Exploring CRISPR-Cas9 HNH-Domain Catalyzed DNA Cleavage Using Accelerated Quantum Mechanical Molecular Mechanical Free Energy Simulation

Richard Van,[†,‡] Xiaoliang Pan,[†] Saadi Rostami,[†] Jin Liu,[§] Pratul K. Agarwal,[∥] Bernard Brooks,[*,‡] Rakhi Rajan,[*,†] and Yihan Shao[*,†]

†*Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Pkwy, Norman, OK 73019, United States.*
‡*Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States.*
§*Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, TX 76107, United States.*
∥*High Performance Computing Center, Oklahoma State University, 106 Math Sciences, Stillwater, OK 74078, United States.*

E-mail: brb@nih.gov; rrajan@ou.edu; yihan.shao@ou.edu

Abstract

The target DNA (tDNA) cleavage catalyzed by the CRISPR Cas9 enzyme is a critical step in the Cas9-based genome editing technologies. Previously, the tDNA cleavage from an active SpyCas9 enzyme conformation was modeled by Palermo and coworkers [Nierzwicki *et al.*, Nat. Catal. 5, 912 (2022)] using *ab initio* quantum mechanical molecular mechanical (ai-QM/MM) free energy simulations, where the free energy barrier was found to be more favorable than that from a pseudoactive enzyme conformation. In this work, we performed ai-QM/MM simulations based on another catalytically active conformation (PDB 7Z4J) of the Cas9 HNH domain from cryo-electron microscopy experiments.For the wildtype enzyme, we acquired a free energy profile for the tDNA cleavage that is largely consistent with the previous report. Furthermore, we explored the role of the active-site K866 residue on the catalytic efficiency by modeling the K866A mutant, and found that the K866A mutation increased the reaction free energy barrier, which is consistent with the experimentally observed reduction in the enzyme activity.

Introduction

CRISPR-Cas9 technology has been an indispensable tool in biotechnology[1,2] and agricultural[3,4] research, owing to its versatility and precision in editing DNA sequences. Following the recent FDA approval, two gene therapies have become available for the treatment of the sickle cell disease.[5,6] Meanwhile, dozens of other gene therapies are currently under various stages of clinical trials.[7]

The capability of Cas9 to precisely edit DNAs stems from its gene targeting mechanism, whereby a single guide RNA (gRNA) molecule[8,9] steers the search for complementary genetic sequences; when a complementary match is found, Cas9 uses its two endonuclease domains (HNH and RuvC) to catalyze two independent DNA cuts, cleaving both strands of the DNA, thereby introducing a double strand DNA break. However, off-target editing can occur when base pair mismatches between the guide region of the gRNA and DNA are tolerated by Cas9, resulting in unintended genetic modifications. To address this, considerable advancements have been made to adapt the Cas9 protein for improved fidelity and reduced cleavage of mismatched DNAs.[10–18]

In general, advancements in developing safer Cas9 variants for gene editing have been facilitated by several complimentary techniques. Firstly, structural characterization of Cas9[19–30] provided a wealth of molecular details governing nucleic acid binding,[16,23,31–33] conformational dynamics,[22,34] catalysis,[35,36] and other nuclease activities.[37,38] Secondly, fluorescent resonance energy transfer (FRET) microscopy[39,40] and other kinetic studies[41–44] have been instrumental in advancing our knowledge of various steps in the Cas9 gene editing process.

Last but not the least, computational studies[23,44–55] have been equally useful in improving our mechanistic understanding of Cas9 activity. The readers are referred to (i) recent reviews[54,55] on the computational modeling of DNA recognition, conformational activation of nuclease domains, and release of cleavage products; and (ii) quantum mechanics/molecular mechanics (QM/MM) free energy simulations from the Palermo lab[56] to investigate the RuvC nuclease activity – non-target-strand DNA (ntDNA) cleavage. Hereby we briefly summarize key computational modeling efforts on the Cas9 HNH nuclease activity – target-strand DNA (tDNA) cleavage – that will be the main focus of this work.

The Streptococcus pyogenes Cas9 (SpyCas9) protein is composed of 1,368 residues, making a bi-lobed structure comprising a nuclease lobe (NUC) and a recognition lobe (REC), each divided into several subdomains (Figure 1). The enzyme has several conformational states as shown in Figure 1a: apo form of SpyCas9 (PDB 4CMP[19]); gRNA-bound (binary) state (PDB 4ZT0[57]); and gRNA/DNA-bound (ternary) state. Among the resolved structures of the ternary complex (Figure 1b), some are pseudo-active, a regulatory state prior to catalysis (*e.g.* PDB 5F9R[21]), while others are of post-cleavage state (e.g. PDBs 6O0Y,[26] 7S4X,[28] and 7Z4J[30]).

The experimental wildtype SpyCas9-HNH catalyzed tDNA hydrolysis rate ranges from 0.67 to 4.3 to over 700 s^{-1},[40–42] which corresponds to a free energy barrier of 13.6–17.8 kcal/mol within the transition state theory. To model SpyCas9 catalyzed tDNA hydrolysis, advanced computational tools such as QM/MM simulations have been employed by the Warshel group,[58,59] Cisneros/Liu groups,[60] and Palermo group[61] to describe the underlying

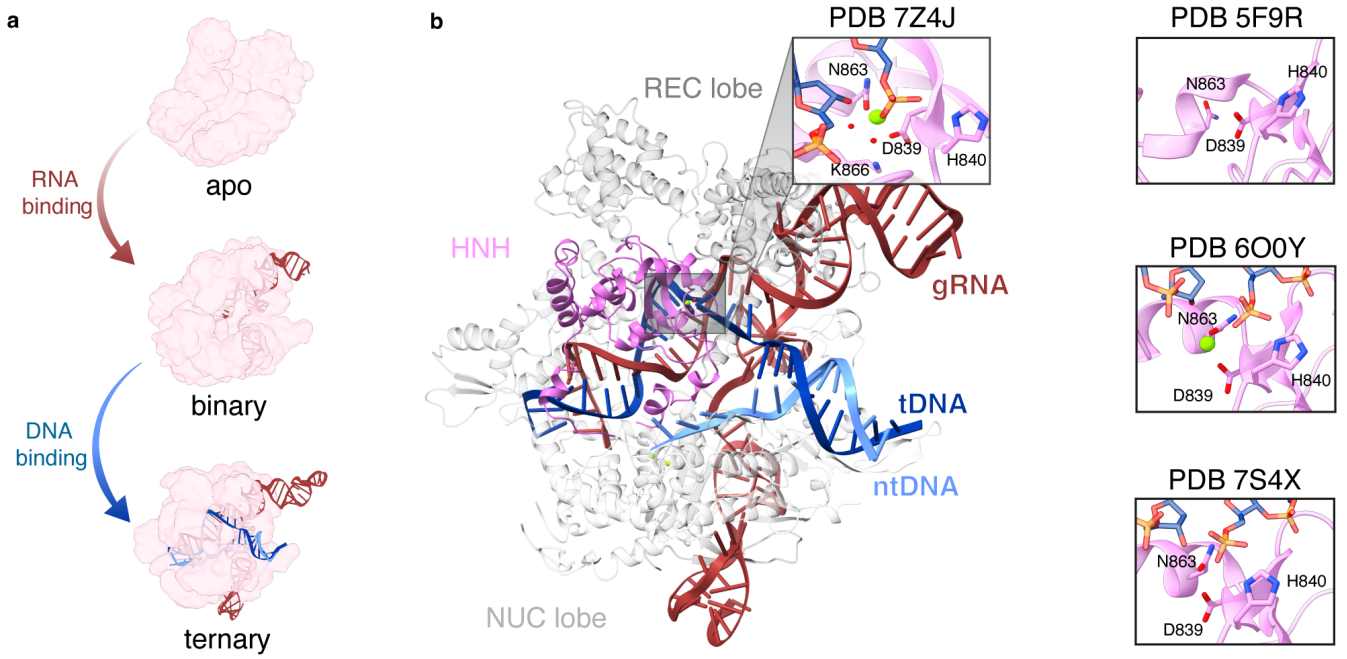

Figure 1: Overview of SpyCas9 conformational states. **(a)** A guide-RNA molecule (red) binds to the inactive Cas9 enzyme (pink) to form the binary complex. Upon recognizing the target DNA sequence, the DNA strands separate, then the ternary complex undergoes conformational changes to position the complementary DNA strand (target DNA; tDNA; dark blue) and non-complementary DNA strand (non-target DNA; ntDNA; light blue) in the two nuclease domains (HNH and RuvC), respectively. **(b)** Several HNH domain conformations have been studied, notably PDB 5F9R,[21] PDB 6O0Y,[26] PDB 7S4X,[28] and PDB 7Z4J.[30] In this study, PDB 7Z4J ternary complex, bound to gRNA, tDNA, and ntDNA, is utilized as the initial model for this study. The HNH domain is highlighted in magenta, along with the REC lobe (dark gray) and NUC lobe (light gray).

bond-breaking and forming events. Among these studies, there is a consensus (Figure 2) that it involves (i) a proton transfer, where a general base (H840) abstracts a proton from a water molecule, and (ii) a nucleophilic attack by the deprotonated water (*i.e.* hydroxide anion) on the phosphorous atom of the scissile phosphate, which causes the severance of the P–O3′ bond.

These studies, however, offered vastly different insights into the detailed mechanism, especially whether the deprotonated/nucleophilic water molecule comes from the first or second solvation shell of the Mg^{2+} ion. While the Warshel group examined the first-solvation-shell-water mechanism (Figure 2a) in their initial modeling (using a homology model for the Cas9/DNA/gRNA complex),[58] they later concluded that the second-solvation-shell-water mechanism (Figure 2b) is energetically more favorable (with a free energy barrier of 16.8 *vs* 21.0 kcal/mol) through modeling based on a cryo-EM structure (PDB 5Y36).[59] Adopting active site structures that combine multiple cryo-EM/X-ray crystallography structures (PDB: 6O0Y[26] and 5F9R[21]) and modeling trajectories, Cisneros, Liu, and coworkers arrived at an opposite conclusion, where the first-solvation-shell-water mechanism (with an energy barrier of 14.3 kcal/mol) is more feasible than the second-solvation-shell-water mechanism.[60] Meanwhile, the Palermo group started from the same cryo-EM structure (6O0Y) and reported a free energy barrier of 17.01 kcal/mol for a second-solvation-shell-water mechanism (Figure 2c).[61]

Such discrepancies in mechanistic details highlight the first and foremost challenge in QM/MM modeling of enzyme reactions: one has to start from a presumed chemically active state structure (*i.e.* pre-cleavage complex), and the choice of such a conformational state could strongly affect the predicted reaction pathway and the associated free energy

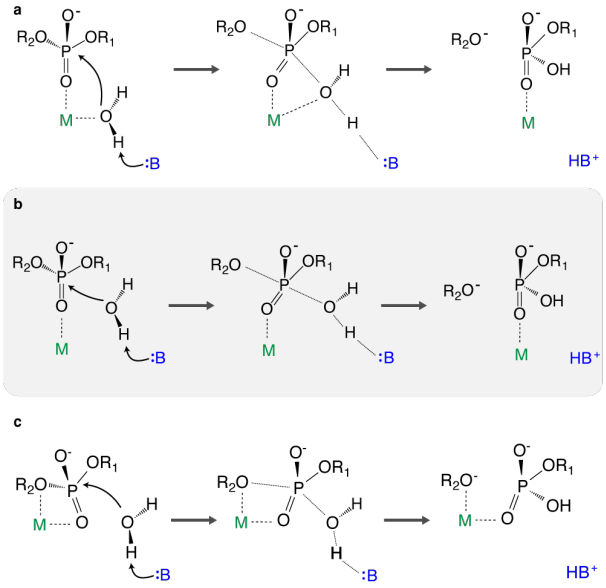

Figure 2: Schematic of previously proposed cleavage mechanisms for SpyCas9 catalyzed tDNA cleavage. It involves a general base abstracting a proton from a water molecule, increasing its nucleophilicity, then attacking the scissile phosphate and breaking the phosphate oxygen bond. However, the nucleophilic water could originate from the first-solvation shell (**a**) or the second-solvation shell (**b** and **c**) of the metal coordination sphere. Mechanism **c** also involves an additional coordination of DNA O3′ to the metal. Mechanisms **a** and **b** were explored by Warshel and colleagues[58,59] and by Cisneros, Liu, and colleagues.[60] Mechanism **c** was investigated by Palermo and coworkers[61] and is the focus of this work.

profile. Besides the coordination of the nucleophilic water, the active state structures from above QM/MM modeling also

differed in whether residues 860–867 (SDKNRGKS) adopt a helix or loop structure and whether the Mg^{2+} ion binds to the O3′ atom of the scissile phosphate, with the latter contributing to the difference in the two water-in-second-solvation-shell mechanisms (Figure 2b and Figure 2c). In principle, multiple conformational states can be catalytically active, with the net reaction rate being an ensemble average.[62] To further complicate the issue, the weight for each reaction pathway in the ensemble could vary substantially upon the mutation of enzyme residues or the introduction of different substrates (matched and mismatched DNAs). Ideally, one should perform a series of free energy simulations [for (i) multiple active state structures/reaction mechanisms, (ii) wildtype and mutant enzymes, and (iii) matched/mismatched DNAs] to gain a comprehensive mechanistic understanding of Cas9 HNH-domain mediated tDNA cleavage. Unfortunately, such free energy simulations, which require the evaluation of *ab initio* QM/MM energy/forces for millions of configurations for the enzyme-substrate complex along the entire bond breaking and formation pathway, would each take 10^5–10^6 CPU hours, rendering a thorough investigation infeasible.

