

The dual role of histidine as general base and recruiter of a third metal ion in HIV-1 RNase H

Simon L. Dürr^{1,2†}, Olga Bohuszewicz¹, Reynier Suardiaz^{1 ‡}, Pablo G. Jambrina^{1 ¶}, Christine Peter², Yihan Shao³, Edina Rosta^{1*}

¹Department of Chemistry, King's College London, London, SE1 1DB, United Kingdom ²Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany ³Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Parkway, Norman, OK 73019–5251, USA

KEYWORDS: RNase H, QM/MM, nuclease, phosphodiester cleavage, HIV-1

ABSTRACT: RNase H is a prototypical example for two metal ion catalysis in enzymes. An RNase H activity is present in the HIV-1 reverse transcriptase but also in many other nucleases such as *Homo sapiens* (*Hs*) or *Escherichia coli* (*Ec*) RNase H. The mechanism of the reaction has already been extensively studied based on the *Bacillus halodurans* (*Bh*) RNase H crystal structures, most recently using time-resolved X-Ray crystallography. However, kinetic and mutation experiments with HIV-1, *Hs* and *Ec* RNase H implicate a catalytic histidine in the reaction that is not present in *Bh* RNase H, and the protonation of the leaving group also remains poorly understood. We use quantum mechanics/molecular mechanics (QM/MM) calculations combining Hamiltonian replica exchange with a finite-temperature string method to study the cleavage of the ribonucleic acid (RNA) backbone of a DNA/RNA hybrid catalyzed by the HIV-1 RNase H with a focus on the proton transfer pathway and the role of the histidine. The reported pathway is consistent with kinetic data obtained with mutant HIV-1, *Hs* and *Ec* RNase H, the calculated pK_a values of the DEDD residues and crystallographic studies. The overall reaction barrier of ~18 kcal mol⁻¹, encountered in the first step, matches the slow experimental rate of ~1-100 min⁻¹. Using Molecular dynamics (MD) calculations we are able to sample the recently identified binding site for a third transient divalent metal ion in the vicinity of the scissile phosphate in the product complex. Our results account for the experimental observation of a third metal ion facilitating product release in an *Aquifex aeolicus* RNase III crystal structure and the *Bh* RNase H *in crystallo* reaction. Based on our data we are able to show that the third ion and the histidine are key to product release as had been hypothesized.

Introduction

The formation and cleavage of phosphodiester bonds is one of the most essential reactions carried out by enzymes. The general mechanistic features of phosphodiester cleavage are universally conserved in ribozymes, polymerases, nucleases, and a wide variety of other nucleic acid processing enzymes such as integrases, transposases or topoisomerases (See ref¹ and references therein). The phosphodiester cleavage generally follows an associative mechanism and utilizes the well-established two metal ion mechanism.² Two divalent ions are typically bound by one bridging aspartate coordinating both metal ions, another aspartate and a glutamate (referred to as a DDE motif).

Enzymes of the nuclease family are found in all branches of life. Amongst many processes conferred by the nucleases are DNA replication, proofreading, RNA maturation and RNA interference.¹ One of the particularly well-studied nucleases is RNase H. The RNase H fold is conserved in many enzymes with nucleolytic activity including transposases, the argonaute proteins, and the Cas9 enzyme.^{1,3-5}

RNase H is also an essential part of the HIV-1 enzyme machinery as HIV-1 reverse transcriptase (RT) depends on both a polymerase and a RNase H active site localized on two domains for its full functionality. The RNase H subdomain (residues 1021 – 1144 of the gag polyprotein, Figure 1B) is located on the p66 subunit of RT where it catalyzes the removal of the RNA strand in the RNA/DNA hybrid generated by the polymerase.⁶ RNase H was found essential for proliferation of HIV-1.⁷ In higher eukaryotes RNase H1 is thought to be involved in generating RNA primers for replication and the prevention of genomic instability by removal of RNA nucleotides in R-Loops.^{8,9}

Studies of the catalytic mechanism of *Bacillus halodurans* (*Bh*) RNase H¹⁰⁻¹³ and human (*Hs*) RNase H¹⁴ showed the importance of the coordinated movement of two Mg²⁺ ions, metal ion preference of Mn²⁺ and Mg²⁺ over Ca²⁺, and possible proton pathways. Recently, Samara and Yang¹⁰ suggested that a third transient Mg²⁺ ion participates in the catalytic reaction, similarly to polymerases during the reverse reaction.^{10,15,16} Additionally, Samara and Yang¹⁰ also implicate monovalent ions in the *Bh* RNase H catalysis.

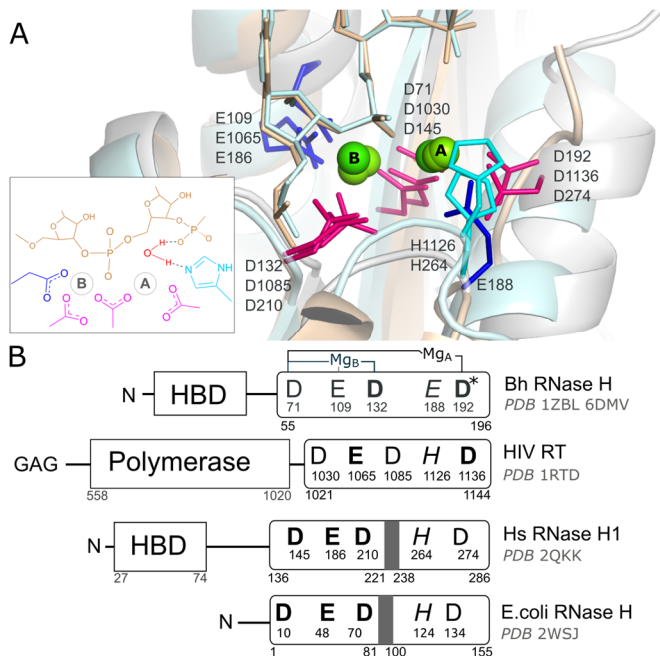


Figure 1 A) Superposition of *Bh* RNase H *in crystallo* reactant after 40ns (PDB 6DMV¹⁰, olive cartoon), HIV-1 QM/MM reactant model (gray cartoon, model based on PDB 1RTD¹⁷/1ZBL³) and *Hs* RNase H (PDB 2QKK¹⁸, light cyan cartoon). The inset shows positioning of the nucleophile and the active site geometry. Respective residues were colored in magenta (Asp), blue (Glu), cyan (His) and green (Mg^{2+}). B) Comparison of the selected RNase H family members and their catalytic sites. Binding of DNA/RNA hybrid is supported by a hybrid binding domain (HBD) or basic protrusion (dark gray). Residues shown in italic, indicate reduced activity when substituted. Residues shown in bold, indicate no activity for the E→Q, D→N or H→A mutation. *partially active in Mn^{2+} , HBD hybrid binding domain, GAG group specific antigen.

Several key details about the reaction mechanism are still not fully consistent with available data. Kinetic experiments indicate a crucial role of a catalytic histidine in the HIV-1, *Hs* and *E. coli* (*Ec*) RNase H reaction (H1126, H264, H124 in Figure 1).^{19,20} All earlier computational studies of the RNase H catalytic reaction were either conducted with the *Bh* model lacking this crucial histidine or the histidine was not included in the QM region.¹²⁻¹⁴ In similar nucleases with nucleolytic activity, such as CRN-4, TREX1, 3'hExo, or RNase T^{1,21-24}, a conserved histidine residue (part of the conserved DEDDh motif²⁵) is thought to deprotonate the water nucleophile and for one of these enzymes, namely the ϵ -subunit of *Ec* DNA pol III²⁶, QM/MM calculations support this claim.²⁷

To address several of the outstanding questions regarding the HIV-1 RNase H catalytic mechanism, namely the nature of the general base, the identity of the proton donor of the leaving group, and the role of the third metal ion, we employed classical molecular dynamics (MD) simulations and mixed quantum/classical (QM/MM) simulations utilizing the B3LYP density functional level of theory. We determined the multidimen-

sional free energy surface underlying the catalytic reaction of HIV-1 RNase H with two metal ions bound in the active site by combining Hamiltonian replica exchange (RE), umbrella sampling (US), and finite-temperature string method calculations. In addition, we studied the cleavage product resulting from the action of RNase H using classical MD simulations to investigate the role of a third metal ion. Our simulations showed a reaction pathway that differs from those of earlier computational studies but are consistent with kinetic and thio-substitution experiments for *Ec* RNase H and *Hs* RNase H. The main differences to other studies of the RNase H catalytic mechanism are the nature of the general base and the source of the proton for leaving group protonation. Based on our MD simulations we propose that a third divalent cation, necessary for the displacement of the cleaved RNA strand, is recruited from the bulk with the aid of the histidine. In a structural comparison with the catalytic motifs of other nucleases, we are able to show that a conserved histidine plays a dual role in many nucleolytic reactions irrespective of the exact metal ion coordination of the two central Mg^{2+} ions in the active site.

