nature catalysis

Article

https://doi.org/10.1038/s41929-022-00848-6

Principles of target DNA cleavage and the role of Mg^{2+} in the catalysis of CRISPR–Cas9

Received: 6 February 2022

Accepted: 25 August 2022

Published online: 06 October 2022

Check for updates

Łukasz Nierzwicki[1], Kyle W. East[2], Jonas M. Binz[3], Rohaine V. Hsu[1], Mohd Ahsan[1], Pablo R. Arantes[1], Erin Skeens[2], Martin Pacesa[3], Martin Jinek[3], George P. Lisi[2] ✉ and Giulia Palermo[1,4] ✉

At the core of the CRISPR–Cas9 genome-editing technology, the endonuclease Cas9 introduces site-specific breaks in DNA. However, precise mechanistic information to ameliorate Cas9 function is still missing. Here, multimicrosecond molecular dynamics, free energy and multiscale simulations are combined with solution NMR and DNA cleavage experiments to resolve the catalytic mechanism of target DNA cleavage. We show that the conformation of an active HNH nuclease is tightly dependent on the catalytic Mg^{2+}, unveiling its cardinal structural role. This activated Mg^{2+}-bound HNH is consistently described through molecular simulations, nuclear magnetic resonance (NMR) and DNA cleavage assays, revealing also that the protonation state of the catalytic H840 is strongly affected by active site mutations. Finally, ab initio quantum mechanics (density functional theory)/molecular mechanics simulations and metadynamics establish the catalytic mechanism, showing that the catalysis is activated by H840 and completed by K866, thus rationalizing DNA cleavage experiments. This information is critical to enhancing the enzymatic function of CRISPR–Cas9 towards improved genome editing.

CRISPR–Cas9 is a genome-editing tool that has revolutionized basic and applied sciences[1]. At the core of this technology, the endonuclease Cas9 associates with a guide RNA structure to recognize and cleave matching sequences of DNA[2]. DNA targeting starts with the recognition of a short protospacer adjacent motif, initiating double strand separation, whereby one strand (the target strand) base pairs with the guide RNA to form an RNA:DNA heteroduplex, while the other non-target strand is displaced. Then, two catalytic domains, HNH and RuvC, catalyse cleavage of the target and non-target strands of DNA, respectively (Fig. 1a). In this complex biophysical process, the nuclease function of Cas9 is an essential step, yet its characterization has remained incomplete.

Biochemical and computational studies have revealed that RuvC cleaves the DNA non-target strand through a two-metal (Mg^{2+})-dependent catalytic mechanism[3–5]. In contrast, the catalysis of target strand cleavage by the HNH nuclease is not understood. Biochemical studies indicated that a single Mg^{2+} ion catalyses phosphodiester bond cleavage[3]. However, HNH displays a complex conformational landscape, which regulates the process of DNA binding and culminates in the docking of HNH at the cleavage site on the target strand[6–8]. While this is a precondition for target strand cleavage, two distinct states were reported displaying HNH in close proximity to the cleavage site. Homology with the T4 endonuclease VII (endo VII)[9] suggested that D861 and D839 coordinate the catalytic Mg^{2+} and form a triad with the catalytic H840 (Fig. 1b)[6,8,10]. This conformation was supported by structures capturing HNH in various non-catalytic states and in the absence of Mg^{2+} ions (Supplementary Fig. 1)[3,4,11,12]. A cryo-electron microscopy (cryo-EM) structure of a catalytically dead Cas9 also supported this conformation of HNH[13]. This structure (Protein Data Bank (PDB) 5Y36), solved at 5.2 Å resolution, was obtained including Mg^{2+} ions in the experimental buffer, but the location of Mg^{2+}

[1]Department of Bioengineering, University of California Riverside, Riverside, CA, USA. [2]Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA. [3]Department of Biochemistry, University of Zürich, Zurich, Switzerland. [4]Department of Chemistry, University of California Riverside, Riverside, CA, USA. ✉e-mail: george_lisi@brown.edu; giulia.palermo@ucr.edu

https://doi.org/10.1038/s41929-022-00848-6

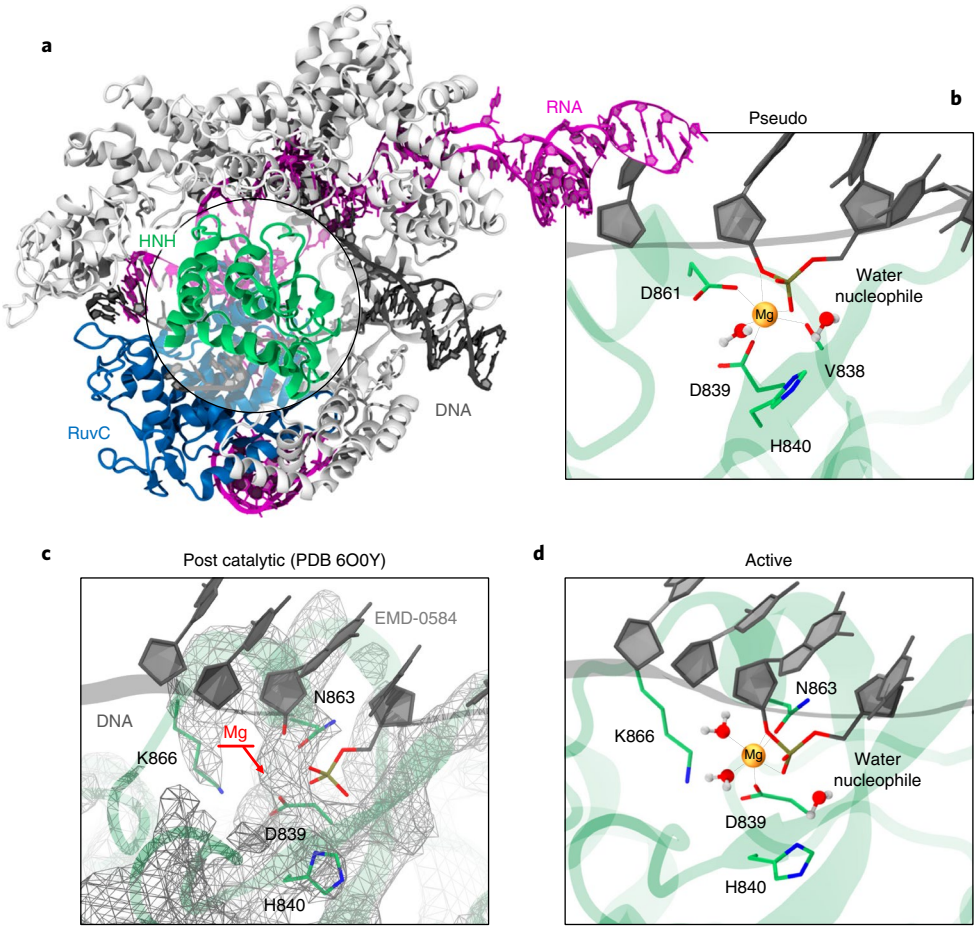

Fig. 1 | Overview of the *Streptococcus pyogenes* (Sp) CRISPR–Cas9 system. **a**, X-ray structure of the CRISPR–Cas9 system (PDB 5F9R)[12]. The Cas9 protein is shown as ribbons, highlighting its catalytic domains HNH (green) and RuvC (blue), in complex with RNA (magenta) and DNA (black). **b**, Close-up view on the HNH catalytic site, displaying the D839 and D831 residues coordinating Mg^{2+} and forming a catalytic triad with H840. This configuration of the catalytic core (referred to as pseudo-active) arises from structures capturing HNH in the absence of Mg^{2+} (for example, PDB 5F9R) and from homology with the T4 endonuclease VII (ref. [9]) (Supplementary Fig. 1). **c**,**d**, Catalytic core from the cryo-EM structure EMD-0584 (PDB 6O0Y)[14] capturing HNH in the presence of Mg^{2+} after target strand cleavage (**c**) and the model of the catalytic site before DNA cleavage (**d**). In this configuration (referred to as active) N863 coordinates Mg^{2+} in place of D861. The atomic coordinates of HNH are shown as cartoon (green), while the electronic density is shown as wireframes (grey). The EMD-0584 map displays a visible density in the position of Mg^{2+} (**c**, indicated using an arrow), in agreement with the EMD-24838 map[15] (Supplementary Fig. 2), enabling location of the Mg^{2+} ion in the catalytic state (**d**).

was not determined experimentally. The HNH site was reconstructed on the basis of X-ray structures obtained in the absence of Mg^{2+} (refs. [4,11]), and on the structure of endo VII (ref. [9]), ultimately suggesting that D839 and D861 coordinate Mg^{2+}.

