]	–3.6	4.7	–4.0	4.5
	K810[141]	–8.5	14.2	–11.3	14.4
	R820	–5.6	2.0	–6.0	2.1
	R832[142]	–6.2	1.7	–6.5	1.7
	K855[141]	–7.6	10.2	–3.1	17.9
	R859[141]	–8.0	1.3	–8.3	1.3
	K890[148]	–3.8	3.2	–4.3	3.3
	K896	–4.1	12.8	–5.5	11.8
	K902[143]	–6.7	9.3	–7.9	9.4
	R905[143]	–5.4	3.1	–8.9	1.1
RuvC-III (L2 Loop)	K913	–7.9	2.4	–10.2	4.8
	K918	–2.7	8.5	–3.7	8.2
	R919[141]	–4.1	6.9	–5.0	7.0
RuvC-III	R925[143]	–3.8	5.3	–4.1	5.2
	K929	–3.9	2.2	–4.1	2.7

Figure 6. (A) Residues with intermolecular interaction effects proposed by the EDA calculations. Residues are shown in red licorice with corresponding residue names and numbers in bold text. The active site's residues are displayed in ball-and-stick, and the residue names and numbers are in italic text. The hydrogen atoms are not shown for clarity. **(B)** List of residues with

significant change in inter-molecular interaction energies between the Matched and MM5 systems. The threshold for the selection is $\Delta\Delta E_{Intermol.\ Interact.} \geq |5|$ kcal mol^{-1}.

A detailed analysis of the interactions between the HNH active site (including the fifth residue from PAM) and some of the residues identified from the EDA revealed an interesting finding about the stabilization of the MM5 system by those residues. A hydrogen bond between the free phosphoryl oxygen of dG(+5) and the backbone of V838 was found to be pivotal for keeping all these residues connected to the active site through a network of hydrogen bonds for MM5, while it was absent in the Matched system (see **Figure 7**, **Figures S19**–**S22**). Especially, in the case of R780 in **Figure 7**, in addition to the hydrogen bond between dG(+5) and V838, the interaction between D809 and R780 is critical to maintaining the stabilization of MM5 by R780. The same residue mutation to Alanine (R780A) has been shown to work well towards off-target containing CRISPR-Cas9 complexes in the literature.[141] However, its single mutations or combination with other residue mutations has not been studied in detail.

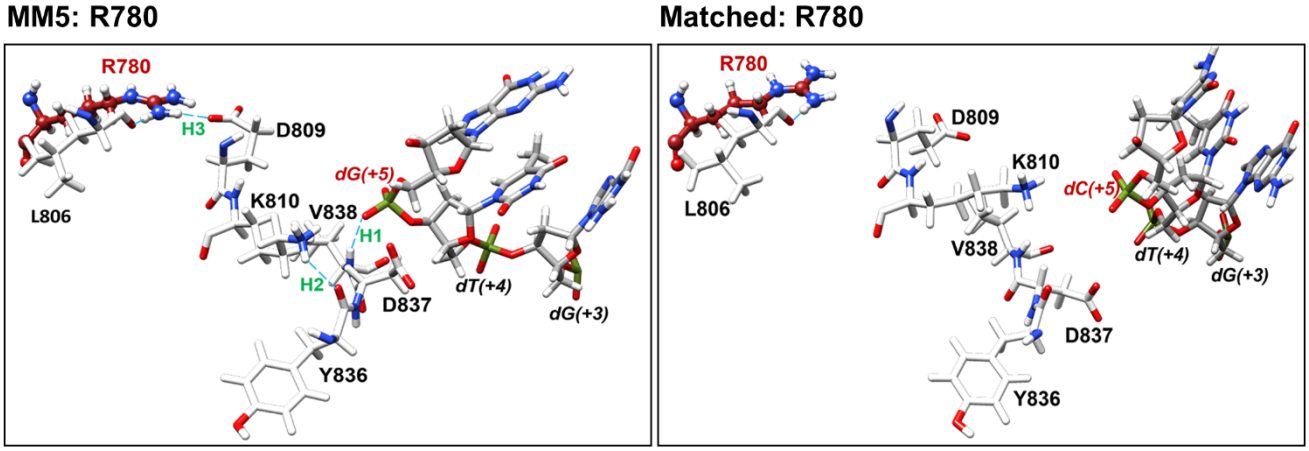


Figure 7. Residue R780 (found from EDA analysis) shows interactions with the mismatched region of t-DNA and corresponding interaction in the Matched system.

