ACS OMEGA

Open Access

This article is licensed under CC-BY-NC-ND 4.0

http://pubs.acs.org/journal/acsodf

Article

Probing Electrostatic Interactions in DNA-Bound CRISPR/Cas9 Complexes by Molecular Dynamics Simulations

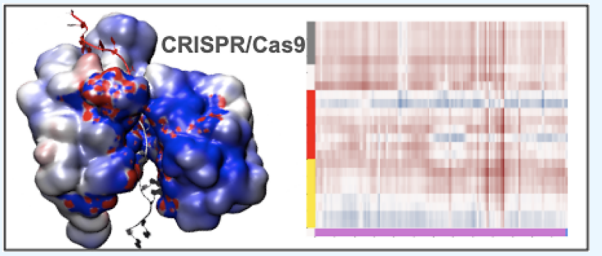
Seyedeh Hoda Fatemi Abhari* and Rosa Di Felice*

Cite This: *ACS Omega* 2024, 9, 44974−44988

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Engineered protein mutations may be exploited to tune molecular interactions in the cellular environment. Here, we have explored the structural consequences of different Cas9 mutations in genome-editing CRISPR/Cas9 systems by means of Molecular Dynamics simulations. We have characterized mutation-induced structural changes and their implications for changes in protein−DNA, DNA−RNA, and DNA−DNA interactions. We present the analysis of multiple trajectories over the cumulative time scale of 7.7 μs, focusing on triple mutations that have been associated with enhancement of genome editing specificity, as well as control mutations. We find that the structural changes induced by the protein mutations are consistent with decreasing the strength of the interaction between Cas9 and the nontarget DNA strand. We discuss the implications of this finding for genome editing specificity.



INTRODUCTION

The CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein) system is an adaptive immune system found in bacteria that provides defense against invading viruses. Its discovery as a programmable nucleotide manipulation method has revolutionized the field of gene editing, leading to Emmanuelle Charpentier and Jennifer Doudna being awarded the 2020 Nobel Prize in Chemistry. The CRISPR/Cas system is categorized into two classes, namely Class 1 and Class 2, with Class 2 further divided into types II, V, and VI. A Class 1 system consists of multiple Cas proteins that form an interference complex, allowing for coordinated binding and processing of the target. In contrast, a Class 2 system is characterized by a single, multidomain CRISPR RNA (crRNA)-binding protein that performs functions analogous to the entire effector complex found in Class 1 systems.[1]

CRISPR/Cas9 falls into Class 2, specifically type II. It consists of a molecular structure comprising a single guide RNA (sgRNA) and a Cas9 protein, which functions as a 160 kDa DNA-endonuclease enzyme.[2] The CRISPR/Cas9 system includes two essential noncoding RNA components, the CRISPR RNA (crRNA) and the transactivating crRNA (tracrRNA), which concomitantly form the fused single guide RNA (sgRNA). This sgRNA is crucial in guiding the Cas9 protein and facilitating its proper folding into an active conformation. Within the sgRNA, the tracrRNA module serves as a scaffold that interacts with the Cas9 protein, while the crRNA module provides the sequence specificity required for target DNA cleavage.[3−5] The crRNA consists of a 20-nucleotide (-nt) long "spacer" or "guide" sequence at the 5′-end and a "repeat" sequence at the 3′-end. The spacer sequence is complementary to the target DNA site, and the repeat segment forms a duplex with tracrRNA.[6] The Cas9 protein, in coordination with the sgRNA, recognizes a specific DNA sequence by the protospacer adjacent motif (PAM), a trinucleotide sequence found immediately downstream of the target site. The presence of a PAM sequence is crucial for the binding and initiation of the DNA cleavage process.[7]

Upon recognition of the target DNA sequence and PAM, the CRISPR/Cas9 system forms an R-loop structure. The R-loop consists of a hybrid duplex formed by the target DNA strand (hereby labeled as tDNA) and the RNA strand of the sgRNA, while the nontarget DNA strand (hereby labeled as ntDNA) is displaced.[8−10] The formation of the R-loop allows the Cas9 nuclease domain to cleave both strands of the double-stranded DNA (dsDNA), inducing a double-strand break (DSB). This DSB triggers the cellular DNA repair mechanisms, leading to DNA modifications, such as gene knockout, insertion, or other edits, depending on the downstream repair pathways.[11,12]

The structural analysis of the CRISPR/Cas9 system has been greatly facilitated by the atomic resolution of crystal structures of *Streptococcus pyogenes* Cas9 in complex with RNA and DNA. The Cas9 protein consists of two main lobes: the recognition (REC) lobe and the nuclease (NUC) lobe. The REC lobe, composed of residues 56−718, is responsible for binding to the

Received: May 7, 2024
Revised: October 16, 2024
Accepted: October 21, 2024
Published: October 30, 2024

ACS Publications

© 2024 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

sgRNA, while the NUC lobe encompasses the RuvC domain (residues 1−55, 719−765, and 919−1098), the HNH domain (residues 780−906), and the PAM-interacting (PI) domain (residues 1099−1368).[13]

The CRISPR/Cas9 system is in charge of scanning DNA sequences in search for the PAM, which is specific for a given CRISPR-associated protein. The PAM sequence is 3′-NGG-5′ for Cas9, where N is any base. It is specifically 3′-TGG-5′ in most CRISPR/CAs9 resolved atomic structures available in the protein data bank. Once the PAM is identified, the crRNA binds, by nucleic acid complementary recognition, the target DNA site upstream of the PAM, and Cas9 performs the cleavage of the dsDNA. Specifically, the HNH and RuvC domains of Cas9 perform the cleavage of the tDNA and ntDNA strands, respectively.

Within the Cas9/RNA/DNA complex, the negatively charged RNA/DNA hybrid duplex inserts itself into a positively charged groove located at the interface between the NUC and REC lobes of the Cas9 protein. Notably, the 3′ end of the sgRNA interacts with a positive patch found in the interface between the RuvC and PAM-interacting (PI) domains. Additionally, the positively charged interface between the HNH and RuvC domains can accommodate the unwound portion of the nontarget DNA strand (hereby labeled as untDNA) through electrostatic interactions with the negatively charged backbone of the DNA.[14] Sternberg et al.[15] also reported a correlation between structural transitions in the HNH domain and cleavage of the ntDNA strand. Moreover, insights into the dynamics of Cas9 have been gained through single-molecule Forster Resonance Energy Transfer (FRET) experiments, which revealed the intrinsic flexibility of the HNH domain, which allows it to access the DNA cleavage site.[16,17] In an effort to further comprehend the cleavage mechanism of Cas9, several computational methods, including the empirical valence bond (EVB) method,[18,19] classical and mixed quantum-classical molecular simulations[20−25] have been employed.

Protein engineering to tune electrostatic interactions can play a role in overcoming a current major shortcoming of gene-editing with CRISPR/Cas9, namely the limited specificity. In fact, although this technique has revolutionized basic and applied research since its discovery in 2012,[3,26,27] it is prone to error. One of its major limitations lies in the off-target cleavage of DNA, where the Cas9 enzyme can cleave DNA even when the target DNA sequence does not perfectly match the designed crRNA sequence.[28,29] This drawback significantly hampers the application of CRISPR/Cas9. In 2016, Kleinstiver et al. demonstrated that specific mutations (N497A, R661A, Q695A, and Q962A) in the spCas9 protein reduce off-target binding. Based on this observation, they proposed that achieving a high-fidelity CRISPR/Cas9 system could be possible through precise mutations in spCas9.[30] In the same year, Slaymaker et al. reported an effective enhancement of fidelity in spCas9 proteins in which mutations were inserted at specific sites, based on the analysis of the general structure of Cas9.[14] The authors identified a conserved patch of positive residues in the crystal structure of Cas9 (PDB ID: 4UN3) and suggested that this region could accommodate the untDNA through protein−DNA electrostatic interactions. They further hypothesized that replacing the positive residues with neutral residues in the HNH/RuvC interface, where the untDNA resides, could reduce the electrostatic attraction and require more Watson−Crick base pairing between the tDNA and the sgRNA to prevent rehybridization of the tDNA and ntDNA strands. To test their hypothesis, Slaymaker and colleagues engineered 30 spCas9 mutants by replacing the positively charged residues at the HNH/RuvC boundary with neutral alanine residues and evaluated the specificity of each mutant. They identified the top five single mutations (R780A, K810A, K848A, K855A, and H982A) with improved specificity. They also identified effective multiple mutations. We highlight that most approaches explored to improve cleavage specificity are based on different criteria, either targeting the designed sgRNA or other portions of Cas9. For instance, Akcakaya et al.[31] developed a method called verification of in vivo off-target (VIVO): by proper sgRNA design, they could achieve high specificity, resulting in no observed off-target cleavage. Alternatively, higher specificity of Cas9 was attained by impairing DNA cleavage through the combined mutation L64P+K65P within the loop region of the arginine-rich bridge helix.[32] Despite these significant advancements, further investigation is still needed to achieve the ultimate specificity of the CRISPR/Cas9 method for more reliable and programmable gene-editing applications. In this scenario, protein mutations that modify electrostatic balance remain a valid strategy toward CRISPR/Cas9 specificity.

Molecular Dynamics (MD) is a powerful computational tool that offers valuable insights into the atomic-level structures, dynamics, and energetics of biomolecules. It has emerged as a crucial technique in the study of CRISPR/Cas9,[25] enabling researchers to unravel its intricate features and mechanisms. By employing MD simulations, various aspects of CRISPR/Cas9 have been elucidated, such as the contribution of the nontarget DNA strand in activating the nuclease activity of the HNH domain, which is responsible for cleaving the target DNA strand.[33] Furthermore, MD simulations provide a platform to investigate the recognition process between Cas9 and the single guide RNA, shedding light onto the intricate interplay between these components.[33] In addition to elucidating the nuclease activation and recognition processes, MD simulations allow for the exploration of the allosteric motions of the HNH domain. These dynamics play a pivotal role in the overall function of CRISPR/Cas9. Indeed, the PAM-induced allosteric mechanism, which regulates the activity of Cas9, has been studied in atomic detail using MD simulations.[34] This mechanism provides crucial insights into how the presence of PAM sequences influences the behavior and function of CRISPR/Cas9.

Off-target effects of CRISPR/Cas9 can also be investigated through MD simulations.[22,24] By analyzing the interaction fingerprints between Cas9 and different DNA sequences, including nontarget sites, MD simulations can reveal the potential for unintended DNA cleavage. By simulating CRISPR/Cas9 bound complexes with wild-type and mutated Cas9, one could monitor mutation-induced structural effects that are proxies of interaction changes.

We carried out MD simulations of six distinct DNA-bound CRISPR/Cas9 complexes, over a cumulative simulation time of 6 μs. The six R-loop complexes with wild-type (wt) and mutant Cas9 were selected based on the hypothesis proposed by Slaymaker et al.[14] that specificity can be improved by tuning ntDNA−protein electrostatic interactions. Our goal was to unravel the effects of protein mutations at the atomic level, particularly focusing on the HNH/RuvC interface and the conformation of the unwound ntDNA strand. By monitoring the dynamics and structural changes within these regions, we gained valuable insights into the functional consequences of specific mutations. The five simulated DNA-bound CRISPR/Cas9 complexes with mutant Cas9 included the following mutations:

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

K810A+K1003A+R1060A, K848A+K1003A+R1060A, R780A-K810A-R978A, R780a-k848A-R978A, and K1059A. A visual representation of these mutation-targeted residues is shown in Figure S1 to illustrate their relative positions with respect to the ntDNA. These complexes were selected based on their potential impact on the function and specificity of CRISPR/Cas9.[14] In fact, experimental findings showed that the first two triple mutations exhibited no off-target cleavage activity on selected genes, while maintaining on-target cleavage efficiency. The other mutations explored by us are viable combinations identified in the same experimental work. This study is inspired by our previous investigation[22] of CRISPR/Cas9 complexes with single and double Cas9 mutations, but reaches a deeper impact. In fact, Ray and Di Felice targeted CRISPR/Cas9 complexes with single mutations K855A and H982A and with the combination double mutation K855A+H982A.[14] Instead, in this study we target the triple mutations with the highest cleavage specificity score,[14] along with the needed control complexes.

