

# CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations

Giulia Palermo<sup>a,b,1</sup>, Yinglong Miao<sup>a,c,d</sup>, Ross C. Walker<sup>e,2</sup>, Martin Jinek<sup>f</sup>, and J. Andrew McCammon<sup>a,b,c,d,e,1</sup>

<sup>a</sup>Howard Hughes Medical Institute, University of California at San Diego, La Jolla, CA 92093; <sup>b</sup>Department of Pharmacology, University of California at San Diego, La Jolla, CA 92093; <sup>c</sup>Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093; <sup>d</sup>National Biomedical Computation Resource, University of California at San Diego, La Jolla, CA 92093; <sup>e</sup>San Diego Supercomputer Center, University of California at San Diego, La Jolla, CA 92093; and <sup>f</sup>Department of Biochemistry, University of Zürich, CH-8057 Zürich, Switzerland

Contributed by J. Andrew McCammon, June 1, 2017 (sent for review May 8, 2017; reviewed by James Briggs and Chung Wong)

CRISPR-Cas9 has become a facile genome editing technology, yet the structural and mechanistic features underlying its function are unclear. Here, we perform extensive molecular simulations in an enhanced sampling regime, using a Gaussian-accelerated molecular dynamics (GaMD) methodology, which probes displacements over hundreds of microseconds to milliseconds, to reveal the conformational dynamics of the endonuclease Cas9 during its activation toward catalysis. We disclose the conformational transition of Cas9 from its apo form to the RNA-bound form, suggesting a mechanism for RNA recruitment in which the domain relocations cause the formation of a positively charged cavity for nucleic acid binding. GaMD also reveals the conformation of a catalytically competent Cas9, which is prone for catalysis and whose experimental characterization is still limited. We show that, upon DNA binding, the conformational dynamics of the HNH domain triggers the formation of the active state, explaining how the HNH domain exerts a conformational control domain over DNA cleavage [Sternberg SH et al. (2015) *Nature*, 527, 110–113]. These results provide atomic-level information on the molecular mechanism of CRISPR-Cas9 that will inspire future experimental investigations aimed at fully clarifying the biophysics of this unique genome editing machinery and at developing new tools for nucleic acid manipulation based on CRISPR-Cas9.

protein–nucleic acid interactions | gene regulation | RNA dynamics | enhanced sampling | free energy

Life sciences are undergoing a transformative phase due to an emerging genome editing technology based on the RNA-programmable CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats–CRISPR-associated protein 9) system (1–3). Although this facile genome editing technology is revolutionizing the fields of medicine, pharmaceuticals, and even agriculture with the development of drought-resistant crops, the structural and mechanistic details underlying the CRISPR-Cas9 function remain unclear. The CRISPR-Cas9 function begins with the association of Cas9 with a guide RNA, composed of a CRISPR RNA (crRNA) and a transactivating CRISPR RNA (tracrRNA), which enables the recognition and cleavage of matching sequences in double-stranded DNA (3, 4). Upon site-specific recognition of a Protospacer Adjacent Motif (PAM) within the DNA, the latter binds Cas9 matching one strand with the RNA guide (the target DNA strand, *t*-DNA), whereas the other strand (nontarget DNA, *nt*-DNA) is displaced. Subsequently, two nuclease domains—HNH and RuvC—perform site-specific cleavages of the *t*-DNA and *nt*-DNA strands, respectively.

Structural studies have revealed that Cas9 is a large multidomain protein, composed of a recognition lobe, which mediates the nucleic acid binding through three regions (*REC1–III*), and a nuclease lobe including the RuvC and HNH catalytic cores (Fig. 1) (5). An arginine rich helix (R-rich) bridges the two lobes, whereas the protein C-terminal (Cterm) and PAM-interacting (PI) domains take part in the DNA binding process. In the *apo* state (apo Cas9) (6), Cas9 adopts an auto-inhibited conformation, which undergoes a dramatic structural transition in the *REC1–III* regions upon RNA

association (Cas9:RNA) (7). Although this overall conformation is preserved in the DNA-bound X-ray structures, the HNH catalytic domain adopts different configurations (*SI Appendix*, Fig. S1). In the presence of an incomplete DNA (Cas9:DNA) (8, 9), the catalytic H840 is located ~30 Å from the cleavage site on the *t*-DNA. Conversely, in the presence of the *nt*-DNA strand in the RuvC domain, the HNH domain directs the catalytic H840 toward the scissile phosphate on the *t*-DNA strand (10). This structure, however, does not capture the actual catalytic state, H840 being at ~15 Å from the *t*-DNA, whereas the *nt*-DNA strand and the RuvC active site are at ~5.5 Å apart, therefore likely representing a precatalytic enzymatic state (Cas9:precat).

Although critically informative, structural studies have revealed snapshots of the complex conformational landscape of the system, raising fundamental questions whose answer is necessary for complete mechanistic understanding of the CRISPR-Cas9 system. In particular, the structural transitions leading the protein from the “open” apo state to a “closed” conformation in which the enzyme binds RNA are unknown. Moreover, information about the conformational activation of the HNH domain is elusive. Förster resonance energy transfer (FRET) experiments have shown that the conformational activation of the HNH domain controls DNA cleavage (11, 12), but it is unclear how the protein structural changes in concert with DNA binding would allow its conformational activation toward catalysis. By using molecular

## Significance

The CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats–CRISPR-associated protein 9) system recently emerged as a transformative genome editing technology that is innovating life science, with cutting-edge impact in biomedicine, pharmaceuticals, and agriculture. Nevertheless, the molecular mechanism underlying CRISPR-Cas9 function is still incompletely understood. Here, enhanced molecular dynamics (MD) simulations, probing displacements over long timescales, capture at atomic level specific features that are difficult to reach via conventional MD simulations and via the currently available experimental techniques, clarifying the molecular mechanism of CRISPR-Cas9, with understanding of its activation process. The insights obtained from our molecular simulations provide key reference points for future experimental studies of CRISPR-Cas9 and its applications as a genome editing tool.

Author contributions: G.P., M.J., and J.A.M. designed research; G.P. performed research; G.P., Y.M., and R.C.W. analyzed data; and G.P., Y.M., and J.A.M. wrote the paper.

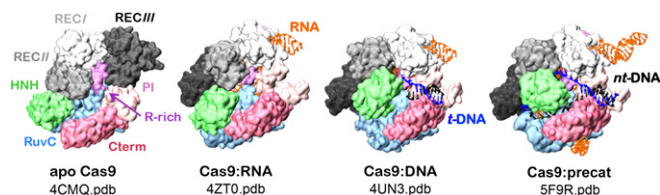
Reviewers: J.B., University of Houston; and C.W., University of Missouri–St. Louis.

Conflict of interest statement: R.C.W. became a full-time employee of GlaxoSmithKline (GSK) in November 2016. This research was conducted prior to R.C.W. joining GSK and was not funded by GSK.

<sup>1</sup>To whom correspondence may be addressed. Email: jmccammon@ucsd.edu or gpalermo@ucsd.edu.

<sup>2</sup>Present address: GlaxoSmithKline, Collegeville, PA 19426-0989.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707645114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707645114/-DCSupplemental).



**Fig. 1.** X-ray structures of the endonuclease Cas9 as apo form (apo Cas9, 4CMQ.pdb) (6), in complex with a guide RNA (Cas9:RNA, 4ZT0.pdb) (7), with an incomplete DNA (Cas9:DNA, 4UN3.pdb) (8) and with both DNA strands in a precatalytic state (Cas9:precat, 5F9R.pdb) (10). Cas9 is shown in molecular surface view, highlighting individual protein domains in different colors. The RNA (orange), target DNA (t-DNA, blue), and nontarget DNA (nt-DNA, black) are shown as ribbons.

dynamics (MD) simulations, we have recently portrayed the conformational dynamics of Cas9 as crystallized in the apo form and in complex with the nucleic acids (13). However, due to the short timescales accessible by conventional MD, this mechanistic study could not fully characterize the long-timescale conformational transitions arising from the interplay between protein and nucleic acids (12, 14).

Here, we report an extensive computational study, based on enhanced MD simulations ( $\sim 15 \mu\text{s}$  in length), that investigates the Cas9 conformational activation, from the apo protein to the final catalytic state via the binary Cas9-guide RNA intermediate and the target DNA-bound complexes. We perform accelerated MD simulations, which have been demonstrated to provide routine access to events occurring on and—in some cases—beyond the millisecond timescale (15, 16), complemented by the application of biasing methods (17), enabling the investigation of intermediate states for which structural characterization is absent. We use a Gaussian-accelerated MD (GaMD) (18) method that accelerates biomolecular conformational transitions between low-energy states by smoothing the potential energy surface with a harmonic boost potential that follows a Gaussian distribution. GaMD has shown robust sampling capability and accurate reconstruction of free-energy profiles for simulations of large biomolecules. It has enabled characterization of ligand binding and conformational changes in G protein-coupled receptors (19) as well as protein folding (20).

In the following, we disclose the conformational changes occurring during CRISPR-Cas9 activation, characterizing the conformational transition of Cas9 from its apo form to the RNA-bound form and suggesting a mechanism for RNA recruitment. We also depict crucial conformational changes of the HNH domain, leading to the characterization of a catalytically competent Cas9 conformation primed for catalysis. Overall, we provide fundamental understanding of the CRISPR-Cas9 mechanistic action with key insights for future experimental studies of CRISPR-Cas9 and its applications as a genome editing tool (21–23).

## Results

Here, GaMD simulations are used to describe the step-by-step dynamics of Cas9 during its activation process (18). To obtain an initial path among the crystallized states and structural information of intermediate starting points that could not be obtained experimentally, we use targeted MD (TMD), in which the root-mean-square deviation (rmsd) between initial and target conformations is reduced (17). We recover the backward Cas9 activation trajectory, from the most complete structure of Cas9 in complex with both nucleic acids (Cas9:precat) up to the apo Cas9, by gradually removing the nucleic acid components, such as following the structural transitions of the protein that would occur upon dissociation of the nucleic acids. Additionally, a putative catalytic state has been obtained by targeting Cas9:precat into a model

structure proposed by Sternberg et al. (11), based on FRET experiments and structural comparisons (*SI Appendix*, Figs. S2 and S3). These simulations, which are detailed in *SI Appendix*, provided starting points, which could not be obtained experimentally, for more rigorous GaMD simulations, while also revealing some key aspects of the protein conformational transitions that we describe below.

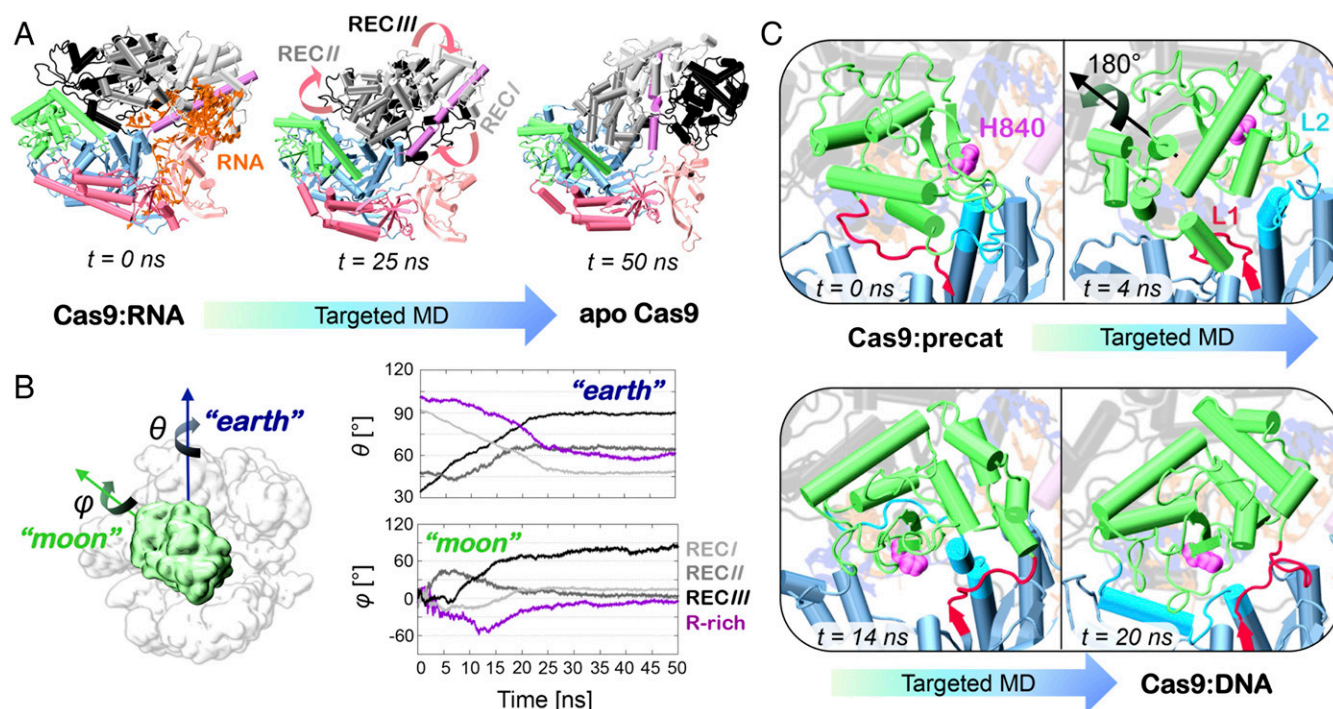
### Conformational Transition of the Apo Cas9 to the RNA-Bound State.

A dramatic conformational change is observed between the apo and RNA-bound structures of Cas9, with a  $\sim 28 \text{ \AA}$  (Fig. 1) (6). To investigate this conformational change, Cas9:RNA was targeted to the apo Cas9 after removing the RNA. TMD captures the process in  $\sim 50 \text{ ns}$ , revealing a conformational rearrangement mainly involving the recognition lobe. The REC I–III regions move in opposite directions relative to each other, leading to the opening of the REC lobe with respect to the nuclease lobe (Fig. 2A and *Movie S1*). An important conformational change is observed in the R-rich helix, which moves together with REC I, in the opposite direction with respect to REC III, flanking the latter in the final apo state. Remarkably, these simulations confirm previous hypotheses based on structural analysis (7) and MD simulations (13) as well as coarse-grained modeling of CRISPR-Cas9 (24), indicating the propensity of the protein domains to move concertedly in different directions as a key factor during nucleic acid association. During the simulations, the Cas9 conformational changes were tracked using a model we refer to as an “earth & moon” model of the system (Fig. 2B and *SI Appendix*, Fig. S4), in which each individual protein domain can rotate around the main protein axis, like a satellite rotation around the earth ( $\theta$  angle) and itself ( $\phi$  angle, details are in *SI Appendix*). In this analysis, we detect a large rotation of REC III with a  $\theta$  angle of  $\sim 60^\circ$  with respect to the protein and  $\phi$  of  $\sim 90^\circ$  with respect to itself, whereas the rotations of the REC II region occur with a smaller extent. The REC I region moves in the opposite direction with respect to REC III and together with the R-rich helix, whereas the nuclease lobe remains stable (*SI Appendix*, Fig. S5) (7, 13, 24).

### Conformational Change of the HNH Domain in the DNA-Bound States.

In the DNA-bound state, the catalytic HNH domain has been crystallized in two opposite configurations (*SI Appendix*, Fig. S1). To investigate the conformational transition of the HNH domain between these two states, we performed TMD simulations targeting Cas9:precat into Cas9:DNA, after removing the nt-DNA strand. As a result, we observe that the HNH catalytic domain rotates by  $\sim 180^\circ$  around itself, reaching its turned configuration by rotating toward the RNA:t-DNA hybrid (Fig. 2C and *Movie S2*). The two hinge regions connecting the HNH and RuvC domains—i.e., the L1 (residues 765–780) and L2 (residues 906–918) loops—also undergo local structural rearrangement, the L2 loop establishing critical H-bonding interactions with the RNA:t-DNA hybrid (*SI Appendix*, Fig. S7). Aiming to understand the role of the nucleic acids during the HNH repositioning, we have simulated the HNH conformational change by targeting the Cas9:precat into Cas9:DNA, after removing both DNA strands (i.e., in the presence of the RNA only) as well as removing the nucleic acids entirely (*SI Appendix*, Fig. S8). As a result, in the absence of interactions between the L2 linker and the RNA:t-DNA hybrid, the L1 and L2 linkers engage in an unphysical overlap that precludes the complete turn of the HNH domain. The latter moves in the opposite direction with respect to what was observed in the presence of the RNA:t-DNA hybrid and with a  $\phi$  angle of only  $\sim 50^\circ$ . These observations hint at a critical role of the L2 loop, which initiates proper conformational transition of the HNH domain by interacting with the RNA:t-DNA hybrid (10). Finally, we also performed TMD simulations of Cas9:precat targeted to Cas9:DNA including all nucleic acid chains (i.e., RNA, t-DNA, and nt-DNA), revealing





**Fig. 2.** (A) Representative snapshots along TMD of Cas9:RNA to the apo Cas9 (shown at time  $t = \sim 0$ ,  $\sim 25$ , and  $\sim 50$  ns). Cas9 is shown in cartoon, color-coded as in Fig. 1. Arrows are used to indicate the conformational changes of the REC-I/III regions and of the R-rich helix, which undergo major conformational transitions. (B) Schematic view of an earth and moon model of the system, where the  $\theta$  and  $\phi$  angles describe the rotation of each individual protein domain (here shown for HNH) with respect to the protein ( $\theta$ ) and themselves ( $\phi$ ) (SI Appendix). The graphs report the time evolution of the  $\theta$  ("earth") and  $\phi$  ("moon") angles for the regions REC-I/III and for the R-rich helix, along TMD. (C) Representative snapshots along TMD of Cas9:precat to Cas9:DNA (shown at  $t = \sim 0$ ,  $\sim 4$ ,  $\sim 14$ , and  $\sim 20$  ns). During the simulation, the HNH domain (green) turns  $\sim 180^\circ$  around itself, whereas the catalytic H840 (magenta) changes configuration facing the RNA:t-DNA hybrid (SI Appendix, Fig. S8). The L1 (red) and L2 (cyan) loops also show remarkable structural transitions.

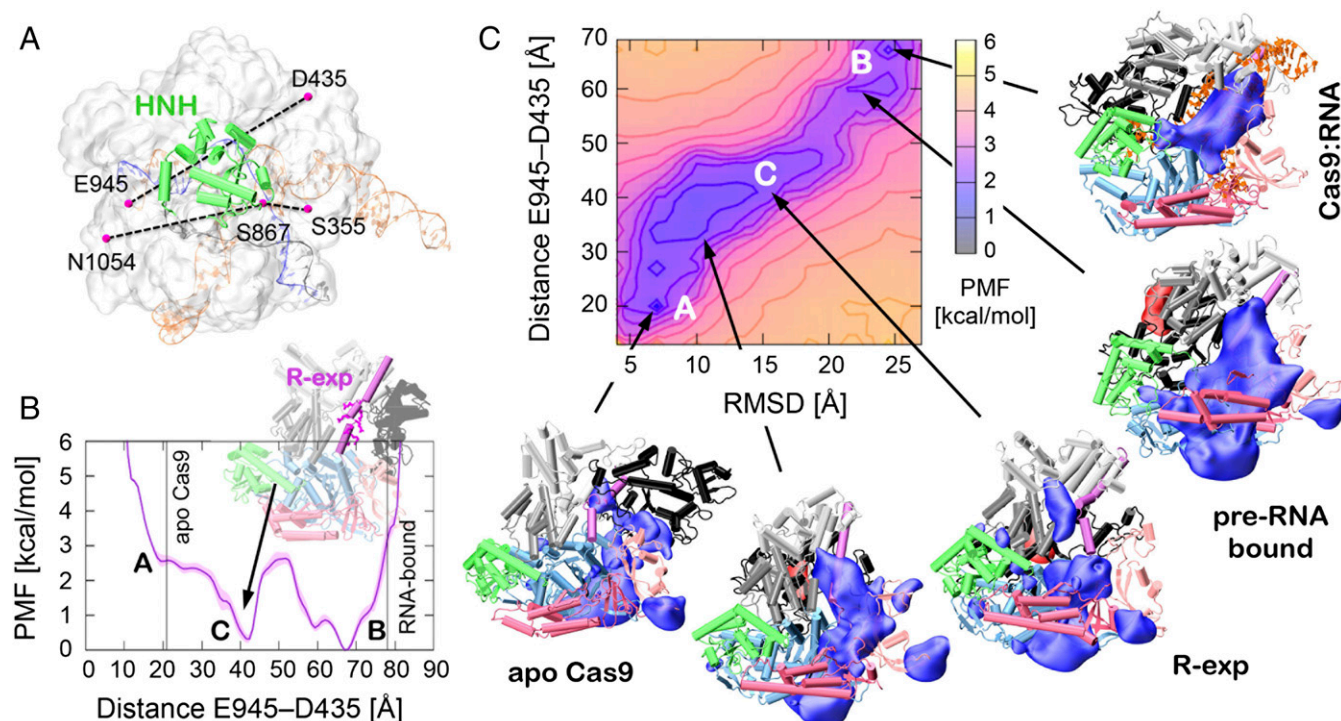
that the presence of the *nt*-DNA within the RuvC groove exerts a steric constraint that hampers the structural transition of HNH ( $\phi$  does not change phase). This indicates that the transition of the HNH domain is unlikely to occur after complete DNA unwinding and relocation for the catalysis, suggesting that HNH repositioning might occur during the process of double-strand separation (7, 11, 13).

