Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations

Giulia Palermo,*,[†,‡] Yinglong Miao,[†,‡] Ross C. Walker,[§,∥] Martin Jinek,[⊥] and J. Andrew McCammon[†,‡,§,∥,#]

[†]Department of Pharmacology, University of California San Diego, La Jolla, California 92093, United States

[‡]Howard Hughes Medical Institute, University of California San Diego, La Jolla, California 92093, United States

[§]San Diego Supercomputer Center, University of California San Diego, 9500 Gilman Drive, MC0505, La Jolla, California 92093-0505, United States

[⊥]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

[∥]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California 92093, United States

[#]National Biomedical Computation Resource, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0608, United States

S *Supporting Information*

ABSTRACT: The CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 system recently emerged as a transformative genome-editing technology that is innovating basic bioscience and applied medicine and biotechnology. The endonuclease Cas9 associates with a guide RNA to match and cleave complementary sequences in double stranded DNA, forming an RNA:DNA hybrid and a displaced non-target DNA strand. Although extensive structural studies are ongoing, the conformational dynamics of Cas9 and its interplay with the nucleic acids during association and DNA cleavage are largely unclear. Here, by employing multi-microsecond time scale molecular dynamics, we reveal the conformational plasticity of Cas9 and identify key determinants that allow its large-scale conformational changes during nucleic acid binding and processing. We show how the "closure" of the protein, which accompanies nucleic acid binding, fundamentally relies on highly coupled and specific motions of the protein domains, collectively initiating the prominent conformational changes needed for nucleic acid association. We further reveal a key role of the non-target DNA during the process of activation of the nuclease HNH domain, showing how the nontarget DNA positioning triggers local conformational changes that favor the formation of a catalytically competent Cas9. Finally, a remarkable conformational plasticity is identified as an intrinsic property of the HNH domain, constituting a necessary element that allows for the HNH repositioning. These novel findings constitute a reference for future experimental studies aimed at a full characterization of the dynamic features of the CRISPR-Cas9 system, and—more importantly—call for novel structure engineering efforts that are of fundamental importance for the rational design of new genome-engineering applications.
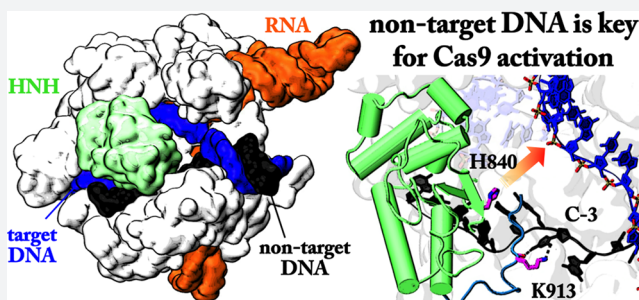
INTRODUCTION

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas is a bacterial immune system that confers protection against invading viruses.[1] In 2012, the discovery that the CRISPR-associated enzyme Cas9 functions as an RNA-programmable DNA endonuclease led to its development as a molecular tool for genome editing.[1,2] The applications of CRISPR-Cas9 technology are poised to improve our understanding of human health and disease and enable safe and efficient gene therapies, while also driving biotechnological advances in areas such as crop engineering and biofuel production.[2−4] In the CRISPR-Cas9 system, the endonuclease Cas9 associates with a guide RNA structure consisting of a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA), and uses its sequence information in the crRNA to recognize and cleave matching sequences in double-stranded DNA.[5] Upon site-specific recognition of a protospacer adjacent motif (PAM) in the target DNA sequence,[6] the DNA binds to Cas9-guide RNA complex, matching the RNA guide with one strand (the target DNA strand, *t*-DNA), while the other strand (non-target DNA, *nt*-DNA) is displaced. Subsequently, Cas9 uses two nuclease domains—HNH and RuvC—to cleave the *t*-DNA and *nt*-DNA strands, respectively.

Atomic-resolution structures revealed that Cas9 adopts a bilobed architecture comprising an α-helical lobe (REC), which mediates the nucleic acid binding, and a nuclease lobe including the RuvC and HNH catalytic domains (Figure 1).[7] An arginine
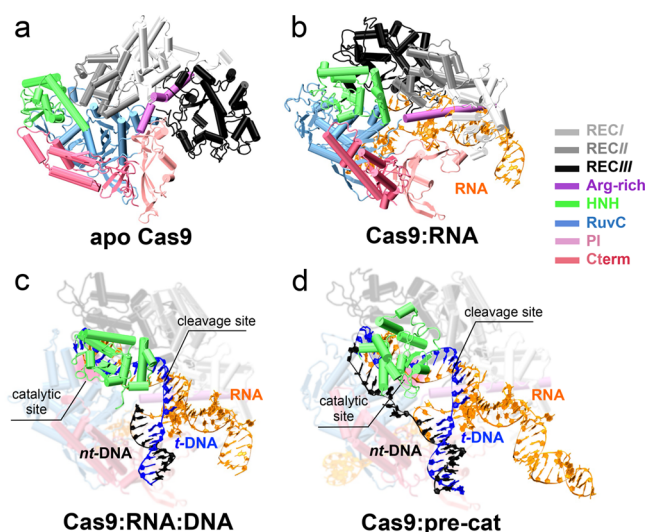
Figure 1. X-ray structures of the apo Cas9 (4CMQ)[5] (a), in complex with RNA (Cas9:RNA, 4ZT0)[8] (b), and in the DNA-bound states (c, d), as captured in the presence of an incomplete DNA (Cas9:RNA:DNA, 4UN3)[9] (c), and in a precatalytic state (Cas9:pre-cat, 5F9R)[11] including both unwound strands (d). Cas9 is shown as cartoons, highlighting individual protein domains with different colors. The RNA (orange), target DNA (*t*-DNA, blue), and non-target DNA (*nt*-DNA, black) strands are in ribbons. In the apo Cas9 and Cas9:RNA structures, the α-helical lobe regions REC*I* (silver), REC*II* (gray), and REC*III* (black) adopt remarkably different configurations; whereas different conformations of the HNH domain (green) are observed in the DNA-bound states. In Cas9:RNA:DNA, the HNH catalytic site (red) points in the opposite direction with respect to the cleavage site in the *t*-DNA, while in Cas9:pre-cat, it repositions itself toward the *t*-DNA, although remaining at a ~18 Å distance from the cleavage site.