Materials and Methods

Initial coordinates: The crystal structure of the catalytically trapped HIV-1 RT DNA/RNA complex by Huang et al.¹⁷ (PDB: 1RTD) with 3.2 Å resolution was used as a starting model. The active site of RNase H was modeled with SWISS-MODEL²⁸ based on the higher resolution structure of D192N *Bh* RNase H catalytic domain by Nowotny et al.³ with 2.2 Å resolution (PDB: 1ZBL) to achieve appropriate coordination of the metal ions and proper orientation of the substrate. The obtained model was compared with the recently published catalysis-ready structure of RNase H in HIV-1 RT (RMSD 0.822 Å, PDB 6BSH).²⁹ The DNA/RNA hybrid of the crystal was replaced with the polypurine tract (PPT) sequence.³⁰ The RNase H reaction has a pH profile with maximum activity at pH 7.5.²⁰ All carboxylate active site residues were therefore considered deprotonated at first and H1126 was modelled as ϵ -protonated, while standard protonation states were used for all the remaining titratable residues. The model was subject to classical MD equilibration of 20 ns and QM/MM minimizations to relax the active site before the restrained energy minimizations along the RC were performed and the relaxed geometries were then finally used for QM/MM simulations. The details about the simulations are shown below.

QM/MM energy minimizations: To carry out the QM/MM energy minimizations, the system was trimmed to a sphere of 20 Å centered on the scissile phosphate group. Atoms further than 15 Å away from that group were kept frozen throughout the simulations. The region treated with molecular mechanics was 4164 atoms in size. The QM region contained 116 atoms

including two Mg^{2+} ions, the RNA backbone of the active site adenosine, the guanosine in 3' direction of the RNA strand and the side chains of residues Q1062, D1030, E1065, D1085, D1136 and H1126. The hydrogen link atoms were placed between the C_{β} and C_{α} bonds for the aspartates, glutamine and histidine or the C_{γ} and C_{β} bonds for the glutamates. Link hydrogens were restrained between the original bond partners and their bond length was scaled by a factor of 0.7261. Twelve water molecules were included in the QM region of which three coordinate Mg_A . To probe for the different proton transfers proposed in the literature¹²⁻¹⁴, we also included all water molecules that make hydrogen bonds to the scissile phosphate, to the *Bh* D132 equivalent D1085, to the O2' and the Q1062 amide oxygen. The charge of this QM region was -2. To probe for a possible new protonation pathway, a water molecule in the solvent channel egressing from the O2'/Q1062 into the bulk solvent was protonated. The total number of atoms therefore was 117 and total charge of the QM region was -1 for the string calculation. All calculations were performed with the CHARMM 27 force field³¹ and charmm40b2³² as the MD engine interfaced with Q-Chem 4.4 using full electrostatic embedding of the MM region.³³ All calculations were carried out with the 6-31+G(d) basis set. Density functional theory calculations and the B3LYP³⁴ exchange-correlation functional were used. The reaction coordinate (RC) was defined by the bond distances between reactive atoms. The RC included the proton transfer (Q_p) and electron transfer (Q_e). Q_{ep} is taken as $Q_{ep} = Q_e + Q_p$. The Q_e coordinate includes r2-r1 (WATnuc:OH2 - RNA:P -RNA:O3', Figure 4) . The Q_p coordinate includes r4-r3 (H1126:NE2 - WATnuc:H - WATnuc:O, and r6-r5 (RNA:O3' - RNA:H2' - E1065:OE1, Figure 4)).

String calculations To determine the reaction free energies and the corresponding minimum free energy pathway of the mechanism, we performed finite temperature string calculations.³⁵ We selected 19 coordinates that define the space for the reaction mechanism (see SI), which include all bond breaking and forming distances for the proton transfers and phosphate cleavage. We performed 60 iterations during the string free energy calculations, running 100 steps of QM/MM dynamics at each iteration using Langevin dynamics with a timestep of 1 fs. Each coordinate was restrained using a force constant of 150 kcal mol⁻¹ Å². The pathway was defined via 26 windows obtained from the QM/MM minimizations. After each iteration, a new pathway was calculated by fitting a polynomial of order 8 at each coordinate, and dividing the resulting 1 dimensional path within the 19 dimensional coordinate space into 26 equal arclengths.¹³ The initial coordinates were swapped using a Metropolis Monte Carlo Hamiltonian replica exchange algorithm after each iteration.¹³

MD simulations of the product complex: MD simulations of the product state were setup with the CHARMM-GUI³⁶ and used the CHARMM 36 force field

due to its better treatment of nucleic acids³⁷. The product state structure from the QM/MM calculations of HIV-1 RT was solvated in a cubic water box with 138 Å in each dimension and 10 Å distance to the protein. The phosphate groups were modelled as completely deprotonated. Periodic boundary conditions were applied. The system was neutralized with 150 mM potassium chloride which has been shown to be optimal for catalysis in *Hs* RNase H²⁰ and in a subsequent step either 0 mM, 6 mM or 50 mM magnesium chloride was added. The concentration of 6 to 10 mM $MgCl_2$ has been shown to be optimal for catalysis in vitro^{20,38} and the concentration of 50 mM is considered inhibitory.³ After minimizing for 10 000 steps, the system was equilibrated for 5 ns with restrained protein backbone atoms to allow for relaxation of side chains and to heat up the system in NVT conditions. The restraints were then lifted and the trajectories for analysis were collected over 100 ns in NPT conditions. A time-step of 2 fs was used. Non-bonded interactions were modeled with scaled 1-4 setting in NAMD2 and a cutoff of 12 Å with switching distance of 10 Å. The Particle Mesh Ewald method was used to evaluate long range electrostatic interactions and the system was coupled to a Langevin thermostat at a temperature of 303.15 K and a Langevin barostat at standard pressure. All systems were run in triplicates. NAMD 2.9³⁹ was used as the MD engine.

Results and Discussion

General scheme of the reaction: Our study does not deviate from the general scheme of two metal ion catalysis. RNase H catalyzed the cleavage of the RNA strand of the DNA/RNA hybrid via two catalytic magnesium ions. The nucleophilic water (WAT_{nuc}) was bound to Mg_A and oriented for nucleophilic attack on the phosphate. The O3' leaving group was bound to Mg_B (Figure 2A). The reaction proceeded via formation of a pentavalent phosphorane transition state (TS1, Figure 2B) after the WAT_{nuc} is deprotonated, a high energy intermediate state (INT, Figure 2C) and a second transition state (TS2, Figure 2D), where the leaving group is protonated.

Deprotonation of WAT_{nuc} by H1126: The order of events for the water deprotonation and the phosphodiester cleavage process was determined with extensive energy minimizations of the QM/MM system followed by finite temperature string free energy calculations. The geometry was optimized along the Q_{ep} RC driving the system from reactant to product. The geometry minimized path was used in the first iteration to determine the minimum free energy pathway in subsequent string free energy calculations. Initially, the first coupled proton transfer and phosphate cleavage step until the formation of the intermediate was probed during minimizations. The Q_e reaction coordinate alone was used to drive the phosphate cleavage via the O3' leaving oxygen, the phosphorous atom and the WAT_{nuc} oxygen. We did not specify *a priori* the WAT_{nuc} deprotonation.

nation mechanism that facilitates the nucleophilic attack. The QM region included the potential proton acceptors: the 3' downstream *pro-R_p* oxygen, the N_ε of H1126 and the *pro-R_p* oxygen of the scissile phosphate.

We consistently found that the N_ε of H1126 acted as a proton acceptor. Additionally, a hydrogen bond was formed between the 3' downstream *pro-R_p* oxygen and the other WAT_{nuc} proton.

To determine the effect of the 3' downstream *pro-R_p* oxygen, H1126 was excluded from the quantum mechanical treatment. In QM/MM minimizations with the smaller QM region, deprotonation was carried out via the 3' downstream *pro-R_p* oxygen. This is consistent with earlier simulations of *Bh* RNase H¹³ that lacks the histidine residue. To verify that the histidine is the stronger general base the product state of the reaction was modelled by moving the WAT_{nuc} proton pointing towards H1126 to the O3' leaving group yielding a product complex with cleaved phosphodiester bond and a monoprotonated phosphate with the hydrogen pointing towards the 3' downstream *pro-R_p* oxygen. In the subsequent unrestrained QM/MM energy minimization of this system, the WAT_{nuc} proton jumped to the N_ε of H1126. This clearly demonstrates the higher basicity of H1126 compared to the 3' downstream *pro-R_p* oxygen.