In recent years, multiple time-step simulations (MTS)[63,64] and machine learning potentials (MLP)[65,66] have emerged as new strategies to substantially enhance the feasibility of *ab initio* QM/MM-quality simulation of enzyme reactions. Empowered by the MTS QM/MM free energy simulation methodology,[63,64] which lowers the cost of ai-QM/MM free energy simulations by several fold, here we revisited the role of chemical environment on the mechanism of SpyCas9 HNH domain catalyzed tDNA cleavage. We were especially interested in the role of K866 on SpyCas9 HNH domain catalyzed tDNA-cleavage, namely how the K866A mutation affects the reaction thermodynamics and kinetics, which has been reported experimentally.[61] Starting from one of the latest post-cleavage cryo-EM structures (PDB 7Z4J),[30] we explored the tDNA cleavage reaction by the wildtype SpyCas9 and the K866A variant. We also decomposed the free energy profile of wildtype SpyCas9 to assess the contributions of individual residues to the reaction-free energy profile. Our free energy simulation and energy decomposition analysis both suggested that K866 helps reduce the free energy barrier of the tDNA cleavage reaction, which was consistent with the experimental and computation reports by Palermo, Lisi, Jinek and coworkers.[61] The leverage of MLPs in modeling Cas9 catalysis will be reported in a separate manuscript.

Materials and Methods

The initial coordinates for the wildtype SpyCas9 model were derived from the Cryo-EM structure of Spy Cas9 bound to gRNA and/ cleaved DNA (PDB: 7Z4J[30]). The pre-catalytic state was assembled by manually bonding the O3′ atom of the cleaved tDNA fragment to the 5′-phosphorous atom of the other tDNA fragment and converting the 5′-phosphate oxygen to water. The MODELLER program[67] was used to add missing protein residues, and homology modeling was performed to add 9 missing nucleotides from the ntDNA. The protonation states of titratable residues were predicted with PROPKA.[68,69] The K866A variant was modeled based on the pre-catalytic wildtype model, followed by a manual conversion of residue K866 to alanine. Our simulations were performed using a modified version of the SANDER program from AmberTools20,[70] Q-Chem 5.2,[71] and an in-house QM/MM interface, QMHub, for DFT/MM calculations.

Molecular Dynamics Simulations. Wildtype and variant systems were modeled with the Amber ff14SB force field[72] for proteins, ff99bsc0 corrections for DNA, and ff99bsc0+χOL3 corrections for RNA. The 12-6-4 vdW potential from Li and Merz was adopted for the Mg^{2+} ion.[73] The ternary complex was solvated in a simulation box with 253,350 TIP3P water molecules[74] and neutralized with 115 counterions. The system was gradually heated from 0 K to 300 K using Langevin dynamics with a friction coefficient of 5 ps and equilibrated for 700 ps. A pressure of 1 atm was maintained with the Berendensen barostat with a relaxation time of 1 ps under periodic boundary conditions. Particle mesh Ewald method[75] was used to treat long-range electrostatic interactions, while van der Waals interactions were handled with a cutoff of 10 Å. The SHAKE algorithm[76] was used to constrain all bonds containing hydrogen atoms, and a time step of 1 fs was used for MD simulations with the leapfrog integrator. After the equilibration, the simulation box size was around 147 Å × 147 Å × 147 Å. Finally, the system was relaxed for 10 ns under NVT ensemble conditions. The classical simulations were performed using GPU-accelerated PMEMD program[77,78] from the Amber20 package.[70] The last frame from the 10 ns trajectory was used as the starting pre-catalytic structure for the subsequent QM/MM simulations in this study.

MTS QM/MM MD Simulations. The QM subsystem is composed of residues D839, N863, H840, DNA nucleotides (G3, T4), Mg^{2+} ion, and three water molecules. Hydrogen-capping atoms were added to maintain the valence of terminal QM atoms, resulting in 73 QM atoms (68 original QM atoms + 5 capping hydrogen atoms). The same QM region was applied to the K866A variant. These atoms were described with the B3LYP functional[79–81] and 6-31G* basis set.[82] Meanwhile, the rest of the system is described as MM charges within the Amber ff14SB force field. The QM/MM-AC method[83] was used to efficiently capture the long-range QM/MM electrostatic interactions (by projecting outer MM charges more than 10 Å away from the QM region onto inner MM atoms). All QM/MM simulations were performed at 300K under the NVT ensemble, utilizing the SIN(R) thermostat[84] with $L(=4)$ sets of thermostat variables coupled to each degree of freedom, where $Q_1 = Q_2 = k_BT(0.05\text{ps})^2$ was adopted for the thermostat coupling parameter with 1 ps^{-1} for the friction coefficient. Accelerated *ab initio* QM/MM MD simulations were carried out with multiple time-step integration.[64] In particular, the QM subsystem is propagated at a 1 fs inner time-step using a semi-empirical energy function, PM3/MM. At each 4 fs outer time-step, correction forces were computed at B3LYP/6-31G*/MM level and applied to the QM subsystem.

Umbrella Sampling. The reaction coordinate was defined as the difference between the $O3'_{DNA} - P_{DNA}$ and $P_{DNA} - O_{WAT}$ distances. A total of 42 umbrella windows were evenly distributed from −1.90 Å to 2.20 Å at an interval of 0.10 Å, with a force constant of 150 kcal Å/mol. Initially, configurations for each window were obtained by gradually pulling the reactant toward the product. A subsequent reverse process returned the system to the reactant, and each window was simulated for 0.5 ps. Independent simulations were initiated from the final reversed coordinates and carried out for 50 ps with configurations saved every 0.1 ps, resulting in 500 configurations per window. The free energy profiles were computed using these configurations with PyMBAR with the

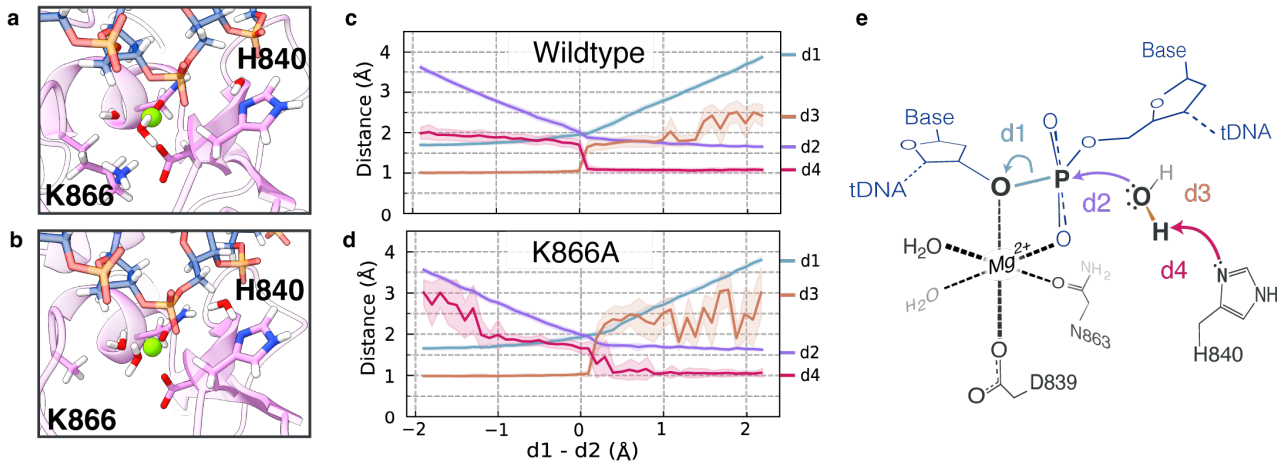

Figure 3: **(a,b)** Pre-catalytic HNH subdomain structure for QM/MM simulations of wildtype SpyCas9 and K866 mutant. **(c,d)** The average distance in each umbrella window for wild-type and K866A. Error bars are shown with a shaded area. **(e)** Key distances in the chemical transformation.

Multistate Bennett acceptance ratio (MBAR[85]) method. The same approach was applied to the SpyCas9 and the K866A mutant complexes.

Free Energy Decomposition by Residue. The free energy profile was decomposed to assess the contributions from different parts of the MM subsystem to the chemical transformation of the QM subsystem.[86] For each window, every other saved configuration was taken (0.2 ps), resulting in 250 configurations in each window for the analysis. Each residue's electrostatic, polarization, and van der Waals contributions were obtained by integrating the force for the corresponding residue atoms interacting with the QM region along the reaction pathway.

Thermodynamic Perturbation Here, the same set of 250 configurations for wildtype SpyCas9 enzyme and K866A variant complexes were adopted. With the reference B3LYP/6-31G*/MM energy values already computed along the original trajectories, all configurations were subjected to additional single-point energy calculations at B3LYP/6-31+G*, ωB97X-D/6-31G*, and ωB97X-D/6-31+G* levels of theory.[87] After that, the pyMBAR package[85] was used to compute the weighted thermodynamic perturbation (wTP) corrected potential of mean force .[88,89]

Results and Discussion

Structure of the Reactant Complex. For the 50 ns MD trajectory for the pre-catalytic complex, the root mean square deviation (RMSD) exhibited a plateau at $\sim$ 2.7 Å (Figure S1a). Root mean square fluctuation (RMSF) analysis indicated similar conformational fluctuations in the pre-catalytic complexes (Figure S1d). A closer inspection at the configurations, however, revealed some differences: for the wildtype system, the Mg^{2+} ion sustains coordination with D839, N863, DG, DT, and two water molecules. In contrast, the K866A system exhibited weaker interactions between the Mg^{2+} ion and almost all of its ligands (Figure S2). Also noticeable was that a slightly longer distance between the phosphorous atom and the nucleophilic water (top of Figure S2) with the K866A variant, thereby making the complex less competent for the chemical transformation.

tDNA Cleavage Reaction Pathway. Figure 3c and Figure 3d shows the evolution of some key distances along the reaction pathway, as catalyzed by the wildtype enzyme and K866A variant, respectively. These included not only the lengths of the scissile O–P bond (d1) and the nascent P–O bond (d2), but also, as a proton transfer is involved in the H840-mediated pathway, the lengths of the broken water O–H bond (d3) and the distance of the transferred proton from the δ-nitrogen atom (d4) of H840, which serves as a general base. These distances were averaged in each simulation window along the reaction pathway, with the shaded region representing the standard deviation within each window.

Figure 3c and 3d demonstrated a concerted reaction, as the phosphodiester bond cleavage occurred in the same simulation windows (with d1-d2 around 0.1 Å for the wildtype and 0.2 Å for K866A) as the proton transfer. Compared to the wildtype, the initial position of nucleophilic water for the K866A variant was further away from the H840 general base (with a d4 distance of $\sim$3.0 Å vs 2.0 Å) and was drawn closer only after d1-d2 reaches ~ -1.2 Å. [The d2 and d4 distributions for the reactant complex differed from those in Figure S2 because of the different description (QM/MM *vs* MM) of the active site.] Additionally, the length of the broken O–H bond (d3) of the K866A variant increased immediately to $\sim$2.5 Å after losing the proton to H840 and fluctuates more substantially between 2.0 Å to 3.0 Å. Figure S3 shows the Mg^{2+}-coordination distances along the reaction pathway of both systems, with the distance of of O3'–Mg^{2+} coordination decreasing to $\sim$ 2.0 Å at the transition state, and then staying at $\sim$ 2.0 Å in the product formation. Compared to the wildtype enzyme reaction pathway, the initial distance of O3' to Mg^{2+} exhibited more fluctuations for the K866A variant ($\sim$ 2.8 Å). This again suggested a more dynamic nature of the HNH active site with the K866A variant, which would overall hinder the chemical transformation.

tDNA Cleavage Reaction Free Energy Profile. The potential of mean force (PMF), as computed using MBAR, was shown in Figure 4 for the wildtype (pink) and K866A (purple) catalyzed reactions. The wildtype system exhibited a reaction free energy of 5.0 kcal/mol, which indicated an endothermic reaction, and a free energy barrier of 20.6 $\pm$ 0.1 kcal/mol at a d1-d2 value of around 0.15Å. This free energy barrier is slightly higher than the aforementioned experimentally derived activation energy of 13.6 – 17.8 kcal/mol.

