### Title

The electronic structure of genome editors from the first principles




### Journal

Electronic Structure, 5(1)

### ISSN

2516-1075



### Authors

Nierzwicki, Łukasz
Ahsan, Mohd
Palermo, Giulia



### Publication Date

2023-03-01

### DOI

10.1088/2516-1075/acb410

Peer reviewed

HHS Public Access

Author manuscript

*Electron Struct*. Author manuscript; available in PMC 2023 March 15.

Published in final edited form as:

*Electron Struct*. 2023 March ; 5(1): . doi:10.1088/2516-1075/acb410.

# The Electronic Structure of Genome Editors from the First Principles

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

[1]Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States

[2]Department of Chemistry, University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States

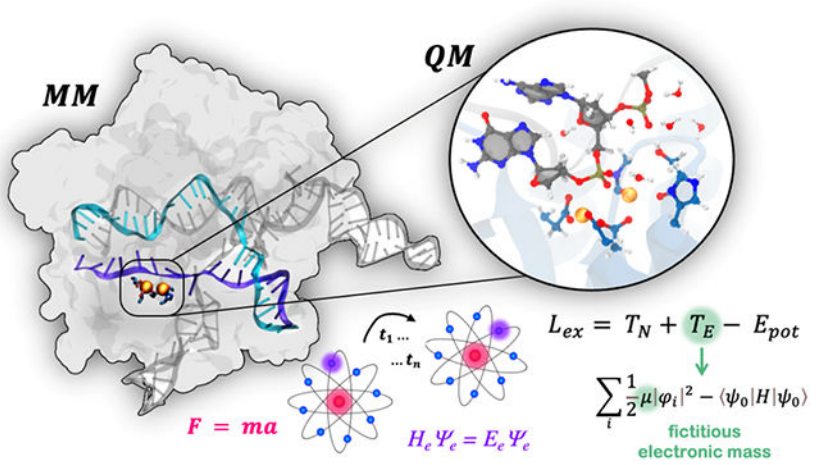
## Abstract

Genome editing based on the CRISPR-Cas9 system has paved new avenues for medicine, pharmaceutics, biotechnology, and beyond. This article reports the role of first-principles (ab-initio) molecular dynamics (MD) in the CRISPR-Cas9 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 characterize 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 of the Cas9 enzyme, providing valuable insights that can serve for the design of genome editing tools with improved catalytic efficiency or controllable activity.

## Graphical Abstract

Competing Interests

The authors declare no competing interests.




## 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 CRISPR-Cas9 system introducing the precise manipulation of nucleic acids.[2]

CRISPR (clustered regularly interspaced short palindromic repeats)-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, 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.

Intense structural studies of the CRISPR-Cas9 reported several crystallographic and 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 (Fig. 1a). HNH cleaves the DNA strand base-pairing the guide RNA, i.e., the so-called target strand (TS), 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]

## 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 (RNase H), with a two-metal dependent site, which mainly exploits $Mg^{2+}$ to perform phosphodiester bond cleavages (Fig. 1b). 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 (Fig. 1c). 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 quantum mechanics (QM) offers the opportunity to unravel the electronic structure deep inside, and to describe how chemical bonds form and

break. By empowering QM with molecular dynamics (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 quantum mechanics/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 (Fig. 2a).[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.

## 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 quantum mechanical 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 quantum mechanics/molecular mechanics (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 (Fig. 2a).[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 ($H_{QM/MM}$). This results in a single hybrid Hamiltonian, $H$:

$$H = H_{QM} + H_{MM} + H_{QM/MM} \quad (1)$$

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:

$$E^{ks}[\{\varnothing_i\}] = T_s[\{\varnothing_i\}] + \int dr\, V_{ext}(r)n(r) + \frac{1}{2}\int dr\, V_H(r)n(r) + E_{xc}[n] + E_{ions}(R^N) \quad (2)$$

where $\varnothing_i$ 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 (GGA) functionals. The last term accounts for the interaction energy due to nuclear charges.

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

$$H_{QMMM} = H^{bonds}_{QM/MM} + H^{vdW}_{QM/MM} + H^{elect}_{QM/MM} \quad (3)$$

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: (i) hydrogen atoms can be placed along the bond to complete the valency requirements of the QM atoms, (ii) a capping potential can be used to cap the boundary atoms, (iii) localized frozen orbitals can be used to define for the boundary atoms of the QM region. The Van der Waals ($H^{vdW}_{QM/MM}$) 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 ($H^{elect}_{QM/MM}$) between QM and MM regions are the most crucial term for the $H_{QM/MM}$ 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).