Recently, the structural determination of the HNH catalytic core captured right after target strand cleavage and in the presence of Mg^{2+}, displayed a different configuration of the catalytic site (Fig. 1c)[14,15]. A first cryo-EM structure (PDB 6O0Y) displayed a visible density in the position of Mg^{2+} (Fig. 1c), enabling the catalytic ion to be located and the catalytic state to be reconstructed. Here, N863 (rather than D861) coordinates Mg^{2+}, forming a catalytic triad with D839 and H840, while D861 points outwards. This configuration was confirmed by other recent cryo-EM structures (for example, PDB 7S4X), revealing the atomic positions of the catalytic site, including the Mg^{2+} ion (Supplementary Fig. 2)[15,16]. Biochemical experiments also showed that the D861A substitution retains DNA cleavage activity, while N863A loses its gene-editing capability[4,17]. These findings suggest that the previously reported state could be an alternative pseudo-active state, advocating also for a possible conformational equilibrium between the two states[17]. In this scenario, studies of the catalytic mechanism inferred information from endo VII (ref. [18]), and were based on a cryo-EM structure of the pseudo-active state[19], sustaining the coordination of Mg^{2+} by both D861 and D839.

Considering this knowledge, fundamental questions remain unanswered. First, the biological relevance of the pseudo-active state in the presence of the Mg^{2+} ions is ambiguous. Indeed, structural evidence supporting the pseudo-active site captured HNH far from the cleavage site, and/or without Mg^{2+} (refs. [3,4,11–13]). Moreover, the possible equilibrium between the two states in the presence of catalytic Mg^{2+} has not been investigated. The molecular details of this equilibrium are central to a complete understanding of the complex conformational landscape of HNH, and to inform the conformational requirements underlying DNA cleavage. Finally, and perhaps more intriguingly, considering the newly reported structural data[14–16], the catalytic mechanism of target strand cleavage, and how it proceeds through the aid of a single Mg^{2+} ion, has not been addressed. This knowledge is important to improve the efficiency of the Cas9 enzyme and can help in overcoming unselective DNA cleavage.

Here, multimicrosecond molecular dynamics (MD) simulations and free energy methods have been combined with extensive ab initio MD and quantum mechanics/molecular mechanics (QM/MM) approaches to establish the molecular details of the equilibrium

https://doi.org/10.1038/s41929-022-00848-6

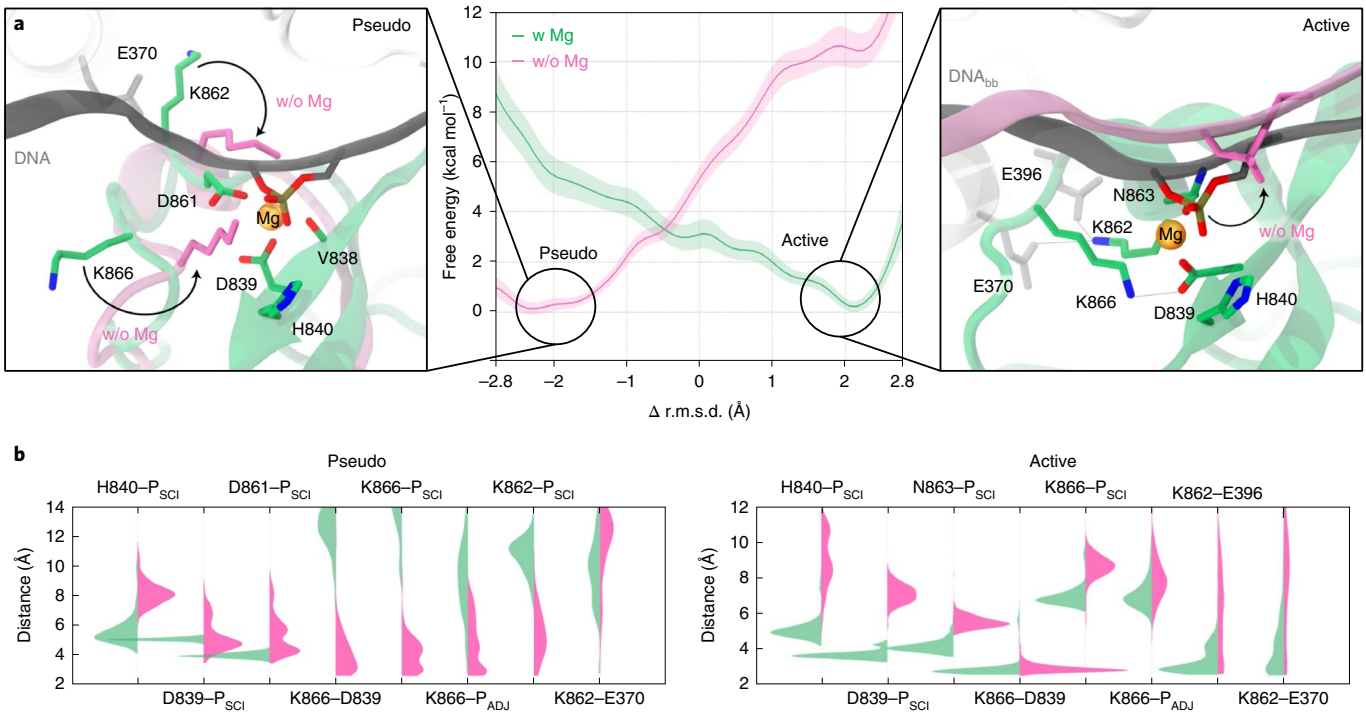

Fig. 2 | Transition of the HNH domain from pseudo-active to active states. **a**, Free energy profiles for the HNH conformational transition in the presence of Mg^{2+} (w Mg, green) and without Mg^{2+} ions (w/o Mg, magenta). Two close-up views show the conformation of the pseudo-active (that is, pseudo) and active states at their energetic minima. The HNH catalytic site bound to Mg^{2+} is shown as in Fig. 1c. Residues in green refer to HNH bound to Mg^{2+} whereas residues in magenta show the conformational change occurring in the absence of Mg^{2+} (also indicated using an arrow). **b**, Probability distributions of critical interaction distances in the presence of Mg^{2+} (green) and absence of Mg^{2+} (magenta) for the pseudo-active and active states at their energetic minima (that is, at −2.5 Å ≤ RC ≤ −1.5 Å and 1.5 Å ≤ RC ≤ 2.5 Å, respectively). The Welch's *t*-test was used to assess the statistical significance of the differences in the distributions of the interaction distances with and without Mg^{2+}. At the confidence level of 95%, we rejected the null hypothesis in favour of the alternative with a *P* value of <0.0001. This observation was true for all cases except for the K866–D839 distance in the active system, where *P* = 0.437, which arises from overlapping distributions.

between active and pseudo-active states, to resolve the mechanism of DNA cleavage. Supported by solution nuclear magnetic resonance (NMR) and biochemical assays, our findings reveal a critical structural role for Mg^{2+}, determining the chemical mechanism and which of the known conformations of HNH is responsible for target DNA cleavage in Cas9.

Results

Molecular preference of the competent HNH

To determine the conformation of the catalytic HNH domain in the presence of Mg^{2+}, and to characterize the relevance of the pseudo-active state, we performed free energy simulations. We investigated the pseudo-active to active transition (and vice versa) using umbrella sampling simulations[20], rigorously sampling the populations of the two states and of their intermediates through ~28 μs of MD runs. The transition was sampled along the difference in root mean square deviation (r.m.s.d.) of the position of the heavy atoms of HNH with respect to both states (used as the reaction coordinate, RC), in the presence and absence of Mg^{2+}. Our classical model describes the metal sites in line with QM/MM simulations, providing the basis for classical and free energy simulations (Supplementary Figs. 3–6).

The simulations show that without Mg^{2+} (w/o Mg^{2+}), the free energy reaches a minimum at an RC of approximately −2.3 Å (Fig. 2a), corresponding to the pseudo-active conformation. By contrast, with Mg^{2+} (w Mg^{2+}) we detected an opposite trend in the free energy profile, which displayed a well-defined minimum at an RC of ~2.2 Å, which is consistent with the active state. Two close-up views show the conformation of the pseudo-active and active states at their energetic minima (Fig. 2a). Hence, the conformational state of HNH critically depends on the presence of Mg^{2+}, which favours the formation of the active conformation, while in the absence of Mg^{2+}, HNH mainly populates the pseudo-active state.

To understand this observation, we analysed the conformational ensembles at the free energy minima. The pseudo-active state at its energetic minimum (at −2.5 Å ≤ RC ≤ −1.5 Å) revealed that, in the presence of Mg^{2+}, the K862 and K866 side chains orient away from the DNA (Fig. 2a, left). Instead, w/o Mg^{2+}, K866 binds the scissile phosphate (P_{SCI}) and its adjacent phosphate (P_{ADJ}), and is stabilized also by the interaction with D839. This is evidenced by the probability distributions of the distances involving K866, K862 and the DNA phosphate groups assuming values of <6 Å (Fig. 2b). This is consistent with the cryo-EM structure of the pseudo-active state, where the K866 side chain locates in proximity to the DNA[13]. Hence, K866 and K862 stabilize the pseudo-active state in the absence of Mg^{2+}.