We present data that portray the intricate molecular mechanisms underlying CRISPR/Cas9 physical properties. Specifically, we identify mutation-induced motions that imply the decrease of protein−DNA electrostatic interaction strength for protein mutants with respect to the wild-type system. This evidence is in line with using protein mutants to change interaction balance so that off-target effects are eventually avoided. Our results contribute to the ongoing efforts to enhance the precision, reliability, and programmability of CRISPR/Cas9 for gene-editing applications.

MATERIALS AND METHODS

Materials: Structure Models of Wild-Type and Mutant CRISPR/Cas9 Complexes. Crystal structures of CRISPR/Cas9 complexes, in various combinations of protein and nucleic acids, are available in the RCSB-PDB database. These structures provide a wealth of information about the conformational changes that occur in Cas9 upon nucleic acid binding, which are essential for elucidating the system's function. For this study, we focused on the crystal structure of the CRISPR/Cas9/DNA ternary complex with PDB ID 4UN3, which was released in 2014.[35] This structure offers valuable insights into the arrangement of the R-loop in the CRISPR/Cas9 system and the significance of PAM recognition for the genome editing function. This structure contains a catalytically inactive Cas9 protein with the H840A mutation. The distance between the nuclease active site and the cleavage site of the DNA in this structure is approximately 30 Å. Specifically, this is measured as the distance between the C_α atom of residue A840 and the scissile P atom of the tDNA strand.

The ntDNA strand in the 4UN3 crystal structure contains only three nucleotides upstream of the PAM section of the DNA, meaning that the unwound ntDNA strand is almost totally absent, due to the difficulty in resolving that portion because of its high flexibility. The upstream ntDNA is significant for the investigation of protein−DNA electrostatic interactions. In fact, it is the untDNA that would be attracted to the positively charged patch at the HNH/RuvC interface. However, no intact untDNA is present in any available PDB structures of Cas9, which presents a significant obstacle to our study. Most deposited structures resolved by either X-ray crystallography or cryogenic electron microscopy (cryo-EM) contain just 0 to 4 nucleotides upstream of the PAM. Only two structures contain a longer ntDNA strand with part of the unwound segment. The cryo-EM structure with PDB ID 5Y36, deposited at the end of 2017 with resolution 5.20 Å, contains 22 ntDNA nucleotides upstream of the AGG PAM. The X-ray structure with PDB ID 5F9R, deposited at the beginning of 2016 with resolution 3.40 Å, contains 9 ntDNA nucleotides upstream of the TGG PAM. When we started this line of research to address electrostatic interactions in CRISPR/Cas9 complexes in response to the experimental work by Slaymaker and co-workers published in January 2016, none of these two structures was available. The X-ray structure with 9 ntDNA nucleotides upstream of the PAM became available soon after. However, the limited length of its untDNA strand does not allow to investigate all relevant ntDNA-Cas9, ntDNA−tDNA and tDNA-RNA electrostatic interactions. Therefore, we generated in silico a model of the CRISPR/Cas9-DNA complex with elongated ntDNA. This wild-type model and derived protein mutant models were used to investigate the balance of electrostatic interactions in DNA-bound CRISPR/Cas9 complexes through structural fingerprints. A general characterization of all the X-ray and cryo-EM structures deposited since 2015 in the PDB database is presented in the Supporting Information (Tables S1 and S2). In Figure 1 we specifically visualize the difference of our model, hereby labeled 4UN3(l), with the structures with PDB IDs 5Y36 and 5F9R. Figure 1 shows a superposition of the three said structures with non-negligible untDNA strands, after alignment of the HNH+RuvCIII portion of Cas9. We note that, despite the

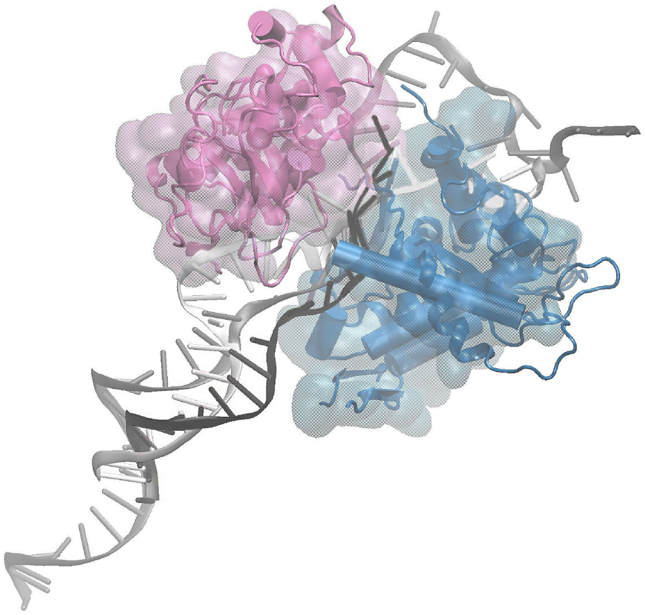


Figure 1. Visual comparison between model 4UN3(l), X-ray structure with PDB ID 5F9R and cryo-EM structure with PDB ID 5Y36, obtained with VMD,[36] after alignment of the backbone of the HNH+RuvCIII part of the complex. The portions HNH, RuvCIII and ntDNA are shown. The partial HNH+RuvCIII root-mean-square deviation of 4UN3(l) and 5F9R relative to 5Y36 is 18.13 and 9.27 Å, respectively. The ntDNA strand is shown in new cartoon drawing method in opaque black, white and gray for 5F9R (9-nt untDNA upstream of PAM), 4UN3(l) (16-nt untDNA upstream of PAM) and 5Y36 (22-nt untDNA upstream of PAM), respectively. The color mauve is used for HNH and the color blue2 is used for RuvCIII: 4UN3(l) is shown in transparent quicksurf drawing method, 5F9R is shown in opaque cartoon drawing method, 5Y36 is shown in glossy new cartoon drawing method. The image clearly shows the HNH/RuvCIII boundary that contains positive residues in wild-type Cas9, where the untDNA is accommodated in the three cases.

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

Table 1. Definition of the DNA-Bound CRISPR/Cas9 Complexes Simulated in This Study[a]

label	mutations	relative efficiency	relative specificity	simulation time (ns)
wt	N/A	reference	reference	1000
mu1	K810A+K1003A+R1060A	=	+	1000
mu2	K848A+K1003A+R1060A	=	+	1000
mu3	R780A+K810A+R976A	0	N/A	1000
mu4	R780A+K848A+R976A	0	N/A	1000
mu5	K1059A	=	=	1000

[a]The symbols = and + indicate equal and higher efficiency/specificity with respect to the wild-type complex, 0 indicates no efficiency/specificity, according to published data.[14] The numbers indicate the residue numbers that were mutated, while the letters on the left and right represent the amino acid names before and after the mutations. The rightmost column indicates the duration of the MD production run.

differences, the untDNA strand consistently spans the HNH/RuvCIII interface.

The wild-type model 4UN3(l) was derived from the crystal structure with PDB ID 4UN3.[22] Among the three RuvC motifs in the 4UN3 structure, RuvCIII is in close proximity to the HNH domain, while the other two motifs do not exhibit significant contact with HNH. Thus, we specifically modeled the elongated ntDNA in 4UN3(l) to be accommodated within the HNH/RuvCIII interface, which includes a positively charged patch of amino acids.[35] The elongated untDNA strand was modeled according to the electrostatic hypothesis,[14] that the positively charged patch in the protein would attract the negatively charged backbone of ntDNA.[14,22] Figure 1 shows that this qualitative behavior has also been detected by X-ray and cryo-EM structure determination. To model the untDNA strand in 4UN3(l) bases were added to the ntDNA in the structure with PDB ID 4UN3, using the complementarity rule to the existing tDNA, such that its shape follows the form of the HNH/RuvCIII boundary while avoiding steric clashes. The comparison of atomic distances obtained from MD trajectories and from electron paramagnetic resonance measurements confirms the reliability of the model.[37] This wild-type model, which contains a 24-mer ntDNA (16 nucleotides upstream of the TGG PAM), was used as a starting configuration for the MD time evolution of the wild-type complex. It was also used to create synthetic mutants and evolve them by MD.

The six investigated CRISPR/Cas9 systems are listed in Table 1: the wild-type complex, four Cas9-triple mutant complexes and one Cas9-single mutant complex.

To prepare the initial mutant structures for MD trajectories, we performed in silico mutagenesis on the wild-type 4UN3(l) structure using PyMOL.[38] The triple mutants *mu1−4* were generated by simultaneously mutating three amino acid residues. In these mutations, we specifically targeted the positively charged Lysines (K) and Arginines (R), substituting them with the neutral Alanine (A). Similarly, in the single mutant K1059A, positively charged Lysine was converted to neutral Alanine. These mutations were chosen to investigate the impact of selected amino acid changes on the system's conformation, which is related to function, guided by their reported effects on specificity and efficiency.[14]

Methods: Molecular Dynamics Simulations. Molecular dynamics simulations were performed using GROMACS 2018.3 and GROMACS 2020.[39] We utilized the *amber14sb:parmbsc1.ff* force field package available in the GROMACS User Contributions. This force field package includes: the Amber ff14SB protein force field[40] and the parmbsc1 force field[41−44] nucleic acids. The parmbsc1 DNA backbone correction[43] is recommended to accurately capture the behavior of the DNA backbone over time scales longer than tens of microseconds. There is no absolute consensus on the force field to be used for RNA.[45,46] The RNA remains folded and far from the HNH/RuvCIII boundary throughout the simulation time in all trajectories, so that it should not affect the electrostatic interactions between ntDNA and Cas9. It could influence the overall interaction balance, which is not the target of this study. In the Supporting Information, we present the results of a benchmarking simulation done with a different nucleic acid force field that includes the specific OL3[47] parameters for RNA and OL15[48] parameters for DNA. This control trajectory *mu1−14SB+OL15+OL3* gives a consistent layout of DNA-Cas9 distance maps (Figure S13) and consistent RNA folding (Figure S14).

There are missing Cas9 residues in the crystal structure with PDB ID 4UN3. We employed I-TASSER,[49] a protein structure prediction tool, to generate complexes with complete protein structures. The initial structures of the simulations contained eight Mg^{2+} ions, whose positions were based on the crystal structure with the PDB ID 4UN3.