As H1126 was very flexible in the MD simulations, it is likely that H1126 shuttles both protons from WAT_{nuc} to the solvent and acts as the general mediator for proton transfers to and from the solvent. It could also play a role in attracting the active site Mg²⁺ ions to the active site, similar to the role of E188 in *Bh*.¹⁰ The *pro-R_p* oxygen of the scissile phosphate was not considered as a possible proton acceptor, as mutational studies showed that thio substitution of the *pro-R_p* oxygen does not decrease the catalytic rate.⁴⁰

Alla and Nicholson²⁰ have proposed a dual functional role of H264 in *Hs* RNase H acting both as a proton acceptor and a facilitator of the product release. The mutation of the catalytic histidine in *Hs* RNase H (H264A) reduces the *k_{cat}* 100-fold²⁰ but has no effect on the *K_m*, strongly supporting the role of histidine as the proton acceptor. A similar effect of the H→A mutation in RNase H can be seen in *Ec*, where the H124A substitution decreases the *k_{cat}* 50-fold and increases *K_m* only by 4-fold.¹⁹ The H1126F mutation in HIV-1 RNase H renders the enzyme inactive.⁴¹ In addition, Tisdale et al.⁷ showed that even the conservative H1126N mutation makes the virus non-infective, corroborating the crucial role of H1126 for RNase H in the virus life cycle. It is likely that the replication of the viral genome is severely impacted by a slower RNase H catalytic rate.

The pH profile of the reaction^{20,42} with maximum activity at pH 7.5 and low activity at low pH is also consistent with the role of the histidine which is protonated at low pH, diminishing the catalytic activity.

As previously argued^{12,13} the effect of the 3' downstream *pro-R_p* oxygen (with sulfur substitution 86% *k_{cat}* decrease⁴³) is most likely an indirect effect, where the 3' downstream *pro-R_p* oxygen increases the catalytic rate by correctly positioning the nucleophile. Another cause of the rate diminishing effect of the 3' downstream *pro-R_p* thio substitution is likely the hindered ability to stabilize a third metal ion (see MD section).

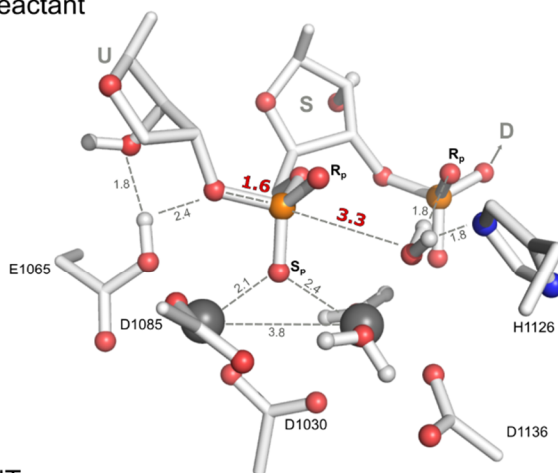
Our QM/MM calculations explain all the aforementioned experimental evidence, evincing that H1126 is likely to act as the proton acceptor in contrast to previous calculations for which the conserved histidine was not included in the QM region.

Protonation of the leaving O3' via E1065: The product of the first reaction step is a cleaved phosphodiester bond with the O3' leaving oxygen tightly bound to Mg_B (Figure 2C). In earlier studies this intermediate has also been observed and found to be highly unstable (Figure 3), and the protonation of the leaving oxygen was found to be essential for the stabilization of the cleavage product.¹³ However, the source of the proton was not established. As we will show below, we found that protonation via the E1065 is an energetically feasible pathway.

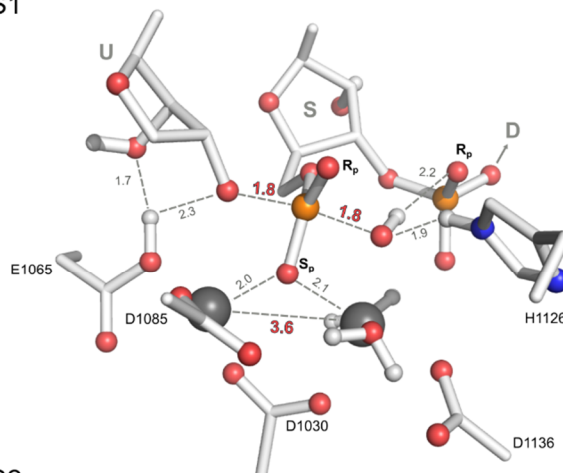
Initially, we tested whether the O3' leaving oxygen could be protonated via a file of water molecules in channel C1 passing under the substrate to the bulk solvent (See Figure 5). Based on evidence from kinetic experiments with mutated O2',⁴⁰ we assumed that the proton could be shuttled to the O3' via the O2' as there is evidence for a role of the O2' in the reaction.⁴⁰

To test this hypothesis a hydronium was placed in the solvent channel. Energy minimizations showed that the proton is transferred to the terminal water molecule in the solvent channel close to the O2', where it is coordinated by a conserved serine (S1086) and a conserved glutamine (Q1062) (see Seq. alignment Fig. S1). With restrained energy minimizations including the distances between the terminal water proton, the O2', the H2' and the O3' distance and the restraints from the above simulations we tried to obtain a potential energy estimate of this pathway. However, we consistently found that the proton was not directly transferred to the O3' but via E1065, which was not included in the restraints. We therefore concluded that E1065 could temporarily store the proton and then transfer it to the O3'. The restraints in the QM/MM minimizations to generate the initial path for the string calculations were modified to account for this pathway by including only the distances between E1065:H, O3', P, O_{nuc}, H_{nuc} and H1126:N_ε in the RC. With this final reaction coordinate a good starting string was constructed conducting iterative QM/MM energy minimizations along the RC. In the string calculations, a protonated E1065 was modelled as reactant. The potential energy barrier of the protonation of E1065 could be determined in separate restrained energy minimizations to be about 16 kcal mol⁻¹.

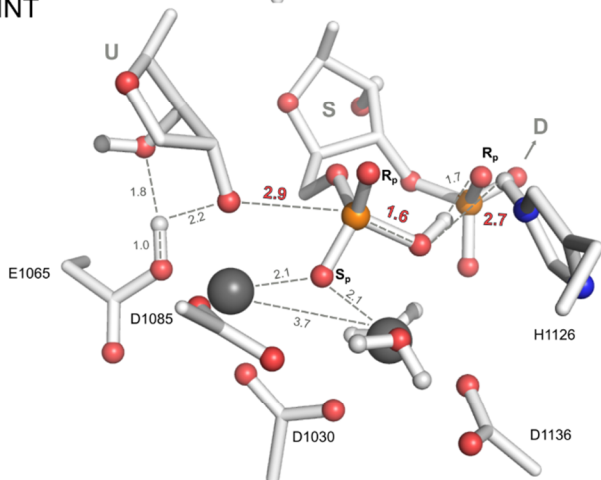
A Reactant



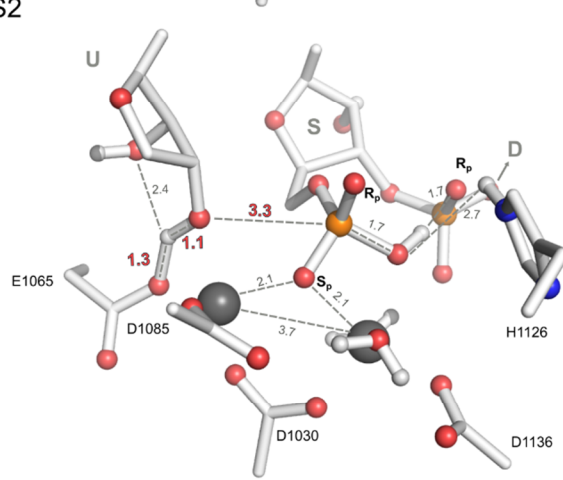
B TS1



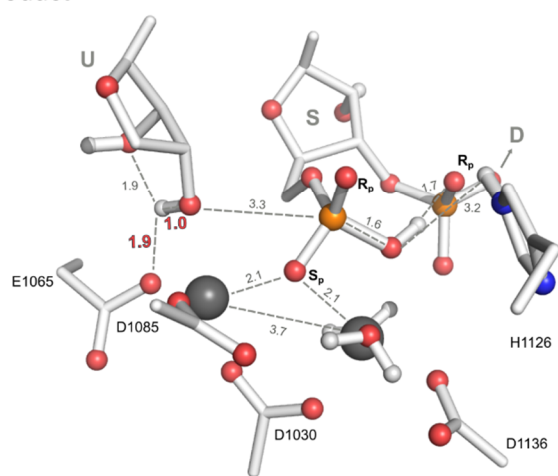
C INT



D TS2



E Product



F Substrate release

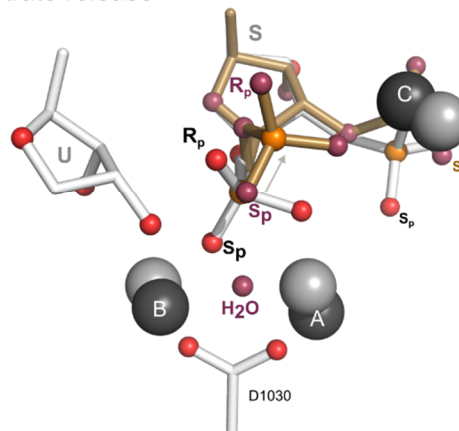


Figure 2: Calculated catalytic mechanism of the HIV-1 RNase H reaction. A) Reactant state B) Phosphorane transition state (TS1) after WAT_{nuc} deprotonation C) Intermediate state with the $\text{O}3'$ leaving oxygen bound to Mg_B D) Proton transfer from E1065 to the $\text{O}3'$ at TS2 E) Product of the HIV-1 RNase H cleavage, the remaining proton of WAT_{nuc} can be abstracted by H1126 after shuttling the first proton to the bulk solvent via sidechain flipping F) Hypothesized substrate release inferred from Aa-RNase III crystal structure⁴⁶ (major conformation scissile phosphate in olive sticks, Mg^{2+} in dark grey spheres, PDB 2NUG) and MD simulation of HIV-1 RT/RNase H product complex at 50 mM Mg^{2+} at 30 ns in replicate 1 (light grey spheres, white sticks). The Mg^{2+} ion C bends the scissile phosphate and the *pro-S_p* oxygen is replaced with a solvent water (sphere in magenta). U is upstream nucleotide, S is the nucleotide directly in the active site containing the scissile phosphate, D is the 3' downstream nucleotide. *pro-S_p* and *pro-R_p* oxygen are abbreviated as *S_p* and *R_p*. An interactive version with 3D view of the reaction (Panels A-E) together with Figure 3 is available online (<https://dev.simondurr.eu/hiv/>).