The active state at its energetic minimum (at 1.5 Å ≤ RC ≤ 2.5 Å) showed that, in the presence of Mg^{2+}, H840, D839 and N863 stably maintain their positions with respect to P_{SCI}, while D839 binds K866 (Fig. 2). In the absence of Mg^{2+}, the catalytic residues move away from the DNA, destabilizing the latter with respect to the catalytic core. With Mg^{2+}, K862 interacts with E396 and E370, while these interactions are lost w/o Mg^{2+}. The active state is thereby stabilized by Mg^{2+}, loss of which leads to unproductive conformations for catalysis.

These findings clarify why structural studies that did not include Mg^{2+} supported a pseudo-active configuration[3,4,11–13]. Indeed, the absence of Mg^{2+} leads to interactions that anchor HNH at the DNA, stabilizing the pseudo-active conformation. Conversely, with Mg^{2+}, HNH shifts its preference towards an active state, which is consistent with cryo-EM findings obtained in the presence of Mg^{2+} (refs. [14–16]).

https://doi.org/10.1038/s41929-022-00848-6

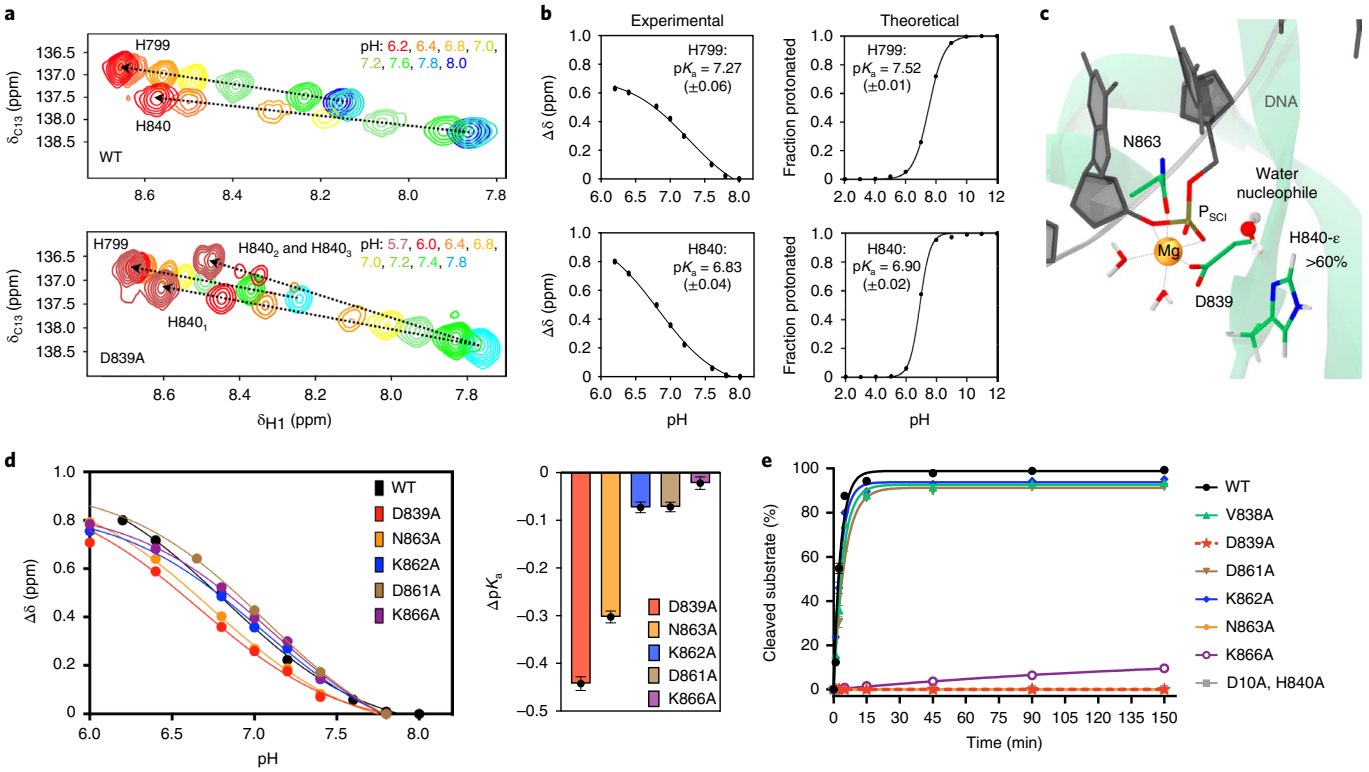

Fig. 3 | Chemical environment enabling the catalysis. a, Titration of the H799 and H840 side chains in the wild-type HNH (WT, top) and in the D893A mutant (bottom), reporting changes in the ^{1}H-$^{13}C_{\varepsilon 1}$ correlation as the histidine nitrogen atoms change protonation state, in the presence of Mg^{2+} and DNA. Resonances are coloured according to the pH values in the key. Chemical shifts of ~8.5 ppm correspond to fully protonated histidine, while those of ~8.2–7.9 ppm refer to partially protonated or deprotonated residues. The trajectories of the chemical shifts are indicated using dashed arrows. In the D893A mutant, H840 shows evidence of two to three conformational states (upper arrow, H840$_{1-3}$). **b**, Fitted titration curves for pK_a determination of H799 and H840 in the WT HNH through solution NMR (experimental) and CpH molecular dynamics (theoretical). A modified version of the Henderson–Hasselbach equation was used for fitting the experimental data. Computational data were obtained by fitting the deprotonated fraction to equation (3). Each data point represents the ensemble population of the protonated/deprotonated states from CpH MD simulation (last ~32 ns) for each pH value. pK_a values are reported, alongside the error from the fit (Supplementary Methods). **c**, Active state of HNH displaying the tautomeric form of H840 protonated on ε (H840-ε), which occurs for >60% of CpH MD at pH 7.4 (Supplementary Fig. 11). **d**, Titration curves for the catalytic H840 in the WT HNH and its mutants, and change in pK_a upon mutation (ΔpK_a). In plots of the former, error associated with variance in NMR chemical shifts is smaller than the data points depicting Δδ. Values of ΔpK_a are depicted as dots with error bars propagated as $E_{(\Delta pKa)} = E_{(mutant)} - E_{(WT)}$, where E is the error determined from fits of the individual titration curves in Prism v.9.0, centred at the calculated ΔpK_a value. Coloured bars are shown as a visual aide to assess the magnitude of ΔpK_a. **e**, In vitro cleavage kinetics of Cas9 HNH mutants on a double-stranded DNA on-target substrate. Lines represent a single exponential fit of each individual time course experiment; each data point represents an average of four independent experiments (n = 4) with standard deviation plotted for each data point.

Chemical environment enabling catalysis

To examine the chemical environment enabling catalysis and to determine the protonation state of the catalytic H840, we performed solution NMR, DNA cleavage experiments and extensive MD simulations (reaching ~54 μs of collective sampling).

Changes in the protonation state of the catalytic histidine side chain were monitored through solution NMR using a construct of the HNH nuclease that was shown to properly represent this domain in full-length CRISPR–Cas9, in the presence of Mg^{2+} and DNA (Methods and Supplementary Fig. 7)[21,22]. Two-dimensional ^{1}H-^{13}C correlation spectra depict strong chemical shift perturbations in the catalytic H840 as the pH of the sample is modulated (Fig. 3a and Supplementary Figs. 8 and 9). We also titrated H799, which does not change on alanine mutation, and is an internal standard. The chemical shift (δ_{H1}) range indicates that H799 and H840 are both partially protonated at pH 7.4. The measured pK_a was ~7.27 and ~6.83 for H799 and H840, respectively, indicating that, at pH 7.4, H799 is 79% protonated and H840 is 57% protonated (Fig. 3b). To further assess the pK_a of H799 and H840 in the full-length CRISPR–Cas9, we performed constant pH (CpH) MD simulations in explicit solvent, in conjunction with replica exchange[23]. The computed pK_a for H799 and H840 in the active state of HNH resulted in values of ~7.52 and ~6.90, respectively (Fig. 3b), which is in very good agreement with the NMR data. Although the determination of the tautomeric populations of H840 by NMR were complicated by fast relaxation of the histidine side chain and possible proton exchange with solvent[24,25], in-depth analysis of the simulations revealed that, at pH 7.4, H840 is likely to assume the neutral tautomeric form protonated on the ε position (H840-ε > 60% of the simulation; Supplementary Fig. 10), resulting in the δ nitrogen in proximity to the nucleophile for activation and at ~5 Å from P$_{SCI}$ (Fig. 3c). CpH MD simulations of the pseudo-active state revealed a shift of the pK_a towards higher values, reducing the fraction of H840-ε to ~50% at pH 7.4. Classical MD of the two tautomeric forms of H840 also showed that H840-ε leads to a stable catalytic site (Supplementary Fig. 11). Conversely, the tautomeric form of H840 protonated on δ results in detachment of the activating nitrogen from P$_{SCI}$ at >6 Å, unlikely for nucleophile activation. Hence, NMR and MD suggest that H840-ε is prone to catalysis, which is critical information for study of the catalytic mechanism (see below).

NMR experiments were also performed introducing alanine mutations of D839, D861, N863, K862 and K866. Each of these HNH mutants showed a shift to a lower pK_a value (Fig. 3d and Supplementary Fig. 9). D839A dramatically reduced the pK_a of the catalytic histidine, suggesting that in this mutant H840 is a weaker base. This mutant hampers cleavage (Fig. 3e and Supplementary Fig. 12)[17] and exhibits substantial

https://doi.org/10.1038/s41929-022-00848-6

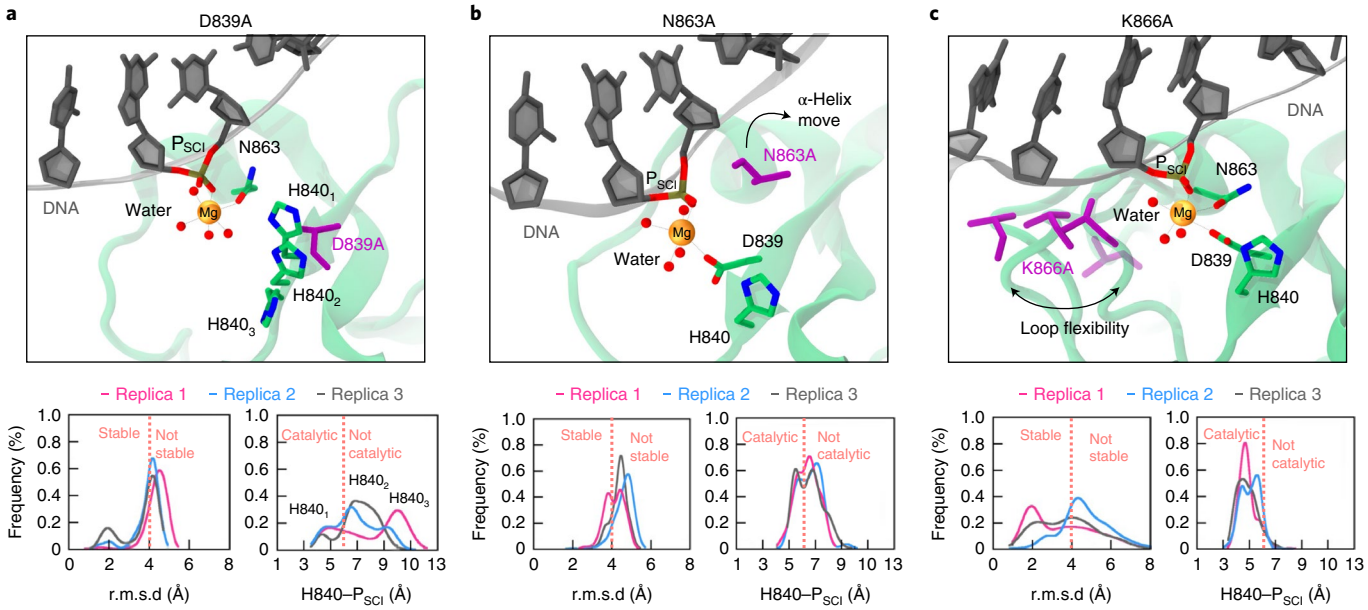

Fig. 4 | Effect of alanine mutations on the catalytic site. Data are shown for the D839A, N863A and K866A mutants in the active state of HNH. Alanine mutations are shown in violet. **a**, D839A affects the conformation of H840, resulting in three main conformations ($H840_{1-3}$). **b**, N863A results in detachment of the S860–K866 α-helix from the catalytic Mg^{2+}, destabilizing the catalytic core. **c**, K866A destabilizes the catalytic core, with flexibility of the S860–D868 loop (four configurations are shown). Bottom graphs: stability of the catalytic site, computed as probability distribution of the root mean square deviation (r.m.s.d.) of the heavy atoms within 8 Å of the catalytic Mg^{2+}; and location of the catalytic H840 with respect to the scissile phosphate (P_{SCI}), computed considering the interatomic distance between H840 (N_δ) and P_{SCI}. Data are reported for three simulation replicas of ~1 μs each. Vertical dashed lines (orange) indicate the cutoffs for the stability of the catalytic site (that is, r.m.s.d. < 4 Å) and for the catalytic function of H840 (that is, H840–P_{SCI} < 6 Å allows the water nucleophile to position between H840 and P_{SCI}, Fig. 3c, while H840–P_{SCI} > 6 Å results in detachment of H840 from the catalytic centre).

structural perturbations, with altered flexibility throughout the protein and near the catalytic site (Supplementary Fig. 8).

In this mutant, H840 populates at least two (or three) states during pH titrations, which manifest as a series of bifurcated NMR resonances in Fig. 3a. This is consistent with classical MD of the D839A mutant in the active state, where H840 adopts three different conformations and detaches from P_{SCI} (Fig. 4a). Hence, the loss of activity in the D839A Cas9 may be due to the ability of this residue to strongly influence the conformation of the adjacent H840. Titration of the N863A mutant also revealed a reduction in the pK_a of H840 (Fig. 3d). This is in line with classical MD of the active state, where N863A results in detachment of the S860–D868 helix from Mg^{2+} (Fig. 4b), inducing instability of the site and affecting the location of H840 with respect to P_{SCI}. This is consistent with NMR data showing structural and dynamic changes at the active site and proximal to the S860–D868 helix (Supplementary Figs. 8, 13 and 14). D861A resulted in a reduced effect on the change in pK_a of H840 (Fig. 3d), supporting that D861 resides outside the catalytic site in the active state, at odds with the pseudo-active conformation where it coordinates Mg^{2+} (Fig. 1c). Finally, K862 and K866 do not affect the pK_a of H840, probably due to their distal location with respect to the catalytic residue. K866A, however, reduces the enzymatic activity (Fig. 3e). In this respect, classical MD of K866A in the active HNH reveals that this mutation induces instability of the S860–D868 helix (Fig. 4c) and increases its flexibility with respect to the wild type (Supplementary Fig. 15), resulting in impeding of the K866–D839 interaction that is crucial for stabilization of the active state (Fig. 2). Remarkably, mutations that retain catalysis (V838A, D861A and K862A) preserve the stability of the catalytic site in the active HNH, with H840 remaining in the vicinity of P_{SCI} (Supplementary Fig. 16).

Overall, the dynamics of the active state agree with NMR and DNA cleavage experiments. Molecular simulations of the pseudo-active state do not provide a rationale for the experimental evidence. Indeed, only V838A preserves the HNH site, maintaining its backbone interactions, while all other mutants result in the detachment of H840 from P_{SCI} (Supplementary Figs. 17 and 18).

Catalytic mechanism of target strand cleavage

To further resolve which of the known HNH conformations is catalytic, and to characterize the cleavage mechanism, we employed ab initio MD and free energy methods, reaching ~700 ps of collective runs[26]. We performed thermodynamic integration[27] with QM/MM simulations[28] of HNH in both the active and pseudo-active states at the density functional theory (DFT)-BLYP level of theory[29,30]. Phosphodiester bond cleavage was studied along the difference in the distance between the breaking and forming P–O bonds (used as RC; Fig. 5). This approach was used on several nucleases[31–34], including the Cas9 RuvC domain[5]. The reactant states were carefully selected from classical MD (Supplementary Fig. 19), and considered the tautomeric form of H840 protonated on ε, as in the NMR and computational studies (see above).

For each system, we collected >200 ps of ab initio MD, obtaining the free energy profiles for the active and pseudo-active states of HNH (Fig. 5a). The chemical step proceeds from the reactants (R) to the products (P), and is separated by a transition state ($TS^{\ddagger}$) maximum. The reaction is activated by H840, acting as a general base, and proceeds through an S_N2-like mechanism. The activation free energy for the chemical step in the active HNH was 17.06 ± 1.22 kcal mol^{-1} (details, cross-validation and error analysis in Supplementary Figs. 20 and 21). This is consistent with the catalytic rate of 4.3 s^{-1} (corresponding to $\Delta G^{\ddagger}$ of ~16/17 kcal mol^{-1}) measured for the HNH catalysis, distinct from the RuvC catalysis and from nucleic acid binding[35].

