IOP Publishing Electron. Struct. 5 (2023) 014003 https://doi.org/10.1088/2516-1075/acb410

Electronic Structure

PAPER

The electronic structure of genome editors from the first principles

Łukasz Nierzwicki[1], Mohd Ahsan[1] and Giulia Palermo[1,2,*]

[1] Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States of America
[2] Department of Chemistry, University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: giulia.palermo@ucr.edu

RECEIVED
21 October 2022

REVISED
22 December 2022

ACCEPTED FOR PUBLICATION
17 January 2023

PUBLISHED
1 February 2023

Keywords: molecular dynamics, QM/MM, CRISPR-cas, RNA, free energy, catalysis

Abstract

Ab-initio molecular dynamics enables following the dynamics of biological systems from the first principles, describing the electronic structure and offering the opportunity to "watch" the evolution of biochemical processes with unique resolution, beyond the capabilities of state-of-the-art experimental techniques. This article reports the role of first-principles (*ab-initio*) molecular dynamics (MD) in the CRISPR-Cas9 genome editing revolution, achieving a profound understanding of the enzymatic function and offering valuable insights for enzyme engineering. We introduce the methodologies and explain the use of *ab-initio* MD simulations to establish the two-metal dependent mechanism of DNA cleavage in the RuvC domain of the Cas9 enzyme, and how a second catalytic domain, HNH, cleaves the target DNA with the aid of a single metal ion. A detailed description of how *ab-initio* MD is combined with free-energy methods—i.e., thermodynamic integration and metadynamics—to break and form chemical bonds is given, explaining the use of these methods to determine the chemical landscape and establish the catalytic mechanism in CRISPR-Cas9. The critical role of classical methods is also discussed, explaining theory and application of constant pH MD simulations, used to accurately predict the catalytic residues' protonation states. Overall, first-principles methods are shown to unravel the electronic structure and reveal the catalytic mechanism of the Cas9 enzyme, providing valuable insights that can serve for the design of genome editing tools with improved catalytic efficiency or controllable activity.

1. Introduction

Genome editing enables the modification of nucleic acids, by deleting, replacing, or inserting desired sequences, to improve biological function and correct diseases [1]. Considered the future of medicine and biotechnology, genome editing will enable to cure at their source plethora of genetic diseases, including cancer and neurodegeneration. Genome editor proteins are at the core of this revolution, with the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 system introducing the precise manipulation of nucleic acids [2].

CRISPR-Cas9 is a part of a prokaryotic immune system that protects bacterial cells against invading foreign DNA. This system is based on a single protein—the endonuclease Cas9—that uses a guide RNA to recognize and cleave foreign DNA sequences [2, 3]. As the guide RNA can be easily switched to target any desired DNA sequence, this system has been widely applied for targeting and manipulating nucleic acids not only in basic bioscience, but also as a successful genome editor for medicinal, pharmaceutical and biotechnological purposes. Due to the widespread utilization of CRISPR-Cas9, an urgent need has emerged to understand its enzymatic function and rationally tune the complex activity to fit the requirements of specific applications. This has been an ideal scenario for first-principles molecular dynamics (MD), enabling us to dig deeply into the electronic structure of CRISPR-Cas9, characterizing its catalytic cycle, and offering fundamental insights for the enzymatic design of improved function.

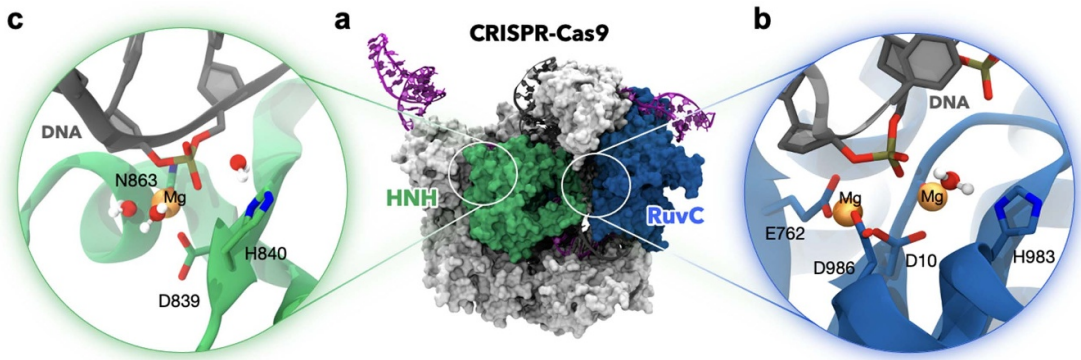

Figure 1. (a) Overview of the *Streptococcus pyogenes* CRISPR-Cas9 system in complex with DNA and RNA (PDB: 5F9R) [11]. The cas9 protein is shown in molecular surface, highlighting the catalytic domains RuvC (blue) and HNH (green). The RNA (magenta) and the DNA (black) are shown as ribbons. (b) Catalytic site of the RuvC domain, displaying a two-metal ion architecture, including two Mg^{2+} ions (orange spheres) surrounded by the E762, D986, D10 carboxylates and the catalytic H983 [15, 16]. (c) Catalytic site of the HNH domain, holding a single catalytic Mg^{2+} ion (orange sphere) coordinated by D839, the N863 backbone, and water molecules [13, 15]. The catalytic H840 is also shown.

Intense structural studies of the CRISPR-Cas9 reported several crystallographic and cryogenic electron microscopy (cryo-EM) structures [4, 5], from a number of species [6–8], characterizing the Cas9 protein in the apo form and in complex with nucleic acids [9–11], and recently also bound to off-target DNA sequences [12]. The most recent structures of the *Streptococcus pyogenes* Cas9 disclose the system in its activated form [13–15], prone to perform double-stranded DNA breaks through two nuclease domains—viz., HNH and RuvC (figure 1(a)). HNH cleaves the DNA strand base-pairing the guide RNA, i.e. the so-called target strand, while RuvC cleaves the other non-target strand (NTS). The nuclease activity of CRISPR-Cas9 is cardinal for genome editing. Indeed, upon site-specific DNA cleavages, the homology-directed DNA repair machinery of the cell fixes the break introducing the desired insertions or modification.

As the CRISPR-Cas9 complex originates from the bacterial immune system, it evolved in a DNA-rich environment avoiding, but often also tolerating, the undesired cleavage of DNA sequences that do not fully match the RNA guide. This phenomenon is part of the so-called 'off-target' effects [17], which can limit the applicability of CRISPR-Cas9 for biomedical applications, resulting in the deletion/insertion of erroneous genes, and henceforth unwanted phenotypes. In order to adapt to this scenario, Cas9 developed a number of allosteric control mechanisms [18, 19], which help limit DNA off-target cleavages. Extensive experimental efforts have established the biophysical function of this genome editor [20], using biochemistry [2, 21, 22], structural biology [4–15], and single-molecule spectroscopy [23–27]. Computational studies based on classical mechanics outlined the conformational changes for nucleic acid binding [28–32], the allosteric effects associated with DNA recognition [33–35], and its specificity [36, 37]. However, characterizing how CRISPR-Cas9 ultimately cleaves DNA is a matter that pertains to its electronic structure and, hence, high-level quantum mechanical (QM) simulations. Understanding the mechanism of DNA cleavage and its underlying electronic structure is critical to control its enzymatic function, and developing new genome editing tools with improved specificity.

In this review article, we provide an overview of the computational methods that have been used to characterize the catalytic mechanism by which CRISPR-Cas9 cuts nucleic acids. The theoretical basis of *ab-initio* methods is explained, followed by their application, providing a direct explanation of their use with real-world examples on genome editing systems. This contribution offers an easy-to-read description of complex computational approaches, including high-level quantum mechanics and first-principles MD simulations, which will help their understanding by a broad audience of chemists and biochemists. Methods and applications reviewer here are valuable to the computational investigation of genome editing systems and RNA guided enzymes through computational methods [38].

2. Unique opportunity for first-principles molecular dynamics

Cas9 is a metal-dependent nuclease, performing DNA cleavages thanks to the aid of metal ions [9]. The RuvC domain displays the structural fold of the RNA ribonuclease H, with a two-metal dependent site, which mainly exploits Mg^{2+} to perform phosphodiester bond cleavages (figure 1(b)). The HNH domain shows structural homology with homing endonucleases (such as the T4 endonuclease VII), performing phosphodiester bond cleavage of the DNA through a single Mg^{2+} ion (figure 1(c)). Although the structural homology of the Cas9 domains with known nucleases was known, the active site chemistry and the specific

role of residues in the catalytic pocket remained ambiguous for a long time. Due to the high flexibility of the complex, early structural studies alone could not provide an unambiguous description of the active site chemistry. Moreover, as the enzyme performs its activity, it is often difficult for structural biology to characterize the rapid events during the cleavage step. As a few experimental techniques enable us to 'watch' the bond breaking and formation along chemical reactions [39], high-level QM offers the opportunity to unravel the electronic structure deep inside, and to describe how chemical bonds form and break. By empowering QM with MD simulations, one can follow the dynamics of the enzymatic reaction from the first-principles (i.e. *ab-initio*), integrating the equation of motion for the electronic degrees of freedom over time. *Ab-initio* MD also offers a reliable description of metal ions, which are so crucial for the Cas9 enzyme and for other genome editors, but whose charge transfer and polarization effects are not completely described by point-charge classical force fields [40].

Despite this power and ever-increasing advance in high-performance computing (HPC), *ab-initio* MD remains computationally expensive and restricted to a few hundreds of atoms. This is a limitation for studying genome editors in their realistic water environment, as they can comprise more than 500 000 atoms. This is solved by combining *ab-initio* MD with mixed QM/molecular mechanics (QM/MM) schemes, treating the enzyme's active site at a high QM level, while the rest of the system in explicit solution is described at the classical MM level (figure 2(a)) [41]. This scheme allows taking into account how the environment impacts the electronic structure of the reactive center, providing a holistic description.

In this scenario, first-principles MD is uniquely positioned not only to achieve a profound understanding of the enzymatic function but to provide valuable insights that can directly serve the design of genome editing enzymes with improved catalytic efficiency or controllable activity. As one of the key goals of CRISPR-Cas9 engineering is to reduce off-target cleavages, knowing how Cas9 cleaves DNA is of the utmost need.

3. Overview of first-principles (*ab-initio*) methods

First-principles (*ab-initio*) MD uses accurate electronic structure calculations to integrate the equation of motion over time, henceforth describing the dynamics of nuclei and electrons [42]. Thanks to the QM description, and its combination with MD simulations, bonds can be formed and broken 'on-the-fly', enabling to study of enzymatic catalysis [43]. The QM description can be achieved using various levels of theory. A good trade-off between accuracy and computational cost is reached with density functional theory (DFT), using functionals of the electron density [44]. Though the application of a mixed QM/MM approach, the reactive center of the enzyme (i.e. active site) can be treated at a QM DFT level, while the rest of the system in explicit solution is described at the classical MM level (figure 2(a)) [45–47]. In the general form of a hybrid QM/MM scheme, the total energy of the system (i.e. the Hamiltonian, H) is achieved through the summation of the Hamiltonians for the quantum (H_{QM}) and classical (H_{MM}) systems and the interaction between the QM and MM regions ($H_{\mathrm{QM/MM}}$). This results in a single hybrid Hamiltonian, H:

$$H = H_{\mathrm{QM}} + H_{\mathrm{MM}} + H_{\mathrm{QM/MM}}. \tag{1}$$

The classical Hamiltonian (H_{MM}) is calculated based on the MM formalism using the force field. The quantum H_{QM} is obtained from the Kohn–Sham energy, described as:

$$E^{\mathrm{ks}}\left[\{\emptyset_i\}\right] = T_s\left[\{\emptyset_i\}\right] + \int \mathrm{d}r\, V_{\mathrm{ext}}(r)\, n(r) + \frac{1}{2}\int \mathrm{d}r\, V_{\mathrm{H}}(r)\, n(r) + E_{\mathrm{xc}}\left[n\right] + E_{\mathrm{ions}}\left(R^{\mathrm{N}}\right) \tag{2}$$

where $\emptyset_i$ is the auxiliary function representing the Kohn–Sham orbitals. The first term (T_s) in the equation represents the kinetic energy, the second term (V_{ext}) denotes the external potential due to the interaction between nuclei and electrons, the third term (V_{H}) represents electrostatic energy due to electron density calculated using Hartree potential, and the fourth term (E_{xc}) is the exchange correlation function calculated here using generalized gradient approximation functionals. The last term accounts for the interaction energy due to nuclear charges.

The $H_{\mathrm{QM/MM}}$ determines the interaction between the QM and MM regions of system. It comprises of terms defining the bonded interaction ($H^{\mathrm{bonds}}_{\mathrm{QMMM}}$) at the boundary, Van der Waals ($H^{\mathrm{vdW}}_{\mathrm{QMMM}}$) and electrostatic interactions ($H_{\mathrm{QMMM}^{\mathrm{elect}}}$) between QM and MM regions,

$$H_{\mathrm{QMMM}} = H^{\mathrm{bonds}}_{\mathrm{QM/MM}} + H^{\mathrm{vdW}}_{\mathrm{QM/MM}} + H^{\mathrm{elect}}_{\mathrm{QM/MM}}. \tag{3}$$

In the equation above, the first term takes care of the covalent bonds which crosses the QM and MM interface. There are three approaches to address these bonds: (a) hydrogen atoms can be placed along the bond to complete the valency requirements of the QM atoms, (b) a capping potential can be used to cap the

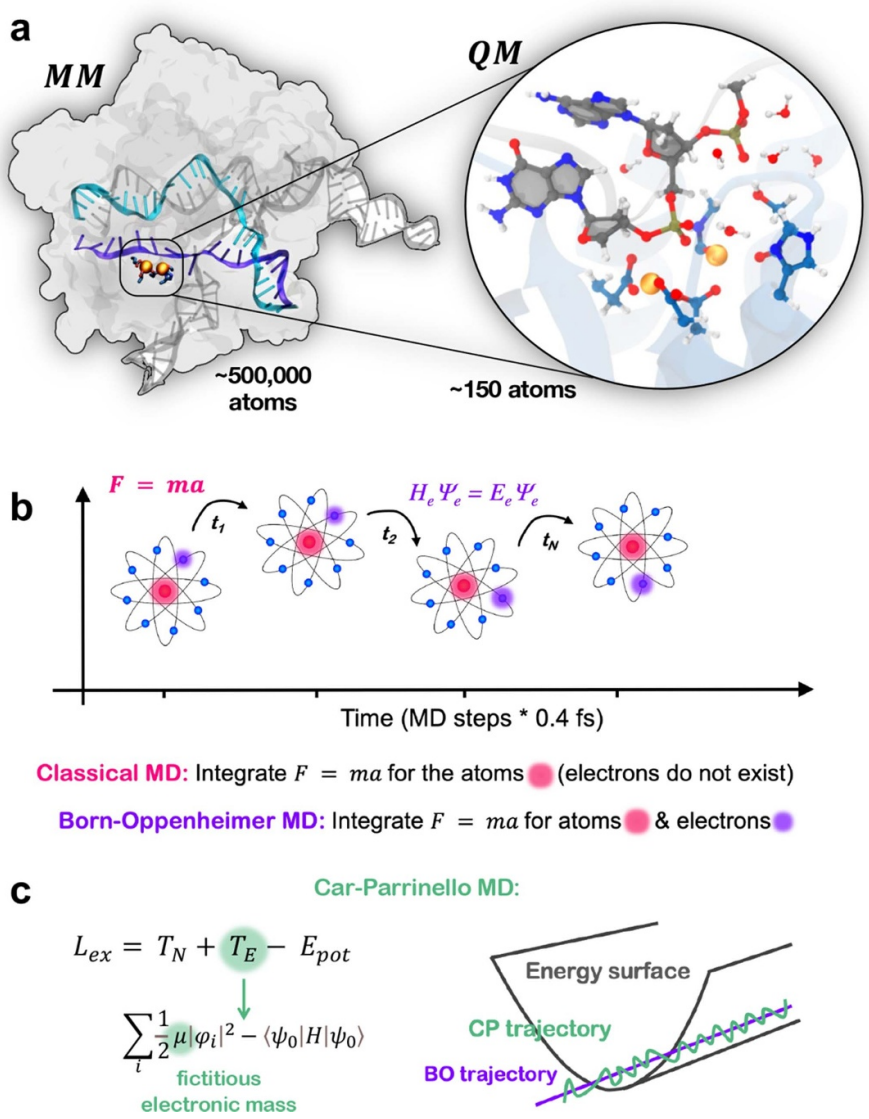

Figure 2. First-principles quantum mechanics/molecular mechanics (QM/MM) approaches. (a) QM/MM partitioning of the CRISPR-Cas9 system. The reactive center is treated at a QM (DFT/BLYP [30, 31]) level of theory, while the remaining of the system is treated at a classical MM force field level (Amber ff12SB [48] + ff99bsc0 for DNA [49] & χOL3 for RNA [50, 51]). (b) Simplified diagram of classical and Born–Oppenheimer MD simulations. In classical MD simulations, the equation of motion $F = ma$ is integrated over time for the atoms only, considering the atoms as balls connected by springs. In Born–Oppenheimer MD simulations, the electronic Schrödinger equation is solved at each time step of the dynamics, computing the forces for the present nuclear configuration. The forces are then used to integrate the equation of motion and propagate the dynamics. (c) In Car–Parrinello MD simulations, an extended Lagrangian (L_{ex}) is used to introduce Newtonian fictitious dynamics for the electronic degrees of freedom. L_{ex} includes the kinetic energy for the nuclei (T_N) and for the electronic degrees of freedom (T_E), as well as the potential energy (E_{pot}). By introducing a *'fictitious'* mass, the electronic degrees of freedom are propagated on the Born–Oppenheimer surface without the need of solving the electronic Schrödinger equation at each time step (as in Born–Oppenheimer MD). Details are given in the text.