rich helix bridges the two lobes, constituting an anchor for the binding of the guide RNA, while the protein C-terminal (Cterm) and PAM-interacting (PI) domains participate in DNA recognition and binding processes. Extensive structural studies of Cas9 have so far revealed different conformational states of the protein in the apo form, as well as bound to the nucleic acids. By comparing X-ray structures of the apo Cas9[5] and the RNA-bound form (Cas9:RNA),[8] a large conformational rearrangement of the α-helical lobe appears to occur upon RNA binding, thereby priming the enzyme for DNA binding. This overall conformation is largely preserved in the DNA-bound states, which exhibit remarkable differences in the configurations adopted by the catalytic HNH domain. In the presence of an incomplete *nt*-DNA strand (Cas9:RNA:D-NA),[9,10] the HNH domain adopts an inactive conformation, pointing the catalytic site in the opposite direction with respect to the cleavage site on the *t*-DNA; whereas a structural repositioning of HNH is observed in a crystal of Cas9 that includes both unwound DNA strands.[11] This structure is thought to depict a precatalytic state of the system (Cas9:pre-cat), given that a distance of ~18 Å separates the catalytic H840 from the cleavage site on the *t*-DNA, suggesting that a further conformational transition is required for the formation of a catalytically competent Cas9. Overall, structural analysis and biochemical experiments have suggested that conformational plasticity of the protein underlies the processes of nucleic acid association and subsequent cleavage.[7,12,13] However, the conformational changes triggering the binding of the nucleic acids remain speculative. It is unclear, in fact, how the

conformational plasticity of Cas9 would allow the transition from an "open" apo state to a "closed" conformation in which the protein binds its guide RNA and DNA target. Moreover, the conformational dynamics of the HNH domain and, importantly, its activation mechanism leading to the formation of a catalytically active Cas9 is not fully understood, particularly with regard to the role of the nucleic acids and their interactions with Cas9 in facilitating this process. Understanding these mechanistic aspects of the CRISPR-Cas9 system is of paramount importance for the rational design of new and more effective genome-editing tools.[14−17]

In an effort to shed light on these unresolved questions, we decided to exploit the power of high performance computing at the petascale and atomistic molecular dynamics (MD) simulations to obtain key insights and relevant biophysical information otherwise inaccessible with the currently available experimental techniques. Extensive MD simulations (>10 μs in total) have been performed, revealing for the first time the conformational plasticity of Cas9 and suggesting the key determinants that allow for the large-scale conformational changes of Cas9 during the association and processing of the nucleic acids. Our studies reveal an important role of the *nt*-DNA in the process of activation of the HNH nuclease domain, suggesting that the presence of the *nt*-DNA strand within the active site cleft of the RuvC domain triggers local conformational changes that favor the formation of a catalytically competent Cas9. As revealed in the following, these outcomes call for novel experimental efforts aimed at a full characterization and improvement of the CRISPR-Cas9 system.

■ RESULTS AND DISCUSSION

Conformational Dynamics of Cas9 over the Nano-to-Microsecond Time Scale. We performed microsecond-length MD simulations on the available X-ray structures, including Cas9 in the apo form (apo Cas9)[5] and in complex with RNA (Cas9:RNA),[8] with an incomplete DNA (Cas9:RNA:DNA)[9] and complete DNA in a precatalytic state (Cas9:pre-cat)[11] (Figure 1). Molecular simulations have been performed in explicit solvent, adopting a protocol similar to that applied to other phosphodiesterases (full details are reported in the Supporting Information).[18]

Analysis of the root mean square fluctuations (RMSF) of individual Cα atoms reveals high flexibility of the protein domains that mediate the nucleic acid binding (α-helical lobe, PI, and Cterm). The protein flexibility is substantially reduced in the Cas9:pre-cat complex (Figure S1), likely due to a stabilizing effect of the bound nucleic acids. This overall stabilization effect is reflected in the time evolution of the root mean square deviation (RMSD) of the protein, indicating an increase of the protein stability upon binding of the nucleic acids (Figure S2).

