JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Cite This: J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Article

pubs.acs.org/jcim

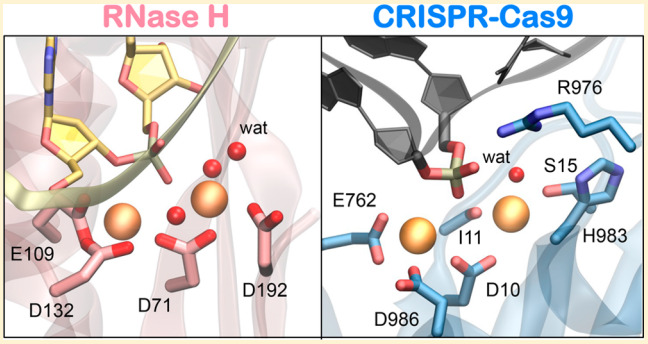
Structure and Dynamics of the CRISPR−Cas9 Catalytic Complex

Giulia Palermo*

Department of Bioengineering, Bourns College of Engineering, University of California Riverside, 900 University Avenue, Riverside, California 92521, United States

Supporting Information

ABSTRACT: CRISPR−Cas9 is a bacterial immune system with exciting applications for genome editing. In spite of extensive experimental characterization, the active site chemistry of the RuvC domain—which performs DNA cleavages—has remained elusive. Its knowledge is key for structure-based engineering aimed at improving DNA cleavages. Here, we deliver an in-depth characterization by using quantum-classical (QM/MM) molecular dynamics (MD) simulations and a Gaussian accelerated MD method, coupled with bioinformatics analysis. We disclose a *two-metal aided* architecture in the RuvC active site, which is poised to operate DNA cleavages, in analogy with other DNA/RNA processing enzymes. The conformational dynamics of the RuvC domain further reveals that an "arginine finger" stably contacts the scissile phosphate, with the function of stabilizing the active complex. Remarkably, the formation of a catalytically competent state of the RuvC domain is only observed upon the conformational activation of the other nuclease domain of CRISPR−Cas9—i.e., the HNH domain—such allowing concerted cleavages of double stranded DNA. This structure is in agreement with the available experimental data and remarkably differs from previous models based on classical mechanics, demonstrating also that only quantum mechanical simulations can accurately describe the metal-aided active site in CRISPR−Cas9. This fully catalytic structure—in which both the HNH and RuvC domains are prone to perform DNA cleavages—constitutes a stepping-stone for understanding DNA cleavage and specificity. It calls for novel experimental verifications and offers the structural foundations for engineering efforts aimed at improving the genome editing capability of CRISPR−Cas9.



INTRODUCTION

CRISPR (clustered regularly interspaced short palindromic repeats) and Cas9 make up a bacterial immune system with exciting applications in the field of genome editing.[1,2] In this system, the endonuclease Cas9 is programmed with single guide RNAs (sgRNA) to site-specifically recognize and cleave any DNA sequence of interest given the presence of a short recognition sequence (protospacer adjacent motif, PAM).[1] The easy programmability of CRISPR−Cas9 led to a frontier genome-editing technology enabling gene manipulation in a variety of biological systems and organisms, encompassing bacteria and eukaryotes, with *in vivo* applicability for the treatment of severe diseases, including cancer, genetic diseases, and brain disorders.[3]

Cas9 is an RNA-guided DNA endonuclease, which generates double-stranded breaks in DNA thanks to the HNH and RuvC nuclease domains (Figure 1A).[4−12] The double stranded DNA binds Cas9 by matching the guide RNA with one strand (the target strand, TS), while the other nontarget strand (NTS) is displaced and also accommodated in the protein framework. In the activated complex of CRISPR−Cas9 (Figure 1B), the HNH and RuvC domains perform concerted cleavages of the TS and NTS, respectively.[13] The most complete X-ray structure of the ternary complex captured an inactivated state of the system, with the HNH catalytic site located ∼16−18 Å from the cleavage site on the TS (Figure 1A).[5] Clearly, the activation of the system toward the catalysis requires a conformational change of the HNH domain. Biochemical experiments and structural comparison with the homologous structure of the T4 endonuclease VII,[14] enabled Sternberg and colleagues to propose a first model for the catalytically active HNH docking at the cleavage site on the TS.[13] Based on this information, our previous computational studies applied extensive molecular dynamics (MD) simulations in conventional[15] and enhanced sampling regime,[16] employing also a specialized supercomputer enabling micro-to-millisecond length simulations (Anton-2),[17,18] with the goal of characterizing the activated structure of CRISPR−Cas9. As a result, ∼16 μs of continuous MD simulations and >20 μs of MD simulations in enhanced sampling regime identified the mechanism of conformational activation of the HNH domain, revealing its approach to the TS for cleavage. Förster resonance energy transfer (FRET) experiments and biochemical

Special Issue: Women in Computational Chemistry

Received: December 31, 2018
Published: February 14, 2019

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

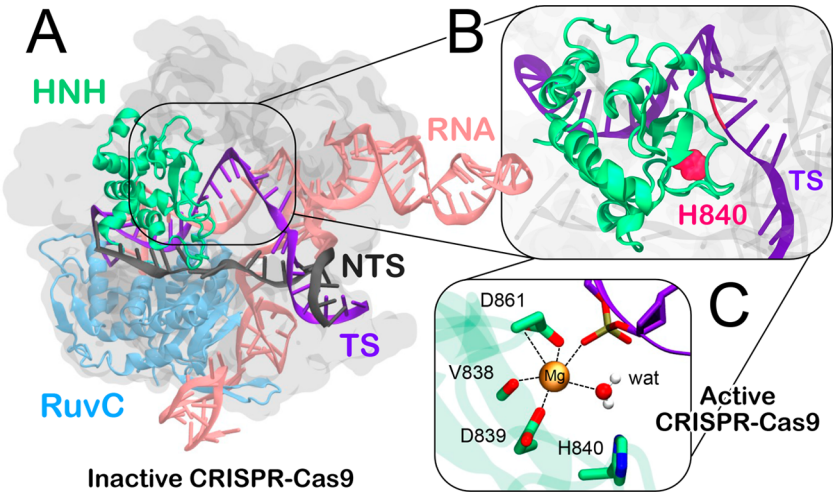


Figure 1. (A) X-ray structure of the inactive CRISPR−Cas9 complex (5F9R.pdb).[5] The Cas9 protein is shown in molecular surface, with the HNH (green) and RuvC (blue) domains shown as cartoons. The RNA (pink), as well as the target strand (TS, violet) and the nontarget strand (NTS, black) of the DNA are shown as ribbons. The structure captures an inactive state of the CRISPR−Cas9 complex, with the catalytic HNH domain located at ∼16−18 Å from the cleavage site on the TS. (B) Close-up view of the activated CRISPR−Cas9 complex derived from molecular simulations[16,17] and FRET experiments,[19−21] displaying the docking of the HNH domain at the cleavage site on the TS, and moving away from the RuvC domain. (C) Site of the activated HNH domain, displaying the coordination of the metal Mg^{2+} ions (gold spheres) in the activated complex.

data[19−21] have cross validated the conformational changes observed through MD simulations, obtaining a consensus structure for the activated CRISPR−Cas9 complex (Figure 1B). This configuration identifies the metal (Mg^{2+}) aided coordination geometry of the HNH active site (Figure 1C) and discloses how the HNH domain docks at the cleavage site on the TS, moving away from the RuvC domain that hosts the NTS.

Although the above-mentioned studies have captured the overall architecture of CRISPR−Cas9 in the activated form, the configuration of the RuvC active site in its catalytically competent state (i.e., prone to perform catalysis) has not been characterized. Indeed, previous models have been based on the inactive complex, which does not account for the activated configuration of the HNH domain.[16,22] This is a critical point, considering that Cas9 cleaves the TS and the NTS concertedly and—only upon activation of HNH—RuvC is supposed to give rise to the formation of a catalytically competent state.[13] Moreover, Cas9 performs the NTS cleavage through a *two-metal (Mg^{2+}) aided* catalysis,[1,23] but the available structures of the CRISPR−Cas9 ternary complex have been captured in the absence of the catalytic Mg^{2+} ions. As such, little is known about the catalytic state of the RuvC nuclease domain in CRISPR−Cas9, and identifying how the divalent metal ions position within the active site for DNA cleavage remains elusive. Knowledge of the metal-aided active site geometry is key for structure-based engineering aimed at improving the catalytic efficiency of the enzyme.

Here, we characterize the structure and dynamics of the RuvC active site upon activation of the HNH domain, through extensive MD simulations. An in-depth characterization has been attained through the use of *ab-intio* quantum mechanics/molecular mechanics (QM/MM) simulations and classical MD in enhanced sampling regime. *Ab-intio* QM/MM simulations have been employed to provide an accurate description of the metal-aided active site, as for other RNA/DNA processing enzymes.[24−26] To broadly explore the conformational dynamics of the active site, we performed classical MD simulations, by employing a Gaussian accelerated MD (GaMD) method.[27] This approach has been shown to allow unconstrained enhanced sampling, providing also solid estimation of the energetics associated with the conformational changes.[28−34] Finally, structure-based comparison and bioinformatics analysis have been used to reveal a consensus with other metal-aided enzymes, further supporting the simulation outcomes.