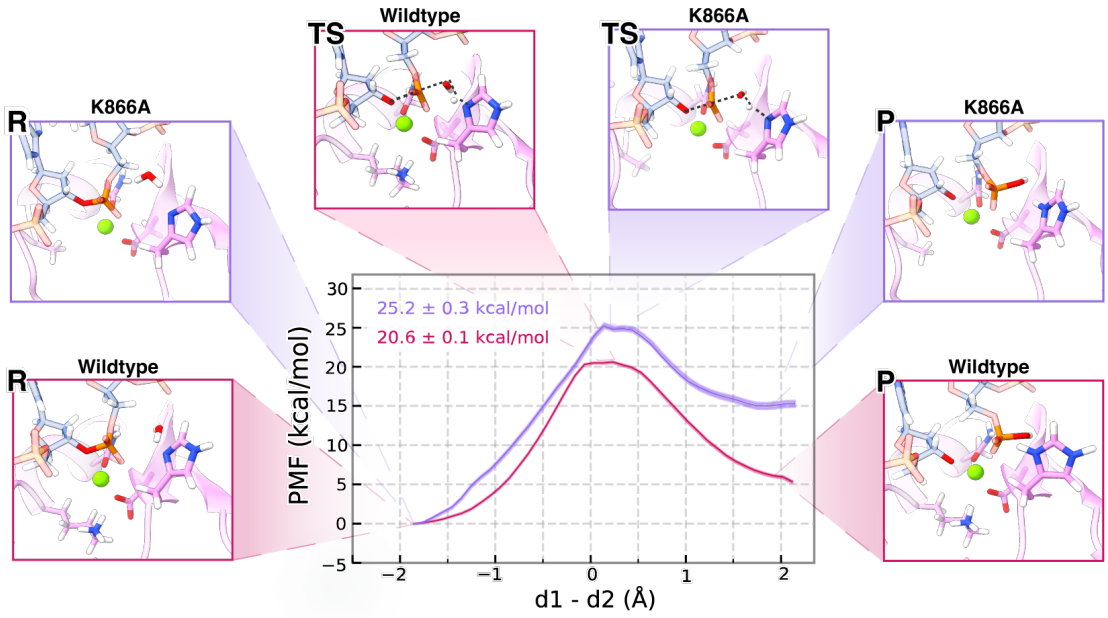

Figure 4: Free energy profiles of HNH domain DNA cleavage in wild-type SpyCas9 (pink) and K866A mutant (purple). The reaction proceeds from the reactant state (**R**) to the transition state (**TS**), followed by the charged product (**P**). Both systems start with the lowest energy reactant state at the reaction coordinate, which is -1.90 Å. The **TS** for K866A is at 0.15 Å, while the wild-type is at 0.25 Å. The product state was chosen from the lowest energy window, which for K866A is at 1.90 Å, while at 2.20 Å for the wildtype.

Meanwhile, the K866A variant was predicted by our simulation to have a product state of this reaction is even less energetically favorable, with its free energy being 15.0 kcal/mol higher than the reactant state. The free energy barrier occurred at a d1-d2 value of around 0.25Å; this slightly later transition state was consistent with an overall more endothermic reaction. The corresponding free energy barrier was found to be 25.2 ± 0.3 kcal/mol, reflecting a 4.6 kcal/mol increase compared to the wildtype enzyme. This indicates that *K866 preferentially stabilizes the transition state over the reactant*. The transition state stabilization is therefore another key role of K866, in addition to its involvement in the protonation of the O3′ atom after the cleavage as suggested by Palermo and coworkers.[61]

Free Energy Contribution from Active-Site Residues. To evaluate the influence of K866 and other active-site residues on the reaction, free energy decomposition analysis was carried out on the wildtype system's free energy profile. The detailed energy contributions of individual residues were summarized in Table 1, which listed twelve residues predicted to affect the free energy barrier most significantly. As shown in the Table, K866 contributed -5.38 kcal/mol to the free energy barrier, which is remarkably consistent with the predicted 4.6 kcal/mol barrier increase of the K866A variant over the wildtype (Figure 4).

In addition to K866, the residues that contributed the most negative values to the free energy profile, *i.e.* effectively increasing the cleavage activity, included D850, E809, R864, V838, D849. Meanwhile, residues that contributed the most positive value to the free energy barrier, *i.e.* effectively decreasing the cleavage activity, were D868, K918, K896, K848, K810, and K855. As expected, this analysis revealed that permanent electrostatics dominated the reaction energetics, with 11 out of these 12 residues carrying charged side chains. A more detailed summary containing all protein residues (with a contribution of 1.0 kcal/mol or higher) can be found in Table S1. In Figure 5, a 3D heatmap of the free energy barrier potential was mapped onto the Cas9 enzyme structure, further highlighting the influence of these key residues.

Experimentally, many of these residues were studied previously. The K810A and K848A mutations were suggested by the Zhang lab[11] to be part of the eSpCas9 variant; the K848A mutation was included in the FeCas9 variant from the Hu lab.[90] More recent experiments by Lisi, Palmero, Batista and coworkers[91] also demonstrated improved enzyme specificity with K810A, K848A, and K855A single mutations. While these residues were suggested to modulate the HNH domain through an allosteric network,[92–94] our free energy decomposition analysis results in Table 1 indicated that these three lysine residues also increased the free energy barrier (K810: 4.56 ± 0.02 kcal/mol; K848: 4.12 ± 0.02 kcal/mol; and K855: 9.51 ± 0.02 kcal/mol, effectively decreasing the cleavage activity). Reaction-wise, losing the positive charges on these residues is thus unlikely to adversely impact the phosphodiester bond breaking of matched DNAs.

Computationally, Cisneros, Liu, and coworkers[60] recently explored how DNA mismatches near the HNH domain affect the catalytic minimum energy pathways. They demonstrated significant structural changes in the local protein environment, disruptions to RNA/DNA interactions, as well as substantial changes to the reaction energetics. Their EDA results highlighted the vital role of not only K810 and K855 in the chemical step, but also that of K896 and K918 (also reported in Table 1).

Effect of the QM Model.

To estimate the effect of using different functionals to compute the free energy profile, configurations along the wildtype and K866A simulation trajectories (at a

Table 1: Energy decomposition analysis of individual residue energy contributions to the free energy barrier (in kcal/mol). The twelve residues with the most influence on the free energy barrier are listed. Additional EDA information can be found for residues (Table S1), DNA (Table S2), and RNA (Table S3).

Residues	$E^{perm.}_{\mathrm{QM-MM}}$	$E^{pol.}_{\mathrm{QM-MM}}$	$E^{vDW.}_{\mathrm{QM-MM}}$	E_{Total}
D850	-10.47 ± 0.03	1.38 ± 0.02	0.00 ± 0.00	-9.09 ± 0.02
E809	-7.32 ± 0.03	-0.18 ± 0.03	0.00 ± 0.00	-7.51 ± 0.03
R864	-6.60 ± 0.04	0.33 ± 0.02	0.00 ± 0.00	-6.27 ± 0.04
V838	-7.34 ± 0.06	2.05 ± 0.03	-0.95 ± 0.04	-6.24 ± 0.07
D849	-6.68 ± 0.03	0.98 ± 0.01	0.00 ± 0.00	-5.70 ± 0.02
K866	-6.39 ± 0.05	1.01 ± 0.03	0.00 ± 0.00	-5.38 ± 0.04
⋮	⋮	⋮	⋮	⋮
D868	3.35 ± 0.01	-0.22 ± 0.01	0.01 ± 0.00	3.13 ± 0.01
K918	3.63 ± 0.01	-0.47 ± 0.01	0.00 ± 0.00	3.16 ± 0.01
K896	4.17 ± 0.01	-0.38 ± 0.00	0.00 ± 0.00	3.79 ± 0.01
K848	4.54 ± 0.02	-0.42 ± 0.01	0.00 ± 0.00	4.12 ± 0.02
K810	4.49 ± 0.02	0.07 ± 0.01	0.00 ± 0.00	4.56 ± 0.02
K855	10.53 ± 0.03	-1.01 ± 0.02	0.00 ± 0.00	9.51 ± 0.02

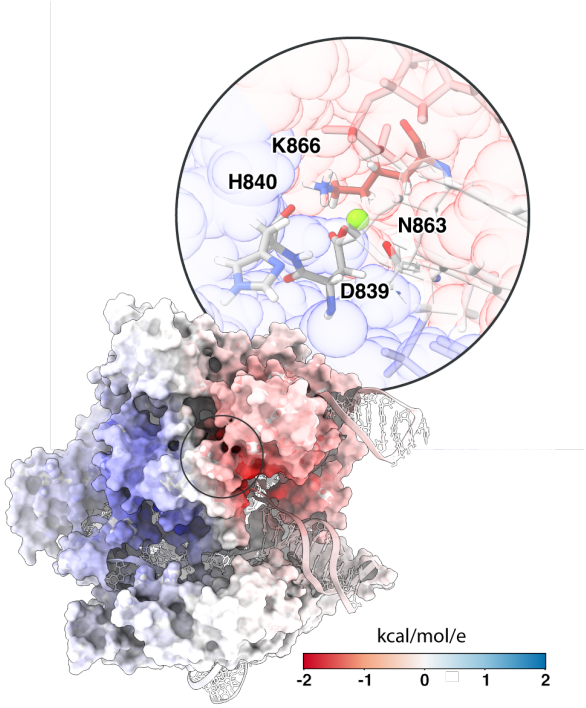

Figure 5: The free energy barrier potential mapped onto the cryo-EM structure of SpyCas9 (*lower left*), with a close-up view of the HNH domain (*upper right*). The negative potential region is marked red (preferring positive-charged residues), whereas the positive potential region is blue.

B3LYP/6-31G* description of the active site) were collected for further weighted thermodynamics perturbation (wTP) to the ωB97X-D/6-31G*, and ωB97X-D/6-31+G* levels of theory. The free energy profiles were smoothened using Gaussian Progress Regression (GPR) curve fitting,[89] with the original B3LYP/6-31G* free energy profile (where the free energy barriers are slightly different from those in Figure 4 due to the use of fewer configurations in the MBAR analysis) and the wTP-corrected ones shown in Figure 6.

Between the B3LYP free energy profiles for the wildtype SpyCas9 mediated reaction, the enlarging of the basis set from 6-31G* to 6-31+G* led to an increase of around 2.0 kcal/mol in the free energy barrier. With the 6-31G* basis, the wTP from the B3LYP functional to ωB97X-D resulted in a 1.4 kcal/mol decrease in the free energy barrier to 19.1 ± 0.8 kcal/mol. A similar free energy barrier of 19.0 ± 1.0 kcal/mol was estimated for the ωB97X-D/6-31+G* level of theory. Thus, it is quite encouraging that the free energy barrier with the ωB97X-D functional agreed better with experimental kinetic rates.

For the K866A mutant, upon the addition of diffuse function to the 6-31G* basis set, the free energy barrier increased only by 0.3 and 1.1 kcal/mol, respectively with B3LYP and ωB97X-D functionals. On the other hand, upon the change of B3LYP to ωB97X-D functional, the free energy barriers lowered substantially more, by 4.8 and 4.0 kcal/mol, respectively, with 6-31G* and 6-31+G* basis sets. Most importantly, for all four functional/basis combinations, the K866A free energy barrier remained to be at least 2.0 kcal/mol higher than the wildtype, which is consistent with experimental observations. This highlights the general reliability of DFT-based QM/MM free energy simulations in the prediction of mutation effects.