To support the possible role of E1065 as a proton transfer, we calculated its pK_a using PROPKA 3.1⁴⁴ obtaining a value of 7.80, significantly larger than its model pK_a of 4.5, which suggests that E1065 could accept a proton throughout the catalytic cycle similar to our finding in the QM/MM energy minimizations (Table S1). In addition, various experimental evidence supports the role of E1065 in the proton transfer. Firstly, the homologous *Ec* E48D mutant is substantially less active with a 95% decrease in k_{cat} .¹⁹ This decrease is most likely connected with suboptimal positioning and change in pK_a . Secondly, it was found that *Ec* E48 is less important for Mg^{2+} binding compared to the three other aspartates as the Mg^{2+} dissociation constant for the *Ec* E48Q mutant is 1.1 mM compared with 0.71 mM for the wild type but 18.8 mM for *Ec* D70N.⁴⁵ As a result of the latter, Kanaya et al.⁴⁵ proposed that the glutamate of the DEDD motif should play an additional role in the catalysis. Our results support that hypothesis, showing the role of E1065 as a mediator in the proton-transfer reaction needed for the stabilization of the leaving group and competition of the cleavage reaction.

In the *in crystallo* reaction in *Bh* RNase H it can also be observed that if there is only potassium present, *Bh* E109 and *Bh* D71 are disordered.¹⁰ Finally, Uchiyama et al.⁴⁰ have measured kinetics of *Ec* RNase H with modified substrate at the O2' position of the nucleotide immediately upstream of the scissile bond which also support the pathway via E1065. Substitution of the O2' into an amino, fluoro or methoxy group or removal of the O2' did not affect the affinity for the substrate but greatly reduced the reaction rate.⁴⁰ E1065 interacts with the O2' through a hydrogen bond both in the crystal and in simulations. Altogether this supports a dual role of E1065 in recognizing the RNA strand in the DNA/RNA hybrid via the hydrogen bond and then by shuttling the proton to the leaving O3'. Also consistent with this hypothesis is the 50% decrease in catalytic activity in *Ec* by mutation of a serine close to the active site (*Ec* S71A, HIV-1 S1086).¹⁹ This serine is conserved in *Bh*, *Hs*, *Ec* and HIV-1 RNase H (Figure S1) and, in our simulations, the proton transfer originates from the terminal water molecule bound to S1086. The binding site for this terminal water is favorable as can be inferred from its presence in multiple crystal structures (e.g. 1ZBL, 2G8I, and 2R7Y, 6DOG) of *Bh* RNase H (Figure 5).¹³

Protonation of D1085, which was previously considered as the most likely protonation pathway for *Bh* RNase H¹³, was not attempted as the pathway via E1065 is in very good agreement with the experimental kinetics. In addition, our pK_a calculations show that E1065 (pK_a 7.80) is more likely to get protonated than D1085 (pK_a 2.71). Thus, E1065 can better stabilize the proton ultimately donated to the O3' alkoxide after the intermediate has formed because of its higher pK_a value. Another contributing factor is that the energetic cost of rotating the D1085 necessary for protonation is not necessary if

protonation occurs via the pre-positioned E1065. The protonation pathway via the *pro*-R_p oxygen of the scissile phosphate is not consistent with available experimental findings wherefore it was not considered as a potential proton mediator in our calculations. Uchiyama et al.⁴⁰ demonstrated that sulfur substitution of the *pro*-R_p oxygen of the scissile phosphate does not decrease activity. A decrease would however be expected for sulfur substitution of the *pro*-R_p oxygen because the phosphorothioate is a weaker base. It can therefore be concluded that protonation via E1065 is the most likely route for the protonation of the leaving group in HIV-1 RNase H and possibly in the other members of the family (*Bh* E109, *Ec* E48, *Hs* E186) in contrast to earlier findings.

Free energy profile of the reaction: There were two maxima in the one-dimensional energy profile projected onto the Q_{ep} RC, corresponding to the two transition states (Figure 4C). The formation of the phosphorane TS (TS1) was the highest point in the free energy profile with ~ 18 kcal mol⁻¹. The intermediate state was ~ 4 kcal mol⁻¹ lower in energy compared to TS1. To form the product via TS2 another energy barrier with ~ 3 kcal mol⁻¹ had to be surmounted to protonate the leaving group. From the plot of the free energy surface on the Q_e and Q_p coordinates it can be seen, that first a proton transfer, then an electron transfer and then again a proton transfer occurs (Figure 3). These barrier heights are in good agreement with experimentally determined barriers calculated from kinetic experiments to 16.7 to 20.5 kcal mol⁻¹ in *Ec*^{43,47} and *Hs* RNase H.^{20,48} The measured k_{cat} with modified substrate in HIV-1 RNase H is 1.01 to 0.11 min⁻¹ corresponding to a barrier of 19.8 to 21.2 kcal mol⁻¹.⁴⁹ The modified substrate is, however, likely to have negatively impacted the k_{cat} for the measurement of HIV-1 RNase H.¹² The k_{cat} values measured with the human and the *Ec* RNase H are therefore considered to be closer to the real value in HIV-1.

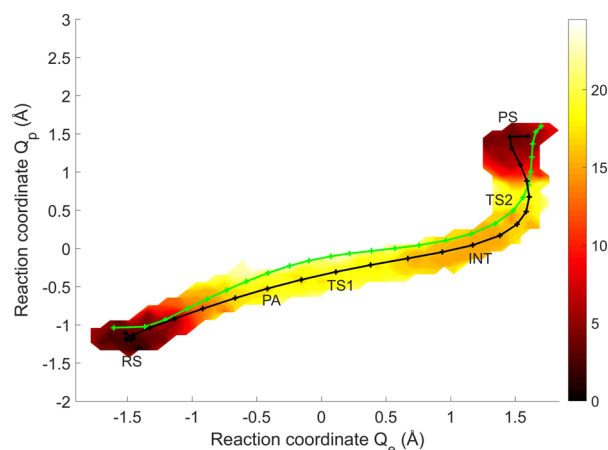


Figure 3: Free energy profile of the reaction from the string simulations with RS reactant state, proton abstraction (PA), transition state 1 (TS1), intermediate (INT), transition state 2 (TS2) and product state (PS). Starting string in green, optimized final string in black. Colorbar indicates ΔG in kcal mol⁻¹. A zoomable and interactive version is available online.