#### Observation of the Step-by-Step Conformational Dynamics of Cas9.

TMD simulations provided an initial pathway for Cas9 conformational activation between the crystallized states and a set of equally distributed frames among them, which could not be obtained experimentally. Along this pathway, 24 equally distributed frames—including the crystallographic states—were subjected to conventional MD and GaMD simulations for a total of  $\sim 15$   $\mu$ s with the aim of rigorously characterizing the conformational dynamics of the complex over a long timescale, made possible through the application of enhanced sampling (SI Appendix) (15, 16, 19). Intermolecular FRET distances have been used to characterize the conformational states of Cas9 during the dynamics (Fig. 3A and SI Appendix, Table S1) (11). The distance between D435 and E945 is indicative of the conformational transition from the apo Cas9 up to the RNA-bound form, whereas the N1054–S867 and S867–S355 distances report on the conformational transition of the HNH domain from the RNA-bound state up to the catalytic state. Notably, the distributions of the FRET distances in the simulations of the apo Cas9, Cas9:RNA, Cas9:DNA, and Cas9:precat are centered on the crystallographic values and also show remarkable agreement with the catalytic state proposed by Sternberg et al. (11) (SI Appendix, Figs. S9 and S10). Potential of mean force (PMF) calculations, which describe the free energy landscape of Cas9, have been performed over the

aggregate trajectories, using the experimental FRET distances as reaction coordinates, enabling comparison with the experimental data (SI Appendix). Similar PMF profiles have been obtained by using different bin sizes and by varying trajectory lengths (SI Appendix), suggesting convergence of the presented GaMD simulations (25). In this regard, for large biomolecular complexes such as CRISPR-Cas9, it remains a major challenge to obtain convergent free-energy profiles from molecular simulations. This requires extensive sampling at the limits of the current computing power and the software's ability to effectively use it, as performed here. In light of this, the free-energy landscape sampled by the present simulations constitutes an approximation. Nevertheless, in agreement with the existing experimental data, our simulation findings provide a highly detailed picture of the conformational transitions of CRISPR-Cas9.

The PMF of Cas9 along the conformational transition from the apo form up to the RNA-bound state reveals three local minima (Fig. 3B), corresponding to the crystallographic apo (A) and RNA-bound (B) states, as well as an intermediate state (C, hereafter referred as R-exp state), in which the R-rich helix becomes solvent-exposed and available for RNA recruitment, consistent with its role in contacting the guide RNA in both the RNA- and DNA-bound X-ray structures (7–10). The RNA-bound form is found at the lowest minimum, being consistent with the stabilization of Cas9 by guide RNA binding. The D435–E945 distance shows a broad distribution in the RNA-bound form and in the closest frames (SI Appendix, Fig. S9), resulting in a minimum at  $\sim 70$  Å. The apo form, constituting the local minimum A, is prone to a conformational transition involving the opening of REC-I/III, favoring the formation of the R-exp state (C), in which the protein is primed for RNA recruitment. This evidence reveals the tendency of the highly flexible apo protein to



**Fig. 3.** (A) Schematic representation of the intermolecular FRET distances between the E945–D435, N1054–S867, and S867–S355 residue pairs used to identify the conformational states of Cas9 (shown on top of Cas9) (11). Distances are measured between the  $\alpha$  atoms. (B) Potential of mean force (PMF) of the E945–D435 distance, describing the conformational transition of Cas9 from the apo form up to the RNA-bound state. Experimental values of the E945–D435 distance are indicated using vertical bars (*SI Appendix, Table S1*). The PMF reveals three local minima, corresponding to the crystallographic apo (A) and RNA-bound (B) states, and to an intermediate state (C), characterized by the solvent exposure of the arginine residues in the R-rich helix (i.e., R-exp state, shown on top of the graph). (C) 2D PMF of the E945–D435 distance in combination with the rmsd with respect to the X-ray structure of the apo Cas9, identifying the path for the conformational transition. Along this path, the electrostatic potential of the Cas9 protein has been computed for selected states, using ensemble averaging over  $\sim 400$  ns GaMD simulations (shown around the PMF). The positive (blue lobes) and negative (red lobes) potential field strengths are captured at an isodensity of  $\pm 200$  kT/e. Cas9 is shown as cartoons, color-coded as in Fig. 1. This shows the formation of a positively charged cavity, upon the occurrence of R-exp state and preceding the formation of the RNA-bound complex (i.e., as pre-RNA bound).

adopt a conformational state prone for RNA binding (13), consistent with the experimental fact that Cas9 mainly exists in the RNA-bound form (3, 4). The R-exp (C) and the RNA-bound (B) states are separated by a free-energy barrier of  $\sim 2.5$  kcal/mol. Considering that RNA binding is supposed to occur along the activation trajectory, it is reasonable to speculate that RNA binding would result in a decrease in the activation energy, favoring the completion of the process. A 2D PMF, calculated using the D435–E945 distance in combination with the rmsd with respect to the X-ray structure of the apo Cas9, identifies the path for the conformational transition (Fig. 3C). Along this path, we have computed the electrostatic potential of the Cas9 protein for selected states, using ensemble averaging over  $\sim 400$ -ns GaMD simulations for each state (*SI Appendix*), with the purpose to understand the role of electrostatics, which is supposed to play a role in attracting and binding RNA. We found the formation of a positively charged cavity upon the occurrence of the R-exp state and preceding the formation of the RNA-bound complex (i.e., pre-RNA bound). This suggests that RNA recruitment might be initiated by the exposure of the R-rich region, while the subsequent formation of a positively charged cavity is poised to accommodate the nucleic acids.

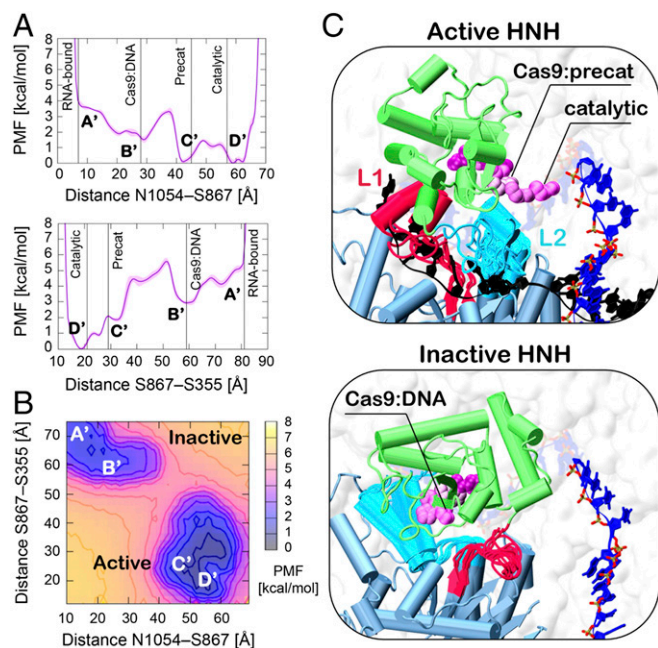
By inspection of the PMF of Cas9 along the conformational transition from the RNA-bound state up to the catalytic Cas9 (Fig. 4A), we observe that the identified minima correspond to the crystallographic states of Cas9:RNA (A'), Cas9:DNA (B'), and Cas9:precat (C') as well as the putative catalytic state (D'), showing a gradual stabilization due to the sequential binding of the RNA and DNA strands. It is noteworthy that both free-energy

profiles, calculated using the N1054–S867 and S867–S355 FRET distances, identify the putative catalytic Cas9 as a low-energy state. In this configuration, Cas9 is prone for the catalysis of both DNA strands. In the HNH active site, H840 stably locates at  $\sim 4$ – $6$  Å from the scissile phosphate in the *t*-DNA strand (Fig. 5 and *SI Appendix, Fig. S12*), allowing the entrance of water molecules, as in a “one-metal aided” architecture (6, 26). Concurrently, the *nt*-DNA strand approaches the RuvC active site with formation of a “two-metal aided” motif (6, 27), in which the binding of divalent cations facilitate the movement of the *nt*-DNA toward the active site (10, 28, 29). However, limitations of the currently available force fields for Mg ions (30) call for structural validation, which could unambiguously clarify the role of metal ions in the catalysis.

The PMFs indicate that the Cas9:DNA and Cas9:precat state constitute two well-separated low-energy states (separated by a free-energy barrier of  $\sim 3.5$  kcal/mol) that define the inactive and active state of the HNH domain, respectively. By inspection of the 2D free-energy profile of the N1054–S867 and S867–S355 distances (Fig. 4B), we observe that the HNH domain samples a significantly large conformational space in both states.

In the active state, the catalytic H840 points toward the scissile phosphate in the *t*-DNA, with the HNH domain assuming a variety of configurations (Fig. 4C) that differ in the position of H840 with respect to the cleavage site in the *t*-DNA. Interestingly, the structural transition of the HNH domain from the precatalytic Cas9 up to the catalytic state is easily accessible (being separated by an energy barrier of  $\sim 1$  kcal/mol; Fig. 4A), due to the high flexibility of the HNH domain. This flexibility is shown to rely on the conformational dynamics of the L1 and





**Fig. 4.** (A) Potential of mean force (PMF) of the N1054–S867 (Top) and S867–S355 (Bottom) FRET distances (defined in Fig. 3A), describing the HNH conformational change from the RNA-bound Cas9 up to the catalytic state. Experimental values of the FRET distances are indicated using vertical bars (11). (B) 2D PMF of the N1054–S867 and S867–S355 distances. The PMFs reveal four local minima, corresponding to the X-ray structures of Cas9:RNA (A'), Cas9:DNA (B'), Cas9:precat (C'), and the putative catalytic state of Cas9 (D') (11). The 2D PMF reveals two well-separated low-energy states, which characterize the inactive and active states of the HNH domain, respectively. (C) Active (Top) and inactive (Bottom) configurations of the HNH domain. Multiple configurations of the two states are superposed, indicating the high flexibility of the L1 (red) and L2 (cyan) loops, which interconnect the HNH (green) and RuvC (blue) domains, as well as the repositioning of the catalytic H840 (magenta) toward the scissile phosphate on the *t*-DNA (blue ribbons). The crystallographic configurations of H840 in the Cas9:precat (5F9R.pdb) (10), Cas9:DNA (4UN3.pdb) (8), as well as the catalytic Cas9 are indicated.

L2 linkers that connect the HNH and RuvC domains. Indeed, by sampling multiple configurations, the interconnecting regions reorient the HNH domain in the active state, driving its catalytic activation. Moreover, the active states of the HNH domain are characterized by the presence of the *nt*-DNA strand within the RuvC cleft, indicating that the activation of the HNH toward cleavage of the *t*-DNA domain requires the presence of the *nt*-DNA within the RuvC cleft (12, 13).

In the inactive states (both Cas9:RNA and Cas9:DNA), the HNH domain samples a distinct set of configurations, in which the catalytic H840 points in the opposite direction with respect to the *t*-DNA and which are separated from the active states (Fig. 4C). This suggests that the transition of the HNH domain from its inactive state (8) to its active state might involve a further critical event leading the process to completion, which is the binding of double-stranded DNA that has not yet been experimentally captured (7, 11, 13). This hypothesis is supported by the above-described TMD simulations of the HNH transition, which indicate that the HNH conformational change is unlikely upon complete DNA unwinding as well as in the absence of both DNA strands.

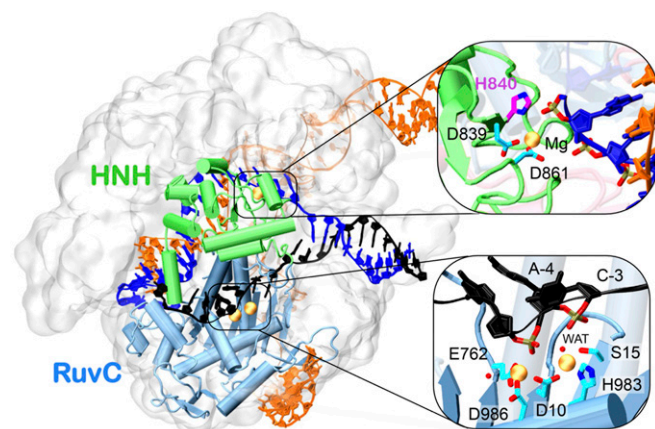
## Discussion

Here, we characterize the dynamics of the activation pathway of the Cas9 endonuclease through extensive enhanced MD simulations (~15  $\mu$ s in length), using a GaMD methodology (18), which allows unconstrained sampling and accurate reconstruction of the

free-energy profiles and probes long timescale events over hundreds of microseconds to milliseconds (19, 20).

We reveal the conformational transition of Cas9 from its apo form to the RNA-bound form, showing that the rearrangement of the REC1-III recognition regions leads to the opening of the REC lobe with respect to the nuclease lobe. This process occurs with solvent exposure of the R-rich helix, which contacts the guide RNA in the RNA- and DNA-bound X-ray structures (7–10) and with formation of a positively charged cavity poised to accommodate the RNA (Fig. 3). This suggests a possible pathway used by Cas9 for RNA recruitment, which is initiated by the exposure of the R-rich region, with subsequent formation of a positively charged cavity poised to accommodate the nucleic acids. Although these outcomes are consistent with the available structural data (7–10), single-molecule FRET experiments will be necessary to reveal the mechanism of RNA recruitment and confirm these mechanistic hypotheses.

GaMD also provides insightful information about the conformational activation of the HNH domain toward the catalysis, identifying a low-energy state that is likely to be productive for catalysis (Figs. 4 and 5) (6, 11). The structural transition of the HNH domain from the precatalytic Cas9 up to the catalytic state is easily accessible, indicating that the high flexibility of the HNH domain controls DNA cleavage, in agreement with previous biochemical evidences (11). This flexibility is shown to rely on the conformational dynamics of the L1 and L2 linkers that connect the HNH and RuvC (Fig. 4C). These linker loops reorient the HNH domain toward the cleavage site on the *t*-DNA, driving its catalytic activation. This result provides dynamic information on how the L1 and L2 linkers may act as “allosteric transducers” facilitating the communication between RuvC and HNH toward the activation for catalysis and thereby ensuring concerted cleavage of both DNA strands (10, 13). As further evidence of the crucial role of the interconnecting loops, we show—by complementing TMD simulations—that the L2 loop triggers the HNH conformational transition by H-bonding the RNA:*t*-DNA hybrid, clarifying also the experimental evidence that the HNH repositioning depends on the formation of the RNA:*t*-DNA hybrid (11). Interestingly, the active states of the HNH domain are characterized by the presence of the *nt*-DNA strand within the RuvC cleft. This finding agrees well with previous



**Fig. 5.** Catalytic state of CRISPR-Cas9, as from GaMD simulations. The protein is shown in molecular surface, highlighting the HNH (green) and RuvC (blue) domains in cartoon view. The RNA (orange), target DNA (*t*-DNA), and nontarget DNA (*nt*-DNA) strands are shown as ribbons. Catalytic Mg ions are shown as gold spheres. Close views of the catalytic sites of the HNH and RuvC domains are reported, showing the typical architectures for the occurrence of a “one-metal ion” (HNH) and a two-metal ion (RuvC) mechanism. Key residues involved in the catalysis are shown as sticks. Water molecules approaching Mg ions are shown as spheres.

computations, suggesting that the interaction between the *nt*-DNA and the L2 linker would favor the approach of the HNH domain toward the scissile phosphate on the *t*-DNA (13). Remarkably, these computations have been confirmed very recently by Dagdas et al. (12), who used single-molecule FRET experiments showing that the docking of HNH in its active configuration requires the presence of the *nt*-DNA. Taken together, these outcomes suggest that an interdependent activation of the RuvC and HNH domains toward concerted cleavage of the DNA strands might be mediated by the interplay between the L1/L2 linkers with the DNA.

GaMD also reveals that the inactive states of the HNH domain, in which the catalytic H840 points in the opposite direction with respect to the *t*-DNA, well-separate and distinguish from the active state (Fig. 4). This suggests that the transition of the HNH domain from its inactive to active state might involve a further critical event leading the process to completion, which is the binding of double-stranded substrate DNA. This hypothesis is supported by the TMD simulations of the HNH transition, which indicate that the HNH conformational change from its inactive to the active configuration is unlikely upon complete DNA unwinding as well as in the absence of both DNA strands. Altogether, these simulations suggest that the HNH repositioning might occur during DNA unwinding, leading to the speculation that the high flexibility of the HNH domain might facilitate somehow double-strand separation (i.e., R-loop formation) (7, 11, 13). In this respect, we recall that the binding of double-stranded DNA has not been experimentally captured yet. As such, our computations call now for novel experiments, which could assess our mechanistic hypotheses and eventually identify the role of the HNH repositioning during DNA binding.

Overall, by using enhanced molecular simulations, we disclose the conformational dynamics of CRISPR-Cas9, with understanding of each stage of the Cas9 activation process, from the apo protein via

the binary Cas9-guide RNA intermediate and the target DNA-bound complexes up to the final catalytic state. Our computational findings provide atomic-level information that will inspire future experimental investigations aimed at fully clarifying the puzzling biophysics of CRISPR-Cas9. This will contribute to rational engineering efforts aimed at designing Cas9 proteins with improved catalytic efficiency or controllable activity (21–23), which is central in the development of more efficient genome-editing tools based on CRISPR-Cas9.

## Method Summary

MD simulations have been based on the X-ray structures of the *Streptococcus pyogenes* apo Cas9 (4CMQ.pdb) (6), Cas9:RNA (4ZT0.pdb) (7), Cas9:DNA (4UN3.pdb) (8), and Cas9:precat (5F9R.pdb) (10). Conventional MD simulations were performed using a well-established protocol used for several other endonucleases (31), using the Amber ff12SB force field, which includes the ff99bsc0 corrections for DNA (32) and the ff99bsc0+ $\gamma$ OL3 corrections for RNA (33, 34). The Åqvist (35) force field has been used for Mg ions. GaMD (18) simulations were performed on the above-mentioned crystallographic structures as well as on a set of intermediate states among them (20 equally distributed frames). To produce initial coordinates of the intermediate states, which could not have been obtained experimentally, a TMD (17) approach has been used. GaMD simulations were performed using the GPU version of AMBER16 (36), whereas TMD simulations were performed with NAMD 2.10 (37). Details of molecular simulations, energetic reweighting, electrostatic calculations, system setup, and analysis are provided in *SI Appendix*.

**ACKNOWLEDGMENTS.** G.P. thanks the J.A.M. group, Prof. Rommie E. Amaro, and Dr. Kendra Hailey for useful discussions. The Swiss National Science Foundation supports this work with Advanced Postdoctoral Grant P300PA\_164698 (to G.P.). Work in the J.A.M. group is supported by NIH, National Science Foundation, Howard Hughes Medical Institute, National Biomedical Computation Resource, and San Diego Supercomputer Center. R.C.W. thanks Intel and NVIDIA for research fellowships. Computing time was provided by the Extreme Science and Engineering Discovery Environment with Grant MCB160059.

- Barrangou R, Doudna JA (2016) Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* 34:933–941.
- Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096.
- Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.
- Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507:62–67.
- Jiang F, Doudna JA (2017) CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys* 46:505–529.
- Jinek M, et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997.
- Jiang F, Zhou K, Ma L, Gressel S, Doudna JA (2015) Structural biology. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* 348:1477–1481.
- Anders C, Niewoehner O, Duerst A, Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513:569–573.
- Nishimasu H, et al. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156:935–949.
- Jiang F, et al. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 351:867–871.
- Sternberg SH, LaFrance B, Kaplan M, Doudna JA (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527:110–113.
- Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *bioRxiv*. doi: [10.1101/122242](https://doi.org/10.1101/122242).
- Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA (2016) Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations. *ACS Cent Sci* 2:756–763.
- Singh D, Sternberg SH, Fei J, Doudna JA, Ha T (2016) Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun* 7:12778.
- Nguyen C, et al. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* 505:427–431.
- Pierce LCT, Salomon-Ferrer R, Augusto F de Oliveira C, McCammon JA, Walker RC (2012) Routine access to millisecond time scale events with accelerated molecular dynamics. *J Chem Theory Comput* 8:2997–3002.
- Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 9:646–652.
- Miao Y, Feher VA, McCammon JA (2015) Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *J Chem Theory Comput* 11:3584–3595.
- Miao Y, McCammon JA (2016) Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor. *Proc Natl Acad Sci USA* 113:12162–12167.
- Miao Y, McCammon JA (2016) Unconstrained enhanced sampling for free energy calculations of biomolecules: A review. *Mol Simul* 42:1046–1055.
- Wright AV, et al. (2015) Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci USA* 112:2984–2989.
- Slaymaker IM, et al. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351:84–88.
- Kleinstiver BP, et al. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523:481–485.
- Zheng W (2017) Probing the structural dynamics of the CRISPR-Cas9 RNA-guided DNA-cleavage system by coarse-grained modeling. *Proteins* 85:342–353.
- Trzesniak D, Kunz AP, van Gunsteren WF (2007) A comparison of methods to compute the potential of mean force. *ChemPhysChem* 8:162–169.
- Biertumpfel C, Yang W, Suck D (2007) Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* 449:616–620.
- Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci USA* 90:6498–6502.
- Zuo Z, Liu J (2016) Cas9-catalyzed DNA cleavage generates staggered ends: Evidence from molecular dynamics simulations. *Sci Rep* 5:37584.
- De Vivo M, Dal Peraro M, Klein ML (2008) Phosphodiester cleavage in ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J Am Chem Soc* 130:10955–10962.
- Casalino L, Palermo G, Abdurakhmonova N, Rothlisberger U, Magistrato A (2017) Development of site-specific Mg(2+)-RNA force field parameters: A dream or reality? Guidelines from combined molecular dynamics and quantum mechanics simulations. *J Chem Theory Comput* 13:340–352.
- Palermo G, et al. (2015) Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc Chem Res* 48:220–228.
- Pérez A, et al. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J* 92:3817–3829.
- Banas P, et al. (2010) Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J Chem Theory Comput* 6:3836–3849.
- Zgarbová M, et al. (2011) Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* 7:2886–2902.
- Åqvist J (1990) Ion-water interaction potentials derived from free energy perturbation simulations. *J Phys Chem* 94:8021–8024.
- Case DA, et al. (2016) *AMBER 2016* (University of California, San Francisco, San Francisco, CA).
- Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802.