Computational Efficiency of the ai-QM/MM Simulation. For tDNA cleavage catalyzed by the wildtype SpyCas9 enzyme and its K866A variant, each free energy calculation required 42 umbrella windows and a 50 ps MTS QM/MM MD simulation per window. The net 2.1 ns simulation (*i.e.* 2,100,000 single-point QM/MM force evaluations) took over 270,000 CPU hours (Table 2). Had DFT energy/force calculations been performed at each 1 fs timestep (instead of at 4 fs time intervals), it would have taken over 1,050,0000 CPU hours. On the other hand, a semi-empirical QM/MM simulation (using a PM3 or AM1-d description of the QM region) would be about 50 times cheaper, at 4700 CPU hours, than our MTS simulation. But such semi-empirical QM/MM simulation would produce much higher free energy barriers that are inconsistent with experimental observations.

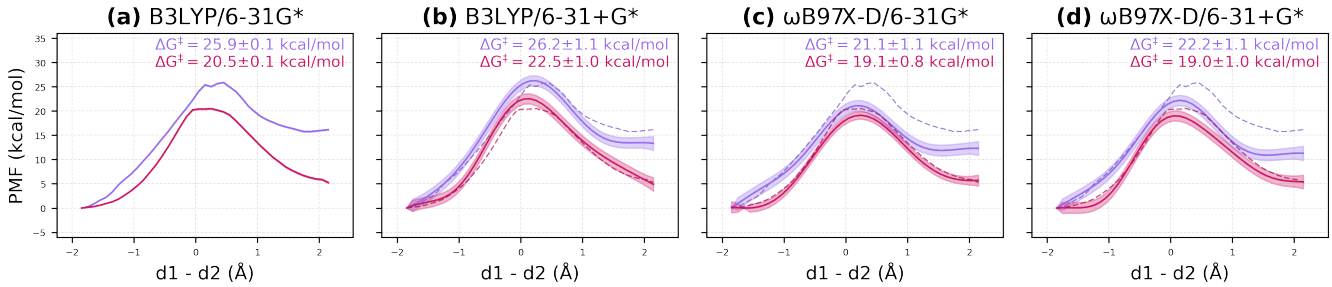

Figure 6: Free energy profiles for wild-type SpyCas9 (red) and K866A variant (purple) catalyzed tDNA cleavage by correcting the reference potential (B3LYP/6-31G*) to B3LYP/6-31+G*, ωB97X-D/6-31G*, and ωB97X-D/6-31+G* levels of theory. Note that the reference free energy profiles (solid lines in panel a and dashed lines in other panels) deviate slightly from the ones in Figure 4, because only 250 configurations (instead of 500) were used from each window in the MBAR free energy calculation.

Table 2: Timings for semi-empirical and DFT-based *ab initio* QM-MM free energy simulations of tDNA cleavage as mediated by wildtype SpyCas9 enzyme or its K866 variant. The number of steps refers to the biased MD simulations in each of the 42 umbrella windows. The time for each simulation step and total CPU time was clocked based on a single core of an Intel Xeon Gold 6130F 2.10GHz CPU.

QM Theory	#(Steps)	Time per step (s)	CPU hours
PM3 or AM1-d	50,000	0.5	4,700
B3LYP	50,000	113.0	1,054,700
B3LYP/6-31G* / PM3	50,000	29.0	270,700

Conclusion

In this study, through *ab initio* QM/MM free energy simulations, the following was achieved:

- The free energy barrier for SpyCas9 H840-mediated tDNA cleavage was predicted to increase by 4.6 kcal/mol going from the wildtype enzyme to the K866A variant. Within the transition state theory, such an increase in the free energy barrier would lower the reaction rate by 3 or 4 orders of magnitude. This is comparable to the reduced kinetic rate observed from biochemical experiments.[61] Such a direct enhancement of catalytic rate complements the previously suggested role of K866 in the protonation of tDNA cleavage product.
- In line with the increase in the reaction-free energy barrier with the K866A mutation, we also observed an energetically less favorable cleavage product. Analysis of key distances along the reaction pathway revealed more substantial structural fluctuations in the reactant and product states of the K866A variant. It increases the flexibility of the HNH domain, thereby destabilizing the ternary complex.
- Our free energy decomposition analysis reaffirmed the key role of the protein permanent electrostatics in modulating the reaction. Furthermore, this analysis allowed us to pinpoint contributions from K866 and other residues of wildtype SpyCas9, especially those in or near the active site, to the free energy barrier. It enhanced our mechanistic understanding of how SpyCas9 catalyzes the tDNA cleavage and provided a systematic way to identify potential residues for mutation (several of which had previously already been explored experimentally).
- In addition to the free energy decomposition analysis, this is also the first application of the multiple timestep ai-QM/MM simulation[64] to complex enzyme reactions, such as SpyCas9 catalysis. It clearly demonstrated the advantage of MTS simulations of accelerating enzyme free energy simulations by several fold without sacrificing the accuracy.

Nevertheless, caution should be taken when tapping into the individual residue energetic contributions in the design of enzyme variants. Firstly, the estimated energetic contributions relate to only the catalytic step. To fully assess the effect of mutating a residue, one should also analyze how the mutation affects the formation of binary and ternary complexes as well as the conformational free energy landscape of the ternary complex, especially the transition from pseudoactive to an active conformation. Secondly, compared to the wildtype, a variant might be subjected to substantial structural (and thus further energetic) perturbations along the catalytic reaction pathway, which can only be modeled accurately by performing ai-QM/MM free energy simulation of the variant.

Acknowledgement The authors acknowledge the financial support from the National Institutes of Health: R01GM135392 (Y. Shao), R01GM148886 (P. Agarwal), R21GM144860 (J. Liu), R35GM153297 (Y. Shao), R44GM133270 (Y. Shao), P30GM145423 (Y. Shao and R. Rajan), and ZIA-HL001051 (B. Brooks). The computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma and the High Performance Computing Center (Pete) at Oklahoma State University.

References

(1) Boubakri, H. (2023) Recent Progress in CRISPR/cas9-Based Genome Editing for Enhancing Plant Disease Resistance. *Gene 866*, 147334, DOI: 10.1016/j.gene.2023.147334.
(2) Eslahi, A., Alizadeh, F., Avan, A., Ferns, G. A., Moghbeli, M., Reza Abbaszadegan, M., and Mojarrad, M. (2023) New Advancements in CRISPR Based Gene Therapy of Duchenne Muscular Dystrophy. *Gene 867*, 147358, DOI: 10.1016/j.gene.2023.147358.
(3) Ravikiran, K. T., Thribhuvan, R., Sheoran, S., Kumar, S., Kushwaha, A. K., Vineeth, T. V., and Saini, M. (2023) Tailoring Crops with Superior Product Quality through Genome Editing: An Update. *Planta 257*, 86, DOI: 10.1007/s00425-023-04112-4.

(4) Yaqoob, H., Tariq, A., Bhat, B. A., Bhat, K. A., Nehvi, I. B., Raza, A., Djalovic, I., Prasad, P. V., and Mir, R. A. (2023) Integrating Genomics and Genome Editing for Orphan Crop Improvement: A Bridge between Orphan Crops and Modern Agriculture System. *GM Crops Food. 14*, 1–20, DOI: 10.1080/21645698.2022.2146952.
(5) Canver, M. C. *et al.* (2015) BCL11A Enhancer Dissection by Cas9-Mediated in Situ Saturating Mutagenesis. *Nature 527*, 192–197, DOI: 10.1038/nature15521.
(6) Frangoul, H. *et al.* (2021) CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med. 384*, 252–260, DOI: 10.1097/01.ogx.0000754392.61396.79.
(7) Khoshandam, M., Soltaninejad, H., Mousazadeh, M., Hamidieh, A. A., and Hosseinkhani, S. (2024) Clinical Applications of the CRISPR/cas9 Genome-Editing System: Delivery Options and Challenges in Precision Medicine. *Genes Dis. 11*, 268–282, DOI: 10.1016/j.gendis.2023.02.027.
(8) Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-Guided Human Genome Engineering via Cas9. *Science 339*, 823–826, DOI: 10.1126/science.1232033.
(9) Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science 337*, 816–821, DOI: 10.1126/science.1225829.
(10) Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science 339*, 819–823, DOI: 10.1126/science.1231143.
(11) Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016) Rationally Engineered Cas9 Nucleases with Improved Specificity. *Science 351*, 84–88, DOI: 10.1126/science.aad5227.
(12) 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., and Doudna, J. A. (2017) Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature 550*, 407–410, DOI: 10.1038/nature24268.
(13) Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F., Inga, A., and Cereseto, A. (2018) A Highly Specific Spcas9 Variant Is Identified by in Vivo Screening in Yeast. *Nat. Biotechnol. 36*, 265–271, DOI: 10.1038/nbt.4066.
(14) Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.-h., Lee, K., Jung, I., Kim, D., Kim, S., and Kim, J.-S. (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun. 9*, 3048, DOI: 10.1038/s41467-018-05477-x.
(15) Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018) Evolved Cas9 Variants with Broad Pam Compatibility and High Dna Specificity. *Nature 556*, 57–63, DOI: 10.1038/nature26155.
(16) Babu, K., Amrani, N., Jiang, W., Yogesha, S., Nguyen, R., Qin, P. Z., and Rajan, R. (2019) Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance. *Biochem. 58*, 1905–1917, DOI: 10.1021/acs.biochem.8b01241.
(17) Guo, M., Ren, K., Zhu, Y., Tang, Z., Wang, Y., Zhang, B., and Huang, Z. (2019) Structural Insights into a High Fidelity Variant of Spcas9. *Cell Res. 29*, 183–192, DOI: 10.1038/s41422-018-0131-6.
(18) Zuo, Z., Babu, K., Ganguly, C., Zolekar, A., Newsom, S., Rajan, R., Wang, Y.-C., and Liu, J. (2022) Rational Engineering of CRISPR-Cas9 Nuclease to Attenuate Position-Dependent Off-Target Effects. *The CRISPR Journal 5*, 329–340, DOI: 10.1089/crispr.2021.0076.
(19) 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., and Doudna, J. A. (2014) Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science 343*, 1247997, DOI: 10.1126/science.1247997.
(20) Nishimasu, H., Ran, F., Hsu, P., Konermann, S., Shehata, S., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014) Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell 156*, 935–949, DOI: 10.1016/j.cell.2014.02.001.
(21) Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., and Doudna, J. A. (2016) Structures of a CRISPR-Cas9 R-Loop Complex Primed for Dna Cleavage. *Science 351*, 867–871, DOI: 10.1126/science.aad8282.
(22) Jiang, F., and Doudna, J. A. (2017) CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys. 46*, 505–529, DOI: 10.1146/annurev-biophys-062215-010822.
(23) Zuo, Z., and Liu, J. (2017) Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Sci. Rep. 7*, 17271, DOI: 10.1038/s41598-017-17578-6.
(24) Swarts, D. C., and Jinek, M. (2018) Cas9 Versus Cas12a/cpf1: Structure–Function Comparisons and Implications for Genome Editing. *WIREs RNA 9*, e1481, DOI: 10.1002/wrna.1481.
(25) Palermo, G. (2019) Structure and Dynamics of the CRISPR–Cas9 Catalytic Complex. *J. Chem. Inf. Model. 59*, 2394–2406, DOI: 10.1021/acs.jcim.8b00988.
(26) Zhu, X., Clarke, R., Puppala, A. K., Chittori, S., Merk, A., Merrill, B. J., Simonović, M., and Subramaniam, S. (2019) Cryo-Em Structures Reveal Coordinated Domain Motions That Govern DNA Cleavage by Cas9. *Nat Struct Mol Biol 26*, 679–685, DOI: 10.1038/s41594-019-0258-2.
(27) Bravo, J. P., Hibshman, G. N., and Taylor, D. W. (2022) Constructing Next-Generation CRISPR–Cas Tools from Structural Blueprints. *Curr. Opin. Biotech. 78*, 102839, DOI: 10.1016/j.copbio.2022.102839.
(28) Bravo, J. P. K., Liu, M.-S., Hibshman, G. N., Dangerfield, T. L., Jung, K., McCool, R. S., Johnson, K. A., and Taylor, D. W. (2022) Structural Basis for Mismatch Surveillance by CRISPR–Cas9. *Nature 603*, 343–347, DOI: 10.1038/s41586-022-04470-1.
(29) Pacesa, M., Lin, C.-H., Cléry, A., Saha, A., Arantes, P. R., Bargsten, K., Irby, M. J., Allain, F. H.-T., Palermo, G., Cameron, P., Donohoue, P. D., and Jinek, M. (2022) Structural Basis for Cas9 Off-Target Activity. *Cell 185*, 4067–4081.e21, DOI: 10.1016/j.cell.2022.09.026.
(30) Pacesa, M., Loeff, L., Querques, I., Muckenfuss, L. M., Sawicka, M., and Jinek, M. (2022) R-Loop Formation and Conformational Activation Mechanisms of Cas9. *Nature 609*, 191–196, DOI: 10.1038/s41586-022-05114-0.
(31) Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014) Dna Interrogation by the CRISPR