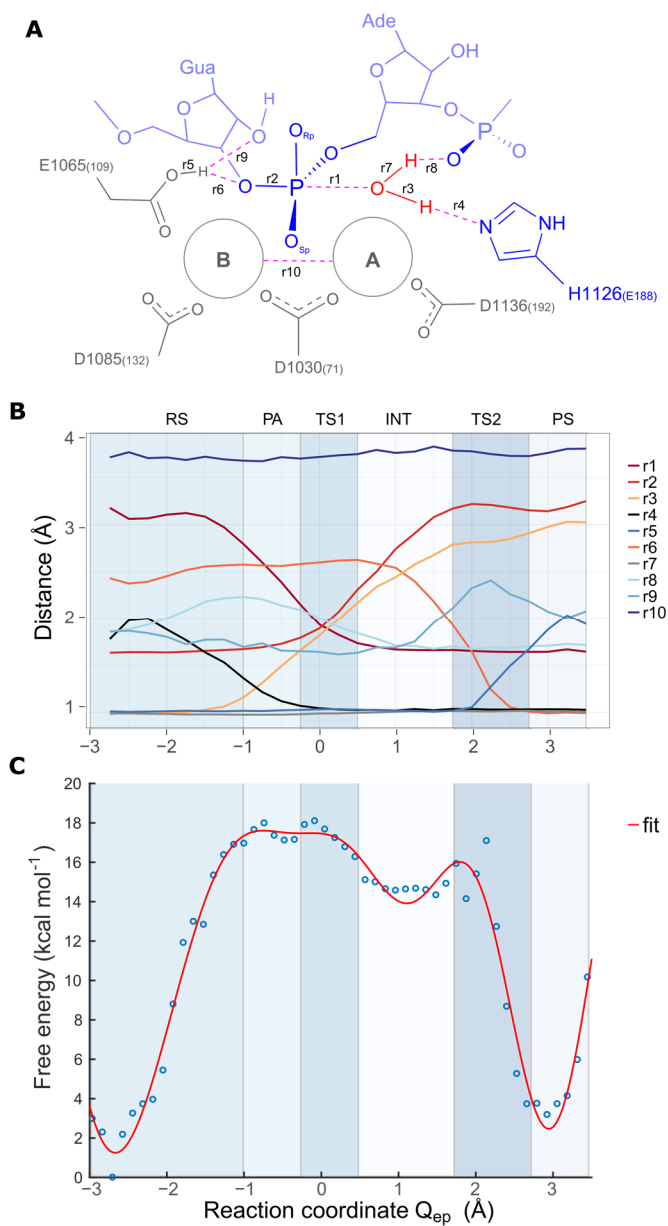


Figure 4: A) Scheme of the catalytic site of the HIV-1 RNase H in the reactant state. Residues in brackets are *Bh* equivalent residues. B) Distance plot of the distances specified in A. Average of distance in the last 1ps in each window is plotted. The shaded rectangles represent RS reactant state, PA proton abstraction, TS1 transition state 1, INT intermediate, TS2 Transition state 2 and PS product state. C) Free energy reaction profile from string calculations projected onto Q_{ep} RC with fitted curve.

Role of monovalent ions: Samara and Yang¹⁰ implicate monovalent potassium ions in the catalysis based on the *in crystallo* analysis of the reaction mechanism in *Bh* RNase H. In *Bh* RNase H, K^+_{w} (site F in Figure 5) coordinates the scissile phosphate and K^+_{U} (site C in Figure 5) together with a transiently bound Mg^{2+}_c activates the nucleophile. *Bh* RNase H likely relies on the cation trafficking to activate the nucleophile because it lacks the histidine residue, that is present in other members of the RNase H family. Therefore, HIV-1, *Hs* and *Ec* RNase might be less reliant on monovalent ions as the histidine activates the nucleophile and therefore replaces the function of the monovalent ions. This is also related to the slightly different active site architecture with *Bh* K197 able to point inwards stabilizing the intermediate, which is not possible for HIV-1 K1137 (Figure S5).

Role of third ion for product release: A third divalent ion was also implicated in the mechanism by Samara and Yang.¹⁰ However, as they deduce importance of the third divalent ion from the *in crystallo* reaction further investigation into the dynamics of the ion recruitment for the real reaction in solution and in the presence of the histidine is still needed. Experimentally, it has been observed that at high Mg^{2+} concentrations (>50mM in *Bh* RNase H) the reaction stalls^{10,20,50-52}, which has led to the hypothesis that a third metal ion could bind to the product or disturb the nucleophilic attack even before it was detected in the *in crystallo* reaction. In their recent work Samara and Yang¹⁰ have shown that the third divalent metal ion coordinates to the scissile phosphate and to the 3' downstream phosphate and that product release can only take place after its departure. Interestingly, H→A mutation in *Hs* or *Ec* RNase H removes the attenuation of the reaction at high Mg^{2+} concentrations, suggesting that the catalytic histidine plays an important role in the recruitment and displacement of the third metal ion.^{20,50} We therefore investigated the product complex of the cleavage reaction in HIV-1 RT/RNase H at the catalytic (6mM), zero and high (50mM) $MgCl_2$ concentrations with classical MD simulations. In multiple 100 ns long MD simulations we found two prominent binding sites for cations in the product complex. The term “binding site” used in this context refers to the ion sampling a small region of space for some extended period of time close to the active site and shall not be indicative of first shell binding via displacement of one of the first shell water ligands of the solvated K^+ or Mg^{2+} ions.

One of the sites (F in Figure 5, Figure S3) was located close to D1085, Mg_A and Mg_B and was populated mostly by potassium ions. This binding site was also occupied during the reaction for *Bh* RNase H crystal structure and was probably populated because of the double negative charge of the scissile phosphate.

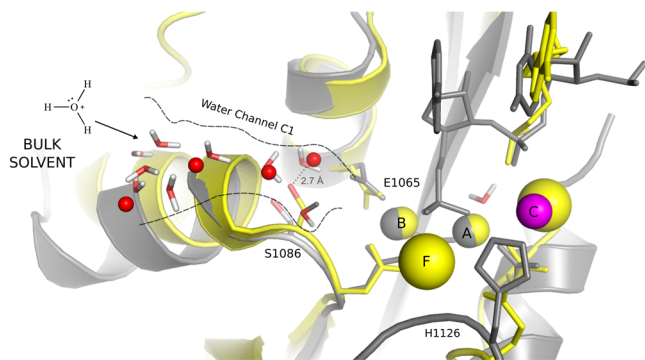


Figure 5 Alignment of MD (gray) to 600 s product structure of *Bh* RNase H (yellow, PDB 6DOG¹⁰). Water channel C1 with water molecules from MD (as sticks) and from crystal (red spheres) structures reaches E1065. Mg²⁺ ion binding sites are also labelled from MD (magenta sphere, location C), from crystal structures (yellow spheres, locations A, B, C, F), and in QM/MM and MD calculations (gray spheres, locations A and B). S1086 keeps the terminal water molecule in place with a hydrogen bond (O-O distance in 6DOG, 2.7 Å)

The second binding site (ion C in Figure 2F, Figure 5, Figure S3) was located between the scissile phosphate and the 3' downstream phosphate, and was populated by Mg²⁺ and K⁺ ions alike. C was also found to be occupied in the product state of the *in crystallo* reaction.¹⁰ The length of the binding events for the ions varied within the replicates from very short binding (~5ns) to quite stable association with the scissile phosphate with retention times of up to 80 ns in binding site C. The relation of potassium vs. magnesium binding is biased towards more potassium ions as they were more abundant in the simulation (150 mM K⁺ vs 50 mM Mg²⁺). As is expected from the classical description of the ions, we do not observe loss of the first solvation shell during the relatively short simulation time. Instead, we observe coordination to the phosphate oxygens mediated via second shell interactions.

The simulations show that the binding of a metal ion in site C causes conformational changes in the RNA strand of the hybrid that disrupt the interaction of the RNA with the DNA strand (Figure S4). Complete dissociation could not be observed likely due to the length of the simulation, the length of the hybrid and/or due to the limitations of the nucleic acid force field. Nevertheless, a general trend for destabilization is visible (measured in terms of RMSD of the active site nucleotide and the two nucleotides in the 3' direction of the RNA strand, Figure S4).

As suggested for the H→A mutation in *Hs* RNase H²⁰, the catalytic histidine H1126 placed on a flexible loop next to the active site seemed to participate in the binding of the third ion. We measured the position of the histidine using its Nα-Cα-Cβ-Cγ dihedral angle, which varied from -45° to 60° for the inwards facing position ready for the catalysis and from -170° to -180° for the outwards position ready for metal ion recruitment and shuttling of the abstracted proton to the solvent (Figure S2). When a third metal ion was in the binding site Mg_C,

we found the H1126 to be almost exclusively facing inwards as a second shell ligand of the third ion (Figure S3). The metal ion dissociated from the binding site shortly after the histidine changed its conformation from in- to outwards in several of the replicas, further supporting the role of the histidine in recruitment of a third ion similarly to that suggested by Alla and Nicholson²⁰ experimentally. This effect is most likely electrostatic, as first shell coordination of a Mg²⁺ or K⁺ by the histidine is not observed in our simulations. This is congruent with findings of low affinity of nitrogens for earth alkali metals.⁵³ In the *Bh in crystallo* reaction, the residue K196 placed at the C-terminus on a flexible loop stabilizes the scissile phosphate. In HIV-1 RNase H a lysine (K1137) is present as well, which, as the C-terminus of the HIV-1 protein extends further, is placed in an alpha helix of HIV-1 RNase H instead of the C-terminal loop of *Bh* RNase H. K1137 in HIV-1 RNase H points away from the active site and is too far away to stabilize the cleaved phosphate (Figure S5).

A possible hypothesis for the function of the third divalent ion in the catalysis and product release can be inferred from Aa-RNase III⁴⁶ and from the *in crystallo* study of the *Bh* RNase H mechanism.¹⁰ Aa-RNase III performs a similar 3'-5' endonucleolytic cleavage like RNase H with the help of two metal ions but the architecture of its active site is slightly different (DDEE motif). In the crystal of the product complex of Aa-RNase III (res. 1.7 Å, PDB: 2NUG), two conformations of the scissile phosphate were observed with different occupancies. The minor conformation with occupancy 0.3 closely resembles the product state of the HIV-1 RNase H reaction from the QM/MM with the scissile phosphate pro-Sp oxygen coordinating both Mg_A and Mg_B. However, the major conformation in the crystal with an occupancy of 0.7 has the scissile phosphate out of the active site, bent away from Mg_A and Mg_B and closer to Mg_C (Figure 2E). The scissile phosphate O_{sp} is replaced by a water molecule. Gan et al.⁴⁶ have proposed that the minor conformation is a model of the complex directly after the reaction took place and the major conformation is the first step towards product release in Aa-RNase III. The product release is facilitated by Mg_C. In the major conformation, the scissile phosphate is a first shell ligand of Mg_C.