## 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:

$$H_e\Psi_e = E_e\Psi_e \tag{4}$$

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 (Fig. 2b). 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 $\varphi_i$.

$$L_{ex} = T_N + T_E - E_{pot} \tag{5}$$

$L_{ex}$ is written as:

$$\begin{aligned} L_{ex} = &\sum\nolimits_I \frac{1}{2} M R_I^2 + \sum\nolimits_i \frac{1}{2}\mu \mid \varphi_i \mid^2 - \langle \psi_0 \mid H \mid \psi_0 \rangle \\ &+ \sum\nolimits_{i,j} \Lambda_{ij} \left[ \{ \int \varphi_i^*(r)\varphi_i(r)dr \} - d_{i,j} \right] \end{aligned} \tag{6}$$

where the Lagrange multipliers $\Lambda_{ij}$ ensure the orthonormality of the wave functions $\varphi_i$ and $\mu$ 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 $E = \langle \psi_0 \mid H \mid \psi_0 \rangle$ 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 (Fig. 2c). 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 (Fig. 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 (Fig. 3a). 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 (Fig. 3a-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 $P_{SCI}$.

## The "Blue Moon Ensemble" to Study 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 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 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 $RC\,(\zeta)$, 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 $\zeta_2$ and $\zeta_1$ along the *RC* can be calculated as:

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

Where $F(\zeta_i)$ is the free energy at point $\zeta_i$. 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 (Fig. 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 transition state ($TS^{‡}$) 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 (ΔF#) 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 transition state 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.

## Metadynamics to Investigate 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 (Fig. 5a). 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 ($CVs$), $\boldsymbol{S}(\boldsymbol{R}) = \boldsymbol{S}_1(\boldsymbol{R})\ldots\boldsymbol{S}_i(\boldsymbol{R})$. The total history dependent potential $V(\boldsymbol{S}, t)$ acting on the system at time $t$ is given by:

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

where $\sigma_i$ is the Gaussian width and (corresponding to the $i^{th}$ $CV$) and $\omega$ is the rate at which the bias potential grows. Several variations of this method led to well-tempered metadynamics,[72] multiple walkers 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 $M_i$ 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 (Fig. 5b).[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 free energy surface (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 (Fig. 5c). As a result, an $S_N2$-like catalytic mechanism was observed, in which H840 extracts the water's proton before the transition state ($TS^{\ddagger}$), 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 (ΔF#) of 17.06 ± 1.22 kcal/mol, and an experimental catalytic rate of 4.3 s-1 (i.e., $\Delta G^{\ddagger}$ ~16/17 kcal/mol).[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 free energy surface (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 modelling are being developed. These include a multiscale modelling in computational chemistry (MiMiC) 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.

## Role of Classical MD 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 p$K_a$ 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 explicit solvent simulations.[88-91] The p$K_a$ is computed from the distributions of the protonation states using the Hill equation:

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

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:

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

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 (Fig. 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 p$K_a$, measured through NMR. Titration curves were built for the HNH catalytic residue H840 and for the neighboring H799, used as a control. The experimentally measured p$K_a$ 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 pKa values of ~6.90 and ~7.52 for H840 and H799, respectively (Fig. 6a). 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; Fig. 6b). 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.

## Conclusions

Here, we reviewed methods and applications of first-principles molecular dynamics 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 molecular dynamics 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.

## 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 No. 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 No. Computer time was also provided by the National Energy Research Scientific Computing Center (NERSC) under Grant No M3807.

## References


(1). Doudna JA The Promise and Challenge of Therapeutic Genome Editing 2020 Nature 578 229–236. [PubMed: 32051598]

(2). Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier EA 2012 Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity Science 337 816–821. [PubMed: 22745249]

(3). Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9 2014 Nature 507 62–67. [PubMed: 24476820]

(4). Jiang F, Doudna JA The Structural Biology of CRISPR-Cas Systems 2015 Curr. Opin. Struct. Biol 30 100–111. [PubMed: 25723899]

(5). Nishimasu H, Nureki O Structures and Mechanisms of CRISPR RNA-Guided Effector Nucleases 2017 Curr. Opin. Struct. Biol 43 68–78. [PubMed: 27912110]

(6). Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O Crystal Structure of Staphylococcus Aureus Cas9. 2015 Cell 162, 1113–1126. [PubMed: 26317473]

(7). Anders C, Niewoehner O, Duerst A, Jinek M Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease. 2014 Nature 513, 569–573. [PubMed: 25079318]

(8). Hirano H, Gootenberg JS, Horii T, Abudayyeh OO, Kimura M, Hsu PD, Nakane T, Ishitani R, Hatada I, Zhang F, Nishimasu H, Nureki O Structure and Engineering of Francisella Novicida Cas9. 2016 Cell 164, 950–961. [PubMed: 26875867]

(9). Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA 2014 Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation Science 343 1247997–1247997 [PubMed: 24505130]

(10). Jiang F, Zhou K, Ma L, Gressel S, Doudna JA STRUCTURAL BIOLOGY. A Cas9-Guide RNA Complex Preorganized for Target DNA Recognition. 2015 Science 348, 1477–1481. [PubMed: 26113724]