Rna-Guided Endonuclease Cas9. *Nature 507*, 62–67, DOI: 10.1016/j.bpj.2013.11.3848.
(32) Newsom, S., Parameshwaran, H. P., Martin, L., and Rajan, R. (2021) The CRISPR-Cas Mechanism for Adaptive Immunity and Alternate Bacterial Functions Fuels Diverse Biotechnologies. *Front. Cell. Infect. Microbiol. 10*, 619763, DOI: 10.3389/fcimb.2020.619763.
(33) Zhong, Z., Li, Z., Yang, J., and Wang, Q. (2023) Unified Model to Predict gRNA Efficiency across Diverse Cell Lines and CRISPR-Cas9 Systems. *J. Chem. Inf. Model. 63*, 7320–7329, DOI: 10.1021/acs.jcim.3c01339.
(34) Sternberg, S. H., LaFrance, B., Kaplan, M., and Doudna, J. A. (2015) Conformational Control of Dna Target Cleavage by CRISPR–Cas9. *Nature 527*, 110–113, DOI: 10.1038/nature15544.
(35) Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012) Cas9–Crrna Ribonucleoprotein Complex Mediates Specific Dna Cleavage for Adaptive Immunity in Bacteria. *Proc. Natl. Acad. Sci. 109*, E2579–E2586, DOI: 10.1073/pnas.1208507109.
(36) Tang, H., Yuan, H., Du, W., Li, G., Xue, D., and Huang, Q. (2021) Active-Site Models of Streptococcus Pyogenes Cas9 in Dna Cleavage State. *Front. Mol. Biosci. 8*, 653262, DOI: 10.3389/fmolb.2021.653262.
(37) Sundaresan, R., Parameshwaran, H. P., Yogesha, S., Keilbarth, M. W., and Rajan, R. (2017) RNA-Independent DNA Cleavage Activities of Cas9 and Cas12a. *Cell Rep. 21*, 3728–3739, DOI: 10.1016/j.celrep.2017.11.100.
(38) Newsom, S. N., Wang, D.-S., Rostami, S., Schuster, I., Parameshwaran, H. P., Joseph, Y. G., Qin, P. Z., Liu, J., and Rajan, R. (2023) Differential Divalent Metal Binding by SpyCas9's RuvC Active Site Contributes to Nonspecific DNA Cleavage. *CRISPR J. 6*, 527–542, DOI: 10.1089/crispr.2023.0022.
(39) Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A., and Ha, T. (2016) Real-Time Observation of Dna Recognition and Rejection by the Rna-Guided Endonuclease Cas9. *Nat. Commun. 7*, 12778, DOI: 10.1038/ncomms12778.
(40) Singh, D., Wang, Y., Mallon, J., Yang, O., Fei, J., Poddar, A., Ceylan, D., Bailey, S., and Ha, T. (2018) Mechanisms of Improved Specificity of Engineered Cas9s Revealed by Single-Molecule Fret Analysis. *Nat. Sruc. Mol. Bio. 25*, 347–354, DOI: 10.1038/s41594-018-0051-7.
(41) Gong, S., Yu, H. H., Johnson, K. A., and Taylor, D. W. (2018) DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep. 22*, 359–371, DOI: 10.2139/ssrn.3155775.
(42) Raper, A. T., Stephenson, A. A., and Suo, Z. (2018) Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J. Am. Chem. Soc. 140*, 2971–2984, DOI: 10.1021/jacs.7b13047.
(43) Stephenson, A. A., Raper, A. T., and Suo, Z. (2018) Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9. *J. Am. Chem. Soc. 140*, 3743–3750, DOI: 10.1021/jacs.7b13050.
(44) Babu, K., Kathiresan, V., Kumari, P., Newsom, S., Parameshwaran, H. P., Chen, X., Liu, J., Qin, P. Z., and Rajan, R. (2021) Coordinated Actions of Cas9 Hnh and Ruvc Nuclease Domains Are Regulated by the Bridge Helix and the Target Dna Sequence. *Biochem. 60*, 3783–3800, DOI: 10.1021/acs.biochem.1c00354.
(45) Palermo, G., Miao, Y., Walker, R. C., Jinek, M., and McCammon, J. A. (2016) Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Cent. Sci. 2*, 756–763, DOI: 10.1021/acscentsci.6b00218.
(46) Palermo, G., Ricci, C. G., Fernando, A., Basak, R., Jinek, M., Rivalta, I., Batista, V. S., and McCammon, J. A. (2017) Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc. 139*, 16028–16031, DOI: 10.1021/jacs.7b05313.
(47) Palermo, G., Miao, Y., Walker, R. C., Jinek, M., and McCammon, J. A. (2017) CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc. Natl. Acad. Sci. 114*, 7260–7265, DOI: 10.1073/pnas.1707645114.
(48) Palermo, G., Chen, J. S., Ricci, C. G., Rivalta, I., Jinek, M., Batista, V. S., Doudna, J. A., and McCammon, J. A. (2018) Key Role of the Rec Lobe during CRISPR–Cas9 Activation by 'sensing', 'regulating', and 'locking' the Catalytic HNH Domain. *Q. Rev. Biophys. 51*, 1–11, DOI: 10.1017/s0033583518000070.
(49) East, K. W., Newton, J. C., Morzan, U. N., Narkhede, Y. B., Acharya, A., Skeens, E., Jogl, G., Batista, V. S., Palermo, G., and Lisi, G. P. (2019) Allosteric Motions of the CRISPR–Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc. 142*, 1348–1358, DOI: 10.1021/jacs.9b10521.
(50) Ricci, C. G., Chen, J. S., Miao, Y., Jinek, M., Doudna, J. A., McCammon, J. A., and Palermo, G. (2019) Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci. 5*, 651–662, DOI: 10.1021/acscentsci.9b00020.
(51) Zuo, Z., and Liu, J. (2020) Allosteric Regulation of CRISPR-Cas9 for Dna-Targeting and Cleavage. *Curr. Opin. Struc. Biol. 62*, 166–174, DOI: 10.1016/j.sbi.2020.01.013.
(52) Nierzwicki, L., Arantes, P. R., Saha, A., and Palermo, G. (2020) Establishing the Allosteric Mechanism in CRISPR-Cas9. *WIREs Comput. Mol. Sci. 11*, e1503, DOI: 10.1002/wcms.1503.
(53) Nierzwicki, L., and Palermo, G. (2021) Molecular Dynamics to Predict Cryo-EM: Capturing Transitions and Short-Lived Conformational States of Biomolecules. *Front. Mol. Biosci. 8*, 641208, DOI: 10.3389/fmolb.2021.641208.
(54) Saha, A., Arantes, P. R., and Palermo, G. (2022) Dynamics and Mechanisms of CRISPR-Cas9 through the Lens of Computational Methods. *Curr. Op. Struct. Biol. 75*, 102400, DOI: 10.1016/j.sbi.2022.102400.
(55) Wang, J., Arantes, P. R., Ahsan, M., Sinha, S., Kyro, G. W., Maschietto, F., Allen, B., Skeens, E., Lisi, G. P., Batista, V. S., and Palermo, G. (2023) Twisting and Swiveling Domain Motions in Cas9 to Recognize Target Dna Duplexes, Make Double-Strand Breaks, and Release Cleaved Duplexes. *Front. Mol. Biosci. 9*, 1072733, DOI: 10.3389/fmolb.2022.1072733.
(56) Casalino, L., Nierzwicki, L., Jinek, M., and Palermo, G. (2020) Catalytic Mechanism of Non-Target Dna Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal. 10*, 13596–13605, DOI: 10/gnr59h.
(57) Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J. A. (2015) A Cas9–Guide Rna Complex Preorganized for Target Dna Recognition. *Science 348*, 1477–1481, DOI: 10.1126/science.aab1452.
(58) Yoon, H., Zhao, L. N., and Warshel, A. (2018) Exploring the Catalytic Mechanism of Cas9 Using Information

Inferred from Endonuclease VII. *ACS Catal. 9*, 1329–1336, DOI: 10.1021/acscatal.8b04324.
(59) Zhao, L. N., Mondal, D., and Warshel, A. (2019) Exploring Alternative Catalytic Mechanisms of the Cas9 Hnh Domain. *Proteins: Struct., Funct., Bioinf. 88*, 260–264, DOI: 10.1002/prot.25796.
(60) Maghsoud, Y., Jayasinghe-Arachchige, V. M., Kumari, P., Cisneros, G. A., and Liu, J. (2023) Leveraging QM/MM and Molecular Dynamics Simulations to Decipher the Reaction Mechanism of the Cas9 HNH Domain to Investigate Off-Target Effects. *J. Chem. Inf. Model. 63*, 6834–6850, DOI: 10.1021/acs.jcim.3c01284.
(61) Nierzwicki, L., East, K. W., Binz, J. M., Hsu, R. V., Ahsan, M., Arantes, P. R., Skeens, E., Pacesa, M., Jinek, M., Lisi, G. P., and Palermo, G. (2022) Principles of Target Dna Cleavage and the Role of Mg2+ in the Catalysis of CRISPR–Cas9. *Nat. Catal. 5*, 912–922, DOI: 10.1038/s41929-022-00848-6.
(62) Nam, K., Shao, Y., Major, D. T., and Wolf-Watz, M. (2024) Perspectives on Computational Enzyme Modeling: From Mechanisms to Design and Drug Development. *ACS Omega 9*, 7393–7412, DOI: 10.1021/acsomega.3c09084.
(63) Nam, K. (2014) Acceleration of Ab Initio QM/MM Calculations under Periodic Boundary Conditions by Multiscale and Multiple Time Step Approaches. *J. Chem. Theory Comput. 10*, 4175–4183, DOI: 10.1021/ct5005643.
(64) Pan, X., Van, R., Epifanovsky, E., Liu, J., Pu, J., Nam, K., and Shao, Y. (2022) Accelerating Ab Initio Quantum Mechanical and Molecular Mechanical (QM/MM) Molecular Dynamics Simulations with Multiple Time Step Integration and a Recalibrated Semiempirical QM/MM Hamiltonian. *J. Phys. Chem. B 126*, 4226–4235, DOI: 10.1021/acs.jpcb.2c02262.
(65) Pan, X., Yang, J., Van, R., Epifanovsky, E., Ho, J., Huang, J., Pu, J., Mei, Y., Nam, K., and Shao, Y. (2021) Machine-Learning-Assisted Free Energy Simulation of Solution-Phase and Enzyme Reactions. *J. Chem. Theory Comput. 17*, 5745–5758, DOI: 10.1021/acs.jctc.1c00565.
(66) Lei, Y.-K., Yagi, K., and Sugita, Y. (2024) Learning QM/MM Potential Using Equivariant Multiscale Model. *J. Chem. Phys. 160*, 214109, DOI: 10.1063/5.0205123.
(67) Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2006) Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinformatics 15*, DOI: 10.1002/0471250953.bi0506s15.
(68) Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput. 7*, 2284–2295, DOI: 10.1021/ct200133y.
(69) Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput. 7*, 525–537, DOI: 10.1021/ct100578z.
(70) Case, D. A. *et al.* AMBER 2020, University of California, San Francisco. 2020.
(71) Epifanovsky, E. *et al.* (2021) Software for the Frontiers of Quantum Chemistry: An Overview of Developments in the Q-Chem 5 Package. *J. Chem. Phys. 155*, 084801, DOI: 10.1063/5.0055522.
(72) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput. 11*, 3696–3713, DOI: 10.1021/acs.jctc.5b00255.
(73) Li, Z., Song, L. F., Li, P., and Merz, K. M. (2020) Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J. Chem. Theory Comput. 16*, 4429–4442, DOI: 10.1021/acs.jctc.0c00194.
(74) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79*, 926–935, DOI: 10.1063/1.445869.
(75) Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: An N Log(n) Method for Ewald Sums in Large Systems. *J. Chem. Phys. 98*, 10089–10092, DOI: 10.1063/1.464397.
(76) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327–341, DOI: 10.1016/0021-9991(77)90098-5.
(77) Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput. 8*, 1542–1555, DOI: 10.1021/ct200909j.
(78) Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput. 9*, 3878–3888, DOI: 10.1021/ct400314y.
(79) Becke, A. D. (1988) Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A 38*, 3098–3100, DOI: 10.1103/physreva.38.3098.
(80) Lee, C., Yang, W., and Parr, R. G. (1988) Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B 37*, 785–789, DOI: 10.1103/physrevb.37.785.
(81) Becke, A. D. (1993) A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys. 98*, 1372–1377, DOI: 10.1063/1.464304.
(82) Hariharan, P. C., and Pople, J. A. (1973) The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta 28*, 213–222, DOI: 10.1007/bf00533485.
(83) Pan, X., Nam, K., Epifanovsky, E., Simmonett, A. C., Rosta, E., and Shao, Y. (2021) A Simplified Charge Projection Scheme for Long-Range Electrostatics in Ab Initio QM/MM Calculations. *J. Chem. Phys. 154*, 024115, DOI: 10.1063/5.0038120.
(84) Leimkuhler, B., Margul, D. T., and Tuckerman, M. E. (2013) Stochastic, Resonance-Free Multiple Time-Step Algorithm for Molecular Dynamics with Very Large Time Steps. *Mol. Phys. 111*, 3579–3594, DOI: 10.1080/00268976.2013.844369.
(85) Shirts, M. R., and Chodera, J. D. (2008) Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys. 129*, 124105, DOI: 10.1063/1.2978177.
(86) Pan, X., Van, R., Pu, J., Nam, K., Mao, Y., and Shao, Y. (2023) Free Energy Profile Decomposition Analysis for QM/MM Simulations of Enzymatic

Reactions. *J. Chem. Theory Comput. 19*, 8234–8244, DOI: 10.1021/acs.jctc.3c00973.