The free energy barrier for the pseudo-active state was ~22.02 ± 1.26 kcal mol^{-1}, in line with studies of the catalytic mechanism based on the pseudo-active configuration[18,19]. This activation barrier is considerably higher than that for the active state (~5 kcal mol^{-1} with no

https://doi.org/10.1038/s41929-022-00848-6

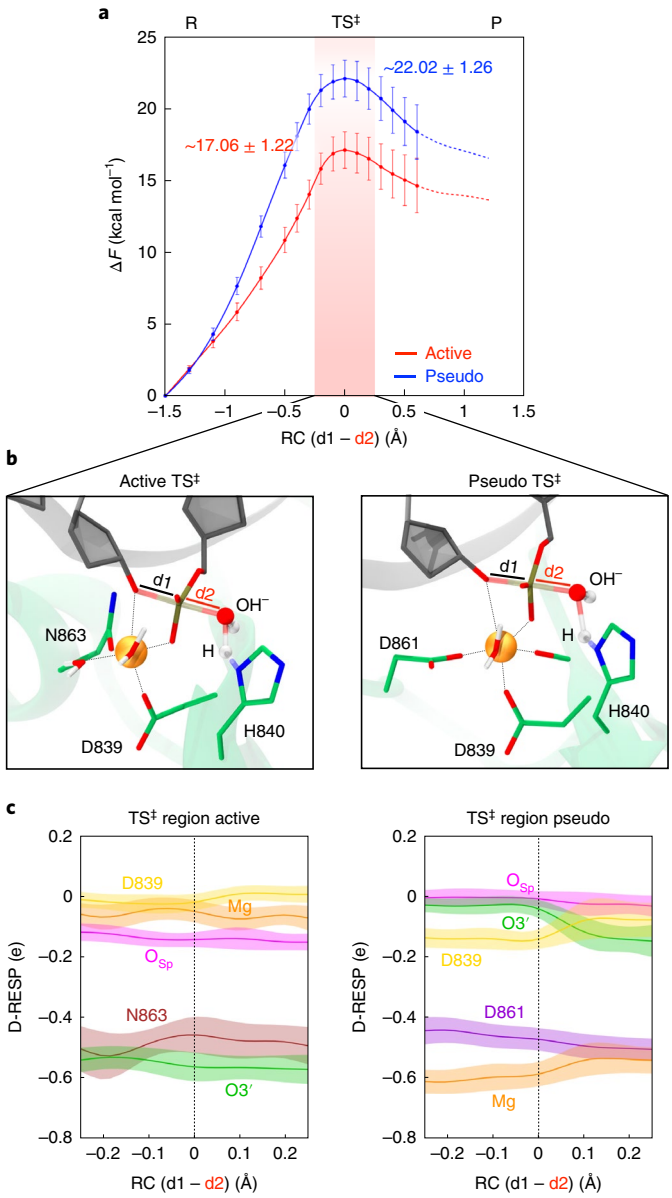

Fig. 5 | Free energy profiles for phosphodiester bond cleavage. a, Free energy profiles (ΔF, in kcal mol^{-1}) for the active (red) and pseudo-active (blue) states of HNH, obtained through QM(BLYP)/MM MD and thermodynamic integration. The difference in distance between the breaking and forming P–O bonds is the reaction coordinate (RC = d1 – d2, shown in **b**). The chemical step evolves from R to P passing through TS‡ (region indicated using a red vertical bar). Note: error bars show standard deviations obtained from error propagation analysis of the primary dataset in which each data point represents the mean from the last ~5 ps of converged ab initio MD. **b**, Close-up view of the TS‡ structure in the active (left) and pseudo-active (right) states of HNH. **c**, D-RESP[28] charges within the TS‡ region (that is, −0.2 Å < RC < 0.2 Å) for the active and pseudo-active states. At each value of the RC, data are presented as mean (solid line) and standard deviation (shaded bands), computed over the last ~5 ps of converged ab initio MD.

overlapping error bars between the energy peaks). This indicates the catalytic preference of the active conformation, and that the catalysis is unlikely to proceed through the pseudo-active state. To understand the origin of this difference in the free energy barrier, we examined the polarization effects in the catalytic centres. We analysed how the dynamical electrostatic potential-derived charges (D-RESP)[28] change at the TS‡ (Fig. 5b,c) and along the chemical steps (Supplementary Fig. 22). At the TS‡, the charge of Mg^{2+} and of the O3′ leaving group display opposite values in the active and pseudo-active conformations of HNH. In the pseudo-active HNH, the charge of Mg^{2+} is lower than that in the active state, mainly due to the coordination of both D839 and D861 carboxylates. D839 also reduces its charge with respect to the active HNH, with increased polarization on Mg^{2+}. The latter loosely binds O3′ in the pseudo-active HNH (Supplementary Fig. 22), with an increase of the O3′ charge compared to the active state. This charge increase reduces the ability of O3′ to be a good leaving group, resulting in a larger barrier at the TS‡.

To fully characterize the catalytic mechanism in the active HNH, we performed QM/MM metadynamics simulations[36]. This method enabled explicit description of the phosphodiester bond cleavage in one dimension (the first collective variable, CV_1) and deprotonation of the water nucleophile in the other dimension (CV_2). With ~120 ps of sampling, the free energy surface confirmed an S_N2-like mechanism (Fig. 6a and Supplementary Figs. 23 and 24). H840 extracts the proton from water right before the free energy peak (the TS‡), which is consistent with the in-depth analysis of thermodynamic integration (Supplementary Fig. 22), and similar to other His-activated nucleases[31,37], including the RuvC domain of Cas9 (ref. [5]). An overall free energy barrier of ~17.38 ± 0.84 kcal mol^{-1} is consistent with our monodimensional profile (Fig. 5) and with kinetic studies[35]. Interestingly, during unbiased QM/MM simulations of the P state, O3′ is protonated by one water molecule coordinating Mg^{2+}. By including the neighbouring K866 in the QM part, K866 releases a proton to the water molecule, which simultaneously protonates the O3′ (Fig. 6 and Supplementary Fig. 25), reaching a final product that is consistent with cryo-EM structures of the postcleavage state (Supplementary Fig 26). Indeed, as shown by CpH MD simulations, the pK_a of K866 is 10.9 (Supplementary Fig 27), making it a suitable proton donor for the basic alkoxide O3′ (pK_a ~16). To further understand the role of K866, we computed the D-RESP charges from ab initio QM/MM simulations in the presence of K866 and its alanine mutant (both included in the QM part). We found that the K866A substitution dramatically increases the negative charge on P_{SCI}, making it a worse electrophile (Supplementary Fig. 28). This analysis also reveals that all three K866 hydrogen atoms are positive in the P state and thereby prone to shuttle towards the negatively charged oxygen atoms of water and the O3′. Hence, K866 creates a favourable electronic environment and is critical for leading the catalysis to completion. This clarifies DNA cleavage experiments, showing that K866A remarkably reduces the enzymatic activity (Fig. 3e).

Discussion

We combined extensive molecular simulations with solution NMR and biochemical experiments to decipher the catalysis of target DNA cleavage in CRISPR–Cas9. Up until now, the structure of the catalytic state has been ambiguous, with two possible conformations (referred to as active and pseudo-active) for which a conformational equilibrium was suggested[17]. Free energy simulations were used to investigate this equilibrium, revealing that Mg^{2+} favours formation of the active state, while its absence leads to the pseudo-active conformation (Fig. 2). Here, D861 and D839 point towards the catalytic core, while N863 orients in the opposite direction, which is consistent with structures of Cas9 obtained without Mg^{2+} ions[3,4,11–13]. This pseudo-active state is stabilized by two second-shell lysines (K866, K862), which anchor the DNA backbone in place of the ion. By contrast, when Mg^{2+} is bound to the catalytic site, N863 reorients to coordinate the ion and engage in catalysis. This is in line with multiple sequence alignments of Cas9 orthologues that show that N863 is conserved, whereas D861 is highly variable[4]. It also agrees with recent cryo-EM structures of the postcatalytic state, obtained in the presence of Mg^{2+} (refs. [14–16]). Hence, Mg^{2+} is important not only for the catalysis, but also for its critical structural role. Accordingly, single-molecule experiments have shown that HNH adopts different states in the presence/absence of Mg^{2+} (ref. [38]). Another single-molecule study showed that HNH is unlikely to transition to the