Principal component analysis (PCA) has been performed in order to characterize the essential degrees of freedom and large-scale collective motions of the Cas9 protein domains in different states. The dynamics of the protein along the first principal mode of motion (principal component 1, PC1)—usually referred to as "essential dynamics"[19]—is shown in Figure 2, where the arrows indicate the direction and relative amplitude of the motions. In the apo state, we observe large amplitude motions of the protein domains directly involved in the process of association with the nucleic acids: the α-helical lobe accommodating the RNA:*t*-DNA hybrid and its counterpart formed by PI and Cterm domains that binds the DNA.
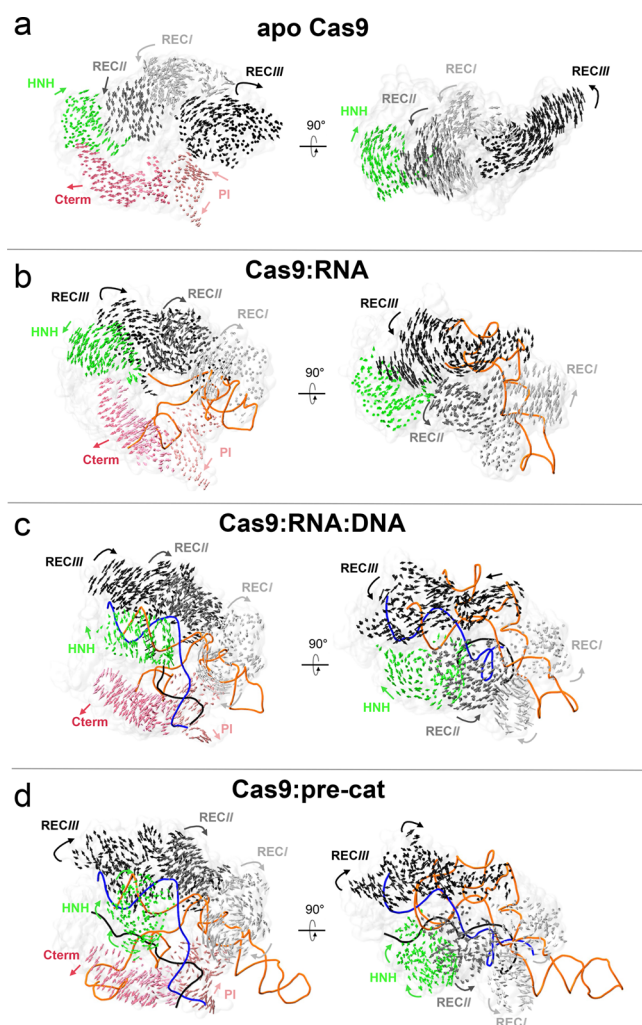
Figure 2. "Essential dynamics", derived from the first principal component (PC1), of the individual protein domains of the apo Cas9 (a), Cas9:RNA (b), Cas9:RNA:DNA (c), and Cas9:pre-cat (d) systems, shown using arrows of sizes proportional to the amplitude of motions. The RNA (orange), target DNA (*t*-DNA, blue), and non-target DNA (*nt*-DNA, black) strands are shown as tubes. For the sake of clarity, the largest amplitude motions are shown, with the Cas9 individual domains color-coded as in Figure 1.

The motions of these domains occur in opposite directions, indicating the tendency toward the "opening" and "closure" of the protein for the accommodation of the nucleic acids (Movie S1). This overall tendency is maintained in the RNA- and DNA-bound states (Movies S2, S3, and S4), although differences are observed in the relative motions of the α-helical domains, due to the conformational transition. By plotting the first versus the second principal components (PC1 vs PC2), we characterize the conformational space sampled by Cas9 in regions in which the protein is "open" and "closed", respectively (Figure 3). This further shows a restriction of the conformational space of Cas9 upon binding of the nucleic acids. Interestingly, in the DNA-bound states, we detect important differences in the dynamics of the HNH domain. In the precatalytic state (Cas9:pre-cat), it moves toward the cleavage site on the target DNA strand, as opposite with respect to the Cas9:RNA:DNA complex, supporting the hypothesis of its dynamic activation toward the catalysis (Figure 4).[11]
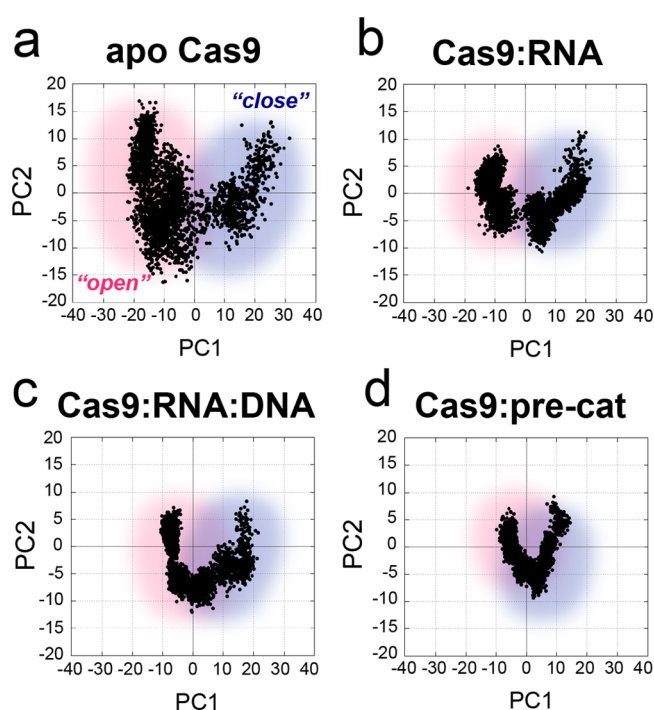

Figure 3. Projections of the first and second principal motions (PC1 vs PC2), derived from MD simulations of the apo Cas9 (a), Cas9:RNA (b), Cas9:RNA:DNA (c), and Cas9:pre-cat (d) systems, characterizing the conformational space sampled by Cas9 into regions in which the protein is "open" (red cloud) and "closed" (blue cloud, Movies S1, S2, S3, and S4).