boundary atoms, (c) localized frozen orbitals can be used to define for the boundary atoms of the QM region. The Van der Waals ($H^{\mathrm{vdW}}_{\mathrm{QM/MM}}$) interaction between QM and MM atoms are calculated by using the 12—6 LJ potential using the coefficients from the MM forcefield. The electrostatic interactions ($H^{\mathrm{elect}}_{\mathrm{QM/MM}}$) between QM and MM regions are the most crucial term for the $H_{\mathrm{QM/MM}}$ Hamiltonian, as the surrounding MM atoms polarize the QM region and establish environmental effects on the QM region. The best approach to obtain the electrostatic term is a fully Hamiltonian electronic embedding scheme [52]. In this method, short range electrostatic effects due to the MM atoms near to the QM region are obtained through a modified coulombic functional, while the electrostatics between the distant MM atoms and QM region is determined by coupling the multipole moments of QM charge with the point charges on MM atoms. The system's dynamics is then investigated by integrating the equation of motion for the electronic degrees of freedom over time through the Born–Oppenheimer or Car–Parrinello methods (vide infra).

4. *Ab-initio* schemes to describe the electronic structure of genome editors

Born–Oppenheimer MD is an established *ab-initio* method, which builds on the approximation that the motion of the electrons and nuclei can be separated [42]. Since the electrons are by orders of magnitude lighter than the nuclei, they move faster and relax rapidly to the ground-state configuration given by the nuclear positions. Hence, the nuclei can be considered stationary points, and their coordinates become parameters in the wave function for the electrons. The electronic Schrödinger equation can be solved for each fixed nuclear configuration:

$$H_{\mathrm{e}}\Psi_{\mathrm{e}} = E_{\mathrm{e}}\Psi_{\mathrm{e}} \tag{4}$$

with E_{e} is the contribution of the electrons to the energy of the system and H_{e} is the electronic Hamiltonian. In Born–Oppenheimer MD, the electronic Schrödinger equation is solved at each time step of the dynamics, computing the forces for the present nuclear configuration (figure 2(b)). This can be computationally demanding in the case of a large number of QM atoms. Born–Oppenheimer MD simulations are commonly employed to characterize the electronic structure of protein/nucleic acid complexes [53, 54], and to perform a careful equilibration prior Car–Parrinello MD simulations.

Car–Parrinello MD introduced Newtonian fictitious dynamics for the electronic degrees of freedom through an extended Lagrangian with coupled equations of motion for both nuclei and electrons [55]. The extended Lagrangian (L_{ex}) includes the kinetic energy for the nuclei (T_{N}) and for the electronic degrees of freedom (T_{E}), as well as the potential energy (E_{pot})that depends on both the nuclear positions (R_I) and the electronic wave functions φ_i,

$$L_{\mathrm{ex}} = T_{\mathrm{N}} + T_{\mathrm{E}} - E_{\mathrm{pot}}. \tag{5}$$

L_{ex} is written as:

$$L_{\mathrm{ex}} = \sum_I \frac{1}{2} M R_I^2 + \sum_i \frac{1}{2}\mu|\varphi_i|^2 - \psi_0|H|\psi_0 + \sum_{i,j}\Lambda_{ij}\left[\left\{\int \varphi_i^*(r)\,\varphi_i(r)\,\mathrm{d}r\right\} - \mathrm{d}_{ij}\right] \tag{6}$$

where the Lagrange multipliers Λ_{ij} ensure the orthonormality of the wave functions φ_i and μ is a *'fictitious'* mass associated with the electronic degrees of freedom; and the potential energy is given by the expectation value of the total (ground state) energy $E = \psi_0|H|\psi_0$ of the system. This enables the electronic degrees of freedom to be treated as *'fictitious'* dynamic variables, which are propagated on the Born–Oppenheimer (i.e. the ground state) surface, without the need of a wave function optimization at each time step as in Born–Oppenheimer MD (figure 2(c)). Car–Parrinello MD has been almost exclusively used for first-principles studies of the CRISPR-Cas9 genome editor, while using Born–Oppenheimer MD to properly equilibrate the system.

Car–Parrinello MD simulations were used to characterize the conformation of the RuvC active site in the presence of Mg^{2+} ions. Early structural data reported the RuvC active site bound to Mn^{2+} ions. In the 4CMQ PDB structure, the D10, D986 and E762 carboxylates coordinate the two metal ions, with the H983 residue also coordinating one metal [9]. Car–Parrinello MD simulations of this structure were performed in the presence of the catalytic Mg^{2+} ions and in replicates of ~40 ps each (figure 3) [56]. As a result, the active site chemistry remarkably changed in the presence of the different metals, as also observed in *ab initio* studies of the CRISPR-associated protein 1, which intervenes in integrating the viral DNA into the bacterial genome [57]. In the Mg^{2+}–bound RuvC site, one water molecule stably locates between H983 and the scissile phosphate on the DNA (figure 3(a)). This suggested that H983 could act as an activator of the catalysis, a hypothesis that was supported by alanine mutations of H983 impeding non-target DNA cleavages [58]. The simulations also revealed that an arginine residue R976, located far away from the active site in the x-ray structure, approached the cleavage site, to stably bind the scissile phosphate (figures 3(a) and (b)). This 'arginine finger' was suggested to stabilize P_{SCI} for catalysis, in line with other Mg^{2+}-aided phosphatases [59]. Remarkably, recent structural work by Bravo and co-workers captured a high-resolution cryo-EM structure of RuvC right after NTS cleavage (PDB: 7S4X at 2.76 Å) [15], confirming that R976 repositions to bind P_{SCI}.

5. The 'blue moon ensemble' to characterize the catalytic mechanism in RuvC

To investigate chemical reactions, *ab-initio* MD is coupled with free energy methods to determine the activation free energy of the chemical step and, within transition state ($TS^{\ddagger}$) theory, the associated rate constant [60, 61]. A popular free energy method is the so-called *'blue moon ensemble'* approach, in association with **Thermodynamic Integration** [62, 63]. The 'blue moon' refers to the fact that chemical

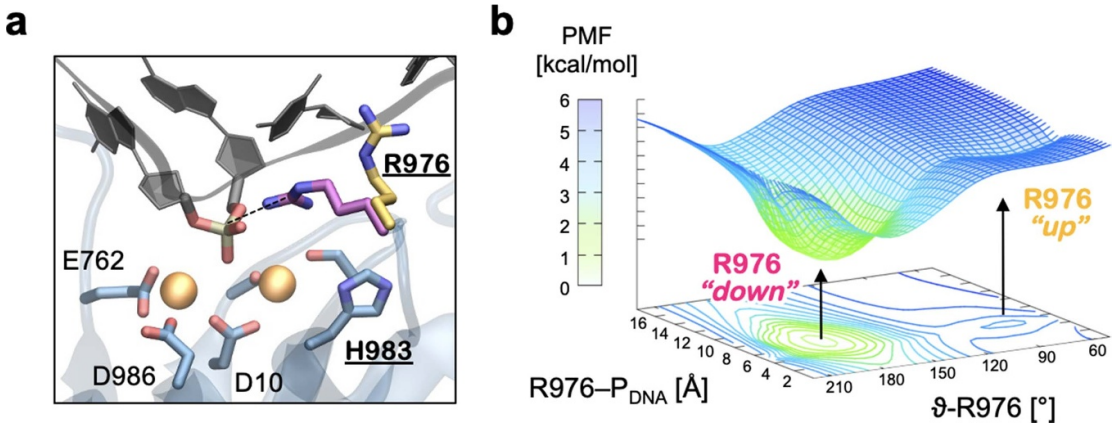

Figure 3. (a) Configuration of the RuvC catalytic site, as arising from QM/MM Car–Parrinello MD simulations (tree replicates of ∼40 ps each) [56]. These simulations consistently revealed that H983 positions in close proximity to the scissile phosphate to act as an activator of the nucleophile. (b) The simulations also showed that the R976 residue, located far away from the active site in the x-ray structure (R976 in yellow), approaches the cleavage site and stably binds the scissile phosphate (R976 in magenta). (c) Free energy landscape describing the conformational change of R976, showing that this residue reaches a stable minimum that corresponds to a 'down' conformation pointing toward the scissile phosphate. The free energy profile is plotted along two coordinates: (a) the distance between the Cζ atom of R976 and the scissile phosphate (R976–P_{DNA}) and (b) the dihedral angle between the Cα − Cβ − Cγ − Cδ atoms of R976 (ϑ–R976). Reprinted with permission from [56]. Copyright (2019) American Chemical Society.

reactions are rare events in the timescales that can be simulated through *ab-initio* MD (i.e. picoseconds), and thereby difficult to observe like a 'blue moon'. This method allows computing the free energy profile along a predefined reaction coordinate (*RC*). In detail, to explore a reaction step, a series of constraints are applied to enable exploration of the configurations along them. Then, the average converged constraint forces are computed and integrated along the given *RC* (ζ), deriving the associated free energy profile. In the case a simple distance (or difference in distances) is used as a *RC*, from the average constraint force *F* at each point along the *RC*, the free energy difference between two points ζ_2 and ζ_1 along the *RC* can be calculated as:

$$F(\zeta_2) - F(\zeta_1) = \int_{\zeta_1}^{\zeta_2} \mathrm{d}\zeta' \frac{\mathrm{d}F}{\mathrm{d}\zeta'} \tag{7}$$

where $F(\zeta_i)$ is the free energy at point ζ_i. This approach has been successfully employed in studies of phosphodiester bond cleavage in several RNA/DNA processing enzymes [54]. The choice of the *RC* is a critical step, since an inappropriate *RC* could lead to an unphysical description of the chemical step. The *RC* might not include all degrees of freedom relevant for the proceeding of the reaction, resulting in the exploration of the free energy surface (FES) along an incorrect pathway. This problem is limited for phosphodiester bond cleavage, where the attacking group linearly opposes the leaving group, and a reduced number of degrees of freedom enter into action. Several studies have shown that using as a simple *RC* the difference in distance between breaking and forming P–O bonds [64–68], one can achieve an unbiased representation of the FES for phosphodiester bond cleavage, obtaining also a fair agreement with the experimental rates.

The catalytic mechanism of non-target DNA cleavage in the RuvC active site was studied using the blue moon ensemble in association with thermodynamic integration (figure 4) [16].

Phosphodiester bond cleavage was studied along the difference in the distance between the breaking and forming P–O bonds (used as *RC*). As a result of these simulations, a S_N2-like associative mechanism was observed, with the critical role of H983. Indeed, H983 acts as a general base, abstracting a proton from the water nucleophile before the TS‡ and thereby activating the nucleophile. This mechanism is similar to what observed in other nucleases, using a histidine residue to activate the nucleophile [65, 69]. The joint dynamics of the two metal ions is critical for this mechanism. Indeed, the two Mg^{2+} ions come closer to each other while reaching the TS‡. In this way, they stabilize the TS‡, and bring together the reactant groups (i.e. leaving group, nucleophile and electrophile). This catalytic step was shown to proceed with an overall Helmholtz free energy ($\Delta F\#$) of ∼16.55 ± 1.22 kcal mol^{-1}, which is consistent with the experimental catalytic rate of 3.5 s^{-1} for the RuvC domain [70] that, employing TS‡ theory and assuming a transmission factor of unity, results in an activation barrier of ∼16/17 kcal mol^{-1}. Taken together, these results reconciled previous experimental evidences, establishing the catalytic role of the conserved H983 and the metal cluster conformation within the RuvC active site.

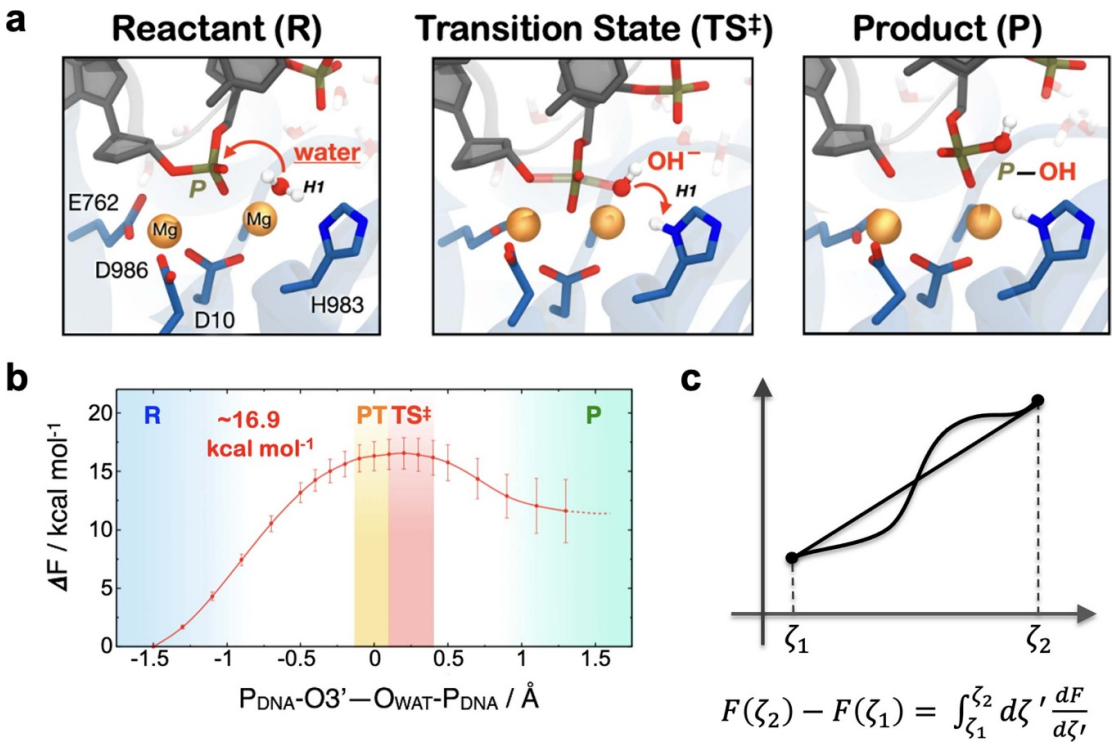

Figure 4. (a) QM/MM study of phosphodiester bond cleavage in the RuvC domain through a thermodynamic integration approach. (b) Free energy profile computed at the QM (DFT/BLYP [30, 31]) level of theory. The reaction is studied along the difference in the distance between the breaking and forming P–O bonds (P_{DNA}-O3′–O_{WAT}-P_{DNA}) used as reaction coordinate. The reaction proceeds from the reactant (R) to the product (P), passing though the transition state ($TS^{\ddagger}$) with an energetic barrier of ∼16.9 kcal mol^{-1}. The chemical step is activated by a proton transfer (PT) from the water nucleophile to H983. (c) Schematic diagram of a thermodynamic integration approach. The free energy difference between two points ζ_2 and ζ_1 is computed by integrating the average constraint forces at each point along the along the reaction coordinate. Reprinted with permission from [16]. Copyright (2020) American Chemical Society.

6. Metadynamics to establish the HNH catalysis

Another possible limitation of thermodynamic integration is that the FES might consist of numerous minima that are separated by barriers much larger than the thermal energies and multiple multidimensional reaction paths might contribute. The **metadynamics** approach is a free energy method that enables the dynamic study of the FES along multiple dimensions [71]. In metadynamics, an external history-dependent bias potential is added to the Hamiltonian of the system as a function of a set of predefined degrees of freedom (collective variables, *CVs*). This allows studying multiple slow processes at once, each described by a separated *CV*. Using metadynamics, one can reduce the complex multidimensional free energy space with a few *CVs*, while obtaining a multidimensional description. The general idea behind the metadynamics history-dependent bias is to enhance the system sampling by discouraging configurations that have already been visited (figure 5(a)). The Hamiltonian of the system, H, is thereby augmented with an external bias potential $V(\boldsymbol{S},\ t)$:

$$H = T + V + V(\boldsymbol{S},\ t)\,. \tag{8}$$

This external bias potential $V(\boldsymbol{S},\ t)$ is constructed as a sum of Gaussians, deposited during the simulation to act on a restricted number of degrees of freedom (CVs), $S(R) = S_1(\boldsymbol{R}) \ldots S_i(\boldsymbol{R})$. The total history dependent potential $V(\boldsymbol{S},\ t)$ acting on the system at time t is given by:

$$V(\boldsymbol{S},\ t) = \int_0^t \mathrm{d}t' \omega \exp\left(-\sum_{i=1}^{d} \frac{(S_i(\boldsymbol{R}) - S_i(\boldsymbol{R}(t'))^2}{2\sigma_i^2}\right) \tag{9}$$

where σ_i is the Gaussian width and (corresponding to the ith CV) and ω is the rate at which the bias potential grows. Several variations of this method led to well-tempered metadynamics [72], multiple walkers

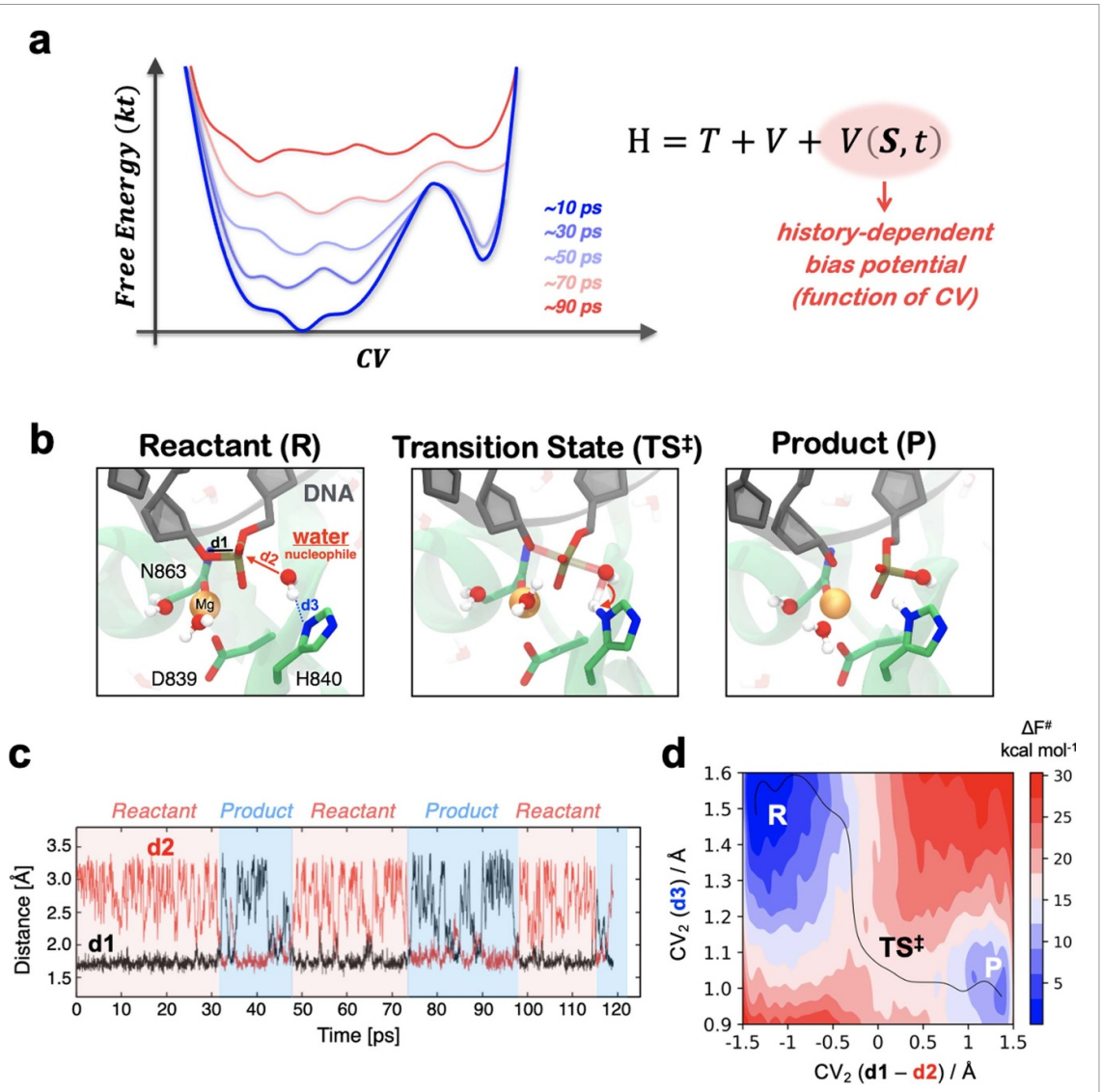

Figure 5. (a) Schematic diagram of a metadynamics approach. An external history-dependent bias potential $V(S,\ t)$ is added to the Hamiltonian H of the system as a function of predefined collective variables (CVs). As the simulation proceeds (i.e. from ∼10 to ∼90 ps), the free energy basin gets filled and the and the free energy surface can be recovered. (b) QM/MM study of phosphodiester bond cleavage in the HNH domain through metadynamics. The reaction is activated by H480, extracting a proton from the water nucleophile and leading to phosphodiester bind cleavage. (c) Time evolution along ∼120 ps of metadynamics of CV_1, i.e. the difference in distance between the breaking bond (d_1 = P_{DNA}–O3′) and forming bond (d_2 = O_{WAT}–P_{DNA}), showing that the simulation visits the reactant and product multiple times. (d) Two-dimensional free energy profile, describing phosphodiester bond cleavage in one dimension (the first collective variable, CV_1) and deprotonation of the water nucleophile in the other dimension (CV_2).

metadynamics [73], and histogram reweighted metadynamics [74], which improve the use of metadynamics for several applications. Metadynamics is widely employed in classical MD simulations for studies of ligand binding, and biophysical characterizations. The use of metadynamics with *ab-initio* MD requires particular attention. Indeed, using a standard metadynamics approach could result in boosting only the collective variable (as in a classical scheme), not coupling the system to the electronic degrees of freedom. To overcome this limitation an extended Lagrangian has been introduced [75]. Briefly, a new set of '*fictitious*' variables $s = \{s_i\}$ is introduced, where each s_i is associated to one of the selected collective variables S_i. The new variable s_i has a '*fictitious*' mass $\boldsymbol{M}_i$ and a velocity $\dot{s}_i$. The dynamics of s_i is then derived though an extended Lagrangian, where the '*fictitious*' kinetic energy and the potential energy are added as a function of s. This approach couples the dynamics of the collective variables to the electronic degrees of freedom, avoiding instabilities and heating of the system. This evades what experts commonly refer to an 'explosion' of the simulation.

Metadynamics was used to characterize the catalytic mechanism of target DNA cleavage in the active conformation of the HNH catalytic site (figure 5(b)) [76]. As noted above, the HNH domain displays high flexibility a complex conformational landscape, which made difficult for the early structural studies alone to

provide an unambiguous description of the active site chemistry. Computational studies thereby inferred structural information from the homologous enzyme endo T4 endonuclease VII [77], or were based on low resolution (i.e. 5.2 Å) cryo-EM structures [78].

However, recent breakthroughs in the structural biology of CRISPR-Cas9 paved the way for novel and unexplored catalytic hypotheses [13–15]. Starting from these new configurations of HNH, obtained at higher resolution, QM/MM metadynamics simulations were applied [76]. A two-dimensional FES was obtained, describing phosphodiester bond cleavage on one dimension (the first CV_1) and the deprotonation of the water nucleophile on the other (CV_2). Through a ∼120 ps metadynamics simulation, the chemical step was sampled from the reactants (R) to products (P) back and forth, which is essential to reconstruct a converged FES (figure 5(c)). As a result, an S_N2-like catalytic mechanism was observed, in which H840 extracts the water's proton before the $TS^{\ddagger}$, similar to what observed in RuvC. Notably, also in this case, the activation free energy for the chemical step was in line with the experimental catalytic rate measured for HNH, with a computed Helmholtz free energy ($\Delta F\#$) of 17.06 ± 1.22 kcal mol^{-1}, and an experimental catalytic rate of 4.3 s^{-1} (i.e. $\Delta G^{\ddagger}$ ∼16/17 kcal mol^{-1}) [70]. The catalytic mechanism also agreed with DNA cleavage experiments [79]. A critical difference that distinguishes HNH from RuvC is that HNH cuts the target DNA through a single Mg^{2+} ion, requiring the support of additional positive charges to stabilize the $TS^{\ddagger}$ [80]. This is attained by K866, which was shown to also engage in the protonation of the DNA O3′, leading target DNA cleavage to completion [76]. In summary, the use of metadynamics and first-principles MD simulations resolved the catalytic mechanism, and the conformation of responsible for target DNA cleavage in CRISPR-Cas9.

Overall, first-principles QM/MM MD simulations have shown to be instrumental in establishing the catalytic mechanism of DNA cleavage in CRISPR-Cas9, achieving a profound understanding of the Cas9 function and offering valuable insights for enzyme engineering. First-principles methods are thereby promising to investigate the catalysis of off-target DNA sequences, which limits the use of CRISPR-Cas9 for biomedical applications, and the catalytic role of alternative metal ions, which remains unmet. Additional free energy methods, such as transition path sampling [81], or the string method [82] could be harnessed in combination with *ab-initio* QM/MM MD to further explore the intricacies of the catalysis in CRISPR-Cas9. These methods are valuable additions to the 'arsenal' of QM/MM methods to study biochemical reactions. It is important to note, however, that the application of *ab-initio* MD for biological systems can be very challenging, especially when combined with complex free-energy methods. *Ab-initio* MD is notoriously very expensive under the computational point of view. It requires massive computational resources that only the world's most advanced HPC architectures can provide. For example, the investigation of the HNH catalysis required collecting an independent sampling of ∼700 ps of *ab-initio* MD. Using a state-of-the-art Intel cluster (i.e. Expanse at the San Diego Supercomputer Center) and using 120 cores, the performance for *ab-initio* MD of a CRISPR-Cas9 system including ∼150 QM atoms (on a total of ∼500 000 classical atoms) was ∼0.6/0.8 ps each day. This clearly shows that reaching appropriate sampling requires time and persistence, often resulting in studies that take two years (or more) of computer simulations. In this scenario, the use of the BLYP [30, 31] DFT functional for the description of the QM part allows sampling the FES in the most exhaustive way possible and in a timely fashion. However, the calculated FES might suffer by an underestimation that is intrinsic to the BLYP level. On the other hand, exhaustive sampling using a hybrid functional (e.g. B3LYP [31, 54]) might be prohibitive, due to the extremely high computational cost. To overcome this limitation, new frameworks for multiscale modeling are being developed. These include a multiscale modeling in computational chemistry approach [83], with high parallelization of both the QM and MM subsystems on HPC architectures. This recent breakthrough will hopefully reduce the computational cost for more accurate QM/MM simulations, implementing also multiple time step approaches.

7. Role of classical molecular dynamics simulations

Classical MD simulations are essential to properly equilibrate the system prior QM/MM simulations and to compute biochemical properties that are foundational to start *ab-initio* MD. A critical property for catalysis is certainly the pK_a of catalytic residues, which can clarify the protonation state of a residue and give an idea on how the reaction could proceed. Starting QM/MM simulations from an erroneous protonation state could lead to invalid catalytic mechanisms. It is thereby essential to establish the protonation state of reactive residues prior QM/MM simulations. Toward this goal, constant pH (CpH) MD simulations are a valuable approach [84]. In this method, the protonation state of an ionizable group can change during the simulation according to the local electrostatic environment and the pH of the solution. The protonation states can be periodically updated using Monte-Carlo sampling following MD steps [85–87], or through λ-dynamics by continuously propagating the motion of a virtual 'λ-particle' between different protonation Hamiltonians in

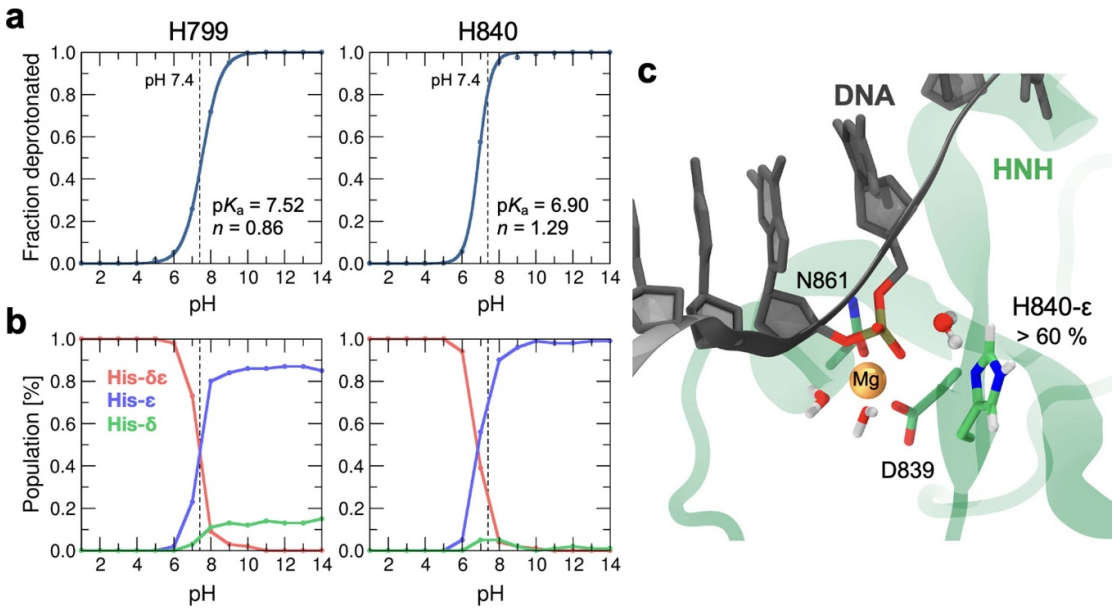

Figure 6. Titration curves (a) and population of states (b) for the H799 and H840 residues of the HNH domain of Cas9, computed though constant pH (CpH) MD simulations. (a) Titration curves (continuous lines) were obtained by fitting the simulation data (points) to equation (9). For each titration curve, the pK_a and the Hill coefficient n are reported. The Hill coefficient assumes values within the 0.5–1.5 range, indicating that the protonation states are properly sampled at each pH value [29]. (b) Population of histidine residues protonated (His-$\delta\varepsilon$, red) and in the two neutral tautomeric forms protonated on δ (His-δ, green) or on ε (His-ε, blue) is computed at each pH value from CpH MD simulations. At pH 7.4 (indicated using a dashed line), the catalytic H840 assumes the neutral tautomeric form protonated on the ε position for >60% of the simulation time. (c) Representative snapshot from CpH MD simulations, showing the HNH catalytic core with H840 protonated on the ε position.

explicit solvent simulations [88–91]. The pK_a is computed from the distributions of the protonation states using the Hill equation:

$$pK_a(i) = \mathrm{pH} - n \log \frac{x_i}{1 - x_i} \tag{10}$$

in which x_i is deprotonated fraction of residue i, and n is the Hill Coefficient. The titration curves can be derived by fitting the deprotonated fraction x_i to equation (9), using the Levenberg–Marquardt nonlinear optimization method:

$$f_{x_i} = \frac{1}{10^{n(pK_a - \mathrm{pH})} + 1}. \tag{11}$$

Good titration curves are characterized by small deviations of each point from the fitted titration curve (i.e. error from the fit) and Hill coefficients between 0.5 and 1.5, which indicates that the protonation states are properly sampled at the simulated pH values [29]. To determine the protonation state of the catalytic histidine residues in CRISPR-Cas9, explicit solvent CpH MD simulations were performed in conjunction with a replica exchange method [23] to enhance the sampling of the protonation states (figure 6).

These simulations were extensively performed, sampling for ∼40 ns at each pH value from 1 to 14, and obtaining an excellent agreement with the experimental pK_a, measured through nuclear magnetic resonance (NMR). Titration curves were built for the HNH catalytic residue H840 and for the neighboring H799, used as a control. The experimentally measured pK_a for H840 was ∼6.83, and ∼7.27 for H799. This indicates that, at pH 7.4, H840 is 57% protonated and H799 is 79% protonated. CpH MD simulations reported an excellent agreement with the NMR measurements, resulting in pK_a values of ∼6.90 and ∼7.52 for H840 and H799, respectively (figure 6(a)). Moreover, in-depth analysis of the protonation states also revealed that at pH 7.4, H840 is likely to be a neutral tautomer protonated on the ε position (H840-ε > 60% of the simulation; figure 6(b)). In this protonation state, the H840 δ nitrogen locates in proximity to the water nucleophile for its activation. This is a critical information, which enabled to start QM/MM simulations from the correct protonation state.

8. Conclusions

Here, we reviewed methods and applications of first-principles MD simulations to unravel the biochemical function of the CRISPR-Cas9 genome editing system. At the core of this technology, the endonuclease Cas9

performs double-stranded DNA cleavages using two catalytic domains. First-principles MD simulations have been instrumental in characterizing the catalytic mechanism of DNA cleavage, offering critical insights for the design of genome editing enzymes with improved catalytic efficiency.

Data availability statement

No new data were created or analyzed in this study.

Acknowledgments

This material is based upon work supported by the National Institute of Health (Grant No. R01GM141329, to GP) and the National Science Foundation (Grant Nos. CHE-1905374 and CHE-2144823, to GP). This work used Expanse at the San Diego Supercomputing Center through allocation MCB160059 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation Grants #2138259, #2138286, #2138307, #2137603, and #2138296. Computer time was also provided by NERSC under Grant Nos. Computer time was also provided by the National Energy Research Scientific Computing Center (NERSC) under Grant No. M3807.

Conflict of interest

The authors declare no competing interests.

ORCID iDs

Łukasz Nierzwicki https://orcid.org/0000-0001-9221-8848
Mohd Ahsan https://orcid.org/0000-0002-0583-4413
Giulia Palermo https://orcid.org/0000-0003-1404-8737

References

[1] Doudna J A 2020 The promise and challenge of therapeutic genome editing *Nature* **578** 229–36
[2] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J A and Charpentier E A 2012 Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity *Science* **337** 816–21
[3] Sternberg S H, Redding S, Jinek M, Greene E C and Doudna J A 2014 DNA interrogation by the CRISPR RNA-guided endonuclease Cas9 *Nature* **507** 62–67
[4] Jiang F and Doudna J A 2015 The structural biology of CRISPR-Cas systems *Curr. Opin. Struct. Biol.* **30** 100–11
[5] Nishimasu H and Nureki O 2017 Structures and mechanisms of CRISPR RNA-guided effector nucleases *Curr. Opin. Struct. Biol.* **43** 68–78
[6] Nishimasu H, Cong L, Yan W X, Ran F A, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F and Nureki O 2015 Crystal structure of Staphylococcus aureus Cas9 *Cell* **162** 1113–26
[7] Anders C, Niewoehner O, Duerst A and Jinek M 2014 Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease *Nature* **513** 569–73
[8] Hirano H *et al* 2016 Structure and engineering of Francisella novicida Cas9 *Cell* **164** 950–61
[9] Jinek M *et al* 2014 Structures of Cas9 endonucleases reveal RNA-mediated conformational activation *Science* **343** 1247997
[10] Jiang F, Zhou K, Ma L, Gressel S and Doudna J A 2015 Structural biology. A Cas9-guide RNA complex preorganized for target DNA recognition *Science* **348** 1477–81
[11] Jiang F, Taylor D W, Chen J S, Kornfeld J E, Zhou K, Thompson A J, Nogales E and Doudna J A 2016 Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage *Science* **351** 867–71
[12] Pacesa M *et al* 2022 Structural basis for Cas9 off-target activity *Cell* **185** 4067–81
[13] Pacesa M, Loeff L, Querques I, Muckenfuss L M, Sawicka M and Jinek M 2022 R-loop formation and conformational activation mechanisms of Cas9 *Nature* **609** 191–6
[14] Zhu X, Clarke R, Puppala A K, Chittori S, Merk A, Merrill B J, Simonović M and Subramaniam S 2019 Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9 *Nat. Struct. Mol. Biol.* **26** 679–85
[15] Bravo J P K, Liu M S, Hibshman G N, Dangerfield T L, Jung K, McCool R S, Johnson K A and Taylor D W 2022 Structural basis for mismatch surveillance by CRISPR–Cas9 *Nature* **603** 343–7
[16] Casalino L, Nierzwicki Ł, Jinek M and Palermo G 2020 Catalytic mechanism of non-target DNA cleavage in CRISPR-Cas9 revealed by *ab initio* molecular dynamics *ACS Catal.* **10** 13596–605
[17] Fu Y, Foden J A, Khayter C, Maeder M L, Reyon D, Joung J K and Sander J D 2013 High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells *Nat. Biotechnol.* **31** 822–6
[18] Nierzwicki Ł, Arantes P R, Saha A and Palermo G 2020 Establishing the allosteric mechanism in CRISPR-Cas9 *WIREs Comput. Mol. Sci.* **11** e1503
[19] Zuo Z and Liu J 2020 Allosteric regulation of CRISPR-Cas9 for DNA-targeting and cleavage *Curr. Opin. Struct. Biol.* **62** 166–74
[20] Chen J S and Doudna J A 2017 The chemistry of Cas9 and its CRISPR colleagues *Nat. Rev. Chem.* **1** 0078
[21] O'Connell M R, Oakes B L, Sternberg S H, East-Seletsky A, Kaplan M and Doudna J A 2014 Programmable RNA recognition and cleavage by CRISPR/Cas9 *Nature* **516** 263–6
[22] Sternberg S H, LaFrance B, Kaplan M and Doudna J A 2015 Conformational control of DNA target cleavage by CRISPR–Cas9 *Nature* **527** 110–3

[23] Dagdas Y S, Chen J S, Sternberg S H and Doudna J A 2017 A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9 *Sci. Adv.* **3** eaao002
[24] Chen J S, Dagdas Y S, Kleinstiver B P, Welch M M, Sousa A A, Harrington L B, Sternberg S H, Joung J K, Yildiz A and Doudna J A 2017 Enhanced proofreading governs CRISPR–Cas9 targeting accuracy *Nature* **550** 407–10
[25] Singh D, Sternberg S H, Fei J, Doudna J A and Ha T 2016 Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9 *Nat. Commun.* **7** 12778
[26] Singh D, Wang Y, Mallon J, Yang O, Fei J, Poddar A, Ceylan D, Bailey S and Ha T 2018 Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis *Nat. Struct. Mol. Biol.* **25** 347–54
[27] Newton M D, Taylor B J, Driessen R P C, Roos L, Cvetesic N, Allyjaun S, Lenhard B, Cuomo M E and Rueda D S 2019 DNA stretching induces Cas9 off-target activity *Nat. Struct. Mol. Biol.* **26** 185–92
[28] Palermo G, Miao Y, Walker R C, Jinek M and McCammon J A 2016 Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations *ACS Cent. Sci.* **2** 756–63
[29] Zuo Z and Liu J 2017 Structure and dynamics of Cas9 HNH domain catalytic state *Sci. Rep.* **7** 17271
[30] Saha A, Arantes P R and Palermo G 2022 Dynamics and mechanisms of CRISPR-Cas9 through the lens of computational methods *Curr. Opin. Struct. Biol.* **75** 102400
[31] Palermo G, Miao Y, Walker R C, Jinek M and McCammon J A 2017 CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations *Proc. Natl Acad. Sci.* **114** 7260–5
[32] Palermo G, Chen J S, Ricci C G, Rivalta I, Jinek M, Batista V S, Doudna J A and McCammon J A 2018 Key role of the REC lobe during CRISPR–Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain *Q. Rev. Biophys.* **51** e9
[33] Nierzwicki Ł, East K W, Morzan U N, Arantes P R, Batista V S, Lisi G P and Palermo G 2021 Enhanced specificity mutations perturb allosteric signaling in the CRISPR-Cas9 HNH endonuclease *eLife* **10** e73601
[34] East K W, Newton J C, Morzan U N, Narkhede Y B, Acharya A, Skeens E, Jogl G, Batista V S, Palermo G and Lisi G P 2020 Allosteric motions of the CRISPR–Cas9 HNH nuclease probed by NMR and molecular dynamics *J. Am. Chem. Soc.* **142** 1348–58
[35] Palermo G, Ricci C G, Fernando A, Basak R, Jinek M, Rivalta I, Batista V S and McCammon J A 2017 Protospacer adjacent motif-induced allostery activates CRISPR-Cas9 *J. Am. Chem. Soc.* **139** 16028–31
[36] Ricci C G, Chen J S, Miao Y, Jinek M, Doudna J A, McCammon J A and Palermo G 2019 Deciphering off-target effects in CRISPR-Cas9 through accelerated molecular dynamics *ACS Cent. Sci.* **5** 651–62
[37] Mitchell B P, Hsu R V, Medrano M A, Zewde N T, Narkhede Y B and Palermo G 2020 Spontaneous embedding of DNA mismatches within the RNA:DNA hybrid of CRISPR-Cas9 *Front. Mol. Biosci.* **7** 39
[38] Palermo G, Casalino L, Magistrato A and McCammon J A 2019 Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations *J. Struct. Biol.* **206** 267–79
[39] Nakamura T, Zhao Y, Yamagata Y, Hua Y and Yang W 2012 Watching DNA polymerase η make a phosphodiester bond *Nature* **487** 196–201
[40] Li P and Merz K M 2017 Metal ion modeling using classical mechanics *Chem. Rev.* **117** 1564–686
[41] Warshel A and Levitt M 1976 Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme *J. Mol. Biol.* **103** 227–49
[42] Marx D and Hutter J 2009 *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (New York: Cambridge University Press)
[43] Carloni P, Rothlisberger U and Parrinello M 2002 The role and perspective of *ab initio* molecular dynamics in the study of biological systems *Acc. Chem. Res.* **35** 455–64
[44] Burke K and Wagner L O 2013 DFT in a nutshell *Int. J. Quantum Chem.* **113** 96–101
[45] Brunk E *et al* 2011 Pushing the frontiers of first-principles based computer simulations of chemical and biological systems *Chim. Int. J. Chem.* **65** 667–71
[46] Gao J 1996 Hybrid quantum and molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry *Acc. Chem. Res.* **29** 298–305
[47] Senn H M and Thiel W 2006 *QM/MM Methods for Biological Systems in Atomistic Approaches in Modern Biology* (Berlin: Springer)
[48] Cheatham T E, Cieplak P and Kollman P A 1999 A modified version of the Cornell *et al*. Force field with improved sugar pucker phases and helical repeat *J. Biomol. Struct. Dyn.* **16** 845–62
[49] Perez A, Marchan I, Svozil D, Sponer J, Cheatham T E 3 R, Laughton C A and Orozco M 2007 Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers *Biophys. J.* **92** 3817–29
[50] Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham T E 3 R, Sponer J and Otyepka M 2010 Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins *J. Chem. Theor. Comput.* **6** 3836–49
[51] Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham T E and Jurecka P 2011 Refinement of the cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles *J. Chem. Theory Comput.* **7** 2886–902
[52] Laio A, VandeVondele J and Rothlisberger U 2002 A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations *J. Chem. Phys.* **116** 6941–7
[53] Palermo G, Stenta M, Cavalli A, Dal Peraro M and De Vivo M 2013 Molecular simulations highlight the role of metals in catalysis and inhibition of type II topoisomerase *J. Chem. Theory Comput.* **9** 857–62
[54] Palermo G, Cavalli A, Klein M L, Alfonso-Prieto M, Dal Peraro M and De Vivo M 2015 Catalytic metal ions and enzymatic processing of DNA and RNA *Acc. Chem. Res.* **48** 220–8
[55] Car R and Parrinello M 1985 Unified approach for molecular dynamics and density-functional theory *Phys. Rev. Lett.* **55** 2471–4
[56] Palermo G 2019 Structure and dynamics of the CRISPR–Cas9 catalytic complex *J. Chem. Inf. Model* **59** 2394–406
[57] Kumar A and Satpati P 2021 Divalent-metal-ion selectivity of the CRISPR-Cas system-associated Cas1 protein: insights from classical molecular dynamics simulations and electronic structure calculations *J. Phys. Chem.* B **125** 11943–54
[58] Nishimasu H, Ran F A, Hsu P D, Konermann S, Shehata S I, Dohmae N, Ishitani R, Zhang F and Nureki O 2014 Crystal structure of Cas9 in complex with guide RNA and target DNA *Cell* **156** 935–49
[59] Nagy G N, Suardíaz R, Lopata A, Ozohanics O, Vékey K, Brooks B R, Leveles I, Tóth J, Vértessy B G and Rosta E 2016 Structural characterization of arginine fingers: identification of an arginine finger for the pyrophosphatase DUTPases *J. Am. Chem. Soc.* **138** 15035–45
[60] Gao J, Ma S, Major D T, Nam K, Pu J and Truhlar D G 2006 Mechanisms and free energies of enzymatic reactions *Chem. Rev.* **106** 3188–209

[61] Garcia-Viloca M, Gao J, Karplus M and Truhlar D G 2004 How enzymes work: analysis by modern rate theory and computer simulations *Science* **303** 186–95
[62] Carter E A, Ciccotti G, Hynes J T and Kapral R 1989 Constrained reaction coordinate dynamics for the simulation of rare events *Chem. Phys. Lett.* **156** 472–7
[63] Ciccotti G, Ferrario M, Hynes J T and Kapral R 1989 Constrained molecular dynamics and the mean potential for an ion pair in a polar solvent *Chem. Phys.* **129** 241–51
[64] Rosta E, Nowotny M, Yang W and Hummer G 2011 Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations *J. Am. Chem. Soc.* **133** 8934–41
[65] Dürr S L, Bohuszewicz O, Berta D, Suardiaz R, Jambrina P G, Peter C, Shao Y and Rosta E 2021 The role of conserved residues in the DEDDh motif: the proton-transfer mechanism of HIV-1 RNase H *ACS Catal.* **11** 7915–27
[66] De Vivo M, Dal Peraro M and Klein M L 2008 Phosphodiester cleavage in ribonuclease h occurs via an associative two-metal-aided catalytic mechanism *J. Am. Chem. Soc.* **130** 10955–62
[67] Casalino L, Palermo G, Rothlisberger U and Magistrato A 2016 Who activates the nucleophile in ribozyme catalysis? An answer from the splicing mechanism of group II introns *J. Am. Chem. Soc.* **138** 10374–7
[68] Borišek J and Magistrato A 2020 All-atom simulations decrypt the molecular terms of RNA catalysis in the exon-ligation step of the spliceosome *ACS Catal.* **10** 5328–34
[69] Cisneros G A, Perera L, Schaaper R M, Pedersen L C, London R E, Pedersen L G and Darden T A 2009 Reaction mechanism of the ε subunit of *E. coli* DNA polymerase III: insights into active site metal coordination and catalytically significant residues *J. Am. Chem. Soc.* **131** 1550–6
[70] Gong S, Yu H H, Johnson K A and Taylor D W 2018 DNA unwinding is the primary determinant of CRISPR-Cas9 activity *Cell Rep.* **22** 359–71
[71] Laio A and Parrinello M 2002 Escaping free-energy minima *Proc. Natl Acad. Sci. USA* **99** 12562–6
[72] Barducci A, Bussi G and Parrinello M 2008 Well-tempered metadynamics: a smoothly converging and tunable free-energy method *Phys. Rev. Lett.* **100** 020603
[73] Raiteri P, Laio A, Gervasio F L, Micheletti C and Parrinello M 2006 Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics *J. Phys. Chem.* B **110** 3533–9
[74] Smiatek J and Heuer A 2011 Calculation of free energy landscapes: a histogram reweighted metadynamics approach *J. Comput. Chem.* **32** 2084–96
[75] Iannuzzi M, Laio A and Parrinello M 2003 Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics *Phys. Rev. Lett.* **90** 4
[76] Nierzwicki L *et al* 2022 Principles of target DNA cleavage and role of Mg2+ in the catalysis of CRISPR-Cas9 *Nat. Catal.* **5** 912–22
[77] Yoon H and Zhao L N 2019 Warshel A 2019 exploring the catalytic mechanism of Cas9 using information inferred from endonuclease VII *ACS Catal.* **9** 1329–36
[78] Zhao L N and Mondal D 2020 Warshel A 2020 exploring alternative catalytic mechanisms of the Cas9 HNH domain *Proteins* **88** 260–4
[79] Zuo Z, Zolekar A, Babu K, Lin V J, Hayatshahi H S, Rajan R, Wang Y-C and Liu J 2019 Structural and functional insights into the bona fide catalytic state of Streptococcus pyogenes Cas9 HNH nuclease domain *elife* **8** e46500
[80] Galburt E A and Stoddard B L 2002 Catalytic mechanisms of restriction and homing endonucleases *Biochemistry* **41** 13851–60
[81] Basner E J and Schwartz E S 2005 How enzyme dynamics helps catalyze a reaction in atomic detail: a transition path sampling study *J. Am. Chem. Soc.* **127** 13822–31
[82] Melo M C R *et al* 2018 NAMD goes quantum: an integrative suite for hybrid simulations *Nat. Methods* **15** 351–4
[83] Olsen J M H, Bolnykh V, Meloni S, Ippoliti E, Bircher M P, Carloni P and Rothlisberger U 2019 MiMiC: a novel framework for multiscale modeling in computational chemistry *J. Chem. Theory Comput.* **15** 3810–23
[84] Chen W, Morrow B H, Shi C and Shen J K 2014 Recent development and application of constant PH molecular dynamics *Mol. Simul.* **40** 830–8
[85] Baptista A M, Teixeira V H and Soares C M 2002 Constant-PH molecular dynamics using stochastic titration *J. Chem. Phys.* **117** 4184–200
[86] Mongan J, Case D A and McCammon J A 2004 Constant PH molecular dynamics in generalized Born implicit solvent *J. Comput. Chem.* **25** 2038–48
[87] Swails J M, York D M and Roitberg A E 2014 Constant PH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation *J. Chem. Theory Comput.* **10** 1341–52
[88] Goh G B, Knight J L and Brooks C L 2012 Constant PH molecular dynamics simulations of nucleic acids in explicit solvent *J. Chem. Theory Comput.* **8** 36–46
[89] Goh G B, Hulbert B S, Zhou H and Brooks C L 2014 Constant PH molecular dynamics of proteins in explicit solvent with proton tautomerism *Proteins Struct. Funct. Bioinform.* **82** 1319–31
[90] Wallace J A and Shen J K 2011 Continuous constant PH molecular dynamics in explicit solvent with PH-based replica exchange *J. Chem. Theory Comput.* **7** 2617–29
[91] Donnini S, Tegeler F, Groenhof G and Grubmüller H 2011 Constant PH molecular dynamics in explicit solvent with λ-dynamics. *J. Chem. Theory Comput.* **7** 1962–78