https://doi.org/10.1038/s41929-022-00848-6

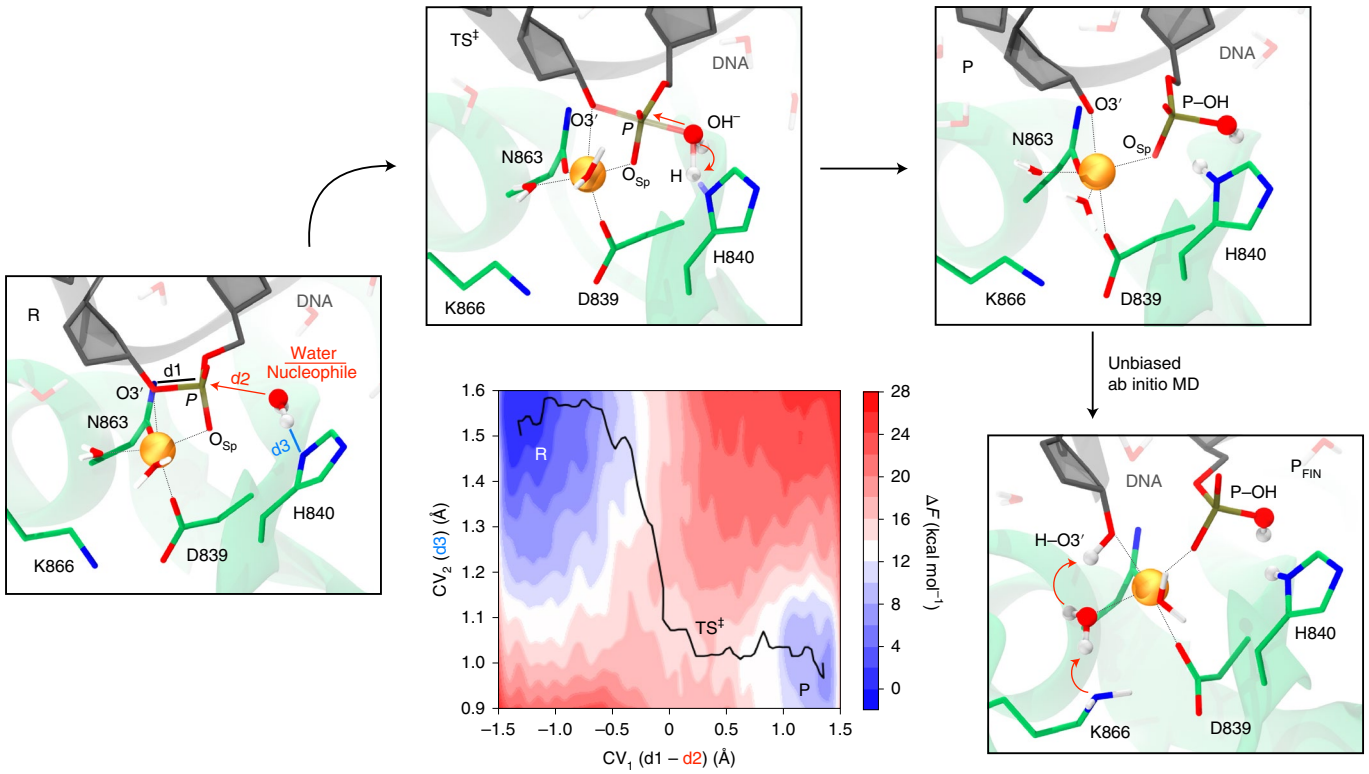

Fig. 6 | Catalytic mechanism of DNA cleavage in the HNH domain of CRISPR–Cas9. Two-dimensional free energy surface for phosphodiester bond cleavage reporting the progress of the chemical step from the R, TS‡ and P along two CVs (shown on the three-dimensional structure of R). CV_1 denotes the nucleophilic attack on scissile phosphate, while CV_2 accounts for the proton transfer from the water nucleophile to H840. The free energy surface was obtained through QM/MM metadynamics. Unbiased QM/MM simulations of the P state reveal that the K866 side chain releases a proton to the water molecule coordinating Mg^{2+}, which protonates the DNA O3′, leading to the final product (P_{FIN}) of DNA cleavage. This clarifies DNA cleavage experiments (Fig. 3e), showing that the K866A substitution remarkably reduces the enzymatic activity.

active state for DNA cleavage without divalent cations[7]. Our findings therefore support a model in which Mg^{2+} shifts the conformational equilibrium of HNH towards the catalytic state.

Solution NMR revealed that at pH 7.4, the pK_a for the catalytic H840 is ~6.83. Molecular simulations using a CpH MD method[23] reported excellent agreement for the pK_a of H840 in the active state (~6.90), and indicated that H840 probably assumes the neutral tautomeric form protonated on ε (Fig. 3c), providing the basis for QM/MM studies of the catalysis. Solution NMR and MD also consistently reported the structure and dynamics of the active HNH in the presence of alanine mutations. Indeed, mutations that disrupt the catalytic activity (D839A, N863A, K866A) more sensibly reduce the pK_a of H840, and alter the location of H840 with respect to P_{SCI} (Figs. 3 and 4), while mutants that preserve the catalysis (D861A, K862A) result in minor conformational effects. The consistency of the molecular simulations with the experiments also suggests harnessing of computations for residual mutations in the active HNH. For instance, R844 (in place of Q844) stably binds P_{SCI}, mimicking the stabilizing role of R976 in the RuvC site (Supplementary Fig. 29)[15,39], which could be exploited to ameliorate HNH function.

Ab initio QM/MM simulations were used to resolve the catalytic mechanism, revealing the catalytic preference of the active conformation (Fig. 5), and that catalysis is unlikely in the pseudo-active state. In the active HNH, an activation barrier for phosphodiester bond cleavage of ~17 kcal mol^{-1} was in line with the experimental catalytic rate[35]. The catalysis proceeds through an S_N2 mechanism, activated by H840 and critically aided by K866. The latter fosters an optimal electronic environment and can intervene in the protonation of the O3′, clarifying why the K866A substitution reduces the enzymatic activity (Fig. 3e). Notably, restriction/homing endonuclease enzymes using a single catalytic metal commonly display a lysine/arginine residue in the position of K866, to complement the divalent metal[40]. In light of this observation, our findings also offer a mechanistic rationale for the role of the additional positively charged residue in other one-metal-dependent enzymes.

Overall, the extensive multiscale approach implemented here resolves the catalytic mechanism, and which of the known conformations of HNH is responsible for target DNA cleavage in CRISPR–Cas9. These findings provide a foundation to ameliorate the function and specificity of CRISPR–Cas9, helping in the development of improved genome-editing tools.

Methods

MD simulations

Molecular simulations were based on the cryo-EM structures EMD-0584 (PDB 6O0Y, at 3.37 Å resolution)[14] and EMD-23838 (PDB 7S4X, at 2.76 Å resolution)[15], which captured the active state of the HNH domain, and on the X-ray structure PDB 5F9R (at 3.40 Å resolution)[12], which was used to model the pseudo-active configuration of HNH[6]. Each system was embedded in explicit water, leading to periodic simulation cells of ~180 × 120 × 140 $Å^3$ and ~340,000 total atoms. The Amber ff12SB force field was employed, including the ff99bsc0 corrections for DNA[41] and the ff99bsc0+χOL3 corrections for RNA[42,43]. The Li and Merz model was used for Mg^{2+} (ref. [44]), describing the metal sites in agreement with QM/MM simulations (Supplementary Fig. 3). An integration time step of 2 fs was used. Temperature control (300 K) was performed via Langevin dynamics, while pressure control was accomplished by coupling the system to a Berendsen barostat[45]. The simulation protocol is described in the Supplementary Methods. Production runs were carried out in

the NVT ensemble on the active and pseudo-active states of the WT Cas9 and of six mutants (that is, V838A, D839A, D861A, K862A, N863A, K866A). The WT Cas9 was also simulated considering two tautomeric forms of H840 (protonated on δ and ε). The active HNH was also simulated in the presence of the Q844R mutation. For each system, three MD replicas of ~1 μs each were performed, for a total of ~54 μs of MD runs. The GPU-empowered version of AMBER 20 (ref. [46]) was used as the MD engine.

Umbrella sampling simulations

The umbrella sampling (US) method was used to compute the free energy profiles associated with the conformational change of the HNH domain from the pseudo-active to active states (and vice versa)[20]. In this method, a number of simulations (US windows) are run in parallel with additional harmonic bias potential applied to selected RCs:

$$V(RC) = \frac{k}{2}\left(RC(t) - RC^*\right)^2 \quad (1)$$

where V(RC) is the value of the bias potential, k is a bias force constant, RC(t) is the value of RC at a given time t and RC* is the reference value of RC. By using different RC* values in each US window, one can sample the biased probability distribution p_b(RC) along the whole RC range of interest. The difference in r.m.s.d. of the positions of the HNH heavy atoms with respect to pseudo-active and active conformations was used as an RC (details in the Supplementary Methods). Two independent sets of US simulations were performed: (1) in the presence of Mg^{2+} ions; and (2) in the absence of Mg^{2+} ions. In both cases, the system was simulated in 14 overlapping windows from RC = −3.5 Å to RC = 3.5 Å using a harmonic restraint with a spring constant of 30 kJ mol^{-1} Å^2. The centre of the harmonic bias potential was distributed along the RC in 14 windows separated by 1.0 Å or 0.5 Å, to allow proper overlapping of the probability distributions. Approximately 1-μs-long trajectories were obtained for each US window, reaching ~14 μs of collective sampling per system (a total of ~28 μs). The free energy profiles were computed using the weighted histogram analysis (WHAM) method[20]. Analysis of the conformational ensembles was performed on the reweighted trajectories. Details in the Supplementary Methods.