Each molecular complex was prepared for simulation by explicitly solvating it with TIP3P water molecules[50] in a cubic periodic box. The box size was set to maintain a minimum distance of 1.2 nm between the solute and the box walls. To neutralize the charge of each system, Na^{+} ions were added to the simulation boxes using GROMACS tools.[39]

After the initial setup, each system was subjected to a standard minimization-equilibration protocol.[37] First, we carried out energy minimization using 50,000 steps of steepest descent. Next, we gradually heated the system in three successive 100 ps-long *NVT* steps at the temperatures of 100, 200, and 300 K. In the 100 K *NVT* equilibration step, initial velocities were generated in random directions, while for each of the following NVT equilibration steps, the velocities were read from the preceding step. After that, we performed a 100 ps-long *NPT* equilibration stage, followed by a final *NPT* production run of 1 μs (Table 1). The time step for each equilibration and production run was set to 0.002 ps. We used the particle-mesh Ewald (PME) method[51] with cubic interpolation and Fourier spacing of 0.16 nm for long-range electrostatic interactions. To maintain a temperature of 300 K and a pressure of 1 bar, we employed V-rescale[52] and Parrinello−Rahman barostat[53] methods, respectively. Replica simulations corroborate our study (Figure S13).

During each simulation, energies and coordinates of the systems were collected every 10 ps for further analysis. We employed GROMACS tools and Pymol software to process and analyze the simulation data. The effective equilibration of the systems was monitored based on the convergence of time-dependent variables. Specifically, we monitored the convergence of the time-dependent root-mean-square deviation (RMSD, eq

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

1), aiming at an oscillating time dependence with no positive or negative slope. Statistical properties of the systems were derived from the collected data after the effective equilibration of the system was achieved, as further detailed below. In eq 1, $\mathrm{RMSD}_i^{(N)}$ stands for the root-mean-square deviation on N particles at instantaneous discretized time labeled by the index i. N can be the total number of atoms in the system N_t (total RMSD) or smaller (partial RMSD). The sum over k_α runs over the N atoms in the portion of the system over which the RMSD is computed, with α spanning the corresponding subset of N_t. The sum over h runs over the three Cartesian components of the position vector $x_{k_\alpha ih}$.

$$\mathrm{RMSD}_i^{(N)} = \sqrt{\frac{\sum_{k_\alpha=1}^{N}\sum_{h=1}^{3}\left(x_{k_\alpha ih} - x_{ih}^{\mathrm{ref}}\right)^2}{N}} \quad (1)$$

To assess structural changes resulting from mutations, we examined various dynamical features, among which the hydrogen bond patterns at the interface between ntDNA and the HNH/RuvCIII boundary. In the H-bond analysis, which was performed using GROMACS, we considered a maximum distance of 0.35 nm between donor and acceptor atoms, and a maximum hydrogen-donor−acceptor angle of 30 degrees. The analysis included hydrogen, nitrogen, and oxygen atoms from both the untDNA and the protein residues of the HNH domain and RuvCIII motif of Cas9.

To shed light onto the conformational states of CRISPR-Cas9, we performed Principal Component Analysis (PCA) for the *wt* and mutated systems using GROMACS. We ran PCA separately on two different subsystems, the DNA (all atoms) and the nuclease domain of Cas9 composed of HNH and RuvCIII (backbone only). Our PCA analysis served two main purposes: (i) to discern the conformational states among the different systems and (ii) to explore the most significant conformational changes exhibited by each system throughout the MD trajectory. We employed two distinct methodologies in our investigation of DNA PCA. First, we computed the covariance matrix of the DNA subsystem based on the concatenated trajectory (6 μs) across all complexes. This approach enabled the identification of primary conformational basins associated with the DNA conformations in different systems. Second, we performed PCA from the DNA covariance matrix of the six individual trajectories (1 μs each), allowing for the detection of key conformational alterations within individual trajectories. To unravel the motions in each trajectory along the primary PC axes, PCA on each trajectory was followed by the extraction of extreme projections of essential motions. Subsequently, we interpolated 200 frames between every pair of extreme conformations using GROMACS. The outcome is illustrated through static visualization of the projections.

Shape analysis on the dsDNA was executed with Curves 5.3[54] for regular MD snapshots every 1 ns, using a computational toolkit, Trj2Shape, that can process the entire trajectory in a high-throughput manner[55] and complies with shape definitions published elsewhere.[56] The shape parameters were averaged over the last 250 ns of each trajectory.

RESULTS

To compute the RMSD from eq 1, a choice for the reference structure and for a desired portion of the system ($\{k_\alpha\}$, N) is necessary. We used the final snapshot of the equilibration stage as the reference structure for each system and calculated the RMSD values as a function of time during the 1 μs production run for different portions of the system.

The 4UN3(l) structure features an elongated ntDNA strand spanning 24 nucleotides. For the following analysis, we partitioned the ntDNA into three distinct segments, as shown in Figure 2, with different inherent flexibilities. The PAM-duplex segment encompasses nucleotides −8 to −1, and forms a duplex structure with the complementary target DNA (tDNA) strand, where nucleotides −1, − 2 and −3 are the PAM sequence. The untDNA comprises the PAM-proximal segment (nucleotides +1 to +8) and the PAM-distal segment (nucleotides +12 to +16), which is situated farther away from the PAM site. The nucleotide numbering conforms to that in our previous study.[22]

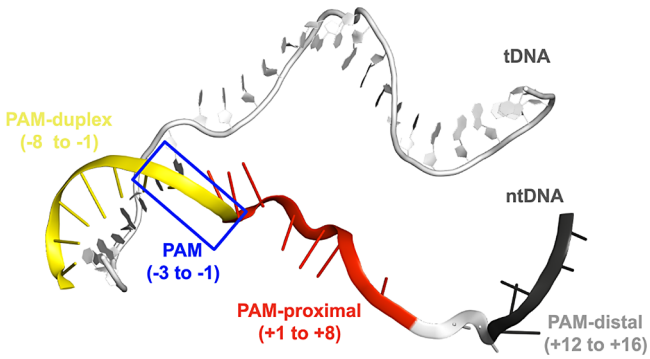


Figure 2. Visualization of the atomic structure of DNA in model 4UN3(l), highlighting the ntDNA portions used in the following analysis. The separation of DNA strands, crucial for base pairing of the tDNA with the gRNA, is initiated by Cas9 binding to a specific three-nucleotide sequence known as the PAM, denoted by the blue box. The PAM-distal region, depicted in gray, represents the more distant part of the ntDNA from PAM. Adjacent to PAM, the red section of ntDNA is the PAM-proximal region. The yellow-shaded area, termed PAM-duplex, indicates the portion of ntDNA engaged in duplex binding with the tDNA.

Mobility of the ntDNA in the Ternary Complexes. The partial RMSD as a function of time, computed on the PAM-duplex, PAM-proximal and PAM-distal portions of the ntDNA strand, is presented in Figure 3 for the wild-type and mutant ternary complexes. While the magnitude of the RMSD depends on the arbitrary choice of the reference structure (see eq 1), the saturation trend is generally consistent. Furthermore, our choice of reference structure is unbiased.

The RMSD of the PAM-duplex segment is systematically smaller and has smaller average fluctuations than those of the PAM-proximal and PAM-distal segments. This evidence points to its higher stability and rigidity, relative to the unwound part of the ntDNA strand. The RMSD of the PAM-duplex segment for all complexes has a magnitude between 0.1 and 0.2 nm, while those of the PAM-proximal and PAM-distal segments reach magnitudes larger than 0.6 nm.

The PAM-duplex RMSD maintains a stable oscillating behavior throughout the entire production run for all complexes, except a short transient for *mu1* at about 300 ns and for *mu3* at about 700 ns. The other ntDNA segments are more irregular. For the PAM-proximal segment, the RMSD is overall characterized by an initial oscillating rise, followed by an oscillating plateau, with the onset of the plateau occurring at different evolution times in different systems. For some systems, we observe more complex patterns that include transients and transitions. Specifically, for *mu2* an initial plateau between 100

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

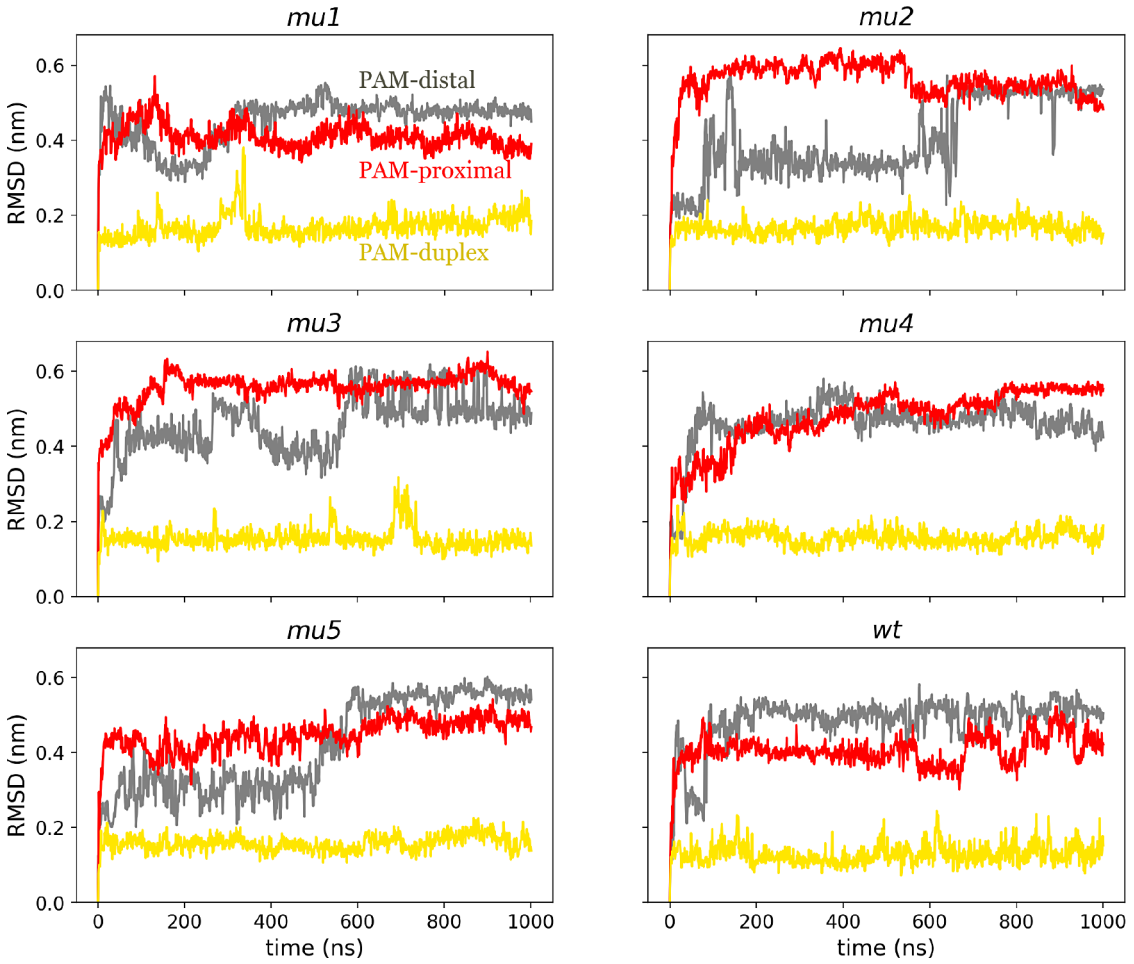


Figure 3. Partial RMSD on the PAM-duplex (yellow), PAM-proximal (red) and PAM-distal (gray) segments of the ntDNA strand, as a function of time over the entire production run of 1 μs, for the six DNA-bound CRISPR/Cas9 complexes simulated in this work.