(11). 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 2016 Science 351 867–871. [PubMed: 26841432]

(12). Pacesa M, Lin C-H, Cléry A, Saha A, Arantes P, Bargsten K, Irby M, Allain F, Palermo G, Cameron P, Donohoue P, Jinek M, Structural Basis for Cas9 Off-Target Activity. 2022 Cell, 185, 4067–4081. [PubMed: 36306733]

(13). Pacesa M, Loeff L, Querques I, Muckenfuss LM, Sawicka M, Jinek M 2022 R-Loop Formation and Conformational Activation Mechanisms of Cas9 Nature 609 191–196. [PubMed: 36002571]

(14). 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. 2019 Nat. Struct. Mol. Biol 26, 679–685. [PubMed: 31285607]

(15). Bravo JPK, Liu MS, Hibshman GN, Dangerfield TL, Jung K, McCool RS, Johnson KA, Taylor DW Structural Basis for Mismatch Surveillance by CRISPR–Cas9 2022 Nature 603 343–347. [PubMed: 35236982]

(16). Casalino L, Nierzwicki Ł, Jinek M, Palermo G Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics 2020 ACS Catal. 10 13596–13605. [PubMed: 33520346]

(17). Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD 2013 High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells Nat. Biotechnol 31 822–826. [PubMed: 23792628]

(18). Nierzwicki Ł, Arantes PR, Saha A, Palermo G 2020 Establishing the Allosteric Mecanism in CRISPR-Cas9 WIREs Comput Mol Sci 11 e1503.

(19). Zuo Z, Liu J 2020 Allosteric Regulation of CRISPR-Cas9 for DNA-Targeting and Cleavage Curr. Opin. Struct. Biol 62 166–174. [PubMed: 32070859]

(20). Chen JS, Doudna JA 2017 The Chemistry of Cas9 and Its CRISPR Colleagues Nat. Rev. Chem 1 0078.

(21). O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA 2014 Programmable RNA Recognition and Cleavage by CRISPR/Cas9 Nature 516 263–266. [PubMed: 25274302]

(22). Sternberg SH, LaFrance B, Kaplan M, Doudna JA 2015 Conformational Control of DNA Target Cleavage by CRISPR–Cas9 Nature 527 110–113. [PubMed: 26524520]

(23). Dagdas YS, Chen JS, Sternberg SH, Doudna JA 2017 A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR-Cas9 Sci. Adv 3 eaao002.

(24). Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA 2017 Enhanced Proofreading Governs CRISPR–Cas9 Targeting Accuracy Nature 550 407–410. [PubMed: 28931002]

(25). Singh D, Sternberg SH, Fei J, Doudna JA, Ha T 2016 Real-Time Observation of DNA Recognition and Rejection by the RNA-Guided Endonuclease Cas9 Nat Commun 7 12778. [PubMed: 27624851]

(26). Singh D, Wang Y, Mallon J, Yang O, Fei J, Poddar A, Ceylan D, Bailey S, Ha T 2018 Mechanisms of Improved Specificity of Engineered Cas9s Revealed by Single-Molecule FRET Analysis Nat. Struct. Mol. Biol 25 347–354. [PubMed: 29622787]

(27). Newton MD, Taylor BJ, Driessen RPC, Roos L, Cvetesic N, Allyjaun S, Lenhard B, Cuomo ME, Rueda DS 2019 DNA Stretching Induces Cas9 Off-Target Activity. Nat. Struct. Mol. Biol 26 185–192. [PubMed: 30804513]

(28). Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA 2016 Striking Plasticity of CRISPR-Cas9 and Key Role of Non-Target DNA, as Revealed by Molecular Simulations ACS Cent. Sci 2 756–763. [PubMed: 27800559]

(29). Zuo Z, Liu J 2017 Structure and Dynamics of Cas9 HNH Domain Catalytic State Sci. Rep 7 17271. [PubMed: 29222528]

(30). Saha A, Arantes PR, Palermo G 2022 Dynamics and Mechanisms of CRISPR-Cas9 through the Lens of Computational Methods Curr. Opin. Struct. Biol 75 102400. [PubMed: 35689914]

(31). Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA 2017 CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations Proc. Natl. Acad. Sci 114 7260–7265. [PubMed: 28652374]

(32). Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, Doudna JA, McCammon JA 2018 Key Role of the REC Lobe during CRISPR–Cas9 Activation by 'Sensing', 'Regulating', and 'Locking' the Catalytic HNH Domain Q. Rev. Biophys 51 e9.

(33). Nierzwicki Ł, East KW, Morzan UN, Arantes PR, Batista VS, Lisi GP, Palermo G 2021 Enhanced Specificity Mutations Perturb Allosteric Signaling in the CRISPR-Cas9 HNH Endonuclease eLife 10 e73601. [PubMed: 34908530]

(34). East KW, Newton JC, Morzan UN, Narkhede YB, Acharya A, Skeens E, Jogl G, Batista VS, Palermo G, Lisi GP 2020 Allosteric Motions of the CRISPR–Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics J. Am. Chem. Soc 142 1348–1358. [PubMed: 31885264]

(35). Palermo G, Ricci CG, Fernando A, Basak R, Jinek M, Rivalta I, Batista VS, McCammon JA 2017 Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9 J. Am. Chem. Soc 139 16028–16031. [PubMed: 28764328]

(36). Ricci CG, Chen JS, Miao Y, Jinek M, Doudna JA, McCammon JA, Palermo G 2019 Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics ACS Cent. Sci 5 651–662. [PubMed: 31041385]

(37). Mitchell BP, Hsu RV, Medrano MA, Zewde NT, Narkhede YB, Palermo G 2020 Spontaneous Embedding of DNA Mismatches Within the RNA:DNA Hybrid of CRISPR-Cas9 Front. Mol. Biosci 7 39. [PubMed: 32258048]

(38). Palermo G, Casalino L, Magistrato A, McCammon JA Understanding the Mechanistic Basis of Non-Coding RNA through Molecular Dynamics Simulations. 2019 J. Struct. Biol 206, 267–279. [PubMed: 30880083]

(39). Nakamura T, Zhao Y, Yamagata Y, Hua Y, Yang W 2012 Watching DNA Polymerase η Make a Phosphodiester Bond Nature 487 196–201. [PubMed: 22785315]

(40). Li P, Merz KM 2017 Metal Ion Modeling Using Classical Mechanics Chem. Rev 117 1564–1686. [PubMed: 28045509]

(41). Warshel A, 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 103 227–249. [PubMed: 985660]

(42). Marx D, Hutter J 2009 Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods (New York: Cambridge Univ. Press)

(43). Carloni P, Rothlisberger U, Parrinello M 2002 The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems Acc. Chem. Res 35 455–464. [PubMed: 12069631]

(44). Burke K, Wagner LO 2013 DFT in a Nutshell Int. J. Quantum Chem 113 96–101.

(45). Brunk E, Ashari N, Athri P, Campomanes P, de Carvalho FF et al. 2011 Pushing the Frontiers of First-Principles Based Computer Simulations of Chemical and Biological Systems Chim. Int. J. Chem 65 667–671.

(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 HM, Thiel W 2006 QM/MM Methods for Biological Systems In Atomistic Approaches in Modern Biology (Berlin Heidelberg: Springer).

(48). Cheatham TE, Cieplak P, Kollman PA 1999 A Modified Version of the Cornell et Al. Force Field with Improved Sugar Pucker Phases and Helical Repeat J. Biomol. Struct. Dyn 16 845–862. [PubMed: 10217454]

(49). Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE 3rd, Laughton CA, Orozco M 2007 Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers Biophys. J 92 3817–3829. [PubMed: 17351000]

(50). Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham TE 3rd, Sponer J, Otyepka M 2010 Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins J. Chem. Theor. Comput 6 3836–3849.

(51). Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham TE, 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 2886–2902. [PubMed: 21921995]

(52). Laio A, VandeVondele J, Rothlisberger U 2002 A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car–Parrinello Molecular Dynamics Simulations J. Chem. Phys 116 6941–6947.

(53). Palermo G, Stenta M, Cavalli A, Dal Peraro M, De Vivo M 2013 Molecular Simulations Highlight the Role of Metals in Catalysis and Inhibition of Type II Topoisomerase J. Chem. Theory Comput 9 857–862. [PubMed: 26588728]

(54). Palermo G, Cavalli A, Klein ML, Alfonso-Prieto M, Dal Peraro M, De Vivo M 2015 Catalytic Metal Ions and Enzymatic Processing of DNA and RNA Acc. Chem. Res 48 220–228. [PubMed: 25590654]

(55). Car R, Parrinello M 1985 Unified Approach for Molecular Dynamics and Density-Functional Theory Phys. Rev. Lett 55 2471–2474. [PubMed: 10032153]

(56). Palermo G 2019 Structure and Dynamics of the CRISPR–Cas9 Catalytic Complex J. Chem. Inf. Model 59 2394–2406. [PubMed: 30763088]

(57). Kumar A, 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, 11943–11954. [PubMed: 34694813]

(58). 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. 2014 Cell, 156, 935–949. [PubMed: 24529477]

(59). Nagy GN, Suardíaz R, Lopata A, Ozohanics O, Vékey K, Brooks BR, Leveles I, Tóth J, Vértessy BG, Rosta E 2016 Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase DUTPases J. Am. Chem. Soc 138 15035–15045. [PubMed: 27740761]

(60). Gao J, Ma S, Major DT, Nam K, Pu J, Truhlar DG 2006 Mechanisms and Free Energies of Enzymatic Reactions. Chem. Rev 106 3188–3209. [PubMed: 16895324]

(61). Garcia-Viloca M, Gao J, Karplus M, Truhlar DG 2004 How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. Science, 303 186–195. [PubMed: 14716003]

(62). Carter EA, Ciccotti G, Hynes JT, Kapral R 1989 Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events Chem. Phys. Lett 156 472–477.

(63). Ciccotti G, Ferrario M, Hynes JT, Kapral R 1989 Constrained Molecular Dynamics and the Mean Potential for an Ion Pair in a Polar Solvent Chem. Phys 129 241–251.

(64). Rosta E, Nowotny M, Yang W, Hummer G 2011 Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from Quantum Mechanics/Molecular Mechanics Simulations J. Am. Chem. Soc 133 8934–8941. [PubMed: 21539371]

(65). Dürr SL, Bohuszewicz O, Berta D, Suardiaz R, Jambrina PG, Peter C, Shao Y, Rosta E 2021 The Role of Conserved Residues in the DEDDh Motif: The Proton-Transfer Mechanism of HIV-1 RNase H ACS Catal. 11 7915–7927.

(66). De Vivo M, Dal Peraro M, Klein ML 2008 Phosphodiester Cleavage in Ribonuclease H Occurs via an Associative Two-Metal-Aided Catalytic Mechanism J. Am. Chem. Soc 130 10955–10962. [PubMed: 18662000]

(67). Casalino L, Palermo G, Rothlisberger U, 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 10374–10377. [PubMed: 27309711]

(68). Borišek J, Magistrato A 2020 All-Atom Simulations Decrypt the Molecular Terms of RNA Catalysis in the Exon-Ligation Step of the Spliceosome ACS Catal. 10 5328–5334.

(69). Cisneros GA, Perera L, Schaaper RM, Pedersen LC, London RE, Pedersen LG, Darden TA 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 1550–1556. [PubMed: 19119875]

(70). Gong S, Yu HH, Johnson KA, Taylor DW 2018 DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. Cell Rep. 22 359–371. [PubMed: 29320733]

(71). Laio A, Parrinello M 2002 Escaping Free-Energy Minima Proc. Natl. Acad. Sci. U. S. A 99 12562–12566. [PubMed: 12271136]

(72). Barducci A, Bussi G, Parrinello M 2008 Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. Phys. Rev. Lett 100 020603. [PubMed: 18232845]

(73). Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M 2006 Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics J. Phys. Chem. B 110 3533–3539. [PubMed: 16494409]

(74). Smiatek J, Heuer A 2011 Calculation of Free Energy Landscapes: A Histogram Reweighted Metadynamics Approach J. Comput. Chem 32 2084–2096. [PubMed: 21500219]

(75). Iannuzzi M, Laio A, Parrinello M 2003 Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics Phys. Rev. Lett 90 4.

(76). Nierzwicki L, East KW, Binz MB, Hsu RV, Ahsan M, Arantes PR, Skeens E, Pacesa M, Jinek M, Lisi GP, Palermo G 2022 Principles of Target DNA Cleavage and Role of Mg2+ in the Catalysis of CRISPR-Cas9 Nat. Catal *in press*.

(77). Yoon H, Zhao LN, Warshel A 2019 Exploring the Catalytic Mechanism of Cas9 Using Information Inferred from Endonuclease VII ACS Catal. 9 1329–1336. [PubMed: 34046245]

(78). Zhao LN, Mondal D, Warshel A 2020 Exploring Alternative Catalytic Mechanisms of the Cas9 HNH Domain Proteins 88 260–264. [PubMed: 31390092]

(79). Zuo Z, Zolekar A, Babu K, Lin VJ, Hayatshahi HS, Rajan R, Wang Y-C, Liu J. 2019 Structural and Functional Insights into the Bona Fide Catalytic State of Streptococcus Pyogenes Cas9 HNH Nuclease Domain. Elife 8:e46500. [PubMed: 31361218]

(80). Galburt EA, Stoddard BL 2002 Catalytic Mechanisms of Restriction and Homing Endonucleases Biochemistry 41 13851–13860. [PubMed: 12437341]

(81). Basner EJ, Schwartz ES How Enzyme Dynamics Helps Catalyze a Reaction in Atomic Detail: A Transition Path Sampling Study. 2005 J. Am. Chem. Soc 127, 13822–13831. [PubMed: 16201803]

(82). Melo MCR, Bernardi RC, Rudack T, Scheurer M, Riplinger C, Phillips JC, Maia JDC, Rocha GB, Ribeiro JV, Stone JE, Neese F, Schulten K, Luthey-Schulten Z NAMD Goes Quantum: An Integrative Suite for Hybrid Simulations. 2018 Nat. Methods 15, 351–354. [PubMed: 29578535]

(83). Olsen JMH, Bolnykh V, Meloni S, Ippoliti E, Bircher MP, Carloni P, Rothlisberger Un2019 MiMiC: A Novel Framework for Multiscale Modeling in Computational Chemistry J. Chem. Theory Comput 15 3810–3823. [PubMed: 30998344]

(84). Chen W, Morrow BH, Shi C, Shen JK 2014 Recent Development and Application of Constant PH Molecular Dynamics Mol. Simul 40 830–838. [PubMed: 25309035]

(85). Baptista AM, Teixeira VH, Soares CM 2002 Constant-PH Molecular Dynamics Using Stochastic Titration. J. Chem. Phys 117 4184–4200.