Overall, we reveal the *two-metal aided* architecture in the RuvC active site, which is poised to operate the catalysis in synchrony with the activated HNH domain. The activated state is in agreement with the available experimental data on CRISPR−Cas9 and remarkably differs from previous models based on the empirical classical force field.[22] In this respect, the activated state identified here displays key analogies with the other two-metal aided enzymes,[23,35] indicating the essential need of *ab-initio* QM/MM MD for an accurate description of the metal-aided active sites in CRISPR−Cas9. Based on our results, the residues coordinating the metal ions in the HNH and RuvC active sites could be essential for the catalysis and, more broadly, for the selectivity of the system against off-target sequences. These hypotheses call now for further experimental validation, adding to the existing knowledge. As well, this fully activated structure—in which both the HNH and RuvC domains are prone to perform DNA cleavages—offers the basis for novel structure-based engineering of both active sites aimed at improving the genome editing capability of CRISPR−Cas9.

METHODS

Structural Models. MD simulations have been based on the most complete X-ray structure of Cas9 in complex with RNA and DNA (5F9R.pdb),[5] which constitutes an inactivated state of the system, and on the activated CRISPR−Cas9 complex obtained combining previous MD simulations and single molecule FRET experiments.[13,16] The catalytic H983 has been protonated in the epsilon position, as from previous studies.[4,16,22] Two model systems have been obtained: (i) activated and (ii) inactivated CRISPR−Cas9, which have been embedded in explicit waters, leading to periodic simulation

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

cells of ~180 Å × 120 Å × 140 Å and a total of ~300 000 atoms for each system.

Classical MD Simulations. The model systems have been equilibrated through conventional MD, using the Amber ff12SB force field, which includes the ff99bsc0 corrections for DNA[36] and the ff99bsc0+χ*OL3* corrections for RNA.[37,38] The Åqvist[39] model has been employed for Mg^{2+} ions, as in previous studies of two-metal aided enzymes.[24,25,40] An integration time step of 2 fs has been used. Hydrogen atoms were added assuming standard bond lengths and were constrained to their equilibrium position with the SHAKE algorithm. Temperature control (300 K) has been performed via Langevin dynamics,[41] with a collision frequency γ = 1. Pressure control was accomplished by coupling the system to a Berendsen barostat,[42] at a reference pressure of 1 atm and with a relaxation time of 2 ps. The simulations were carried out with the following protocol. The systems were subjected to energy minimization to relax the water molecules and the counterions, keeping the protein, as well as the RNA, DNA, and Mg^{2+} ions fixed with harmonic position restraints of 300 kcal/mol·Å^2. Then, the systems were heated from 0 to 100 K in the canonical ensemble (NVT), by running two NVT simulations of 5 ps each, imposing position restraints of 100 kcal/mol·Å^2 on the above-mentioned elements of the considered systems. The temperature was further increased up to 200 K in ~100 ps of MD in the isothermal−isobaric ensemble (NPT), in which the restraint was reduced to 25 kcal/mol·Å^2. Subsequently, all restraints were released and the temperature of the system was raised up to 300 K in a single NPT simulation of 500 ps. After ~1.1 ns of equilibration, ~10 ns of NPT production was carried out allowing the density of the system to stabilize around 1.01 g/cm^3. Finally, ~100 ns of MD simulations have been collected in the NVT ensemble. All simulations have been performed with the GPU version of AMBER 16.[43] The well-equilibrated systems have been used for QM/MM and GaMD simulations.

***Ab-Initio* QM/MM Simulations.** QM/MM simulations have been performed on the activated CRISPR−Cas9 complex. The QM part was composed of the RuvC active site including the Mg^{2+} ions and the surrounding protein residues that coordinate the metal ions (i.e., E762, D986, D10, S15), as well as the H983 residue, part of the DNA nucleobases C-3 and A-4, and six water molecules (Figure S1). A total of 105 atoms have been described at the QM DFT/BLYP[44,45] level, while the remaining MM part was treated using the force field described above. The valence of the terminal QM atoms was saturated by the addition of 13 capping hydrogen atoms. QM/MM simulations have been performed using the CPMD code.[46] The wave functions were expanded in a plane wave basis set up to a cutoff of 75 Ry in a QM cell of dimensions ~26 Å × 22 Å × 26 Å. The interactions between the valence electrons and ionic cores were described with norm-conserving Martins−Troullier pseudopotentials.[47] Dispersion interactions have been described by correcting the exchange-correlation functional through the use of properly parametrized dispersion corrected atom-centered potentials (DCACPs).[48] The QM part was treated as an isolated system, and electrostatic interactions between periodic images were decoupled by the scheme of Tuckerman.[49] Notably, a rigorous Hamiltonian treatment of the electrostatic interaction between the QM and MM regions was used.[50] The QM/MM protocol consisted of an initial optimization of the wave function, followed by ~6 ps of careful equilibration carried out with Born−Oppenheimer (BO) MD in the canonical (NVT) ensemble using an integration time step of 20 au (~0.48 fs). The temperatures of the QM and MM subsystems were kept constant at 300 K using a Nosé−Hoover thermostat.[51,52] After this initial phase, Car−Parrinello (CP)[53] QM/MM simulations were carried out with a time step of 5 au (~0.12 fs) and a fictitious electron mass of 600 au. A total of ~40 ps of *ab-initio* QM/MM MD were carried out. In this respect, we note that although *ab-initio* MD is limited to the picosecond time scale, changes in the coordination geometry of metal ions can be observed over less than 10 ps.[54,55] Considering also that *ab-initio* MD remarkably increases in computational cost when dealing with a high number of QM atoms (as in the present case, including 105 QM atoms and 13 capping hydrogen atoms), the ~40 ps simulation time performed here provides a reliable estimation of the active site chemistry.

Gaussian Accelerated Molecular Dynamics (GaMD) Simulations. Accelerated MD (aMD) is an enhanced sampling method that adds a boost potential to the potential energy surface (PES), effectively decreasing the energy barriers and thus accelerating transitions between low-energy states.[56] We applied a novel and robust implementation of aMD—i.e., a Gaussian aMD (GaMD)—which allows simultaneous sampling and free energy reweighting, using harmonic functions to construct a boost potential that is adaptively added to the PES.[27] The boost potential follows a near-Gaussian distribution, which allows for accurate reweighting using cumulant expansions up to the second order. This extends the use of aMD to big-data analysis, in which a large statistical noise can make prohibitive the standard reweighting procedure.[30,57]

Considering a system with N atoms at positions $\vec{r} = \{\vec{r}_1, ..., \vec{r}_N\}$, when the system potential $V(\vec{r})$ is lower than a threshold energy E, the energy surface is modified by adding a boost potential as

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}), \quad V(\vec{r}) < E \tag{1}$$

$$\Delta V(\vec{r}) = \frac{1}{2}k(E - V(\vec{r}))^2 \tag{2}$$

where k is the harmonic force constant. The two adjustable parameters E and k are automatically determined by applying the following three criteria. First, for any two arbitrary potential values $V_1(\vec{r})$ and $V_2(\vec{r})$ found on the original energy surface, if $V_1(\vec{r}) < V_2(\vec{r})$, ΔV should be a monotonic function that does not change the relative order of the biased potential values, i.e., $V_1^*(\vec{r}) < V_2^*(\vec{r})$. Second, if $V_1(\vec{r}) < V_2(\vec{r})$, the potential difference observed on the smoothened energy surface should be smaller than that of the original, i.e., $V_2^*(\vec{r}) - V_1^*(\vec{r}) < V_2(\vec{r}) - V_1(\vec{r})$. By combining the first two criteria and plugging in the formula of $V^*(\vec{r})$ and ΔV, we obtain

$$V_{max} \le E \le V_{min} + 1/k \tag{3}$$

where V_{min} and V_{max} are the system minimum and maximum potential energies. To ensure that eq 4 is valid, k has to satisfy $k \le {}^1/_{V_{max}} - V_{min}$. By defining $k \equiv k_0 {}^1/_{V_{max}} - V_{min}$, then $0 < k \le 1$. Third, the standard deviation of ΔV needs to be small enough (i.e., narrow distribution) to ensure accurate reweighting using cumulant expansion to the second order: $\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \le \sigma_0$, where V_{avg} and σ_V are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV and σ_0 as a user-specified upper limit (e.g., 10 k_BT) for accurate reweighting. When E is set to