CpH molecular dynamics

CpH MD simulations were performed in explicit solvent, in conjunction with a replica exchange method to enhance the sampling of the protonation states[23]. pH values were sampled from 1 to 14, with exchanges between adjacent replicas every 200 fs, reaching ~40 ns in each replica. The pK_a was computed from the distribution of the protonation states using the Hill equation:

$$pK_a(i) = pH - n\log\frac{x_i}{1 - x_i} \quad (2)$$

in which x_i is the deprotonated fraction of residue i and n is the Hill coefficient. The titration curves were derived by fitting the deprotonated fraction x_i to equation (3), using the Levenberg–Marquardt non-linear optimization method.

$$f_{x_i} = \frac{1}{10^{n(pK_a - pH)} + 1} \quad (3)$$

Good titration curves display small deviations of each point from the fitted titration curve, and Hill coefficients between 0.5 and 1.5, indicating that the protonation states are properly sampled at the simulated pH values[23]. By using this method, the pK_a of H840 and H799 in the active and pseudo-active states of HNH was computed. The pK_a of H840 in the active HNH was also computed in the presence of the Q844R mutant. Furthermore, CpH MD simulations were carried out to compute the pK_a of K866 in the active state. For each system, CpH MD built on ~560 ns of MD sampling (that is, ~40 ns for 14 replicas), resulting in a total of ~3.4 μs of CpH MD simulations. Details in the Supplementary Methods.

QM/MM simulations

QM/MM simulations were performed on the active and pseudo-active states of HNH. In the active state, the QM part included the Mg^{2+} ion and its coordinating residues (D839, N863), the catalytic H840, the DNA bases G3, T4 and C5, and nine water molecules. In the pseudo-active state, the QM part consisted of the Mg^{2+} ion and its coordinating V838, D839 and D861, H840, the DNA bases G3, T4 and C5, and eight water molecules. For both systems, capping hydrogens were used to saturate the valence of the terminal QM part, resulting in 123 and 111 QM atoms for the active and pseudo-active systems, respectively. The catalytic H840 was simulated in the neutral tautomeric form protonated at the ε position, as indicated by classical and CpH MD (Supplementary Figs. 10 and 11) and NMR data (Fig. 3). The QM part was described at the QM DFT/BLYP[29,30] level, and the MM part was treated using the force field above. QM/MM simulations were performed using the CPMD code[47]. The wavefunctions were expanded in a plane-wave basis set up to a cutoff of 75 Ry in a QM cell of dimensions ~26 × 28 × 26 Å^3. A rigorous Hamiltonian treatment of the electrostatic interaction between the QM and MM regions was used[28]. The temperatures of the QM and MM subsystems were kept constant at 300 K using a Nosé–Hoover thermostat[48,49]. Car–Parrinello QM/MM simulations[50] were performed with a time step of 5 a.u. (~0.12 fs) and a fictitious electron mass of 600 a.u. (that is, ~0.33 AMU). On ~40 ps of unconstrained ab initio MD, the Mg^{2+}-bound configurations were used to start for free energy simulations (see below).

Thermodynamic integration

Thermodynamic integration and the 'blue moon ensemble' method[27] were used to compute the free energy profiles for phosphodiester bond cleavage in the active and pseudo-active states of HNH. In this approach, the average converged constraint forces are computed and integrated along a given RC, deriving the associated free energy profile. The difference in distance between the breaking and forming P–O bonds was used as the RC, as in previous studies of DNA/RNA processing enzymes[5,31–34]. Starting from an RC = −1.5 Å, we sampled in 18 sequential windows along the RC, with a resolution of 0.2 Å (0.1 Å in the region in the vicinity of the TS‡). Each window was simulated for ≥8 ps, reaching convergence of the constraint force, and collecting ~144 ps of ab initio MD for the active and pseudo-active states. To estimate the error associated with hysteresis, we also computed the backward free energy pathways, resulting in eight additional sampling windows and ~64 ps of ab initio MD for each system. The statistical error at each point of the free energy profiles (both forwards and backwards) was computed by error propagation analysis. The overall error on the free energy barrier was estimated as the sum of the statistical error and the error due to hysteresis between the forward and backward pathways. Full details are given in the Supplementary Methods. Altogether, considering the forward and backward pathways, and unconstrained runs, we collected ~218 ps of ab initio QM/MM MD for each system.

Metadynamics

The catalytic mechanism of DNA cleavage in the active CRISPR–Cas9 was characterized through QM/MM simulations and metadynamics[36]. This method applies an external history-dependent bias potential to the Hamiltonian of the system as a function of a set of predefined degrees of freedom (CVs). We used two CVs: (1) the difference in distance between the breaking and forming P–O bonds; and (2) the distance between the H840 δ nitrogen and the hydrogen of the water nucleophile. This enabled mapping the free energy for phosphodiester bond cleavage on one dimension and characterizing

nucleophile activation on a second dimension. An extended Lagrangian version of the method was used for proper coupling with the QM/MM simulations. For both CVs, the mass value of the fictitious particles was 16 AMU (a.u./Bohr)^2, while the value of the force constant was 0.24 Hartree/Bohr^{-2}. The height of the Gaussian terms was 0.5 kcal mol^{-1}, which ensures sufficient accuracy for reconstructing the free energy surface. The width of the Gaussian terms was 0.05 Å, in line with the oscillations of the CVs in a free Car–Parrinello QM/MM run. The Gaussian function deposition rate was set to 24 fs. A total of ~5,000 Gaussians were deposited, collecting ~120 ps. During this time, the chemical step crossed from the reactants to products multiple times (Supplementary Fig. 23), properly sampling the conformational space. The statistical error associated with the free energy for phosphodiester bond cleavage was estimated by average blocking, that is, by computing and averaging the standard deviation of the mean free energy in blocks of ~10 ps of the converged metadynamics simulation. Full details are given in the Supplementary Methods.

NMR spectroscopy

The WT HNH domain and the active site alanine mutations (D839A, H840A, D861A, K862A, N863A, K866A) of Cas9 were expressed and purified as previously described[21]. Samples of ^{13}C- and ^{15}N-labelled HNH were expressed in Rosetta(DE3) cells in M9 minimal media containing MEM vitamins, $MgSO_4$ and $CaCl_2$ with $^{15}NH_4Cl$ and U-^{13}C-dextrose (Cambridge Isotopes) as the sole nitrogen and carbon sources. NMR data were collected in a buffer containing 20 mM HEPES, 80 mM KCl, 1 mM DTT and 7.5% (v/v) D_2O at pH 7.4. Samples contained a background of 10 mM Mg^{2+} and 5 mM DNA (5′-GGACCATAGGATGGTCC-3′). Backbone amide fingerprints and chemical shifts of WT HNH and variants were determined via the sensitivity-enhanced HSQC (hsqcetf3gpsi2 pulse sequence) on a 600 MHz Bruker NEO spectrometer with the ^{15}N dimension centred at 117 ppm. Histidine ε_1 side chain chemical shifts were measured via HMQC (hmqcphpr pulse sequence) using NMR samples that were successively buffer exchanged into NMR buffer at selected pH values between 6 and 8. The ^{13}C dimension centred at 130 ppm and the ε_1 side chain chemical shifts of HNH were assigned by mutagenesis of H840 to alanine. H840 pK_a values were determined from fitting the ^{1}H chemical shift trajectories to a modified Henderson–Hasselbach equation. Details are given in the Supplementary Methods.

SpCas9 in vitro cleavage assay

Single guide (sg) RNAs and SpCas9 protein used in cleavage assays were produced as described previously[11]. Fluorescently labelled target oligonucleotides were synthesized by Integrated DNA Technologies, with HPLC purification. Cleavage substrates were prepared by mixing complementary oligonucleotides in a 1:2 molar ratio of target strand to non-target strand, heating to 95 °C for 5 minutes and slowly cooling to room temperature. Cas9 protein was first pre-incubated with the sgRNA in a 1:1.25 molar ratio and incubated for 10 minutes at room temperature in the reaction buffer containing 20 mM HEPES pH 7.5, 250 mM KCl, 5 mM $MgCl_2$ and 1 mM DTT. The binary complex was rapidly mixed with ATTO-532-labelled double-stranded (ds) DNA substrate, to yield final concentrations of 1.67 µM protein and 66.67 nM substrate in a 7.5 µl reaction. Reactions were incubated at 37 °C and time points were collected at 1, 2.5, 5, 15, 45, 90 and 150 minutes. Substrate cleavage was stopped by addition of 2 µl STOP buffer mix, consisting of 250 mM EDTA, 0.5% SDS and 20 µg of proteinase K. Reaction products were separated on a 16% denaturing 7 M urea–PAGE gel and imaged using a Typhoon FLA 9500 gel imager. The average from four independent experiments was fit to a single exponential function for every time point (Fig. 3e). The sgRNA sequence used in cleavage experiments was GGAAUCCCUUCUGCAGCACCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUG. The dsDNA cleavage substrate target strand sequence was 5′-[Atto-532]CGACAATACCAGGTGCTGCAGAAGGGATTCCAGATGAGC-3′. The cleavage substrate non-target strand sequence was 5′-GCTCATCTGGAATCCCTTCTGCAGCACCTGGTATTGTCG-3′.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates of the optimized computational models are available in figshare with the identifier https://doi.org/10.6084/m9.figshare.19158080. NMR resonance assignments for the HNH nuclease are available in the BMRB entry 27949. All other data are available from the authors upon reasonable request. Source data are provided with this paper.