(86). Mongan J, Case DA, McCammon JA 2004 Constant PH Molecular Dynamics in Generalized Born Implicit Solvent J. Comput. Chem 25 2038–2048. [PubMed: 15481090]

(87). Swails JM, York DM, Roitberg AE 2014 Constant PH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation J. Chem. Theory Comput 10 1341–1352. [PubMed: 24803862]

(88). Goh GB, Knight JL, Brooks CL 2012 Constant PH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent J. Chem. Theory Comput 8 36–46. [PubMed: 22337595]

(89). Goh GB, Hulbert BS, Zhou H, Brooks CL 2014 Constant PH Molecular Dynamics of Proteins in Explicit Solvent with Proton Tautomerism Proteins Struct. Funct. Bioinforma 82 1319–1331.

(90). Wallace JA, Shen JK 2011 Continuous Constant PH Molecular Dynamics in Explicit Solvent with PH-Based Replica Exchange J. Chem. Theory Comput 7 2617–2629. [PubMed: 26606635]

(91). Donnini S, Tegeler F, Groenhof G, Grubmüller H 2011 Constant PH Molecular Dynamics in Explicit Solvent with λ-Dynamics J. Chem. Theory Comput 7 1962–1978. [PubMed: 21687785]

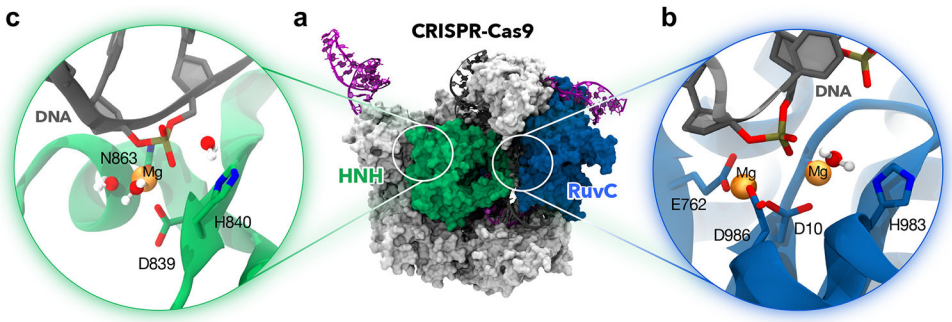



**Figure 1.**

**(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.

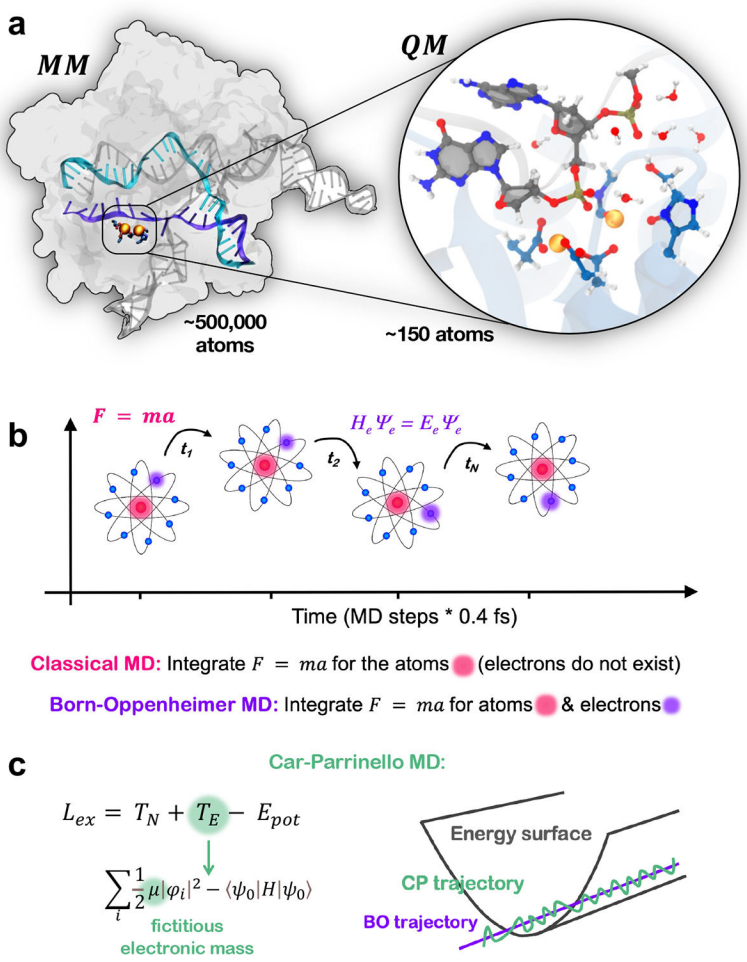



**Figure 2. First-principles Quantum Mechanics/Molecular Mechanics (QM/MM) approaches.**

**(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. $\zeta_2$ includes the kinetic energy for the nuclei $\zeta_1$ and for the electronic degrees of freedom ($T_N$), as well as the potential energy $L_{ex}$. 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.

