



# Dynamics and mechanisms of CRISPR-Cas9 through the lens of computational methods

Aakash Saha<sup>1,a</sup>, Pablo R. Arantes<sup>1,a</sup> and Giulia Palermo<sup>1,2</sup>

## Abstract

The clustered regularly interspaced short palindromic repeat (CRISPR) genome-editing revolution established the beginning of a new era in life sciences. Here, we review the role of state-of-the-art computations in the CRISPR-Cas9 revolution, from the early refinement of cryo-EM data to enhanced simulations of large-scale conformational transitions. Molecular simulations reported a mechanism for RNA binding and the formation of a catalytically competent Cas9 enzyme, in agreement with subsequent structural studies. Inspired by single-molecule experiments, molecular dynamics offered a rationale for the onset of off-target effects, while graph theory unveiled the allosteric regulation. Finally, the use of a mixed quantum-classical approach established the catalytic mechanism of DNA cleavage. Overall, molecular simulations have been instrumental in understanding the dynamics and mechanism of CRISPR-Cas9, contributing to understanding function, catalysis, allostery, and specificity.

## Addresses

<sup>1</sup> Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA, 92512, United States

<sup>2</sup> Department of Chemistry, University of California Riverside, 900 University Avenue, Riverside, CA, 92512, United States

Corresponding author: Palermo, Giulia ([giulia.palermo@ucr.edu](mailto:giulia.palermo@ucr.edu))

 (Saha A.)

 (Arantes P.R.)

 (Palermo G.)

<sup>a</sup> These authors contributed equally to this work.

**Current Opinion in Structural Biology** 2022, 75:102400

This review comes from a themed issue on **Macromolecular Assemblies**

Edited by **Alan Brown** and **Franca Fraternali**

For complete overview of the section, please refer the article collection - **Macromolecular Assemblies**

Available online 8 June 2022

<https://doi.org/10.1016/j.sbi.2022.102400>

0959-440X/© 2022 Elsevier Ltd. All rights reserved.

## Introduction

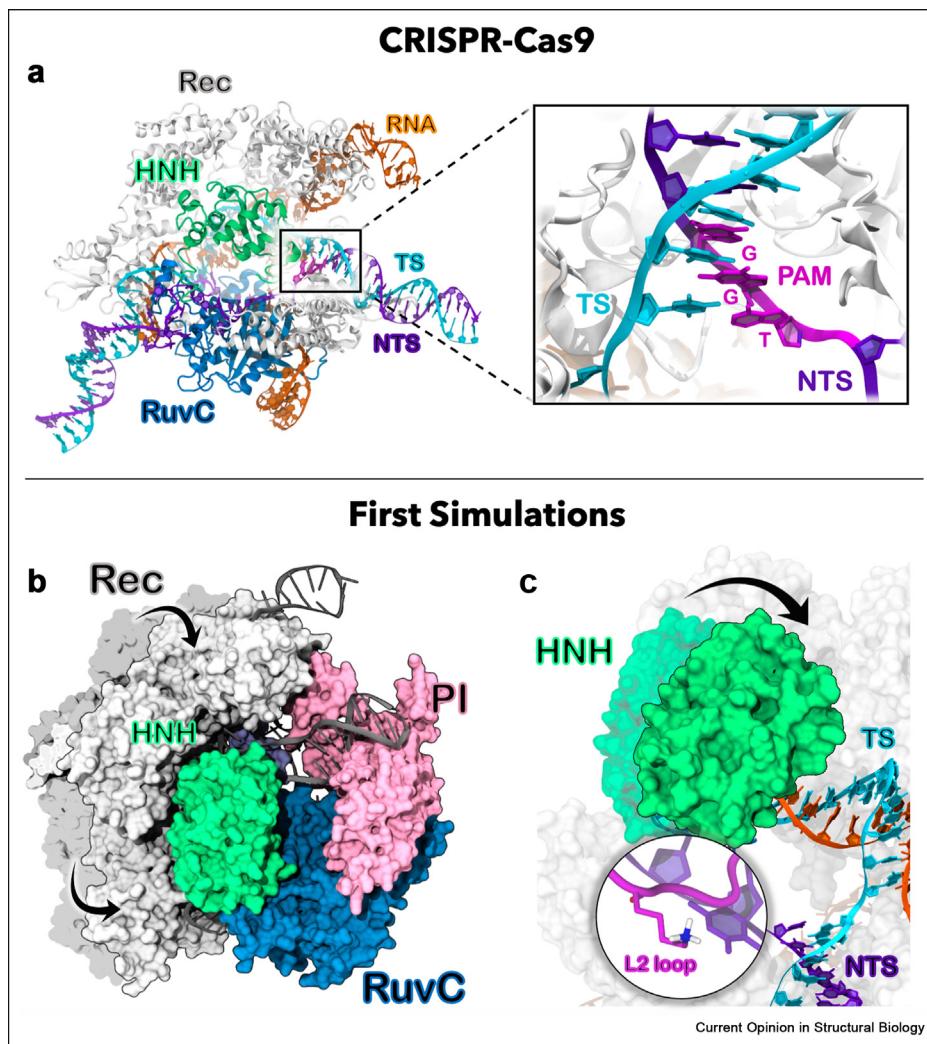
In 2012, the precise manipulation of nucleic acids became a reality with the introduction of the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 technology [1]. Since then, ceaseless developments across basic and applied sciences have

made CRISPR-Cas9 the 'gold standard' for genome-editing applications, driving innovations across life sciences and beyond. The applicability of this technique also permeated into a myriad of different other fields, including plant biotechnology, biofuel production, and many more, establishing CRISPR-Cas9 as the 'genetic scissor' that tailors nature's gene pool [2].

CRISPR-Cas9 is an integral part of the bacterial adaptive immune system that confers protection against invading viruses. In this system, foreign sequences of viral DNA are incorporated into the bacterial genome for the degradation of invading DNA by the Cas9 endonuclease. Cas9 is an RNA-guided enzyme that leverages the sequence specificity of the guide RNA to bind and cleave complementary DNA sequences (Figure 1a) [1]. Upon recognition of a short protospacer adjacent motif (PAM), Cas9 binds the DNA by matching the guide RNA with one DNA strand (target strand, TS), while the other strand ("non-target" strand, NTS) is displaced [3]. The discovery that Cas9 can be reprogrammed with a single-guide RNA to recognize any DNA sequence next to PAM paved the way for advanced genome editing.