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

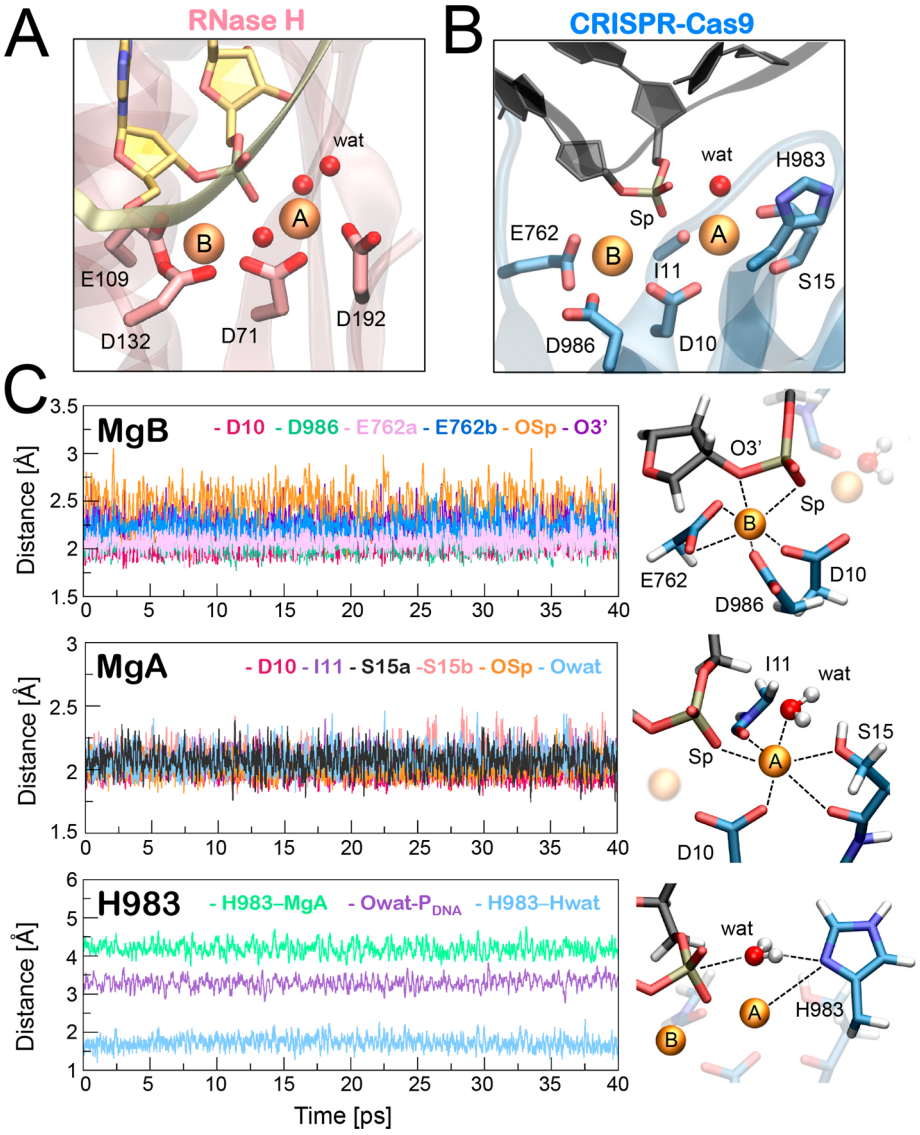


Figure 2. (A) Active site of the RNA Ribonuclease H (RNase H)[64] displaying a two-metal aided architecture, characterized by the highly conserved DEDD (or DDE) motif. The two Mg^{2+} ions (A and B, gold) and the water molecules (red) are shown as spheres. (B) RuvC active site of CRISPR−Cas9 as arising from *ab-initio* QM/MM MD simulations. The active site adopts a catalytically competent two-metal aided configuration. One water molecule locates in close proximity to the scissile phosphate and interacts with H983. The oxygen *pro*-S (Sp) jointly coordinates the two metals, as required for the two-metal aided catalysis.[23,25,26,65] (C) Time evolution along ∼40 ps of *ab-initio* QM/MM MD of the coordination distances for the ligands of MgB (top) and MgA (center), as well as of the interaction network established by H983 and by the water nucleophile (bottom). For each plot, the distances are shown on the 3D structure in the right panel.

the lower bound, $E = V_{min}$, according to eq 4, k_0 can be calculated as

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V}\frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right) \quad (4)$$

Alternatively, when the threshold energy E is set to its upper bound $E = V_{min} + 1/k$, k_0 is

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right)\frac{V_{max} - V_{min}}{V_{avg} - V_{min}} \quad (5)$$

if k_0'' is calculated between 0 and 1. Otherwise, k_0 is calculated using eq 4, instead of being set to 1 directly as described in the original paper.[27]

Based on extensive tests on the CRISPR−Cas9 system performed in our previous work,[16,58] the system threshold energy has been set to $E = V_{max}$ for all GaMD simulations. The boost potential has been applied in a *dual-boost* scheme, in which two acceleration potentials are applied simultaneously to the system: (i) the torsional terms only and (ii) across the entire potential. A time step of 2 fs has been used. The maximum, minimum, average, and standard deviation values of the system potential (V_{max}, V_{min}, V_{avg}, and σ_V) has been obtained from an initial ∼12 ns NPT simulation with no boost potential. GaMD simulations have been applied on the activated CRISPR−Cas9 complex, in four simulation replicas. Two additional GaMD simulations have been performed on the inactivated complex. Each GaMD simulation proceeded with a ∼ 50 ns run, in which the boost potential has been updated every 1.6 ns, thus reaching equilibrium values. Finally, ∼400 ns of GaMD simulations have been carried out in the NVT *ensemble*, for a total of ∼2.4 μs of GaMD. The choice of

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

this simulations length (i.e., ~400 ns per replica) has shown to exhaustively explore the conformational space and to enable convergent derivation of the free energy profiles in our previous study.[16]

Potential of Mean Force (PMF) Calculations. The PMF is calculated upon accurate reweighting of the GaMD simulations using cumulant expansion to the second order. For simulations of a biomolecular system, the probability distribution along a selected reaction coordinate $A(r)$ is written as $p^*A(r)$, where r denotes the atomic positions $\{r_1,\cdots,r_n\}$. Given the boost potential $\Delta V(r)$ of each frame, $p^*A(r)$ can be reweighted to recover the canonical ensemble distribution, $p(A)$, as

$$p(A_j) = p^*(A_j)\frac{\langle e^{\beta\Delta V(r)}\rangle_j}{\sum_{j=1}^{M}\langle e^{\beta\Delta V(r)}\rangle_j} \quad j = 1, ..., M \tag{6}$$

where M is the number of bins, $\beta = 1/k_BT$, and $\langle e^{\beta\Delta V(r)}\rangle_j$ is the ensemble-averaged Boltzmann factor of $\Delta V(r)$ for simulation frames found in the jth bin. To reduce the energetic noise, the ensemble-averaged reweighting factor can be approximated using cumulant expansion:

$$\langle e^{\beta\Delta V}\rangle = \exp\left\{\sum_{k=1}^{\infty}\frac{\beta^k}{k!}C_k\right\} \tag{7}$$

where the first three cumulants are given by

$$\begin{aligned} C_1 &= \langle\Delta V\rangle, \\ C_2 &= \langle\Delta V^2\rangle - \langle\Delta V^2\rangle = \sigma_{\Delta V}^2 \\ C_3 &= \langle\Delta V^3\rangle - 3\langle\Delta V^2\rangle\langle\Delta V\rangle + 2\langle\Delta V^3\rangle \end{aligned} \tag{8}$$

As shown earlier,[27,30,57] when the boost potential follows near-Gaussian distribution, cumulant expansion to the second order provides the more accurate reweighting compared with the exponential average and Maclaurin series expansion methods. Finally, the reweighted free energy is calculated as

$$F(A_j) = -1\left(\frac{1}{\beta}\right)\ln p(A_i) \tag{9}$$

Energetic reweighting has been computed using as reaction coordinates the distance between the terminal carbon of R976 and the scissile phosphate (R976−P_{DNA}) and the dihedral angle between the Cα−Cβ−Cγ−Cδ atoms of R976 (ϑ-R976). Since accurate energetic reweighting needs to be performed on a solid statistical ensemble, PMF calculations have been performed over the aggregate trajectories of four simulation replicas (i.e., ~400 ns × 4 replicas = ~1.6 μs). Moreover, the trajectories have been saved every 1 ps (i.e., 50 steps), leading to 400 000 frames per replica and a total of 1 600 000 frames used for energetic reweighting. Energetic reweighting has also been performed over the individual runs (Figure S2), showing the similarity of the PMF plots computed over independent replicas. Error bars have been computed as the standard error of the PMF values calculated for the individual runs, as in our previous study.[16] Convergence of the PMF has been assessed by computing the location of the minimum as a function of time in the individual runs (Figure S3).

Bioinformatics Analysis. Sequence alignment has been performed considering the primary sequence of the RuvC domain of the *Streptococcus pyogenes* (*Sp*),[5] *Staphylococcus aureus* (*Sa*),[59] *Francisella Novicida* (*Fn*),[60] *Actinomyces naeslundii* (*Ana*),[4] and *Campylobacter jeuni* (*Cj*)[61] CRISPR−Cas9 systems. The unified bioinformatics analysis tool Multiseq,[62] implemented in the Visual Molecular Dynamics[63] code has been used to perform sequence alignment. Structural superimposition has been performed using the structures deposited in the Protein Data Bank under the 5F9R,[5] 5AXW,[59] 5B2O,[60] 4OGC,[4] and 5X2G[61] accession codes.