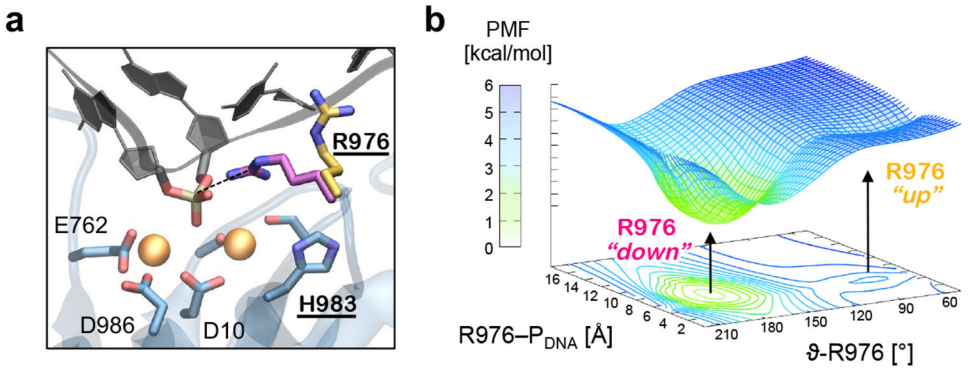



**Figure 3.**

**(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: (i) the distance between the Cζ atom of R976 and the scissile phosphate (R976–$P_{DNA}$) and (ii) the dihedral angle between the Cα—Cβ—Cγ—Cδ atoms of R976 (ϑ-R976). Adapted with permission from Palermo G. (2019).[56] Copyright 2019 American Chemical Society, https://pubs.acs.org/doi/10.1021/acs.jcim.8b00988

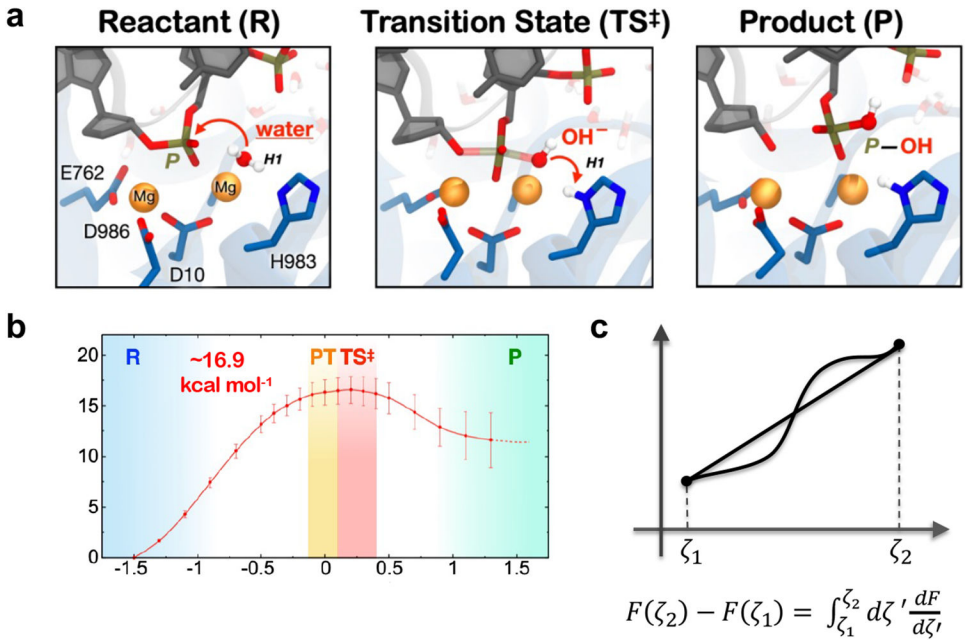



**Figure 4.**

**(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 ($TS^{‡}$) 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 ($T_E$) and ($E_{pot}$) is computed by integrating the average constraint forces at each point along the along the reaction coordinate. Adapted with permission from Casalino L. et al. (2020).[16] Copyright 2020 American Chemical Society, https://pubs.acs.org/doi/10.1021/acscatal.0c03566

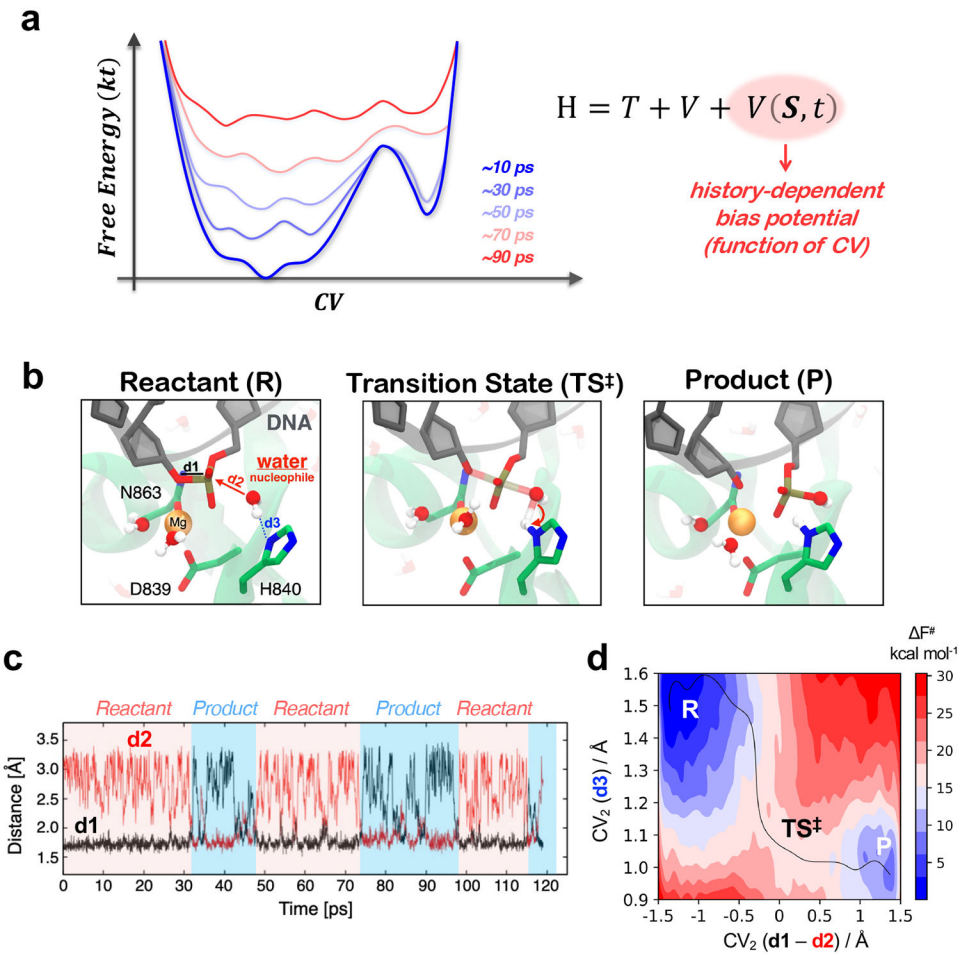



**Figure 5.**

**(a)** Schematic diagram of a metadynamics approach. An external history-dependent bias potential $V(\boldsymbol{S},t)$ 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 ($CV_2$).

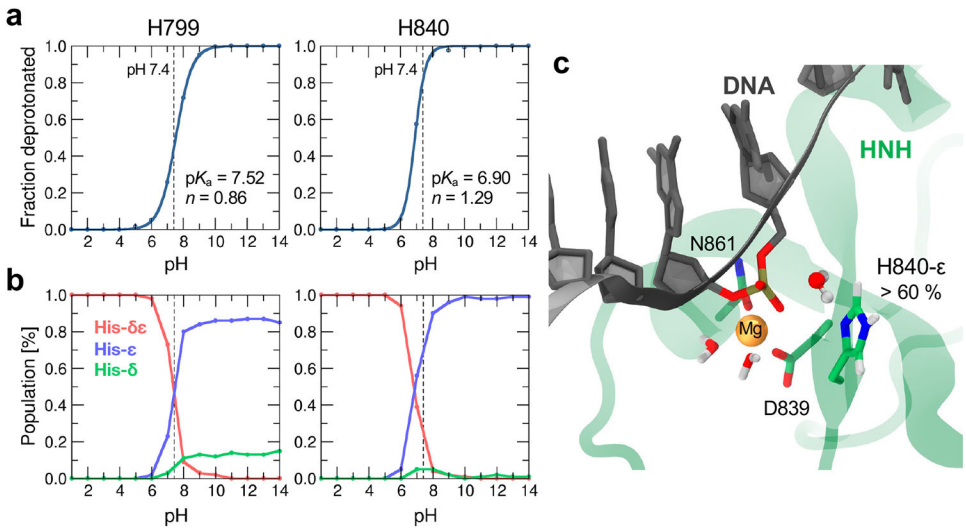



**Figure 6.**

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 p$K_a$ 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-δε, 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.