(87) Chai, J.-D., and Head-Gordon, M. (2008) Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys. 10*, 6615, DOI: 10.1039/b810189b.

(88) Li, P., Jia, X., Pan, X., Shao, Y., and Mei, Y. (2018) Accelerated Computation of Free Energy Profile at ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semi-Empirical Reference Potential. I. Weighted Thermodynamics Perturbation. *J. Chem. Theory Comput. 14*, 5583–5596, DOI: 10.1021/acs.jctc.8b00571.

(89) Hu, W., Li, P., Wang, J.-N., Xue, Y., Mo, Y., Zheng, J., Pan, X., Shao, Y., and Mei, Y. (2020) Accelerated Computation of Free Energy Profile at Ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semiempirical Reference Potential. 3. Gaussian Smoothing on Density-of-States. *J. Chem. Theory Comput. 16*, 6814–6822, DOI: 10.1021/acs.jctc.0c00794.

(90) Yin, J., Liu, M., Liu, Y., Wu, J., Gan, T., Zhang, W., Li, Y., Zhou, Y., and Hu, J. (2019) Optimizing Genome Editing Strategy by Primer-Extension-Mediated Sequencing. *Cell Discov. 5*, 18, DOI: 10.1038/s41421-019-0088-8.

(91) Wang, J., Skeens, E., Arantes, P. R., Maschietto, F., Allen, B., Kyro, G. W., Lisi, G. P., Palermo, G., and Batista, V. S. (2022) Structural Basis for Reduced Dynamics of Three Engineered HNH Endonuclease Lys-to-Ala Mutants for the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Associated 9 (CRISPR/Cas9) Enzyme. *Biochem. 61*, 785–794, DOI: 10.1021/acs.biochem.2c00127.

(92) Ray, A., and Di Felice, R. (2020) Protein-Mutation-Induced Conformational Changes of the DNA and Nuclease Domain in CRISPR/Cas9 Systems by Molecular Dynamics Simulations. *J. Phys. Chem. B. 124*, 2168–2179, DOI: 10.1021/acs.jpcb.9b07722.

(93) Nierzwicki, L., East, K. W., Morzan, U. N., Arantes, P. R., Batista, V. S., Lisi, G. P., and Palermo, G. (2021) Enhanced Specificity Mutations Perturb Allosteric Signaling in CRISPR-Cas9. *eLife 10*, e73601, DOI: 10.7554/elife.73601.

(94) Ergünay, T., Ayhan, , Celen, A. B., Georgiadou, P., Pekbilir, E., Abaci, Y. T., Yesildag, D., Rettel, M., Sobhiafshar, U., Ogmen, A., Emre, N. T., and Sahin, U. (2022) Sumoylation of Cas9 at Lysine 848 Regulates Protein Stability and Dna Binding. *Life Sci. Alliance 5*, e202101078, DOI: 10.26508/lsa.202101078.

Supplementary Information for

Exploring CRISPR-Cas9 HNH-Domain Catalyzed DNA Cleavage Using Accelerated Quantum Mechanical Molecular Mechanical Free Energy Simulation

Richard Van,[†,‡] Xiaoliang Pan,[†] Saadi Rostami,[†] Jin Liu,[§] Pratul K. Agarwal,[‖] Bernard Brooks,[*,‡] Rakhi Rajan,[*,†] and Yihan Shao[*,†]

†*Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Pkwy, Norman, OK 73019, United States.*
‡*Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States.*
§*Department of Pharmaceutical Sciences, University of North Texas System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, TX 76107, United States.*
‖*High Performance Computing Center, Oklahoma State University, 106 Math Sciences, Stillwater, OK 74078, United States.*

E-mail: brb@nih.gov; rrajan@ou.edu; yihan.shao@ou.edu

Contents

Supplementary Methods

Pre-Catalytic State Models

The initial coordinates for the pre-catalytic state of *Streptococcus pyogenes* Cas9 (SpyCas9) were based on a post-catalytic state cryo-EM structure by Jinek and coworkers (PDB 7Z4J). Resolved at 2.99 Å resolution, this structure captures SpyCas9 bound to 18-nucleotide complementary DNA substrate and sgRNA. Both active sites contain a metal ion, where the HNH domain is seen just after tDNA cleavage, while RuvC catalysis resulted in the ntDNA 3′-product being unresolved. Since ntDNA cleavage requires full-length ntDNA, missing nucleotides were modeled based on the ntDNA backbone coordinates of PDB 5F9R (3.40 Å resolution). The nucleotide sequence was corrected, and missing atoms were added in the leap. Since both DNA strands were obtained after phosphodiester bond cleavage, the O3′-P bond in the catalytic states was reconstructed following minimization. The O3′ (chain C) atom served as the initial coordinates for nucleophilic water and was manually renamed to a water residue (WAT).

The K866A variant was prepared from the pre-catalytic wildtype SpyCas9. The mutation was introduced by manually deleting side-chain atoms of K866, and backbone atoms were renamed to alanine (ALA). Then, tleap was used to add missing side chain atoms.

Quantum Mechanics/Molecular Mechanics (QM/MM) Simulations

In QM/MM MD simulations, a quantum mechanical (QM) description is used for atoms involved in the chemical reaction, while point-charge classical force fields describe the rest of the system. This multi-scale approach allows the QM subsystem to be dynamically explored by integrating the equations of motion for the electronic degrees of freedom concerning time while considering the contributions of the environment (enzyme and solvation effects). Coupled with free energy methods, constructing a free energy profile can aid in determining the free energy barrier of the chemical step and associated rate constants (within transition state theory).

QM/MM simulations were conducted at 300K under the NVT ensemble to simulate the DNA cleavage mechanism of the wild type and the K866A mutant. The QM region consisted of the coordinating residues (D839, N863), general-base (H840), DNA nucleotides (G3, T4), Mg^{2+} ions, and three water molecules (2 coordinating to Mg^{2+}; 1 nucleophilic). The valence of the terminal QM atoms was saturated with capping hydrogen atoms, resulting in a total of 73 atoms (68 QM atoms + 5 capping hydrogen atoms). The QM part was described with the B3LYP functional and 6-31G* basis set, while the rest of the system was described as MM charges within classical force fields. The particle mesh Ewald method was used to describe the MM-MM electrostatic interactions, while van der Waals interactions were handled with a cutoff of 10 Å. The QM/MM-AC method was used to capture the long-range QM/MM electrostatic interactions efficiently (by projecting outer MM charges more than 10 Å away from the QM region onto inner MM atoms).

The multiple time step (MTS) protocols used a 1 fs step for the inner steps (at PM3/MM level), while a 4 fs step for the outer steps (at B3LYP/6-31G(d) level). QM/MM MD simulations were performed using our in-house QM/MM interface, QMHub, a modified version of the SANDER program from AmberTools20, and Q-Chem 5.2 for DFT/MM calculations.

Umbrella sampling was used to model the H840-mediated DNA cleavage reaction of wildtype and K866A variants. This approach uses a reaction coordinate to bias the system's potential along independent simulation windows. The mechanism was studied by defining the reaction coordinate as the difference distance between the atoms breaking atoms $O3'_{DNA} - P_{DNA}$ and bond-forming atoms $P_{DNA} - O_{WAT}$. A total of 42 windows were evenly distributed in intervals of 0.10 Å to cover the range of

-1.90 Å to 2.20 Å. The reactant state is found at -1.90 Å, where the distance of $O3'_{DNA} - P_{DNA}$ atoms is shorter than the distance between $P_{DNA} - O_{WAT}$. A harmonic potential was set for each window with a force constant of 150 kcal/mol. Each umbrella window was simulated for 50 ps, and configurations were saved every 0.5 ps, resulting in a total of 21,000 configurations for each system. The PyMBAR package calculated the free energy profile using the Multistate Bennett acceptance ratio (MBAR) method.

Energy Decomposition of the Free Energy Profile

Free energy contributions from each residue can be obtained by post-processing ai-QM/MM free energy simulation trajectories. Specifically, the electrostatic, polarization, and van der Waals contributions from each residue to the free energy barrier can be obtained by mean force integration along the minimum accessible energy pathway.

Energy Decomposition of the QM/MM free energy profile of wildtype Cas9 provided the interaction energy contributions, which were further divided into individual residue contributions. Configurations were used in thermodynamic integration to generate a free energy profile. The two analyses show nearly identical free energy profiles within statistical uncertainty (Figure S5a). The TI-integrated QM/MM free energy profile was divided into contributions from QM/MM energy terms (i.e., gas-phase, permanent electrostatics, polarization, and van der Waals). The corresponding contributions to the reaction-free energy barrier and reaction-free energy are shown in Figure S5b. These free energy components were further divided into each MM atom. Residues then grouped the atomic contributions from the enzyme. At the same time, the ones from the solvent and counterions were grouped by their distances to the QM region into near- and far-solvent contributions. The residues that contributed to the free energy barrier by at least 0.3 kcal/mol were listed in Table S1.

Supplementary Figures

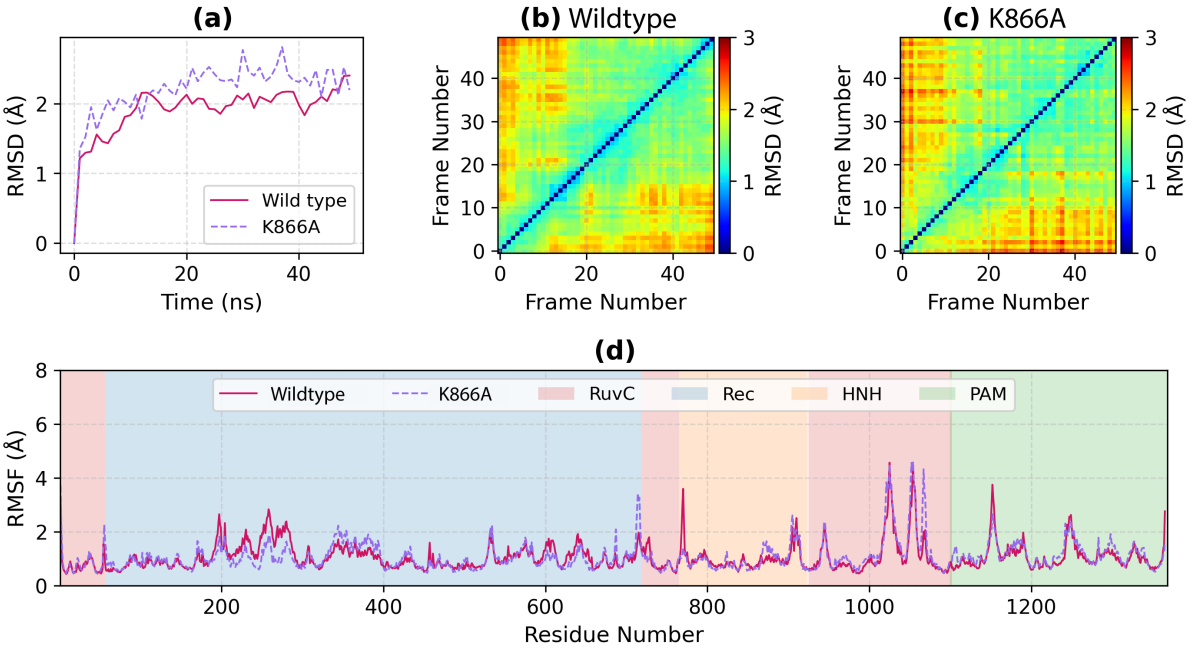

Figure S1: (a) 1-D RMSD, (b, c) 2-D RMSD, and (d) RMSF fluctuations along classical 50 ns MD simulation trajectories of wildtype Cas9 and K866A enzymes in their pre-catalytic configurations in complex with gRNA and DNA.