Our per-residue contribution analysis indicates that multiple residues in SpCas9 stabilize the HNH catalytic site in both the Matched and MM5 systems, although the stabilizing effect is greater in MM5 (see **Table S7**). For instance, most residues in the BH region (R63, R69, R74, R75, K76, R78, and K92) have a larger stabilizing contribution in the MM5 system compared to the Matched. By comparing interactions of SpCas9 with t-DNA and sgRNA in the Matched and MM5 systems, we observed that these residues stabilize the t-DNA-sgRNA hybrid in both cases (see **Figure S23**). These results uncover sites that could be potential candidates for mutations to explore effects on the off-target removal. Charpentier and coworkers[149] also have shown that the bridge helix is essential for R-loop formation and that R63, R66, and R70 reduce SpCas9 specificity by stabilizing the R-loop in the presence of mismatches. Thus, mutations of these residues would destabilize mismatch-containing systems as another approach for the off-target effect removal. Interestingly, the EDA results reveal that R66 and R70 have a significant stabilizing effect on the MM5, while simultaneously causing destabilizing effects on the Matched (see **Figure 6B**). In addition, our findings are also consistent with other studies focusing on these BH's residues mutations in the literature regarding SpCas9 specificity.[149, 150]

The HypaCas9 variant proposed by Chen et al.[36] involves four amino acid substitutions (N692A/M694A/Q695A/H698A) located on the PAM distal REC-III domain of SpCas9. They suggest that the mutation of residues within REC-III involved in RNA–DNA heteroduplex recognition, such as those mutated in HypaCas9 or SpCas9-HF1, prevents transitions by the REC-II domain. This more tightly traps the HNH domain in the conformational checkpoint in the presence of mismatches. Our EDA approach also revealed several other residues in the REC-III domain (**Tables S6** & **S7**), which would be interesting to study further related to their hypothesis. Although the residue's selection is based on the PAM proximal single mismatch MM5, the mentioned studies support our method and the possible activity of these residues' mutations toward other mismatch-containing (especially PAM distal mismatches) systems as well. However, further studies are needed to confirm the activities of these mutations concerning their effects on the SpCas9 specificity.

CONCLUSIONS

We have used classical molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) simulations to study the catalytic cleavage reaction of the t-DNA at the HNH domain of the SpCas9 using a recently discovered catalytic-active structure of this enzyme in complex with sgRNA and DNA. Based on the MD results, the second coordination shell water could also be considered the nucleophile in addition to the metal-bound water. To better understand the impact of sgRNA and t-DNA complementarity on the catalysis process, we also designed a mismatched structure (MM5) with a C to G mismatch at the fifth position from the t-DNA's PAM region. Calculated QM/MM results show that the nucleophilic attack by a second coordination shell water with the reaction energy of 32.6 kcal mol^{-1} is not energetically feasible. Calculated reaction energies for the Matched and MM5 systems with the attacking water bound to the Mg^{2+} ion (termed fist-shell water) are –15.4 and –10.6 kcal mol^{-1}, respectively, suggesting a structural effect of the t-DNA mismatch on the catalytic function of SpCas9. The calculated energy barriers for the cleavage reaction by the Matched and the MM5 systems are 14.3 and 24.3 kcal mol^{-1}, respectively. Additionally, the ESP charges of the attacking water and its non-covalent interactions with the active site residues show that the reactant of the Matched is more favorable than the MM5. Combined with the QM/MM energy barriers and reaction energies for the Matched and MM5, results of the energy decomposition analysis (EDA) show that the non-bonded intermolecular interactions between the SpCas9 and the residues of the active site in the TS and product of the Matched are considerably more stabilizing than the MM5. This shows that the amino acid residues of the SpCas9 have stabilizing contributions to the reactant–TS and reactant–product pathways of both systems. Still, this facilitating contribution is significantly larger for the Matched structure. Our EDA results also suggest that residues R66, R70, K253, K263, R780, R783, K810, R832, K855, R859, K890, K896, K902, R905, and K913 can be good targets for the mutation. Ten of these residues, namely R66, R70, R780, R783, K810, R832, K855, R859, K902, and R905, have been studied individually or in combination with other residues. Taken together, our results suggest that K253 and K263 in the REC-II, R820, and K896 in the HNH, and K913 and K918 in the RuvC-III region may be promising candidates for further computational/experimental investigation.