RESULTS

Configuration of the *Two-Metal Aided* Active Site. Biochemical data and structural comparison with homologous nucleases, suggested that a two-metal aided mechanism[23] is responsible of the catalysis within the RuvC active site. Indeed, the RuvC domain shares the structural fold of the RNA Ribonuclease H (RNase H),[64] which is the prototypical two-metal aided enzyme (Figure 2A). Within its active site, two catalytic Mg^{2+} ions are coordinated by a highly conserved sequence composed by the DEDD residues (or DDE), which are a fingerprint for the two-metal aided architecture in DNA and RNA processing enzymes.[35] The catalysis requires that the *pro*-Sp oxygen of the scissile phosphate—but not the *pro*-Rp oxygen—jointly coordinates the two Mg^{2+} ions, enabling an in-line S_N2-like nucleophilic attack on the scissile phosphorus atom.[23,25,26,65] In order to assess the structure and dynamics of the RuvC active site, as well as the coordination sphere of the metal ions, we performed *ab-initio* QM/MM MD simulations. Through this approach, we characterized the RuvC active site geometry at the QM BLYP level, while the rest of the CRISPR−Cas9 complex in explicit solution has been treated at the classical MM level (the QM/MM partitioning is shown in Figure S1). This enabled us to more accurately characterize the active site chemistry and—importantly—the metal ion coordination geometry, which is often not accurately described by the empirical force fields.[66] We carried out ~40 ps of QM/MM *ab-initio* MD simulations, using the Car−Parrinello (CP) method.[53] As a result of the CP MD simulations, the RuvC active site shows a catalytically competent two-metal aided architecture (Figure 2B). In detail, MgB is coordinated by the carboxylate functionality of E762, D986, and D10 and by the DNA O3′. The *pro*-Sp oxygen of the scissile phosphate also coordinates MgB. This acidic structural motif is consistent with the requirements for the catalysis[23,25,26,65] and with the highly conserved DDE motif, which ensures the metal-aided function.[35] MgA results coordinated by D10, I11, and S13, as well as by the *pro*-S oxygen of the scissile phosphate, as in other two-metal aided enzymes.[35] The *pro*-S oxygen of the scissile phosphate locates in between the two metals and favoring an in-line S_N2-like nucleophilic attack. One QM water molecule coordinates MgA and locates in close proximity to the scissile phosphate, as a possible nucleophile. Figure 2C reports the time evolution along *ab initio* MD of the coordination ligands of MgB (top graph) and MgA (middle graph), as well as of the interaction network established by H983 and the water nucleophile (bottom graph, data are in Table S1). Remarkably, H983 locates away from MgA and binds the QM water molecule approaching the scissile phosphate. This configuration well agrees with the requirements for the catalysis.[21,35,73,74] Indeed, H893 has been experimentally suggested to act as a general base deprotonating the water nucleophile,[11] as observed in other nucleases.[35] In this respect, the *two-metal ion* mechanism supports His as activators of the catalysis, as arising from several experimental

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

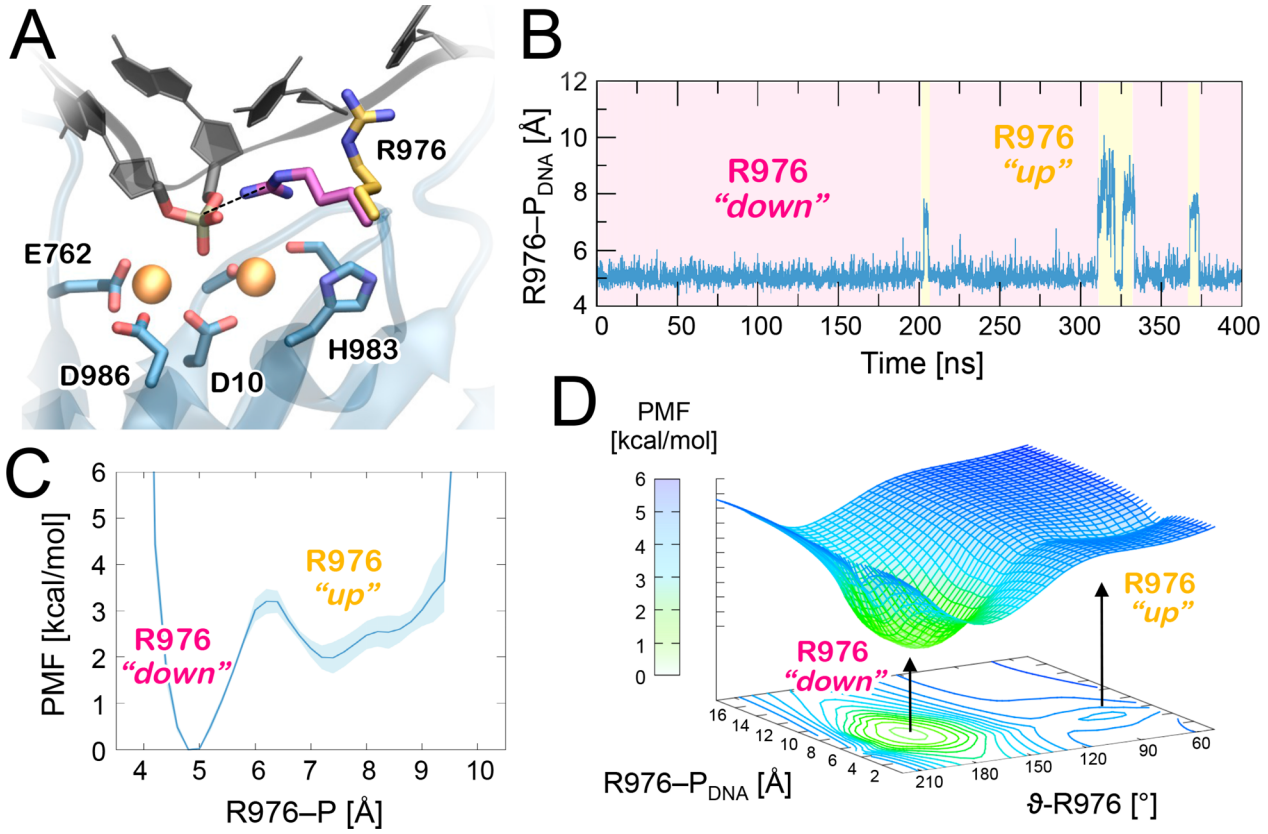


Figure 3. (A) Catalytic site of the RuvC domain in the CRISPR−Cas9 complex, displaying two conformations of R976. In the activated CRISPR−Cas9 complex, R976 assumes a "down" conformational state (magenta), bringing its positive side chain in close contact to the scissile phosphate. In the inactive state of the CRISPR−Cas9 complex, R976 moves away from the scissile phosphate and assumes an "up" conformation (yellow). (B) Time evolution of distance between the terminal carbon of R976 and the scissile phosphate (R976−P_{DNA}) along ∼400 ns of GaMD simulations of the activated CRISPR−Cas9 complex (replicas are in Figure S4). Time windows indicate the "down" (magenta) and "up" (yellow) conformations of R976. (C) PMF of the R976−P_{DNA} distance computed over the aggregated data from four simulation replicas for ∼3.2 μs of total sampling. The PMF is expressed in kilocalories per mole. (D) Two-dimensional PMF of the R976−P_{DNA} distance in combination with the dihedral angle between the Cα−Cβ−Cγ−Cδ atoms of R976 (ϑ-R976).

studies[35] and also previously shown in QM/MM investigations of the DNA polymerase III.[67] Further support to the active site chemistry identified here arises from the crystal structure of the RuvC resolvase enzyme,[68] which displays a remarkable identity to the Cas9 RuvC domain. In the RuvC resolvase, H143 corresponds to the catalytic H983 in Cas9, while the Mg^{2+} coordinating ligands D7, E70, and D146 correspond to the Cas9 residues D10, E762, and D986, respectively.

Dynamics of the Activated State. Here, accelerated MD simulations have been employed to investigate the dynamics of the RuvC active site in the activated CRISPR−Cas9 complex. Four simulation replica of ∼400 ns each have been carried out in enhanced sampling regime by using a GaMD method. This enabled to enhance the configurational sampling, broadly exploring the conformational dynamics of the RuvC domain in the activated state, providing also the estimation of the energetics associated with the conformational changes.[28−34]

As a result, we found that in the activated CRISPR−Cas9 complex, the RuvC catalytic site shows a remarkable difference with respect to the X-ray structure of the inactive state (Figure 3A). Indeed, in the activated CRISPR−Cas9, the R976 residue approaches the cleavage site bringing its positive side chain in close contact to the scissile phosphate (magenta in Figure 3A). In this configuration, the R976 side chain points "down" toward the active site, as opposite to the crystallographic structure of the inactivated CRISPR−Cas9 complex, where R976 moves away from the scissile phosphate, assuming an "up" conformational state (yellow in Figure 3A). During GaMD simulations of the activated CRISPR−Cas9 complex, the "down" conformational state of R976 is well maintained (Figure 3B). The R986 side chain also adopts the "up" conformational state, which results however poorly populated, as also observed in the simulation replicas (Figure S4). PMF calculations—which are commonly used to describe the free energy landscape of conformational transitions—have been performed to estimate the free-energy landscape associated with the R976 rearrangements in the activated CRISPR−Cas9 complex. The PMF computed over the distance between the terminal carbon of R976 and the scissile phosphate (R976−P_{DNA}, Figure 2C) reveals that the R976 "down" state constitutes the lowest minimum of the free energy surface. The PMF has been computed over the aggregated data from four simulation replicas, as well us over the individual runs (Figure S2), revealing the similarity of the PMF computed over independent replicas. Error bars have been computed as the standard error of the PMF values calculated for the individual runs, revealing that the uncertainty is low at the level of the lowest minimum (i.e., corresponding to the "down" conformation), therefore sustaining the reliability of the identified minimum. As well, convergence tests have been performed by computing the PMF as a function of the simulated time in the individual replicas (Figure S3).[16] A two-

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

dimensional PMF of the R976−P_{DNA} distance in combination with the dihedral angle between the Cα−Cβ−Cγ−Cδ atoms of R976 (ϑ-R976) has also been computed, confirming that R976 preferentially locates in a well-defined minimum corresponding to the "down" conformational state, pointing toward the scissile phosphate (Figure 2D, the PMF on ϑ-R976 computed for the individual runs and convergence tests are reported in Figures S2−3). Overall, the simulations depict a consensus scenario for the activated complex, where R976 stably points its positive side chain in proximity of the scissile phosphate.