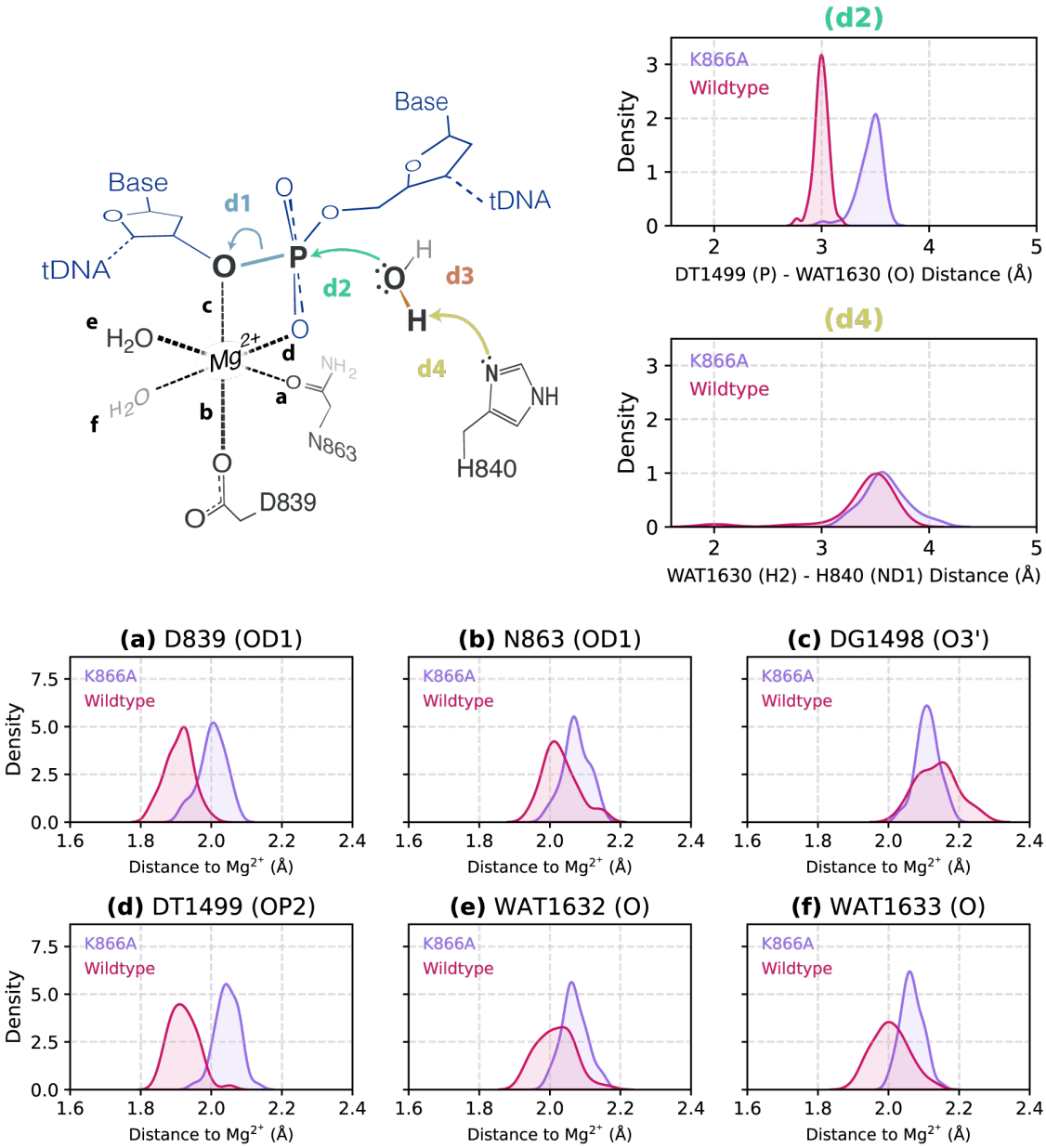

Figure S2: Key distances in the HNH subdomain of the reactant complex. Distribution of d2, d4 and P–N_δ(H840) distances. (a-f) Distribution of distance Mg^{2+} to its ligands.

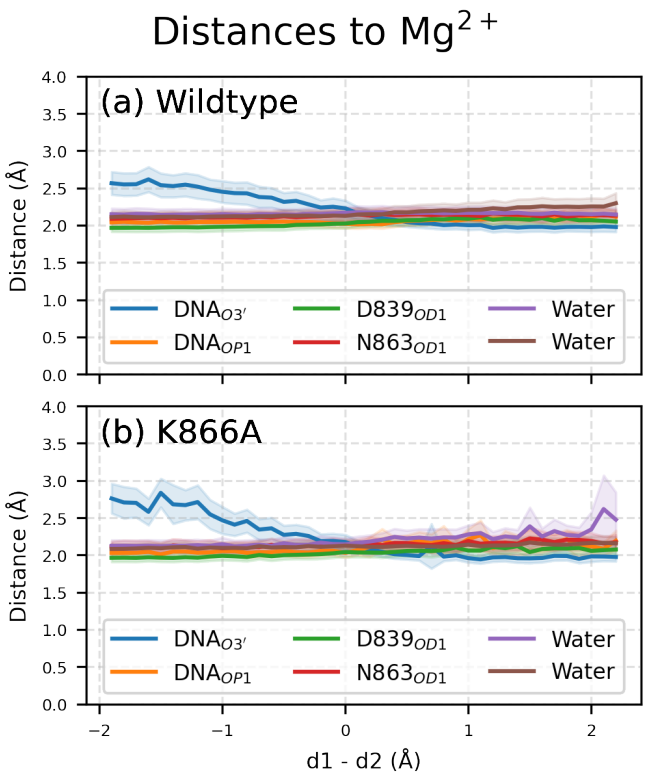

Figure S3: Key Mg^{2+}-coordination distances along the reaction coordinate (d1 - d2) for, (a) wildtype, and (b) K866A variant.

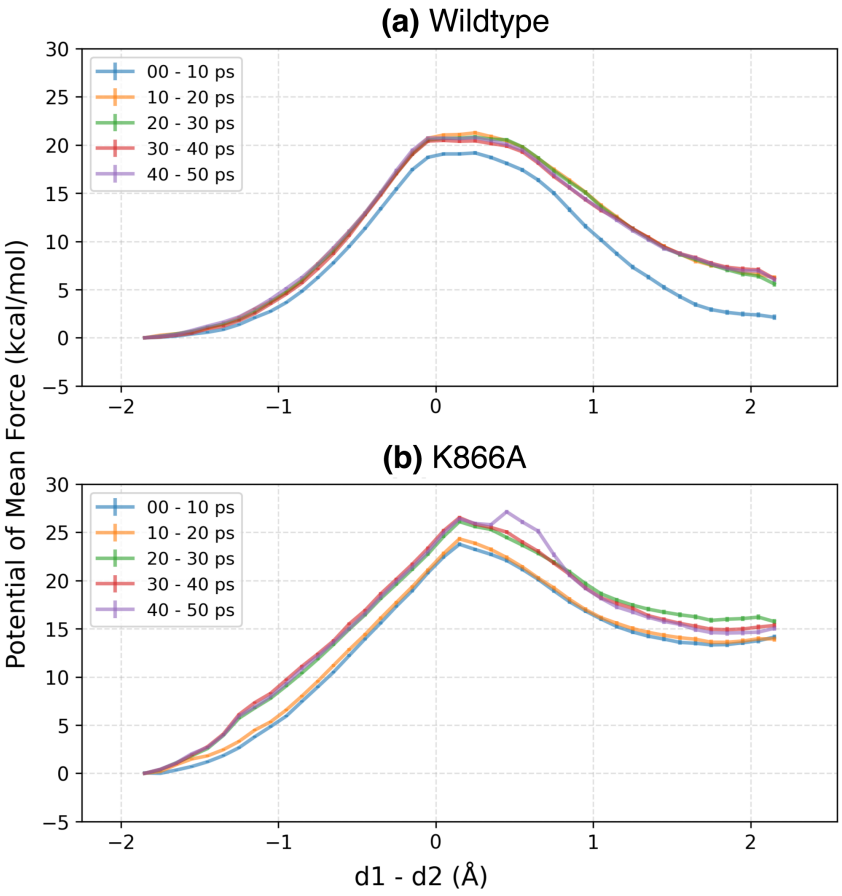

Figure S4: Convergence of the QM/MM free energy profile for (a) wildtype SpyCas9 and (b) K866A variant.

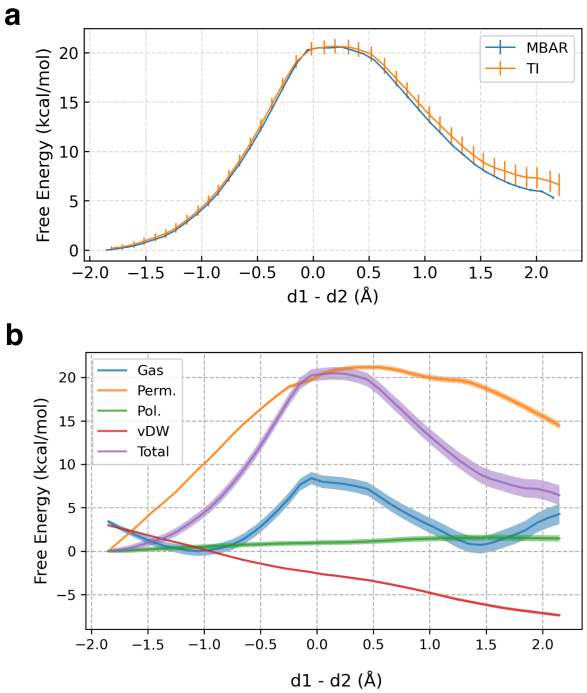

Figure S5: Energy decomposition analysis of the free energy profile shows (a) TI-calculated free energy profile is nearly identical to MBAR, and (b) energy contributions.

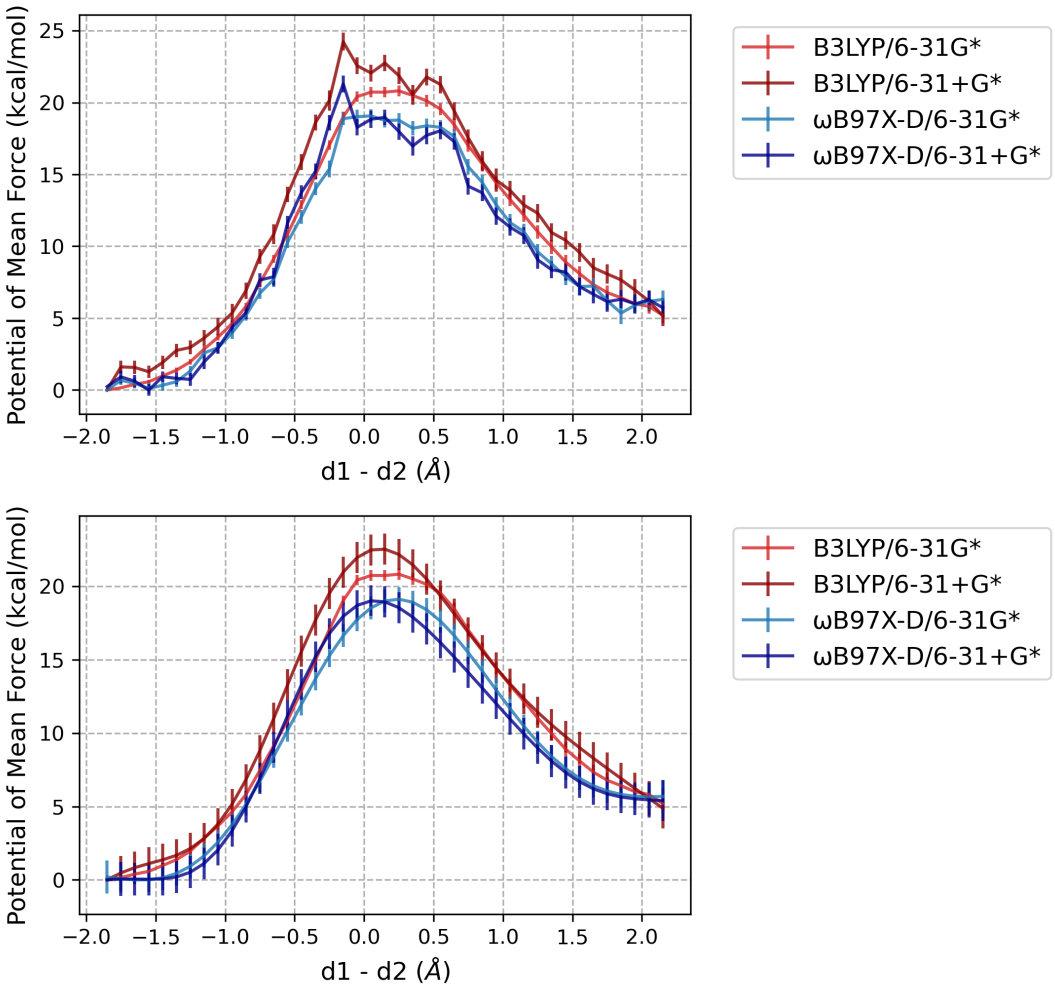

Figure S6: Raw and smoothened free energy profiles from weighted thermodynamic perturbation calculations for the wildtype SpyCas9 catalyzed tDNA cleavage. The smoothened profiles are the same in Figure 6.