Structural studies of the *Streptococcus pyogenes* Cas9 (SpCas9) disclosed snapshots of the Cas9 conformational landscape [4], while biophysical approaches harnessing single-molecule spectroscopy revealed its large-scale dynamics [5–9]. In this scenario, molecular dynamics (MD) simulations have been digging into the mechanistic function, investigating the conformational transitions of the Cas9 protein and its interplay with nucleic acids. All-atom MD simulations exploited the power of modern exascale architectures and advanced GPU cards to synergistically apply a variety of theoretical approaches. This included classical and novel accelerated MD methods, which extend the spatio-temporal limits of molecular simulations, mixed quantum/classical (QM/MM) approaches, cryo-EM refinement techniques, as well as graph theory-derived analysis methods. This has created a multiscale framework for the investigation of the biophysics of CRISPR-Cas9, contributing to the clarification of function, catalysis, allostery, and specificity. Here, we showcase how this multiscale computational approach contributed to the dynamics and mechanism of the CRISPR function.

Figure 1



**Top panel:** (a) Overview of the *Streptococcus pyogenes* CRISPR-Cas9 system, with a close-up view of the PAM recognition region [10]. The Cas9 protein is shown as ribbons, highlighting the recognition lobe (Rec, gray), and the HNH (green), and RuvC (blue) nuclease domains. The RNA (orange) base pairs the DNA target strand (TS, cyan), while the DNA "non-target" strand (NTS, violet) is displaced. **Bottom panel:** Outcomes of the first molecular dynamics simulations [11]. (b) Large-scale motions of the Rec lobe from an open-to-closed conformational state are indicated using two arrows on the Cas9 molecular surface. (c) Conformational change of the HNH domain toward the DNA TS (shown using an arrow), aided by interactions of the L2 loop with the DNA NTS.

### First all-atom molecular simulations

The very first all-atom simulations of CRISPR-Cas9 focused on the conformational dynamics of Cas9 as an apoprotein and bound to RNA and DNA [11]. The essential dynamics of the protein revealed the tendency of large-scale motions of the recognition lobe (Rec) for the accommodation of the nucleic acids (Figure 1b). Intriguingly, the catalytic HNH domain exhibited a 'striking plasticity', suggesting fast conformational transitions for nucleic acid cleavages. The HNH conformational activation toward the DNA TS was observed to be highly dependent on the binding of the NTS within the RuvC groove. Indeed, the NTS was shown to establish a

number of interactions with the L2 loop of HNH, facilitating the docking of this domain at the site of cleavage (Figure 1c). This key role of the NTS was later confirmed by single-molecule Förster Resonance Energy Transfer (smFRET) experiments, reporting that the presence of the NTS within the RuvC core is critical for the docking of HNH at the TS for cleavage [5]. It is worth noting that, on account of the initial unavailability of high-resolution structural models, this early study had to rely on difficult interpretation and refinement of low-resolution ( $\sim 20$  Å) negative stain EM structures [12]. Nevertheless, its findings were remarkably corroborated by experimental studies [5], thus cementing all-atom

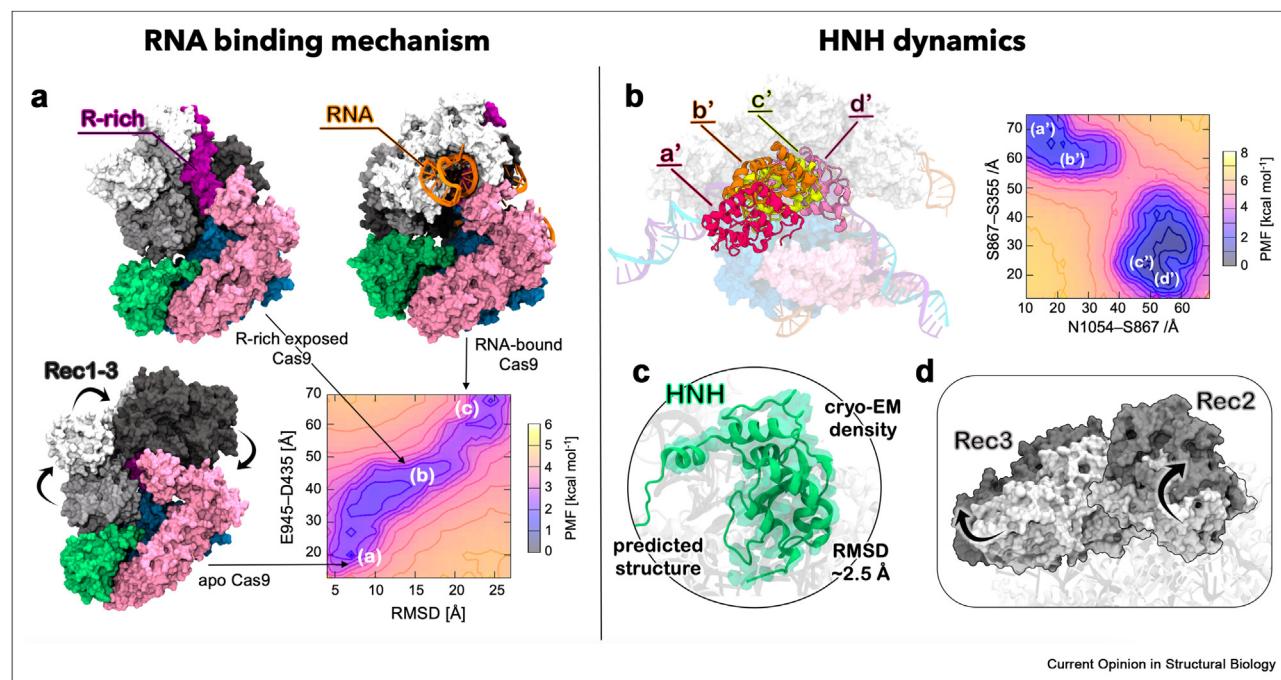
MD simulations to produce reliable predictive models of the CRISPR-Cas9 function.