GaMD simulations have also been performed on the inactivated CRISPR−Cas9 complex, in two replicas of ∼400 ns each, enabling a comparison with the activated state and assessing the conformational dynamics of R976 in the crystallographic state.[5] The simulations reveal that R976 does not stably locate in proximity of the scissile phosphate. On the contrary, R976 approaches and establishes interactions with the backbone oxygen atoms of the Q910, L911, and K913 residues of the HNH domain (Figure 4A and S5). Figure 4B illustrates the probability distributions of the distances between the R976 terminal nitrogen atom (NH1) and the backbone oxygen atoms of the Q910, L911, and K913 residues, computed from the GaMD simulations of the inactive (yellow) and active (green) states of CRISPR−Cas9. As a result, in the inactive state, the R976 side chain interacts with the Q910 backbone, while also establishing interactions with L911 and K913. These interactions contribute in stabilizing R976 in the up conformational state in the inactive configuration of the CRISPR−Cas9 complex. Contrariwise, in the activated CRISPR−Cas9, R976 moves away from the Q910, L911, and K913 residues of the HNH domain, with the distances probability distributions remarkably separated from the inactivated state. These outcomes suggest that the interaction of R976 with the 908−913 residues of the HNH domain sequestrate R976 from the RuvC active site in the inactive state. Contrariwise, upon activation, when the HNH domain approaches the TS and moves away from the RuvC active site, R976 becomes free to contact the scissile phosphate. This evidence is in agreement with the experimental fact that the activated configuration of the HNH functionality is an essential requirement for the formation of a catalytically competent RuvC active site.[13] Accordingly, the tight interdependence in the conformational dynamics of HNH and RuvC is at the basis of the domains activation for concerted cleavages of the two DNA strands.[69]

Key Role of the Arginine Finger in the Two-Metal Aided Enzymes. The approach of R976 to the scissile phosphate of the active Cas9 is of particular interest in light of the function that the so-called "arginine fingers" exert in the active site of other Mg-aided phosphatases.[70] Indeed, by contacting the scissile phosphate, the arginine fingers confer proper structural stabilization of the active site for efficient catalysis. Arginine residues have also been shown to be key for the function of DNA/RNA processing enzymes that share with Cas9 a two-metal aided architecture. In type II topoisomerase (topo II), which controls the topology of DNA by cleaving and religating both strands of a DNA duplex, R781 contacts the nucleophilic Y782 and the scissile phosphate, assuming a "down" conformation (Figure 5A, central panel).[55,71] The experimental work of Schmidt and colleagues has shown that the closeness of R781 to both the DNA strand and Y782 is functional to the catalysis, as it stabilizes the DNA 5′ prior phosphodiester bond cleavage and lowers the pK_a of the

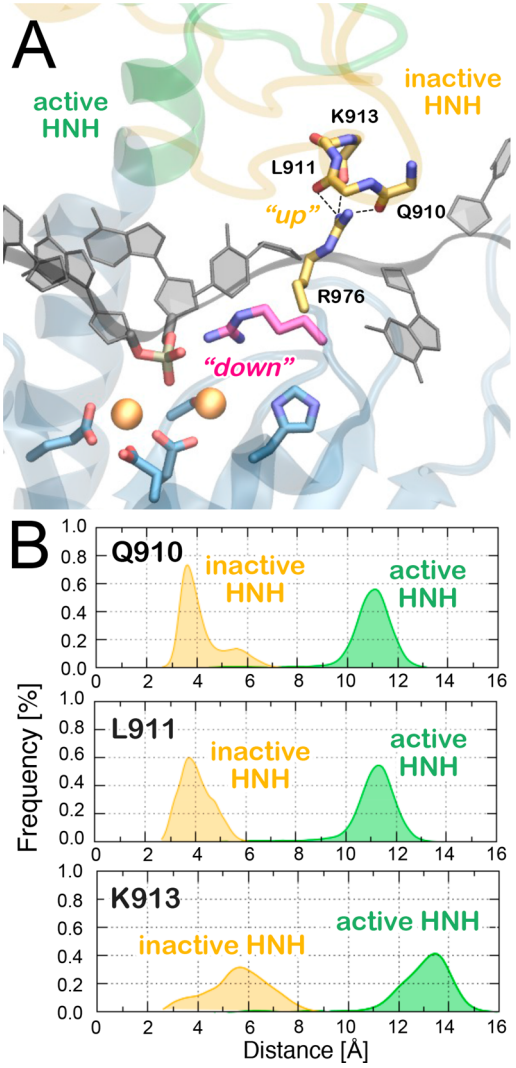


Figure 4. (A) Configuration of R976 in the inactive CRISPR−Cas9 complex (R976 in yellow), compared to the configuration of R976 in the activated complex (R976 in magenta). The inactive configuration of the HNH domain (shown as ribbons in yellow) is superposed to the activated configuration (ribbons, green), showing that in the active conformation, HNH moves away from the RuvC active site. (B) Probability distributions of the distances between the R976 terminal nitrogen atom (NH1) and the backbone oxygen atoms of the Q910, L911, and K913 residues, computed over GaMD simulations of the inactive (yellow) and active (green) CRISPR−Cas9 complexes (data are aggregated over the simulation replicas).

attacking Y782. In the polymerase−η (pol−η), which repairs DNA lesions by extending damaged strands, an arginine residue (i.e., Arg61) also locates upstream the active site (Figure 5A, right panel). In the X-ray structure of the reactant state of pol−η, Arg61 points its positive side chain toward the dATP α-phosphate, assuming a "down" conformational state for the 85% of occupancy, as opposite to what observed in the product state (where Arg61 assumes the "up" conformation).[72] Recent computational investigations provided a rational for this conformational change, suggesting that in the "down" conformation, Arg61 stabilizes the formation of a catalytic complex for the polymerase reaction.[73] Remarkably, MD simulations of the activated pol−η show that Arg61 mainly adopts a "down" conformational state, rarely transitioning in the "up" conformational state, similar to what observed in the