Interestingly, the main binding site for the third ion and the scissile phosphate in the HIV-1 RNase H MD simulations are in the same position as in the minor conformation of the Aa-RNase III crystal, suggesting similarity between the Aa-RNase III and RNase H product release mechanism. The stabilization through the third ion likely has stabilized this state in Aa-RNase III enough in order to be detected by conventional crystallography.

Attempts to crystallize the *Bh* RNase H product complex with 100% occupation of a third Mg²⁺ have not been successful⁵² and in the *in crystallo* reaction the third ion was only detectable in 500mM Mn²⁺ or in 5mM Mg²⁺/200mM Li⁺ buffer. It was found that the concentration

of Mg^{2+} needed for optimal catalysis is higher than the concentration to fully occupy the A and B sites.¹⁰ In analogy to our observations, Samara and Yang found the third ion to bind to the product state.¹⁰ In the crystal it can be seen that with Mg_C the 5' phosphate product is shifted by 1 Å.¹⁰

The suggested role of the histidine in the recruitment of the third ion is consistent with the observation that for H→A mutation the attenuation of RNase H activity at high Mg^{2+} concentration is lost.^{20,50} *Bh* E188A mutant, which has a lower activity, also displays no attenuation at high Mg^{2+} concentrations.³ Nowotny and Yang⁵² have attributed the lower activity of *Bh* E188A to an abnormally high affinity to the cleavage product which supports the interpretation that a third metal ion is important for the product release. In *Hs* RNase H264A mutation increases the rate constant for the product release 12-fold which further supports the role of histidine in recruitment of the third ion.²⁰

Similarity to other nucleases

The RNase H topology is shared between many catalytic cores of enzymes that process nucleic acids. We therefore strategically mined the protein data bank for similar catalytic motifs to obtain a picture of the generality of the catalytic mechanism using the histidine and a third ion in cleaving nucleic acids. We used VMD to automatically select structures based on certain criteria.⁵⁴ As criteria we used the presence of a histidine within 5 Å of two Mg^{2+} , Ca^{2+} or Mn^{2+} ions in all structures with resolution better than 2.5 Å and manually inspected the resulting structures. From our analysis, we could determine that indeed catalytic sites of various nucleases are similarly composed of glutamates, aspartates and histidines as has been described before,^{1,55} but we were able to further reason the positioning of some of the residues based on the results presented in this article and the recently available *in crystallo* study of the reaction¹⁰ as well as the QM/MM dynamics of the Cas9 active site.⁵⁶ All of the identified enzymes belong to the RNHL superfamily.⁵⁵

While the DEDD motif is relatively common among nucleases^{1,58}, we identified three different subclasses of this catalytic motif with histidine based on the PDB mining. Two with Glu bound to metal B (RNase H-like, RuvC-like, Figure 6AC) and the other with Glu bound to metal A (RNase T-like, Figure 6B).

Both the RNase H-like motif as well as the RNase T-like motif have been commonly referred to as DEDD motifs. The RuvC-like motif is referred to as DDE motif.⁵⁶ For the RNase T-like structures, Zuo and Deutscher²⁵ established the DEDDh nomenclature to differentiate DEDD nucleases that have histidine from those with a tyrosine (DEDDy) in their active site. It is well known that the catalytic mechanism of these nucleases is strongly related to the coordination of the catalytic magnesium ions^{13,18}, therefore the RNase H-like and RNase T-like motifs are not the same DEDDh motif even

though the position of the histidine is very similar (Figure 6). In both types, the histidine is placed on a loop segment coordinating the nucleophile (Figure S2).²⁴

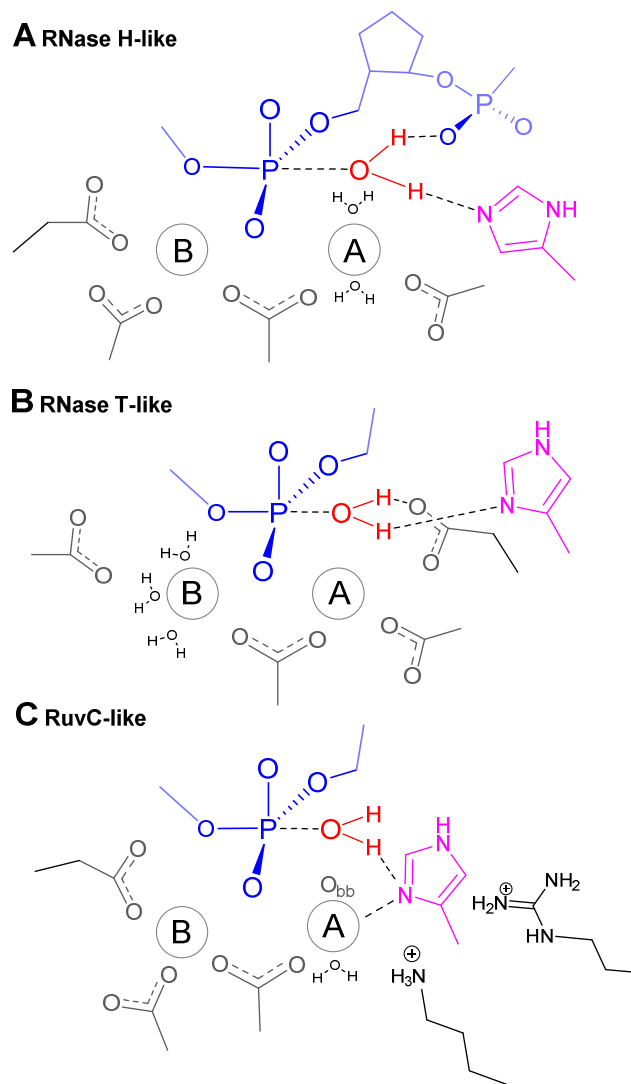


Figure 6: Comparison of the three main types of catalytic motifs identified in the PDB search and from literature. A) In RNase H-like, Glu coordinates a solvent excluded Mg_B and stores proton for leaving group protonation. His abstracts the proton and facilitates product release. Example PDBs: 1RTD, 2QKK, 6BSH. B) In RNase T-like enzymes, the Glu is shifted to Mg_A aiding in WAT_{nuc} positioning. Mg_B is coordinated by three water molecules⁵⁷ of which one is indirectly coordinated by an Asp. The leaving group is most likely protonated from one of the water molecules. Example PDBs: 4KAZ, 1W0H, 1J53, 3NH1, 2O4G, 3CG7. C) In RuvC-like enzymes, an additional Lys/Arg is present which we consider to be atypical for a general RNase H but present in *Bh* RNase H (K196). Mg_A is coordinated by a backbone (bb) oxygen.⁵⁶ Example PDBs: 4LD0, 4EP4, 4UN3, 1HJR, 3HVR

Differences in the catalytic mechanism are mainly associated with Mg_B , which is penta-coordinated in the RNase H-type and hexa-coordinated in the RNase T-like, most likely due to different specificities for DNA and

RNA substrates. In contrast in the RNase T-like enzymes Mg_A is penta-coordinated.²⁷ For the first two types, the pK_a of the glutamate is high (pK_a 7.5-9 as calculated with PROPKA 3.1, see Table S1) due to the relatively nonpolar micro environment around the pocket with the protonated oxygen. The high pK_a for the active site Glu supports the role of the glutamate as proton donor (RNase H-like, RuvC like) or in stabilizing the WAT_{nuc} (RNase T-like) as it is significantly more basic than a standard Glu ($pK_a \sim 4.5$). In QM/MM simulations of the mechanism of the ϵ -subunit of DNA pol III it was shown that the Glu in the RNase T-like motif can pick up the second proton from the WAT_{nuc} after cleavage of the scissile bond.²⁷

We show that in the RNase H-like enzymes protonation occurs via the Glu coordinated to Mg_B . However, this pathway is not possible in the canonical DEDDh motif in the RNase T-like motif due to the different position of the Glu (Figure 6). In the RNase T-like enzymes, there is an Asp exclusively coordinated to Mg_B via a water molecule but its pK_a likely is much closer to a standard Asp ($pK_a \sim 3$, see Table S1), making protonation via this residue unlikely. The proton for the protonation of the leaving group thus likely originates from any of the up to three water molecules of the first solvation shell of Mg_B .⁵⁹ For example in the RNase T structure solved by Hsiao et. al.⁵⁹ Mg_B is solvated with 3 water ligands as opposed to zero water ligands in the RNase H-like and in the RuvC-like structures.⁵⁶ This mechanism is also consistent with the lowering of pK_a of Mg associated water molecules.⁶⁰ In the QM/MM simulation of the RNase T-like ϵ -subunit of the pol III the protonation of the leaving O3' remained elusive, but pathways via the scissile phosphate were ruled out based on the calculated energy barriers of >30 kcal mol⁻¹.²⁷ Thus, the protonation via the solvent ligands in the RNase T-like motif is the only viable alternative. One of the waters in the RNase T-like motif is in a similar position to the glutamate in the RNase H-like enzymes.