## Conformational changes underlying RNA binding

The structure of the Cas9 endonuclease is composed of two main lobes. The recognition (Rec) lobe accommodates nucleic acids, while the nuclease lobe performs DNA cleavages through the catalytic HNH and RuvC domains (Figure 1a) [4]. Structural investigations of Cas9 indicate remarkable conformational changes upon RNA and DNA binding [13,14]. For investigating the large-scale transition of Cas9 from its apo form to the binary RNA-bound state, biased MD simulations were combined with a Gaussian accelerated MD (GaMD) method [15]. Specifically, biased simulations were used to recover the conformational change from the apo to RNA-bound form. Then,  $\sim 20$   $\mu$ s of unconstrained GaMD simulations were applied to the obtained pathway to gain an enhanced sampling of the

intermediate states. This enabled exploring large-scale conformational motions typically occurring over longer timescales (i.e.,  $\mu$ s to ms) while also reconstructing the associated free energy landscape. The simulations revealed a remarkable conformational change of the Rec lobe, involving oppositely directed movements of the Rec1-3 domains to allocate nucleic acids (Figure 2a) [16]. A critical intermediate state was identified during this transition, characterized by an arginine-rich helix exposed to the solvent. This helical structure, which bridges the recognition and nuclease lobes, was suggested to be critical for the recruitment of the guide RNA. This was indeed supported by electrostatic analysis of the intermediate, revealing the formation of a positively charged cavity poised to accommodate RNA. It is notable that subsequent experimental studies demonstrated that the arginine residues in this bridging helix influence the guide RNA and target DNA binding [17] and that the mutations of the bridging helix improve the specificity of Cas9 [17,18].

Figure 2



**Left panel.** (a) Mechanism of RNA binding. A dramatic conformational change is observed from the apo Cas9 (bottom, left) to the RNA-bound state (top, right), involving oppositely directed movements of the Rec1-3 domains to allocate nucleic acids (shown using arrows) [16]. An intermediate state, characterized by the solvent exposure of an arginine-rich helix (magenta), was suggested to be crucial for RNA recruitment. The free energy surface associated with the protein conformational change is also shown (bottom, right). **Right panel.** Conformational dynamics of the HNH domain. (b) Conformational (left) and energetic (right) landscape of the HNH domain, characterizing the three energetic minima (a'-c') in agreement with existing structural and smFRET studies [6,21]. The fourth energetic minimum (d') corresponds to a putative active state, where HNH docks at the DNA cleavage site. (c) Comparison between the predicted structure [16] and the cryo-EM density of the active HNH (EMD-0584) [25] reported  $\sim 2.5$   $\text{\AA}$  difference in RMSD (computed through rigid-body docking of the predicted structure into the cryo-EM map) [26]. (d) Outward movement of the Rec2-3 domains observed during molecular dynamics simulations of the HNH activation [24]. Adapted with permission from Palermo et al. [16] Copyright 2017 National Academy of Sciences.

## Formation of the catalytic complex

DNA binding induces the formation of several conformational states in Cas9, characterized by different orientations of the highly flexible HNH domain (Figure 2b) [5]. This ability of HNH to rapidly change conformation initially impeded structural characterization of the active state. The early DNA-bound structures of Cas9 displayed an inactive conformation of HNH [19,20], a so-called ‘conformational checkpoint’ between DNA binding and cleavage [5]. Subsequently, a precatalytic conformation was also structured, reporting the catalytic site at  $\sim 15$  Å from the DNA TS [10].

To broadly sample the conformational dynamics of HNH upon DNA binding, once again, GaMD appealed as a robust method [16]. Based on a  $\sim 20$   $\mu$ s sampling, a set of energetic minima corresponded to the available structures and smFRET experiments (minima a’-c’ in Figure 2b) [21]. The simulations also identified a putative active conformation of Cas9, which was found to be thermodynamically stable and in agreement with the available biophysical data (minimum d’ in Figure 2b) [21]. This investigation was followed by an independent study, corroborating similar collective motions of the HNH domain [22]. To fully understand the mechanism of activation, continuous simulations of the HNH transition were then obtained using Anton-2 [23], known as the world’s fastest supercomputer for MD simulations [24].

The dynamical docking of HNH at the cleavage site was characterized over multi- $\mu$ s long simulations, identifying an active conformation that confirmed the initial model. This theoretical model also remarkably agreed with the active state determined two years later through cryo-EM [25], reporting an average RMSD for the HNH domain of  $2.47 \pm 0.14$  Å (Figure 2c) [26].

The long timescale simulations further indicated that the Rec lobe undergoes an extensive conformational change. Specifically, a major opening was observed for the Rec2 (displaying an overall translation of  $\sim 8$ – $10$  Å relative to the starting position) and Rec3 (translating by  $\sim 5$  Å) domains, enabling docking of HNH at the catalytic site (Figure 2d) [24]. This large-scale movement of Rec2-3 was also observed in a series of very recent cryo-EM structures of Cas9 along with its conformational activation [14,27,28].

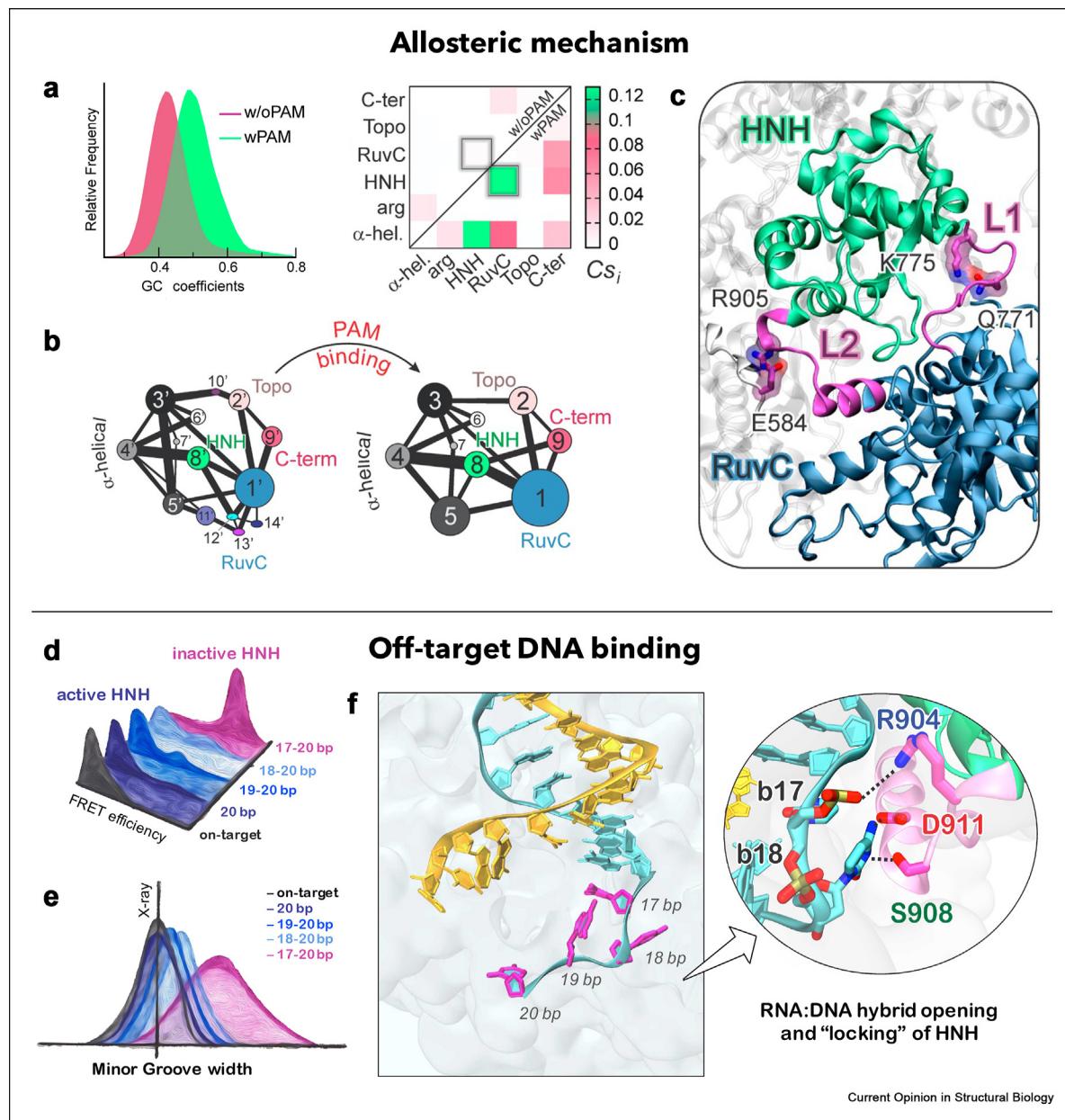
## Allosteric regulation

The CRISPR-Cas9 complex is an intriguing allosteric system. In the allosteric regulation, substrate binding at a region different from the catalytic site activates the protein function. Biochemical studies suggested that the binding of the PAM recognition sequence could allosterically activate the concerted catalytic function of the spatially distant HNH and RuvC nucleases [3]. All-

atom MD simulations revealed that the binding of PAM induces a population shift and highly coupled motions of HNH and RuvC (Figure 3a), in line with a typical allosteric response [29] and biochemical studies [30]. PAM binding also induced the formation of an optimal allosteric network, with a stronger connection between domains, when compared to the model without PAM (Figure 3b). This evidence suggested that PAM acts as an ‘allosteric effector’ of the Cas9 function. Calculation of the allosteric pathways revealed that the ‘cross-talk’ between the HNH and RuvC catalytic domains flows through the L1/L2 loops, regarded as the ‘signal transducers’ (Figure 3c) [10]. A recent cryo-EM study showed that L1/L2 dynamically repositions the HNH domain for TS cleavage while relocating the NTS in the catalytically active RuvC groove [28]. Furthermore, experimental modification of the ‘signal transducers’ led to the development of the LZ3-Cas9 variant with improved specificity [31]. Mutations of the central node residues, through which the majority of allosteric pathways pass, also reported increased specificity, as observed for the K775A and R905A mutations in the enhanced specificity (eCas9) [32] and Hyper Accurate (HypaCas9) [6] variants, respectively.

Biophysical studies also revealed that the flexibility of HNH could facilitate the signaling transfer from the DNA recognition region to the cleavage sites [5,6,11,21]. For characterizing the signal transmission, NMR relaxation dispersion experiments detected slow dynamical motions in core residues of HNH [33]. MD simulations showed that these residues form a contiguous pathway connecting the recognition region to the HNH and RuvC catalytic sites. This provided a path for the allosteric transmission and a mechanistic explanation for biophysical experiments [6,21], describing how the HNH dynamics communicate the DNA binding information (from Rec) to the nuclease sites. Interestingly, such interdomain signaling was found to be specific in the SpCas9, and is replaced by faster (ns) dynamics of a thermophilic variant of Cas9 [34]. The HNH allosterism was then investigated in the presence of three lysine-to-alanine mutations (K810A, K848A, and K855A) that are critical to enhancing Cas9 specificity [32]. The mutations were found to interrupt the main allosteric pathway connecting Rec to RuvC [35]. Notably, the disruption of the communication signal was found to compare to the order of specificity improvement observed in the three single mutants (K855A  $>$  K848A  $\sim$  K810A), with the mutation achieving the highest specificity also strongly perturbs the signal transfer. This suggested a direct link between changes in the allosteric network and the increase in the Cas9 specificity. Finally, it was also observed that the allosteric inhibition shifts the conformational ensemble toward a less catalytically competent state [36].

Figure 3



**Top panel.** Dynamic allosteric in CRISPR-Cas9 [30]. **(a)** Cas9 in complex with PAM (wPAM) displays an increase in the generalized correlations (GC) with respect to the system without PAM (w/oPAM). The matrix on the right shows that coupled motions of HNH and RuvC are activated in the presence of PAM. **(b)** Upon PAM binding, an optimal allosteric network is formed, with a stronger connection between domains (shown using thicker bonds), when compared to Cas9 without PAM. **(c)** Allosteric pathways between HNH and RuvC flowing through the L1/L2 signal transducers [10]. Central nodes of the signaling – R905, E584, K775, Q771 – are also shown. **Bottom panel.** Binding of off-target DNA sequences. **(d)** Schematic illustration of single-molecule data, showing one to three DNA mismatches at PAM-distal ends (positions 18 to 20) allow HNH to assume an active state, while four mismatches (positions 17 to 20) lead to an inactive state of HNH. Original data in Ref [6]. **(E)** Schematic illustration of the RNA: DNA minor groove width from molecular simulations, showing a substantial increase in the presence of four mismatches with respect to the crystallographic value. Original data in Ref [39]. **(F)** Opening of the RNA: DNA hybrid with four DNA mismatches. A close-up view shows the interactions between the DNA TS and the L2 loop, effectively ‘locking’ the HNH conformational dynamics. Adapted with permission from Palermo et al. (2017) [30], Copyright 2017 American Chemical Society, <https://pubs.acs.org/doi/10.1021/jacs.7b05313>; and from Ricci et al. (2019) [39], Copyright 2019 American Chemical Society, <https://pubs.acs.org/doi/full/10.1021/acscentsci.9b00020>. Further permissions related to the material excerpted should be directed to the American Chemical Society.