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

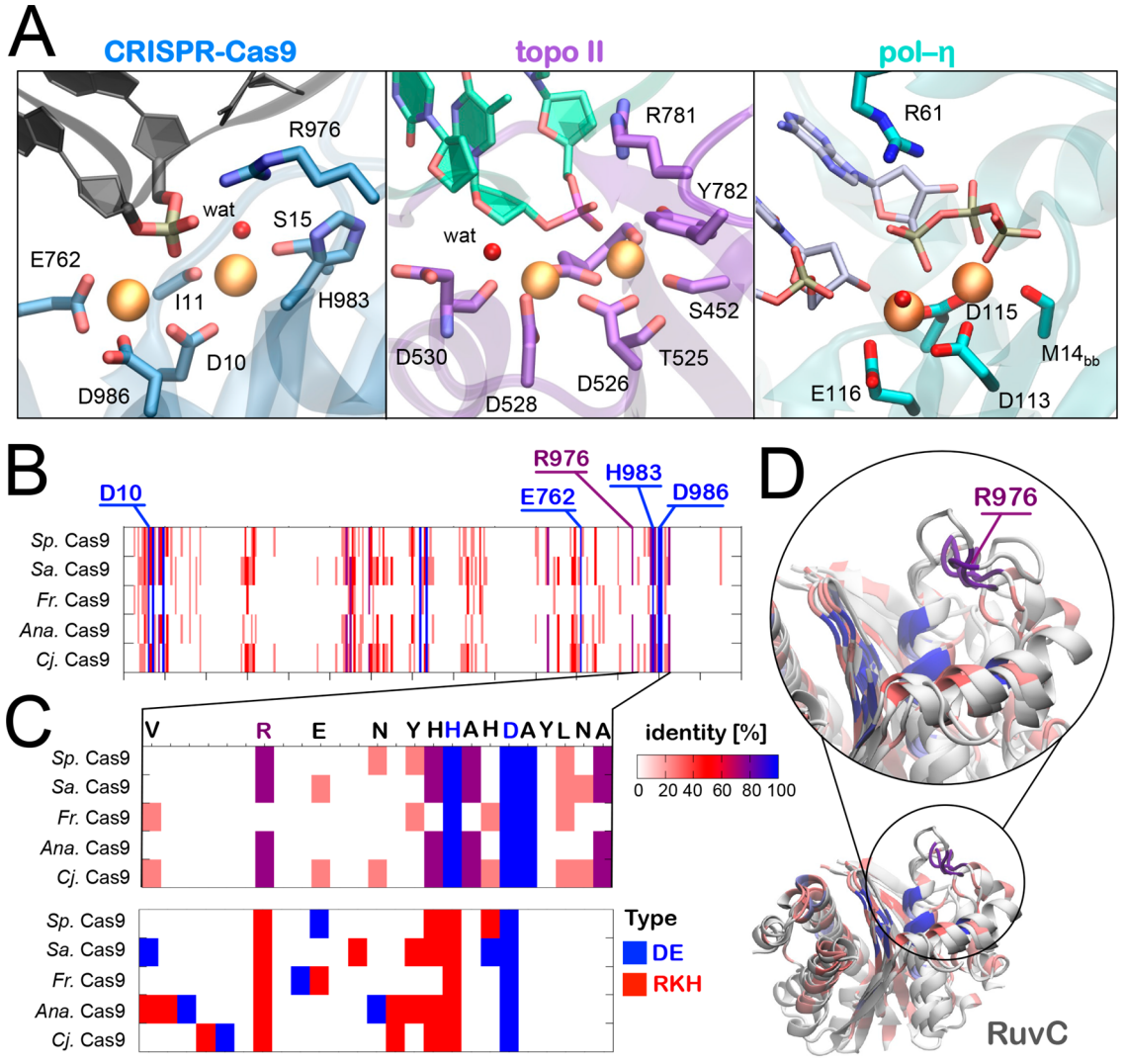


Figure 5. (A) Active site of the RuvC domain in the activated CRISPR–Cas9 (left panel) in comparison with type II topoisomerase[71] (topo II, central panel) and polymerase–η[72] (pol–η, right panel). (B) Alignment of the primary sequence of the RuvC domain in the *Streptococcus pyogenes* (*Sp*),[5] *Staphylococcus aureus* (*Sa*),[59] *Francisella Novicida* (*Fn*),[60] *Actinomyces naeslundii* (*Ana*),[4] and *Campylobacter jejuni* (*Cj*)[61] CRISPR–Cas9 systems. The sequence identity is color-coded from white to blue to indicate a percentage of sequence identity from 0 to 100%. (C) The region from residues 970–990 is enlarged and color-coded to display the sequence identity (top) and the presence of similar residue types (bottom, red for RKH and blue for D and E). (D) Structural superimposition of the RuvC domain in the investigated systems, with close up view displaying the location of R976 in the five systems. The structures are shown as ribbons, color-coded in agreement with the sequence identity scale (i.e., white no identity; blue full identity).

present study for CRISPR–Cas9 (Figures 3 and 5A, left panel). In CRISPR–Cas9, DNA cleavage assays by capillary electrophoresis have shown that the R976A mutation reduces the cleavage of the NTS.[6] This experimental evidence pinpoints a critical role of R976 in the catalysis of the NTS. In light of this and considering that in the above-mentioned two-metal aided enzymes, arginine residues in the "down" conformational state exert an important stabilization role, it is reasonable to speculate that also in CRISPR–Cas9 the "down" conformation of R976 confers stabilization to the NTS prior phosphodiester bond cleavage. Intrigued by the role of the arginine fingers in DNA and RNA processing enzymes, we have been seeking more broadly for analogies in the two-metal aided enzymes, by considering the database of two-metal aided enzymes crystallized at high resolution recently collected.[74] We found that an arginine residue upstream of the nucleic acids resembling the position of R976 in the CRISPR–Cas9 complex is present in different enzymatic classes (EC), encompassing enzymes such as the restriction endonuclease *Bam*HI,[75] the Exonuclease 1,[76] and several DNA/RNA polymerases[77−81] (Table S2). Remarkably, in the majority of the enzymes crystallized at high resolution (i.e., < 3 Å) in a catalytically competent state, the arginine residue assumes the "down" conformation pointing toward the processed phosphate. This evidence further supports our computational model of the activated CRISPR–Cas9 complex. It is important to note, however, that in-depth computational studies and experimental data assessing the importance for the catalysis of arginine residues upstream the two-metal ions motif have been performed only for CRISPR–Cas9, topo II, and pol–η (Figure 5A). For the remaining enzymes reported in Table S2 experimental assays are absent. However, the observed common feature in the structures of the two-metal aided enzymes suggests an important role of arginine fingers in their

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

active sites, which calls for future experimental validation to investigate the role of arginine residues upstream the two-metal ions motif. Finally, to seek the presence of R976 in the Cas9 enzymes from different organisms, we performed sequence alignment of the RuvC domain considering the *Streptococcus pyogenes* (*Sp*, object of this study),[5] *Staphylococcus aureus* (*Sa*),[59] *Francisella Novicida* (*Fn*),[60] *Actinomyces naeslundii* (*Ana*)[4] and *Campylobacter jejuni* (*Cj*)[61] organisms (Figure 5B−C, the full sequence alignment is in Figure S6). We found that, although the overall sequence identity of the RuvC domains is low (i.e., < 30%), the active site residues directly responsible for the catalysis (i.e., D10, E762, H983, and D986) show 100% of sequence identity. R976 also shows very high sequence identity (80%), being present all Cas9 systems but in the *Fn* Cas9, where it is replaced by a lysine (retaining a positively charged side chain, like arginine). Remarkably, structural alignment of the X-ray structures reveals that the RuvC domain maintains its characteristics $\beta\beta\alpha$−metal binding fold (in spite of the low sequence identity) and shows that R976 locates in the same position upstream of the active site (Figure 5D). Overall, the DNA cleavage assays assessing the importance of R976 in the cleavage of the NTS[6] and the similarity with the two-metal aided enzymes, support our computational model of the activated CRISPR−Cas9 complex. As well, the stabilization role exerted by the arginine fingers in other Mg-aided phosphatases[70] and in the two-metal aided enzymes topo I[55,71] and pol−η[72,73] suggests a similar function in CRISPR−Cas9, where R976 could stabilize the formation of a catalytic complex.

DISCUSSION

The CRISPR−Cas9 system attains genome editing thanks to two nuclease domains—HNH and RuvC—which perform concerted cleavages of double stranded DNA.[13] Over the last years, the integration of structural data,[5] extensive MD simulations,[16,17] single molecule FRET, and biochemical experiments[13,21] enabled the identification of a catalytically active CRISPR−Cas9 complex (Figure 1B). In this structure, the catalytic HNH domain docks at the cleavage site on the TS, while the RuvC active site hosts the NTS, prone for double stands cleavages. Here, this activated complex has been used as a basis for the investigation of the structure and dynamics of the RuvC active site through extensive MD simulations. *Ab-initio* QM/MM simulations, accelerated MD, and bioinformatics analysis enabled an in-depth characterization of the RuvC active site, revealing its two-metal aided architecture.

The description of the metal aided active site has been attained through *ab-initio* QM/MM simulations, offering a more accurate description of the coordination geometry with respect to the empirical force field models.[66] As a result, the two catalytic Mg^{2+} ions are coordinated by a highly conserved sequence composed by the DDE residues, which characterizes the *two-metal-aided* architecture in DNA and RNA processing enzymes (Figure 2A−B).[23,35] The scissile phosphate locates between the two catalytic Mg^{2+} ions, which are jointly coordinated by the *pro*-Sp oxygen, enabling an in-line S_N2-like nucleophilic attack on the scissile phosphorus atom.[23,25,26,65] The catalytic water molecule coordinates MgA and stably locates in close proximity to the scissile phosphate and to the catalytic H893 (Figure 2C, bottom panel). The latter is suggested to act as a general base for the activation of the water nucleophile,[11] as also shown in other two-metal aided enzymes.[35] The identified active site geometry is in good agreement with the canonical two-metal aided sites, which have been structurally determined in the prototypical RNase H[64] (Figure 2A) and in other DNA/RNA processing enzymes.[21,73] Moreover, the architecture observed in the present study holds the main prerequisites for attaining the metal aided phosphodiester bond cleavage[23,25,26,65] and remarkably differs from the previous models. In a previous computational study, force field-based MD runs (carried out over ∼40 ns) attempted to identify the active site chemistry of the RuvC domain in the inactive crystal structure (5F9R.pdb).[22] The authors have shown that the catalysis is likely to occur at position −4, in agreement with the current study and other previous experimental investigations.[82] However, the identified catalytic site shows that the *pro*-Rp and *pro*-Sp oxygen atoms coordinate the MgA and MgB ions, respectively, leading to a structure that is unlikely to give rise to phosphodiester bond cleavage. This unlikely configuration is also observed in a model derived from the empirical reconstruction of a cryo-EM map solved at 5.2 Å resolution,[6] making this structure unavailable for further investigations. Contrariwise, the active site geometry identified in the present study reveals a joint coordination of the two Mg^{2+} ions by the *pro*-Sp oxygen of the scissile phosphate, favoring an in-line S_N2-like nucleophilic attack.[23,25,26,65] Notably, the previous studies exclusively employed classical and force field-based models. Hence, the obtained active site geometry might have been affected by the well-known limitations of the classical force field for Mg^{2+} ions.[66] Moreover, the previous classical study has been based on the inactivated structure of the CRISPR−Cas9 complex.[22] This is a critical point, because the activated configuration of the HNH functionality is an essential requirement for the formation of a catalytically competent RuvC active site.[13]

Through the application of unconstrained enhanced sampling, by using a Gaussian accelerated MD (GaMD) method,[27] we broadly explored the conformational dynamics of the RuvC domain at the classical MD level. As a result, we found that in the activated CRISPR−Cas9 complex, the R976 residue approaches the cleavage site in the RuvC domain bringing its positive side chain in close contact to the scissile phosphate (Figure 3A). The R976 residue stably maintains this configuration throughout the dynamics. The approach of R976 to the scissile phosphate of the active Cas9 is of particular interest in comparison with other Mg^{2+}-aided phosphatases[70] and DNA/RNA processing enzymes—such as type II topoisomerase (topo II)[55,71] and the polymerase−η (pol−η)[72,73]—that share with Cas9 a similar two-metal aided architecture. In these enzymes, the presence of an arginine finger upstream of the *metal aided* site exerts a functional role, stabilizing the formation of a catalytically active complex (Figure 5A). Structural comparison with other two-metal aided enzymes belonging to different enzymatic classes, as well as the sequence alignment of the RuvC domain from different CRISPR−Cas9 organisms (Figure 5B-D), revealed that the presence of an arginine finger upstream of the active site is a conserved feature. These analogies support our computational model of the activated CRISPR−Cas9 complex. Moreover, DNA cleavage assays have shown that the R976A mutation in CRISPR−Cas9 reduces the cleavage of the NTS,[6] indicating a key role of the residue for the activity. In light of this experimental evidence, the stabilization role exerted by the arginine fingers contacting the scissile phosphate in other Mg-aided phosphatases[70−72] suggests a similar function in

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

CRISPR−Cas9, where R976 could stabilize the formation of a catalytic complex. MD simulations of the inactive CRISPR−Cas9 complex revealed that R976 moves away from the active site in the inactive CRISPR−Cas9 complex (Figure 3A). In this complex, R976 is shown to interact with the HNH domain and is unable to properly locate to contact the scissile phosphate on the NTS. Notably, in the inactive configuration of CRISPR−Cas9, the HNH domain locates away from the TS and contacts the NTS within the RuvC core (Figures 1A and 4). Upon activation, the HNH domain is detached from the NTS and approaches the TS, resulting in a loss of structural constraint on the RuvC active site. This results in newly formed interactions—such as the approach of R976 to the scissile phosphate—which contribute in the formation of a catalytically competent state in the RuvC domain. This evidence is in agreement with the experimental fact that, in order to perform concerted cleavages of double stranded DNAs, both the nuclease domains have to assume an activated configuration.[13] As well, the conformational activation of HNH is an essential requirement for the formation of a catalytically active RuvC.[13,69]

Overall, the present computational study discloses the structure and dynamics of the catalytically competent CRISPR−Cas9 system, which is in good agreement with the available experimental evidence. This study therefore provides an important contribution for the structure-based engineering CRISPR−Cas9 toward improved and/or controlled DNA cleavages. Moreover, the identified structure constitutes a stepping-stone for future QM/MM investigations of the metal-aided catalysis, currently ongoing in our laboratory. This will ultimately characterize the chemistry of the CRISPR−Cas9 system.

CONCLUSIONS

Here, we disclosed the structure and dynamics of the catalytically competent CRISPR−Cas9 system through the use of extensive molecular simulations. We combined an *ab-intio* QM/MM approach and accelerated MD simulations, coupled with bioinformatics analysis, to obtain an in-depth characterization of the RuvC active site in the activated complex. We disclosed a canonical two-metal aided architecture in the RuvC active site, which is poised to operate the catalysis of the NTS, in analogy with other DNA/RNA processing enzymes.[23,35] In the active site, the two catalytic Mg^{2+} ions are coordinated by a highly conserved sequence composed by the DDE residues, while the DNA scissile phosphate locates between the two divalent ions, enabling an in-line S_N2-like nucleophilic attack. Remarkably, the formation of a catalytically competent state of the RuvC domain is observed upon the conformational activation of the other nuclease domain of Cas9—i.e., the HNH domain—such justifying concerted cleavages of double stranded DNA.[13] The use of a recently developed accelerated MD methodology—i.e., a Gaussian accelerated MD (GaMD)—further enabled broad exploration of the conformational dynamics of the RuvC domain over longer time scales. As a result, an arginine finger is found to stably contact the scissile phosphate, as also observed in other Mg^{2+}-aided phosphatases[70] and DNA/RNA two-metal aided enzymes,[55,71−73,83] with the function of stabilizing the formation of a catalytically active complex. Overall, the agreement with the available experimental studies on CRISPR−Cas9 and the analogies found with other two-metal aided enzymes support the model proposed in the present study, offering a reliable description of the active site chemistry. Notably, this structure remarkably differs from previous models based on classical mechanics, and demonstrates that only *ab-initio* QM/MM MD can accurately describe the metal-aided chemistry in CRISPR−Cas9. *Ab-initio* QM/MM approaches will therefore be necessary for completing the characterization of the metal-aided catalysis and for describing the active site of the HNH nuclease functionality. This research is currently ongoing in our laboratory, building on our experience in *ab-initio* MD and QM/MM simulations.[24,54,84,85] Finally, the catalytically active state identified here offers the basis for novel experimental investigations aimed at fully clarifying the role of the residues that coordinate the Mg^{2+} ions in the catalytic sites. Based on our results, we speculate that the residues coordinating the metal ions in the active sites of the HNH (D861, V838, D839; Figure 1C) and RuvC (E762, D986, D10, I11, S15; Figure 2A) domains could be essential for the catalysis and, more broadly, for the selectivity of the system against off-target sequences. This in-depth understanding of the CRISPR−Cas9 catalytic complex also poses the basis for further development of novel CRISPR−Cas9 genome editing tools with improved specificity and catalytic efficiency.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00988.

Supplementary figures, tables, and references (PDF)

AUTHOR INFORMATION

Corresponding Author

*Email: giulia.palermo@ucr.edu.

ORCID

Giulia Palermo: 0000-0003-1404-8737

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Janice S. Chen and Dr. Lorenzo Casalino for useful discussions, and Prof. Victor Rodgers for proofreading the manuscript. This research was supported by initial complement funding from the University of California Riverside to G.P. Computer time was provided by the Extreme Science and Engineering Discovery Environment (XSEDE) through the grant TG-MCB160059.

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

(4) 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.

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(5) 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.
(6) Huai, C.; Li, G.; Yao, R. J.; Zhang, Y.; Cao, M.; Kong, L. L.; Jia, C. Q.; Yuan, H.; Chen, H. Y.; Lu, D. R.; Huang, Q. Structural Insights into DNA Cleavage Activation of CRISPR−Cas9 system. *Nat. Commun.* **2017**, *8*, 1375.
(7) 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.
(8) 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.
(9) 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.
(10) Hirano, S.; Nishimasu, H.; Ishitani, R.; Nureki, O. Structural Basis for the Altered PAM Specificities of Engineered CRISPR−Cas9. *Mol. Cell* **2016**, *61*, 886−94.
(11) 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.
(12) Nishimasu, H.; Shi, X.; Ishiguro, S.; Gao, L.; Hirano, S.; Okazaki, S.; Noda, T.; Abudayyeh, O. O.; Gootenberg, J. S.; Mori, H.; Oura, S.; Holmes, B.; Tanaka, M.; Seki, M.; Hirano, H.; Aburatani, H.; Ishitani, R.; Ikawa, M.; Yachie, N.; Zhang, F.; Nureki, O. Engineered CRISPR−Cas9 Nuclease with Expanded Targeting Space. *Science* **2018**, *61*, 1259−1262.
(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) Biertumpfel, C.; Yang, W.; Suck, D. Crystal Structure of T4 Endonuclease VII Resolving a Holliday Junction. *Nature* **2007**, *449*, 616−620.
(15) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. Striking Plasticity of CRISPR−Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Cent. Sci.* **2016**, *2*, 756−763.
(16) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR−Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7260−7265.
(17) Palermo, G.; Chen, J. S.; Ricci, C. G.; Rivalta, I.; Jinek, M.; Batista, V. S.; Doudna, J. A.; McCammon, J. A. Key Role of the REC Lobe During CRISPR−Cas9 Activation by "Sensing", "Regulating" and "Locking" the Catalytic HNH domain. *Q. Rev. Biophys.* **2018**, *51*, No. e9.
(18) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis* **2014**, 41−53.
(19) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A.; et al. Enhanced Proofreading Governs CRISPR−Cas9 Targeting Accuracy. *Nature* **2017**, *550*, 407−410.
(20) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A. A Conformational Checkpoint Between DNA Binding And Cleavage By CRISPR−Cas9. *Sci. Adv.* **2017**, *3*, No. eaao0027.
(21) Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR−Cas9 Activity. *Cell Rep.* **2018**, *9*, 359−371.
(22) Zuo, Z.; Liu, J. Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Sci. Rep.* **2016**, *5*, 37584.
(23) 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.
(24) 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.
(25) 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.
(26) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 8934−8941.
(27) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584−3595.
(28) Sibener, L. V.; Fernandes, R. A.; Kolawole, E. M.; Carbone, C. B.; Liu, F.; McAffee, D.; Birnbaum, M. E.; Yang, X.; Su, L. F.; Yu, W.; Dong, S.; Gee, M. H.; Jude, K. M.; Davis, M. M.; Groves, J. T.; Goddard, W. A., 3rd; Heath, J. R.; Evavold, B. D.; Vale, R. D.; Garcia, K. C. Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. *Cell* **2018**, *174*, 672−687.
(29) Miao, Y.; McCammon, J. A. Mechanism of the G-protein Mimetic Nanobody Binding to a Muscarinic G-Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 3036−3041.
(30) Miao, Y.; McCammon, J. A. Graded Activation and Free Energy Landscapes of a Muscarinic G Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12162−12167.
(31) Chuang, C. H.; Chiou, S. J.; Cheng, T. L.; Wang, Y. T. A Molecular Dynamics Simulation Study Decodes the Zika Virus NS5Methyltransferase Bound to SAH and RNA Analogue. *Sci. Rep.* **2018**, *8*, 6336.
(32) Wang, Y. T.; Chan, Y. H. Understanding the Molecular Basis of Agonist/Antagonist Mechanism of Human Mu Opioid Receptor through Gaussian Accelerated Molecular Dynamics method. *Sci. Rep.* **2017**, *7*, 7828.
(33) Liao, J. M.; Wang, Y. T. In Silico Studies of Conformational Dynamics of Mu Opioid Receptor Performed using Gaussian Accelerated Molecular Dynamics. *J. Biomol. Struct. Dyn.* **2019**, *37*, 166.
(34) Salawu, E. O. The Impairment of TorsinA's Binding to and Interactions With its Activator: An Atomistic Molecular Dynamics Study of Primary Dystonia. *Front Mol. Biosci* **2018**, *5*, 64.
(35) Yang, W. Nucleases: Diversity of Structure, Function and Mechanism. *Q. Rev. Biophys.* **2011**, *44*, 1−93.
(36) 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.
(37) 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.
(38) 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.
(39) Aqvist, J. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(40) Casalino, L.; Palermo, G.; Abdurakhmonova, N.; Rothlisberger, U.; Magistrato, A. 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.* **2017**, *13*, 340−352.

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(41) Turq, P.; Lantelme, F.; Friedman, H. L. Brownian Dynamics - its Application to Ionic-Solutions. *J. Chem. Phys.* **1977**, *66*, 3039−3044.
(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(43) Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham, I., T. E; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2016*, University of California, San Francisco, 2016.
(44) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.
(45) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(46) Parrinello, M.; Andreoni, W.; Curioni, A. *CPMD*; IBM Corporation and Max-Planck Institut: Armonk, New York and Stuttgart, Germany, 2000.
(47) Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. II. Operators for Fast Iterative Diagonalization. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 8861−8869.
(48) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Optimization of Effective Atom Centered Potentials for London Dispersion Forces in Density Functional Theory. *Phys. Rev. Lett.* **2004**, *93*, 153004.
(49) Martyna, G. J.; Tuckerman, M. E. A Reciprocal Space Based Method for Treating Long Range Interactions in Ab Initio and Force-Field-Based Calculations in Clusters. *J. Chem. Phys.* **1999**, *110*, 2810−2821.
(50) Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car−Parrinello Molecular Dynamics Simulations. *J. Chem. Phys.* **2002**, *116*, 6941−6947.
(51) Hoover, W. G. Canonical Dynamics - Equilibrium Phase-space Distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695−1697.
(52) Nose, S. An Extension of the Canonical Ensemble Molecular-Dynamics Method. *Mol. Phys.* **1986**, *57*, 187−191.
(53) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.
(54) Ma, Z.; Palermo, G.; Adhireksan, Z.; Murray, B. S.; von Erlach, T.; Dyson, P. J.; Rothlisberger, U.; Davey, C. A. An Organometallic Compound Displays a Unique One-Stranded Intercalation Mode that is DNA Topology-Dependent. *Angew. Chem., Int. Ed.* **2016**, *128*, 7441−7444.
(55) 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.
(56) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(57) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10*, 2677−2689.
(58) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Molecular Mechanism of Off-Target Effects in CRISPR−Cas9. *BiorXiv* **2018**, DOI: 10.1101/421537.
(59) Nishimasu, H.; Cong, L.; Yan, W. X.; Ran, F. A.; Zetsche, B.; Li, Y. Q.; Kurabayashi, A.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Staphylococcus aureus Cas9. *Cell* **2015**, *162*, 1113−1126.
(60) Hirano, H.; Gootenberg, J. S.; Horii, T.; Abudayyeh, O. O.; Kimura, M.; Hsu, P. D.; Nakane, T.; Ishitani, R.; Hatada, I.; Zhang, F.; Nishimasu, H.; Nureki, O. Structure and Engineering of Francisella Novicida Cas9. *Cell* **2016**, *164*, 950−961.
(61) Yamada, M.; Watanabe, Y.; Gootenberg, J. S.; Hirano, H.; Ran, F. A.; Nakane, T.; Ishitani, R.; Zhang, F.; Nishimasu, H.; Nureki, O. Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR−Cas9 Systems. *Mol. Cell* **2017**, *65*, 1109−1121.
(62) Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. MultiSeq: Unifying Sequence and Structure Data for Evolutionary Analysis. *BMC Bioinf.* **2006**, *7*, 382.
(63) 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*, 1781−802.
(64) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. Crystal Structures of RNase H Bound to an RNA/DNA Hybrid: Substrate Specificity and Metal-Dependent Catalysis. *Cell* **2005**, *121*, 1005−1016.
(65) Elsasser, B.; Fels, G. Atomistic Details of the Associative Phosphodiester Cleavage in Human Ribonuclease H. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11081−11088.
(66) Li, P. F.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564−1686.
(67) Cisneros, G. A.; Perera, L.; Schaaper, R. M.; Pedersen, L. C.; London, R. E.; Pedersen, L. G.; Darden, T. A. Reaction Mechanism of the Epsilon Subunit of E-coli DNA Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. *J. Am. Chem. Soc.* **2009**, *131*, 1550−1556.
(68) Gorecka, K. M.; Komorowska, W.; Nowotny, M. Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic Acids Res.* **2013**, *41*, 9945−9955.
(69) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. PAM-Induced Allostery Activates CRISPR−Cas9. *J. Am. Chem. Soc.* **2017**, *139*, 16028−16031.
(70) Nagy, G. N.; Suardiaz, R.; Lopata, A.; Ozohanics, O.; Vekey, K.; Brooks, B. R.; Leveles, I.; Toth, J.; Vertessy, B. G.; Rosta, E. Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases. *J. Am. Chem. Soc.* **2016**, *138*, 15035−15045.
(71) Schmidt, B. H.; Burgin, A. B.; Deweese, J. E.; Osheroff, N.; Berger, J. M. A Novel and Unified Two-Metal Mechanism for DNA Cleavage by Type II and IA Topoisomerases. *Nature* **2010**, *465*, 641−644.
(72) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W. Watching DNA Polymerase eta Make a Phosphodiester Bond. *Nature* **2012**, *487*, 196−201.
(73) Genna, V.; Gaspari, R.; Dal Peraro, M.; De Vivo, M. Cooperative Motion of a Key Positively Charged Residue and Metal Ions for DNA Replication Catalyzed by Human DNA Polymerase-eta. *Nucleic Acids Res.* **2016**, *44*, 2827−2836.
(74) Genna, V.; Colombo, M.; De Vivo, M.; Marcia, M. Second-Shell Basic Residues Expand the Two-Metal-Ion Architecture of DNA and RNA Processing Enzymes. *Structure* **2018**, *26*, 40−50.
(75) Viadiu, H.; Aggarwal, A. K. Structure of BamHI Bound to Nonspecific DNA: a Model for DNA Sliding. *Mol. Cell* **2000**, *5*, 889−895.
(76) Orans, J.; McSweeney, E. A.; Iyer, R. R.; Hast, M. A.; Hellinga, H. W.; Modrich, P.; Beese, L. S. Structures of Human Exonuclease 1 DNA Complexes Suggest a Unified Mechanism for Nuclease Family. *Cell* **2011**, *145*, 212−223.
(77) Sawaya, M. R.; Prasad, R.; Wilson, S. H.; Kraut, J.; Pelletier, H. Crystal Structures of Human DNA Polymerase Beta Complexed with Gapped and Nicked DNA: Evidence for an Induced Fit Mechanism. *Biochemistry* **1997**, *36*, 11205−11215.
(78) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Human DNA Polymerase Iota Incorporates dCTP Opposite

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Template G via a G.C + Hoogsteen Base Pair. *Structure* **2005**, *13*, 1569−1577.
(79) Silverstein, T. D.; Johnson, R. E.; Jain, R.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Structural Basis for the Suppression of Skin Cancers by DNA Polymerase eta. *Nature* **2010**, *465*, 1039−1043.
(80) Sharma, A.; Kottur, J.; Narayanan, N.; Nair, D. T. A Strategically Located Serine Residue is Critical for the Mutator Activity of DNA Polymerase IV from Escherichia Coli. *Nucleic Acids Res.* **2013**, *41*, 5104−5114.
(81) Wang, D.; Bushnell, D. A.; Westover, K. D.; Kaplan, C. D.; Kornberg, R. D. Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell* **2006**, *127*, 941−954.
(82) Raper, A. T.; Stephenson, A. A.; Suo, Z. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.* **2018**, *140*, 2971−2984.
(83) 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.
(84) Palermo, G.; Campomanes, P.; Cavalli, A.; Rothlisberger, U.; De Vivo, M. Anandamide Hydrolysis in FAAH Reveals a Dual Strategy for Efficient Enzyme-Assisted Amide Bond Cleavage via Nitrogen Inversion. *J. Phys. Chem. B* **2015**, *119*, 789−801.
(85) Palermo, G.; Branduardi, D.; Masetti, M.; Lodola, A.; Mor, M.; Piomelli, D.; Cavalli, A.; De Vivo, M. Covalent inhibitors of fatty acid amide hydrolase: a rationale for the activity of piperidine and piperazine aryl ureas. *J. Med. Chem.* **2011**, *54*, 6612−6623.

DOI: 10.1021/acs.jcim.8b00988
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX