Structural and functional insights into the 
bona fide catalytic state of Streptococcus pyogenes Cas9 HNH nuclease domain

Zhicheng Zuo¹ 2, Ashwini Zolekar¹, Kesavan Babu³, Victor JT Lin¹, Hamed S Hayatshahi¹, Rakhi Rajan³, Yu-Chieh Wang¹ *, Jin Liu¹ *

¹Department of Pharmaceutical Sciences, UNT System College of Pharmacy, University of North Texas Health Science Center, Fort Worth, United States; ²College of Chemistry and Chemical Engineering, Shanghai University of Engineering Science, Shanghai, China; ³Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, Norman, United States

Abstract The CRISPR-associated endonuclease Cas9 from Streptococcus pyogenes (SpyCas9), along with a programmable single-guide RNA (sgRNA), has been exploited as a significant genome-editing tool. Despite the recent advances in determining the SpyCas9 structures and DNA cleavage mechanism, the cleavage-competent conformation of the catalytic HNH nuclease domain of SpyCas9 remains largely elusive and debatable. By integrating computational and experimental approaches, we unveiled and validated the activated Cas9-sgRNA-DNA ternary complex in which the HNH domain is neatly poised for cleaving the target DNA strand. In this catalysis model, the HNH employs the catalytic triad of D839-H840-N863 for cleavage catalysis, rather than previously implicated D839-H840-D861, D837-D839-H840, or D839-H840-D861-N863. Our study contributes critical information to defining the catalytic conformation of the HNH domain and advances the knowledge about the conformational activation underlying Cas9-mediated DNA cleavage.

Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease Cas9 from Streptococcus pyogenes (SpyCas9) has become a gene-editing tool that holds an immense promise for the development of novel therapeutic approaches for human diseases (Cong et al., 2013; Jinek et al., 2012; Knott and Doudna, 2018; Mali et al., 2013). Two, magnesium (Mg²⁺) ion-dependent, nuclease domains (i.e. HNH and RuvC domains) in Cas9 cleave the target DNA strand (tDNA) complementary to the guide region of a dual single-guide RNA (sgRNA) and the non-target DNA strand (ntDNA), respectively (Gasiunas et al., 2012; Jinek et al., 2012). The conformational and mechanistic knowledge of Cas9 activation to achieve DNA cleavage is essential for the rational engineering of Cas9 to possibly ensure minimal off-target effects while retaining high gene-editing efficiency (Chen et al., 2017a; Kleinstiver et al., 2016; Slaymaker et al., 2016; Sternberg et al., 2015).

Several structures of the SpyCas9-sgRNA-DNA ternary complex to depict the HNH nuclease domain in a ‘cleavage-competent’ state have been reported (Huai et al., 2017; Palermo et al., 2018; Palermo et al., 2017; Zuo and Liu, 2017). Notably, the amino acid residue D861 in the HNH domain of SpyCas9 pointed towards the catalytic center in the absolute majority of resolved crystal structures (Anders et al., 2016; Anders et al., 2014; Dong et al., 2017; Hirano et al., 2016; Jiang et al., 2015; Jiang et al., 2016; Jinek et al., 2014; Liu et al., 2019; Nishimasu et al., 2014; Zuo et al., 2017; Zuo and Liu, 2017).
The DNA inside human cells provides instructions for all of the processes that happen inside the body. Errors in the DNA may lead to cancer, sickle cell disease, cystic fibrosis, Huntington’s disease, or other genetic disorders. Medical researchers are exploring whether it is possible to replace or repair the faulty DNA (an approach known as gene therapy) to reduce the symptoms, or even cure individuals, of these conditions.

Over the last ten years, a new technology known as CRISPR-Cas9 gene editing has proved to be a reliable and efficient way to make small and precise changes to DNA in living cells. First, an enzyme called Cas9 searches for a segment of target DNA that matches a template molecule the enzyme carries. Cas9 then cuts the target DNA, which is repaired to match a new customized DNA sequence: this changes the genetic information of the cell.

The Cas9 protein is made of a succession of building blocks called amino acids that create long chains which then fold to form the final three-dimensional shape of the enzyme. A region of Cas9 known as the HNH domain is responsible for cutting the target DNA. However, it remains unclear exactly which amino acids within this domain work together to sever the DNA.

Here, Zuo et al. combined computational and experimental approaches to reveal the three-dimensional structure of the Cas9 enzyme when the HNH domain is poised to cut the target DNA. The findings were used to generate a computational model of Cas9 and this model predicted that the HNH domain relies on a group of three amino acids known collectively as D839-H840-N863 to cleave DNA strands.

This knowledge is useful to understand exactly how Cas9 modifies genetic information. Ultimately, this may help to improve CRISPR-Cas9 technology so it could be safely used in gene editing therapies.

DOI: https://doi.org/10.7554/eLife.46500.002
Figure 1. Architecture of the HNH domain βαα-Me fold in different binding forms of Cas9 (a–d) and site-directed mutagenesis experiments identifying potential catalytic residues (e–f). (a–d) βαα-Me fold in the sgRNA-bound state of Cas9 (a), in the intermediate state (b), in the pre-catalytic state (c), and in the pseudoactive state (d). The βαα-Me fold is represented as pink ribbons, and the residues are shown in stick models and colored by atom type (C, dark green; N, blue; O, red). If present, the bound Mg$^{2+}$ ion is depicted as a magenta sphere, and only the tDNA phosphate-sugar backbone is shown.

Zuo et al. eLife 2019;8:e46500. DOI: https://doi.org/10.7554/eLife.46500
displayed for clarity. The location of the Cas9 D861 is highlighted by an arrow, and the dashed lines denote hydrogen bonds or coordinative bonds. (e) The expression and DNA-editing activity of the wild-type and D861A variants of Cas9 paired with an sgRNA sequence that targets the egfp gene in HEK293T-EGFP cells. (f) The expression and DNA-editing activity of the wild-type and indicated variants of Cas9 paired with an sgRNA sequence that targets the egfp gene in HEK293T-EGFP cells. The retention of EGFP expression reflected the loss of activity of Cas9 protein in the cells.

DOI: https://doi.org/10.7554/eLife.46500.008

The following source data and figure supplements are available for figure 1:

Source data 1. Numeric data for activity quantification of Cas9 and its variants (XLS).
DOI: https://doi.org/10.7554/eLife.46500.009

Figure supplement 1. The DNA sequencing analysis of vectors for expressing different Cas9 variants and the flow cytometry analysis of HEK293T-EGFP cells with different Cas9-sgRNA expression vectors.
DOI: https://doi.org/10.7554/eLife.46500.004

Figure supplement 2. Plasmid cleavage activity of SpyCas9WT.
DOI: https://doi.org/10.7554/eLife.46500.005

Figure supplement 3. Cleavage of radioisotope-labeled oligo DNA substrate by SpyCas9WT, SpyCas9D861A and SpyCas9N863A.
DOI: https://doi.org/10.7554/eLife.46500.006

Figure supplement 4. The architecture of the HNH domain ββα-Me motif in the apo structure of SpyCas9 (a) and structural modeling of SpyCas9 with Mg2+ ion bound at the catalytic center (b–c).
DOI: https://doi.org/10.7554/eLife.46500.007

Figure supplement 5. The architecture of the HNH domain ββα-Me motif in the apo structure of Actinomyces naeslundii Cas9 (AnaCas9) (a) and in that of partial dsDNA-bound Staphylococcus aureus Cas9 (SauCas9) (b).
DOI: https://doi.org/10.7554/eLife.46500.008

Results and discussion

Intrigued by the above paradoxical findings between the structural and functional experiments, we performed molecular modeling and molecular dynamics (MD) simulations to further investigate the residues that may participate in the catalysis of tDNA cleavage. We first examined the apo-state crystal structure of Cas9 (Jinek et al., 2014). One noticeable feature in this apo-Cas9 structure is that the α-helical element in the HNH domain ββα-metal (ββα-Me) fold appears to pose a unique conformation with N863 pointing toward the catalytic center (Figure 1a–1d). This inward orientation of N863 was distinct from the inward orientation of D861 observed in the Cas9 structures co-crystalized with sgRNA and/or DNA (Figure 1a–1d). However, the N-terminal of the α-helical segment is disordered in this apo-Cas9 structure (Figure 1a–1d), suggesting a high conformational flexibility around this region. We completed the disordered regions of the apo-Cas9 structure by homology modeling and performed molecular dynamics (MD) simulation with a Mg2+ ion placed in the catalytic ββα-Me motif (Figure 1b–1c). As a result, the Mg2+ ion formed an octahedral coordination with D839, N863, and four water molecules (Figure 1c), which closely resembled the Mg2+ coordination in the X-ray crystal structure of Actinomyces naeslundii Cas9 (AnaCas9, Figure 1d) (Jinek et al., 2014). Encouraged by this finding, we next applied the above Mg2+-bound α-helical conformation in our optimized pseudoactive Cas9 complex structure (Zuo and Liu, 2017) (Figure 2a–2b) and performed MD-based refinement on the entire structure (see details in Materials and methods). In our final structure, D839, H840 and N863 on the ββα-Me motif formed the catalytic triad that was poised for cleaving the tDNA. The Mg2+ ion was coordinated with the residues D839 and N863 of SpyCas9, a scissile phosphate (the pro-Sp oxygen and leaving O3’ engaged), and two water molecules in a strained geometry (Figure 2c). This structure highly resembled the catalytic configuration present in the X-ray structure of T4 Endo VII/DNA complex (Biertümpfel et al., 2007) (Figure 2c), indicating the N863 of Cas9 could be engaged in the formation of the catalytic center for tDNA cleavage.

To validate the role of N863 in SpyCas9 function, we examined the gene-editing activities of an N863A variant and a D861A/N863A double-mutant variant using cell-based assays. Both SpyCas9 variants virtually lost the entire gene-editing capability compared to the wild-type protein (Figure 1f and Figure 1g). Consistently, our in vitro assays also showed an effective loss of tDNA cleavage activity for the N863A variant, essentially producing only nicked products in the plasmid cleavage assay (Figure 1h) and causing the disappearance of the 23-
Figure 2. Comparison of the active and pseudoactive Cas9-nucleic acid complex structures and proposed mechanism for DNA cleavage activation of Cas9. (a–b) Zoomed-in view (a) and zoomed-out view (b) of the HNH domain docked onto the tDNA and REC lobe in the optimized pseudoactive state. (c) Close-up view of the catalytic configuration of the T4 Endo VII (N62D) ββα-Me motif complexed with a DNA substrate. (d–f) Zoomed-in views (d, f) and zoomed-out view (e) of the HNH domain docked onto the tDNA and REC lobe in the catalytically active state. (g) Schematic diagram of the proposed mechanism underlying Cas9 HNH domain conformational activation.

Figure 2 continued on next page
nucleotide tDNA cleavage product in the oligo assay (Figure 1—figure supplement 3). We note that the DNA nicks produced in the cell-based assay could be efficiently repaired by the cellular machinery, thereby causing the apparent loss of gene-editing capability in the above N863A variants. Together, our results clearly show N863, instead of D861, is the functional catalytic residue. Some studies have also suggested the possibility of N863 in forming the catalytic center of the Cas9 HNH domain (Nishimasu et al., 2014; Yoon et al., 2019). In contrast to the observation from another in silico model that was recently published (Yoon et al., 2019), our computational model indicated that D861 and N863 are unlikely to be simultaneously involved in the coordination of Mg\(^{2+}\) ion for cutting the tDNA. Besides N863, the functional significance of H840 (acting as a general base) has been confirmed by experiments (Jinek et al., 2012; Nishimasu et al., 2014). Another putative active residue in the HNH domain is D839, which could be inferred from the structures of homologous nucleases (Yang, 2008), despite lacking experimental data to support its catalytic role. Here, we discovered that the D839A mutation substantially compromised the gene-editing activity of Cas9 (Figure 1f), which is consistent with our simulation studies (Zuo and Liu, 2017) and directly supports the significance of D839 for Cas9 activity. Collectively, our structural and functional data corroborate the essence of D839, H840 and N863 for the cleavage function of the HNH domain.

Our MD simulations also excluded the likelihood of D837 and N854 (at the two ends of the \(\beta\)-sheet of the \(\beta\)\(\bar{\alpha}\)-Me motif) to function as the catalytic residues previously postulated to be directly engaged in the coordination of the Mg\(^{2+}\) ion (Chen and Doudna, 2017b; Huai et al., 2017). The simulations showed that the side chain of N854 is \(~10\) Å distant from the tDNA, making it unlikely be a residue that directly mediates catalysis. However, an activity assay demonstrated the partial loss of activity in the N854A variant (Nishimasu et al., 2014), suggesting that N854 may play an auxiliary role in catalysis. According to our MD models and the solved Cas9 structures (Figure 1a–1d), N854 may stabilize the activated conformation of the HNH domain by making several intra-molecular hydrogen bonds within the \(\beta\)\(\bar{\alpha}\)-Me motif. Moreover, our structural analyses revealed that the SpyCas9 N854 is spatially and functionally equivalent to the AnaCas9 N597 and Staphylococcus aureus Cas9 (SauCas9) N571 (Figure 1—figure supplement 5), indicating this stabilization role may exist in multiple Cas9 orthologs (Jinek et al., 2014; Nishimasu et al., 2015).

The D837A Cas9 moderately lost gene-editing activity (Figure 1f). From the solved Cas9 structures and our simulations, D837 may contribute to the structural integrity of the \(\beta\)\(\bar{\alpha}\)-Me motif,
although only one hydrogen bond involving its carboxyl group appear to be established (Figure 1a–1d). It is also possible that D837 may aid in proper positioning of the tDNA relative to the HNH domain (as described below).

The comparison between the active and pseudoactive Cas9-nucleic acid complex structures revealed distinct interactions that stabilize these two conformational states. In the pseudoactive conformation, R864 appeared to stabilize the catalytic core by forming ionic and hydrogen-bonding interactions within the ββα-Me motif (Figure 2a), as observed in various bound forms of Cas9 (Figure 1a–1d), whereas K866 is engaged to E371 to form a salt bridge (Figure 2b). In contrast, the corresponding residue of R864 is surrogated by K866 in the active state (Figure 2d), which is in line with the structures of AnaCas9 (K609) and SauCas9 (K583) (Figure 1—figure supplement 5), while R864, on the other hand, is tethered to a negative pocket lined by E370, E396 and D406 on the REC1 domain (Figure 2e). Further structural and sequence analysis indicated that the arginine residue, which is at an equivalent position of R864 in the pseudoactive state, is highly conserved across the ββα-Me nucleases beyond Cas9 (Figure 2—figure supplement 1), whereas both the Type II-A and II-C Cas9 orthologs feature an invariant lysine residue (Figure 2—figure supplement 2) that is spatially equivalent to K866 in the active state. Therefore, a basic amino acid residue (Lys/Arg) seemed to be required at the catalytic core for the ββα-Me superfamily. Moreover, K862 formed aliphatic-aromatic or cation-π interactions with F405 in the pseudoactive state (Figure 2b), whereas it appeared to interact with D837 and/or the +5 phosphate (relative to the protospacer adjacent motif) in one of our simulations for the active state (Figure 2—figure supplement 3). Despite these differences in the interactions with the REC1 domain, other regions of the HNH domain, especially the helical segment preceding the ββα motif, showed similar patterns of binding to the REC2 domain in the pseudoactive and active states (Figure 2f and Figure 2—figure supplement 4).

The removal of the interfacial Mg$^{2+}$ ion from our models substantially destabilized the pseudoactive and active conformations of the Cas9-nucleic acid complex by ~600 kcal/mol (Supplementary file 1), supporting the indispensable role of the Mg$^{2+}$ ion for Cas9 conformational activation that has been previously demonstrated (Dagdas et al., 2017; Osuka et al., 2018; Zuo and Liu, 2017). The pseudoactive conformation (denoted as N863-OUT hereafter) appeared to be unique to SpyCas9 since the corresponding asparagine residues in other ββα-Me nucleases, regardless of the apo or bound form, typically oriented toward the catalytic center (Figure 1—figure supplement 5 and Figure 2—figure supplement 1). End-state free energy calculations revealed that the N863-OUT SpyCas9 complex is thermodynamically more stable than its catalytically active (N863-IN) form (Supplementary file 1). Long time-scale MD simulations based on the isolated HNH domain of Cas9 also showed that the N863-IN conformation is remarkably flexible and tends to transition to the N863-OUT conformation, but not vice versa (Figure 2—figure supplement 5). These results, at least partially, explain why the N863-OUT SpyCas9 conformation was captured in the majority of experimental structures (Anders et al., 2016; Anders et al., 2014; Hirano et al., 2016; Huai et al., 2017; Jiang et al., 2015; Jiang et al., 2016; Jinek et al., 2014; Nishimasu et al., 2014).

To further enhance the confidence in our MD-derived structure models, QM/MM simulations were performed. The metal centers shown in Figure 2a and d were optimized by DFTB3 QM/MM simulations (Figure 2—figure supplement 6). Compared to our MD simulations, we identified very limited differences from the QM/MD simulations in terms of the active site configuration. Encouragingly, this observation is consistent with our most recent benchmark work (Zuo and Liu, 2018a) demonstrating the good performance of the used nonbonded Mg$^{2+}$ ion model (Liao et al., 2017) in maintaining challenging enzyme metal centers.

Based on the findings reported here and in the previous studies (Dagdas et al., 2017; Osuka et al., 2018; Sternberg et al., 2015; Yang et al., 2018), we present a working model to illustrate how the activation of SpyCas9-mediated DNA cleavage is achieved (Figure 2g). The binding of DNA to Cas9-sgRNA induces the interconversion of the HNH domain between the checkpoint intermediate (I) state and the pre-cleavage (P) state. By bridging a gap between the catalytic center and the opposite scissile phosphate, the ambient Mg$^{2+}$ ion subsequently facilitates a conformational change to transition the HNH domain from the P state to the tDNA docked pseudoactive (D, N863-OUT) or active (D’, N863-IN) state. Before the cleavage reaction occurs, both states exist and the conformational population of the two states reaches an equilibrium. The lower-energy pseudoactive state has larger conformational population than the active state. Once the irreversible DNA cleavage...
reaction rapidly occurs, the active state conformational population decreases rapidly. The equilibrium between these two states is lost and the conformational population begins to move from the pseudoactive state to the active state until the reaction is completed. The coexistence of multiple Cas9 ternary complexes has important implications for deciphering the biphasic kinetics of Cas9-catalyzed DNA cleavage that has been shown recently by Raper et al. (2018). Taking their results into considerations, we surmise that the early, fast phase characterizes a direct transition from the precatalytic state to the active state, while the slow transition from the pseudoactive to the active state likely contributes to the late, slow phase. Our conclusion would benefit from structural determination of the activated Cas9 ternary complex with metal ions.

Meanwhile, we note that the possibility cannot be ruled out that the N863-IN conformation may be stably formed when enough Mg\(^{2+}\) ions are present in the solvent. In other words, the N863-IN conformation could be induced by Mg\(^{2+}\) even without engaging a DNA substrate. To test this hypothesis, we have performed additional simulations on the pre-cleavage Cas9 complex (Jiang et al., 2016) with multiple Mg\(^{2+}\) ions positioned around N863, but were unable to observe the trend toward the N863-IN conformation over \(\mu\)s time scales (data not shown). Consistently, the HNH domain in 4UN3 (Figure 1b, right panel) and other crystals (like 5B2T, 5B2S, 5FQ5, 5FW1, 5FW2, 5FW3 and 6Al6) remained in the N863-OUT conformation even though as high as eight Mg\(^{2+}\) ions were cocrystallized (Anders et al., 2016; Hirano et al., 2016; Nishimasu et al., 2014; Olieric et al., 2016; Shin et al., 2017). Therefore, the presence of metal ions alone appears to be insufficient for the HNH domain conformational activation. Nevertheless, further experimental and computational research is needed to elucidate the transition process between the two states shown herein as well as the chemical principles defining the requirement for asparagine instead of aspartic acid.

**Conclusion**

Overall, our study has delineated a molecular framework underlying the catalytic conformation of the HNH nuclease domain of SpyCas9. The findings presented here advance our knowledge of conformational activation that enables Cas9-mediated DNA cleavage, set an important foundation for future studies to further understand the structure-function relationships of Cas9, and facilitate the rational design of Cas9 variants in the future.

### Materials and methods

**Key resources table**

| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|-----------------------------------|-------------|---------------------|-------------|------------------------|
| **Cell line**<br> (H. sapiens)    | HEK-293T    | ATCC                | CRL-3216 (RRID: CVCL_0063) | N/A         |
| Antibody                         | Cas9 (7A9-3A3) mouse monoclonal Ab | Cell Signaling Technology | 14697 (RRID: AB_2750916) | WB (1:500), nonfat milk blocking |
| Antibody                         | Anti-ACTIN mouse monoclonal Ab (C4) | MP Biomedicals | SKU08691002 (RRID: AB_2335127) | WB (1:5000), nonfat milk blocking |
| Transfected construct (Synthetic) | pLenti CMV GFP Puro | Addgene | 17448 | Campeau et al., 2009 Aug 6;4 (8):e6529. |
| Transfected construct (Synthetic) | lentiCRISPR-EGFP sgRNA 1 (WT Cas9) | Addgene | 51760 | Shalem et al., Science 334 (6166):84–7. |
| Transfected construct (Synthetic) | lentiCRISPR-EGFP sgRNA 1 (D861A Cas9) | Site-directed mutagenesis (Wang lab) | N/A | N/A |
| Transfected construct (Synthetic) | lentiCRISPR-EGFP sgRNA 1 (D837A Cas9) | Site-directed mutagenesis (Wang lab) | N/A | N/A |

*Continued on next page*
| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|-----------------------------------|-------------|--------------------|-------------|------------------------|
| Transfected construct (Synthetic) | lentiCRISPR-EGFP sgRNA 1 (D839A Cas9) | Site-directed mutagenesis (Wang lab) | N/A | N/A |
| Transfected construct (Synthetic) | lentiCRISPR-EGFP sgRNA 1 (N863A Cas9) | Site-directed mutagenesis (Wang lab) | N/A | N/A |
| Transfected construct (Synthetic) | lentiCRISPR-EGFP sgRNA 1 (N863A/D861A Cas9) | Site-directed mutagenesis (Wang lab) | N/A | N/A |

**Cell culture**

Human embryonic kidney 293T (HEK293T) cells (ATCC, Manassas, VA) as a subclone of the HEK293 cell line were cultured in DMEM (Thermo Fisher Scientific, Carlsbad, CA) containing 10% fetal bovine serum (FBS; Thermo Fisher Scientific, Carlsbad, CA) at 37°C. All cells were periodically tested using the MycoAlert mycoplasma detection kit (Lonza, Walkersville, MD) and free of mycoplasma. The HEK293T cells were used to established EGFP-expressing cells by the lentivirus-mediated transduction of pLenti-CMV-GFP-Puro expression plasmid (Addgene, Cambridge, MA) into the cells followed by the selection of single-cell clones that stably express EGFP and fluorescence green (Campeau et al., 2009). The stable clone A2 was used in this study.

**CRISPR-Cas9-mediated gene editing**

For testing Cas9-mediated editing of the egfp gene in the HEK293T-EGFP (A2) cells, an EGFP-targeting sgRNA sequence (EGFP sgRNA1: 5’GGGCGAGGAGCTGTTCACCG3’) was cloned into a lentiviral plasmid (Addgene, Cambridge, MA) and resulted in a construct of a one-vector system for co-expression of sgRNA and wild-type SpyCas9 (Addgene, Cambridge, MA) (Shalem et al., 2014). The site-directed mutagenesis was performed to specifically introduce mutations into the cas9 gene open reading frame (ORF) in the expression construct to generate the expression vectors of different Cas9 variants with the EGFP sgRNA sequence. After mutagenesis, the DNA sequencing of each expression construct was performed to confirm the mutations of the Cas9 gene ORF. HEK293T-EGFP (A2) cells transduced with the Cas9 and sgRNA expression constructs were selected using 5 μg/ml puromycin for two weeks prior to the downstream analysis to determine the editing efficiencies of different Cas9 variants.

**Immunoblotting**

The general procedure for immunoblotting was described in previously published reports (Wang et al., 2008; Zolekar et al., 2018). The primary antibody against SpyCas9 (catalog# ab191468) was obtained from Abcam (Cambridge, MA). HRP-conjugated secondary antibodies were from Jackson ImmunoResearch Laboratories (West Grove, PA).

**Flow cytometry analysis**

For determining fluorescence intensity and quantifying the percentages of EGFP fluorescence-positive cells in cell samples, samples (~5 x 10⁵ cells per sample) harvested and resuspended in phosphate-buffered saline (PBS) were analyzed using a SH800Z cell sorter (Sony Biotechnology, San Jose, CA).

**Site directed mutagenesis, and protein overexpression and purification**

The protein variants, SpyCas9<sub>D861A</sub> and SpyCas9<sub>N863A</sub> were produced by sequence independent cloning method (SLIC) using SpyCas9<sup>WT</sup> template plasmid (Addgene: pMJ806) (Jinek et al., 2012) and mutagenic primers (Supplementary file 2) (Scholz et al., 2013). Sequence confirmed clones were transformed into Escherichia coli Rosetta strain 2 (DE3) for protein expression. Overexpression and protein purification were carried out using previously published protocols (Babu et al., 2019; Jinek et al., 2012).
sgRNA and substrate DNA preparation

The template for in vitro transcription of sgRNA (98-nucleotide long) contained a 20 nt long spacer as previously described (Babu et al., 2019; Nishimasu et al., 2014). The protocols used for in vitro transcription and sgRNA annealing were as reported previously (Babu et al., 2019). For creating target DNA, a 30 nt long protospacer flanked by a PAM (GGG) was introduced into pUC19 (Babu et al., 2019).

Plasmid DNA cleavage assay

The proteins were diluted in 20 mM HEPES pH 7.5, 150 mM KCl, and 2 mM TCEP, and the cleavage assays were carried out in a final volume of 10 μL. The reaction mix contained 20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM MgCl2, 5% (v/v) glycerol, 0.5 mM TCEP, 100 ng of substrate plasmid, 50 nM SpyCas9, and 60 nM sgRNA (protein and RNA at a ratio of 1:1.2 molar). The reaction mixture was incubated at 37°C and stopped at different time points (15 s, 30 s, 1 min, 2.5 min, 5 min, 7.5 min, 10 min, 15 min, 30 min, 45 min, and 60 min) by the addition of 50 mM EDTA and 1% SDS. The reaction products were resolved on 1% agarose gel and products were visualized by ethidium bromide staining. The gel was imaged using a BioRad ChemiDoc MP apparatus. To quantify the cleavage activities, each gel image was analyzed using the ImageJ software (Schneider et al., 2012). The bands of nicked (N), linear (L), and supercoiled (SC) DNA were quantified and designated as $I_{N}$, $I_{L}$, and $I_{SC}$ respectively. The nicked, linear and total activity (TA) was calculated using the following equations:

\[
\text{Nicked(\%)} = \frac{I_{N}}{I_{N} + I_{L} + I_{SC}} \times 100
\]

\[
\text{Linear(\%)} = \frac{I_{L}}{I_{N} + I_{L} + I_{SC}} \times 100
\]

\[
\text{TA(\%)} = \frac{I_{N} + I_{L}}{I_{N} + I_{L} + I_{SC}} \times 100
\]

For each reported data point, average values were obtained from a minimum of three replications that were performed using proteins produced from two independent preparations to account for variations in active protein fraction between different preparations. Standard deviation (SD) and standard error of mean (SEM) were calculated based on the number of replications using the following equations:

\[
SD = \sqrt{\frac{\sum(R - R_{AV})^2}{n - 1}}
\]

\[
SEM = \frac{SD}{\sqrt{n}}
\]

Radioactive assay

Two separate oligo DNA strands used for the radioactive assay were ordered from Integrated DNA technology (IDT, Supplementary file 2). Target (T) and non-target (NT) strands were mixed at equimolar concentration in the annealing buffer (30 mM HEPES pH 7.5, 100 mM potassium acetate) and heated at 95°C for 2 min and allowed for slow cooling. The annealed duplex oligo DNA was 5’ end labeled with $^{32}$P (γ-$^{32}$P ATP purchased from PerkinElmer) using T4 polynucleotide kinase (New England Biolabs). The labeled oligo DNA was purified using BioSpin column P-30 (BioRad). The reaction buffer contained 20 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM MgCl2, 5% (v/v) glycerol, 0.5 mM TCEP. Approximately 5 nM of labeled oligo duplex was incubated with 250 nM SpyCas9 and 300 nM sgRNA (protein and RNA at a ratio of 1:1.2 molar) at 37°C and stopped at different time points (15 min, 30 min, and 60 min) using EDTA at 10 mM final concentration. Then the samples were treated with proteinase K (New England Biolabs) for 15 min at 50°C to remove SpyCas9. This was followed by addition of equal volume of loading dye (2X concentration is 20 mM EDTA, 95%
Molecular modeling and molecular dynamics simulations of apo-Cas9
The initial coordinates of the apo-state SpyCas9 were taken from the Protein Data Bank (PDB) under accession number 4CMP (solvated at 2.6 Å resolution) (Jinek et al., 2014). This X-ray structure contains two Cas9 monomers, and the molecule B was considered for modeling here (Figure 1—figure supplement 4a). The disordered regions were built up with the tool SWISS-MODEL (Waterhouse et al., 2018) and the missing heavy atoms and hydrogens were added by using the leap program within AMBER16 (Salomon-Ferrer et al., 2013). The complete structure was then solvated in a cubic water box with a minimal thickness of 13.5 Å from each edge, leading to a periodic boundary box of 138 × 153 ×126 Å³. The system was neutralized by Na⁺, and additional NaCl was added to generate a physiological ionic strength of 150 mM. The resulting simulation box contains ~230,000 atoms.

The above system was simulated by the CUDA-accelerated version of AMBER16 pmemd engine (pmemd.cuda; Salomon-Ferrer et al., 2013) using the amber force field ff14SBonlysc for protein, the TIP3P model for water (Jorgensen et al., 1983), and the Joung-Cheatham parameter sets for monovalent ions (Joung and Cheatham, 2008). The non-bonded interactions were truncated at 10 Å, and the long-range electrostatics were calculated through the particle mesh Ewald (PME) summation method (Darden et al., 1993), with a grid spacing of 1 Å. The bonds involving hydrogens were constrained via the SHAKE algorithm (Miyamoto and Kollman, 1992), allowing use of 2-fs time step of integration. After thorough energy minimization, the system underwent slow heating over 50 ps from 0 K to the target 310.15 K in the isothermal-isochoric (NVT) ensemble, followed by a 20-ns equilibration in the isothermal-isobaric (NPT) condition. The protein backbone atoms were restrained in the heating and equilibration stages. Finally, the production run was performed under the NPT ensemble without restraints, extending up to 100 ns. The pressure was controlled at 1.013 bar via the Monte Carlo barostat, and the temperature was maintained at 310.15 K through the Langevin thermostat implemented in AMBER16.

The final structural snapshot from the above simulation was then extracted, and a Mg²⁺ ion was placed at its HNH domain active center to set up the Mg²⁺-bound system (Figure 1—figure supplement 4) by reference to the AnaCas9 crystal structure bound with a Mg²⁺ ion (Figure 1—figure supplement 5a) (Jinek et al., 2014). Also, extra Mg²⁺ ions were introduced into the system to obtain a physiological concentration of 5 mM. The parameter set developed by Li et al. (Li et al., 2013b) was selected for Mg²⁺. In the equilibration stage, the distances between the Mg²⁺ and the coordinating oxygens on D839 and N863 was restrained to 2.1 Å (i.e., the experimental ion-oxygen distance; Zheng et al., 2008). The production run without restraints was extended to 50 ns.

Molecular dynamics refinement of the pseudoactive Cas9 complex
The starting structure of the pseudoactive Cas9 complex (Figure 1d, right panel) was obtained from our recent work (Zuo and Liu, 2017), which was derived by employing the unbiased, brute-force MD simulations on the crystal structure of Cas9-sgRNA-DNA (PDB code: 5F9R) that was captured in the pre-cleavage state (Jiang et al., 2016). The structural model has been validated by different experiments, yet the Mg²⁺ ion at the HNH domain catalytic center appeared to lose one critical coordination bond with the leaving group O3’ of the scissile phosphate (Zuo and Liu, 2017) (Figure 2—figure supplement 7), as compared to the homologous T4 Endo VII structure complexed with a DNA junction (Biertu¨ mpfel et al., 2007) (Figure 2c). We reasoned that this issue may be due to the deficiency with the simple point-charge Mg²⁺ model used. Most recently, we systematically evaluated the performance of all four types of non-bonded Mg²⁺ ion models in terms of maintaining a challenging metal center configuration in a nuclease system (Nowotny et al., 2005). Our benchmark calculations demonstrated that the multisite models based a 12-6-4 Lennard-Jones potential (Li and Merz, 2014; Liao et al., 2017), which take charge-induced dipole effects into account, are the only ones that are capable of reproducing the experimental coordination patterns (Zuo and Liu, 2018a). Accordingly, the 12-6-4 type multisite model (Jorgensen et al., 1983) (here the midC4 set) was
considered for the Cas9 complex simulation, along with the TIP4PEw model for water, the Joung-Cheatham parameter sets for monovalent ions (Joung and Cheatham, 2008), and the amber force fields ff14SBonlysc, ff99bsc0_chiol3, ff99bsc0_OL15 for protein, RNA and DNA, respectively. Basically, the complex system was set up and simulated following the above protocol for the apo-Cas9 systems. The generated simulation box is approximately $109 \times 145 \times 166$ Å$^3$ sized, containing ~282,000 atoms. With different random seed numbers, two parallel simulations were carried out by using the latest AMBER18 (Salomon-Ferrer et al., 2013) that enables GPU calculations of the 12-6-4 ion potential. The simulation length was set to 200 ns each.

**Molecular modelling and molecular dynamics simulations of the active Cas9 complex**

The initial model for the active Cas9 complex was constructed by replacing the $\alpha$-helical segment of the $\beta$$\beta$xt-Me motif in the optimized pseudoactive Cas9 complex (Figure 1a) with the corresponding part in the Mg$^{2+}$-bound apo-Cas9 structure (Figure 1—figure supplement 4c). The pseudoactive Cas9 complex structure was taken from the above production simulation near 100 ns (i.e. about half of the simulation time). The Mg$^{2+}$-bound apo-Cas9 structure from the simulation trajectory was selected based on the observation of reasonable bonding with the connecting residues and minimal steric clashes after replacement of the $\alpha$-helical segment. After sufficient energy minimization, the structural model was subjected to multi-stage equilibration: an initial 20-ns relaxation of the $\alpha$-helical segment and surrounding residues, an another 20-ns equilibration with the inter-atomic distances within the metal center retrained relative to the T4 Endo VII system (Biertümpfel et al., 2007), followed by an additional 20-ns equilibration with the restraints gradually released. Subsequently, two independent replicas were performed (250 ns/run) under the same simulation conditions set for the pseudoactive system above.

**Molecular dynamics simulations of isolated Cas9 HNH domain**

Additional MD simulations were performed to investigate the relative stability of the two conformational states (i.e. N863-IN and N863-OUT) of the $\alpha$ structure element containing N863 and D861. The starting coordinates were taken from the respective structure models above, and only the HNH domain of Cas9 (residues 781 to 905) was included in our simulation to enhance the sampling efficiency. Each isolated HNH domain in the two states was immersed in a truncated octahedral water box, with a minimal thickness of 14.5 Å. The ionic centration was set to 100 mM by adding an appropriate number of K$^+$ and Cl$^-$ ions in the aqueous solution. The amber force field ff14SBonlysc and the TIP3P model were used for describing the protein and the water molecules, respectively, and the parameter sets for the monovalent ion were derived from the work by Joung and Cheatham (2008). For each system, five independent simulations were performed under the NPT ensemble with different initial velocities, using a timestep of 2-ps. Each replica was extended to ~10 us, yielding a total of ~50 us of sampling for each system.

**Hybrid quantum mechanics/molecular mechanics simulations**

The semiempirical DFTB3 QM/MM simulations were further implemented to improve the reliability of our MD-derived structure models. DFTB3 is the third-order variant of density functional theory tight binding (DFTB) that is formulated in a DFT framework (Gaus et al., 2012). According to the extensive benchmark calculations, DFTB3 in its current form is most reliably for structural properties, including for fairly complex bimetallic motifs in diverse metalloenzymes (particularly the phosphoryl-transfer enzymes) (Gaus et al., 2012; Lu et al., 2015; Lu et al., 2016; Roston et al., 2018). The QM region includes the catalytic Mg$^{2+}$ ion, the protein residues that coordinate the metal ion (i.e., D839 and D861/N863), the general base H840, part of the scissile phosphate and nearby atoms on the target strand, and the water molecules surrounding the metal ions, H840, and the scissile phosphate (cf. Figure 2a and d). Only the side chains of protein and the backbone of DNA are kept in the QM region, and link atoms are added between the Ca and Cβ atoms for the amino acids or between the C4′ and C5′ atoms for the nucleotides. The partitioning results in a total of 72 and 75 QM atoms for the pseudoactive and active Cas9 models, respectively. The dummy complex for the Mg$^{2+}$ ions employed in pure MD simulations is changed back to the realistic single-atom form. The MM part of the protein and nucleic acids are described using the same AMBER force fields as mentioned above,
and the water molecules are described with the TIP3P model. After the stages of energy minimization and slow heating, each system was subjected to two parallel 1,000-ps QM/MM simulations performed with the AMBER program.

**Computation of conformational free energies**

The free energies of the Cas9-nucleic acid complex conformers were estimated through the endpoint Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) approach (Miller et al., 2012). Compared to the alternative Molecular mechanics-Poisson Boltzmann Surface Area (MM-PBSA), MM-GBSA has been proven to be give comparable or even better accuracy in ranking ligand binding affinities as well as in calculating the relative stability of multiple conformations of a biomolecular system, though MM-PBSA is physically more rigorous (Li et al., 2013a; Zuo and Liu, 2016a; Zuo et al., 2018b). For each state, the MM-GBSA calculations were performed over an ensemble of 2000 snapshots extracted from the last 50 ns of the simulation trajectories using the program MMPBSA.py in AmberTools16. The pairwise GB model of Hawkins, Cramer, and Truhlar (GB\textsuperscript{HCT}) (Hawkins et al., 1995; Hawkins et al., 1996) was used, with the parameters described by Tsui and Case (2000). The default values of the surface tension and the offset to correct the non-polar contribution to the solvation free energy were adopted and the salt concentration was set to 150 mM. Following our recent works (Zuo and Liu, 2016a; Zuo et al., 2018b), the two water molecules closest to the Mg\textsuperscript{2+} ion at the HNH domain active center were retained as part of the complex, considering the importance of the interfacial water for binding. The entropic contribution was not taken into account due to high computational demand and potential convergence problem, yet omission of this term does not qualitatively affect the results as previously suggested (Hou et al., 2011; Li et al., 2013a; Zuo et al., 2018b; Zuo et al., 2016b).

**Acknowledgements**

We appreciate the assistance from Dr. Qinghua Liao (Uppsala University) in preparing the Mg\textsuperscript{2+} ion model for MD simulations and thank the Texas Advanced Computing Center (TACC) at the University of Texas at Austin and the University of North Texas (UNT) High Performance Computing Services (a division of the University Information Technology with additional support from UNT Office of Research and Economic Development) for providing us with computational powers that our study has leveraged.

**Additional information**

**Funding**

| Funder                            | Grant reference number | Author          |
|-----------------------------------|------------------------|-----------------|
| Shanghai Municipal Education      | Program for Professor  | Zhicheng Zuo    |
| Commission                        | of Special Appointment |
|                                   | at Shanghai Institutions|
|                                   | of Higher Learning     |
| National Science Foundation       | MCB-1716423            | Rakhi Rajan     |
| National Institute of General     | P20GM103640            | Rakhi Rajan     |
| Medical Sciences                  |                        |                 |
| University of North Texas (UNT)   | Start-up Fund and      | Yu-Chieh Wang   |
| Health Science Center             | Faculty Pilot Grant    |                 |
| University of North Texas (UNT)   | Start-up Fund and Basic| Jin Liu         |
| Health Science Center             | Research Seed Grant    |                 |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

**Author contributions**

Zhicheng Zuo, Conceptualization, Data curation, Formal analysis, Validation, Investigation, Visualization, Writing—original draft, Writing—review and editing; Ashwini Zolekar, Victor JT Lin,
Formal analysis, Validation, Investigation; Kesavan Babu, Formal analysis, Validation, Investigation, Writing—review and editing; Hamed S Hayatshahi, Formal analysis; Rakhi Rajan, Formal analysis, Validation, Investigation, Visualization, Writing—review and editing; Yu-Chieh Wang, Conceptualization, Formal analysis, Validation, Investigation, Visualization, Writing—original draft, Writing—review and editing; Jin Liu, Conceptualization, Formal analysis, Supervision, Validation, Investigation, Visualization, Writing—original draft, Project administration, Writing—review and editing

Author ORCIDs
Zhicheng Zuo https://orcid.org/0000-0003-2761-5841
Hamed S Hayatshahi https://orcid.org/0000-0001-8639-8130
Rakhi Rajan https://orcid.org/0000-0002-8719-4412
Yu-Chieh Wang https://orcid.org/0000-0002-7445-4561
Jin Liu https://orcid.org/0000-0002-1067-4063

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.46500.023
Author response https://doi.org/10.7554/eLife.46500.024

Additional files
Supplementary files
• Supplementary file 1. Free energies of Cas9 ternary complexes.
  DOI: https://doi.org/10.7554/eLife.46500.019
• Supplementary file 2. Primers used to construct Cas9 variants.
  DOI: https://doi.org/10.7554/eLife.46500.020
• Transparent reporting form
  DOI: https://doi.org/10.7554/eLife.46500.021

Data availability
The data generated or analysed during this study are included in the manuscript and supporting files. Source data files have been provided for Figures 1 and 2.

References
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. DOI: https://doi.org/10.1038/nature13579, PMID: 25079318
Anders C, Bargsten K, Jinek M. 2016. Structural plasticity of PAM recognition by engineered variants of the RNA-Guided endonuclease Cas9. Molecular Cell 61:895–902. DOI: https://doi.org/10.1016/j.molcel.2016.02.020, PMID: 26990992
Babu K, Amrani N, Jiang W, Yogesha SD, Nguyen R, Qin PZ, Rajan R. 2019. Bridge Helix of Cas9 modulates target DNA cleavage and mismatch tolerance. Biochemistry 58:1905–1917. DOI: https://doi.org/10.1021/acs.biochem.8b01241, PMID: 30916546
Biertümpfel C, Yang W, Suck D. 2007. Crystal structure of T4 endonuclease VII resolving a Holliday junction. Nature 449:616–620. DOI: https://doi.org/10.1038/nature06152, PMID: 17873859
Campeau E, Ruhl VE, Rodier F, Smith CL, Rahmberg BL, Fuss JO, Campisi J, Yaswen P, Cooper PK, Kaufman PD. 2009. A versatile viral system for expression and depletion of proteins in mammalian cells. PLOS ONE 4:e6529. DOI: https://doi.org/10.1371/journal.pone.0006529, PMID: 19657394
Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA. 2017a. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550:407–410. DOI: https://doi.org/10.1038/nature24268, PMID: 28931002
Chen JS, Doudna JA. 2017b. The chemistry of Cas9 and its CRISPR colleagues. Nature Reviews Chemistry 1:0078. DOI: https://doi.org/10.1038/s41570-017-0078
Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823. DOI: https://doi.org/10.1126/science.1231143, PMID: 23287718
Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A. 2017. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Science Advances 3:eaa0027. DOI: https://doi.org/10.1126/sciadv.aao0027, PMID: 28808686
Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98:10089–10092.

Dong D, Guo M, Wang S, Zhu Y, Wang S, Xiong Z, Yang J, Xu Z, Huang Z. 2017. Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature 546:436–439. DOI: https://doi.org/10.1038/nature22377, PMID: 28440866

Gasiorunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. PNAS 109:E2579–E2586. DOI: https://doi.org/10.1073/pnas.1208507109, PMID: 22949671

Gaus M, Cui Q, Elstner M. 2012. DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). Journal of Chemical Theory and Computation 7:931–948. DOI: https://doi.org/10.1021/ct1006845, PMID: 23204947

Hawkins GD, Cramer CJ, Truhlar DG. 1995. Pairwise solute descreening of solute charges from a dielectric medium. Chemical Physics Letters 246:122–129. DOI: https://doi.org/10.1016/0009-2614(95)01082-K

Hawkins GD, Cramer CJ, Truhlar DG. 1996. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. The Journal of Physical Chemistry 100:19824–19839. DOI: https://doi.org/10.1021/jp961710n

Hirano S, Nishimatsu H, Ishitani R, Nureki O. 2016. Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. Molecular Cell 61:886–894. DOI: https://doi.org/10.1016/j.molcel.2016.02.018, PMID: 26990991

Hou T, Wang J, Li Y, Wang W. 2011. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations. Journal of Chemical Information and Modeling 51:69–82. DOI: https://doi.org/10.1021/ci100275a, PMID: 21117705

Hui C, Li G, Yao R, Zhang Y, Cao M, Kong L, Jia C, Yuan H, Chen H, Lu D, Huang Q, Dr L. 2017. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. Nature Communications 8:1375. DOI: https://doi.org/10.1038/s41467-017-01496-2, PMID: 29123204

Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. 2015. A Cas9-guide RNA complex preorganized for target DNA recognition. Science 348:1477–1481. DOI: https://doi.org/10.1126/science.aab1452, PMID: 26113724

Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science 351:867–871. DOI: https://doi.org/10.1126/science.aad8282, PMID: 26841432

Jiang F, Liu J-J, Osuna BA, Xu M, Berry JD, Rauch BJ, Nogales E, Bondy-Denomy J, Doudna JA. 2019. Temperature-Responsive competitive inhibition of CRISPR-Cas9. Molecular Cell 73:601–610. DOI: https://doi.org/10.1016/j.molcel.2018.11.016

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821. DOI: https://doi.org/10.1126/science.1225429

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 343:1247997. DOI: https://doi.org/10.1126/science.1247997, PMID: 24505130

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 79:926–935. DOI: https://doi.org/10.1063/1.445869

Joung IS, Cheatham TE. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. The Journal of Physical Chemistry B 112:9020–9041. DOI: https://doi.org/10.1021/jp0800164

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529:490–495. DOI: https://doi.org/10.1038/nature16526, PMID: 26735016

Knott GJ, Doudna JA. 2018. CRISPR-Cas guides the future of genetic engineering. Science 361:866–869. DOI: https://doi.org/10.1126/science.aat5011, PMID: 30166482

Li CH, Zuo ZC, Su JG, Xu XJ, Wang CX. 2013a. The interactions and recognition of cyclic peptide mimetics of tat with HIV-1 TAR RNA: a molecular dynamics simulation study. Journal of Biomolecular Structure & Dynamics 31:276–287. DOI: https://doi.org/10.1080/07391102.2012.698248, PMID: 22943434

Li P, Roberts BP, Chakravorty DK, Merz KM. 2013b. Rational design of particle mesh ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. Journal of Chemical Theory and Computation 9:2733–2748. DOI: https://doi.org/10.1021/ct400146w, PMID: 23914143

Li P, Merz KM. 2014. Taking into account the Ion-induced dipole interaction in the nonbonded model of ions. Journal of Chemical Theory and Computation 10:289–297. DOI: https://doi.org/10.1021/ct400751u, PMID: 24659926

Liao Q, Pabis A, Strolobel, Kamerlin SCL. 2017. Extending the nonbonded cationic dummy model to account for ion-Induced dipole interactions. The Journal of Physical Chemistry Letters 8:5408–5414. DOI: https://doi.org/10.1021/acs.jpclett.7b02358, PMID: 29022713

Liu L, Yin M, Wang M, Wang Y. 2019. Phage AcrIIA2 DNA mimicry: structural basis of the CRISPR and Anti-CRISPR arms race. Molecular Cell 73:611–620. DOI: https://doi.org/10.1016/j.molcel.2018.11.011

Lu X, Gaus M, Elstner M, Cui Q. 2015. Parametrization of DFTB3/3OB for magnesium and zinc for chemical and biological applications. The Journal of Physical Chemistry B 119:1062–1082. DOI: https://doi.org/10.1021/acs.jpc.1506557, PMID: 25178644
Lu X, Fang D, Ito S, Okamoto Y, Ovchinnikov V, Cui Q. 2016. QM/MM free energy simulations: recent progress and challenges. *Molecular Simulation* **42**:1056–1078. DOI: https://doi.org/10.1080/08927022.2015.1132317, PMID: 27563170

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**:823–826. DOI: https://doi.org/10.1126/science.1232033, PMID: 23287722

Miller BR, McGee TD, Swails JM, Homersey N, Gohlke H, Roitberg AE. 2012. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8**:3314–3321. DOI: https://doi.org/10.1021/ct300418h, PMID: 26605738

Miyamoto S, Kollman PA. 1992. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**:952–962. DOI: https://doi.org/10.1002/jcc.540130805

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmnae I, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**:935–949. DOI: https://doi.org/10.1016/j.cell.2014.02.001, PMID: 24529477

Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O. 2015. Crystal Structure of Staphylococcus aureus Cas9. *Cell* **162**:1113–1126. DOI: https://doi.org/10.1016/j.cell.2015.08.007, PMID: 26317473

Nowotny M, Gaidamakov SA, Crouch RJ, Yang W. 2005. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**:1005–1016. DOI: https://doi.org/10.1016/j.cell.2005.04.024, PMID: 15989951

Olieric V, Weinert T, Finke AD, Anders C, Li D, Olieric N, Borca CN, Steinmetz MO, Caffrey M, Jinek M, Wang M. 2016. Data-collection strategy for challenging native SAD phasing. *Acta Crystallographica Section D Structural Biology* **72**:421–429. DOI: https://doi.org/10.1107/S2059798315024110

Osuka S, Isomura K, Kaijimoto S, Komori T, Nishimasu H, Shima T, Nureki O, Uemura S. 2018. Real-time observation of flexible domain movements in CRISPR-Cas9. *The EMBO Journal* **37**:e96941. DOI: https://doi.org/10.15252/embj.201796941

Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. 2017. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *PNAS* **114**:7260–7265. DOI: https://doi.org/10.1073/pnas.1706451114, PMID: 28652374

Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, Doudna JA, McCammon JA. 2018. Key role of the REC lobe during CRISPR-Cas9 activation by ‘sensing’, ‘regulating’, and ‘locking’ the catalytic HNH domain. *Quarterly Reviews of Biophysics* **51**:e9. DOI: https://doi.org/10.1017/S0301385518000070

Raper AT, Stephenson AA, Suo Z. 2018. Functional insights revealed by the kinetic mechanism of CRISPR/Cas9. *Journal of the American Chemical Society* **140**:2971–2984. DOI: https://doi.org/10.1021/jacs.7b13047, PMID: 29942507

Roston D, Lu X, Fang D, Demapan D, Cui Q. 2018. Analysis of Phosphoryl-Transfer enzymes with QM/MM free energy simulations. *Methods in Enzymology* **607**:53–90. DOI: https://doi.org/10.1016/bs.mie.2018.05.005, PMID: 30149869

Salomon-Ferrer R, Case DA, Walker RC. 2013. An overview of the amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3**:198–210. DOI: https://doi.org/10.1002/wcms.1121

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH image to ImageJ: 25 years of image analysis. *Nature Methods* **9**:671–675. DOI: https://doi.org/10.1038/nmeth.2089, PMID: 22930834

Scholz J, Besir H, Strasser C, Suppmann S. 2013. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. *Biopolymers* **56**:53–90. DOI: https://doi.org/10.1002/bip.21060, PMID: 23918188

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**:84–87. DOI: https://doi.org/10.1126/science.1247005, PMID: 24336571

Shin J, Jiang F, Liu JJ, Bray NL, Rauch BJ, Baik SH, Nogales E, Bondy-Denomy J, Corn JE, Doudna JA. 2017. Disabling Cas9 by an anti-CRISPR DNA mimic. *Science Advances* **3**:e1600260. DOI: https://doi.org/10.1126/sciadv.1600260, PMID: 28706995

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**:84–88. DOI: https://doi.org/10.1126/science.aad5227, PMID: 26628643

Sternberg SH, LaFrance B, Kaplan M, Doudna JA. 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**:110–113. DOI: https://doi.org/10.1038/nature15544, PMID: 26524520

Tsui V, Case DA. 2000. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers* **56**:275–291. DOI: https://doi.org/10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E, PMID: 11754341

Wang YC, Kulp SK, Wang D, Yang CC, Sargeant AM, Hung JH, Kashida Y, Yamaguchi M, Chang GD, Chen CS. 2008. Targeting endoplasmic reticulum stress and akt with OSU-03012 and gefitinib or erlotinib to overcome resistance to epidermal growth factor receptor inhibitors. *Cancer Research* **68**:2820–2830. DOI: https://doi.org/10.1158/0008-5472.CAN-07-1336, PMID: 18431370

Waterhouse A, Bertoni M, Henriet S, Studer G, Tauriello G, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**:823–826. DOI: https://doi.org/10.1126/science.1232033, PMID: 23287722

Waterhouse A, Bertoni M, Henriet S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Beldorf L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**:W296–W303. DOI: https://doi.org/10.1093/nar/gky427
Yang W. 2008. An equivalent metal ion in one- and two-metal-ion catalysis. Nature Structural & Molecular Biology 15:1228–1231. DOI: https://doi.org/10.1038/nsmb.1502
Yang W. 2011. Nucleases: diversity of structure, function and mechanism. Quarterly Reviews of Biophysics 44:1–93. DOI: https://doi.org/10.1017/S0033583510000181, PMID: 20854710
Yang M, Peng S, Sun R, Lin J, Wang N, Chen C. 2018. The conformational dynamics of Cas9 governing DNA cleavage are revealed by Single-Molecule FRET. Cell Reports 22:372–382. DOI: https://doi.org/10.1016/j.celrep.2017.12.048, PMID: 29320734
Yang H, Patel DJ. 2017. Inhibition mechanism of an Anti-CRISPR suppressor AcrIIA4 targeting SpyCas9. Molecular Cell 67:117–127. DOI: https://doi.org/10.1016/j.molcel.2017.05.024
Yoon H, Zhao LN, Warshel A. 2019. Exploring the catalytic mechanism of Cas9 using information inferred from endonuclease VII. ACS Catalysis 9:1329–1336. DOI: https://doi.org/10.1021/acscatal.8b04324
Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W. 2008. Data mining of metal ion environments present in protein structures. Journal of Inorganic Biochemistry 102:1765–1776. DOI: https://doi.org/10.1016/j.jinorgbio.2008.05.006, PMID: 18614239
Zolekar A, Lin VJT, Mishra NM, Ho YY, Hayatshahi HS, Parab A, Sampat R, Liao X, Hoffmann P, Liu J, Emmite KA, Wang YC. 2018. Stress and interferon signalling-mediated apoptosis contributes to pleiotropic anticancer responses induced by targeting NGLY1. British Journal of Cancer 119:1538–1551. DOI: https://doi.org/10.1038/s41416-018-0265-9, PMID: 30385822
Zuo Z, Weng J, Wang W. 2016b. Insights into the inhibitory mechanism of D13-9001 to the multidrug transporter AcrB through molecular dynamics simulations. The Journal of Physical Chemistry B 120:2145–2154. DOI: https://doi.org/10.1021/acs.jpcb.5b11942
Zuo Z, Smith RN, Chen Z, Agharkar AS, Snell HD, Huang R, Liu J, Gonzales EB. 2018b. Identification of a unique Ca2+-binding site in rat acid-sensing ion channel 3. Nature Communications 9:2082. DOI: https://doi.org/10.1038/s41467-018-04424-0, PMID: 29802295
Zuo Z, Liu J. 2016a. Cas9-catalyzed DNA cleavage generates staggered ends: evidence from molecular dynamics simulations. Scientific Reports 5:37584. DOI: https://doi.org/10.1038/srep37584, PMID: 27874072
Zuo Z, Liu J. 2017. Structure and dynamics of Cas9 HNH domain catalytic state. Scientific Reports 7:17271. DOI: https://doi.org/10.1038/s41598-017-17578-6, PMID: 29222528
Zuo Z, Liu J. 2018a. Assessing the performance of the nonbonded Mg2+ Models in a Two-Metal-Dependent Ribonuclease. Journal of Chemical Information and Modeling 59:399–408. DOI: https://doi.org/10.1021/acs.jcim.8b00627, PMID: 30521334

Zuo et al. eLife 2019;8:e46500. DOI: https://doi.org/10.7554/eLife.46500

17 of 17