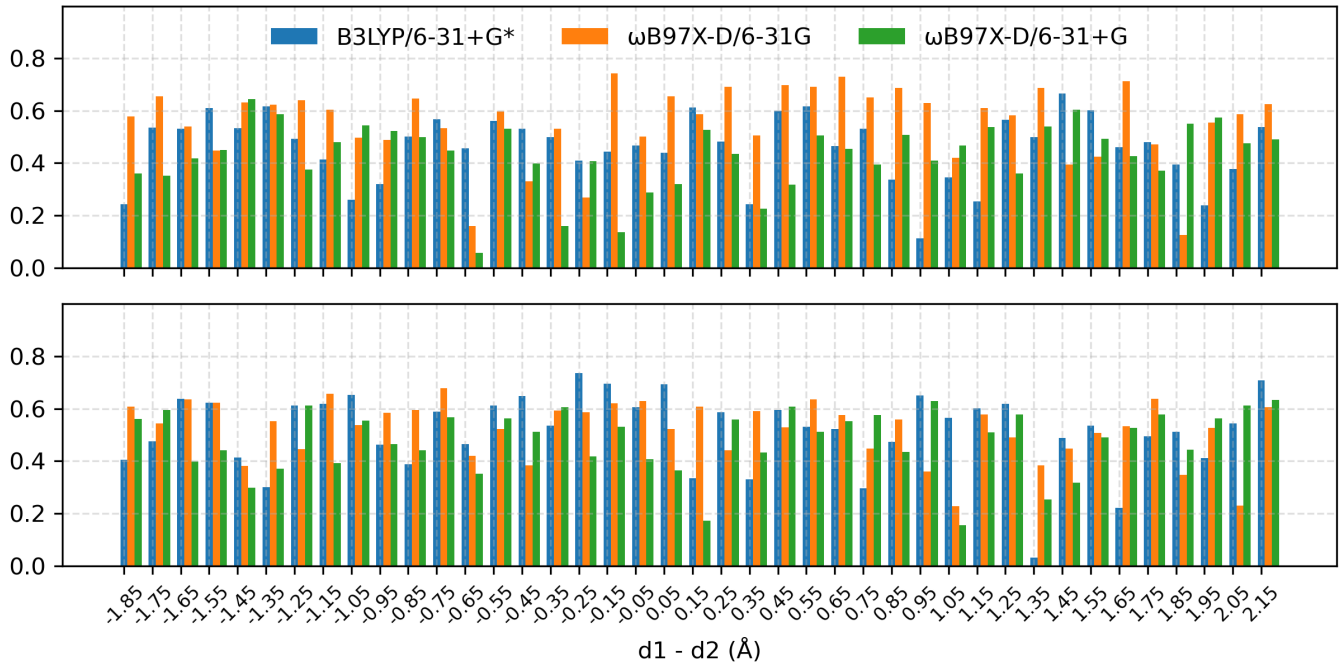

Figure S7: Reweighting entropies for weighted thermodynamic perturbations to the reaction free energy profiles of (top) wildtype SpyCas9 and (bottom) K866 variant.

Supplemental Tables

Table S1: Decomposition of the Free Energy Profile for Individual Residue Contribution (kcal/mol). Residues that changed the free energy barrier by at least 2.0 kcal/mol are listed.

Residues	$E^{perm.}_{\mathrm{QM-MM}}$	$E^{pol.}_{\mathrm{QM-MM}}$	$E^{vDW.}_{\mathrm{QM-MM}}$	E_{Total}
D850	-10.47 ± 0.03	1.38 ± 0.02	0.00 ± 0.00	-9.09 ± 0.02
E809	-7.32 ± 0.03	-0.18 ± 0.03	0.00 ± 0.00	-7.51 ± 0.03
R864	-6.60 ± 0.04	0.33 ± 0.02	0.00 ± 0.00	-6.27 ± 0.04
V838	-7.34 ± 0.06	2.05 ± 0.03	-0.95 ± 0.04	-6.24 ± 0.07
D849	-6.68 ± 0.03	0.98 ± 0.01	0.00 ± 0.00	-5.70 ± 0.02
K866	-6.39 ± 0.05	1.01 ± 0.03	0.00 ± 0.00	-5.38 ± 0.04
E60	-3.05 ± 0.02	0.21 ± 0.01	0.00 ± 0.00	-2.83 ± 0.01
R71	-2.90 ± 0.01	0.18 ± 0.01	0.00 ± 0.00	-2.72 ± 0.01
K401	-2.30 ± 0.00	0.06 ± 0.00	0.00 ± 0.00	-2.24 ± 0.00
D853	-2.39 ± 0.01	0.24 ± 0.00	0.00 ± 0.00	-2.16 ± 0.01
R403	-2.21 ± 0.01	0.13 ± 0.00	0.00 ± 0.00	-2.08 ± 0.00
R400	-2.21 ± 0.01	0.17 ± 0.00	0.00 ± 0.00	-2.04 ± 0.01
E396	2.25 ± 0.01	-0.23 ± 0.00	0.00 ± 0.00	2.02 ± 0.01
R919	2.40 ± 0.01	-0.28 ± 0.00	0.00 ± 0.00	2.12 ± 0.01
D406	2.50 ± 0.01	-0.16 ± 0.00	0.00 ± 0.00	2.34 ± 0.01
D861	2.54 ± 0.01	-0.18 ± 0.01	0.00 ± 0.00	2.36 ± 0.01
G1366	2.63 ± 0.01	-0.27 ± 0.00	0.00 ± 0.00	2.36 ± 0.01
R780	2.90 ± 0.01	-0.16 ± 0.00	0.00 ± 0.00	2.74 ± 0.00
D868	3.35 ± 0.01	-0.22 ± 0.01	0.01 ± 0.00	3.13 ± 0.01
K918	3.63 ± 0.01	-0.47 ± 0.01	0.00 ± 0.00	3.16 ± 0.01
K896	4.17 ± 0.01	-0.38 ± 0.00	0.00 ± 0.00	3.79 ± 0.01
K848	4.54 ± 0.02	-0.42 ± 0.01	0.00 ± 0.00	4.12 ± 0.02
K810	4.49 ± 0.02	0.07 ± 0.01	0.00 ± 0.00	4.56 ± 0.02
K855	10.53 ± 0.03	-1.01 ± 0.02	0.00 ± 0.00	9.51 ± 0.02

Table S2: Decomposition of the Free Energy Profile for Individual DNA Contribution (kcal/mol). Residues that changed the free energy barrier by at least 1.0 kcal/mol are listed.

Residues	$E^{perm.}_{\mathrm{QM-MM}}$	$E^{pol.}_{\mathrm{QM-MM}}$	$E^{vDW.}_{\mathrm{QM-MM}}$	E_{Total}
C1500	-14.84 ± 0.07	0.07 ± 0.04	0.37 ± 0.01	-14.39 ± 0.06
T1501	-2.11 ± 0.01	-0.01 ± 0.01	0.03 ± 0.00	-2.09 ± 0.01
C1505	-1.38 ± 0.00	0.10 ± 0.00	0.00 ± 0.00	-1.28 ± 0.00
T1504	-1.21 ± 0.00	0.07 ± 0.00	0.00 ± 0.00	-1.14 ± 0.00
T1507	-1.15 ± 0.00	0.11 ± 0.00	0.00 ± 0.00	-1.05 ± 0.00
C1502	-1.01 ± 0.01	0.03 ± 0.00	-0.02 ± 0.00	-1.00 ± 0.00
G1496	1.47 ± 0.00	-0.08 ± 0.00	0.00 ± 0.00	1.39 ± 0.00
A1495	1.49 ± 0.00	-0.06 ± 0.00	0.00 ± 0.00	1.43 ± 0.00
C1497	1.84 ± 0.01	-0.18 ± 0.01	0.00 ± 0.00	1.67 ± 0.01
G1498	4.94 ± 0.04	-0.52 ± 0.02	0.00 ± 0.00	4.42 ± 0.03
T1499	7.94 ± 0.04	-0.72 ± 0.03	0.00 ± 0.00	7.22 ± 0.03

Table S3: Decomposition of the Free Energy Profile for Individual gRNA Contribution (kcal/mol). Residues that changed the free energy barrier by at least 1.0 kcal/mol are listed.

Residues	$E^{perm.}_{\mathrm{QM-MM}}$	$E^{pol.}_{\mathrm{QM-MM}}$	$E^{vDW.}_{\mathrm{QM-MM}}$	E_{Total}
A1381	-4.34 ± 0.01	0.48 ± 0.01	0.00 ± 0.00	-3.86 ± 0.01
G1380	-3.65 ± 0.01	0.44 ± 0.01	0.00 ± 0.00	-3.21 ± 0.01
A1379	-3.50 ± 0.01	0.34 ± 0.00	0.00 ± 0.00	-3.16 ± 0.01
A1378	-2.23 ± 0.00	0.17 ± 0.00	0.00 ± 0.00	-2.06 ± 0.00
U1382	-1.99 ± 0.01	0.21 ± 0.00	0.00 ± 0.00	-1.78 ± 0.01
A1377	-1.49 ± 0.00	0.10 ± 0.00	0.00 ± 0.00	-1.40 ± 0.00
G1383	-1.34 ± 0.01	0.10 ± 0.00	0.00 ± 0.00	-1.24 ± 0.00
U1376	-1.13 ± 0.00	0.07 ± 0.00	0.00 ± 0.00	-1.05 ± 0.00
C1387	1.11 ± 0.01	-0.10 ± 0.00	0.00 ± 0.00	1.02 ± 0.01
C1389	1.34 ± 0.01	-0.13 ± 0.01	0.00 ± 0.00	1.21 ± 0.01
A1416	1.32 ± 0.00	-0.09 ± 0.00	0.00 ± 0.00	1.23 ± 0.00
U1393	1.40 ± 0.00	-0.08 ± 0.00	0.00 ± 0.00	1.32 ± 0.00
U1414	1.58 ± 0.00	-0.10 ± 0.00	0.00 ± 0.00	1.48 ± 0.00
A1415	1.67 ± 0.00	-0.10 ± 0.00	0.00 ± 0.00	1.57 ± 0.00
A1420	1.69 ± 0.01	-0.08 ± 0.00	0.00 ± 0.00	1.62 ± 0.00
U1392	2.16 ± 0.00	-0.12 ± 0.00	0.00 ± 0.00	2.04 ± 0.00
U1391	3.13 ± 0.01	-0.14 ± 0.00	0.00 ± 0.00	2.99 ± 0.01
G1388	3.34 ± 0.01	-0.26 ± 0.01	0.00 ± 0.00	3.07 ± 0.01
G1390	4.73 ± 0.01	-0.25 ± 0.01	0.00 ± 0.00	4.48 ± 0.01

TOC Graphic