DATA AND SOFTWARE AVAILABILITY

All simulations and analyses employed via third-party software are described and referenced in the Computational Methods section. The EDA and LICHEM software programs are available at the Cisneros Research Group GitHub: https://github.com/CisnerosResearch/AMBER-EDA and https://github.com/CisnerosResearch/LICHEM. Additional ESI (ESI.zip) includes initial coordinates of all the studied systems.

ASSOCIATED CONTENT

Supporting information

The Supporting Information is available free of charge at.

Additional details of molecular dynamics, binding enthalpies, *k*-means clustering, QM/MM path optimization, ESP charges, and energy decomposition analysis (PDF).

Additional ESI (ESI.zip) includes initial coordinates of all the studied systems, supplementary movie for the HNH reaction path for Matched and MM5, and the negative imaginary frequency for the approximate TSs of the Matched and MM5.

ACKNOWLEDGMENTS

This work was supported by R15HL147265 and R21GM144860 to JL, and by R01GM108583 to GAC. We acknowledge the use of computational resources provided by the High-Performance Computing Center at the University of North Texas (Talon3) and The University of North Texas CASCaM high-performance clusters, supported by NSF grants CHE-1531468 and OAC-2117247.

CONFLICTS

The authors declare no conflicts of interest.

REFERENCES

1. Makarova, K. S.; Grishin, N. V.; Shabalina, S. A.; Wolf, Y. I.; Koonin, E. V., A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **2006**, 1, 1-26.

2. Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S., Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014**, 343, 1247997.

3. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A., Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **1987**, 169, 5429-5433.

4. Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S. D., Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology-SGM* **2005**, 151, 2551-2561.

5. Brouns, S. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J.; Snijders, A. P.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; Van Der Oost, J., Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **2008**, 321, 960-964.

6. Mojica, F. J.; Díez-Villaseñor, C.; García-Martínez, J.; Almendros, C., Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology-SGM* **2009**, 155, 733-740.

7. Terns, M. P.; Terns, R. M., CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **2011**, 14, 321-327.

8. Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V., Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109, E2579-E2586.

9. Koonin, E. V.; Makarova, K. S., Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2019**, 374, 20180087.

10. Mojica, F. J. M.; Montoliu, L., On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals. *Trends Microbiol.* **2016**, 24, 811-820.

11. Ding, Y.; Li, H.; Chen, L. L.; Xie, K., Recent advances in genome editing using CRISPR/Cas9. *Front. Plant Sci. (New Haven, CT, U. S.)* **2016**, 7, 703.

12. Wang, F.; Wang, L.; Zou, X.; Duan, S.; Li, Z.; Deng, Z.; Luo, J.; Lee, S. Y.; Chen, S., Advances in CRISPR-Cas systems for RNA targeting, tracking and editing. *Biotechnol. Adv.* **2019**, 37, 708-729.

13. Pickar-Oliver, A.; Gersbach, C. A., The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **2019**, 20, 490-507.

14. Yip, B. H., Recent advances in CRISPR/Cas9 delivery strategies. *Biomolecules* **2020**, 10, 839.

15. Wei, T.; Cheng, Q.; Farbiak, L.; Anderson, D. G.; Langer, R.; Siegwart, D. J., Delivery of tissue-targeted scalpels: opportunities and challenges for in vivo CRISPR/Cas-based genome editing. *ACS Nano* **2020**, 14, 9243-9262.

16. Goell, J. H.; Hilton, I. B., CRISPR/Cas-Based Epigenome Editing: Advances, Applications, and Clinical Utility. *Trends Biotechnol.* **2021**, 39, 678-691.

17. Rao, M. J.; Wang, L., CRISPR/Cas9 technology for improving agronomic traits and future prospective in agriculture. *Planta* **2021**, 254, 1-16.

18. Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Terns, R. M.; Terns, M. P.; White, M. F.; Yakunin, A. F.; Garrett, R. A.; van der Oost, J.; Backofen, R.; Koonin, E. V., An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **2015**, 13, 722-736.

19. Andersson, A. F.; Banfield, J. F., Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **2008**, 320, 1047-1050.

20. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, 337, 816-821.

21. Koonin, E. V.; Makarova, K. S.; Zhang, F., Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **2017**, 37, 67-78.

22. Makarova, K. S.; Wolf, Y. I.; Iranzo, J.; Shmakov, S. A.; Alkhnbashi, O. S.; Brouns, S. J. J.; Charpentier, E.; Cheng, D.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Scott, D.; Shah, S. A.; Siksnys, V.; Terns, M. P.; Venclovas, C.; White, M. F.; Yakunin, A. F.; Yan, W.; Zhang, F.; Garrett, R. A.; Backofen, R.; van der Oost, J.; Barrangou, R.; Koonin, E. V., Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **2020**, 18, 67-83.

23. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F., Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, 339, 819-823.

24. Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J., RNA-programmed genome editing in human cells. *Elife* **2013**, 2, e00471.

25. Konermann, S.; Brigham, M. D.; Trevino, A. E.; Joung, J.; Abudayyeh, O. O.; Barcena, C.; Hsu, P. D.; Habib, N.; Gootenberg, J. S.; Nishimasu, H.; Nureki, O.; Zhang, F., Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **2015**, 517, 583-588.

26. Murugan, K.; Babu, K.; Sundaresan, R.; Rajan, R.; Sashital, D. G., The revolution continues: newly discovered systems expand the CRISPR-Cas toolkit. *Mol. Cell* **2017**, 68, 15-25.

27. Fogarty, N. M. E.; McCarthy, A.; Snijders, K. E.; Powell, B. E.; Kubikova, N.; Blakeley, P.; Lea, R.; Elder, K.; Wamaitha, S. E.; Kim, D.; Maciulyte, V.; Kleinjung, J.; Kim, J.-S.; Wells, D.; Vallier, L.; Bertero, A.; Turner, J. M. A.; Niakan, K. K., Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **2017**, 550, 67-73.

28. Adli, M., The CRISPR tool kit for genome editing and beyond. *Nat Commun* **2018**, 9, 1-13.

29. Jinek, M., Krzysztof; Chylinski, Ines; Fonfara, M; Hauer, Jennifer; Doudna, and Emmanuelle Charpentier. 2012."A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.". *Science*, 816-21.

30. Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A., A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* **2015**, 348, 1477-1481.

31. Geny, S.; Pichard, S.; Brion, A.; Renaud, J.-B.; Jacquemin, S.; Concordet, J.-P.; Poterszman, A. Tagging Proteins with Fluorescent Reporters Using the CRISPR/Cas9 System and Double-Stranded DNA Donors. In *Multiprotein Complexes*; Springer: 2021, pp 39-57.

32. 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.

33. Tsai, S. Q.; Joung, J. K., Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nat. Rev. Genet.* **2016**, 17, 300-312.

34. Wu, S.-S.; Li, Q.-C.; Yin, C.-Q.; Xue, W.; Song, C.-Q., Advances in CRISPR/Cas-based gene therapy in human genetic diseases. *Theranostics* **2020**, 10, 4374.

35. Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A., A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Science advances* **2017**, 3, eaao0027.

36. Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A., Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **2017**, 550, 407-410.

37. 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.

38. Nishimasu, H.; Ran, F. A.; Hsu, Patrick D.; Konermann, S.; Shehata, Soraya I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, 156, 935-949.

39. Anders, C.; Bargsten, K.; Jinek, M., Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol. Cell* **2016**, 61, 895-902.

40. Dong, D.; Guo, M.; Wang, S.; Zhu, Y.; Wang, S.; Xiong, Z.; Yang, J.; Xu, Z.; Huang, Z., Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **2017**, 546, 436-439.

41. Yang, H.; Patel, D. J., Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9. *Mol. Cell* **2017**, 67, 117-127.e5.

42. Liu, L.; Yin, M.; Wang, M.; Wang, Y., Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race. *Mol. Cell* **2019**, 73, 611-620.e3.

43. 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.

44. Shin, J.; Jiang, F.; Liu, J.-J.; Bray, N. L.; Rauch, B. J.; Baik, S. H.; Nogales, E.; Bondy-Denomy, J.; Corn, J. E.; Doudna, J. A., Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci Adv* **2017**, 3, e1701620.

45. Jiang, F.; Liu, J.-J.; Osuna, B. A.; Xu, M.; Berry, J. D.; Rauch, B. J.; Nogales, E.; Bondy-Denomy, J.; Doudna, J. A., Temperature-Responsive Competitive Inhibition of CRISPR-Cas9. *Mol. Cell* **2019**, 73, 601-610.e5.

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

47. Palermo, G.; Chen, J. S.; Ricci, C. G.; Rivalta, I.; Jinek, M.; Batista, V. S.; Doudna, J. A.; McCammon, J. A., Key role of the REC lobe during CRISPR–Cas9 activation by 'sensing','regulating', and 'locking'the catalytic HNH domain. *Q. Rev. Biophys.* **2018**, 51.

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

49. Cotton, F. A.; Hazen, E. E.; Legg, M. J., Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme—thymidine 3′, 5′-bisphosphate—calcium ion complex at 1.5-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, 76, 2551-2555.

50. Li, C. L.; Hor, L. I.; Chang, Z. F.; Tsai, L. C.; Yang, W. Z.; Yuan, H. S., DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. *EMBO J.* **2003**, 22, 4014-4025.

51. Biertumpfel, C.; Yang, W.; Suck, D., Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **2007**, 449, 616-620.

52. Yang, W., An equivalent metal ion in one-and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* **2008**, 15, 1228-1231.

53. Yang, W., Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011**, 44, 1-93.

54. Yoon, H.; Zhao, L. N.; Warshel, A., Exploring the Catalytic Mechanism of Cas9 Using Information Inferred from Endonuclease VII. *ACS Catal.* **2019**, 9, 1329-1336.

55. Zhao, L. N.; Mondal, D.; Warshel, A., Exploring alternative catalytic mechanisms of the Cas9 HNH domain. *Proteins* **2020**, 88, 260-264.

56. Nierzwicki, Ł.; East, K. W.; Binz, J. M.; Hsu, R. V.; Ahsan, M.; Arantes, P. R.; Skeens, E.; Pacesa, M.; Jinek, M.; Lisi, G. P.; Palermo, G., Principles of target DNA cleavage and the role of Mg2+ in the catalysis of CRISPR–Cas9. *Nat. Catal.* **2022**, 5, 912-922.

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

58. 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, e46500.

59. 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.

60. 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, e46500.

61. Babu, K.; Kathiresan, V.; Kumari, P.; Newsom, S.; Parameshwaran, H. P.; Chen, X.; Liu, J.; Qin, P. Z.; Rajan, R., Coordinated actions of Cas9 HNH and RuvC nuclease domains are regulated by the bridge helix and the target DNA sequence. *Biochemistry* **2021**, 60, 3783-3800.

62. Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A., Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **2016**, 351, 867-871.

63. 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, 1247997.

64. Schafmeister, C.; Ross, W.; Romanovski, V., LEaP. *University of California, San Francisco* **1995**.

65. 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.

66. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, 11, 3696-3713.

67. Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham III, T. E., Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, 12, 4114-4127.

68. Zgarbová, M.; Otyepka, M.; Šponer, J. i.; Mládek, A. t.; Banáš, P.; Cheatham III, T. E.; Jurecka, P., Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **2011**, 7, 2886-2902.

69. Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., AMBER 2018, Univ. *California, San Fr* **2018**.

70. Zwanzig, R., Nonlinear generalized Langevin equations. *J. Stat. Phys.* **1973**, 9, 215-220.

71. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **1992**, 32, 523-535.

72. Gillespie, D. T., The chemical Langevin equation. *J. Chem. Phys.* **2000**, 113, 297-306.

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

74. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, 103, 8577-8593.

75. Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013**, 9, 3878-3888.

76. Roe, D. R.; Cheatham III, T. E., PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, 9, 3084-3095.

77. Likas, A.; Vlassis, N.; Verbeek, J. J., The global k-means clustering algorithm. *Pattern Recognit* **2003**, 36, 451-461.

78. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, 33, 889-897.

79. Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A., Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, 30, 211-243.

80. Wang, J.; Hou, T.; Xu, X., Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. *Curr. Comput. Aided Drug Des.* **2006**, 2, 287-306.

81. Homeyer, N.; Gohlke, H., Free energy calculations by the molecular mechanics Poisson− Boltzmann surface area method. *Mol. Inform.* **2012**, 31, 114-122.

82. Miller III, B. R.; McGee Jr, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA. py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012**, 8, 3314-3321.

83. Li, C. H.; Zuo, Z. C.; Su, J. G.; Xu, X. J.; Wang, C. X., The interactions and recognition of cyclic peptide mimetics of Tat with HIV-1 TAR RNA: a molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* **2013**, 31, 276-287.

84. Zuo, Z.; Liu, J., Cas9-catalyzed DNA cleavage generates staggered ends: evidence from molecular dynamics simulations. *Sci. Rep.* **2016**, 6, 1-9.

85. Zuo, Z.; Weng, J.; Wang, W., Insights into the inhibitory mechanism of D13-9001 to the multidrug transporter AcrB through molecular dynamics simulations. *J. Phys. Chem. B* **2016**, 120, 2145-2154.

86. Zuo, Z.; Smith, R. N.; Chen, Z.; Agharkar, A. S.; Snell, H. D.; Huang, R.; Liu, J.; Gonzales, E. B., Identification of a unique Ca2+-binding site in rat acid-sensing ion channel 3. *Nat. Commun.* **2018**, 9, 1-11.

87. Zuo, Z.; Liu, J., Structure and dynamics of Cas9 HNH domain catalytic state. *Scientific reports* **2017**, 7, 1-13.

88. Naseem-Khan, S.; Berger, M. B.; Leddin, E. M.; Maghsoud, Y.; Cisneros, G. A., Impact of Remdesivir Incorporation along the Primer Strand on SARS-CoV-2 RNA-Dependent RNA Polymerase. *J. Chem. Inf. Model.* **2022**, 62, 2456-2465.

89. Maghsoud, Y.; Dong, C.; Cisneros, G. A., Computational Characterization of the Inhibition Mechanism of Xanthine Oxidoreductase by Topiroxostat. *ACS Catal.* **2023**, 6023-6043.

90. Maghsoud, Y.; Dong, C.; Cisneros, G. A., Investigation of the Inhibition Mechanism of Xanthine Oxidoreductase by Oxipurinol: A Computational Study. *J. Chem. Inf. Model.* **2023**.

91. Kratz, E. G.; Walker, A. R.; Lagardère, L.; Lipparini, F.; Piquemal, J. P.; Andrés Cisneros, G., LICHEM: A QM/MM program for simulations with multipolar and polarizable force fields. *J. Comput. Chem.* **2016**, 37, 1019-1029.

92. Gökcan, H.; Vázquez-Montelongo, E. A.; Cisneros, G. A., LICHEM 1.1: recent improvements and new capabilities. *J. Chem. Theory Comput.* **2019**, 15, 3056-3065.

93. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

94. Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W., Tinker 8: software tools for molecular design. *J. Chem. Theory Comput.* **2018**, 14, 5273-5289.

95. Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, 10, 6615-6620.

96. Chai, J.-D.; Head-Gordon, M., Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, 128, 084106.

97. Kratz, E. G.; Duke, R. E.; Cisneros, G. A., Long-range electrostatic corrections in multipolar/polarizable QM/MM simulations. *Theor. Chem. Acc.* **2016**, 135, 1-9.

98. Fang, D.; Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A. s., Toward a deeper understanding of enzyme reactions using the coupled ELF/NCI analysis: application to DNA repair enzymes. *J. Chem. Theory Comput.* **2013**, 9, 2156-2160.

99. Singh, U. C.; Kollman, P. A., An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, 5, 129-145.

100. Eyring, H., The activated complex in chemical reactions. *J. Chem. Phys.* **1935**, 3, 107-115.

101. Evans, M. G.; Polanyi, M., Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.* **1935**, 31, 875-894.

102. Dzib, E.; Cabellos, J. L.; Ortíz-Chi, F.; Pan, S.; Galano, A.; Merino, G., Eyringpy: A program for computing rate constants in the gas phase and in solution. *Int. J. Quantum Chem.* **2019**, 119, e25686.

103. E. Dzib, A. Q., F. Ortiz-Chi, G. Merino, Eyringpy 2.0. *Cinvestav, Merida, Yucatan* **2021**.

104. Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W., Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010**, 132, 6498-6506.

105. Lu, T.; Chen, F., Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, 33, 580-592.

106. Swails, J.; Hernandez, C.; Mobley, D. L.; Nguyen, H.; Wang, L.-P.; Janowski, P., ParmEd. *URL: https://github. com/ParmEd/ParmEd* **2010**.

107. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, 97, 10269-10280.

108. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The REd. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, 12, 7821-7839.

109. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y., RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, 39, W511-W517.

110. Wang, F.; Becker, J.-P.; Cieplak, P.; Dupradeau, F.-Y. RED Python: Object oriented programming for Amber force fields. In Abstracts of Papers of the American Chemical Society, 2014; AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA: 2014; Vol. 247.

111. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, 25, 1157-1174.

112. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, 25, 247-260.

113. Graham, S. E.; Syeda, F.; Cisneros, G. A. s., Computational prediction of residues involved in fidelity checking for DNA synthesis in DNA polymerase I. *Biochemistry* **2012**, 51, 2569-2578.

114. Dewage, S. W.; Cisneros, G. A., Computational analysis of ammonia transfer along two intramolecular tunnels in Staphylococcus aureus glutamine-dependent amidotransferase (GatCAB). *J. Phys. Chem. B* **2015**, 119, 3669-3677.

115. Walker, A. R.; Cisneros, G. A. s., Computational simulations of DNA polymerases: detailed insights on structure/function/mechanism from native proteins to cancer variants. *Chem. Res. Toxicol.* **2017**, 30, 1922-1935.

116. Cui, Q.; Karplus, M., Catalysis and specificity in enzymes: a study of triosephosphate isomerase and comparison with methyl glyoxal synthase. *Adv. Protein Chem.* **2003**, 66, 315-372.

117. Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J., Preorganization and reorganization as related factors in enzyme catalysis: the chorismate mutase case. *Chem. - Eur. J.* **2003**, 9, 984-991.

118. Senn, H. M.; O'Hagan, D.; Thiel, W., Insight into enzymatic C− F Bond formation from QM and QM/MM calculations. *J. Am. Chem. Soc.* **2005**, 127, 13643-13655.

119. Cisneros, G. A.; Perera, L.; Schaaper, R. M.; Pedersen, L. C.; London, R. E.; Pedersen, L. G.; Darden, T. A., Reaction mechanism of the ε subunit of E. coli DNA polymerase III: insights into active site metal coordination and catalytically significant residues. *J. Am. Chem. Soc.* **2009**, 131, 1550-1556.

120. Fang, D.; Lord, R. L.; Cisneros, G. A., Ab initio QM/MM calculations show an intersystem crossing in the hydrogen abstraction step in dealkylation catalyzed by AlkB. *J. Phys. Chem. B* **2013**, 117, 6410-6420.

121. Fang, D.; Cisneros, G. A. s., Alternative pathway for the reaction catalyzed by DNA dealkylase AlkB from ab initio QM/MM calculations. *J. Chem. Theory Comput.* **2014**, 10, 5136-5148.

122. Torabifard, H.; Cisneros, G. A., Insight into wild-type and T1372E TET2-mediated 5hmC oxidation using ab initio QM/MM calculations. *Chem. Sci.* **2018**, 9, 8433-8445.

123. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, 25, 1605-1612.

124. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, 14, 33-38.

125. Dennington, R.; Keith, T. A.; Millam, J. M., GaussView 6.0. 16. *Semichem Inc.: Shawnee Mission, KS, USA* **2016**.

126. Raper, A. T.; Stephenson, A. A.; Suo, Z., Functional insights revealed by the kinetic mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.* **2018**, 140, 2971-2984.

127. Liu, M.-S.; Gong, S.; Yu, H.-H.; Jung, K.; Johnson, K. A.; Taylor, D. W., Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **2020**, 11, 3576.

128. Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **2014**, 507, 62-67.

129. Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A., Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **2015**, 527, 110-113.

130. Singh, D.; Sternberg, S. H.; Fei, J.; Doudna, J. A.; Ha, T., Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **2016**, 7, 1-8.

131. Jiang, F.; Doudna, J. A., CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys* **2017**, 46, 505-529.

132. Singh, D.; Wang, Y.; Mallon, J.; Yang, O.; Fei, J.; Poddar, A.; Ceylan, D.; Bailey, S.; Ha, T., Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* **2018**, 25, 347-354.

133. Bravo, J. P.; Liu, M.-S.; Hibshman, G. N.; Dangerfield, T. L.; Jung, K.; McCool, R. S.; Johnson, K. A.; Taylor, D. W., Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **2022**, 603, 343-347.

134. 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.

135. Yourik, P.; Fuchs, R. T.; Mabuchi, M.; Curcuru, J. L.; Robb, G. B., Staphylococcus aureus Cas9 is a multiple-turnover enzyme. *RNA* **2019**, 25, 35-44.

136. Aboelnga, M. M.; Wetmore, S. D., Unveiling a single-metal-mediated phosphodiester bond cleavage mechanism for nucleic acids: a multiscale computational investigation of a human DNA repair enzyme. *J. Am. Chem. Soc.* **2019**, 141, 8646-8656.

137. Hu, Q.; Jayasinghe-Arachchige, V. M.; Zuchniarz, J.; Prabhakar, R., Effects of the metal ion on the mechanism of phosphodiester hydrolysis catalyzed by metal-cyclen complexes. *Front Chem* **2019**, 7, 195.

138. Kamerlin, S. C. L.; Sharma, P. K.; Prasad, R. B.; Warshel, A., Why nature really chose phosphate. *Q. Rev. Biophys.* **2013**, 46, 1-132.

139. O'Ferrall, R. M., Relationships between E 2 and E 1c B mechanisms of β-elimination. *J. Chem. Soc. B* **1970**, 274-277.

140. Lassila, J. K.; Zalatan, J. G.; Herschlag, D., Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annu Rev Biochem* **2011**, 80, 669-702.

141. 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.

142. Wang, J.; Skeens, E.; Arantes, P. R.; Maschietto, F.; Allen, B.; Kyro, G. W.; Lisi, G. P.; Palermo, G.; Batista, V. S., 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* **2022**, 61, 785-794.

143. Zheng, L.; Shi, J.; Mu, Y., Dynamics changes of CRISPR-Cas9 systems induced by high fidelity mutations. *Phys. Chem. Chem. Phys.* **2018**, 20, 27439-27448.

144. Kleinstiver, B. P.; Pattanayak, V.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Zheng, Z.; Joung, J. K., High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **2016**, 529, 490-495.

145. Zuo, Z.; Babu, K.; Ganguly, C.; Zolekar, A.; Newsom, S.; Rajan, R.; Wang, Y. C.; Liu, J., Rational Engineering of CRISPR-Cas9 Nuclease to Attenuate Position-Dependent Off-Target Effects. *CRISPR J* **2022**, 5, 329-340.

146. Vakulskas, C. A.; Dever, D. P.; Rettig, G. R.; Turk, R.; Jacobi, A. M.; Collingwood, M. A.; Bode, N. M.; McNeill, M. S.; Yan, S. Q.; Camarena, J.; Lee, C. M.; Parka, S. H.; Wiebking, V.; Bak, R. O.; Gomez-Ospina, N.; Pavel-Dinu, M.; Sun, W. C.; Bao, G.; Porteus, M. H.; Behlke, M. A., A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **2018**, 24, 1216-1224.

147. Casini, A.; Olivieri, M.; Petris, G.; Montagna, C.; Reginato, G.; Maule, G.; Lorenzin, F.; Prandi, D.; Romanel, A.; Demichelis, F.; Inga, A.; Cereseto, A., A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **2018**, 36, 265-271.

148. Lee, J. K.; Jeong, E.; Lee, J.; Jung, M.; Shin, E.; Kim, Y.-h.; Lee, K.; Jung, I.; Kim, D.; Kim, S., Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* **2018**, 9, 1-10.

149. Bratovič, M.; Fonfara, I.; Chylinski, K.; Gálvez, E. J.; Sullivan, T. J.; Boerno, S.; Timmermann, B.; Boettcher, M.; Charpentier, E., Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat. Chem. Biol.* **2020**, 16, 587-595.

150. Babu, K.; Amrani, N.; Jiang, W.; Yogesha, S.; Nguyen, R.; Qin, P. Z.; Rajan, R., Bridge helix of Cas9 modulates target DNA cleavage and mismatch tolerance. *Biochemistry* **2019**, 58, 1905-1917.

FOR TABLE OF CONTENTS ONLY