Collectively, these findings indicate that the allosteric regulation in CRISPR-Cas9 can be leveraged to improve the system's specificity. Indeed, engineering of the 'signal transducers' in the LZ3-Cas9 variant [31], mutation of the allosteric nodes in the eCas9 and HypaCas9 variants [6,32], as well as perturbation of the allosteric signaling through lysine-to-alanine mutations achieves an enhancement of specificity against off-target effects [32]. This highlights the potential of modulating the dynamically driven allostery of Cas9 to improve its function for therapeutic intervention. Most importantly, computational methods can identify critical allosteric hotspots, which can be targeted through mutational studies [35]. Controlling allosteric networks in proteins is indeed a useful strategy for regulating their function and the design of potential drugs [37], as shown for a number of allosteric enzymes, including protein kinases, ubiquitinating enzymes, and transcription factors [38], as well as large protein/nucleic acid complexes, such as the nucleosome core particle and the spliceosome [29].

### Mechanistic insights into off-target effects

Considering the widespread application of the CRISPR-Cas9 genome-editing technology, a severe issue limiting its therapeutic implementation is the off-target effects [2]. At the molecular level, this issue refers to the cleavage of DNA sequences unmatching the guide RNA, resulting in undesirable phenotypes. MD simulations delved into this phenomenon by introducing base pair mismatches within the RNA: DNA hybrid of CRISPR-Cas9 [39]. These studies were inspired by single-molecule experiments, showing that Cas9 binding and catalysis are extremely sensitive to the positioning of base-pair mismatches [5–8]. Indeed, mismatches until nine base pairs from the PAM-proximal end have a dramatic impact on DNA binding and cleavage, while PAM-distal mismatches compromise catalysis but still allow stable DNA binding. Notably, DNA containing one to three mismatches at the PAM-distal end (close to Rec3) enables sufficient flexibility in HNH to assume an active catalytic conformation. On the other hand, in the presence of four mismatches, HNH mainly assumes an inactive state (Figure 3d).

Molecular simulations revealed a mechanism for this inactivation. Indeed, four mismatches at the PAM-distal end induced an extended opening of the minor groove in the RNA: DNA hybrid (Figure 3e), which enables the DNA TS to form novel interactions with the L2 loop of the HNH domain (Figure 3f) [39]. This binding 'locks' HNH in its inactive configuration and decreases its ability to change conformation for DNA cleavage. Conversely, internal mismatches within the RNA: DNA structure are readily embedded in the hybrid, not significantly affecting the interaction with the protein [40]. These findings were recently corroborated by several X-ray crystal structures capturing Cas9 bound to

a number of off-target substrates [41]. Structural evidence also pointed out conformational changes of the Rec lobe (Rec2 and Rec3) with respect to HNH in the presence of DNA mismatches at the PAM-distal end. In this respect, MD simulations have shown that interactions between the Rec3 region and the on-target RNA: DNA complex stabilize the hybrid structure [39]. Interestingly, off-target binding at the PAM-distal end leads to conformational changes of Rec3, with destabilization of the heteroduplex. Thus, MD simulations, single-molecule and structural experiments complemented each other toward understanding off-target binding in CRISPR-Cas9.

### Catalytic mechanism of DNA cleavage

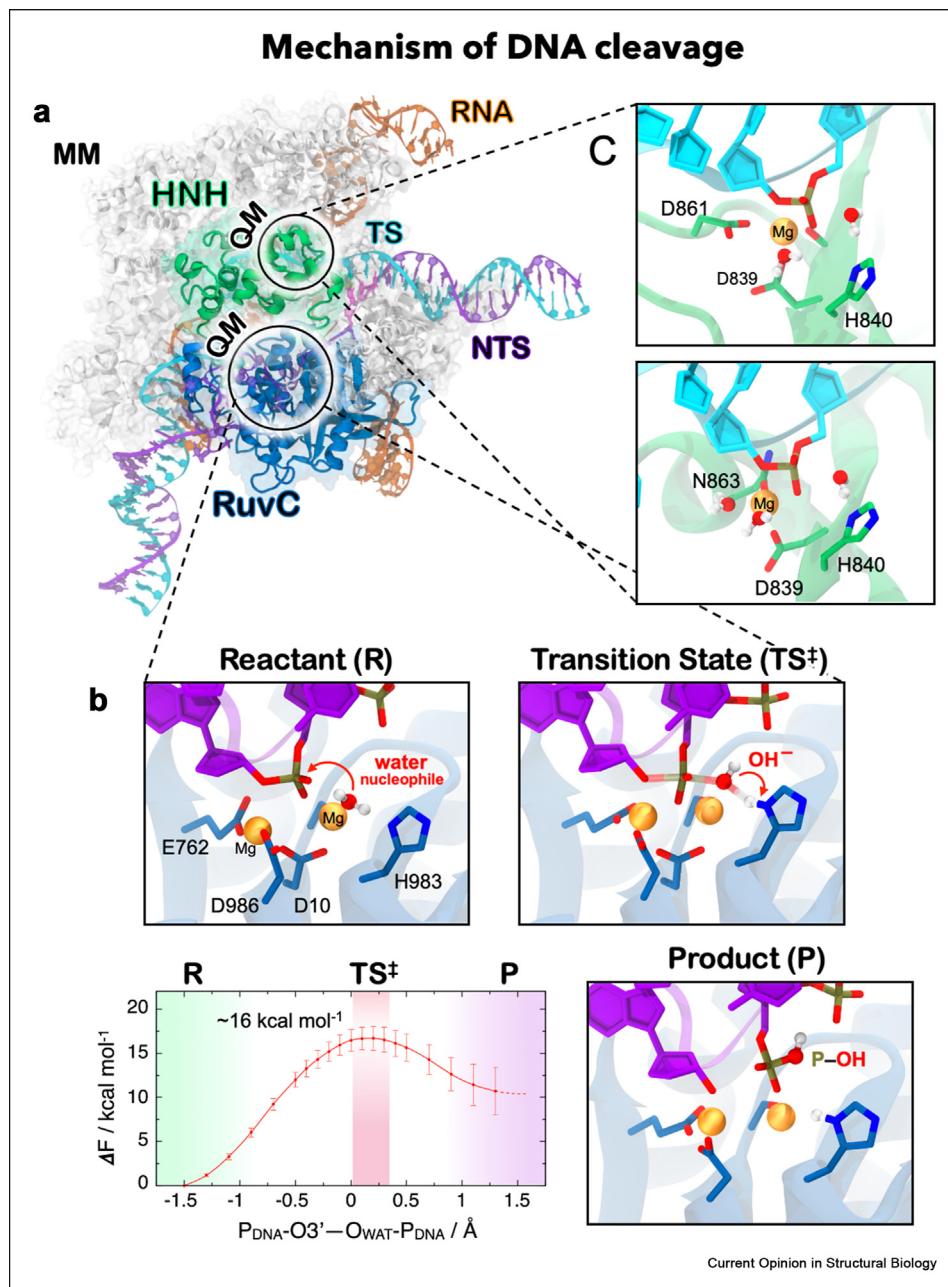
Cas9 is a metal-dependent nuclease that uses  $Mg^{2+}$  to cleave DNA [1]. While the RuvC domain cleaves the NTS using two  $Mg^{2+}$  ions, the HNH nuclease performs TS cleavage depending on a single  $Mg^{2+}$  ion. Studies of the catalytic mechanism required quantum mechanics/molecular mechanics (QM/MM) approach that allows the capturing of the formation and breakage of bonds. Briefly, the reactive center is treated at QM Density Functional Theory (DFT) level, while the rest of the system in explicit solution is described at the classical MM level (Figure 4a). This method was combined with *ab-initio* MD and free energy methods to follow the dynamics along with the enzymatic reaction in both the RuvC and HNH domains of Cas9.

Studies of the RuvC catalysis have been based on a metal-bound structure of Cas9 obtained in the presence of alternative metal ions (i.e.,  $Mn^{2+}$ ) [12]. This catalytic core displayed a highly conserved carboxylate motif (D10, D986, and E762) coordinating the two metals. *Ab initio* QM/MM simulations revealed a conformational rearrangement of the  $Mg^{2+}$ -bound RuvC active site, which entails the relocation of H983 to act as a general base (Figure 4b) [42]. Then, DNA cleavage proceeds through an associative catalytic mechanism, critically aided by the cooperative dynamics of the two  $Mg^{2+}$  ions [43].

This proposed mechanism was in line with the catalytic rate for RuvC ( $3.5\text{ s}^{-1}$  corresponding to  $\Delta G$  of  $\sim 16/17\text{ kcal mol}^{-1}$ ) [44] and clarified previous experimental evidence, which could not fully establish the catalytic role of the conserved H983 and the metal cluster conformation [12,20]. Moreover, the H983A mutation hampers NTS cleavage [20], supporting H983 acts as a nucleophilic activator. Finally, the identified mechanism was substantiated by the analysis of other metal-dependent nucleases, suggesting a similar catalytic strategy for genome editing and recombination.

HNH cleaves the DNA TS upon a complex conformational rearrangement [16]. Because of the lack of

Figure 4



Catalytic mechanism of DNA cleavage in CRISPR-Cas9. (a) QM/MM partitioning of CRISPR-Cas9. The HNH and RuvC catalytic cores are treated at the QM (DFT) level of theory, while the rest of the system is described at the classical MM level. (b) Structural and energetic properties of the catalysis of nontarget DNA cleavage within the RuvC active site. The catalysis proceeds from the reactant (R) to the transition state (TS<sup>‡</sup>) and product (P) through an associative mechanism activated by H983 [40]. The computed free energy barrier of ~16 kcal mol<sup>-1</sup> is in line with the experimental  $k_{\text{cat}}$  of ~3.5 s<sup>-1</sup> [41]. (c) Configurations of the HNH active site. Early structural models suggested that D861 and D839 could coordinate Mg<sup>2+</sup> and form a catalytic triad with H840 (top panel) [13,19,20]. Recent structural data of HNH right after TS cleavage suggest that N863 coordinates Mg<sup>2+</sup> in place of D861 (bottom panel) [25,28].

structural data on the catalytically competent HNH, studies of the catalytic mechanism used the information inferred from the T4 endonuclease VII analog [45]. It was suggested that D861, D839, and N863 coordinate Mg<sup>2+</sup> for catalysis, whose coordination sphere gets

saturated by the nucleophilic water (Figure 4c, top panel). This conformation of HNH is supported by several X-ray structures capturing HNH in various noncatalytic states, where D861 points toward D839 [13,19,20]. More recently, the structural determination

of HNH right after TS cleavage displayed a different configuration of the active site (Figure 4c, bottom panel) [25,28]. Here, N863 (rather than D861) coordinates Mg<sup>2+</sup> and forms a catalytic triad with D839 and H840. This configuration is supported by DNA cleavage assays, showing that the D861A substitution retained DNA cleavage activity while N863A lost gene-editing capability [46]. In light of these new experiments, the catalytic mechanism could remarkably differ and grants further investigations currently ongoing in our lab.

## Conclusions

Here, we showcased the role of state-of-the-art computer simulations in characterizing the dynamics and mechanism of the CRISPR-Cas9 genome-editing system. A multiscale approach harnessing classical and enhanced simulations, *ab-initio* MD and mixed QM/MM approaches, was used to delve into the biomolecular function, catalysis, allosteric, and selectivity of this system. Inspired by single-molecule experiments, structural data, and biochemical studies, molecular simulations have been instrumental in understanding the mechanism of action of CRISPR-Cas9, providing useful insight for rational engineering. Future computational studies will bridge the scales from quantum mechanical approaches to enhanced simulations and cryo-EM processing methods to push the boundaries of molecular simulations and tackle the most exciting problems in the biophysics of genome editing.

## Author contribution

AS, PRA, and GP wrote the manuscript. PRA and GP created the graphics. GP created the schematic illustrations through handmade painting and digital manipulation, in Figure 3d–e. GP conceived this research.

## Conflict of interest statement

Nothing declared.

## Acknowledgements

This material is based upon work supported by the National Institutes of Health (Grant No. R01GM141329) and by the National Science Foundation (Grant No. CHE-1905374). Computer time for MD simulations has been awarded by XSEDE (Grant No. TG-MCB160059) and by NERSC (Grant No. M3807).

## References

Papers of particular interest, published within the period of review, have been highlighted as:

\* of special interest  
\*\* of outstanding interest

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.** *Science* 2012, **337**: 816–821.
2. Doudna JA: **The promise and challenge of therapeutic genome editing.** *Nature* 2020, **578**:229–236.

This review article discusses the scientific, technical and ethical aspects of using CRISPR technology for therapeutic applications.

3. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA: **DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.** *Nature* 2014, **507**:62–67.
4. Jiang F, Doudna JA: **CRISPR–Cas9 structures and mechanisms.** *Annu Rev Biophys* 2017, **46**:505–529.
5. Dagdas YS, Chen JS, Sternberg SH, Doudna JA: **A conformational checkpoint between DNA binding and cleavage by CRISPR-cas9.** *Sci Adv* 2017, **3**, eaao002.
6. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA: **Enhanced proofreading governs CRISPR–Cas9 targeting accuracy.** *Nature* 2017, **550**:407–410.
7. Singh D, Sternberg SH, Fei J, Doudna JA, Ha T: **Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9.** *Nat Commun* 2016, **7**:12778.
8. Singh D, Wang Y, Mallon J, Yang O, Fei J, Poddar A, Ceylan D, Bailey S, Ha T: **Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis.** *Nat Struct Mol Biol* 2018, **25**:347–354.
9. Newton MD, Taylor BJ, Driessens RPC, Roos L, Cveticic N, Allyjaun S, Lenhard B, Cuomo ME, Rueda DS: **DNA stretching induces Cas9 off-target activity.** *Nat Struct Mol Biol* 2019, **26**: 185–192.
10. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA: **Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage.** *Science* 2016, **351**:867–871.
11. Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA: **Striking plasticity of CRISPR-cas9 and key role of non-target DNA, as revealed by molecular simulations.** *ACS Cent Sci* 2016, **2**: 756–763.
12. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, et al.: **Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.** *Science* 2014, **343**:1247997. 1247997.
13. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA: **A Cas9-guide RNA complex preorganized for target DNA recognition.** *Science* 2015, **348**:1477–1481.
14. Cofsky JC, Soczek KM, Knott GJ, Nogales E, Doudna JA: **\* CRISPR-Cas9 bends and twists DNA to read its sequence.** *Nat Struct Mol Biol* 2022, **29**:395–402, <https://doi.org/10.1101/2021.09.06.459219>.

This study employed cryo-EM and biochemical experiments to characterize the conformations of Cas9 through its activation process. Structural data revealed large-scale conformational changes of the recognition lobe to allocate DNA binding.

15. Miao Y, Feher VA, McCammon JA: **Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation.** *J Chem Theor Comput* 2015, **11**:3584–3595.
16. Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA: **CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations.** *Proc Natl Acad Sci USA* 2017, **114**:7260–7265.
17. Bratović M, Fonfara I, Chylinski K, Gálvez EJC, Sullivan TJ, Boerno S, Timmermann B, Boettcher M, Charpentier E: **Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches.** *Nat Chem Biol* 2020, **16**:587–595.
18. Babu K, Amrani N, Jiang W, Yugesha SD, Nguyen R, Qin PZ, Rajan R: **Bridge helix of Cas9 modulates target DNA cleavage and mismatch tolerance.** *Biochemistry* 2019, **58**:1905–1917.
19. Anders C, Niewohner O, Duerst A, Jinek M: **Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.** *Nature* 2014, **513**:569–573.
20. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, 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.
21. Sternberg SH, LaFrance B, Kaplan M, Doudna JA: **Conformational control of DNA target cleavage by CRISPR–Cas9.** *Nature* 2015, **527**:110–113.

22. Zuo Z, Liu J: **Structure and dynamics of Cas9 HNH domain catalytic state.** *Sci Rep* 2017, **7**.

23. Shaw DE, Grossman JP, Bank JA, Batson B, Butts JA, Chao JC, Deneroff MM, Dror RO, Even A, Fenton CH, et al.: **Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer.** In *SC14: international conference for high performance computing, networking, storage and analysis.* IEEE; 2014:41–53.

24. Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, Doudna JA, McCammon JA: **Key role of the REC lobe during CRISPR–Cas9 activation by ‘sensing’, ‘regulating’, and ‘locking’ the catalytic HNH domain.** *Q Rev Biophys* 2018, **51**:e9.

25. Zhu X, Clarke R, Puppala AK, Chittori S, Merk A, Merrill BJ, Simonović M, Subramaniam S: **Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9.** *Nat Struct Mol Biol* 2019, **26**:679–685.

26. Nierwinski Ł, Palermo G: **Molecular dynamics to predict cryo-EM: capturing transitions and short-lived conformational states of biomolecules.** *Front Mol Biosci* 2021, **8**:120.

This study compared the structure of an activated CRISPR-Cas9 obtained through molecular simulations (REF. 16) with the cryo-EM density of the activated system (EMD-0584, REF. 25). Rigid-body docking of the predicted structure within the cryo-EM density reported a ~2.5 Å RMSD difference for the HNH domain. More broadly, the study showed the useful capability of molecular dynamics simulations to predict short-lived conformational states, which are often difficult to capture through experimental techniques.

27. Pacesa M, Jinek M: **Mechanism of R-loop formation and conformational activation of Cas9.** *bioRxiv* 2021, <https://doi.org/10.1101/2021.09.16.460614>.

In this article, cryo-EM structures of intermediate binding states of CRISPR-Cas9 reveal that the Rec2 and Rec3 domains undergo substantial conformational changes. The Rec2 domain is shown to substantially open-up to enable DNA binding. This opening of Rec2 is in line with the large-scale conformational change of Rec2 observed during molecular dynamics, reporting an outward translation of ~8–10 Å with respect to the starting position (REF. 24).

28. Bravo JPK, Liu M-S, Hibshman GN, Dangerfield TL, Jung K, McCool RS, Johnson KA, Taylor DW: **Structural basis for mismatch surveillance by CRISPR–Cas9.** *Nature* 2022, **603**: 343–347.

Here, kinetic experiments and cryo-EM structures were used to characterize the binding of off-target DNA sequences. Structural data revealed that the L1/L2 loops dynamically reposition the HNH domain for TS cleavage, in line with their role of allosteric transducers (REFs. 10 and 30). High flexibility of the Rec2 domain is also observed.

29. Arantes PR, Patel AC, Palermo G: **Emerging methods and applications to decrypt allosteric in proteins and nucleic acids.** *J Mol Biol* 2022, <https://doi.org/10.1016/j.jmb.2022.167518>.

This review article discusses the established and most innovative approaches to elucidate allosteric in large protein/nucleic acid systems. Allosteric models and applications are showcased for three paradigmatic examples of allosteric: (i) the nucleosome core particle, (ii) the CRISPR-Cas9 genome editing system, and (iii) the spliceosome. The review includes fundamental aspects and practical examples to target a broad audience of scientists interested in the field.

30. Palermo G, Ricci CG, Fernando A, Basak R, Jinek M, Rivalta I, Batista VS, McCammon JA: **Protospacer adjacent motif-induced allosteric activates CRISPR-cas9.** *J Am Chem Soc* 2017, **139**:16028–16031.

31. Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, Zhang F: **Highly parallel profiling of Cas9 variant specificity.** *Mol Cell* 2020, **78**:749–800.

32. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F: **Rationally engineered Cas9 nucleases with improved specificity.** *Science* 2016, **351**:84–88.

33. East KW, Newton JC, Morzan UN, Narkhede YB, Acharya A, Skeens E, Jogl G, Batista VS, Palermo G, Lisi GP: **Allosteric motions of the CRISPR–cas9 HNH nuclease probed by NMR and molecular dynamics.** *J Am Chem Soc* 2020, **142**: 1348–1358.

This study combined molecular simulations with solution NMR to study the allosteric role of the HNH dynamics in CRISPR-Cas9. A millisecond timescale dynamic pathway was found to span HNH from the RuvC nuclease to the DNA recognition region. This provided a route for the allosteric transduction and a mechanistic rationale to biophysical experiments (REFs 6 and 21), clarifying how the HNH dynamics could transfer the DNA binding information to the cleavage sites.

34. Belato BH, D'Ordine AM, Nierwinski Ł, Arantes PR, Jogl G, Palermo G, Lisi GP: **Structural and dynamic insights into the HNH nuclease of divergent Cas9 species.** *J Struct Biol* 2022, **214**:107814.

35. Nierwinski Ł, East KW, Morzan UN, Arantes PR, Batista VS, Lisi GP, Palermo G: **Enhanced specificity mutations perturb allosteric signaling in the CRISPR-cas9 HNH endonuclease.** *Elife* 2021, **10**, e73601.

This paper shows that mutations enhancing the specificity of Cas9 also alter the system's allosteric communication. It suggests that modulating the allosteric signaling could be a useful strategy to decrease off-target effects. The paper also reports a computational protocol to identify allosteric hotspots that can be targeted through mutational studies for improving the system's function.

36. Li X, Wang C, Peng T, Chai Z, Ni D, Liu Y, Zhang J, Chen T, Lu S: **Atomic-scale insights into allosteric inhibition and evolutionary rescue mechanism of *Streptococcus thermophilus* Cas9 by the anti-CRISPR protein AcrIIA6.** *Comput Struct Biotechnol J* 2021, **19**:6108–6124.

37. Dokholyan NV: **Controlling allosteric networks in proteins.** *Chem Rev* 2016, **116**:6463–6487.

38. Papaleo E, Saladino G, Lambrughi M, Lindorff-Larsen K, Gervasio FL, Nussinov R: **The role of protein loops and linkers in conformational dynamics and allostery.** *Chem Rev* 2016, **116**:6391–6423.

39. Ricci CG, Chen JS, Miao Y, Jinek M, Doudna JA, McCammon JA, Palermo G: **Deciphering off-target effects in CRISPR-cas9 through accelerated molecular dynamics.** *ACS Cent Sci* 2019, **5**:651–662.

40. Mitchell BP, Hsu RV, Medrano MA, Zewde NT, Narkhede YB, Palermo G: **Spontaneous embedding of DNA mismatches within the RNA:DNA hybrid of CRISPR-cas9.** *Front Mol Biosci* 2020, **7**:39.

41. Pacesa M, Lin CH, Cléry A, Bargsten K, Irby MJ, Allain FHT, Cameron P, Donohoue PD, Jinek M: **Structural basis for Cas9 off-target activity.** *bioRxiv* 2021, <https://doi.org/10.1101/2021.11.18.469088>.

This study reported a series of crystallographic structures bound to bona fide off-target substrates. Structural data show that DNA mismatches within the RNA:DNA hybrid result in minor effects on the protein/nucleic acid interactions, while PAM-distal mismatches induce an unpairing of the RNA:DNA hybrid, with conformational changes in the Rec2-3 and HNH domain. These structural data corroborate early predictions obtained through molecular dynamics simulations (REFs 39–40).

42. Palermo G: **Structure and dynamics of the CRISPR–cas9 catalytic complex.** *J Chem Inf Model* 2019, **59**:2394–2406.

43. Casalino L, Nierwinski Ł, Jinek M, Palermo G: **Catalytic mechanism of non-target DNA cleavage in CRISPR-cas9 revealed by ab initio molecular dynamics.** *ACS Catal* 2020, **10**:13596–13605.

Here, high-level quantum mechanical simulations described the chemical mechanism of nontarget DNA cleavage within the RuvC domain. A mixed quantum-classical (QM/MM) approach was combined with free energy methods to describe the dynamical pathways and energetic properties of phosphodiester bond cleavage. The study revealed an associative mechanism activated by H983, in agreement with DNA cleavage experiments (REF. 20) and kinetic data (REF. 44). A common catalytic strategy for genome editing and recombination was also suggested.

44. Gong S, Yu HH, Johnson KA, Taylor DW: **DNA unwinding is the primary determinant of CRISPR–cas9 activity.** *Cell Rep* 2018, **22**:359–371.

45. Yoon H, Zhao LN, Warshel A: **Exploring the catalytic mechanism of Cas9 using information inferred from endonuclease VII.** *ACS Catal* 2019, **9**:1329–1336.

46. Zuo Z, Zolekar A, Babu K, Lin VJ, Hayatshahi HS, Rajan R, Wang Y-C, Liu J: **Structural and functional insights into the bona fide catalytic state of *Streptococcus pyogenes* Cas9 HNH nuclease domain.** *Elife* 2019, **8**, e46500.