Correlated Motions of Individual Protein Domains Mediate Nucleic Acid Association. To detect the presence of possible dynamic correlations among different protein domains of Cas9, we performed extensive correlation analyses, including Pearson cross-correlation coefficients (CC_{ij}) and generalized correlation (GC_{ij})[20] that allows for capture of both linear and nonlinear correlations. Results showed that the motions of the Cas9 domains mediating the nucleic acid binding (i.e., α-helical lobe, PI, and Cterm domains) are highly coupled in the apo and RNA-bound forms. They become less correlated upon complete DNA binding, which stabilizes the complex structure (Figures S4 and S5). In order to identify the interdependent motions of the protein regions moving in lockstep (i.e., as characterized by a $CC_{ij} > 0$) or showing opposite motions ($CC_{ij} < 0$), we have computed a per-residue correlation score (Cs_i), which is a measure of the number and intensity of the correlated and anticorrelated motions for each residue (full details in the Supporting Information).[21] Per-residue Cs_i have been accumulated over each protein domain and plotted as a two-by-two matrix, such detailing the interdomain correlations (Figure 5). In the apo form, the Cas9 protein domains show specific patterns of correlated/anticorrelated motions with respect to each other, which characterize their tendency to move concertedly in different directions. Particularly interestingly, the α-helical lobe shows opposite motions of regions REC*II* and REC*III*, while the regions REC*I* and REC*II* move in lockstep, with the Arg-rich helix moving in the same direction as the α-helical region REC*I*. This evidence, together with the fact that α-helical *III* is the most anticorrelated region of Cas9 (Figure S6), well reflects the structural analysis performed by Jiang et al.[8] Indeed, the Cα−Cα vector map of the apo Cas9 versus Cas9:RNA (Figure S7)
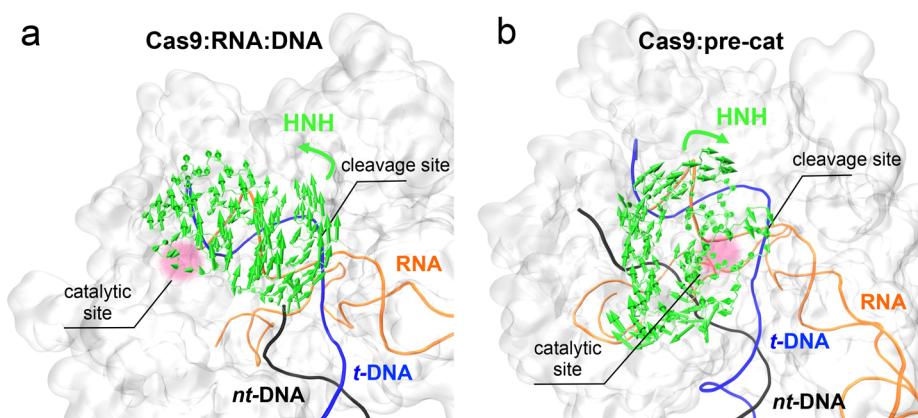
Figure 4. "Essential dynamics" (i.e., first principal component, PC1) of the HNH domain plotted on the protein molecular surface of the Cas9:RNA:DNA (a) and Cas9:pre-cat (b) systems. The RNA (orange), target DNA (*t*-DNA, blue), and non-target DNA (*nt*-DNA, black) strands are shown as tubes. In Cas9:pre-cat, the HNH domain moves toward the *t*-DNA, approaching the catalytic site (indicated using a red cloud) to the cleavage site, in contrast to Cas9:RNA:DNA.
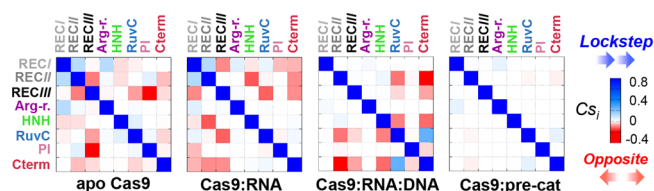

Figure 5. Two-by-two matrices of the accumulated per-residue correlation scores (Cs_i, reported as a normalized frequency), calculated for each protein domain of the four simulated systems. This identifies interdependent domain motions, occurring in lockstep (blue) or in opposite direction (red) with respect to each other (full details in the Supporting Information). The protein sequence is shown along the axes, highlighting individual protein domains with different colors.

suggests that a substantial rearrangement of RECIII would occur in the opposite direction with respect to RECI−II domains during the structural transition of Cas9 from the apo state up to RNA binding. The dynamics of RECIII also anticorrelates with respect to RuvC, PI, and Cterm domains, showing how opposite motions take part in the "opening" and

"closure" of the protein (Movie S1). Although this overall Cs_i pattern is preserved in the RNA-bound state, differences are observed in the RECI−III domains indicative of the conformational transition. Upon DNA binding (Cas9:RNA:DNA), correlations are lost within the α-helical lobe, while its nuclease counterpart shows both correlated and anticorrelated motions involving the catalytic domains that are likely to preclude their activation. This communication supports the hypothesis that the configurational activation of the HNH domain is assisted by an allosteric "crosstalk" with the RuvC domain.[13] The precatalytic state is characterized by weak correlations, as also detected via visual inspection of the CC_{ij}/GC_{ij} correlation matrices (Figures S4 and S5).

Overall, characteristic motions of the individual protein domains, as well as their "essential dynamics", indicate their propensity to move concertedly in different directions, thus allowing the structural transition leading to the nucleic acid association. Indeed, the observed global tendency toward the "opening" and "closure" of the protein relies on coupled
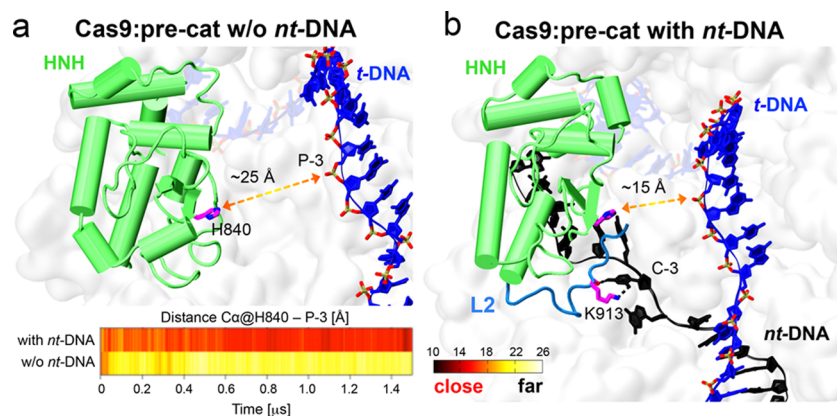
Figure 6. Representative snapshots of Cas9:pre-cat without the non-target DNA (w/o *nt*-DNA) (a) and with the *nt*-DNA (b), from MD simulations. In the absence of the *nt*-DNA, the catalytic H840 moves to a distance of ~25 Å from the scissile phosphate P-3 on the target DNA (*t*-DNA). In the presence of the *nt*-DNA, H840 approaches P-3 at ~15 Å distance. Concurrently, the K913 residue in the L2 loop forms H-bonds with C-3 in the *nt*-DNA. The protein is shown in molecular surface, highlighting the HNH domain (green) and the L2 loop (blue, right panel) as cartoon. The *t*-DNA (blue) and *nt*-DNA (black) strands are shown as ribbons. The RNA is omitted for the sake of clarity. Key protein residues (e.g., H840 and K913) are shown as sticks. The bottom graph reports time evolution of the distance between H840 (Cα atom) and the scissile phosphate P-3, during MD simulations of Cas9:pre-cat with *nt*-DNA (top) and without *nt*-DNA (bottom), color-coded according to the scale on the right.

motions of the individual Cas9 domains, and underlies the prominent conformational changes of the binding process.

Key Role of the Non-target DNA in the Activation of the HNH Domain. To date, extensive biochemical studies have suggested that a tight interplay between Cas9 and the nucleic acids plays a role in the activation of the HNH domain, whose conformational dynamics, in turn, directly controls the cleavage of the double stranded DNA.[13] However, although it has been shown that the presence of the RNA:t-DNA hybrid is critical for the HNH conformational activation,[13] it is unclear how the nt-DNA would be involved in this mechanism. To shed light on this unresolved question, we performed molecular simulations of Cas9:pre-cat in the absence of the nt-DNA (Cas9:pre-cat w/o nt-DNA), thus clarifying the effect of the nt-DNA on the dynamics of the precatalytic state. During the simulations, the HNH domain moves far apart from the catalytic site on the t-DNA, with the catalytic H840 reaching a distance of ~25 Å (initially ~18 Å) from the scissile phosphate P-3 (Figure 6, Movie S5). This overall motion of the HNH domain is reflected in its "essential dynamics" (Figure S8), and is opposite to what was observed in the presence of the nt-DNA, where the catalytic domain moves toward the cleavage site on the t-DNA strand and stabilizes at a distance of ~15 Å from the scissile P-3 (Figures 4, 6). These molecular simulations strongly suggest the presence of the nt-DNA as a key factor for the conformational activation of the HNH domain. With the aim of deciphering the molecular determinants connecting the presence of the nt-DNA to the approaching of the HNH domain to the cleavage site, we looked at interactions, on the atomic scale, established by the nt-DNA during the dynamics of Cas9:pre-cat. We identified an extended network of H-bonding interactions with the hinge regions L1 (residues 765−780) and L2 (residues 906−918) that link the HNH and RuvC domains (Figure S9). In detail, while the interaction between the nt-DNA with L1 is characteristic of the X-ray structure, as well as being conserved during MD, stable H-bonding interactions between C-3 in the nt-DNA and K913 in the L2 loop occur over the time scale of ~0.75 μs, thus stabilizing the HNH catalytic site a distance of ~15 Å from the scissile phosphate on the t-DNA. The contacts of the nt-DNA with the L1 and L2 loops further result in an overall stabilization of the system, justifying the occurrence of correlated motions of lower intensity with respect to the system lacking the nt-DNA (Figure S8).

These findings suggest that the nt-DNA plays a key role in the conformational activation of the HNH domain toward the catalysis of the t-DNA strand. Indeed, the occurrence throughout the dynamics of the above-described interactions clarifies how the nt-DNA positioning within the RuvC groove would trigger local conformational changes, which result in the approach of the HNH active site to the scissile P-3 and formation of a catalytically competent Cas9.[11]

Interestingly, the structural superpositions of the 5F9R (Cas9:pre-cat)[11] X-ray structure, which includes the nt-DNA, with the 4UN3 (Cas9:RNA:DNA)[9] and 4ZT0 (Cas9:RNA)[8] structures reveal a steric clash between the nt-DNA in Cas9:pre-cat and the HNH domain in the other structures, with the L1 and L2 loops precluding the binding of the nt-DNA within the active site cleft of the RuvC domain (Figure S10). These observations suggest that nt-DNA binding within the RuvC cleft would occur during or upon HNH repositioning to the precatalytic state (Cas9:pre-cat). Concurrently, key interactions between the nt-DNA with the L2 loop—as

revealed from MD simulations—would trigger the last step of conformational activation with the approach of the catalytic H840 toward the cleavage site in the t-DNA. Overall, considering that the HNH domain repositioning—as observed in the available crystal structures[8−11]—is accompanied by the reorientation of the L1 and L2 loops, our simulations detail how the communication between RuvC—hosting the nt-DNA—and HNH relies on the interaction between the interconnected regions and the nt-DNA strand. This information depicts the mechanics of the communication between the HNH and RuvC catalytic domains, revealing that their allosteric "crosstalk"[11,13] is dependent on the presence of the nt-DNA strand. Indeed, the interaction between this latter with the L1 and L2 loops explains how these hinge regions would act as "signal transducers" during the catalytic activation.[13] These identified allosteric effects are of particular interest in light of the wider scope of computational biophysics in deciphering protein allostery.[22−24] Besides this, and more importantly, the interplay between L2 and the nt-DNA strand—which has been here identified as a key element for the activation process—calls for novel mutagenesis and kinetic experiments, in an effort to structurally engineer Cas9 for achieving higher efficiency.

Conformational Plasticity of HNH. With the final goal of gaining a more comprehensive picture of the conformational plasticity of the HNH domain, we performed MD simulations of the Cas9:RNA:DNA and Cas9:pre-cat systems—which differ in the orientation of the HNH domain—after removing the nucleic acids. In both systems, the HNH domain displays a high conformational mobility, as indicated by particularly high RMSF values (Figure S11). During the dynamics of Cas9:RNA:DNA, the HNH domain undergoes a conformational shift toward the direction of the configuration observed in the X-ray structure of the RNA-bound state (Figure S12),[8] while an opposite conformational transition is observed in Cas9:pre-cat. These outcomes reveal that a high conformational flexibility is intrinsic to the HNH domain and constitutes the necessary element that allows HNH repositioning during nucleic acid binding and processing. Moreover, this striking conformational plasticity suggests that the HNH domain can adopt multiple conformational states during its activation process, in agreement with Förster resonance energy transfer experiments showing that the HNH domain exists in a conformational equilibrium between the inactive and active states.[13] The observed conformational plasticity of HNH is a key insight that has to be considered in light of the critical role of the nt-DNA strand during the HNH activation, as revealed from MD simulations, and considering the inability of the nt-DNA to bind the RuvC without HNH repositioning (Figure S10).[8,9,11] By taking together previous experimental observations and our molecular simulations, it is tempting to speculate that the conformational changes in the HNH domain might intervene or facilitate the process of DNA binding and double strand separation (i.e., R-loop formation). This hypothesis calls for additional experimental investigations aimed at clarifying the binding mode of DNA prior to unwinding.

■ CONCLUSIONS

In summary, long time scale MD simulations, performed over the multi-microsecond time scale (>10 μs), reveal the conformational plasticity of the CRISPR-Cas9 system and identify the key dynamic determinants underlying the large-scale conformational changes that occur during the nucleic acid

association and processing. We show how the "closure" of the protein, which accompanies the nucleic acid binding, fundamentally relies on highly coupled and specific motions of the protein domains that collectively initiate the prominent conformational changes necessary for the nucleic acid association. In light of the experimental observations of a tight interplay between Cas9 and the nucleic acids,[11,13] we reveal that the activation of the HNH domain for catalysis of the *t*-DNA cleavage depends on the presence of the *nt*-DNA strand within the groove of the RuvC domain, thus identifying the *nt*-DNA strand as a key determinant for the conformational activation. Our simulations show that the presence of the *nt*-DNA within the RuvC groove triggers local conformational changes that result in a shift of the HNH active site toward the cleavage site on the *t*-DNA for catalysis. Moreover, the major evidence of a critical role, of the interaction between the L2 loop—which connects the nuclease domains—and the *nt*-DNA strand calls for novel experimental efforts (i.e., mutagenesis and kinetic experiments) aimed at assessing how protein structure engineering at the level of the L2 loop could help in the design of more efficient Cas9. Finally, a remarkable conformational plasticity of the HNH domain is identified over the dynamics, suggesting the flexibility as a necessary element for HNH repositioning. Overall, by performing extensive molecular simulations of the CRISPR-Cas9 system, we provide for the first time a dynamic picture and an atomic-level explanation of the conformational plasticity of this unique genome-editing engine. The novel insights arising from molecular simulations provide a foundation for further experimental studies aimed at a full characterization of the dynamic features of the CRISPR-Cas9 system.

■ MATERIALS AND METHODS

Structural Models. MD simulations have been performed on four model systems of the Cas9 in apo form (apo Cas9)[5] and in complex with RNA (Cas9:RNA),[8] with an incomplete DNA (Cas9:RNA:DNA)[9] and in a precatalytic state (Cas9:pre-cat)[11] including both DNA strands. These model systems have been prepared using the crystallographic coordinates of the *Streptococcus pyogenes* apo Cas9 (4CMQ),[5] Cas9:RNA (4ZT0),[8] Cas9:RNA:DNA (4UN3),[9] and Cas9:pre-cat (5F9R),[11] solved at 3.09, 2.50, 2.58, and 3.40 Å resolution, respectively. In order to study the effect of the *nt*-DNA on the dynamics of the precatalytic state, a fifth model system has been built, deleting the *nt*-DNA strand from Cas9:pre-cat (Cas9:pre-cat w/o *nt*-DNA). Moreover, with the purpose of studying the conformational dynamics of HNH in the absence of the nucleic acids, two additional model systems have been built deleting the nucleic acids from the DNA bound states (i.e., Cas9:RNA:DNA and Cas9:pre-cat, corresponding to the PDB codes 4UN3 and 5F9R), which differ in the orientation of the HNH domain. A total of 7 model systems have been embedded in explicit waters, leading to orthorhombic periodic simulation cells of $107 \times 158 \times 138$ Å^3 (apo Cas9, for a total of ~220 K atoms), ~$148 \times 107 \times 140$ Å^3 (Cas9:RNA, ~210 K atoms), ~$144 \times 108 \times 146$ Å^3 (Cas9:RNA:DNA, ~216 K atoms), ~$180 \times 116 \times 139$ Å^3 (Cas9:pre-cat with and w/o *nt*-DNA, ~270 K atoms each), and ~$136 \times 103 \times 144$ Å^3 (Cas9:RNA:DNA and Cas9:pre-cat w/o nucleic acids, ~190 K atoms each). Full details are reported in the Supporting Information.

MD Simulations. The above-mentioned model systems have been equilibrated and production runs have been performed using the Amber ff12SB force field, which includes the ff99bsc0 corrections for DNA[25] and the ff99bsc0+χOL3 corrections for RNA.[26,27] The Åqvist[28] force field parameters for the Mg ions been employed, which favor an octahedral coordination for the Mg ion. The here employed computational protocol has been previously employed in studies on similar protein/nucleic acid systems,[18,29,30,31] performing the catalysis of the DNA via a "two-metal aided" mechanism,[32] as suggested for Cas9.[5] The TIP3P model has been employed for waters.[33] A salt concentration of 0.08 mM of NaCl has been considered, in agreement with the experimental conditions of cleavage assays.[9,34] Hydrogen atoms were added assuming standard bond lengths and were constrained to their equilibrium position with the RATTLE[35] algorithm. All MD simulations have been performed with NAMD 2.10.[36] MD simulations have been performed in the isothermal−isobaric (NPT) ensemble by using a time step of 2 fs. The systems have been coupled to a Langevin thermostat at 298 K and barostat at 1 atm. Periodic boundary conditions were applied. The particle mesh Ewald (PME)[37] method was used to evaluate long-range electrostatic interactions, and a cutoff of 12 Å was used to account for the van der Waals interactions. All the simulations were carried out with the following protocol. First, the systems were subjected to energy minimization by using the steepest descent algorithm. Then, the systems were thermalized up to physiological temperature in the canonical ensemble (NVT) using a Langevin bath in three consecutive steps: (1) the solvent was first equilibrated over ~10 ps of MD, slowly increasing the temperature from 0 to 100 K and maintaining both the protein and the nucleic acids fixed; (2) the temperature was further increased up to 200 K over ~10 ps of MD, while keeping fixed only the coordinates of backbone atoms of the protein/nucleic acid complex; (3) constraints were released, and the systems were simulated for ~25 ps of MD to reach the temperature of 298 K. Then, we switched to the NPT statistical ensemble, performing ~100 ps of MD at 298 K. After this initial phase, equilibration runs were carried out in the NPT statistical ensemble, obtaining ~40 ns of MD at 298 K. Production runs have been carried out reaching ~1.5 μs for each system, for a total of >10 μs of classical MD (i.e., ~1.5 μs × 7 systems). Coordinates of the systems were collected every 10 ps for a total of ~150,000 up to 160,000 frames for each run.

Analysis of the Results. PCA has been employed to capture the essential motions of the simulated systems. In PCA, the covariance matrix of the protein Cα atoms is calculated and diagonalized to obtain a reduced set of coordinates (eigenvectors) to describe the system motions. Each eigenvector—also called principal component (PC)—is associated with an eigenvalue corresponding to the mean square fluctuation of the system with its trajectory projected to that eigenvector. By sorting the eigenvectors according to their eigenvalues, the first principal component (PC1) corresponds to the system's largest amplitude motion, and the dynamics of the system along PC1 is usually referred to as "essential dynamics".[19] In this work, each structure arising from the MD trajectories is projected into the collective coordinate space defined by the first two eigenvectors (PC1 and PC2), thus allowing the characterization of the conformational space sampled by Cas9 during MD. Importantly, in order to identify differences in the essential structural-dynamic properties of Cas9, each simulated system has been superposed onto the same reference structure (i.e., considering as a reference the RuvC and Cterm domains that do not show relevant

conformational differences among the crystallized states) and aligned to allow projection into the same collective coordinate space. PCA has been performed using the GROMACS 4.4.5 suite of analysis codes.[38] Specifically, the g_covar program has been employed for the construction and diagonalization of the covariance matrix. Subsequently, the program g_anaeig has been used to analyze and visualize the eigenvectors. Figure 2 has been produced using the Normal Mode Wizard (NMWiz) plugin of the Visual Molecular Dynamics (VMD) molecular visualization program.[39] Full details are reported in the Supporting Information.

Correlation Analyses. The cross-correlation matrix CC_{ij}— based on Pearson coefficients—between the fluctuations of the $C\alpha$ atoms relative to their average positions has been used in order to identify the coupling of the motions between the protein residues. In addition, we performed generalized-correlation (GC_{ij})[20] analysis, which is independent of the relative orientation of the atomic fluctuations and allows capturing nonlinear correlations. Cross-correlation score (Cs_i) coefficients have also been calculated, as a measure of the number and intensity of the correlated and anticorrelated motions for each residue (full details in the Supporting Information).[21] Full details on correlation analyses are reported in the Supporting Information.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.6b00218.

Materials and methods, results, references, figures, and complete captions for the movies (PDF)

Movie S1, "essential dynamics" of the apo Cas9 (MPG)

Movie S2, "essential dynamics" of the Cas9:RNA system (4ZT0) (MPG)

Movie S3, "essential dynamics" of the Cas9:RNA:DNA system (4UN3) (MPG)

Movie S4, "essential dynamics" of the Cas9:pre-cat system (5F9R) (MPG)

Movie S5, MD simulations of Cas9:pre-cat without *nt*-DNA 5F9R) (MPG)

Movie S6, MD simulations of Cas9:pre-cat including both DNA strands (5F9R) (MPG)

AUTHOR INFORMATION

Corresponding Author
*E-mail: gpalermo@ucsd.edu.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) 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.

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

(3) Wright, A. V.; Nunez, J. K.; Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **2016**, *164*, 29−44.

(4) Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157*, 1262−1278.

(5) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014**, *343*, 1247997.

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

(7) Jiang, F.; Doudna, J. A. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **2015**, *30*, 100−111.

(8) Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **2015**, *348*, 1477−1481.

(9) 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.

(10) 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.

(11) Jiang, F. G.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K. H.; 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.

(12) Vazquez Reyes, C.; Tangprasertchai, N. S.; Yogesha, S. D.; Nguyen, R. H.; Zhang, X.; Rajan, R.; Qin, P. Z. Nucleic Acid-Dependent Conformational Changes in CRISPR-Cas9 Revealed by Site-Directed Spin Labeling. *Cell Biochem. Biophys.* **2016**, DOI: 10.1007/s12013-016-0738-5.

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

(14) Wright, A. V.; Sternberg, S. H.; Taylor, D. W.; Staahl, B. T.; Bardales, J. A.; Kornfeld, J. E.; Doudna, J. A. Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2984−2989.

(15) Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Topkar, V. V.; Nguyen, N. T.; Zheng, Z. L.; Gonzales, A. P. W.; Li, Z. Y.; Peterson, R. T.; Yeh, J. R. J.; Aryee, M. J.; Joung, J. K. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **2015**, *523*, 481−485.

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

(17) Hirano, S.; Nishimasu, H.; Ishitani, R.; Nureki, O. Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol. Cell* **2016**, *61*, 886−894.

(18) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.* **2015**, *48*, 220−228.

(19) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.

(20) Lange, O. F.; Grubmuller, H. Generalized correlation for biomolecular dynamics. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 1053−1061.

(21) Ricci, C. G.; Silveira, R. L.; Rivalta, I.; Batista, V. S.; Skaf, M. S. Allosteric Pathways in the PPAR gamma-RXR alpha nuclear receptor complex. *Sci. Rep.* **2016**, *6*, 19940.

(22) Guo, J.; Zhou, H. X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116*, 6503−6515.

(23) Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **2016**, *116*, 6370−6390.

(24) Bowerman, S.; Wereszczynski, J. Chapter Seventeen − Detecting Allosteric Networks Using Molecular Dynamics Simulation. *Methods Enzymol.* **2016**, *578*, 429−447.

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

(26) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.

(27) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.

(28) Aqvist, J. Ion-Water interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.

(29) Casalino, L.; Palermo, G.; Rothlisberger, U.; Magistrato, A. Who Activates the Nucleophile in Ribozyme Catalysis? An Answer from the Splicing Mechanism of Group II Introns. *J. Am. Chem. Soc.* **2016**, *138*, 10374−10377.

(30) Palermo, G.; Stenta, M.; Cavalli, A.; Dal Peraro, M.; De Vivo, M. Molecular Simulations Highlight the Role of Metals in Catalysis and Inhibition of Type II Topoisomerase. *J. Chem. Theory Comput.* **2013**, *9*, 857−862.

(31) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester cleavage in ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J. Am. Chem. Soc.* **2008**, *130*, 10955−10962.

(32) Steitz, T. A.; Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6498−6502.

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(34) O'Connell, M. R.; Oakes, B. L.; Sternberg, S. H.; East-Seletsky, A.; Kaplan, M.; Doudna, J. A. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **2014**, *516*, 263.

(35) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24−34.

(36) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(37) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

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

(39) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.