In the RuvC-like enzymes the DEDD motif is not observed (only DDE) and the last Asp of the DEDD motif is possibly replaced by a His (e.g in *S. pyogenes* (*Spy*) Cas9 RuvC H983). This is supported by the fact that H \rightarrow D mutation in *Tth* RuvC is shown to activate the enzyme,⁵ which would reconstitute the DEDD motif of RNase H. This indicates the strong similarity of the RuvC-type and the RNase H-type. In contrast to this finding, other studies report that the affinity of nitrogen (e.g., in histidine or imino side chains of nucleic acids) for Mg^{2+} is low⁵³ and in the QM/MM ab initio MD of the Cas9 active site it can be seen, that the stable distance of H983 to Mg_A is around 4.2 Å, which is not directly coordinating. We can reproduce the same distance if we align our QM/MM reactant structure to the inactive *Spy* Cas9 (PDB 4UN3) RuvC active site (Figure S6). In the H \rightarrow D mutation for *Tth* RuvC most likely the adjacent H982, which is also conserved, can act as a replacement general base.

Neither H982 nor H983 is placed on a flexible loop as is the histidine in the RNase H-type or the RNase T-type.²⁴ The RuvC-type histidines will likely be able to abstract the proton and act as general base but not as third ion recruiter or facilitator of product release. However, there is a conserved lysine (*Spy* K974, *Tth* K112) and a conserved arginine (*Spy* R976) in close vicinity of the active site. This lysine (2-4 residues in direction of the N-terminus from the arginine) and the arginine are conserved among various Cas9 family members such as from *Staphylococcus Aureus*, *Francisella Novicida*, *Actinomyces naeslundii* and *Campylobacter jejuni* (Figure S7). The lysine is very mobile (no sidechain density for *Tth* RuvC K112 PDB 4LD0) and most likely carries out a similar task such as K196 in *Bh*.¹⁰ In Gaussian Accelerated MD simulations of the activated Cas9 complex it has also been shown that R976 is more stable in a down conformation pointing in the active site.⁵⁶ K196 in *Bh* stabilizes the negatively charged cleaved product until arrival of the third Mg^{2+} which then possibly facilitates the product release.¹⁰ We postulate that one positively charged residue (K or R) is needed to facilitate catalysis and removal of the product in the RuvC-like type.

Combinations of these general motifs are also possible such as in *Bh* RNase H. Here, the leaving group protonation most likely is RNase H-like, but the histidine, which is missing, is partly replaced by the substrate and transient metal ions^{10,13} and partly by the lysine as in the RuvC-like motif.

Conclusion

HIV-1 RNase H catalyzes the cleavage of the DNA/RNA hybrid bound to the HIV-1 RT enzyme dependent on a non-canonical DEDDh motif for coordination of two catalytic Mg^{2+} ions. Our QM/MM calculations, in agreement with prior work, found that the reaction proceeds via two transition states that separate the two protonation events: (i) deprotonation of the nucleophilic water and (ii) protonation of the leaving group. Beyond this, here we identified a new proton transfer mechanism, both for the deprotonation of the nucleophilic water and for the protonation of the leaving group, that is able to account for the multitude of experimental results not only in RNase H but in many other nucleases. We identified a conserved histidine residue, which has a dual role as activator of the nucleophile and recruiter of a third divalent metal ion in the product complex. The histidine is structurally and functionally conserved across several other enzymes with a RNase or DNase function, and homologous mutational and computational studies in nucleases with a different positioning of the same active site residues (i.e the canonical DEDDh motif) support our mechanistic results. We highlight differences and similarities of the canonical DEDDh motif in 3'-5' exonucleases such as the ϵ subunit of polymerase III or RNase T with the non-canonical motif found in various RNases H in the discussion of our results.

We also identified a highly conserved glutamate residue as the most likely proton donor for the leaving group phosphate O3', providing evidence why a glutamate and not a fourth aspartate is present in the DEDD motif in RNase H.

Furthermore, to shed light on the post-reaction product release process, we performed MD simulations to identify the role of active-site metal ions and the conserved histidine residue. We sampled the possible metal ion binding sites of the cleaved product state for transiently bound mono- and divalent cations in multiple, 100 ns-long MD simulations. Our simulations show that a third metal cation is recruited from the bulk finding two potential cation binding sites that are consistent with those identified experimentally by Samara and Yang¹⁰ and Gan et. al.⁴⁶. Our results suggest that the third Mg²⁺ binding site plays a key role in the product release based on the cleaved product stability in our MD simulations, aided also by the conserved histidine, which has not been reported before. Our results confirm that this mechanism is likely more general for other nucleases, based on structural similarity with *Bh* RNase H, HIV-1 RNase H and *Aa*-RNase II crystallographic data.

Through the novel understanding of structural and mechanistic roles of key conserved residues, we are able to generalize the nucleic acid cleavage reaction mechanism to a wide array of biomedically very important nucleases, aiding structural and functional understanding into this ubiquitous biochemical process.

ASSOCIATED CONTENT

Supporting Information including

- Figures S1-5, Geometries from the QM/MM MD, QM/MM distance file,
- Pymol 1.8.6 session file with alignment of MD ion binding sites, PDB 2nug and *Bh* RNase H crystal structures.
- Pymol 1.8.6 session file with alignment of RNaseH-type, RNaseT-type and RuvC-type active sites
- Table S1 with results from PROPKA 3.1 calculations
- Source code of interactive Figure of the free energy profile and the geometries from the QM/MM. Figure is available for direct viewing on <https://dev.simonduerr.eu/hiv/>.

AUTHOR INFORMATION

Corresponding Author

*edina.rosta@kcl.ac.uk

Present Addresses

† Laboratory of computational biochemistry and chemistry, SB ISIC, Ecole polytechnique fédérale de Lausanne, CH-1015 Lausanne, Switzerland

‡ School of Chemistry, University of Bristol, Bristol, BS81TS United Kingdom

¶ Departamento de Química Física, Facultad de Ciencias Químicas, Universidad de Salamanca, 37008 Salamanca, Spain

Funding Sources

E.R acknowledges funding from EPSRC (EP/R013012/1), BBSRC (BB/N007700/1), and ERC (project 757850 BioNet). S.D received scholarships from Erasmus+ and Studienstiftung des deutschen Volkes. P. G. J Acknowledges funding by Fundacion Salamanca City of Culture and Knowledge ("Programme for attracting scientific talent to Salamanca"). E.R and R.S acknowledge funding from the EC Marie Curie IEF.

Notes

ORCID

S.L.D 0000-0002-4304-8106

R.S 0000-0002-1035-9020

P.G.J 0000-0001-8846-3998

E.R 0000-0002-9823-4766

ACKNOWLEDGMENT

Computing time was provided on Rosalind HPC cluster by King's College London and the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>) via HecBioSim (funded by EPSRC grant EP/R029407/1).

ABBREVIATIONS

Hs Homo sapiens, *Ec* Escherichia coli, *Bh* Bacillus halodurans, *Tth* Thermus thermophilus, *Spy* Streptococcus pyogenes, TS transition state, RS reactant state, PA proton abstraction, INT intermediate, PS product state, HBD hybrid binding domain, GAG group specific antigen.

REFERENCES

- (1) Yang, W. *Q. Rev. Biophys.* **2011**, *44* (1), 1–93.
- (2) Steitz, T. A.; Steitz, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6498–6502.
- (3) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. *Cell* **2005**, *121* (7), 1005–1016.
- (4) Jiang, F.; Doudna, J. A. *Annu.Rev.Biophys* **2017**, *46*, 505–529.
- (5) Górecka, K. M.; Komorowska, W.; Nowotny, M. *Nucleic Acids Res.* **2013**, *41* (21), 9945–9955.
- (6) Katz, R.; Skalka, A. M. *Annu. Rev. Biochem.* **1994**, *63* (1), 133–173.
- (7) Tisdale, M.; Schulze, T.; Larder, B. A.; Moelling, K. *J Gen Virol* **1991**, *72* (Pt 1), 59–66.
- (8) Chon, H.; Sparks, J. L.; Rychlik, M.; Nowotny, M.; Burgers, P. M.; Crouch, R. J.; Cerritelli, S. M. *Nucleic Acids Res.* **2013**, *41* (5), 3130–3143.
- (9) Cerritelli, S. M.; Crouch, R. J. *FEBS J.* **2009**, *276* (6), 1494–1505.
- (10) Samara, N. L.; Yang, W. *Nat. Struct. Mol. Biol.* **2018**, *25* (8), 715–721.
- (11) Rosta, E.; Yang, W.; Hummer, G. *J. Am. Chem. Soc.* **2014**, *136* (8), 3137–3144.
- (12) De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Am. Chem. Soc.* **2008**, *130* (33), 10955–10962.
- (13) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. *J. Am. Chem. Soc.* **2011**, *133* (23), 8934–8941.
- (14) Elsässer, B.; Fels, G. *Phys. Chem. Chem. Phys.* **2010**, *12* (36), 11081–11088.
- (15) Yoon, H.; Warshel, A. *Proteins Struct. Funct. Bioinforma.* **2017**, *85* (8), 1446–1453.
- (16) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y.; Yang, W. *Nature* **2012**, *487* (7406), 196–201.
- (17) Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. *Science* **1998**, *282* (5394), 1669–1675.

- (18) Nowotny, M.; Gaidamakov, S. A.; Ghirlando, R.; Cerritelli, S. M.; Crouch, R. J.; Yang, W. *Mol. Cell* **2007**, *28* (2), 264–276.
- (19) Kanaya, S.; Kohara, A.; Miura, Y.; Sekiguchi, A.; Iwai, S.; Inoue, H.; Ohtsuka, E.; Ikehara, M. *J. Biol. Chem.* **1990**, *265* (8), 4615–4621.
- (20) Alla, N. R.; Nicholson, A. W. *FEBS J.* **2012**, *279* (24), 4492–4500.
- (21) Hsiao, Y. Y.; Duh, Y.; Chen, Y. P.; Wang, Y. T.; Yuan, H. S. *Nucleic Acids Res.* **2012**, *40* (16), 8144–8154.
- (22) Cheng, H.-L.; Lin, C.-T.; Huang, K.-W.; Wang, S.; Lin, Y.-T.; Toh, S.-I.; Hsiao, Y.-Y. *Nucleic Acids Res.* **2018**, *46* (22), 12166–12176.
- (23) Cheng, Y.; Patel, D. J. *J. Mol. Biol.* **2004**, *343* (2), 305–312.
- (24) Huang, K.-W.; Hsu, K.-C.; Chu, L.-Y.; Yang, J.-M.; Yuan, H. S.; Hsiao, Y.-Y. *J. Med. Chem.* **2016**, *59* (17), 8019–8029.
- (25) Zuo, Y.; Deutscher, M. P. *Nucleic Acids Res.* **2001**, *29* (5), 1017–1026.
- (26) Hamdan, S.; Carr, P. D.; Brown, S. E.; Ollis, D. L.; Dixon, N. E. *Structure* **2002**, *10* (4), 535–546.
- (27) Cisneros, G. A.; Perera, L.; Schaaper, R. M.; Pedersen, L. C.; London, R. E.; Pedersen, L. G.; Darden, T. A. *J. Am. Chem. Soc.* **2009**, *131* (4), 1550–1556.
- (28) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; Beer, T. A. P. De; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. **2018**, *46* (May), 296–303.
- (29) Tian, L.; Kim, M.-S.; Li, H.; Wang, J.; Yang, W. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (3), 507–512.
- (30) Abbondanzieri, E. A.; Bokinsky, G.; Rausch, J. W.; Zhang, J. X.; Le Grice, S. F. J.; Zhuang, X. *Nature* **2008**, *453*, 184–189.
- (31) Foloppe, N.; Mackerell, A. D. *Nucleic Acids Res.* **2000**, *21* (2), 86–104.
- (32) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Cafisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614.
- (33) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuš, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. D.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Mol. Phys.* **2015**, *113* (2), 184–215.
- (34) Becke, A. D. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (35) Weinan, E.; Ren, W.; Vanden-Eijnden, E. *J. Phys. Chem. B* **2005**, *109* (14), 6688–6693.
- (36) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.
- (37) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Mackerell, A. D. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257–3273.
- (38) Goldschmidt, V.; Didierjean, J.; Ehresmann, B.; Ehresmann, C.; Isel, C.; Marquet, R. *Nucleic Acids Res.* **2006**, *34* (1), 42–52.
- (39) Phillips, J. C.; Braun, R.; Wang, W. E. I.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Poincaré, H. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.
- (40) Uchiyama, Y.; Miura, Y.; Inoue, H.; Ohtsuka, E.; Ueno, Y.; Ikehara, M.; Iwai, S. *J. Mol. Biol.* **1994**, *243* (4), 782–791.
- (41) Schatz, O.; Cromme, F. V.; Grüninger-Leitch, F.; Grice, S. F. J. *Le. FEBS J.* **1989**, *257* (2), 311–314.
- (42) Kashiwagi, T.; Jeanteur, D.; Haruki, M.; Katayanagi, K.; Kanaya, S.; Morikawa, K. *Protein Eng.* **1996**, *9* (10), 857–867.
- (43) Hamki, M.; Tsunaka, Y.; Morikawa, M.; Iwai, S.; Kanaya, S. *Biochemistry* **2000**, *39* (45), 13939–13944.
- (44) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295.
- (45) Kanaya, S.; Oobatake, M.; Liu, Y. *Biochemistry* **1996**, *271* (51), 32729–32736.
- (46) Gan, J.; Shaw, G.; Tropea, J. E.; Waugh, D. S.; Court, D. L.; Ji, X. *Mol. Microbiol.* **2008**, *67* (1), 143–154.
- (47) Kanaya, E.; Kanaya, S. *Eur. J. Biochem.* **1995**, *231* (3), 557–562.
- (48) Lima, W. F.; Wu, H.; Nichols, J. G.; Prakash, T. P.; Ravikumar, V.; Crooke, S. T. *J. Biol. Chem.* **2003**, *278* (50), 49860–49867.
- (49) Shaw-Reid, C. A.; Feuston, B.; Munshi, V.; Getty, K.; Krueger, J.; Hazuda, D. J.; Parniak, M. A.; Miller, M. D.; Lewis, D. *Biochemistry* **2005**, *44* (5), 1595–1606.
- (50) Keck, J. L.; Goedken, E. R.; Marqusee, S. *J. Biol. Chem.* **1998**, *273* (51), 34128–34133.
- (51) Ho, M.-H.; De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Am. Chem. Soc.* **2010**, *132* (39), 13702–13712.
- (52) Nowotny, M.; Yang, W. *EMBO J.* **2006**, *25* (9), 1924–1933.
- (53) Leonarski, F.; D'Ascenzo, L.; Auffinger, P. *Nucleic Acids Res.* **2017**, *45* (2), 987–1004.
- (54) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.
- (55) Majorek, K. A.; Dunin-Horkawicz, S.; Steczkiewicz, K.; Muszewska, A.; Nowotny, M.; Ginalska, K.; Bujnicki, J. M. *Nucleic Acids Res.* **2014**, *42* (7), 4160–4179.
- (56) Palermo, G. *J. Chem. Inf. Model.* **2019**, DOI: 10.1021/acs.jcim.8b00988.
- (57) Hsiao, Y. Y.; Yang, C. C.; Lin, C. L.; Lin, J. L.; Duh, Y.; Yuan, H. S. *Nat Chem Biol* **2011**, *7* (4), 236–243.
- (58) Nowotny, M.; Figiel, M. In *Human Immunodeficiency Virus Reverse Transcriptase: A Bench-to-Bedside Success*; LeGrice, S. F., Goette, M., Eds.; Springer Science, 2013; pp 77–95.
- (59) Hsiao, Y. Y.; Fang, W. H.; Lee, C. C.; Chen, Y. P.; Yuan, H. S. *PLoS Biol.* **2014**, *12* (3).
- (60) Pontius, B. W.; Lott, W. B.; von Hippel, P. H. *Proc. Natl. Acad. Sci. USA* **1997**, *94* (6), 2290–2294.
- (61) Chojnacki, S.; Cowley, A.; Lee, J.; Foix, A.; Lopez, R. *Nucleic Acids Res.* **2017**, *45* (W1), W550–W553.
- (62) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. *Science* **2014**, *343* (6176), 1247997–1247997.
- (63) Nishimasu, H.; Cong, L.; Yan, W. X.; Ran, F. A.; Zetsche, B.; Li, Y.; Kurabayashi, A.; Ishitani, R.; Zhang, F.; Nureki, O. *Cell* **2015**, *162* (5), 1113–1126.
- (64) Hirano, H.; Gootenberg, J. S.; Horii, T.; Abudayyeh, O. O.; Kimura, M.; Hsu, P. D.; Nakane, T.; Ishitani, R.; Hatada, I.; Zhang, F.; Nishimasu, H.; Nureki, O. *Cell* **2016**, *164* (5), 950–961.
- (65) Yamada, M.; Watanabe, Y.; Gootenberg, J. S.; Hirano, H.; Ran, F. A.; Nakane, T.; Ishitani, R.; Zhang, F.; Nishimasu, H.; Nureki, O. *Mol. Cell* **2017**, *65* (6), 1109–1121.e3.

Table of Contents Artwork