Table 2. Time-Resolved Average Root Mean Square Deviation, Relative to the Reference Structure Specified in the Section "Materials and Methods"[a]

	PAM-duplex		PAM-proximal		PAM-distal	
	Δt (ns)	$R \pm \delta R$ (nm)	Δt (ns)	$R \pm \delta R$ (nm)	Δt (ns)	$R \pm \delta R$ (nm)
wt	100−1000	0.130 ± 0.021	100−550	0.402 ± 0.018	125−1000	0.504 ± 0.027
			575−675	0.359 ± 0.017		
			725−1000	0.430 ± 0.040		
mu1	100−250	0.159 ± 0.020	200−1000	0.406 ± 0.026	350−1000	0.483 ± 0.018
	400−1000	0.172 ± 0.022				
mu2	100−1000	0.170 ± 0.021	100−500	0.595 ± 0.018	200−500	0.340 ± 0.027
			600−1000	0.538 ± 0.024	750−1000	0.525 ± 0.025
mu3	100−650	0.155 ± 0.019	200−1000	0.569 ± 0.019	100−250	0.422 ± 0.026
	750−1000	0.141 ± 0.016			400−550	0.388 ± 0.029
					650−1000	0.510 ± 0.042
mu4	100−1000	0.158 ± 0.020	800−1000	0.553 ± 0.009	500−1000	0.471 ± 0.024
mu5	100−1000	0.159 ± 0.021	250−550	0.439 ± 0.026	100−500	0.310 ± 0.035
			650−1000	0.484 ± 0.019	650−1000	0.554 ± 0.017

[a] R is the average RMSD over the indicated time interval Δt, and δR is the corresponding standard deviation. The initial 100 ns are discarded from the estimation of average values. The average values were computed with a time resolution of 1 ps for the wild-type system and 10 ps for the mutants.

and 500 ns is followed by a second plateau with a smaller average RMSD after 600 ns. The abrupt drop at about 600 ns is potentially indicative of a structural transition. For the *wt* complex, we note multiple abrupt changes after 550 ns, which are also potentially associated with structural transitions. The PAM-distal RMSD reveals even more abrupt changes for all simulated complexes. In Table 2 we have compiled these observations in a quantitative way. Overall, the values reported in Table 2 are consistent with a high flexibility of the single-stranded segments PAM-proximal and PAM-distal, and a higher

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

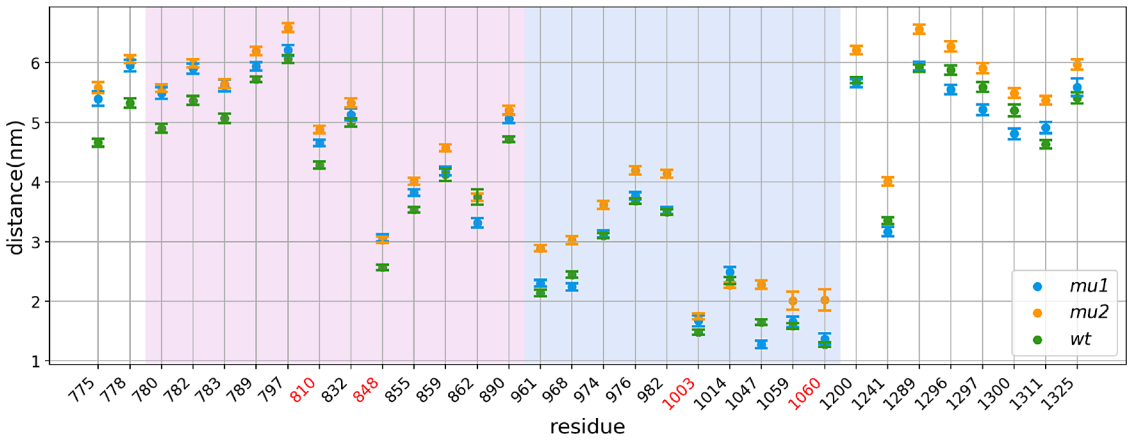


Figure 4. Time-average distance, over the last 250 ns, between the geometric center of the PAM-distal segment of the ntDNA and the 32 amino acids at the HNH/RuvCIII boundary that are positively charged in the *wt* complex. The shaded areas in mauve and blue represent residues in the HNH domain and RuvCIII motif, respectively. The red font on the horizontal axis is used for the residues that are mutated in complexes *mu1* and *mu2*.

rigidity of the double-stranded segment PAM-duplex. This observation is also supported by the larger and more variable standard deviations detected for the single-stranded portions. In fact, for PAM-proximal and PAM-distal, the RMSD standard deviation ranges between 0.01 and 0.04 nm, while for PAM-duplex, it is stable around 0.02 nm.

To further investigate the structural modifications of untDNA during the simulations, we performed clustering analysis on each of the six MD trajectories, with a resolution of one frame every 1 ns. To extract clusters of similar conformations, we employed GROMACS tools and the Linkage method.[57] A cutoff value of 0.14 nm was used to generate RMSD matrices for the untDNAs, after testing different cutoff values.[22] For the *wt* system, we observed one major cluster populated with 90% of the trajectory snapshots, which indicates no drastic changes in the structure of the unwound ntDNA over the simulated time span. For the mutant systems, we obtained multiple major clusters, indicating possible structural transitions. In *mu1*, the two largest clusters include 25 and 50% of the structures, related to the time spans of 240−500 and 500−1000 ns. In *mu2*, the two largest clusters represent a population of 35 and 33% of the entire trajectory, in the time spans 180−560 and 660−1000 ns, respectively. In *mu3*, we identified five significant clusters, with the two most populated clusters spanning 18 and 44% of the trajectory, in the time spans 375−550 and 560−1000 ns, respectively. In *mu4*, two main clusters were observed with population percentages of 11 and 74%, spanning the time ranges of 100−230 ns and 250−1000 ns. In *mu5*, two main clusters were identified with population percentages of 26 and 32%, covering periods of 200−490 and 675−1000 ns, respectively. The first and second most populated clusters of all mutants are associated with separated time intervals. This evidence is overall consistent with the RMSD statistics reported in Table 2 and with Figure 3 (see also Figure S2).

Based on the above data, average quantities presented in the following sections were computed over the last 250 ns of the trajectories, during which all systems are stationary.

Motions of the ntDNA Relative to the HNH/RuvCIII Boundary. We identify such motions based on analysis of distance distributions and contact maps.

The HNH/RuvCIII interface of the *wt* CRISPR/Cas9 complex contains 32 positively charged amino acids that may electrostatically attract the negatively charged backbone of untDNA strand.[14,22] To elucidate the functional role of these native positively charged amino acids, we inspected protein−DNA distances. Specifically, we evaluated the average distance, in the last 250 ns of the trajectories, between each ntDNA segment and each of these 32 amino acids, for the wild-type and mutant bound complexes. We present here a comparison of wild-type versus mutant distances. Namely we compare the distances between each of the 32 amino acids and the ntDNA segments in the wild-type system to those in the systems where some of the amino acids within the *wt* positive patch are mutated to neutral amino acids. No significant changes upon mutation were detected for the distances of the PAM-duplex and PAM-proximal segments to the selected 32 amino acids (not shown). However, the PAM-distal segment undergoes remarkable changes for what concerns its distance from the selected 32 amino acids.

The average distances between the geometric center of the PAM-distal segment and the geometric center of each of the 32 residues that are positively charged in the *wt* complex are presented in Figure 4 for *wt*, *mu1* and *mu2*. Figures S3 and S4 extend this analysis to the other mutants and the crystal structure. We point out a systematic distance increase upon mutation for *mu2*, with 30 out of 32 boundary residues showing an increased distance. The maximum increase, observed for residue 775, is 0.92 Å. This is mostly true also for *mu1*, with some exceptions. Specifically, in *mu1*, 24 out of 32 residues encounter an increased distance from the PAM-distal, with a maximum increase of 0.74 Å observed at residue 775. These observations suggest a detachment of the PAM-distal segment from the HNH/RuvCIII boundary in mutated complexes compared to the wild-type complex, consistent with a reduction in the protein-ntDNA electrostatic interaction.

A deeper examination reveals that most of the positive residues are in closest proximity to the PAM-proximal segment. However, four residues (1003, 1047, 1059, and 1060) in the RuvCIII motif, two of which are neutralized in the *mu1* and *mu2* structures, are closest to the PAM-distal segment. Interestingly, despite the neutralized residues, the average distances between the PAM-distal segment and these four residues do not increase in the trajectory of the *mu1* system. In other words, we infer that the ntDNA, and in particular its distal portion, experiences the protein mutations in a nonlocal manner.

The *Cas*9 protein's ability to cleave the ntDNA is attributed to the presence of four catalytic residues, namely D10, E762, H983, and D986.[58] These residues, which initiate the cleavage process,

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

are closest to the PAM-proximal segment of the ntDNA. To probe the interactions of these catalytic residues with the ntDNA and how they are affected by the selected mutations, we computed the distances between the geometric centers of ntDNA segments and the geometric centers of the catalytic residues in all systems. The results for the PAM-proximal segment are illustrated in the histograms shown in Figure 5. The data in Figure 5 reveal a consistent pattern where the average distance between the PAM-proximal segment and the four catalytic residues is the smallest for the *mu4* complex, followed by the *mu2* complex. These two complexes are the only ones, within the group of six systems investigated in this work, that contain the K848A mutation. Hence, we can hypothesize that the K848A mutation plays a role in tuning protein-ntDNA distances and interactions. We remark here that *mu2* has the same efficiency as *wt* and higher specificity, while *mu4* is not effective. Thus, it is difficult to unambiguously associate this particular mutation with specificity changes.

To further examine the interactions between the ntDNA strand and the HNH domain and RuvCIII motif, we computed time-averaged contact maps for the distances between the geometric centers of the ntDNA nucleotides and of each residue in HNH and RuvCIII, over the last 300 ns of the trajectories. We present the results for *mu1* and *mu2* in Figure 6a,b as difference contact maps, obtained by subtracting the contact map of the wild-type system from the contact map of each mutant system. Positive values in the difference contact maps indicate an increase in distance between the nucleotide-residue pairs in the mutant structures compared to the wild-type structure. Conversely, negative values represent a decrease of distances in the nucleotide-residue pairs due to the mutations. A distance increase (decrease) can be read as an index of weaker (stronger) electrostatic interactions in the mutant with respect to the wild-type. Similar difference contact maps for all mutants are depicted in Figure S5.

We remind the reader that among the four triple mutants that we have simulated, *mu1* and *mu2* exhibit enhanced specificity relative to the *wt* complex, while *mu3* and *mu4* are control systems that lack cleavage efficiency. In the difference contact maps of both enhanced specificity mutants, we observe a remarkable predominance of large positive values in the top part of the maps, which corresponds to the PAM-distal segment of the ntDNA strand. This is true all along the horizontal axis, which indicates that the PAM-distal segment is overall displaced from the entire HNH and RuvCIII portion and, in particular, from the HNH/RuvCIII boundary. The effect is weaker in the control triple mutants *mu3* and *mu4* (S5). The single mutant *mu5* is characterized by a larger displacement from RuvCIII than from HNH (S5).

Within the PAM-distal segment, we also find a few negative values, suggesting a decrease in the distance between certain nucleotides and specific residues. Upon closer examination, we observe that most of these negative values are associated in *mu1* and *mu2* with the distances between the PAM-distal segment and the loop structures within the RuvCIII motif, which are not related to electrostatic interactions.

The difference contact maps of all mutants consistently show a high number of negative values in the region corresponding to the PAM-proximal segment. This finding is somewhat surprising given that some of the mutated residues are in close proximity to the PAM-proximal segment. Intuitively, one could expect that the neutralization of these nearby positive residues would lead to a decrease in interaction between the mutated domain of the protein and the backbone of the PAM-proximal segment, resulting in an increase in the distance between them. However, the difference contact maps obtained from the MD trajectories of the mutant complexes contradict this expectation. This

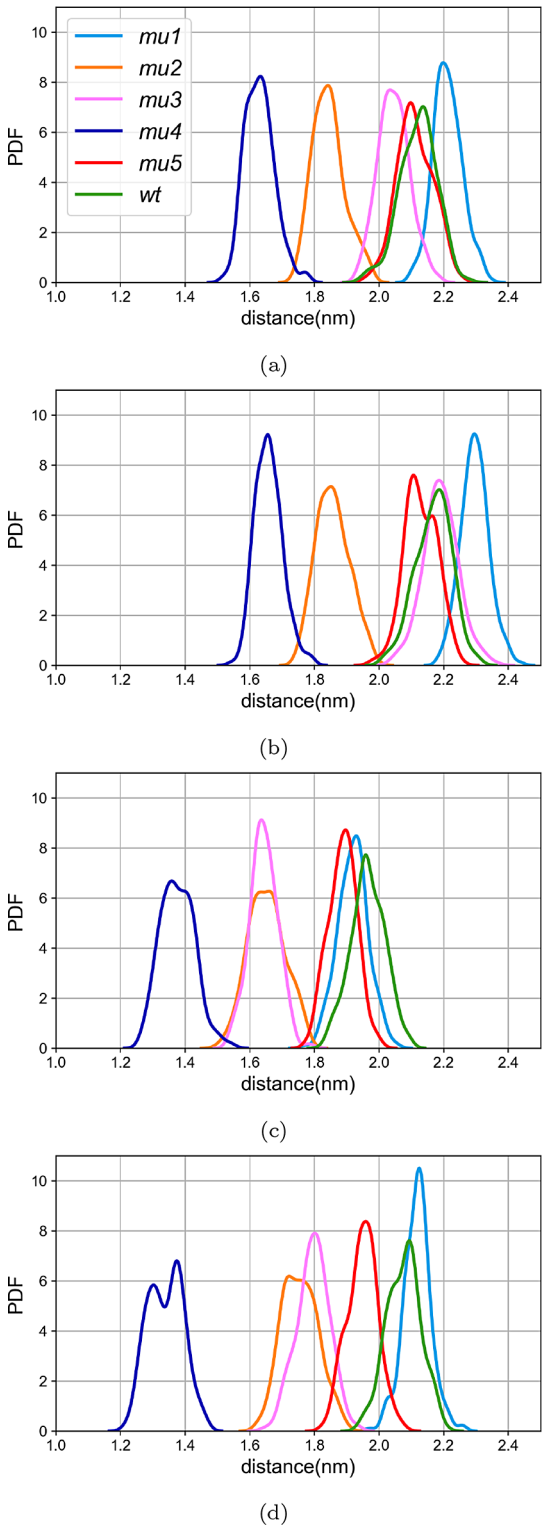


Figure 5. Distribution of the distance of the PAM-proximal geometric center from each of the four catalytic residues: D10 (a), E762 (b), H983 (c), and D986 (d), over the last 250 ns of simulated trajectories. The vertical axis represents the probability density function (PDF), namely the probability of detecting a certain distance (horizontal axis) in the collected snapshots.

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

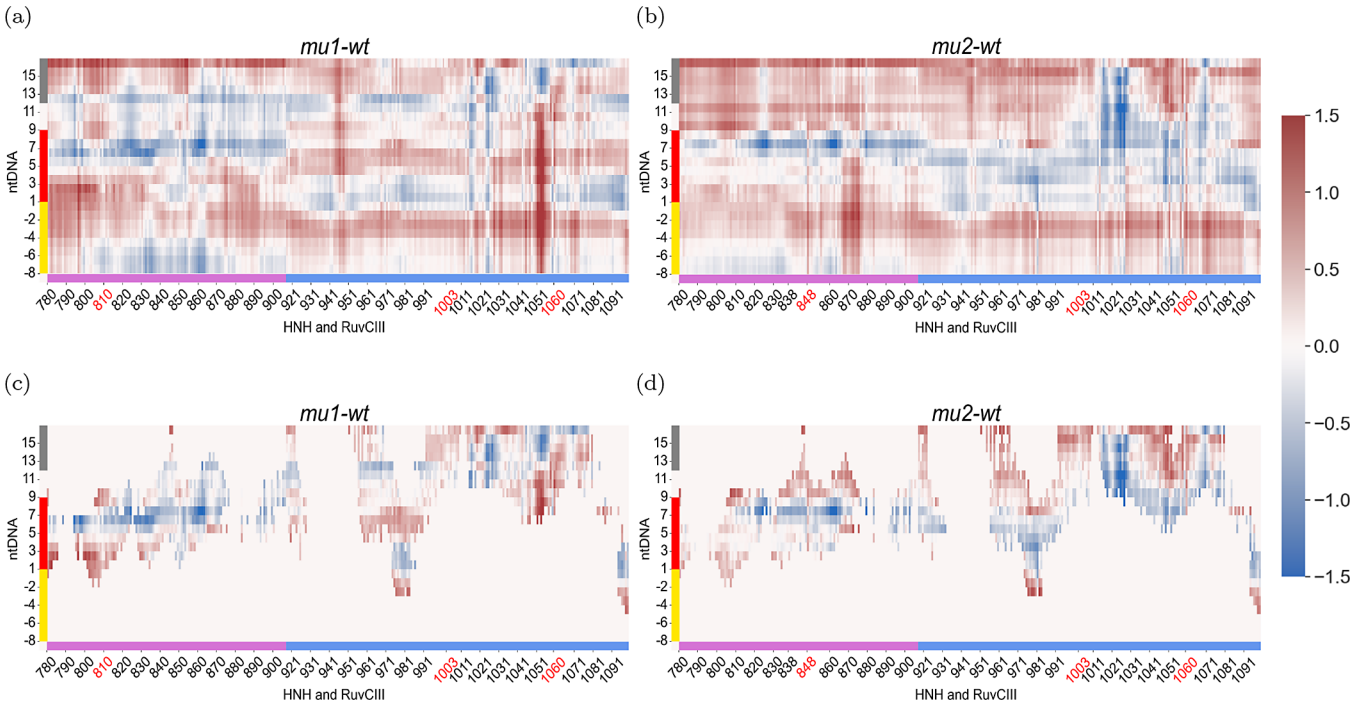


Figure 6. (a,b): Difference contact maps that illustrate changes in ntDNA distance from the HNH/RuvCIII boundary upon mutation, in *mu1* (a) and *mu2* (b). Each pixel represents the variation in average distances between ntDNA and HNH/RuvCIII residues in mutant structures compared to the wild-type structure, over the last 300 ns of the trajectories. (c,d): Modified difference contact maps that blur out regions where the amino acid−base distance is smaller than 1.5 nm. This visual representation is based on selectively obscuring noncritical sections of the contact maps to accentuate significant regions. Red (blue) pixels signify an increase (decrease) in distance due to the mutation. The units on the color bar are in nanometers. Color-coded axes indicate specific regions of the complexes: on the horizontal axis, mauve and blue represent the HNH domain and RuvCIII motif; on the vertical axis, yellow, red, and gray denote the PAM-duplex, PAM-proximal and PAM-distal portions of the ntDNA strand, respectively.

suggests that the effect of the mutations on the interaction between the mutated domain and the PAM-proximal segment is more complex than initially anticipated, and characterized by nonlocal effects.

In Figure 7 we report the distribution of distance differences that appear in the maps of Figure 6a,b. We note that the

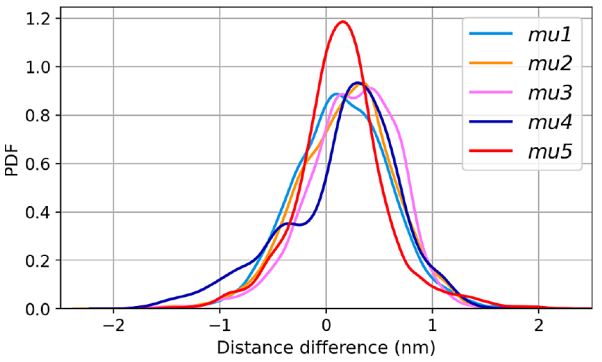


Figure 7. Distribution of the distance differences that appear in Figure 6a,b, over the last 300 ns of the trajectories.

distribution curve obtained from the *mu5* trajectory exhibits a peak near zero distance difference. The peak is more intense and narrower than the broad and occasionally bimodal peaks observed in the other mutants. This observation is consistent with a comparatively minor effect of the single mutation in *mu5* with respect to the triple mutant *mu1*, *mu2*, *mu3* and *mu4*. This trend aligns with the known similarities in specificity and efficiency between *mu5* and the *wt* structure. Importantly, all the plots in Figure 7 have a larger area-under-the-plot at positive values of the distance difference than at negative values, indicating an overall increase in the distance between HNH/RuvCIII and ntDNA residues due to mutations.

To gain more insights into mutation-induced atomic displacements, we focused on specific sections of the contact maps. We extracted these sections based on distances below 1.5 nm between ntDNA and HNH/RuvCIII residues in wild-type or mutant Cas9 variants. The resulting modified difference contact maps, as shown in Figure 6c,d, effectively highlight zones indicative of potential interactions between ntDNA and HNH/RuvCIII residue pairs, blurring out irrelevant structural portions. Figure S6 illustrates similar modified difference contact maps for all the mutants. Subsequently, the highlighted areas were subdivided into subareas that correspond to distinct portions of the ntDNA and HNH/RuvCIII domains. For each of these subdivisions, we calculated the average distance differences, illustrating the outcome through a bar plot (Figure 8). The term "average distance difference" refers to the mean of nonzero pixel values within each highlighted area. While the contact maps are presented in Figure 6 only for the enhanced-specificity mutants, the bar plot in Figure 8 includes all simulated mutants. The bar plot provides a concise representation of the average distance differences between distinct portions of the ntDNA and the HNH/RuvCIII protein portion, excluding long-distance residues.

The PAM duplex consistently displays noteworthy positive changes in the distance of the ntDNA from both HNH and RuvCIII domains across all mutants. In *mu5*, the average distance between PAM-duplex and RuvCIII undergoes the smallest deviation from the *wt* structure among other pairs. The

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

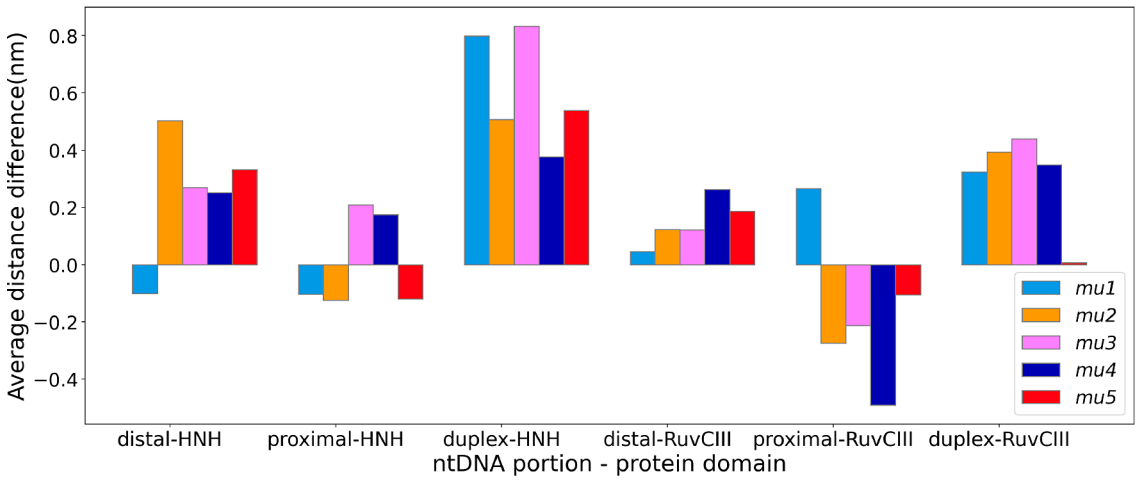


Figure 8. Average distance differences, with respect to the wild-type complex, between ntDNA segments and nuclease Cas9 portions, in mutant DNA-bound CRISPR/Cas9 systems simulated in this work. Distal, proximal and duplex stand for PAM-distal, PAM-proximal and PAM-duplex. The color code to identify different mutants is the same as in Figures 2−4.

PAM-distal segment exhibits a consistent distance increase from both HNH and RuvCIII across all mutants, except for a decrease observed in *mu1* concerning the distance from HNH. Conversely, the PAM-proximal segment manifests varied behavior across the mutant variants, with predominantly negative average distance differences. Overall, the data presented in Figure 8 indicate a mutation-induced geometrical separation between the ntDNA and the HNH+RuvCIII domain of Cas9.

Hydrogen Bond Patterns. The plots in Figure 9 show a remarkable effect of the mutations on the hydrogen bonding profile between the HNH/RuvCIII residues and the untDNA nucleotides. In particular, as shown in Figure 9a, the *mu1* and *mu2* systems have more hydrogen bonds between the HNH domain and the untDNA than the wild-type complex. Conversely, the same mutant systems have fewer ntDNA-RuvcIII hydrogen bonds than the wild-type complex, resulting in a fairly invariant total number of hydrogen bonds, as indicated in Figure 9c. In fact, in Figure 9c the cyan and green curves have a sizable overlap, especially during the final 300 ns. Similarly, also the orange and green curves have a sizable overlap.

On the other hand, the single mutant *mu5* and the control triple mutants *mu3* and *mu4* display a lower total number of hydrogen bonds between the untDNA and residues in both HNH and RuvCIII. The plot in Figure 9c reveals a gradual loss of hydrogen bonds during the time span of the trajectory, particularly pronounced in *mu3* and *mu5*. See also Figure S7 for a different visual representation. Additionally, Table S3 lists the specific HNH/RuvCIII residues involved in H-bonding with untDNA.

In summary, we find that the hydrogen bond variation patterns do not closely follow the patterns in the residue-nucleotide average distances. Moreover, the loss of hydrogen bonds cannot be strictly associated with specificity gains.

Conformational Transitions through Principal Component Analysis. The results of PCA applied to the cumulative trajectory of the HNH+RuvCIII subsystem are illustrated in Figure 10. Examination of the eigenvalue plot in the lower left corner of Figure 10 reveals that the five largest eigenvalues are notably greater than the rest, indicating that these five principal components capture the majority of the variance in the system, accounting for 67% of the total. Our analysis focuses on PC1 and

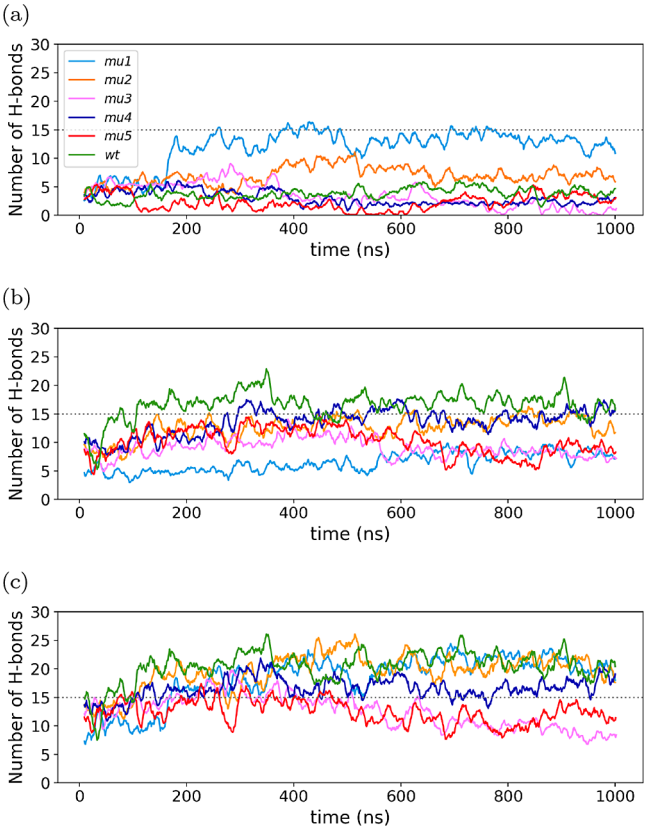


Figure 9. (a) Number of hydrogen bonds between the unwound ntDNA segment and the HNH domain of the Cas9 protein, as a function of time. (b) Number of hydrogen bonds between the unwound ntDNA segment and the RuvCIII motif of the Cas9 protein, as a function of time. (c) Number of hydrogen bonds between the unwound ntDNA segment and the union of HNH and RuvCIII, as a function of time. See also Figure S7.

PC2, which have the largest eigenvalues and capture 39% of the total variance. These components are the most informative in terms of capturing the dominant modes of the conformational dynamics of the simulated CRISPR-Cas9 systems.

We projected the subtrajectory of HNH+RuvCIII onto the two-dimensional subspace defined by PC1 and PC2. The

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

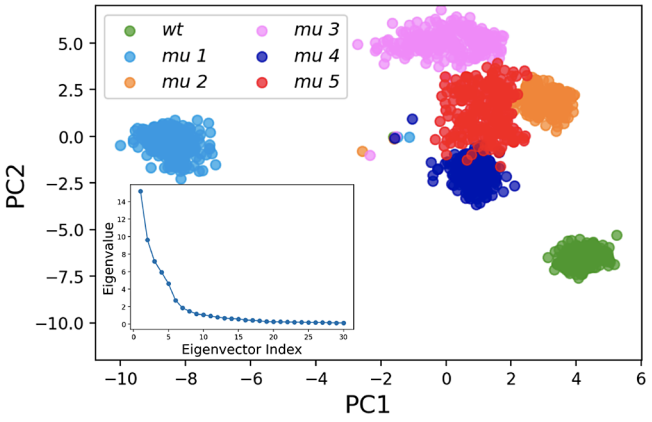


Figure 10. Projection of the cumulative 6-μs trajectory on PC1 and PC2 obtained from the covariance matrix computed on the backbone of the HNH+RuvCIII portion of the entire CRISPR/Cas9 bound complexes. The inset shows the decreasing eigenvalues of the said covariance matrix, as a function of the eigenvalue index. The steep decrease and following saturation of this plot indicate that few principal components represent the global system behavior. In particular, the five largest eigenvalues capture the entire dynamics, and the two largest eigenvalues represent 39% of the total variance.

resulting plot, depicted in Figure 10 for the last 250 ns of simulation, shows each system represented by a distinct color. Clear separation between structural configurations of different complexes is observed in the clusters. Notably, the cluster representing the *wt* conformational states is distinctly separated from the cluster of *mu1* states in the HNH+RuvCIII projected subtrajectory along PC1. We also note that the HNH+RuvCIII conformational states of mutants, specifically *mu1*, *mu3*, and *mu5* exhibit higher scattering compared to the *wt* structure in either or both PC1 and PC2, hinting to their energetically less stable nature.

Further insights are gained by examining motions spanning the continuum between the extremes along PC1 and PC2. The extreme states of the HNH+RuvCIII subsystem are visualized as spectra of states, ranging from blue to yellow to denote the states with the smallest to largest PC values (Figures 11 and S8). Along PC1 (Figure 11), notable conformational changes occur within residues 940−950 and 1048−1058 in the RuvCIII domain, as well as across the entire HNH domain. Specifically, residues 1048−1058, located close to the untDNA, exhibit conformational alterations parallel to the nearby untDNA segment. The disparity between the state spectra of the *mu1* and *wt* configurations is particularly pronounced, given their positions at opposing ends of the PC1 axis.

Similarly (see Figure S8), spectra of extended states along the PC2 direction highlight significant conformational changes within regions encompassing residues 940−950, 1020−1030, and 1048−1058. Along PC2, the conformational dynamics of residues 1048−1058, situated near the untDNA segment, are observed to change perpendicular to the untDNA segment.

Examining the principal conformational variances of the DNA captured by PC1 using the first method described in the methods section, we observed significant manifestations in the untDNA region, followed by residues 9−12 of the tDNA. This elucidated notable discrepancies in DNA conformation between the wild-type and mutant systems (Figure 12).

Upon examination of the spectra depicting extreme conformations of DNA along the PC1 direction for each system, using the second method explained in the methods

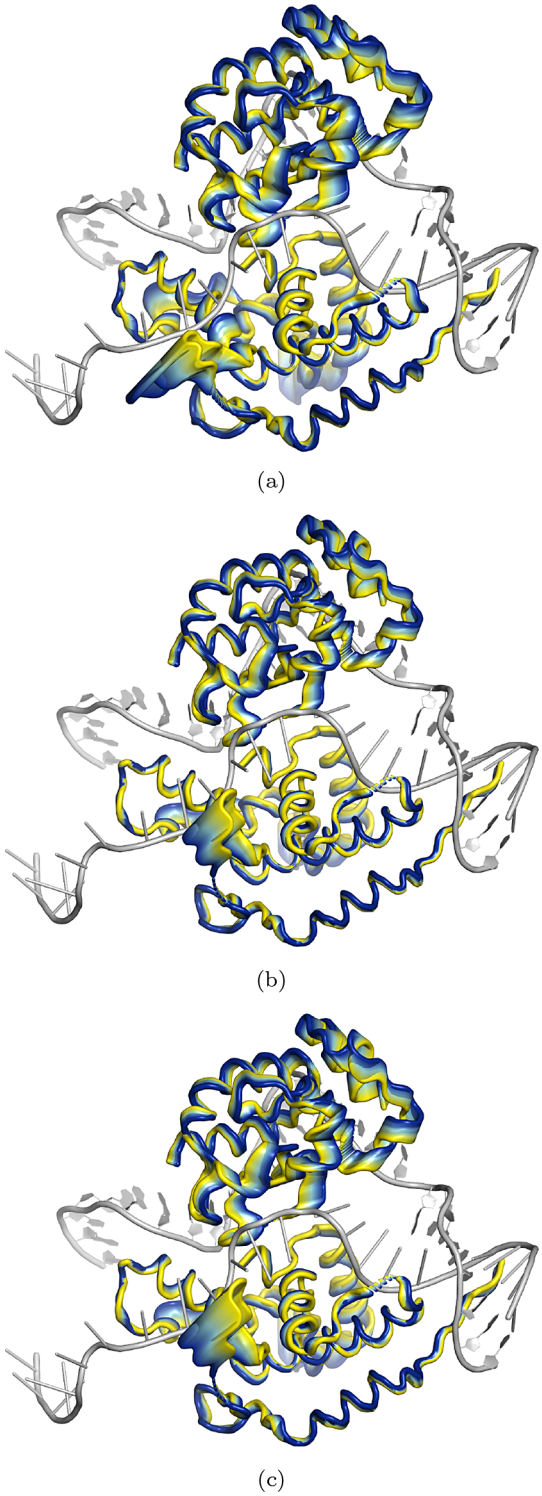


Figure 11. Static visualization of the projection of three trajectories on the first largest principal component, PC1, of the HNH+RuvCIII covariance matrix: (a) *mu1*; (b) *mu2*; (c) *wt*. The color scale from blue to yellow represents the protein domain in states with PC1 values from smallest to largest. The protein domain is shown in cartoon representation. The gray color is used for DNA. The tDNA strand is shown in new ribbon representation and the ntDNA strand is shown in new cartoon representation in which a base is represented with a stick perpendicular to the tubular backbone. The DNA structure is taken from the initial frame of the *mu1* trajectory.

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

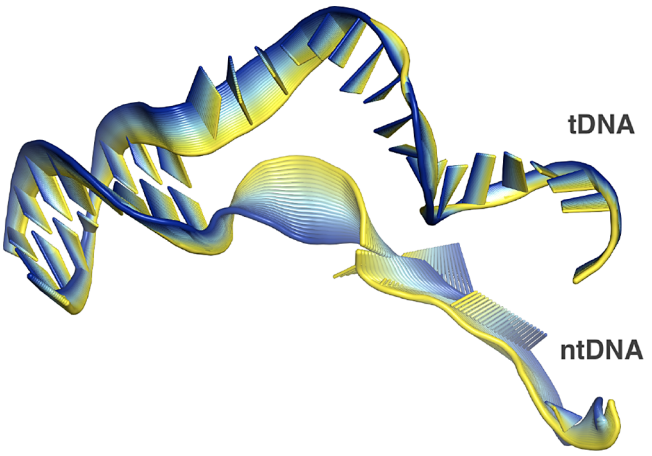


Figure 12. Static visualization depicting the cumulative 6-μs trajectory of all DNA subsystems projected onto the first principal component, PC1, of the DNA covariance matrix. The color gradient, ranging from blue to yellow, indicates states with PC1 values from smallest to largest. The representation of DNA is depicted in a cartoon format. Labels indicating tDNA and ntDNA are included within the figure.

section, we observed notable alterations within the entirety or portions of the untDNA region in the mutants (see Figure S9). Notably, the mutant system denoted as *mu3* exhibited the most subtle changes within the untDNA. Conversely, PC1 of the wild-type system revealed a discernible conformational shift in the PAM-proximal section; however, this change was comparatively less pronounced than in the mutants.

dsDNA Shape Analysis. The plots in Figure 13 report selected shape features of the dsDNA segment along the dsDNA sequence, averaged over the last 250 ns. The horizontal axis represents the base-pair (bp) index in the dsDNA, with bp number seven denoting the closest base pair to the unwound ntDNA. There are six intrabase-pair helix parameters and six interbase-pair helix parameters that describe the DNA shape for a given atomic structure, plus four parameters that describe the position and orientation of a base with respect to the origin, and the minor groove width (MGW).[59] To focus our attention on a smaller number of parameters, we selected for Figure 13 the shape features that are most relevant for modeling transcription factor-DNA binding specificity.[56,60] In Table 3, on the other hand, we expand our attention to all the base-pair features.

The helical twist (Figure 13a) is not constant along the sequence and deviates from the ideal B-DNA value of 36 degrees. The trend is very similar for all the triple mutants for bp's 1−5, while edge bp's 6 and 7 show more variability. This is consistent with the fact that bp's 6 and 7 are closest to the unfolded part of DNA. The sequence-average value over all bp's is 35.3°, 35.4°, 34.0°, 34.3°, 35.6° and 35.2° in *mu1*, *mu2*, *mu3*, *mu4*, *mu5* and *wt*, respectively, systematically smaller than the ideal value by 1−6%. At bp's 6 and 7, *wt* and *mu5* have the largest values of the helical twist.

The Roll in Figure 13b is fairly uniform in all structures in bp's 1−3 and then becomes scattered. System *wt* (*mu5*) has negative Roll values at bp's 3, 4, 5 (5, 6, 7).

The dsDNA's time-average MGW is shown in Figure 13d. The *wt* structure has a MGW maximum in the center of the duplex DNA segment, while all the mutants have a MGW minimum. MGW minima are usually associated with electrostatic potential wells,[61] which may control protein−DNA

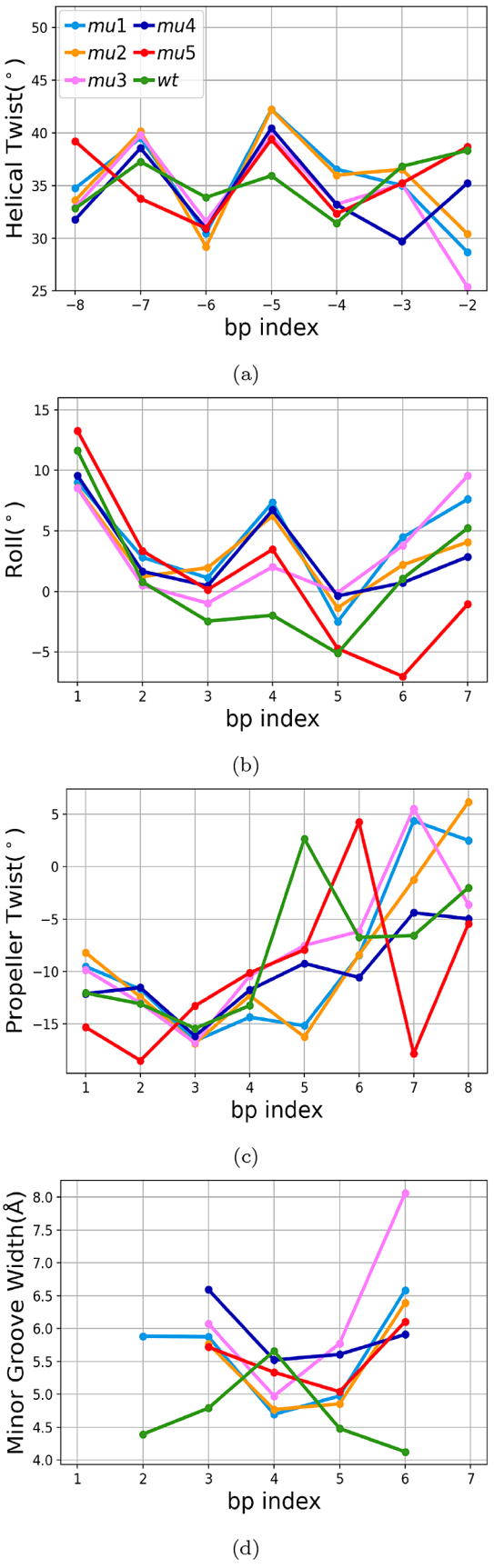


Figure 13. Selection of helix shape parameters as a function of the base-pair index along the dsDNA sequence: (a) helical twist in degrees; (b) roll in degrees; (c) propeller twist in degrees; (d) minor groove width in Å.

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

Table 3. Euclidean Distance between Mutant and Wild-Type Curves for the Intra-Base-Pair (Upper Six Rows, White Background) and Inter-Base-Pair (Lower Six Rows Part, White Background) Shape Parameters[a]

	mu1	*mu2*	*mu3*	*mu4*	*mu5*
opening (°)	12.41	12.63	10.58	7.66	10.15
ProT (°)	21.70	21.73	16.35	13.20	20.61
buckle (°)	45.30	41.68	37.65	34.77	38.49
stretch (Å)	0.40	0.37	0.38	0.22	0.21
stagger (Å)	1.10	0.94	0.75	1.04	1.13
shear (Å)	1.16	0.97	1.07	0.83	0.81
HT (°)	13.51	12.396	14.124	9.760	8.766
roll (°)	3.26	2.650	4.390	2.779	2.280
tilt (°)	9.44	7.422	7.059	9.736	5.697
slide (Å)	1.92	1.94	1.82	1.49	2.19
rise (Å)	1.20	1.15	0.96	0.93	0.96
shift (Å)	3.61	3.52	2.03	2.29	3.41
ΣA (°)	**105.6**	**98.5**	**90.1**	**77.9**	**86.0**
ΣB (Å)	**9.4**	**8.9**	**7.0**	**6.8**	**8.7**

[a]The bold cells report the sum of all the angle values (ΣA) and the sum of all the length values (ΣB), quantifying the total variations of each mutant complex from the wild-type complex.

interactions, in particular interactions that involve charged amino acids.

In the propeller twist plots, Figure 13c, we observe that the propeller twist of the first half of the dsDNA is relatively similar for all structures except *mu5*. However, the propeller twist is more scattered in the second half of the dsDNA, closer to the untDNA.

Overall, these trends do not denote a clear behavior of the shape parameters as a function of the sequence position. On the other hand, they do not even reveal any neat trend induced by the mutations. The remaining four interbase-pair and five intrabase-pair shape parameters are reported in Figures S10 and S11, respectively, and do not add substantial interpretation.

Table 3 summarizes the DNA shape analysis. The complexes *mu1* and *mu2* present the largest cumulative changes in the duplex shape, which accompany the significant ntDNA−protein changes discussed above.

SUMMARY AND DISCUSSION

The CRISPR/Cas9 gene editing technique, while powerful, is susceptible to off-target cleavage events characterized by RNA-DNA mismatches. Various strategies are under investigation to enhance specificity, including tailored mutations of the Cas9 protein. Specifically, mutations of conserved positively charged residues at the HNH/RuvCIII nuclease boundary have been proposed. The rationale behind these mutations is to reduce the positive charge of the boundary, thereby weakening the electrostatic attraction between the negatively charged ntDNA backbone and Cas9. This, in turn, would facilitate the rehybridization of DNA strands when the hybrid RNA/DNA duplex includes energetically unfavorable mismatches, potentially minimizing off-target effects.[22] Electrostatic potential maps computed from our trajectories confirm this intuition (see Supporting Figure S12).

In this study, we conducted a comprehensive analysis of Molecular Dynamics (MD) trajectories involving CRISPR/Cas9/DNA ternary complexes. These complexes featured both wild type (*wt*) and mutant proteins, with the mutations strategically targeting conserved positively charged residues at the HNH/RuvCIII boundary.[14] The complexes comprised the Cas9 protein, the single guide RNA (sgRNA), the target DNA (tDNA) strand, and the elongated nontarget DNA (ntDNA) strand. Our investigation sheds light on the impact of protein mutations on the protein and nucleic acid conformations. Specifically, we focused our attention on mutation-induced changes in the time-dependent structural quantities that characterize the relative spatial arrangement between Cas9 and the ntDNA strand.

Our analysis reveals conformational changes in the ntDNA of mutant complexes, resulting in an increase of the distance between the ntDNA and HNH/RuvCIII residues. This increase is a proxy for interaction weakening.

The observed patterns of protein-ntDNA distance changes reveal nonlocal effects. In fact, mutations in specific Cas9 residues do not directly influence their very distance from the ntDNA. Instead, the explored mutations have an indirect effect on the distance of the PAM-distal ntDNA from other, native, residues at the HNH/RuvCIII boundary. In other words, we detect mutation-induced allosteric motions that are conducive to Cas9-ntDNA interaction decay. Allosteric effects are presented here as observations, without aiming at their deep understanding, which would require the simulation of many more Cas9 mutations. This does not compromise our main conclusion that Cas9 mutations alter the interaction balance between different portions of the DNA-bound CRISPR/Cas9 complex.

Inspection of the trajectories also unveils mutation-induced effects in the H-bond patters. Specifically, we have shown that the number of H-bonds between the ntDNA strand and the RuvCIII domain is reduced in mutants with respect to the *wt* complex. This reduction of H-bond contacts is particularly noteworthy because the RuvCIII is the Cas9 portion that carries out ntDNA cleavage. On the other hand, our data do not reveal direct correlation between H-bond formation/disruption and specificity changes.

Overall, the results of this work corroborate the hypothesis that protein mutations orchestrate the energy balance between different parts of DNA-bound CRISPR/Cas9 complexes.[14,22] Because the energy balance is a key factor in the function of composite systems as the DNA-bound CRISPR/Cas9 complexes, this strategy may work to engineer more functional systems, in particular, more specific systems. While other strategies are explored to engineer specificity,[31,32] the stage is not mature to exclude any of them, and the combination of different approaches may eventually be beneficial. Future studies aimed at probing the energy landscapes with metadynamics or other accelerated sampling methods[62−64] would be desirable in order to shed more light on how mutation-induced changes of electrostatic interactions determine free-energy pathways that ultimately control stability and related functionality.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c04359.

Available X-ray and cryo-EM structures of spCas9 (Tables S1 and S2); residues involved in hydrogen bonding with untDNA (Table S3); HNH and RuvCIII with bold targeted residues for mutation (Figure S1); mean distance between geometric center of the PAM-distal and the 32 positively charged residues (Figures S3 and S4); differ-

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988

ence contact maps illustrating alterations in ntDNA distances from the HNH/RuvCIII boundary resulting from mutations (Figures S5, S6, and S13); static visualization of the projection of trajectories on principal components for DNA and HNH+RuvCIII (Figures S8 and S9); representative structures from clustering of untDNA (Figure S2); distribution curves of the relative H-bond count of each mutant complex with respect to the wild-type complex (Figure S7); interbase-pair and intrabase-pair helix shape parameters (Figures S10 and S11); surface representation of the difference in electrostatic potential (Figure S12); and RNA representative structures of *mu1* and *mu1-ol3* (Figure S14) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Seyedeh Hoda Fatemi Abhari − *Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0009-0005-0815-1801; Email: fatemiab@usc.edu

Rosa Di Felice − *Departments of Physics and Astronomy and Quantitative and Computational Biology, University of Southern California, Los Angeles, California 90089, United States; CNR Institute of Nanoscience, Modena 41125, Italy;* orcid.org/0000-0002-7772-6550; Email: difelice@usc.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c04359

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the Center for Advanced Research Computing (CARC) at the University of Southern California for providing computing resources that have contributed to the research results reported within this publication (URL: https://carc.usc.edu), and the Defense Advanced Research Projects Agency (DARPA) for Contract No. HR001122C0063.

REFERENCES

(1) Makarova, K. S.; Wolf, Y. I.; Iranzo, J.; Shmakov, S. A.; Alkhnbashi, O. S.; Brouns, S. J.; Charpentier, E.; Cheng, D.; Haft, D. H.; Horvath, P.; et al. Evolutionary classification of CRISPR−Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **2020**, *18*, 67−83.

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

(3) 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.

(4) Zeng, Y.; Cui, Y.; Zhang, Y.; Zhang, Y.; Liang, M.; Chen, H.; Lan, J.; Song, G.; Lou, J. The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic acids research* **2018**, *46*, 350−361.

(5) Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **2011**, *471*, 602−607.

(6) Rueda, F. O.; Bista, M.; Newton, M. D.; Goeppert, A. U.; Cuomo, M. E.; Gordon, E.; Kröner, F.; Read, J. A.; Wrigley, J. D.; Rueda, D.; et al. Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat. Commun.* **2017**, *8*, 1610.

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

(8) Jore, M. M.; Lundgren, M.; Van Duijn, E.; Bultema, J. B.; Westra, E. R.; Waghmare, S. P.; Wiedenheft, B.; Pul, Ü.; Wurm, R.; Wagner, R.; et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology* **2011**, *18*, 529−536.

(9) Szczelkun, M. D.; Tikhomirova, M. S.; Sinkunas, T.; Gasiunas, G.; Karvelis, T.; Pschera, P.; Siksnys, V.; Seidel, R. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9798−9803.

(10) 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.

(11) Xue, C.; Greene, E. C. DNA repair pathway choices in CRISPR-Cas9-mediated genome editing. *Trends in Genetics* **2021**, *37*, 639−656.

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

(13) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. 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.

(14) 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.

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

(16) Osuka, S.; Isomura, K.; Kajimoto, S.; Komori, T.; Nishimasu, H.; Shima, T.; Nureki, O.; Uemura, S. Real-time observation of flexible domain movements in CRISPR−Cas9. *EMBO journal* **2018**, *37*, No. e96941.

(17) Yang, M.; Peng, S.; Sun, R.; Lin, J.; Wang, N.; Chen, C. The conformational dynamics of Cas9 governing DNA cleavage are revealed by single-molecule FRET. *Cell reports* **2018**, *22*, 372−382.

(18) Zhao, L. N.; Mondal, D.; Warshel, A. Exploring alternative catalytic mechanisms of the Cas9 HNH domain. *Proteins: Struct., Funct., Bioinf.* **2020**, *88*, 260−264.

(19) Yoon, H.; Zhao, L. N.; Warshel, A. Exploring the catalytic mechanism of Cas9 using information inferred from endonuclease VII. *ACS catalysis* **2019**, *9*, 1329−1336.

(20) Nierzwicki, L.; East, K.; Binz, J.; Hsu, R.; Ahsan, M.; Arantes, P.; Skeens, E.; Pacesa, M.; Jinek, M.; Lisi, G.; Palermo, G. Principles of target DNA cleavage and the role of Mg2+ in the catalysis of CRISPR−Cas9. *Nature Catalysis* **2022**, *5*, 912−922.

(21) Saha, A.; Arantes, P.; Palermo, G. Dynamics and mechanisms of CRISPR-Cas9 through the lens of computational methods. *Curr. Opoin. Struct. Biol.* **2022**, *75*, No. 102400.

(22) Ray, A.; Di Felice, R. Protein-Mutation-Induced Conformational Changes of the DNA and Nuclease Domain in CRISPR/Cas9 Systems by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2020**, *124*, 2168−2179.

(23) Ray, A.; Di Felice, R. Molecular Simulations have Boosted Knowledge of CRISPR/Cas9: A Review. *J. Self-Assem. Mol. Electron.* **2019**, *7*, 45−72.

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

(25) Palermo, G.; Ricci, C. G.; McCammon, J. A. The invisible dance of CRISPR-Cas9. *Phys. Today* **2019**, *72*, 30−36.

(26) Watters, K. E.; Kirkpatrick, J.; Palmer, M. J.; Koblentz, G. D. The CRISPR revolution and its potential impact on global health security. *Pathogens and Global Health* **2021**, *115*, 80−92. PMID: 33590814

(27) Wang, J.; Doudna, J. CRISPR technology: A decade of genome editing is only the beginning. *Science* **2023**, *379*, No. eadd8643.

(28) Zischewski, J.; Fischer, R.; Bortesi, L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology Advances* **2017**, *35*, 95−104.
(29) Guo, C.; Ma, X.; Gao, F.; Guo, Y. Off-target effects in CRISPR/Cas9 gene editing. *Front. Bioeng. Biotechnol.* **2023**, *11*, No. 1143157.
(30) Kleinstiver, B.; Pattanayak, V.; Prew, M.; Tsai, S.; Nguye, N.; Zheng, Z.; Joung, J. High-fidelity CRISPR−Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **2016**, *529*, 490−495.
(31) Akcakaya, P.; Bobbin, M.; Guo, J.; et al. In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature* **2018**, *561*, 416−419.
(32) Babu, K.; Amrani, N.; Jiang, W.; Yogesha, S.; Nguyen, R.; Qin, P. Z.; Rajan, R. Bridge Helix of Cas9Modulates Target DNA Cleavage and Mismatch Tolerance. *Biochemistry* **2019**, *58*, 1905−1917.
(33) Bhattacharya, S.; Satpati, P. Insights into the Mechanism of CRISPR/Cas9-Based Genome Editing from Molecular Dynamics Simulations. *ACS Omega* **2023**, *8*, 1817−1837.
(34) Palermo, G.; Gravina Ricci, C.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V.; McCammon, J. PAM-induced allostery activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139*, 16028.
(35) 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.
(36) Humphrey, W.; Dalke, A.; Schulten, K. VMD−Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(37) Tangprasertchai, N. S.; Di Felice, R.; Zhang, X.; Slaymaker, I. M.; Vazquez Reyes, C.; Jiang, W.; Rohs, R.; Qin, P. Z. CRISPR−Cas9 mediated DNA unwinding detected using site-directed spin labeling. *ACS Chem. Biol.* **2017**, *12*, 1489−1493.
(38) Schrödinger, L.; DeLano, W. *PyMOL*. http://www.pymol.org/pymol.
(39) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(40) Maier, J.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(41) Cheatham, T. I.; Cieplak, P.; Kollman, P.; et al. A modified version of the Cornell. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(42) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophysical journal* **2007**, *92*, 3817−3829.
(43) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55−58.
(44) Esmaeeli, R.; Bauzá, A.; Perez, A. Structural predictions of protein−DNA binding: MELD-DNA. *Nucleic Acids Res.* **2023**, *51*, 1625−1636.
(45) Vangaveti, S.; Ranganathan, S.; Chen, A. Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *WIREs RNA* **2017**, *8*, No. e1396.
(46) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E1346−E1355.
(47) Zgarbová, M.; Otypka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. I.; Jurečka, 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*, 2886−2992.
(48) Galindo-Murillo, R.; Robertson, J.; Zgarbová, M.; Šponer, J.; Otypka, M.; Jurečka, P.; Chatham, T. I. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114−4127.
(49) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7−8.
(50) MacKerell, A. J.; et al. Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(51) Sagui, C.; Darden, T. A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure* **1999**, *28*, 155−179.
(52) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(53) Parrinello, M.; Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Physical review letters* **1980**, *45*, 1196.
(54) Lavery, R.; Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct.* **1989**, *6*, 655−667.
(55) Lara-Gonzalez, S.; Dantas Machado, A.; Rao, S.; Napoli, A.; Birktoft, J.; Di Felice, R.; Rohs, R.; Lawson, C. The RNA Polymerase a Subunit Recognizes the DNA Shape of the Upstream Promoter Element. *Biochemistry* **2020**, *59*, 4523−4532.
(56) Zhou, T.; Shen, N.; Yang, L.; Abe, N.; Horton, J.; Mann, R.; Bussemaker, H. J.; Gordan, R.; Rohs, R. Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 4654−4659.
(57) Ross, G. Algorithm AS 15: single Linkage cluster analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **1969**, *18*, 106−110.
(58) Casalino, L.; Nierzwicki, L.; Jinek, M.; Palermo, G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal.* **2020**, *10*, 13596−13605.
(59) Dickerson, R. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* **1989**, *17*, 1797.
(60) Zhou, T.; Yang, L.; Lu, Y.; Dror, I.; Dantas Machado, A.; Ghane, T.; Di Felice, R.; Rohs, R. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* **2013**, *41*, W56.
(61) Rohs, R.; West, S.; Sosinsky, A.; Liu, P.; Mann, R.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248−1253.
(62) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(63) Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics* **2020**, *2*, 200−212.
(64) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16*, 163−199.

https://doi.org/10.1021/acsomega.4c04359
ACS Omega 2024, 9, 44974−44988