References

1. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
2. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
3. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
4. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
5. Casalino, L., Nierzwicki, Ł., Jinek, M. & Palermo, G. Catalytic mechanism of non-target DNA cleavage in CRISPR-Cas9 revealed by ab Initio molecular dynamics. *ACS Catal.* **10**, 13596–13605 (2020).
6. Palermo, G., Miao, Y., Walker, R. C., Jinek, M. & McCammon, J. A. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc. Natl Acad. Sci. USA* **114**, 7260–7265 (2017).
7. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **3**, eaao0027 (2017).
8. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110–113 (2015).
9. Biertümpfel, C., Yang, W. & Suck, D. Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **449**, 616–620 (2007).
10. Zuo, Z. & Liu, J. Structure and dynamics of Cas9 HNH domain catalytic state. *Sci. Rep.* **7**, 17271 (2017).
11. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
12. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
13. Huai, G. et al. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat. Commun.* **8**, 1375 (2017).
14. Zhu, X. et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **26**, 679–685 (2019).
15. Bravo, J. P. K. et al. Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **603**, 343–347 (2022).
16. Pacesa, M. et al. R-loop formation and conformational activation mechanisms of Cas9. *Nature* **609**, 191–96 (2022).
17. Zuo, Z. et al. Structural and functional insights into the bona fide catalytic state of *Streptococcus pyogenes* Cas9 HNH nuclease domain. *eLife* **8**, e46500 (2019).
18. Yoon, H., Zhao, L. N. & Warshel, A. Exploring the catalytic mechanism of Cas9 using information inferred from endonuclease VII. *ACS Catal.* **9**, 1329–1336 (2019).

19. Zhao, L. N., Mondal, D. & Warshel, A. Exploring alternative catalytic mechanisms of the Cas9 HNH domain. *Proteins* **88**, 260–264 (2019).
20. Kästner, J. Umbrella sampling. *WIREs Comput. Mol. Sci.* **1**, 932–942 (2011).
21. East, K. W. et al. Allosteric motions of the CRISPR–Cas9 HNH nuclease probed by NMR and molecular dynamics. *J. Am. Chem. Soc.* **142**, 1348–1358 (2020).
22. Nierzwicki, Ł. et al. Enhanced specificity mutations perturb allosteric signaling in the CRISPR-Cas9 HNH endonuclease. *eLife* **10**, e73601 (2021).
23. Swails, J. M. & Roitberg, A. E. Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. *J. Chem. Theory Comput.* **8**, 4393–4404 (2012).
24. Hansen, A. L. & Kay, L. E. Measurement of histidine pK_a values and tautomer populations in invisible protein states. *Proc. Natl Acad. Sci. USA* **111**, 1705–1712 (2014).
25. Shimahara, H. et al. Tautomerism of histidine 64 associated with proton transfer in catalysis of carbonic anhydrase. *J. Biol. Chem.* **282**, 9646–9656 (2007).
26. Brunk, E. et al. Pushing frontiers of first-principles based computer simulations of chemical and biological systems. *Chimia (Aarau)* **65**, 667–671 (2011).
27. Carter, E. A., Ciccotti, G., Hynes, J. T. & Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **156**, 472–477 (1989).
28. Laio, A., VandeVondele, J. & Rothlisberger, U. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.* **116**, 6941–6947 (2002).
29. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
30. Lee, C., Yang, W. & Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
31. Dürr, S. L. et al. The role of conserved residues in the DEDDh motif: the proton-transfer mechanism of HIV-1 RNase H. *ACS Catal.* **11**, 7915–7927 (2021).
32. Casalino, L., Palermo, G., Rothlisberger, U. & Magistrato, A. Who activates the nucleophile in ribozyme catalysis? An answer from the splicing mechanism of group II introns. *J. Am. Chem. Soc.* **138**, 10374–10377 (2016).
33. Borišek, J. & Magistrato, A. All-atom simulations decrypt the molecular terms of RNA catalysis in the exon-ligation step of the spliceosome. *ACS Catal.* **10**, 5328–5334 (2020).
34. Palermo, G. et al. Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.* **48**, 220–228 (2015).
35. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep.* **22**, 359–371 (2018).
36. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
37. Cisneros, G. A. et al. 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 (2009).
38. Wang, Y. et al. Real-time observation of cas9 postcatalytic domain motions. *Proc. Natl Acad. Sci. USA* **118**, e2010650118 (2021).
39. Palermo, G. Structure and dynamics of the CRISPR–Cas9 catalytic complex. *J. Chem. Inf. Model.* **59**, 2394–2406 (2019).
40. Galburt, E. A. & Stoddard, B. L. Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry* **41**, 13851–13860 (2002).
41. Perez, A. et al. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* **92**, 3817–3829 (2007).
42. Banas, P. et al. Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theor. Comput.* **6**, 3836–3849 (2010).
43. Zgarbova, M. et al. 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 (2011).
44. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational design of particle mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput.* **9**, 2733–2748 (2013).
45. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
46. Case, D. A. et al. AMBER 2020 (Univ. of California, San Francisco, 2020).
47. Parrinello, M., Andreoni, W. & Curioni, A. CPMD (IBM Corporation and Max-Planck Institute, 2000).
48. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
49. Nosé, S. An extension of the canonical ensemble molecular dynamics method. *Mol. Phys.* **57**, 187–191 (1986).
50. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).

Acknowledgements

This material is based on work supported by the National Institute of Health (grant no. R01GM141329, to G.P.) and the National Science Foundation (grant no. CHE-1905374, to G.P.). G.P.L. is supported by the National Science Foundation (grant no. MCB-2143760). This work was also supported in part by the National Institute of Health (grant no. R01GM136815 to G.P. and G.P.L.). M.J. acknowledges support from the Swiss National Science Foundation (31003A_182567). M.J. is an International Research Scholar of the Howard Hughes Medical Institute and Vallee Scholar of the Bert L & N Kuggie Vallee Foundation. Computer time for MD has been awarded by XSEDE under grant no. TG-MCB160059 and by NERSC under grant no. M3807 (to G.P.).

Author contributions

L.N. performed molecular simulations and analysed data. K.W.E. and E.S. performed NMR experiments. J.M.B. and M.P. performed DNA cleavage experiments. P.R.A., R.V.H. and M.A. performed molecular simulations. M.J. supervised DNA cleavage experiments. G.P.L. supervised NMR experiments. G.P. conceived this research, supervised computational studies and wrote the manuscript, with critical input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41929-022-00848-6.

Correspondence and requests for materials should be addressed to George P. Lisi or Giulia Palermo.

Peer review information *Nature Catalysis* thanks Quanjiang Ji, Priyadarshi Satpati, Jeong-Yong Suh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Giulia Palermo

Last updated by author(s): Aug 19, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	AMBER20 (for classical simulations), CPMD4.1 (for ab-initio molecular dynamics and mixed quantum/classical approaches)
Data analysis	AMBER20 (for classical simulations), CPMD4.1 (for ab-initio molecular dynamics and mixed quantum/classical approaches), NMRPipe, NMRFAM-SPARKY

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Atomic coordinates of the optimised computational models are available in figshare with the identifier DOI: 10.6084/m9.figshare.19158080. NMR resonance assignments for the HNH nuclease are available in the BMRB entry 27949. All other data is available from the authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size of all simulations was determined with the aim of reaching converged results. In molecular dynamics simulations, convergence is based on ergodicity, indicating that all the microstates are accessible over the simulation time. Convergence of structural properties from the ensemble population was assessed in comparison with the experimentally known value. For the free energy simulations the sample size was chosen for the forces or energy profiles reaching convergence.
Data exclusions	Analysis of molecular dynamics simulations has been performed on the entire ensemble population. For the calculation of the free energy profiles, the initial part of the simulations was excluded with the rationale of considering the converged trajectories. Convergence of the forces and energy profiles are reported in the supplementary information.
Replication	Molecular simulations were performed in triplicate, with results from all the three replicates reported in the main and supplementary figures. DNA cleavage data arise from four independent experiments (n = 4), with data presented as the average and standard deviation.
Randomization	To maintain the randomization, each replica of molecular dynamics simulations was started with random velocities. The replicates starting from the different initial velocities converged to similar results.
Blinding	Not applicable. Blinding couldn't be implemented in our case because the identity (the chemical characteristics) of each sample needed to be known in the major part of the simulations and experiments that we performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging