Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

DATA SCIENCE AND COMPUTATION

Ciclo 33

Settore Concorsuale: 03/D1 - CHIMICA E TECNOLOGIE FARMACEUTICHE, TOSSICOLOGICHE E NUTRACEUTICO - ALIMENTARI

Settore Scientifico Disciplinare: CHIM/08 - CHIMICA FARMACEUTICA

MOLECULAR SIMULATIONS AND MODELING OF PHARMACEUTICALLY RELEVANT RNA-PROCESSING ENZYMES

Presentata da: Jacopo Manigrasso

Coordinatore Dottorato

Andrea Cavalli

Supervisore

Marco De Vivo

Co-supervisore

Andrea Cavalli

Esame finale anno 2022

Table of Contents

Table of Figures

Choose Your Future.
Choose Life!

Acknowledgments

During these four years, many people have contributed to my personal and scientific growth. Certainly, I feel the need to deeply thank my mentor, Dr. Marco De Vivo. Thank you for having involved me in challenging, timely and multidisciplinary projects. I am grateful for the time you spent supervising all my work, from helping me staying focused on projects' goals – keeping always an eye on future perspectives – to teaching me how to clearly deliver ideas. At times, it has been tough, but I really enjoyed this experience!

I am also really grateful to Dr. Marco Marcia, for having shared with me his own experience, expertise, as well as mentoring time. It has been great to deliver scientific projects together in such a multidisciplinary environment, taking advantage of different perspectives and ideas.

Nonetheless, I would like to thank prof. Giulia Palermo, for having hosted me in her laboratories, at the University of California - Riverside. Thank you for having given me this opportunity, for your great scientific and professional advice, and your always kind willingness.

And, "finally", thank you all folks! There is no need to mention anyone, specifically. I shared with you all almost every moment of my life here, in such a cramped and dirty *città vecchia* that lives in the dim light, still offering some of the warmest avenues I have ever seen. From simple stairs in front of the sea, to comfortable Bordeaux-red velvet-covered seats, and fresh marble armrest. Sincerely, I feel those moments and their stories deeply close to my heart.

This thesis is dedicated to those moments of dim light, and all of our stories.

Abstract

The two-metal-ion architecture is a structural feature found in a variety of RNA-processing metalloenzymes or ribozymes (RNA-based enzymes), which control the biogenesis and the metabolism of vital RNAs, including non-coding RNAs (ncRNAs). Notably, such ncRNAs are emerging as key players for the regulation of cellular homeostasis, and their altered expression has been often linked to the development of severe human pathologies, from cancer to mental disorders. Accordingly, understanding the biological processing of ncRNAs is foundational for the development of novel therapeutic strategies and tools. Here, we use state-of-the-art molecular simulations, complemented with X-ray crystallography and biochemical experiments, to characterize the RNA processing cycle as catalyzed by two two-metal-ion enzymes: the group II intron ribozymes and the RNase H1. We show that multiple and diverse cations are strategically recruited at and timely released from the enzymes' active site during catalysis. Such a controlled cations' trafficking leads to the recursive formation and disruption of an extended two-metal-ion architecture that is functional for RNA-hydrolysis – from substrate recruitment to product release. Importantly, we found that these cations' binding sites are conserved among other RNA-processing machineries, including the human spliceosome and CRISPR-Cas systems, suggesting that an evolutionarily-converged catalytic strategy is adopted by these enzymes to process RNA molecules. Thus, our findings corroborate and sensibly extend the current knowledge of two-metal-ion enzymes, and support the design of novel drugs targeting RNA-processing metalloenzymes or ribozymes as well as the rational engineering of novel programmable gene-therapy tools.

Chapter 1. Biological Functions and Therapeutic Relevance of RNA Molecules

1.1 The Rise of Non-Coding RNA

According to one of the long-standing principles of cellular biology,[1] the ribonucleic acid (RNA) is a fundamental component of living cells, being essential for the expression of genes in all domains of life. Indeed, RNA comes in many shapes to support the information transfer from genes to proteins in the cells. One crucial RNA molecule for this process is the well-known messenger RNA (mRNA), which is transcribed from the genomic DNA and functions as a ready-to-read frame for the protein translation machinery. Moreover, two additional RNA molecules, the ribosomal and transfer RNAs (rRNA and tRNA, respectively), together contribute to proteins' synthesis. Specifically, the rRNA constitutes the molecular platform for the assembly of the ribosome, the macromolecular structure in which amino acids polymerization occurs.[2] As well, the tRNA supports such polymerization, deciphering the mRNA's information content by recognizing specific nucleotides triplets, and favoring the recruitment of the correct amino acid for the proper elongation of peptides.[2]

In line with their functions, the mRNA, rRNA, and tRNA molecules are categorized into two main classes. While the mRNA is a protein-coding transcript, both rRNA and tRNA are also known as non-coding RNAs, since they do not encode any information for protein sequence.[3] Surprisingly, despite the clear importance of proteins for the regulation of cellular homeostasis, only a small amount of the human genome encodes protein-coding transcripts. Indeed, the vast majority of genes encodes for ncRNAs (~70%).[4] In recent years, several efforts have been made to identify and characterize the types and functions of these ncRNAs, both in humans as well as in fungi, viruses, and bacteria. Importantly, there is nowadays evidence of regulatory ncRNAs that are functionally involved in the

modulation of crucial cellular processes, such as translation, splicing, gene expression, immunity, and metabolic protein regulation, as well as viral and bacterial infections.[3,5] Thus, these pieces of knowledge have extended the importance of RNA from its central role as a DNA-to-protein information-transferring linker to a more sophisticated molecule, which can directly modulate the cellular homeostasis by interacting with a heterogeneous plethora of binding partners – as much as like is observed for proteins. Accordingly, the discovery of ncRNAs represents a crucial breakthrough in life sciences, paving the way for the understanding of previously uncharacterized metabolic pathways and possibly expanding the number of druggable targets for the treatment of severe human diseases, as shown by the growing interest in the development of RNA-targeted small molecule drugs.[6–8]

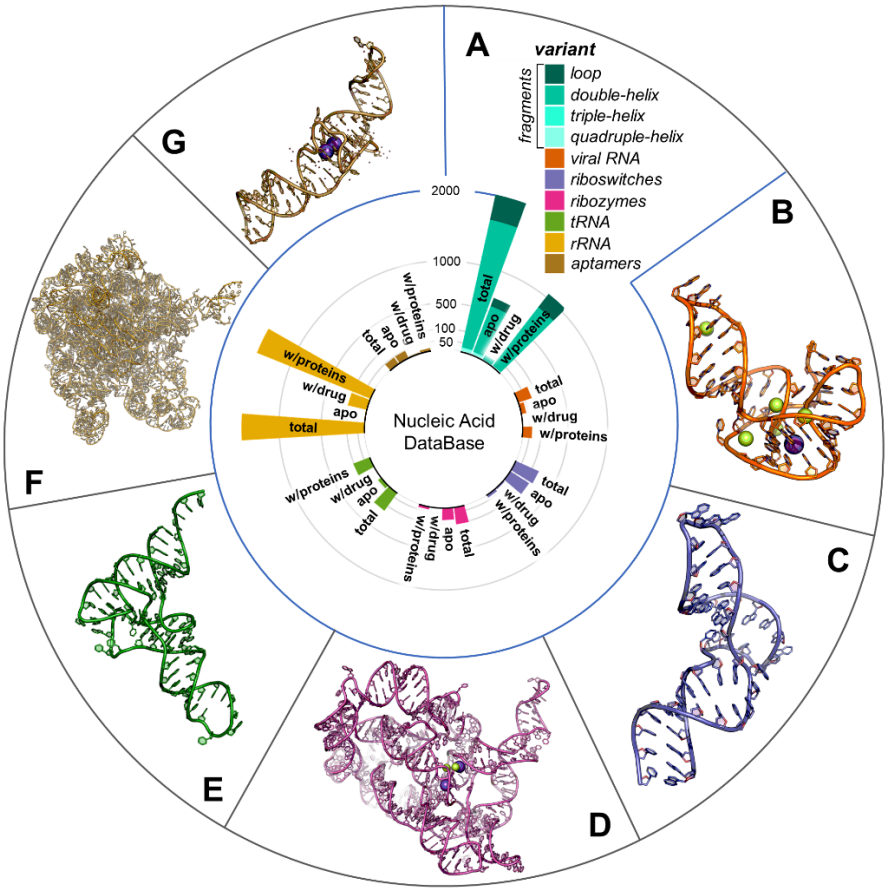


Figure 1.1: Experimentally determined RNA structures in the nucleic acid database. ***(A)*** *The number of structures in the nucleic acid database for apo RNA, RNA-drug, and RNA-protein complexes is reported as a bar plot for RNA fragments (cyan to dark-cyan), viral RNA (orange), riboswitches (purple), ribozymes (pink), tRNA (green), rRNA (yellow), and RNA aptamers (brown).* ***(B–G)*** *Representative structures are shown as cartoons for viral RNA* ***(B)****, riboswitches* ***(C)****, ribozymes* ***(D)****, tRNA* ***(E)****, rRNA* ***(F)****, and RNA aptamers* ***(G)****. Divalent (green), and monovalent (purple) cations are represented as spheres, when present in the original structure.*

One of the most characterized human ncRNAs are the so-called microRNAs (miRNAs). These are transcribed in the cells' nucleus by the RNA Polymerase II as pri-miRNA of ~3000 nucleotides, which are then processed by the RNase III Drosha to form a pre-miRNA of ~80 nucleotides that is translocated into the cytoplasm. Here, the pre-miRNA is cleaved by the RNase III Dicer to synthesize the biologically active miRNA, a double-stranded RNA of ~20 nucleotides.[9] At this point, these short miRNAs can pair a complementary sequence located at the 3' untranslated regions (3'UTRs) of their target mRNA. As a consequence of this binding event, the translation of the target mRNA is suppressed, such that the expression of the mRNA-encoded protein is inhibited. It has been shown that more than 60% of humans mRNAs express miRNA binding site at their 3'UTRs, and that one single miRNA can complementary target multiple mRNAs, suggesting a complex and broad mode of action of such miRNAs.[10] Importantly, an altered expression of tissue-specific miRNA has been related to the development of several human diseases, including many types of cancer and severe mental disorders.[5] As such, miRNAs have been recently recognized as an attractive and promising target for the development of novel therapeutic agents – including small molecule drugs – to treat such disorders. For instance, the small-molecule Targaprimir-515 was recently reported to selectively target the miR-515 hairpin precursor to sensitize HER2-negative cancer cells to Herceptin. Additionally, the biogenesis of the well-recognized tumorigenic miR-21 has been targeted with small molecule drugs, including some neomycin-aminoacid-artificial nucleobase conjugates.

As opposite to short miRNAs, the so-called long-non-coding RNA (lncRNA) are mammalians transcripts longer than 200 nucleotides that do not encode for proteins. Intuitively, because of the weak boundaries defining this class, lncRNAs gather together very different kinds of transcripts, from both functional and structural perspectives. In general, lncRNAs were initially believed to act via promiscuous low specificity mechanisms primarily connected only to their transcription in *cis*.[11] However, in the last five years, it has emerged that lncRNAs possess discrete tertiary structure motifs involved in the establishment of selective interactions with partner

proteins, which can occur in multiple ways.[11] In fact, lncRNAs may be characterized by highly ordered tertiary architectures, as observed for protein systems, but may also adopt less compact conformations, characterized by disordered single-stranded motifs and preserving folded protein binding sites at their structural periphery. However, despite the likely diversity of lncRNAs molecular architectures, these structures can be surprisingly related to those of proteins in terms of main structural features for functionality, as exemplified by the remarkable similarity found between two-metal-ion-centered ribozymes and proteins. These considerations suggest that ncRNA structures are often playing a very important role in regulating critical cellular processes, such as chromatin modification, transcriptional and post-transcriptional regulation.[12] Indeed, the altered expression of lncRNAs has been related to several human pathologies, including several types of cancer as well as mental disorders, like schizophrenia and Alzheimer's disease.[13] Notably, the importance of lncRNAs in the development of such diseases is further underscored by the recent efforts to target these ncRNAs with several drug design approaches, including the development of antisense nucleotides.[14]

Another well-characterized class of ncRNAs is that of bacterial riboswitches. These aptamers are constituted by folded but flexible RNA strands located at the 5′-untranslated region of the bacterial mRNA. Riboswitches act as molecular sensors that react upon recognition and binding of specific endo- or exo-metabolites, such as metal ions or small organic molecules. These binding events induce specific riboswitch conformational changes, which modulate the translation of the downstream mRNA, affecting the response to environmental stress and preserve bacterial cell's homeostasis. Notably, it has been shown that riboswitches control vital metabolic pathways in well-known bacterial pathogens, like *Bacillus Subtilis*, *Escherichia Coli*, and *Staphylococcus Aureus*, to mention some of them.[15,16] Thus, it is not surprising that – in the era of bacterial multidrug resistance – several research efforts have been focused on the design and development of riboswitches' inhibitors able to promote the dysregulation of bacterial metabolic pathways, ultimately functioning as antibiotics. By way of example, synthetic mimics of the

flavin nucleotide have been designed to block the activity of the target riboflavin riboswitches, resulting in the inhibition of growth of several bacterial cells, including *C. Difficile* and *E. Coli*.[17,18]

1.2 The ribozymes: catalytic RNA molecules

A particular class of ncRNA is represented by the so-called ribozymes, which are RNA molecules that directly catalyze the transesterification of 3',5'-phosphodiester bonds. These RNAs were characterized for the first time in the early '80s, and for their discovery, the Noble Price in Chemistry was awarded in 1989 to the two ribozymes' pioneers Tom Cech and Sid Altman. In 1981, the first one observed that one RNA sequence contained in the mRNA of the protozoan *Tetrahymena thermophila* could excise itself from the mRNA.[19] In 1983, Sid Altman discovered that the protein component of the protein-RNA complex RNase P was not essential for RNA cleavage.[20] Since then, several kinds of ribozymes have been found in the cellular transcripts of all domains of life, supporting the importance of the role they fulfill in living cells, as discussed below.

Catalytic RNAs can be classified according to several principles. The most convenient categorization for the scope of this thesis can be based upon ribozymes' catalyzed reaction and comprises two main classes: self-cleaving and self-splicing ribozymes. To the first class belong relatively small ribozymes whose catalytic cycle is limited to a single step nucleolytic reaction. In general, this follows an S_N2-like mechanism, in which the 2'-OH nucleophile attacks the adjacent 3'-phosphate to cleave the 3',5'-phosphodiester bond, resulting in the formation of 2',3'-cyclic phosphate and the release of 5'-OH RNA products. This class includes Hammerhead, Hairpin, Varkud satellite, HDV ribozymes, Twister, Pistol, and the glmS ribozymes. Certainly, several sub-classifications can be made according to particular reaction mechanisms adopted by each catalytic RNA. For example, the Hammerhead ribozyme has been proposed to follow a general acid-base catalytic mechanism, in which the nucleophile $2'OH^{C17}$ is activated by $N2^{G12}$, which

functions as a general base. Upon activation, the nucleophile can attack 3',5'-phosphodiester bond between C17 and C1, which is depolarized by the 2'OH^{G8}, acting as a general acid. This mechanism has been validated via both experimental and computational investigations.[21,22] These last, suggested that one Mg^{2+} ion is transiently recruited to lower down the pk_a of 2'OH^{G8}, which otherwise would be too high to depolarize the 3',5'-phosphodiester bond.[23] Several other catalytic mechanisms are adopted by other small self-cleaving ribozymes, and a detailed discussion of them is out of the scope of this thesis. In general, the wide variety of precise catalytic strategies adopted by relatively short RNA sequences, and their abundance in living organisms, is representative of the importance that RNA might have played in the early stages of life's evolution.

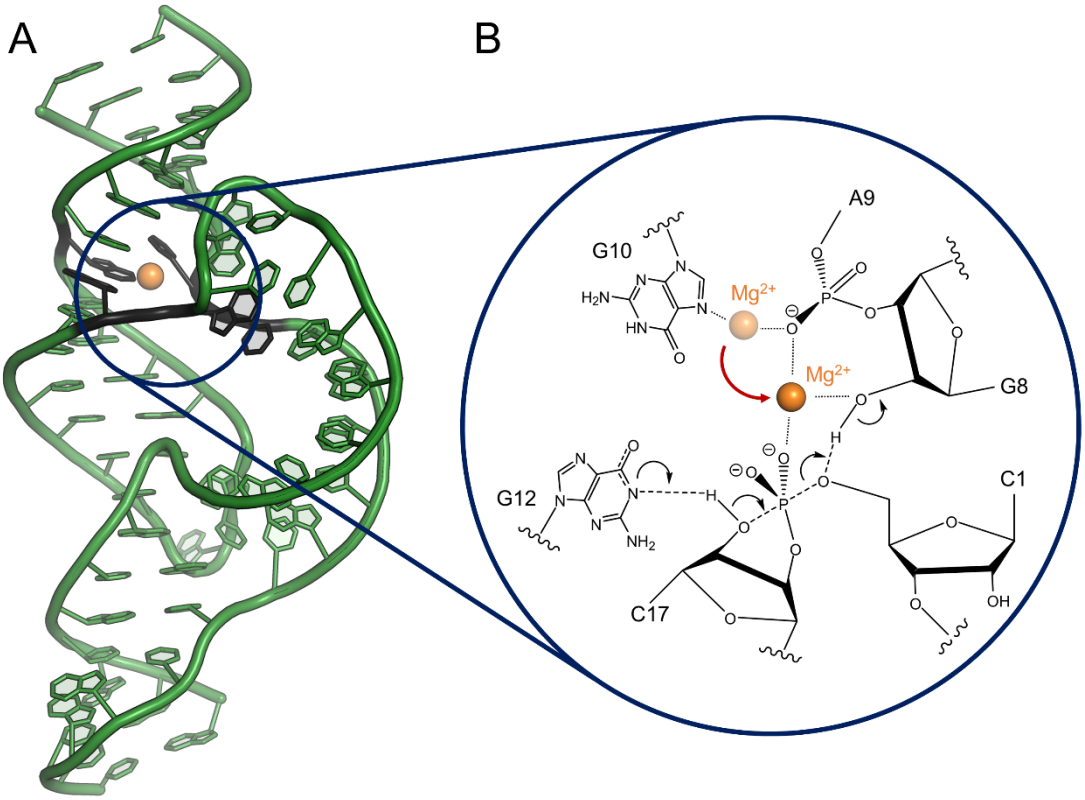


Figure 1.2: The catalytic mechanism of the Hammerhead ribozyme follows a metal-ion-aided acid-base strategy. **(A)** The *three-dimensional structure of the hammerhead ribozyme (PDBid: 3ZP8) is represented as a green cartoon, while the active site residues are highlighted in black, in the blue circle.* ***(B)*** *The catalytic mechanism is reported. Notably, one divalent metal ion is suggested to favor the activation of the reactants for catalysis.*

As opposite to one-step catalysis adopted by self-cleaving RNAs, self-splicing ribozymes follow a two-step catalytic mechanism in which two consecutive S_N2-like transesterification reactions are performed. Among the most studied self-splicing ribozymes are the group II introns, which catalyze their self-excision from the host RNA sequence and the subsequent re-joining of the resulting two pieces of RNA (exons).[24,25] These catalytic RNAs are found in many plants, fungi, and bacteria, where they crucially regulate the expression of genes, being thus vital for those organisms.[26,27] For this reason, they represent an attractive target for the development of novel drugs as antimicrobial agents.[28] Moreover, since the catalytic cycle of group II introns is reversible,[29,30] these ribozymes can also catalyze reverse splicing, a foundational reaction for the biological process known as intron invasion.[31] During this, the intron recognizes a specific host RNA (or DNA) sequence and inserts itself into it, ultimately spreading across and altering the genome of the host.[32] Owing to this functional property, group II introns are also considered promising biotechnological tools that can be engineered and programmed to insert themselves at specific sites of the host genome and alter the expression of pathogenic genes,[33] with promising applications for gene therapy, as like as CRISPR-Cas systems.

The biological and catalytic activity of group II introns is strictly linked to their structural properties, which have been extensively characterized for the family of group IIC introns (hereafter referred to as introns). These introns are long RNA sequences, made up of more than 400 nucleotides that are folded in a highly ordered 3-dimensional architecture, in which six different domains interact with each other to form the catalytically competent conformation.[34–37] Domain 1 (D1) is the largest domain, folds in a shell-like structure, and functions as a central hub for the assembly of the other five domains.[34] Notably, several long-range contacts guarantee the proper and stable folding of the D1. Among them, the so-called z-anchor is established between the helices I(ii) and IC and support the reciprocal positioning of the stems.[38] Another crucial structural motif of the D1 is the five-way junctions, comprised of the T-loop substructure (U31, G32, A33, G34, A245, and

A35; numbering as in PDBid: 4FAQ), which helps the correct orientation of the longest stems of D1 and set the stage for the overall folding of this domain.[38] Nonetheless, the D1 comprises also the Exon Binding Site 1 (EBS1) sequence, which is responsible for the recognition of and pairing with the Intron Binding Sequence 1 (IBS1) presented by the complementary exon.[34] The formation of such interaction is crucial for the intron to recognize its excision/insertion site.[39] Accordingly, in more complex and highly-evolved introns, more than one EBS is found at the D1 to enhance the selectivity of the exon sequence recognition.[40]

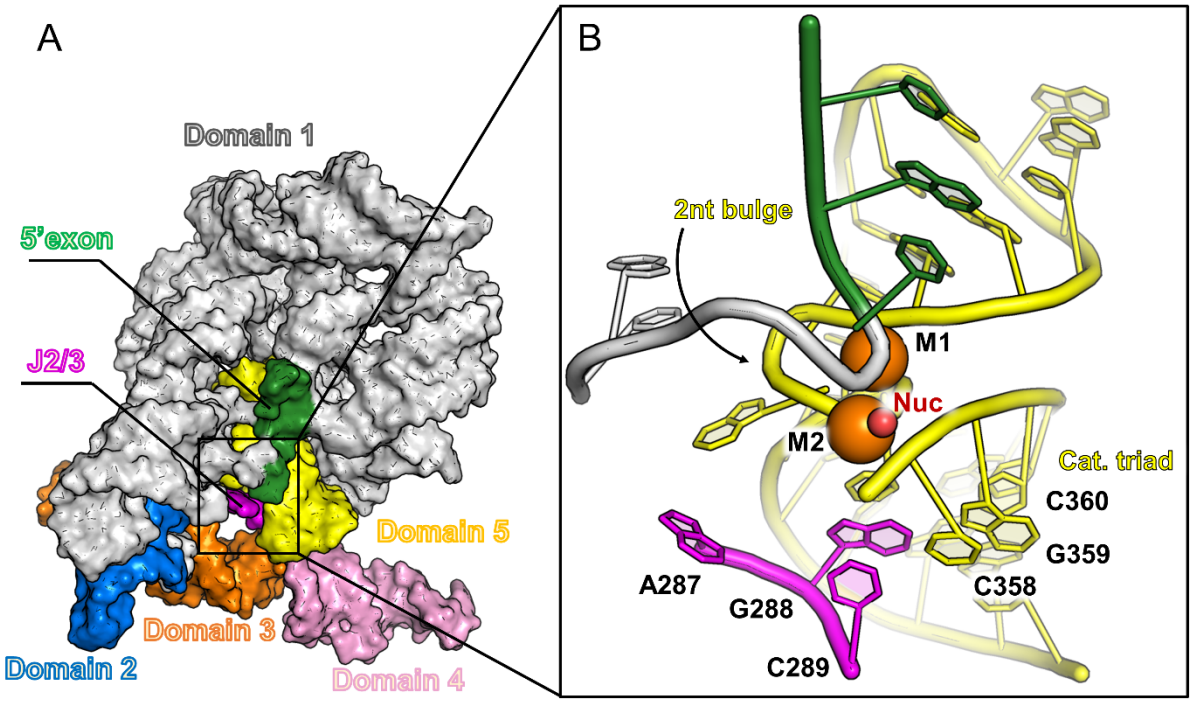


Figure 1.3: The highly ordered 3D architecture of group II intron set the stage for catalysis. ***(A)*** *The 3D structure of the Oceanobacillus Hiyeiensis group II intron is reported as surface (PDBid: 4FAQ). The Domain 1 (grey), Domain 2 (blue), Domain 3 (orange), Domain 4 (pink), Domain 5 (yellow), together with the J2/3 junction (purple), and the 5' exon (green) are highlighted.* ***(B)*** *The active site of the group II intron, including the catalytic triad and the two-nucleotide bulge (yellow), as well as the J2/3 (purple) and the 5' exon (green) are shown as cartoons. The two catalytic M1-M2 ions (orange) and the nucleophilic water (red) are represented as spheres.*

Domain 2 (D2) and domain 3 (D3) are accommodated at the basis of the D1, with which they interact through a series of minor groove triples and ribose zippers.[38,41] The D2 is fundamental to orient the domain 6 (D6), which brings the branch-point Adenosine, the nucleophile of the first splicing step.[42] On the other hand, the D3

functions as a catalytic effector.[43] Most importantly, a series of three nucleotides connecting the D2 and D3 form the so-called J2/3 junction (A287, G288, C289), which projects into the core of the intron.[35] Here, the J2/3 forms the catalytic triple-helix site together with the nucleotides of Domain 5 (D5). Notably, the nucleotides of the J2/3 are among the most conserved of the intron, together with those of the so-called catalytic triad (C358, G359, C360) found at the D5.[35] Such a highly-structured active site is docked into the D1, where it is stabilized via the D1-D5 tertiary contacts κ–κ', λ–λ', and ζ–ζ', and by the presence of the D4, which sits at the bottom of the D5 ensuring its docking.[34] Importantly, the positioning of the D5 supports its structural folding, which is crucial for catalysis. Indeed, D5 conformation is characterized by a stem and a terminal loop that are separated by one flexible two-nucleotide bulge, which is fundamental to recruit two catalytic Mg^{2+} ions that are ultimately responsible for the alignment and activation of the reactants during the splicing.[37]

Group II introns' self-splicing proceeds with two consecutive transesterification reactions catalyzed via the so-called two-metal-ions mechanism. In the first splicing step, the 3',5'-phosphate bond bridging the intron and the 5'-exon (5'-splice site) is aligned at the catalytic site, where is directly coordinated by both the Mg^{2+} ions (M1-M2), sitting at ~4Å from each other.[35] Specifically, the M1 also coordinates the 3'-oxygen of the 5'-exon, and contributes to the depolarization of the phosphodiester bond. Instead, the M2 coordinates the nucleophile, promoting its deprotonation to favor the nucleophilic attack at the scissile phosphate.[44] At this point, the transesterification reaction occurs via an S_N2-like mechanism, proceeding with the formation of a pentacovalent, planar transition state, and resulting in the release of the free 5'-exon.[44] The second step of splicing occurs after the crucial structural rearrangement of the intron,[35] which leads to the alignment of the 3'-splice site at the active site (see *Chapter 3*). Upon alignment, the same S_N2-like transesterification reaction occurs with inverted chemistry. That is, the metal M1 now activates the nucleophile 3'-oxygen of the 5'-exon, while the metal M2 depolarizes the scissile phosphate at the 3'-splice site. Notably, group II introns

splicing can follow two different pathways: the hydrolytic or the branching path. At the end of the splicing, the reaction products will depend on the nucleophile of the first reaction step. Indeed, if the first-step nucleophile is a water molecule, the hydrolytic pathway is followed and, thus, the spliced 3',5'-exons and the free linear intron are released.[45,46] On the other hand, if the first-step nucleophile is the branching Adenosine (located at the D6), the splicing proceeds via the branching pathway, resulting in the release of the spliced 3',5'-exons, and the free lariat intron.[47,48]

Remarkably, the splicing cycle catalyzed by the group II intron has several and profound similarities to that operated by the human spliceosome, the macromolecular machinery that is responsible for the maturation of protein-coding mRNA in human cells. Indeed, both these molecular systems share the active site architecture, in which a two-metal-ion-based set of transesterification reactions is directly catalyzed by RNA molecules.[25,38] This evidence has been supported at first by several chemical biology experiments, and, most recently, by the release of X-ray and Cryo-EM models of both the group II introns and the spliceosome,[25] respectively. Thus, because of the chemical, structural and functional similarities, the group II introns are considered to be evolutionary ancestors of the human spliceosome. Accordingly, group II introns represent a great molecular system to model the spliceosome functional properties. Indeed, the characterization of the principles that regulate group II intron catalysis can be of crucial importance to decipher the biological mechanism underlying the spliceosome activity,[49] which is essential for protein synthesis (see *Chapter 5*). Moreover, altered splicing functions have been related to the development of several human diseases, including Duchenne[50] and Becker[51] muscular dystrophy as well as the spinal muscular atrophy (SMA),[52,53] among the others.[54] As a matter of fact, the human spliceosome is the target of several FDA-approved Antisense Nucleotide-based therapies,[55,56] and the first small molecule drug targeting the human spliceosome has been approved as well for the treatment of SMA, recently.[57,58] Thus, structural and functional insights of the group II intron and the human spliceosome can ultimately support the

development of novel therapeutic approaches to target altered splicing functions observed in severe human diseases.

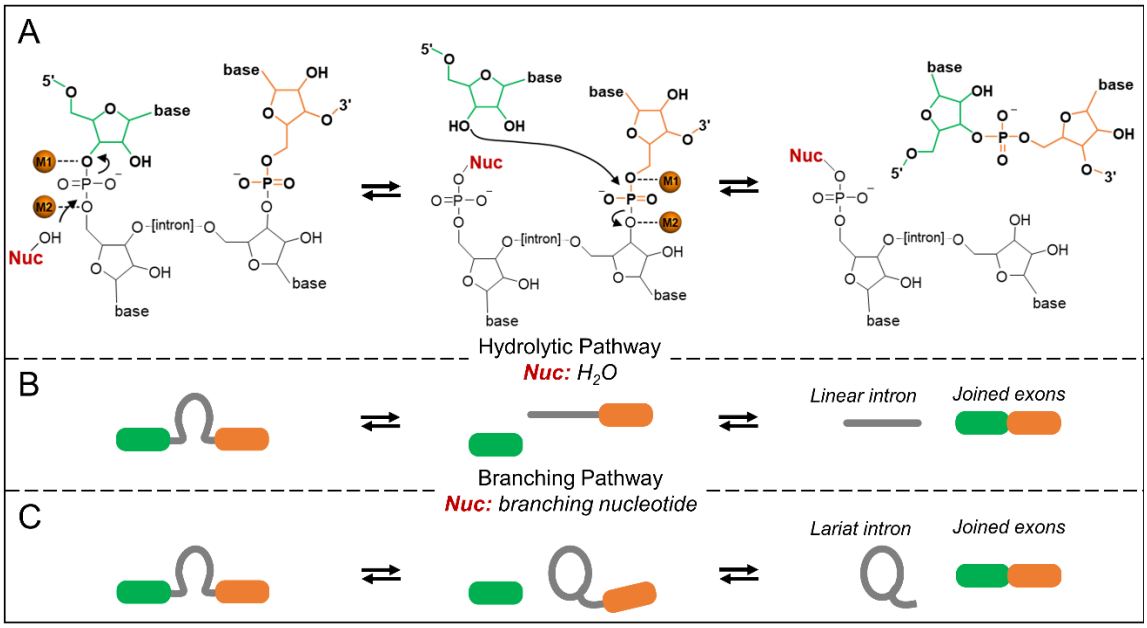


Figure 1.4: The splicing cycle catalyzed by group II introns. ***(A)*** *Two reversible transesterification reaction are catalyzed at the active site of the intron to join the 5'- and the 3'-exon. Notably, a structural rearrangement is needed to release the first-step reaction products and align the second-step reactants. Most importantly, the final splicing products depend on the nature of the nucleophile. Indeed, if the nucleophile is a water molecule* ***(B)****, the splicing finish with the release of a linear intron. On the other hand, if the nucleophile is the branching Adenosine* ***(C)****, the splicing follows the branching mechanism, and a lariat intron is released.*

1.3 The enzymatic processing of RNA molecules

As discussed in the previous paragraphs, RNA is an essential component of living cells. Thus, several enzymes have structurally and functionally evolved to carefully control RNA biogenesis and metabolism. Two main chemical reactions are at the basis of these essential biochemical processes: the polymerization and the hydrolysis of RNA strands. The first one is catalyzed, e.g., by the so-called DNA-dependent RNA polymerases, which are thus responsible for the synthesis of both coding and non-coding RNA transcripts.[59] On the other hand, RNA nucleases (RNases) catalyze the hydrolysis of RNA molecules, and are thus involved in a broad class of cellular processes, from DNA replication to RNA processing, silencing, and interference. Specifically, the Ribonuclease H (RNase H) represents a prototypical example of an RNA-processing enzyme. RNase H belongs to the

retroviral integrase superfamily, and hydrolyzes RNA strands when found in RNA:DNA hybrids.[60] For this reason, it plays a crucial role in removing the RNA primers used to start the synthesis of the Okazaki fragments during DNA replication.[61] RNase H has been the first enzyme of its class for which the 3-dimensional structure has been determined.[62,63] RNase H most conserved structural core is characterized by five stranded β-sheets and a variable series of α-helices that are arranged to form the cleft in which the RNA:DNA hybrid is accommodated. Here, four invariant residues form the so-called DEDD motif, which directly coordinates two catalytic Mg^{2+} ions. The catalytic mechanism adopted by RNase H to perform RNA hydrolysis is similar to that discussed for group II introns, and it has been extensively characterized by means of both biochemical, structural, and computational evaluations. Indeed, high-level-of-theory quantum mechanics/molecular mechanics (QM/MM) calculations have shown that the catalytic strategy adopted by the RNase H follows an associative two-metal-aided mechanism, in which the two catalytic Mg^{2+} ions support the stabilization of both the transition state and the leaving group.[64] Notably, similar calculations have also shown that the presence of two Mg^{2+} ions is a strict catalytic requirement, as the replacement of these metals with either Ca^{2+} or Na^{+} ions impairs the RNase H catalysis.[65] Additionally, equilibrium molecular dynamics simulations have been employed to characterize the effect of Mg^{2+} concentration on RNase H catalytic mechanism, showing for the first time that additional metal ions can be transiently recruited at the enzyme's active site during catalysis.[66] Taken together, these studies have contributed to the characterization of the catalytic cycle operated by RNase H, and have underscored the functional importance of the two-metal-ion-based mechanism for the general processing of RNA molecules.

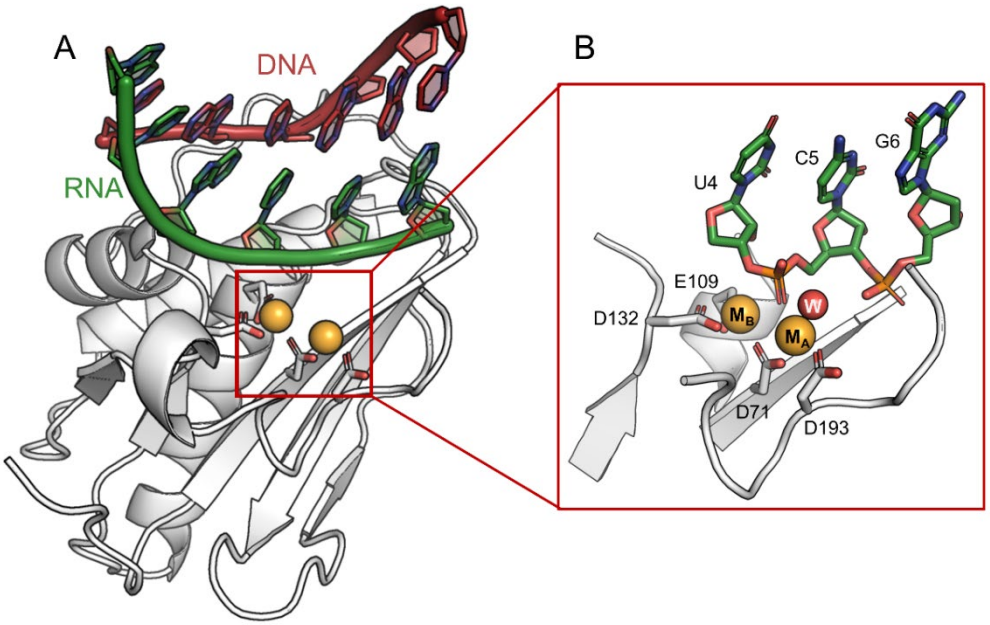


Figure 1.5: The two-metal-ion architecture is a structural feature found in several nucleic acids' processing metalloenzymes. ***(A)*** *The structural of the prototypical metalloenzyme RNase H (white) in complex with an RNA:DNA hybrid (green:red) is highlighted as cartoon.* ***(B)*** *The catalytic DEDD motif (white) and the reaction RNA substrate (green) are reported as sticks, while the catalytic metals M1-M2 (orange) and the nucleophilic water (red) are highlighted as spheres.*

1.4 Extended two-metal-ion architectures for RNA processing

The two-metal-ion mechanism used by both Group II introns and RNase to process RNA molecules was proposed in 1993 by Steitz and Steitz building upon the X-ray structures of alkaline phosphatase and the exonuclease domain of the DNA polymerase-I.[67] These structures captured the active site of both the enzymes, where two divalent metals (M1-M2), sitting at ~4Å from each other, directly coordinate the reaction substrates. Accordingly, Steitz and Steitz proposed that both the ions are involved in the alignment of the substrate while, specifically, M1 is responsible for the activation of the nucleophile and M2 promotes the stabilization of both the transition state and the reaction products.[67] Outstandingly, this mechanism has been confirmed by several other structural and biochemical evidence obtained for a wide variety of nucleic acid processing enzymes, like DNA and RNA polymerases,[59] exo-, and endo-nucleases,[68,69] and also topoisomerases.[70] Indeed, these enzymes share an extraordinarily conserved catalytic core, in which two structured divalent metal ions are precisely positioned to directly participate in

catalysis.[69] As a matter of fact, the importance of this evolutionarily conserved two-metal-ion cluster is further shown by the number of drugs targeting this structural architecture to modulate the activity of metalloenzymes involved in the development of severe human diseases, such as cancer or viral infections.[71]

However, recent structural, biophysical, and computational investigations have shown that an extended two-metal-ion architecture may be required to finely tune the enzymatic polymerization or hydrolysis of both RNA and DNA molecules.[59] For instance, recent X-ray structures have shown that two second-shell monovalent metal ions (K1-K2) are found at the active site of group II intron ribozymes, where they form a heteronuclear metal cluster together with M1-M2.[72] Most importantly, K1-K2 ions are absent in the *in-crystallo* intermediate obtained in between the two steps of intron's splicing (so-called toggled state), when the intron has to release the first step products and recruit the second step reactants.[72] Thus, K1-K2 ions may be transiently released from the active site to favor splicing progression, suggesting a functional role of K1 and K2 dynamics for catalysis.[72] In support of this hypothesis, K1- and K2-like elements (i.e., second-shell basic residues) are also found in other RNA/DNA processing enzymes, such as BamHI, Polymerase-η (Pol-η), Exo-λ, and RuvC.[73] Indeed, structural analyses of these enzymes have shown that second-shell basic residues are often positioned in similar locations as of K1 and K2 in group II introns.[73] For instance, the Lys61 and Lys126 were found to occupy K1- and K2-like sites in BamH1, respectively. Similar results were obtained analyzing the structures of Pol-η, in which the Lys231 and Lys224 located likewise intron's K1 and K2 ions, respectively.[73] As well, in Exo-λ, K1- and K2-like elements corresponded to the second-shell basic residues Lys121 and Arg28, respectively.[73] The importance of the dynamics of these structurally-conserved and strategically-positioned K1- and K2-like elements have been also characterized via computational investigations. Specifically, using equilibrium MD and enhanced sampling simulations, it was shown that the altered dynamics of Pol-η K1-like second-shell basic induce the distortion of reaction substrates, disrupting the Michaelis−Menten complex, and, ultimately, impairing the catalysis.[73] Similar

positively charged second-shell residues have been recently shown, via MD simulations, to play an active role for catalysis in other nucleic-acid processing metalloenzymes, as in the case of λ-exonuclease, dUTPases, and human exonuclease 1 enzymes.[74–76]

Remarkably, a transiently formed heteronuclear metal cluster has been recently captured at the active site of RNase H. Indeed, *in-crystallo* RNase H catalytic intermediates have revealed that multiple K^+ and Mg^{2+} ions locate at the active site during different stages of catalysis, suggesting that their controlled trafficking may be functional for RNA hydrolysis (see *Chapter 4*).[72] Additional metal ions have been also found in the proximity of the reaction center of several DNA polymerases and exonucleases.[59] Such positive ions at the metal-aided catalytic site during the processing of nucleic acids were suggested to contribute actively to the overall catalytic process, although their exact functional and dynamic role remains only partially understood.[77–79] Several biophysical and computational studies have focused on the characterization of the role of such additional metals for the enzymatic processing of nucleic acids. For example, multi-microsecond MD simulations have revealed that a third Mg^{2+} ion is involved in the release of reaction products from the active site of Pol-η.[80] A similar role has been proposed for a divalent metal during the catalytic cycle of Exo-1. Indeed, extensive equilibrium and non-equilibrium MD simulations have shown that, upon substrate hydrolysis, a divalent metal (MgC) is recruited at the vicinity of the Exo-1 active site by the structurally-conserved second-shell residue Glu89.[76] Most importantly, the recruitment of the MgC leads to destabilization of the reaction products and triggers their release, as a consequence of the direct interaction between the additional metal and the 5'-monophosphate nucleotide leaving group.[76]

In general, biophysical and computational data have shown that additional metal ions can be transiently recruited at the reactive center of RNA and DNA processing enzymes to form extended two-metal-ion architectures. These are remarkably conserved among functionally diverse enzymes, which might have thus structurally evolved to follow a similar catalytic strategy.

1.5 Aim of this research project

The two-metal-ion mechanism for the processing of RNA molecules has been proposed in 1993,[67] and, in the following years, it has been confirmed to be a general catalytic strategy adopted by several and diverse metalloenzymes involved in both DNA and RNA biogenesis and metabolism.[59,68] However, recent structural, biophysical and computational evidence support the hypothesis of an extended two-metal-ion architecture for the processing of nucleic acids. That is, heteronuclear metal-ion clusters, including but not limited to the two catalytic divalent metals M1-M2, have been captured at the active site of both metalloenzymes and RNA-based enzymes.[35,72] This suggests that they might have convergently evolved to guarantee the functional processing of nucleic acids. In this context, here we focused on the characterization of the metal-aided catalytic cycle of RNA-processing ribozymes and metalloenzymes.

First, we complemented state-of-the-art MD simulations and free energy calculations with X-ray crystallography and splicing assays to show that the dynamic formation and timely disruption of the catalytic heteronuclear metal-ion cluster are fundamental to guarantee the progression of the splicing cycle of group II intron ribozymes.[81] In brief, we revealed that the protonation of one evolutionary-conserved catalytic nucleotide (C358) induces the release of a structured K^+ ion from the reactive center of the ribozyme. This unbinding event, timely controlled upon the first splicing step, ensures the structural rearrangement of the intron to favor the release of the first step products, and the recruitment of second step reactants. Remarkably, our results were supported with and corroborated by four new X-ray structures and splicing assays of three intron constructs bringing a mutation at position 358. These experiments confirmed that the splicing is inhibited when C358 is mutated in non-protonable nucleotides. Building upon these results, we propose a general mechanism for splicing reactions shared between evolutionarily related splicing machineries, such as group II introns and the human spliceosome.

Second, we used equilibrium and non-equilibrium MD simulations to characterize the dynamic events that control the trafficking of additional metal ions at the two-Mg^{2+} active site of RNase H.[82] We show that, prior RNA hydrolysis, one K^+ is recruited by the structurally-conserved Glu188 to favor substrates' alignment. Notably, upon catalysis, this heteronuclear metals' cluster is disrupted, ultimately leading to the spontaneous binding of a third Mg^{2+} in the proximity of the leaving group. This binding event induces the release of the reaction products, improving the reaction turnover rate. Notably, we also show that the controlled trafficking of cation at the reactive center of RNase H is affected by cations' concentration in the buffer. Our findings are in agreement with the recently published biophysical data on RNase H catalysis, and suggest that the controlled trafficking of metals at the RNase H active site is designed and controlled to aid catalysis. The structural analysis of several other nucleic acids' processing enzymes suggests our findings may be extended to a variety of metalloenzymes, including CRISPR-Cas systems.

Last, we comment on how computational investigations, opportunely coupled with experimental data, can be used to deliver predictive mechanistic and structural hypotheses for large macromolecular complexes.[49] Outstandingly, we show that our prediction on the catalytic cycle of the human spliceosome, built upon computational and experimental insights from introns' splicing, has been proven to be reliable and effective. Indeed, the recent release of two Cryo-EM models of the splicesomal machineries confirmed that the controlled formation and disruption of the catalytic heteronuclear metal cluster is functional for spliceosome catalysis, in line with our hypothesis. Remarkably, this outcome underscores the reliability of current state-of-the-art computational methods, which can be complemented and supported by experimental results to push forward the advances of life sciences, from the engineering of programmable enzymes for gene therapy to structures refinement and drug design.

Chapter 2. Theory and Methods

2.1 Molecular Dynamics Simulations

Molecular Dynamics (MD) is an approach used in computational science to investigate the evolution of the structural and energetic properties of a given molecular system over time. The molecules are modeled according to molecular mechanics (MM), following the principles of the Born-Oppenheimer approximation.[83] Thus, each atom is represented as a rigid bead (or particle), neglecting the contribution of electrons. The connections between the particles (i.e., the chemical bonds) are treated as springs, oscillating around equilibrium values. The physicochemical properties of the particles (e.g., atom hybridization, radius, charge) and their bonds (e.g. lengths, angles, and torsionals) are empirically pre-determined via either experiments or high-level-of-theory computations. These parameters are collected into a potential energy function, the so-called force field (e.g., equation 2.1), which also includes non-bonded terms accounting for attractive and repulsive Van der Waals and Coulomb interactions. Thus, given a particular configuration of interacting particles, the potential energy of the system is directly determined by solving the force field equation.

$$
\begin{aligned}
V(r^N) = & \sum_{i \in bonds} k_{bi}(l_i - l_i^0)^2 + \sum_{i \in angles} k_{ai}(\theta_i - \theta_i^0)^2 \\
& + \sum_{i \in torsions} \sum_{n} \frac{1}{2} V_i^n [1 \\
& + \cos(n\omega_i - n\gamma_i)] \\
& + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \varepsilon_{ij} \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}
\end{aligned}
\tag{2.1}
$$

Building upon these principles, MD simulations use classical Newtonian mechanics to generate a series of time-dependent configurations of the system. A general MD workflow can be summarized with the following scheme. *First*, at time zero (t=t0), the spatial coordinates, the velocities, and the momenta of the particles included in the system are given as input. Importantly, in case MD is used to characterize a biological macromolecule, its atomic coordinates are obtained via either experimental determination (NMR, X-ray, CryoEM) or by computational modeling. Both velocities and momenta are randomly assigned to each atom according to the Maxwell distribution. *Second*, the force field equation is solved, and the force acting on each particle is determined as the negative gradient of the potential energy.

$$F = -\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \tag{2.2}$$

Third, the acceleration is computed from the forces according to the Newtonian equation of motion. *Fourth*, the position, velocity, and momentum of each particle are updated. This set of operations is repeated iteratively at every time step (dt), until the desired simulation time is reached. At the end of the cycle, MD simulations generate a time-dependent series of molecular conformations, which are collectively called a trajectory.

One of the main advantages of using MD for the characterization of molecular properties is that, according to the ergodic hypothesis, if the simulation time is sufficiently long to sample the entire space of configurations, the time average of one property as computed along the trajectory can be considered equal to its ensemble average. In other words, the property's values computed from the simulation corresponds to those experimentally measured under the macroscopic wet-lab conditions. For this reason, MD simulations are often used to visualize and characterize complex biophysical phenomena, like protein folding,[84] large structural

and conformational changes of macromolecular complexes,[85] and ligand binding-unbinding events.[86]

Importantly, two main aspects have to be taken into account when setting up and performing MD simulations: *(i)* the force field reliability and *(ii)* the computational costs. As discussed above, the force field is the set of equations that regulates the particles' interactions during the simulations. Accordingly, the more reliable is the force field, the more accurate will be the results of the simulations. Nowadays, proteins' force field development and parametrization have converged towards high-standard results, owing to years of research focused on the understanding and characterization of proteins' dynamics.[87] On the other hand, the refinement of RNA's force field is still currently ongoing, and significant improvements have been obtained in the last few years, as discussed in *Paragraph 2.2*.

Another important issue is the convergence of the simulations. Indeed, complex structural transition usually occurs in milliseconds timescales,[88–91] while routine equilibrium MD simulations are often limited to microseconds timescale – in spite of recent technological advances.[92–95] This implies that the simulations may not entirely cover the conformational space, such that the ergodic hypothesis cannot be satisfied. In other words, the accessibility of rarely populated states, and thus the sampling of the so-called rare events, could be limited when performing equilibrium molecular dynamics. For this reason, several enhanced sampling MD approaches have been developed to overcome this issue. These methods build upon one main idea: an external source of potential energy is coupled to the equilibrium MD in order to bias the sampling of the system's conformational space using reasonable computational resources. One of these methods is known as Metadynamics, and is discussed in the *Paragraph 2.3*.

2.2 RNA Force Fields: Recent Improvements and Current Limitations

Molecular-mechanics force fields approximate the interaction between particles. Depending on the model's resolution, force fields can be categorized into two main classes: all-atom and coarse-grained. Intuitively, according to all-atom representations, each atom of the system is defined as a model particle. On the other hand, coarse-grained modeling allows reducing the cost of simulating a big system by grouping certain sets of atoms into so-called beads, each of them representing a given structural feature (e.g., RNA sugar, base, and backbone). Coarse-grained models can effectively reproduce the physicochemical properties of even large RNA molecules.[96] However, neglect of the atomistic details can lead to significant disadvantages in understanding RNA structure/function relationship, and thus, in this thesis, we will focus on all-atom MD simulations for RNA-targeted drug discovery.

Several recent efforts have improved the accuracy of all-atom classical RNA force fields. One improvement of the AMBER99 force field is the refinement of the RNA backbone torsion angles. Backbone torsion angles control the flexibility of RNA molecules and are more common in RNAs than in proteins (six vs. two angles per residue). The accuracy of MD simulations of both RNA and DNA double helices was thus improved by reparametrizing the α/γ torsion terms of the AMBER-parmbsc0 force field based on quantum-mechanics/molecular-mechanics (QM/MM) calculations.[97] Similarly, the RNA glycosidic χ torsion angle was reparametrized in the AMBERχ force field series,[98,99] which corrected the formation of high-anti ladder-like RNA structures.[99] Beyond the AMBER force fields, the CHARMM community developed a set of novel RNA parameters to improve the reliability of their MD models.[100] One interesting CHARMM advance is the development of the first polarizable force field for nucleic acids, based on the classical Drude oscillator model.[101] This Drude-2017 force field accounts for charge

transfer between charged groups, such as those occurring in RNA-metal interactions, essential for RNA tertiary folding and stabilization.

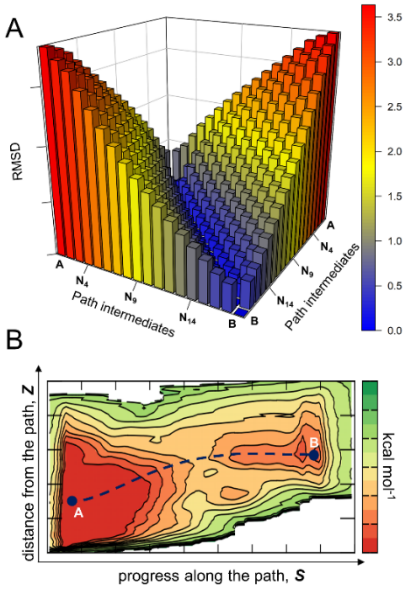
Despite continuous advances in the development of RNA force fields, none of the currently available force fields are a general-purpose resource. Rather, each provides advantages when dealing with particular RNA targets and specific research questions. The main criticisms of conventional force fields include an imbalanced distribution of RNA torsionals (see below) as sampled with AMBER-parmbsc0 with χOL3 corrections.[102] The same set of parameters also failed to correctly identify the most thermodynamically stable structure for three different RNA tetraloops during ns-to-μs-long MD simulations.[103] An overestimation of base-stacking has also been reported.[104] However, while polarizable force fields are being continuously refined, they are still not as computationally efficient as classical force fields.[105] This may limit their current application to small-to-medium RNA systems, where converged simulations can be obtained relatively quickly. These force field imbalances are prompting the development of novel approaches, and we expect further refinements to improve the representation of RNA structural dynamics, as discussed elsewhere.[105–107]

2.3 Rare Events and Free Energy Calculations

Biological molecules, with particular regards to RNAs, are flexible systems and can sample several metastable states,[108] which are often separated by high energy barriers.[109] The transitions between these states are thus limited, and usually occur at longer timescales than those routinely simulated with equilibrium MD.[109] For this reason, such rare events are not properly sampled in classical MD simulations, and several methods have been developed to enhance the MD sampling and favor the characterization of rare transitions.

Metadynamics[110] (MetaD) is an out-of-equilibrium computational approach that aims to enhance MD sampling by adding an external bias to the potential energy of the system under investigation. Specifically, the MetaD bias is applied on selected

degrees of freedom, dubbed collective variables (CVs), which are assumed to fundamentally represent the process under investigation. By way of example, for a reasonably simple problem such as NaCl dissociation, one collective variable could be readily identified as the distance between the Na^+ and Cl^- atoms.[110] However, the choice of the CV, or the set of CVs, could not be straightforward in the case of complex structural transitions. Accordingly, several CVs, more sophisticated than geometrical descriptors, have been designed to address such challenging transitions. For instance, the so-called Path-CV aims to measure the progress along pathways connecting two different configurations of the same system.[111] Assuming that these configurations, A and B, are defined as a set of atomic coordinates (e.g., two conformational states of one RNA molecule), the distance between the states A and B can be monitored using several metrics, including the root-mean-square-deviation (RMSD). Accordingly, it is possible to build a pathway composed of a set of intermediates N, equally spaced between A and B in the RMSD space, which progressively connect A and B (Figure 2.1A). As a result, one variable (S, eq. 2.3) can be defined to monitor the progression of the simulated system (R) over the pathway of N states connecting A and B, while a second variable (Z, eq. 2.4) can trace the distance of the sampled conformations from the reference path (Figure 2.1B).[111]



$$S(R) = \frac{1}{N-1} \frac{\sum_{i=1}^{N}(i-1)e^{-\lambda(R-R_{(i)})^2}}{\sum_{i=1}^{N} e^{-\lambda(R-R_{(i)})^2}} \quad (2.3)$$

$$Z(R) = -\frac{1}{\lambda}\ln\left(\sum_{i=1}^{P} e^{-\lambda\left(R-R_{(i)}\right)^2}\right) \quad (2.4)$$

Figure 2.1: Definition of a pathway between two different configurations of a system. ***(A)*** *The RMSD matrix is plotted for an ideal pathway. Notably, the pathway's intermediates are equally spaced in the RMSD space.* ***(B)*** *Example of the free-energy surface as estimated with metadynamics calculation using S and Z as collective variables.*

Upon the selection of the CV, a history-dependent bias is deposited into the CV space to discourage the system from sampling already explored configurations. Specifically, the bias potential is written as a sum of Gaussians,[110] according to the equation 2.5:

$$V(S,t) = \int_0^t dt' \omega e \left(\frac{\left(S_i(R) - S_i\left(R_{(t')}\right)\right)^2}{2\sigma_i^2} \right) \quad (2.5)$$

Here, S is the collection of *i* CVs, R are the coordinates of the system, σ_i is the Gaussian's width for the *i*th CV, and ω is the ratio between the Gaussian's height (W) and its deposition time (τ).

According to the metadynamics algorithm, the Gaussians are sequentially deposited into the CV space, and when the convergence is reached the free energy landscape can be reconstructed by summing all the Gaussians:

$$F(S) = -\frac{1}{\beta} \ln \left(\int dR \delta \left(S - S(R) e^{-\beta U_{(R)}} \right) \right) \quad (2.6)$$

However, one crucial drawback of metadynamics is that it is difficult to detect and verify the convergence of the simulations. That is, despite the CV space has been fully filled with the Gaussian-shaped potential (i.e., convergence has been reached), such potential is still added into the system, which in turn may be pushed to explore physically unreliable or not relevant configurations.[112] Notably, the so-called well-tempered MetaD (wt-MetaD) formalism has been designed to address this issue.[113] In wt-MetaD, the following external bias is applied during the simulations:

$$V(S,t) = \Delta T \ln \left(1 + \frac{\omega N(S,t)}{\Delta T} \right) \quad (2.7)$$

Thus, the Gaussian-shaped bias deposited in a given point of the CV space decreases together with the inverse of the simulation time, as the height of the Gaussian is reduced the height according to the scaling factor:

$$W = \omega\tau_G \, e^{-\frac{V(S,t)}{k_b \Delta T}} \tag{2.8}$$

Following this approach, however, the bias potential converges only to a fraction of the free energy, which, for long time range, corresponds to:

$$V(S, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \tag{2.9}$$

Thus, for ΔT equal to zero, equilibrium MD is carried out, while tuning ΔT between zero and infinite allows regulating the sampling of the free energy landscape.

2.4 Integrating Experiments and Simulations to Gain Biophysical Reliability

To overcome some of the above-mentioned limitations, especially those related to the need for exhaustive configurational sampling (i.e., long simulation time), MD can often be synergistically integrated with experimental data, such as in solution structural data. Since the 1990s, for example, MD simulations have been combined with NMR experiments to characterize the structural dynamics of HIV-TAR elements. Indeed, Puglisi et al.[114] used NMR data to restrain MD simulations and explore the binding of arginine derivatives with TAR hairpin. They showed that the structural rearrangement of U23, A27, and U38 helps the proper positioning of G26 in order to establish the hydrogen-bond network with the ligand, with implications for HIV replication. With a similar approach, Aboul-ela et al.[115] confirmed the importance of the structural rearrangement of TAR RNA for the binding of the arginine residue that emanates from the viral protein Tat. Importantly, NMR-

informed MD has also been used to analyze and interpret NMR data retrospectively and shed light on undetectable effects, like the perturbation of chemical shifts caused by the ion RNA binding.[116] Indeed, the ion atmosphere plays a significant role in the binding of HIV-TAR with Tat by modulating the electrostatic of the binding partners to help their recognition, as reported by Do and collaborators.[117]

Reweighting is another way to inform MD with NMR data. With reweighting, the MD-pictured conformational landscape is reconstructed a posteriori to match a given experimental observable, such as residual dipolar couplings (RDCs).[118] With this approach, the dynamics of HIV-TAR-1 have been explored beyond the detection limits of traditional NMR.[119] This has revealed that the ligand-unbound hairpin samples a broad range of conformations, including some conformers recognized by known ligands. This suggests that TAR-ligand interaction can occur via conformational selection. This recognition-and-binding mechanism has been used to identify and develop TAR binders as HIV replication inhibitors,[120,121] highlighting the importance of RNA structural dynamics for the selective binding of small molecules. A recently proposed novel approach to integrating NMR data with MD simulations is based on the Shannon maximum-entropy principle.[102] This framework includes the error associated with the experimental data, creating a more robust model that considers account measurement uncertainties related to the ensemble- and time-averaged observables. The authors reconstructed the conformational landscape of four RNA tetraloops and highlighted the limitations of the current RNA force field. They showed that the distribution of an RNA torsional (C2-α) as computed in MD is markedly different from that returned by NMR data. This highlights the importance of integrating experimental data into the MD framework to test and improve the accuracy of molecular simulations.

Chapter 3. Visualizing Group II Intron Dynamics Between the First and Second Steps of Splicing

3.1 Abstract

Group II introns are ubiquitous self-splicing ribozymes and retrotransposable elements evolutionarily and chemically related to the eukaryotic spliceosome, with potential applications as gene-editing tools. Recent biochemical and structural data have captured the intron in multiple conformations at different stages of catalysis. Here, we employ enzymatic assays, X-ray crystallography, and molecular simulations to resolve the spatiotemporal location and function of conformational changes occurring between the first and the second step of splicing. We show that the first residue of the highly-conserved catalytic triad is protonated upon 5'-splice-site scission, promoting a reversible structural rearrangement of the active site (toggling). Protonation and active site dynamics induced by the first step of splicing facilitate the progression to the second step. Our insights into the mechanism of group II intron splicing parallel functional data on the spliceosome, thus reinforcing the notion that these evolutionarily-related molecular machines share the same enzymatic strategy.

3.2 Introduction

Self-splicing group II intron ribozymes are essential regulators of gene expression in all domains of life and they share evolutionary origins and enzymatic properties with the spliceosome, the eukaryotic machinery that catalyzes nuclear splicing of mRNA precursors[25,122]. Spliced group II introns are active retrotransposable elements that contribute to genomic diversification with potential applications in medicine and gene editing[33,123]. Therefore, elucidating the mechanism of group II intron catalysis is crucial for understanding key steps in gene expression and RNA maturation, and to develop therapeutic and biotechnological tools.

The current understanding of group II intron self-splicing mechanism derives from biochemical and cell biology studies[45,124–126] and from 3D structures of introns from various phylogenetic classes.[34,127–131] These studies have provided detailed molecular insights on intron folding[132] and high-resolution molecular snapshots of the Oceanobacillus iheyensis group IIC intron trapped in various conformations throughout the catalytic cycle.[35,37,133–135]

The intron catalytic site comprises a highly conserved triple helix formed by nucleotides of the so-called catalytic triad (in domain D5, C358-G359-C360), two-nucleotide bulge (D5, A376-C377), and J2/3 junction (between D2 and D3, A287-G288-C289, all numbering from the crystallized form of the O. iheyensis intron, i.e. PDB id: 4FAQ; Supplementary Figures A.1A-B). The site also harbors a metal-ion cluster formed by two divalent (M1-M2) and two monovalent (K1-K2) ions (Supplementary Figure A.1A). These ions participate directly in catalysis,[35,36,133,134] which occurs via a series of nucleophilic SN2 reactions (Figure 3.1). In the first step of splicing, depending on whether the intron follows a hydrolytic or a transesterification mechanism, respectively,[46] a water molecule or the 2'-OH group of a bulged adenosine in D6, activated by M2 and by the triple helix, attack the 5'-splice junction of the precursor (5e-I-3e), forming an intron/3'-exon intermediate (I-3e), in which the scissile phosphate is coordinated by K2. In the second step of

splicing, the 5'-exon (5e), activated by M1, performs a nucleophilic attack on the 3'-splice junction, releasing ligated exons (5e-3e) and a linear or lariat form of the excised intron (I; Figure 3.1). The latter can then further reverse splice into cognate or non-cognate genomic DNA, in processes known as retrohoming or retrotransposition.[136,137] Crystal structures of the pre- and post-hydrolytic states are available for the first and second steps of splicing, allowing precise localization of reactants,[35,130] and computational studies have elucidated energetics and dynamics of the related reaction chemistry.[44,138]

However, a key aspect of the group II intron splicing cycle that remains largely uncharacterized is the transition between the splicing steps, when the intron must release products of the first reaction and recruit substrates of the second splicing event. Biochemical and structural studies suggest that, after the first step of splicing, the intron rearranges at the K1 binding site, transiently adopting a specific inactive conformation (aka the toggled conformation), in which G288 (in the J2/3 junction) and C377 (in the two-nucleotide bulge) disengage from their triple helix with the catalytic triad of nucleotides in D5, thereby disrupting the catalytic metal center (Supplementary Figure A.1A).[134,139] Parallel studies also suggest that group II intron conformational changes may be triggered by protonation of active site nucleotides during the splicing cycle.[140] Specifically, the N1 atom of adenosines (N1A) and the N3 atom of cytosines (N3C) can undergo large pKA shifts in folded DNA or RNA and thereby serve as proton donors/acceptors, much like histidine residues in proteins.[141] Consistent with this, functional studies on the spliceosome suggest that protonation within the U6 intramolecular stem-loop (ISL), which is analogous to the group II intron two-nucleotide bulge and catalytic triad, antagonizes binding of catalytic metal ions and induces transient base-flipping during splicing. [142,143]

To understand the transition between the first and second step of splicing, here we probe the group II intron active site by mutagenesis, enzymatic assays, crystallography, and molecular dynamics modeling. We find that, immediately after the first step of splicing, protonation of a conserved nucleobase within the catalytic triad promotes the spontaneous release of K1 and induces intron toggling.

Consistent with this finding, intron mutants that cannot be protonated have defects in the second step of splicing. Our group II intron data have parallels with functional studies on the nuclear spliceosome, suggesting that protonation and toggling are common mechanistic strategies that are adopted by both these splicing machines.

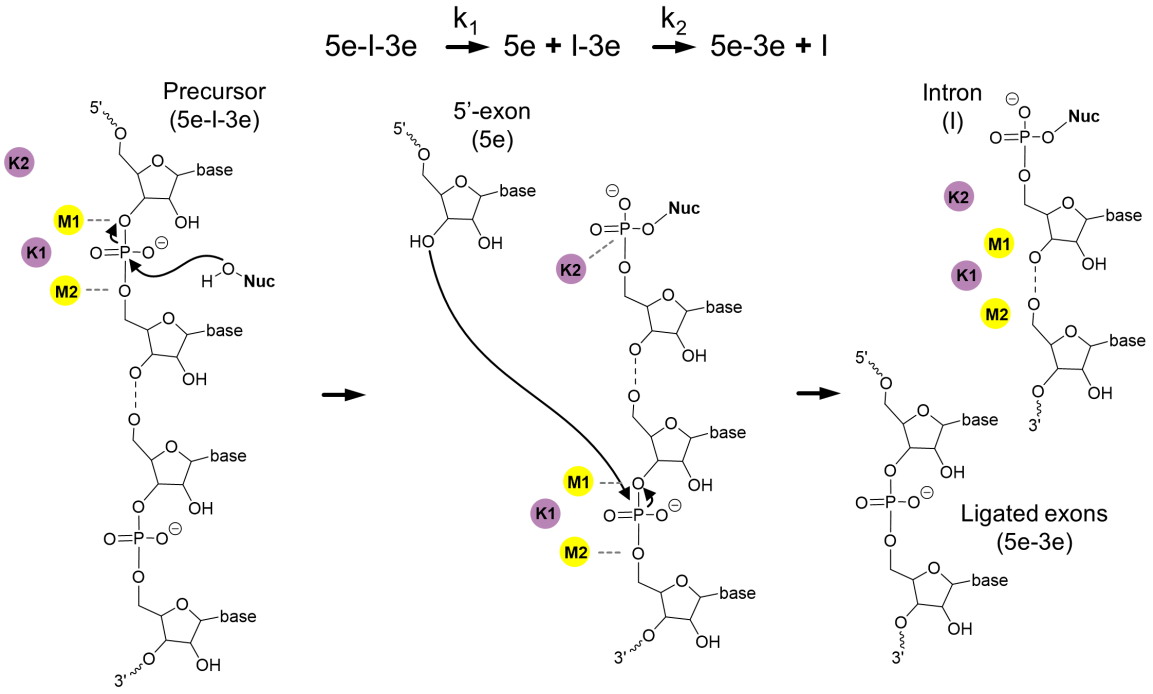


Figure 3.1: Group II Intron splicing mechanism. *Schematics of the splicing reaction and sketch of the chemical mechanism of catalysis by group II introns. k1 is the rate constant of the first and k2 of the second step of splicing. Kinetic rate constants of all constructs are reported in Supplementary Table A.2. Black arrows indicate nucleophilic attacks; gray dotted lines indicate interactions between oxygen atoms of the scissile phosphate groups and catalytic metal ions; Nuc indicates the reaction nucleophile.*

3.3 A Catalytic Residue May Become Protonated During Splicing

Because crystal structures of distinct states of the O. iheyensis group II intron are available, we first analyzed these structures using continuum electrostatics to obtain an initial qualitative approximation of the pK_A values of active site nucleotides (Supplementary Table A.1). Using nonlinear Poisson-Boltzmann calculations, we noted that the pKA value of most residues remains unchanged (Supplementary Table A.1). By contrast, the computed pK_A value of C358 (catalytic triad) shifts between the pre-hydrolytic state (pK_A ~ 4.5 in PDB id: 4FAQ) and the so-called toggled state that forms after the first step of splicing16 (pK_A ~ 7.2 in PDB id: 4FAU). Although these values are qualitative due to the influence of geometrical changes and uncertainties in the definition of the grid and dielectric constants, the Poisson-Boltzmann calculations suggest that C358 has different protonation states along the splicing trajectory (Supplementary Figure A.1C). Consistent with these findings, nucleotide position 358 in other introns can be occupied by an adenine or a cytidine, i.e. bases that can be protonated, but this same position never varies to guanidine or uracil, i.e. bases that cannot be protonated.[144]

Computational studies on the O. iheyensis group II intron immediately after 5'-exon hydrolysis have identified proton transfer pathways from the reaction nucleophile into the bulk solvent involving up to five water molecules (corresponding to a migration distance of ~15 Å).[44] Although less efficient than direct proton transfer, such chains of water molecules enable a proton to shuttle from the nucleophile to the N3 atom of C358 ($N3^{C358}$), which is exposed within the same solvent-filled cavity at a distance of 9.8 Å in the structure of the pre-hydrolytic state (PDB id: 4FAQ)[145] (Supplementary Figure A.2A). Moreover, our hybrid quantum (DFT/BLYP)/classical simulations show that, once a proton is positioned at the N3 atom, C358 remains stably protonated for over 15 ps (see Methods section and Supplementary Figure A.2B-C). Taken together, our observations from continuum electrostatics and quantum mechanical simulations, the specific evolutionary conservation pattern of C358, and its key structural role in the pre-hydrolytic state suggested that C358 plays a direct role in group II intron catalysis.

3.4 Non-protonatable Mutants Show Second Step Splicing Defects

To explore the functional role of C358 in reaction chemistry, we created O. iheyensis splicing precursor constructs[35] in which C358 was replaced with A, G, or U. In addition, to maintain the structural integrity of the catalytic triple helix, we isosterically replaced the two partners of C358, i.e. its Watson-Crick pairing partner (position 385) and its J2/3 triple-helical partner (position 289) (Supplementary Figure A.1D). After incorporating the resulting triple base mutations (C289A/C358A/G385U, aka the A-mutant; C289G/C358G/G385C, aka the G-mutant; and C289U/C358U/G385A, aka the U-mutant), we monitored effects on splicing kinetics.

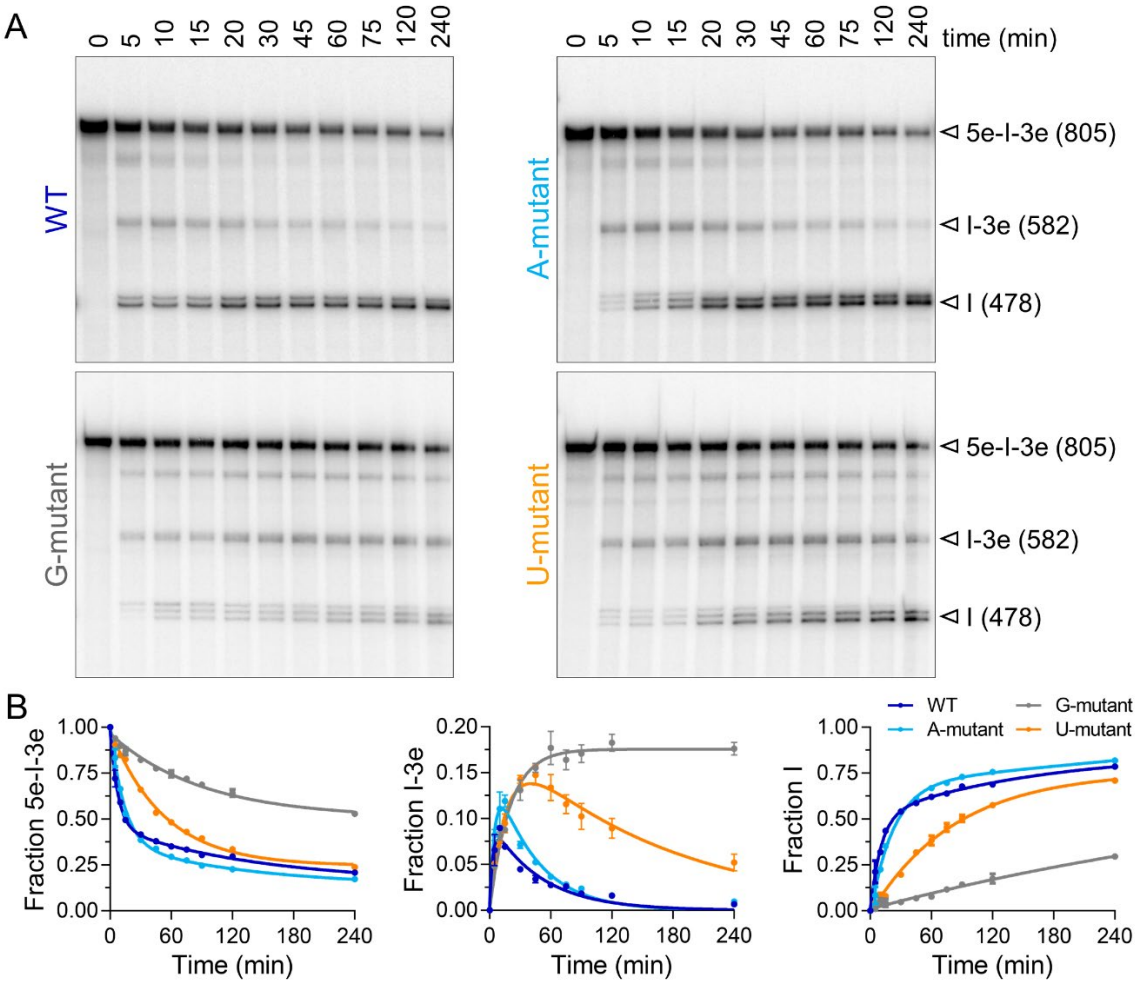


Figure 3.2: Kinetics of intron mutants. ***A)*** *Representative splicing kinetics of wild type intron and A-, G- and U-mutants. Precursors are indicated as 5e-I-3e (nt length in parenthesis). Intermediate (I-3e) and linear intron (I) migrate as double or triple bands because of cryptic cleavage sites, as explained previously16.* ***B)*** *Evolution of the populations of precursor (5e-I-3e, left panel), intermediate (I-3e, middle panel), and linear intron (I, right panel) over time. Error bars represent standard errors of the mean (s.e.m.) calculated from n = 3 independent experiments.*

We found that the A-mutant – which can be protonated at position 358 – splices at rates comparable to wild type, whereas the G- and the U-mutants – which cannot be protonated at position 358 – have splicing defects. Specifically, in the presence of near-physiological potassium and magnesium concentrations, the first splicing step of the G-mutant is ~12-fold slower and that of the U-mutant ~7-fold slower than in wild type. Moreover, the second splicing step of the G-mutant is ~48-fold slower and that of the U-mutant ~8-fold slower than in wild type (Figure 3.2A-B and Supplementary Table A.2). Most remarkably, both G- and U-mutants show accumulation of linear intron/3'-exon intermediate (I-3e), which indicates that these mutants stall after the first step of splicing and have difficulty progressing into the second step (Figure 3.2B, middle panel). These defects are comparable to those of other intron mutants designed to perturb the catalytic triad, such as ai5γ intron double mutants that carry G or U mutations at the nucleotide position analogous to O. iheyensis residue 358 and compensatory mutations of its corresponding Watson-Crick pair.[146] Finally, the splicing defects of our triple mutants are comparable to those of other O. iheyensis group II intron mutants designed to impair toggling, such as the C377G mutant reported in previous studies.[35] In this way, our enzymatic data connect defects in the transition between the two steps of splicing to specific active site mutations that prevent protonation on C358.

3.5 The Mutants Are Structurally Intact but Do Not Toggle

To understand the splicing defects of the G- and U-mutants at the molecular level, we inserted the corresponding mutations into the previously described Oi5eD1-5 construct[35] and visualized the mutant active site by X-ray crystallography.

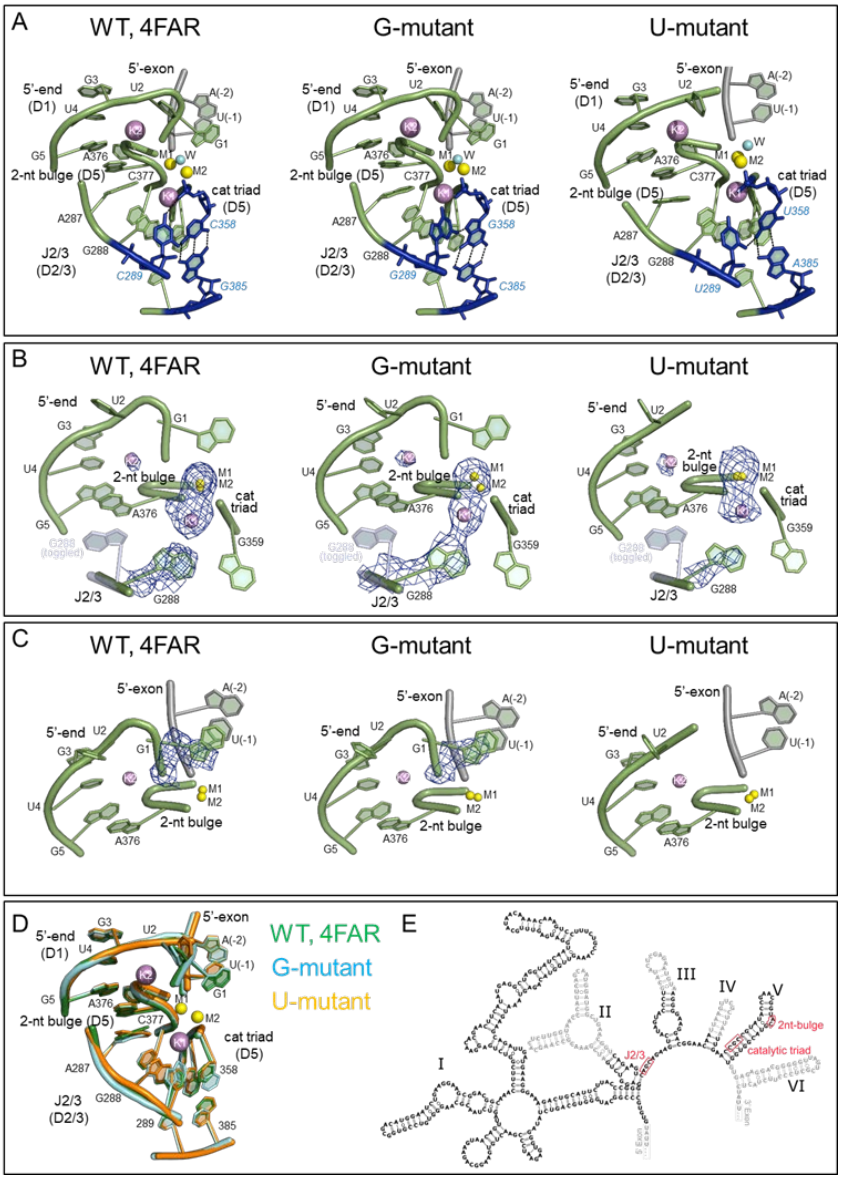


Figure 3.3: Crystal structures of the intron in potassium and magnesium. A) *From left to right: Active site of wild type (WT; PDB id: 4FAR), G-mutant, and U-mutant in the state following 5'-exon hydrolysis. The intron is depicted as a cartoon representation in light green, the 5'-exon is in gray, and the mutated triplex is in blue (black dotted lines indicate H-bonds forming the triple helix).* ***B)*** *Fo-Fc simulated-annealing electron density omit-maps calculated by omitting J2/3 residues (nt 287-289) and the M1-M2-K1-K2 metal cluster from each of the structures depicted in panel A. Positive electron density peaks within 3 Å from the omitted atoms are depicted as blue mesh at a contour level of 3 σ. The toggled conformation of G288 (from PDB id: 4FAX) is depicted as a semi-transparent light blue cartoon representation for reference.* ***C)*** *Fo-Fc simulated-annealing electron density omit-maps calculated by omitting G1 from each of the structures depicted in panel A. Positive electron density peaks within 3 Å from the omitted atoms are depicted as blue mesh at a contour level of 3 σ. G1 is unresolved in the structure of the U-mutant.* ***D)*** *Superposition of the active sites of the structures depicted in panel A.* ***E)*** *Secondary structure of the wild-type O. iheyensis group II intron with catalytic site elements highlighted by red boxes and nucleotides not present in the crystal structures in grey.*

First, we determined crystal structures of the G- and the U-mutants in the presence of potassium and magnesium at 3.4 and 3.6 Å resolution, respectively (Table A.3). Both mutants have a folded structure similar to that of the post-hydrolytic state of the wild type intron after the first step of splicing (PDB id.: 4FAR; root mean square deviation (RMSD)4FAR-Gmutant = 0.49 Å, RMSD4FAR-Umutant = 0.43 Å; Figures 3.3A and 3.3D). Importantly, both mutant structures adopt the triple-helical configuration that corresponds with that of the wild type intron structure (Figure 3.3A and Supplementary Figure A.1D). The Fo-Fc simulated-annealing electron density omit-maps calculated by omitting the J2/3 residues and the catalytic metal cluster reveal strong electron density signal for the triple helix conformer, as in wild type (total peak height for the nucleobase of G288 = 8.9 σ and 6.7 σ for the G- and the U-mutants, respectively; maximum peak height for the metals = 9.5 σ and 6.6 σ for the G- and the U-mutants, respectively; Figure 3.3B). Moreover, the Fo-Fc maps calculated by omitting the first intron nucleotide (G1) show that the 5'-splice junction has undergone cleavage in both mutants during the crystallization process (Figure 3.3C). In summary, the similarity of these mutant structures with that of wild type suggests that, despite some reductions in rate, the first step of splicing is structurally and mechanistically unaffected by the G and U mutations.

We then determined the crystal structures of the G- and the U-mutants in the presence of sodium and magnesium at 3.2 and 3.3 Å resolution, respectively (Table A.3). In this case, both mutants adopt overall structures similar to wild type (PDB id.: 4FAX; RMSD4FAX-Gmutant = 3.9 Å, RMSD4FAX-Umutant = 0.75 Å; Figure 3.4A). However, the detailed architecture of the active site differs significantly from wild type under sodium conditions. For wild type, these conditions induce a rotation of the backbone in the J2/3 region, which breaks the triple helix structure and generates the so-called toggled conformation that is implicated in the transition between the first and the second step of splicing16 (Figures 3.4B-C). By contrast with wild type, the G- and the U-mutants in sodium maintain the triple helix configuration, as revealed by the Fo-Fc maps calculated by omitting the J2/3

residues and the catalytic metals (total peak height for the triple helix conformer of the G288 nucleobase = 7.5 σ in the G-mutant and = 9.3 σ in the U-mutant; Figures 3.4B-C). Therefore, these structures show that the G- and U-mutants are unable to adopt the toggled conformation, which may explain their tendency to stall after the first step of splicing. Taken together, the enzymatic and structural data suggest that C358 protonation and active site toggling facilitate the rearrangement of the intron active site between the two steps of splicing.

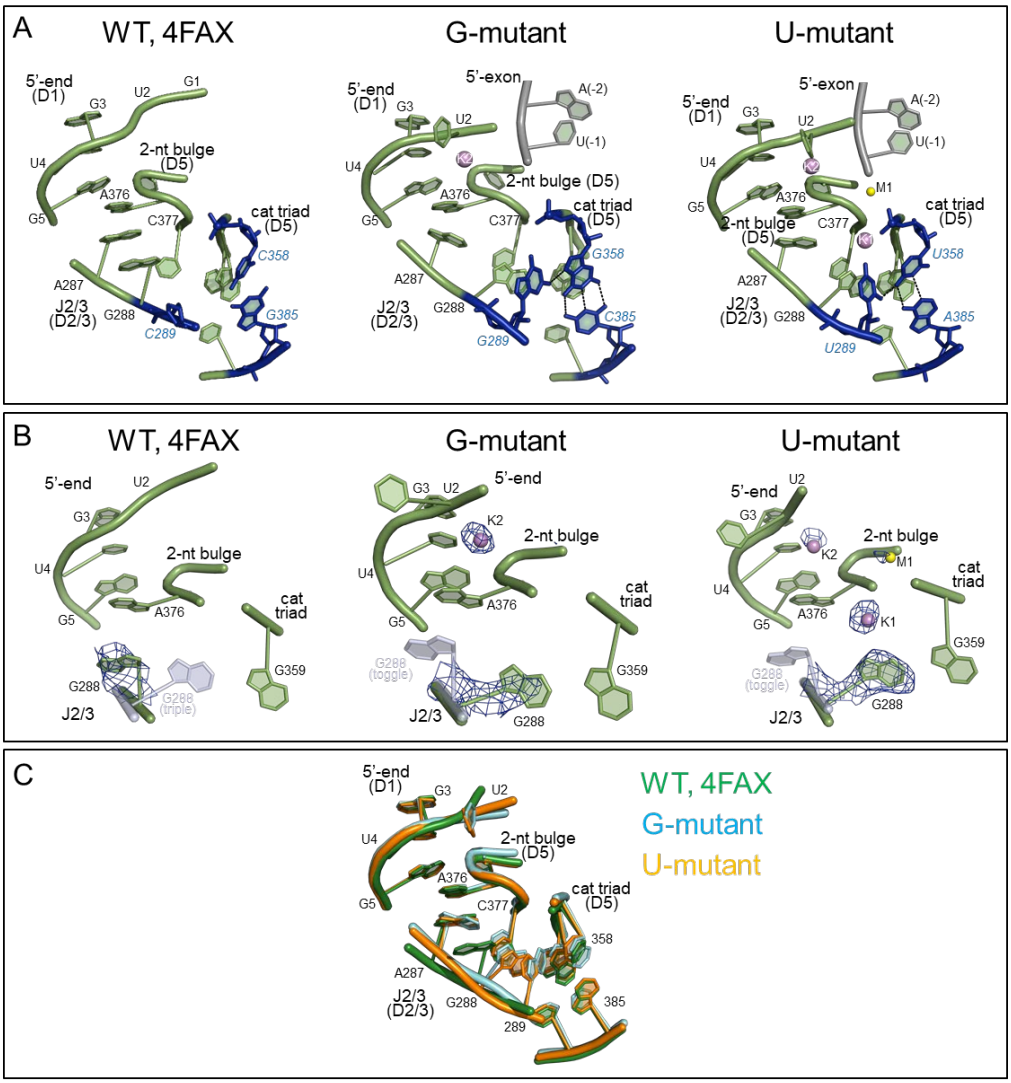


Figure 3.4: Crystal structures of the intron in sodium and magnesium. ***A)*** *From left to right: Active site of wild type (WT; PDB id: 4FAX), G-mutant, and U-mutant. Active site elements are depicted as in Figure 2A.* ***B)*** *Fo-Fc simulated-annealing electron density omit-maps calculated by omitting J2/3 residues (nt 287-289) and the M1-M2-K1-K2 metal cluster from each of the structures depicted in panel A. Positive electron density peaks within 3 Å from the omitted atoms are depicted as blue mesh at a contour level of 3 σ. For the wild type structure (left), the triple helix conformation of G288 is depicted as semi-transparent light blue cartoon representation, the M1-M2 metals as semi-transparent yellow spheres, and the K1-K2 metals as semi-transparent violet spheres (all using coordinates from PDB id: 4FAR). Analogously, the toggled conformation of G288 and unresolved metals of the cluster are depicted as semi-transparent representations for the G- and U-mutant (middle and right panels, using coordinates from PDB id: 4FAX for G288 and 4FAR for unresolved metals).* ***C)*** *Superposition of the active sites of the structures depicted in panel A.*

3.6 Scission of the 5'-exon Disrupts the Catalytic Metal Cluster

To establish how C358 protonation and active site toggling are mechanistically connected, and to understand the chain of events that regulate active site rearrangement, we performed force-field-based molecular dynamics (MD) simulations. We used a flexible nonbonded approach for the metal center (see Methods section), followed by comparative analyses of multiple systems built using the published structures of the wild type O. iheyensis group II intron captured at different stages of catalysis[35,37] and the structures of the G- and U-mutants. These structures represent the highest resolution crystallographic data available for group II introns and display the most detailed architecture of an intron active site, including all metals and first splicing step reactants.[35,37]

We initially investigated the dynamics of the wild type intron in the pre-hydrolytic state (PDB id: 4FAQ; two classical MD simulations for ~ 600 ns and ~ 1.2 μs, respectively). We observed that, shortly after equilibration (~25 ns), K1 shifted closer to the N7 atom of G288 ($N7^{G288}$), which was concomitant with the weakening of the K1 interaction with $O5'^{G359}$ observed in the crystal structure (dK1-$N7^{G288}$ = 2.98 ± 0.27 Å in the simulations, dK1-$N7^{G288}$ = 4.3 Å in PDB id: 4FAQ, Figure 3.5 and Supplementary Figure A.3). In both simulations, the system was structurally stable, especially nucleotides within the active site (domain D5 and junction J2/3). This was reflected in the average RMSD = 1.95 ± 0.27 Å (Supplementary Figure A.3) and the fact that catalytic triad residues maintained positions observed in the crystal structure (dM1-M2 4.24 ± 0.04 Å in the simulations, dM1-M2 = 4.3 Å in PDB id: 4FAQ). These simulations suggest that the pre-hydrolytic configuration does not have a tendency to undergo structural rearrangements.

We then investigated the dynamics of the wild type intron after 5' exon hydrolysis, thus considering the post-hydrolytic state (PDB id: 4FAR) in protonated (three simulations, ~350 ns each) and non-protonated (six simulations, ~750 ns each) configurations (Supplementary Figure A.4). In these simulations, the overall

structural fold was stably maintained, with an averaged RMSD of 4.72 ± 0.65 Å. As in the simulations of the pre-hydrolytic state, K1 shifted closer to the $N7^{G288}$ after equilibration (~25 ns, dK1-$N7^{G288}$ = 2.82 ± 0.15 Å in the simulations, dK1-$N7^{G288}$ = 4.4 Å in PDB id: 4FAR). However, none of the post-hydrolytic systems were able to release the products of the first step of splicing. For example, the scissile phosphate (SP) appears locked by the K2 ion in the proximity of the active site and the nucleobase of G1 remains stably coordinated to M1-M2 (Supplementary Figure A.4). These observations suggest that the post-hydrolytic crystal structures used in these simulations may represent an unproductive low energy configuration of the intron that is not directly relevant to the pre-second step splicing configuration.

To address this issue, we modeled an active site state of the wild type intron that would provide an improved starting point for simulations. We started with the structure of the pre-hydrolytic state (PDB id: 4FAQ), broke the scissile bond, and inverted the stereochemistry of the scissile phosphate (further modeling details in the Methods section and in Supplementary Figure A.5). This state represents the intron immediately after the first step of splicing, where the scissile phosphate has just been cleaved but is still coordinated by M1 and M2 (Supplementary Figure A.5). Also for this 'cleaved' state, we simulated both protonated and non-protonated forms of C358 (two simulations per system, ~600 ns per simulation). In all cases, the system showed considerable stability, with an overall RMSD of 4.61 ± 0.81 Å. During these simulations, the K1-$N7^{G288}$ interaction was formed and initially preserved. Moreover, the scissile phosphate was not sequestered by K2 outside the active site. In other words, the distance between the scissile phosphate and M2 was constantly maintained at dSP-M2 = 3.23 ± 0.10 Å (Supplementary Figure A.5). Intriguingly, in the protonated state, after ~20 ns of simulation, a water molecule bridged $O6^{G288}$ and M2, such that these two atoms became closer to each other (dM2-O6 = 5.71 Å in PDB id: 4FAQ; dM2-O6 = 4.75 ± 0.23 Å in the simulations; Figure 3.6 and Supplementary Figure A.5). Concomitantly, the value of dM1-M2 increased from 4.31 Å to 5.05 ± 0.13 Å (Figure 3.6 and Supplementary Figure A.5). Importantly, at this point, the coordination shell of K1 was perturbed, and the K1-

N7G288 interaction broke, leading to the spontaneous release of K1 from the active site into the bulk solvent after just additional ~30 ns (Figure 3.6 and Supplementary Figure A.5). Notably, these events occurred also in the non-protonated state, although less promptly. In this case, the initial conformational changes occurred after ~200 ns, with K1 released soon after, at ~250 ns. Interestingly, in all cases, the conformational ensemble of the active site after the release of K1 differed from the characteristic triple helix configuration.

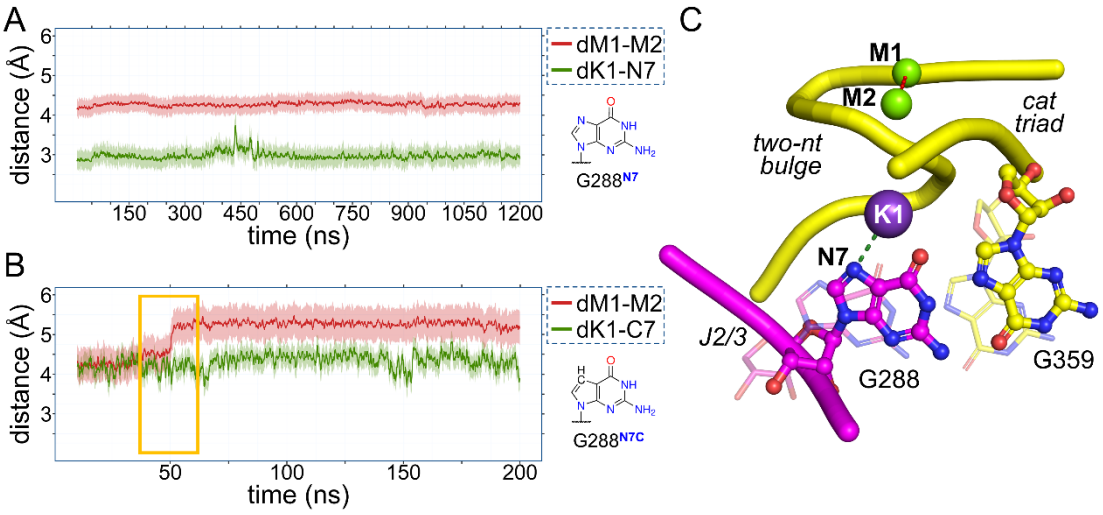


Figure 3.5: Importance of the K1 interaction with N7G288. ***A)*** *Changes during MD simulations of the pre-hydrolytic state for key structural descriptors dK1-N7G288 (green trace; light green shading indicates the standard deviation of the corresponding distance) and dM1-M2 (red trace). The pre-hydrolytic state is stable and does not undergo any structural rearrangement, but it forms the K1-N7 interaction rapidly after the equilibration phase.* ***B)*** *MD simulations of the N7-deaza state (structure of N7-deaza-G in the inset). The absence of the K1-N7 interaction alters the conformation of G288, eventually leading to active site misfolding (dM1-M2 = 5.28 ± 0.12 Å, red trace).* ***C)*** *Graphical representation of the K1-N7 interaction modeled from the simulations of the pre-hydrolytic state. The J2/3 junction (purple, ball and stick representation), M1 and M2 (green spheres), and the catalytic triad and two-nucleotide bulge (yellow, ball and stick representation, backbone as ribbon) are highlighted, together with the key descriptors reported in panels A and B (dotted lines represent dK1-N7G288 and dM1-M2). Disruption of K1-N7 causes G288 and G359 to move from their triple helix conformation (i.e. from the pre-hydrolytic state, solid colors) to a state in which the triple helix is broken (i.e. as simulated for the N7-deaza state, semi-transparent representation).*

To specifically monitor the triple helix geometry, we used the following two geometrical parameters: (i) the distance between the O2 atom of C289 (J2/3) and the N4 atom of G358 (D5, catalytic triad) (d289-358), which adopts values ≤ 3 Å in the triple helix configuration and > 3 Å when the triple helix is disrupted; and (ii) the angle α between the nucleobases plains of C358 and its Watson-Crick pair G385, which adopts values ≤ 0.35 rad in the triple helix configuration and > 0.35 rad when the triple helix is disrupted (Supplementary Figure A.6). Indeed, d289-358 = 2.7 Å and α = 0.17 rad in the crystallized pre-hydrolytic state (PDB id: 4FAQ), which

harbors K1 and adopts the triple helix configuration. Notably, though, after K1 release in our MD simulations, d289-358 reached average values of ~ 4.88 ± 1.05 Å and α reached average values of ~ 0.63 ± 0.14 rad in the protonated state (~ 3.07 ± 0.14 Å and ~ 0.47 ± 0.10 rad in the non-protonated state, respectively), suggesting that the triple helix is destabilized and the active site may toggle under these conditions (Supplementary Figure A.6).

Finally, we also simulated the crystallized G- and U-mutants in the cleaved and post-hydrolytic states (eight simulations, ~600 ns each; Supplementary Figures A.7 and A.8). We noted that the K1-N7G288 interaction was not stably formed in the mutants, preventing K1 release (Supplementary Figures A.7 and A.8). Thus, the triple helix was stabilized in its crystallographic conformation.

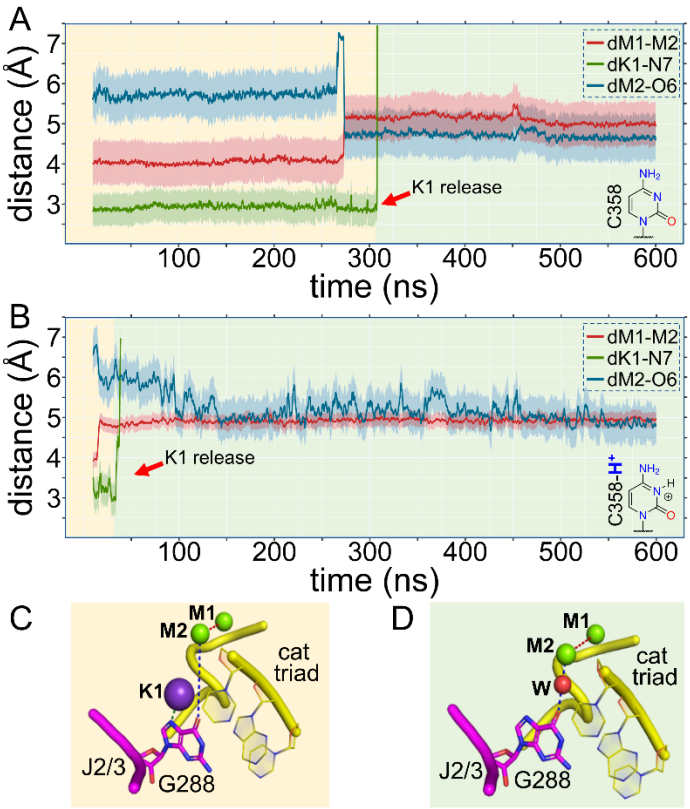


Figure 3.6: Protonation of N3C358 favors K1 release. *Evolution of dM1-M2 (red trace), dK1-N7 (green trace), and dM2-O6 (blue trace) over the course of MD simulations of:* ***A)*** *the cleaved state, and* ***B)*** *the cleaved-H+ state. Shading around the traces indicates the s.d. of the corresponding distance Schematic structures of non-protonated and protonated cytidine groups are depicted in the bottom right of each panel. K1 is released after ~300 ns and ~ 40 ns from the two states, respectively (red arrow).* ***C)*** *Structure of the active site at the beginning of the simulations after the formation of K1-N7G288.* ***D)*** *Structure of the active site at the end of the simulations, after K1 release. K1 is released after the solvation of the active site (water molecule depicted as a red sphere, W).*

For example, in the simulations, d289-358 = 2.67 ± 0.24 Å and α = 0.19 ± 0.11 rad for the G-mutant (α = 0.15 rad in PDB id 6T3K) and d289-358 = 1.95 ± 0.23 Å and α = 0.33 ± 0.13 rad for the U-mutant (α = 0.31 rad in PDB id=6T3R) (Supplementary Figures A.7 and A.8). These simulations suggest that the G- and U-mutants are unlikely to toggle in their cleaved form.

Taken together, these data suggest that K1 is stably bound to the active site in the pre-hydrolytic state of the wild type intron, but is spontaneously released from the active site immediately after 5'-exon hydrolysis. The release of K1 breaks the catalytic triple helix, and the intron begins sampling the toggled conformation. Such a rearrangement is significantly favored by protonation of $N3^{C358}$, and it does not occur in the G- and U-mutants, which cannot be protonated.

3.7 The K1-$N7^{G288}$ Interaction Stabilizes the Intron Active Site

Interestingly, in the simulations of the wild type intron described above, but not in the simulations of the mutants, K1 establishes a stable interaction with N7G288 within a very short time after equilibration (Figures 3.5-3.6 and Supplementary Figures A.4-A.8). Moreover, simulations of the cleaved state immediately after 5'-exon hydrolysis showed that interaction with N7G288 is a necessary step for releasing K1 from the active site (Figure 3.6 and Supplementary Figure A.5). Importantly, an N7-deaza mutation at position G288 was shown to impair the first step of splicing34. These observations suggest that the K1- N7G288 interaction may be structurally and functionally important for splicing. To test this hypothesis, we modeled the N7-deaza mutation at G288 in the pre-hydrolytic state (PDB id: 4FAQ), and we tested the importance of the K1-N7G288 interaction for the proper folding of the active site. Three classical MD simulations of these in silico mutants (~200 ns each) showed that the loss of the K1-N7G288 interaction irreversibly destabilized the triple helix, causing separation of M1-M2 (averaged dM1-M2 = 5.28 ± 0.12 Å, Figure 3.5B) and eventually leading to the unfolding of the active site.

These data suggest that the K1-N7G288 interaction plays a crucial role in preventing premature release of K1 and consequent disruption of the triple helix.

3.8 Toggling Energetics Agree With Catalytic Rate Constants

To appropriately sample and semi-quantitively evaluate the energetics associated with intron toggling, we used path-metadynamics (MtD).[147] We performed enhanced sampling MtD simulations starting from either the cleaved protonated or non-protonated wild type models and terminating at the toggled state (referred to as the the $cH^+ \rightarrow T$ and the $c \rightarrow T$, respectively; see details in the Methods section). The reference path involves exclusively the J2/3 junction, which rearranges as defined from structural data,[35] and employs two collective variables that trace i) the progress of the system along the reference path (variable S), and ii) the distance of the sampled conformations from the reference path (variable Z). In this way, MtD simulations sample the conformational space to find the lowest energy path for the conformational change under investigation. Notably, the non-bonded metal cluster M1-M2-K1-K2 and its extended coordination shell at the catalytic site can freely explore conformational space during these simulations.

Mechanistically, in simulations where C358 was protonated, the system first sampled a large, deep free energy minimum that contained multiple isoenergetic conformational states. While A287 freely explored the conformational space, C358 protonation disrupted the canonical WC base pairing with G385, leading to C358 rotation (d289-358 = 5.85 ± 1.94 Å and α = 0.34 ± 0.08 rad, state A, Figure 3.7). This spontaneous rearrangement promoted hydration of the K1 binding site, with consequent prompt release of K1 to the bulk solvent, disruption of the hydrogen bond contacts between C358 and C289, and further separation of these two residues (d289-358 = 10.33 ± 1.75 Å and α = 1.24 ± 0.18 rad, state B, Figure 3.7). In this conformation, the flexibility of J2/3 nucleotides was enhanced, allowing G288 and C289 to rotate out of the active site and to stack with A287, thus enabling the disruption of the C377-C360 base pair (Toggled state, Figure 3.7). The computed energetic barrier for this overall transition ($\Delta G^{\ddagger}_{cH^+-T}$) was ~20 kcal mol^{-1}, while the final metastable toggled state had a value of about +5 kcal mol^{-1} relative to the triple helix conformer (Figure 3.7).

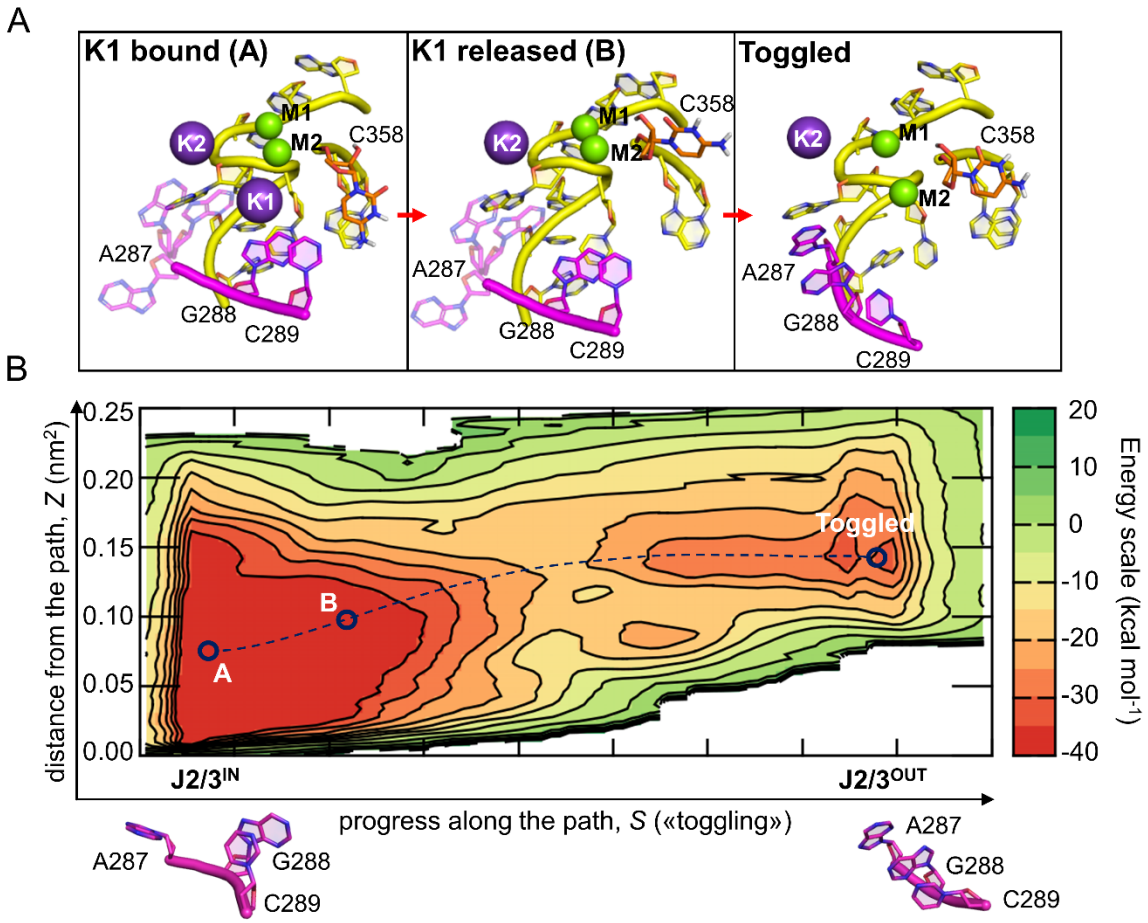


Figure 3.7: Energetics associated with intron toggling in the protonated state. A) *Structural architecture of the active site for the intermediate A, intermediate B, and toggled states identified by the MtD simulation reported in panel B.* ***B)*** *Path MtD free energy landscape of the cleaved-H⁺ state. The intermediate A, intermediate B, and toggled states are indicated along the MtD trajectory (dark blue dotted line). The energy scale is indicated in kcal mol⁻¹ on the right. The conformations of the J2/3 junction in state A and in the toggled state are represented at the bottom of the figure.*

Nucleotides within J2/3 also rearranged in the non-protonated configuration, albeit with different dynamics and higher energy barriers. Indeed, with the spontaneous rotation of A287, the intron rearranged into the first intermediate state (d289-358 = 2.97 ± 0.21 Å and α = 0.17 ± 0.08 rad, state A', Supplementary Figure A.9, which is the lowest free energy minimum), in which K1 is more exposed to the bulk water. K1 is then released simultaneously to the partial rotation of G288. This led to the disruption of the triple helix and the formation of a second intermediate state (d289-358 = 7.97 ± 1.19 Å and α = 0.31 ± 0.16 rad, state B', Supplementary Figure A.9). Finally, the stacking of A287 and C289 with G288, together with the rotation of C377, completed the conformational rearrangement and formed the final toggled state (Supplementary Figure A.9). The computed free energy barrier for this

transition ($\Delta G^{\ddagger}_{c\text{-}T}$) was ~25 kcal mol^{-1}, which is therefore less favorable than that of the protonated intron ($\Delta G^{\ddagger}_{cH^{+}\text{-}T}$ ~20 kcal mol^{-1}). Importantly, the final metastable toggled state had a value of about +5 kcal mol-1 relative to lowest free energy minimum (state A', Supplementary Figure A.9). These computed activation barriers are a good match with empirical values calculated using the experimental splicing rate constants inserted into the Eyring–Polanyi equation[148,149] ($k_1 = 0.031 \pm 0.003$ $min^{-1} \rightarrow \Delta G^{\ddagger}_1 = 22.8$ kcal mol^{-1} ; $k_2 = 0.026 \pm 0.003$ $min^{-1} \rightarrow \Delta G^{\ddagger}_2 = 22.9$ kcal mol^{-1}). This result corroborates our proposed toggling mechanism, indicating that conformational rearrangements of the intron active site between the catalytically active triple helix configuration and the toggled structure captured crystallographically16 are compatible with catalysis.

3.9 Discussion

By combining structural, enzymatic, and computational methods, we have elucidated the molecular mechanism for the transition between the two steps of group II intron splicing and we have described the dynamic behavior of the intron active site as it moves through the splicing process (Figure 3.8).

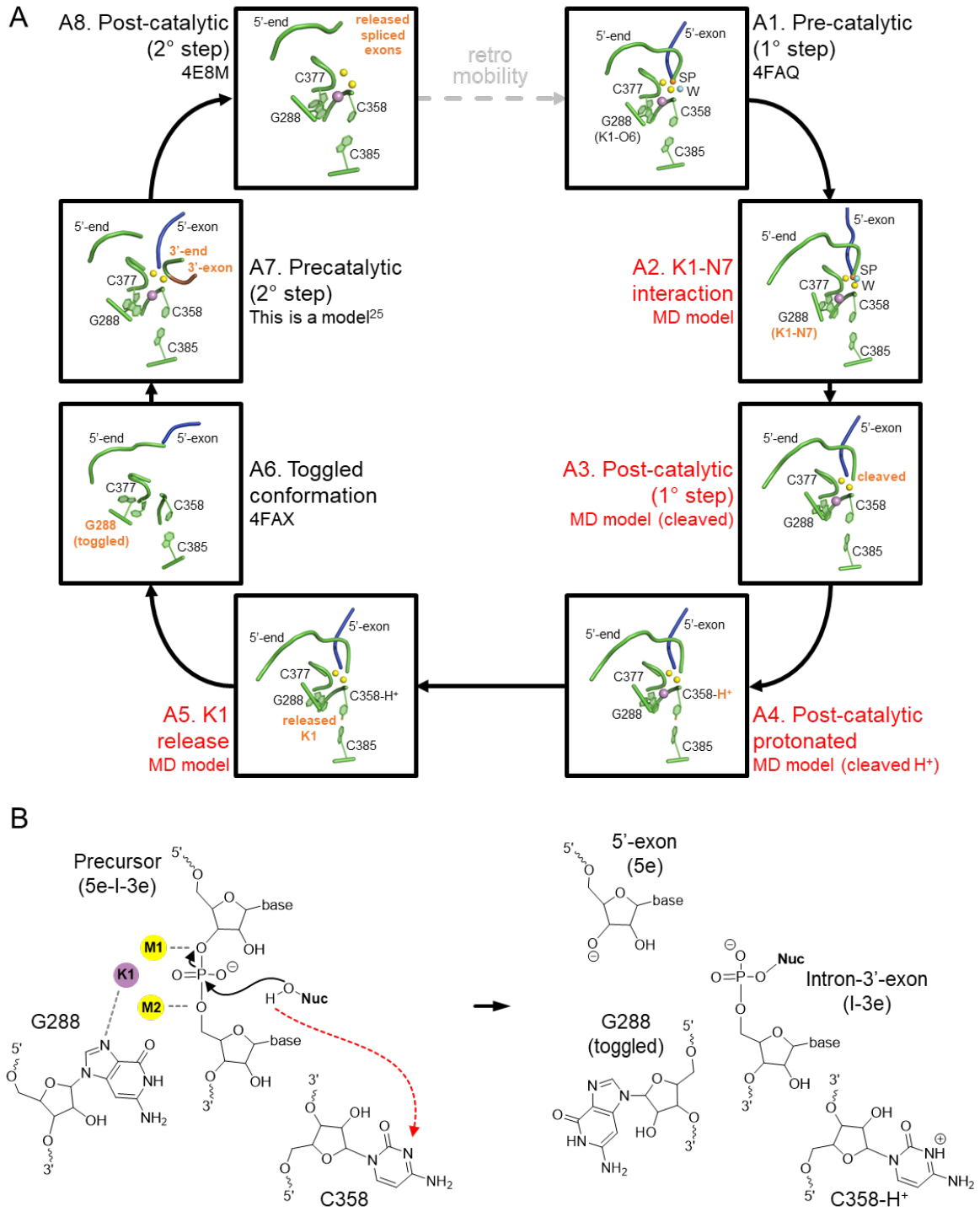


Figure 3.8: Revised group II intron splicing cycle. A) *Snapshots of the intron active site derived from the crystal structures and the MD simulations. In the pre-catalytic state, which corresponds to the pre-hydrolytic state of the first step of splicing, a water molecule (light blue sphere, W) or the 2'-OH of a bulged adenosine are poised for nucleophilic attack on the scissile phosphate (orange sphere, SP) (PDB id. 4FAQ). In this state, the intron established the K1-N7^{G288} interaction, and the nucleophile cleaves the intron-5'-exon junction. Immediately after hydrolytic cleavage, the proton released by the nucleophile initially to the bulk solvent is shuttled to C358 in the catalytic triad. C358 protonation favors release of the metal ions cluster and toggling of J2/3 junction (PDB id. 4FAX). This conformational rearrangement likely prompts the release of the products of the first step of splicing, subsequent rearrangement of D6, and reverse toggling of J2/3 to reconstitute the active site and align second step reactants. Finally, cleavage of intron-3'exon junction leads to the release of free linear intron (PDB id. 4E8M) and spliced exons. The free intron is still an active ribozyme, which can retrotranspose into target genomic DNA and re-initiate a new splicing cycle (dashed grey arrow). Relevant intron motifs are shown as cartoon representations in green. The 5'-exon is in blue, the 3'-exon in brown. K1 is shown as a violet sphere, M1 and M2 as yellow spheres. K2 is not shown for clarity. Intron states described in this work are labeled red, and differences between consecutive panels are indicated as bold orange labels.* ***B)*** *Sketches of the intron active sites corresponding to panels A1/2 (precatalytic state with K1-N7 interaction, left) and A5/6 (K1 release and toggling, right) drawn in the same style as in Figure 1A. For clarity, only the 5'-splice site is represented. The red dotted arrow indicates the proton transfer pathway from the nucleophile (Nuc) to N3^{C358}.*

In the pre-hydrolytic state, the group II intron adopts the triple helix conformation, which coordinates the heteronuclear metal cluster M1-M2-K1-K2.[35,36] M2 and the phosphate backbone of C358 deprotonate the reaction nucleophile for catalyzing the scission of the 5'-splice-site. At this stage, the proton released into bulk solvent by the reaction nucleophile[44] is transferred to the N3 atom on the C358 nucleobase, either via specific proton transfer pathways, as previously proposed[44] (an illustration of one possible transfer pathway is reported in Supplementary Figure A.2A) or via simple diffusion through the solvent. Independent on the exact proton transfer mechanism, hybrid quantum-classical simulations suggest that C358 remains stably protonated on N3, never exchanging its proton with surrounding water in the quantum region (Supplementary Figure A.2B-C). Indeed, it is remarkable that position 358 is often occupied by an adenosine (which is readily protonated) in the majority of group II introns and in the spliceosome, but it never varies to G or U (which are nucleobases that cannot be protonated). C358 protonation thus emerges as a previously-unrecognized event that stimulates the progression of the intron towards the second step of splicing. Indeed, when we experimentally replaced C358 with adenosine, splicing was unaffected. But when we replaced C358 with either a G or a U, the mutated intron accumulated linear intron/3'-exon intermediate (I-3e), indicating a defect in the progression onto the second step of splicing. Notably, and in line with the MD simulations of protonated and non-protonated wild type intron in the cleaved post-hydrolytic state (see below), splicing is not completely inhibited in the mutants, suggesting that

protonation accelerates splicing but it is not essential. These differences in kinetics likely constitute a phenotypic advantage for the intron, which has preserved protonatable residues at position 358 throughout evolution. When position 358 is occupied by a G or U residue, steric or electrostatic perturbations may also contribute to the observed splicing defects. The extent of such perturbations may be qualitatively inferred from the behavior of the A-mutant, which – despite being protonatable at position 358 – contains a bulkier purine substitute. Remarkably, splicing defects of the A-mutant are minimal (~2-fold, Figure 3.2 and Supplementary Table A.2) and crucially, they do not lead to accumulation of an I-3e intermediate (Figure 3.2 and Supplementary Table A.2), suggesting that the defects of the G- and U-mutants predominantly derive from their inability to become protonated at position 358. The crystal structures of these mutants additionally confirm that these constructs preserve triple helix architecture, so any perturbation of their active site must be minimal. Finally, the crystallographic data suggest that the G- and U-mutants do not sample the toggled conformation, thus impairing a critical rearrangement of the intron active site between the two steps of splicing.

Molecular dynamics (MD) simulations enabled us to dissect the precise sequence of events that lead from 5'-splice site cleavage to toggling. Most importantly, we observed that 5'-exon hydrolysis induces a spontaneous and prompt release of K1 from the active site, as previously hypothesized.[35] This key event, which happens only in the post-hydrolytic but not in the pre-hydrolytic state, is much favored by protonation of C358, which induces fast K1 release (just after ~50 ns when protonated, as compared to ~250 ns in the non-protonated state, with an energetic barrier of $\Delta G^{\ddagger}_{cH^{+}\text{-}T} \sim 20$ kcal mol^{-1} and $\Delta G^{\ddagger}_{c\text{-}T} \sim 25$ kcal mol^{-1}, respectively). These observations reveal that K1 is a highly dynamic ion, despite its tight coordination to nearly all active site residues in the catalytic triple helix configuration, and it plays an extremely crucial role during splicing. The interaction of K1 with $N7^{G288}$, in the J2/3 junction, is particularly important for stabilizing the intron active site in the catalytically competent configuration and for controlling the binding and release of

K1 within the active site along the splicing cycle. These results explain why $N7^{G288}$-deaza mutants are defective in splicing.[150]

As a result of K1 release, the triple helix conformation becomes unstable. Under such circumstances, G288 toggles out of the active site, undergoing backbone rotations that expose the Watson-Crick face of guanosine to functional groups in D3 and the Hoogsteen face to a cavity that is likely occupied by D6 and by 3'-splice junction residues.[35,127,129,130] In this conformation, G288 is thus optimally placed to promote key interactions that facilitate the second step of splicing (see below). MtD simulations show that the energy required for such conformational toggling is compatible with catalytic rate constants. Importantly, mutants that are defective for toggling, either because they cannot be protonated or because their triple helix is stable even under conditions where the wild type toggles (i.e. the G- and U-mutants described here, and the C377G mutant studied previously),[35] fail to progress onto the second step of splicing.

Toggling of the J2/3 junction and progression to the second step of splicing is also likely to involve A287 (nucleotide γ). In our simulations, A287 establishes a canonical WC interaction with the second nucleobase of the intron (U2; $d_{U2\text{-}A287}$ = 2.13 ± 0.32 Å; Supplementary Figure A.10), which was maintained as long as K1 remained in the active site, but which was broken when K1 left the active site and the intron toggled. Such findings suggest that G288 toggling may be needed to release A287 from U2. This process would ensure the formation of the essential γ-γ' interaction, in which A287 pairs with its partner nucleotide γ' in D6 during the second step of splicing.[127,151,152] After recruiting D6 via A287, the toggled intron would then reestablish the catalytic triple helix conformation by reverse toggling, explaining how the first and second steps of splicing are mechanistically connected. Based on the simulations, the toggled state is ~5 kcal mol^{-1} higher in free energy compared to the triple helix state, suggesting that reverse toggling is energetically inexpensive. It is therefore tempting to speculate that protonation and toggling also occur at the end of the second step, which might favor the release of the splicing product, and reduce the chances of spliced exons reopening.[136,137] In either case,

these processes may be further facilitated by participation of the intron-encoded maturase protein.

The idea of a rearrangement involving J2/3 and formation of a transiently inactive intermediate is compatible with the mechanism of splicing via the branching pathway. Indeed, biochemical data and recent crystal structures of lariat introns show that the hydrolytic and transesterification pathways occur at the same active site, involve positioning of the reaction nucleophile (the proton donor) in the exact same structural position compared to nucleotide 358, and follow the same reaction chemistry.[37,46,127,129,130] Indeed, G- and U-mutations at the 358-equivalent position of the lariat-forming ai5γ intron from *S. cerevisiae* (A816G and A816U) display splicing defects similar to our G- and U-mutants.[146] Moreover, a protonation-dependent structural rearrangement mechanism is strongly supported by functional data obtained on the spliceosome, which is evolutionarily and chemically analogous to the group II intron.[25,122] In the spliceosome, the last G (G52 in yeast) of the conserved ACAGAGA box in U6 snRNA corresponds to intron G288.[144] This residue is in close proximity to the branch site,[153] it interacts with the 5'-splice site, and it undergoes a rearrangement between the splicing steps[154] in a process that is modulated by protein subunits (i.e. Prp8, Prp16)[155,156] and potassium ions.[73,157] Such reorganization of G52 facilitates the release of the 5'-end of spliceosomal introns from the active site after the first splicing step, while also favoring the recruitment of the 3'-splice junction into the active site for the second step of splicing.[154] These rearrangements could be induced by protonation of nucleotides of the U6 ISL, which are analogous to the group II intron two-nucleotide bulge and catalytic triad because their protonation antagonizes binding of the catalytic metal ions to the spliceosome and induces a transient base-flipping conformational change.[142,143] Furthermore, G52 mutations in the spliceosome have an inhibitory effect on the second step of splicing,[158] similar to the effects we described for G288 in the group II intron in this and in previous work.[136,159] Finally, during the splicing cycle, the spliceosome adopts transiently inactive states, possibly similar to the group II intron inactive toggled conformation, to avoid processing non-ideal pre-mRNA substrates.[160] In the

light of these structural and functional analogies between the intron and the spliceosome, it seems plausible that conformational toggling and dynamics of catalytic metal ions in the active site may regulate spliceosomal activation, too.

In summary, through the integration of four X-ray structures of active site mutants and *in vitro* splicing assays with multi-microsecond classical molecular simulations and free energy calculations, we have elucidated the dynamical behavior and determined the functional role of structural rearrangements within the group II intron active site, showing how they contribute to the mechanism of RNA splicing. We have determined that critical dynamic processes are triggered by protonation of a highly-conserved catalytic residue, thereby promoting the transition between the first and the second steps of splicing. Importantly, the resulting mechanism explains the apparent paradox of how and why a tightly bound metal ion cluster can be broken and reformed during the catalytic cycle, thereby promoting a directional sequence of coordinated chemical reactions. These findings may help in future engineering of complex, RNA-based enzymes for use as biotechnological tools and gene-specific therapeutics.[33,71]

3.9 Methods

Cloning and mutagenesis

The constructs of *Oceanobacillus iheyensis* group II intron used in this work are the pOiA wild type and the OiD1-5 crystallization constructs of the *Oceanobacillus iheyensis* group II intron.[35] All mutagenesis experiments were performed using the PfuUltra II Hotstart PCR Master Mix (Agilent). The restriction enzymes *Cla*I and *Bam*HI used for template linearization were purchased from NEB. All constructs were confirmed by DNA sequencing (W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University, and Eurofins).

In vitro transcription and purification

Following restriction with the appropriate endonucleases at 37°C overnight, the intron was transcribed *in vitro* using T7 polymerase.[35] For crystallization purposes,[133,134] it was then purified under non-denaturing conditions,[161] rebuffered and concentrated to 80 μM in 10 mM $MgCl_2$ and 5 mM sodium cacodylate pH 6.5. For splicing studies, the intron was radiolabeled during transcription, purified in a denatured state,[35] and subsequently refolded.

Splicing assays

Purified radiolabeled intron precursor was refolded by denaturation at 95°C for 1 minute in the presence of 40 mM Na-MOPS pH 7.5, and cooled at room temperature for 2 minutes. Subsequently, the appropriate monovalent ions were added to a final concentration of 150 mM. Finally, $MgCl_2$ was added to a final concentration of 5 mM to start the splicing reaction. The refolded precursor samples were incubated at 37°C. 1 μL aliquots of the splicing reactions taken at specific time-points were quenched by the addition of 20 μL gel loading solution containing urea and chilled on ice. The samples were analyzed onto a denaturing 5 % (w/v) polyacrylamide gel.

The kinetic rate constants were calculated using the Prism 6 package (GraphPad Software).

Crystallization

The natively-purified intron was mixed with a 0.5 mM spermine solution in 10 mM $MgCl_2$ and 5 mM sodium cacodylate pH 6.5, and with the crystallization buffer in a 1:1:1 volume ratio.[133] Crystals were grown at 30°C by the hanging drop vapor diffusion method using 2 μL sample drops and 300 μL crystallization solution in a sealed chamber (EasyXtal 15-Well Tool, Qiagen). Crystals were harvested after 2 – 3 weeks. Crystals were cryo-protected in a solution containing the corresponding crystallization buffers supplemented with 25 % EG and immediately flash-frozen in liquid nitrogen. The crystallization solutions used to solve the structures of the excised intron presented in this work were composed of: (i) 50 mM Na-HEPES pH 7.0, 100 mM magnesium acetate, 150 mM potassium chloride, 10 mM lithium chloride, 4% PEG 8000 for the G-mutant in potassium and magnesium (PDB id: 6T3K), (ii) 50 mM Na-HEPES pH 7.0, 100 mM magnesium acetate, 150 mM potassium chloride, 10 mM lithium chloride, 4% PEG 8000 for the U-mutant in potassium and magnesium (PDB id: 6T3R), (iii) 50 mM Na-HEPES pH 7.0, 100 mM magnesium acetate, 150 mM sodium chloride, 4% PEG 8000 for the G-mutant in sodium and magnesium (PDB id: 6T3N), and (iv) 50 mM Na-HEPES pH 7.0, 100 mM magnesium acetate, 150 mM sodium chloride, 4% PEG 8000 for the U-mutant in sodium and magnesium (PDB id: 6T3S).

Structure determination

Diffraction data were collected with an X-ray beam wavelength of 0.979 Å and at a temperature of 100 K at beamlines 24ID-C and E (NE-CAT) at the Argonne Photon Source (APS), Argonne, IL, and processed with the Rapid Automated Processing of Data (RAPD) software package (https://rapd.nec.aps.anl.gov/rapd/) and with the XDS suite.[162] The structures were solved by molecular replacement

using Phaser in CCP4[163] and the RNA coordinates of PDB entry 4FAR (without solvent atoms) as the initial model.[35,133,134] The models were improved automatically in Phenix[164] and Refmac5,[163] and manually in Coot,[165] and finally evaluated by MolProbity.[166] The figures depicting the structures were drawn using PyMOL.[167] Stereo images of selected regions of the electron density are reported in Supplementary Figure 11.

pK_A calculations

We used continuum electrostatics calculations based on the nonlinear Poisson-Boltzmann equation to estimate the pK_A of C358 in the pre-hydrolytic state (PDB id. 4FAQ) and in the toggled state (PDB id. 4FAU). Calculations were performed with DelPhiPKa,[168] using a pH range from 0 to 14 with a pH interval of 0.5, a dielectric constant for RNA $\varepsilon_{RNA} = 4$, and a dielectric constant for solvent $\varepsilon_{solvent} = 80$. Metals were not considered in the calculations.

Structural models for molecular dynamics (MD) simulations

We have used ten systems for MD simulations: *(i)* The pre-hydrolytic state, a wild type system modeled on PDB id: 4FAQ;[35] *(ii)* The N7-deaza state, a pre-hydrolytic state in which $N7^{G288}$ was replaced by a carbon atom; *(iii)* The cleaved state, a pre-hydrolytic state in which the phosphodiester bond between the intron and the 5'-exon was broken introducing an oxygen atom and inverting the stereochemical configuration of the scissile phosphate, and in which Ca^{2+} ions were replaced with Mg^{2+} ions; *(iv)* The cleaved-H^+ state, a cleaved state protonated on $N3^{C358}$; *(v)* The post-hydrolytic state, a wild type system modeled on PDB id: 4FAR; *(vi)* The post-hydrolytic H^+ state, a post-hydrolytic state protonated at $N3^{C358}$; *(vii)* The post-hydrolytic G-mutant, modeled on the structure of the G-mutant in potassium and magnesium; *(viii)* The cleaved G-mutant, a cleaved state carrying the C289G/C358G/G385C triple mutations; *(ix)* The post-hydrolytic U-mutant, modeled on the structure of the G-mutant in potassium and magnesium; *(x)* The

cleaved U-mutant, a cleaved state carrying the C289U/C358U/G385A triple mutations. Each system was hydrated with a 12-Å layer of TIP3P[169] water molecules, and the ions concentration was set to the same used for crystallization.[35] All the crystallized ions and water molecules were considered for model building. The final models are thus enclosed in a box of ~145·125·144 $Å^3$, containing ~220,000 water molecules, resulting in a total of ~250,000 atoms for each system.

MD simulation set up

The AMBER-ff12SB (ff99+bsc0+χOL3)[97] was used for the parametrization of the RNA. Nucleotide G288 in the N7-deaza model, nucleotide C358 in the cleaved-H^+ and post-hydrolytic H^+ models, and both 5'- and 3'- terminal nucleotides in all models were parametrized with the general amber force field (i.e. GAFF),[170] and their atomic charges were derived with RESP procedure.[171] We used the Joung-Cheatham parameters[172] for the monovalent metal ions, while the divalent metal ions were parametrized according to Li and Merz.[173] In the simulations, we have used ionic concentrations of 100 mM for magnesium ions and 150 mM for potassium ions, in line with the crystallization conditions of the intron (see above). The two catalytic metal ions were modeled using a flexible nonbonded approach based on the atoms in molecules partitioning scheme.[80,174,175] All MD simulations were performed with Gromacs 5.1.4.[176] The integration time step was set to 2 fs, while the length of all covalent bonds was set with the P-LINCS algorithm.[177] A temperature of 310K was imposed using a velocity-rescaling thermostat[178] with a relaxation time τ = 0.1 ps, while pressure control was achieved with Parrinello-Rahman barostat[179] at reference pressure of 1 atm with τ = 2 ps. Periodic boundary conditions in the three directions of Cartesian space were applied. Particle Mesh Ewald method, with a Fourier grid spacing of 1.6 Å, was used to treat long-range electrostatics. All the systems were subjected to the same simulation protocol. To relax the water molecule and the ions, energy minimization was carried out. At this stage, active core ions (M1, M2, K1, K2, K4)[36] along with the RNA backbone were kept fixed with harmonic positional restraints of 500 kcal/mol·$Å^2$. Subsequently, the

systems were heated up from 0 to 310 K with an NVT simulation of ~1 ns with the same positional restraints used in the energy minimization. A second NVT of ~1 ns was then performed at a fixed temperature (310 K), halving the positional restraints. Additionally, ~1 ns of NPT simulation was performed with 100 kcal/mol·Å^2 residual restraints on the backbone and the core ions to allow partial backbone relaxation. Finally, different production runs were performed in the NPT ensemble for each system. We collected overall more than 15 μs of MD trajectories, specifically: (i) ~1.8 μs for the pre-hydrolytic system, 2 replicas; (ii) ~600 ns for the N7-deaza system, 3 replicas; (iii) ~1.2 μs for the cleaved system, 2 replicas; (iv) ~1.2 μs for the cleaved-H^+ system, 2 replicas; (v) ~4.5 μs for the post-hydrolytic system, 6 replicas; (vi) ~1 μs for the post-hydrolytic H^+ state, a post-hydrolytic state protonated at $N3^{C358}$, 3 replicas; (vii) ~1.2 μs for the post-hydrolytic G-mutant, modelled on the structure of the G-mutant in potassium and magnesium, 2 replicas; (viii) ~1.2 μs for the cleaved G-mutant, a cleaved state carrying the C289G/C358G/G385C triple mutations, 2 replicas; (ix) ~1.2 μs for the post-hydrolytic U-mutant, modelled on the structure of the G-mutant in potassium and magnesium, 2 replicas; (x) 1.2 μs for the cleaved U-mutant, a cleaved state carrying the C289U/C358U/G385A triple mutations, 2 replicas. For each system, statistics were collected after the systems reached the equilibration (i.e. stabilization of the RMSD of the nucleic acid backbone), thus discarding the first 25 ns of the trajectories.

Metadynamics simulations

The reference path was built upon the different conformations of the nucleotides U285 to A290 in the pre-hydrolytic and toggled states (PDB id: 4FAQ and 4FAX, respectively).[35] The two structures were used to generate 30 interpolated intermediates through the MolMov morphing server.[180] Each intermediate was subjected to energy minimization, and 16 snapshots were chosen to build the path. Each node of the path (i.e. intermediate structure) is equally spaced with a distance of ~ 0.32 Å. According to Branduardi et al.,[111] we defined two-path collective

variables (pCV): (i) *S*, which defines the progress along the reference path; (ii) *Z*, which measures the distance from the reference path. To sample the free energy landscape, we used adaptative-width metadynamics as implemented in Plumed,[147,181] in which the width of the gaussian was determined by the fluctuation of S and Z over a time interval of 1 ps. A lower-bound limit for the width of the gaussian was set to 0.03 in the appropriate unit for each coordinate. The height of the gaussian was set to 0.3 kJ/mol with an additional frequency of 1 ps. By considering the distance between the nodes of the path, we set a $\lambda = 23.66$ Å^{-2}. We collected: (i) 350 ns, for the transition cleaved state to toggled state (referred to as $c \rightarrow T$); (ii) 200 ns for the transition cleaved-H^+ state to the toggled state (referred to as $cH^+ \rightarrow T$).

Hybrid mechanical/molecular mechanical (QM/MM) simulations

Hybrid quantum mechanical/molecular mechanical (QM/MM) simulations were performed on the structure of the cleaved-H^+ state with CP2K molecular dynamics engine[182] to explore the stability of the protonated form of $N3^{C358}$. The AMBER force field was used for the MM subsystem, whereas Density Functional Theory (DFT) was used to describe the QM atoms. The BLYP functional[183] supplemented by a dispersion correction was employed.[184] The Quickstep algorithm was used to solve the electronic structure problem,[185] employing a double zeta plus polarization (DZVP) basis set[186] to represent the valence orbitals and plane waves for the electron density (320 Ry cutoff). Goedecker-Teter-Hutter (GTH) type pseudopotentials were used for valence-core interactions.[187] Wavefunction optimization was achieved through an orbital transformation method[188] using a threshold of $5 \cdot 10^{-7}$ on the electronic gradient as a convergence criterion. The QM/MM coupling follows the protocol proposed by Laino and collaborators.[188] Simulations were performed in the NVT ensemble (300K), employing a velocity rescaling thermostat[178] in the initial configuration. After about 4.3 ps, a second water molecule was included in the QM

region, and the simulation was restarted to collect 15 ps of simulation time. $N3^{C358}$ remained stably protonated throughout the entire simulation.

Chapter 4. Controlled Trafficking of Multiple and Diverse Cations Prompts Nucleic Acid Hydrolysis

4.1 Abstract

Recent *in crystallo* reaction intermediates have detailed how nucleic acid hydrolysis occurs in the RNA ribonuclease H1 (RNase H1), a fundamental metalloenzyme involved in maintaining the human genome. At odds with the previous characterization, these *in crystallo* structures unexpectedly captured multiple metal ions (K^+ and Mg^{2+}) transiently bound in the vicinity of the two-metal-ion active site. Using multi-microsecond all-atom molecular dynamics and free-energy simulations, we investigated the functional implications of the dynamic exchange of multiple K^+ and Mg^{2+} ions at the RNase H1 reaction center. We found that such ions are timely positioned at non-overlapping locations near the active site, at different stages of catalysis, being crucial for both reactants' alignment and leaving group departure. We also found that this cation trafficking is tightly regulated by variations of the solution's ionic strength and is aided by two conserved second-shell residues, E188 and K196, suggesting a mechanism for the cations' recruitment during catalysis. These results indicate that controlled trafficking of multi-cation dynamics, opportunely prompted by second-shell residues, is functionally essential to the complex enzymatic machinery of the RNase H1. It also opens new catalytic possibilities for other similar metalloenzymes including, but not limited to, CRISPR-Cas9 and group II intron ribozyme.

4.2 Introduction

Recently, structural and biophysical studies have revealed that additional metal ions (*viz.,* K^+ and Mg^{2+}) are transiently engaged in the vicinity of the well-recognized two-metal-ion catalytic site of enzymes like DNA/RNA polymerases,[77,189] endo- and exo-nucleases,[190–192] and even type II topoisomerase.[69,70,193] All these metalloenzymes are fundamental for the expression and maintenance of RNA and DNA within the cell,[193–195] and are often targeted to treat human diseases, from cancer to viral and bacterial infections.[196–198]

Remarkably, transient K^+ and Mg^{2+} ions have been often captured unexpectedly close to conserved second-shell protein residues, which have been thereby proposed to favor metal recognition and recruitment from the bulk.[76] Such positive ions at the metal-aided catalytic site during the processing of nucleic acids were suggested to contribute actively to the overall catalytic process,[59,73,77–79,175,199,200] although their exact functional and dynamic role remains only partially understood.

In this context, the prototypical nucleic acid-processing enzyme RNA ribonuclease H1 (RNase H1, Figure 4.1A), which cleaves phosphodiester bond in RNA:DNA hybrids being responsible for removing the Okazaki fragments during DNA replication,[61] has been extensively characterized via biophysical[62,201–206] and computational[64,66,207] studies.

Nevertheless, only recent data, obtained using time-resolved X-ray crystallography,[72] have shown that the catalysis of RNase H1 is accompanied by the trafficking of both mono (K^+) and divalent (Mg^{2+}) metal ions, transiently bound in the vicinity of the two-metal-ion catalytic site, at different stages of catalysis (Figure 4.1B-C). Indeed, the new X-ray structures of the *Bacillus Halodurans* RNase H1 in complex with an RNA:DNA hybrid substrate have captured the *in crystallo* intermediates of the catalysis (Figure 4.1B-C). The active site displays both catalytic Mg^{2+} ions (M_A-M_B) properly located and coordinated by first-shell conserved carboxylates, referred to as the DEDD motif (D71, E109, D132, D192).

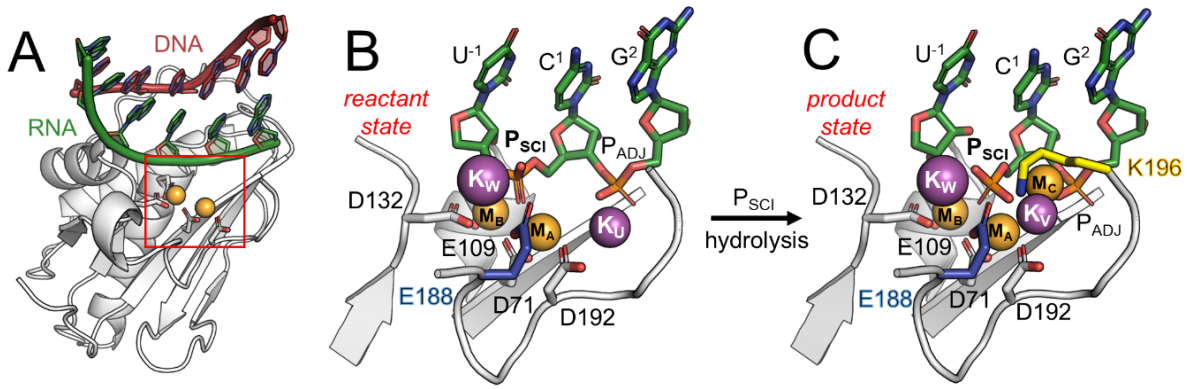


Figure 4.1: RNase H1 catalytic intermediates captured by time-resolved X-ray crystallography.[72] ***(A)*** *Overview of the RNase H1 (white) in complex with an RNA:DNA hybrid (green:red).* ***(B)*** *Catalytic site before RNA hydrolysis (i.e., reactant state, PDBid: 6DO9). The catalytic residues forming the DEDD motif (D71, E109, D132 and D192; white) and the second shell residue E188 (blue) are represented as sticks. RNA nucleotides (green) including the scissile phosphate (P_{SCI}) and its adjacent phosphate (P_{ADJ}) are also shown. The two catalytic magnesium ions M_A-M_B (orange) and the additional potassium ions K_W and K_U (purple) are shown as spheres.* ***(C)*** *Catalytic site upon RNA hydrolysis (i.e., product state, PDBid: 6DOX). Here, K_U is replaced by the third divalent metal ion M_C (orange), or by the monovalent K_V (purple), while the second-shell residue K196 (yellow) can directly interact with the scissile phosphate.*

These novel findings support the possible functional role of such transitory metal ions.[59,73,77–79,175,199] Intriguingly, prior RNA hydrolysis, two K^+ ions are captured at the active site, at different times (Figure 4.1B). The X-ray structure solved 40 seconds (40s) after the *in crystallo* incubation shows that a K^+ ion locates at the "U" site (and is hereafter named K_U, PDBid: 6DMV, Figure B.1A).[72] In this position, K_U binds to the catalytic D192 and the phosphate adjacent (P_{ADJ}) to the scissile group (P_{SCI}). Notably, at this stage, no reaction is observed as the Michaelis-Menten complex is not adequately formed. The distance between Pro-Sp oxygen (O_{Sp}) of the scissile phosphate and the catalytic M_B is 2.6Å (dM_B–O_{Sp} in Figure B.1A, close-up view). Subsequently, after 120s, a second K^+ ion locates at the "W" site (*viz.,* K_W, Figure 4.1B; PDBid: 6DO9),[72] where it is directly coordinated to the second-shell residue E188 and to P_{SCI}, in agreement with previous proposal from classical molecular simulations of this enzyme.[66] Notably, at this stage, the scissile phosphate results properly positioned for the catalysis, profiting of the increased vicinity to M_B (with the dM_B–O_{Sp} = 2.2Å, Figure B.1B, close-up view). Then, upon RNA hydrolysis, while K_W is still present at the active site (Figure 4.1C), the occupancy of K_U decreases with product formation. Indeed, after 360s, a third K^+ ion locates very close to the "U" site, at the "V" site (*viz.,* K_V, Figure 4.1C; PDBid: 6DOX),[72] eventually substituting K_U. Moreover, at high concentrations of divalent metals

[M^{2+}], a third Mg^{2+} ion is captured to directly coordinate the scissile phosphate at the "C" site (*viz.,* M_C, Figure 4.1C; PDBid: 6DPD).[72] The *in crystallo* catalysis has also shown that the RNase H1 activity depends on the ionic strength of the reaction buffer.[72] Indeed, the formation of reaction products increases with the concentration of both K^+ and Mg^{2+} cations (*i.e.,* [K^+] and [Mg^{2+}]), although an excess of ions leads to the so-called "attenuation effect", which reduces RNA hydrolysis.

Interestingly, catalysis is impaired for the single mutants E188A or K196A, in presence of Mg^{2+}.[72] The X-ray structures of such mutants have revealed an altered active site architecture as compared to the wild type (wt) RNase H1 upon RNA hydrolysis, with only partial product formation. In the product structure of the E188A mutant, no K^+ ion locates at the "W" site (Figure B.2A), in contrast to what observed in the wt form, indicating that E188 directly impacts the presence of K^+ at the active site. On the other hand, in the structure of the K196A mutant (Figure B.2B), one K^+ ion is captured at the "W" site, as in the wt RNase H1. Nonetheless, in absence of K196, the products are misfolded due to a slight rotation of the scissile phosphate, in contrast to what was observed in the reactive center of the wt form. Most importantly, no other M_C locates at the active site in this K196A mutant.

Such a wealth of structural and biochemical evidence suggests a functional role of multiple metal ions (K^+ and Mg^{2+}) transiently bound in the vicinity of the two-metal-ion active site in RNase H1, where metal trafficking and exact localization at the reaction center seems critical for catalysis.[72] Indeed, this mechanistic hypothesis would justify the stage-dependent location of additional mono and divalent metals during catalysis, as observed in the *in crystallo* intermediates of RNase H1.

To test this hypothesis and examine the dynamics of this complex step-wise and metal-aided enzymatic process, we applied extensive all-atom molecular dynamics (MD) and free-energy simulations (>27 μs, in total) on multiple systems of RNase H1, at different stages of the catalysis and ionic strengths of the solution. Our results identify an ordered dynamics of multiple cations and their controlled trafficking at the RNase H1 active site as a functional element of the enzymatic strategy to prompt RNA hydrolysis.

4.3 Two Transient K^+ Ions Are Alternatively Located at the Reactants for Catalysis.

According to structural data,[72] the RNase H1 catalytic site can contain three K^+ ions, namely K_U, K_W, and K_V, which are located at the respective binding site "U", "W", and "V" (Figure 4.1). The *in crystallo* data also show that this extended K^+ ion cluster is transiently formed and disrupted during catalysis. Only K_U is initially bound (at 40s, when no reaction is observed, PDBid: 6DMV).[72] Then, in the reactant state, K_U occupancy decreases, while K_W locates at the active site (at 120s, PBBid: 6DO9; at 200s, PDBid: 6DOB). Finally, K_W and K_V are found in the products (at 360s PDBid: 6DOX).[72]

To investigate K^+ ion's dynamical recruitment and binding at the catalytic site, we first performed two MD simulations of ~1 us each, where both K_U and K_W were removed from the reactant state (PDBid: 6DO9). Such simulations were run at the reference and optimal concentration for catalysis (in line with the experimental conditions of 6mM [Mg^{2+}] and 200mM [K^+]). First, we noted that even in the absence of K^+ ions, the overall catalytic architecture is maintained well (see RMSD in Figure B.3). Monitoring the spatial distribution and number of K^+ ions in a sphere of 6 Å centered on the scissile phosphate (P_{SCI}, Figure 4.2A), we found that a transient K^+ ion can be intermittently located at the W-site (*i.e.,* K_W). In this position, K_W sits at ~2.77 ± 0.24 Å from P_{SCI}, (Figure S3), in excellent agreement with X-ray data (PDBid: 6DO9, K_W at = 2.5 Å). On the other hand, K^+ ions do not occupy the "U" site, which remains empty (Figure 4.2A and B.4A).

We also found that the spontaneous binding of K_W is associated with a conformational change of the second-shell residue E188 (Figure 4.2B), which can assume two main orientations, pointing towards the active site as in the X-ray structure (*viz*., E188$_{IN}$), and outside (*viz.,* E188$_{OUT}$) toward the bulk. By monitoring the variation of the torsional angle E188-θ (formed by the E188 atoms N-Cα-Cβ-Cδ, Figure 4.2C, lower panel), we found that E188 mainly samples the E188$_{IN}$ state when K_W locates at the "W" site (shown in the upper panel).

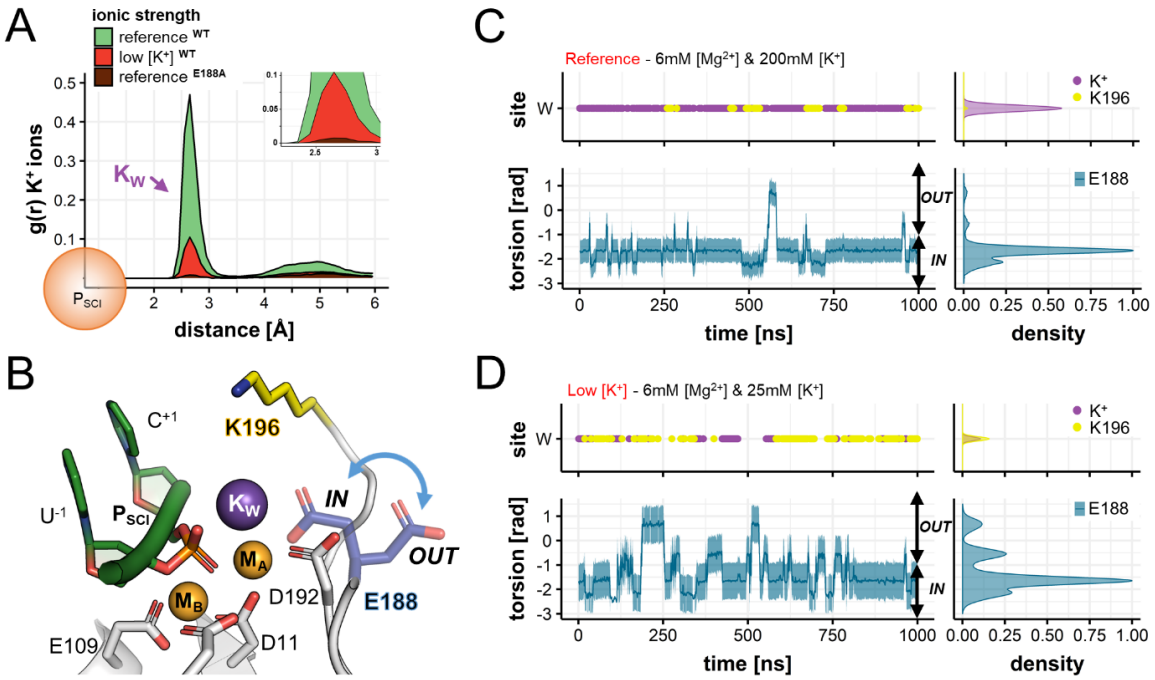


Figure 4.2: Occupancy of positive charges (i.e., K^+ ions and K196) and conformational dynamics of E188 in the reactant state. ***(A)*** *Spatial distribution and number of K^+ ions around the scissile phosphate (P_{SCI}) of wild type RNase H1, computed during MD simulations as the radius of gyration, g(r), for K^+ ions at reference (green) and low (red) K^+ concentrations. In both cases, K^+ ions are mainly found at ~2.77 ± 0.24 Å from the scissile phosphate, thus located at the "W" site. The g(r) for K^+ is also computed during MD simulations of the E188A mutant performed at reference concentration (brown). Here, the "W" site is never occupied by K^+ ions.* ***(B)*** *Conformational change of the second-shell residue E188 (blue arrow), pointing inside the active site ($E188_{IN}$) and outside ($E188_{OUT}$).* ***(C)*** *Occupancy of the "W" site (upper panel) by K^+ ions (purple) and by K196 (yellow); and variation of the torsional angle E188-θ (lower panel, computed between the N-Cα-Cβ-Cδ atoms) along MD simulations at reference [K^+]. Normalized densities are on the right. At optimal ionic strength, the binding of K^+ is favored and E188 is stabilized in the $E188_{IN}$ conformation.* ***(D)*** *Same descriptors as in (C) for MD simulations at low [K^+]. Here, the binding of K_W is disfavored, substituted by K196, and the inner ↔ outer swings of E188 are more frequent.*

On the other hand, E188 samples the $E188_{OUT}$ state when the "W" site is not occupied by K_W. This correlation, which is also observed in the simulation replicas (Figure B.S6), suggests that the inner ↔ outer conformational swing of E188 can favor the recruitment of K^+ ions from the bulk to the "W" site. Importantly, these findings are also in agreement with biochemical data showing that catalysis is impaired in the E188A mutant.[72] While the active site architecture of this mutant is highly similar to the wt RNase H1 (Figure B.2A), the only structural difference is the lack of K_W (PDBid: 6DPO).[72] Accordingly, we further evaluated the role of E188 for the dynamic recruitment of K_W, and performed two additional MD simulations (1us, each). First, we simulated the E188A mutant at optimal ionic strength. As a result, no K^+ ions bind at the active site (Figure 4.2A), at odds with what was observed for the wt RNase H1. This result confirms that the presence of the second-shell E188 promotes the binding of K_W. Then, we performed MD simulations of the

wt RNase H1, restraining the E188 residue in the "out" conformation. Here, the occupancy of the "W" site by K^+ ions was dramatically reduced (Figure B.6C), suggesting that the inner ↔ outer conformational swing of E188 is fundamental to guarantee the recruitment and proper positioning of K_W, which thus appears crucial for catalysis.

As noted above, K^+ ions do not occupy the "U" site during such MD simulations. To further corroborate this observation, a third simulation replica has been performed by locating one K^+ ion at the "U" site, based on the coordinates of the *in crystallo* RNase H1 intermediate obtained as soon as 40s after incubation (Figure B.1A, PDBid: 6DMV).[72] In this simulation, such K^+ ion spontaneously departs from the "U" site (Figure B.4B), just after the equilibration phase, when the P_{SCI} approaches the M_A-M_B reactive center. That is, the distance between M_B and the Pro-Sp oxygen (O_{Sp}) of the scissile phosphate reaches $dM_B–O_{Sp}$ = 1.9 ± 0.13 Å. Notably, these findings align with the occupancy trend observed for K_U in the series of X-ray structures obtained at different times from incubation. Indeed, the occupancy of K_U (K_U^{Occ}) progressively reduces when the scissile phosphate gets properly positioned for catalysis and approaches M_B. In detail, at $dM_B–O_{Sp}$ = 2.6 Å, K_U^{Occ} = 0.35 (at 40s, PDBid:6DMV); at $dM_B–O_{Sp}$ = 2.2 Å, K_U^{Occ} = 0.25 (at 120s, PDBid:6DO9); and at $dM_B–O_{Sp}$ = 2.1 Å, K_U^{Occ} = 0.10 (200s, PDBid:6DOB). Thus, MD simulations agree with structural evidence, supporting that the proper positioning of the substrate favors K_U departure.

Interestingly, the *in crystallo* data also revealed that low [K^+] lead to reduced product formation in RNase H1.[72] We simulated such conditions with three more ~1μs-long MD runs at low K^+ concentration (i.e., 6mM [Mg^{2+}] and 25mM of [K^+], according to experimental conditions), with both "U" and "W" sites empty. This allowed us to investigate whether a diminished K^+ ionic strength could affect the recruitment of K^+ ions at the W site, as well as the dynamical role of E188 in this process. As a result, also at low [K^+], we observed null occupancy of the "U" site (Figures 4.2A and B.4A). Moreover, the recruitment of K_W occurs less frequently compared to the simulations at the optimal [K^+] (paragraph above), as shown by the

decreased occupancy of the "W" site by K^+ ions (Figure 4.2A). Most notably, E188 more frequently samples the $E188_{OUT}$ conformation, pointing toward the solvent, searching for a transient K^+ ion from the bulk (Figure 4.2D, lower panel). In this regard, however, it is worth noting that the lack of K_W is often compensated by the positively charged side chain of the second-shell residue K196, which at times accesses the "W" site, with almost the same frequency of a K^+ ion (Figure 4.2D, upper panel). Overall, MD runs at low [K^+] indicate that when K^+ does not access the "W" site, E188 preferably assumes the $E188_{OUT}$ state, at odds with MD runs at high [K^+], where the $E188_{IN}$ conformation is mostly observed and K^+ occupies the "W" site. These data corroborate that the binding of K_W is associated with the conformational change of E188, which could act as a K_W functional recruiter.

To further understand the influence of K^+ concentration on the E188-mediated recruitment of K_W and to estimate the energetics of such process, we performed free-energy simulations using well-tempered metadynamics at both optimal (200mM) and low (25mM) [K^+] (Figure B.7). The torsional angle E188-θ was used as a collective variable to sample multiple inner ↔ outer conformational swings of this residue in both directions. At optimal [K^+], the free-energy profile for such a conformational switch, when going from $E188_{IN}$ to $E188_{OUT}$ conformations, showed a barrier of ~6.2 kcal mol^{-1}, while at low [K^+] was ~3.7 kcal mol^{-1} (Figure B.7). Additionally, the $E188_{OUT}$ to $E188_{IN}$ conformational swing displayed a barrier of ~3.2 kcal mol^{-1} at optimal [K^+], while at low [K^+] it was ~2.4 kcal mol^{-1} (Figure B.7). Such low barriers are consistent with the conformational changes observed in our equilibrium simulations. Additionally, the slightly smaller barriers at low [K^+], compared to the barrier at high [K^+], is in line with the result from unbiased MD, overall indicating that a more frequent conformational change of E188 may thus favor the recruitment of one K^+ ion, from the bulk to the "W" site.

Finally, four MD simulations of the reactant were performed at high [Mg^{2+}] (80mM [Mg^{2+}] and 75mM [K^+], three MD runs of ~1 us each), which was experimentally shown to reduce RNA hydrolysis through the so-called attenuation effect.[72] In these simulations, despite E188 is stabilized in the $E188_{IN}$ conformation,

no K^+ ions are captured at "W" site (Figure B.8). This is because an additional Mg^{2+} metal ion locates at ~4.53 ± 0.11 Å from the P_{SCI} (Figure B.5), close to the "W" site (*viz.,* M_W) and directly binds E188, closing the accessibility of the "W" site to K^+ ions (Figure B.8E). In this way, this third Mg^{2+} ion hampers the dynamic recruitment of K_W, preventing the formation of the K^+ metal cluster for catalysis at high [Mg^{2+}]. This finding may thus represent an explanation of the attenuation effect.[203,208] Notably, these results were confirmed through multiple MD runs using three different non-bonded models for Mg^{2+} (Figure B.8). Indeed, despite Mg^{2+} parametrization can affect the M_C-E188 binding mode,[209] and the dynamics of E188 (Figure B.8A-D, lower panel), still high [Mg^{2+}] reduces the E188 inner ↔ outer conformational switch, regardless the ions' model. In turn, the conformational preference of E188 prevents the E188-mediated recruitment of K_W for catalysis.

Taken together, MD simulations of the reactant state reveal an ordered dynamics of cations, which agrees well and adds further clarification to recent hypotheses based on experimental data.[72] Accordingly, the catalytic site of RNase H1 initially contains only K_U, which is released just upon the reactants' alignment, so prone to catalysis – in general agreement with the mechanistic proposal by Samara and Yang.[72] In addition, the simulations reveal an inner ↔ outer swing of the second-shell residue E188, which is suggested here to critically favor the recruitment of one additional K^+ ion at the "W" site to form the reactant RNase H1 active site. In this way, this sequence of dynamic events, which we show to be optimally controlled by the ionic strength in solution, leads to a competent catalytic state (as in PDBid: 6DOB).[72]

4.4 Multi K^+ and Mg^{2+} Trafficking Enhances Product Formation.

We performed MD simulations of RNase H1 in the product state to evaluate if and how the ionic distribution in the vicinity of the active site is altered upon RNA hydrolysis. The starting configuration included both catalytic metals M_A-M_B and the additional K_W, as observed after 600s of incubation when the RNA hydrolysis reached a plateau (PDBid: 6DOG).[72] MD simulations have been performed in analogy with those of the reactant state (previous section), considering an initial reference, optimal ionic strength of 6mM [Mg^{2+}] and 200mM [K^+].

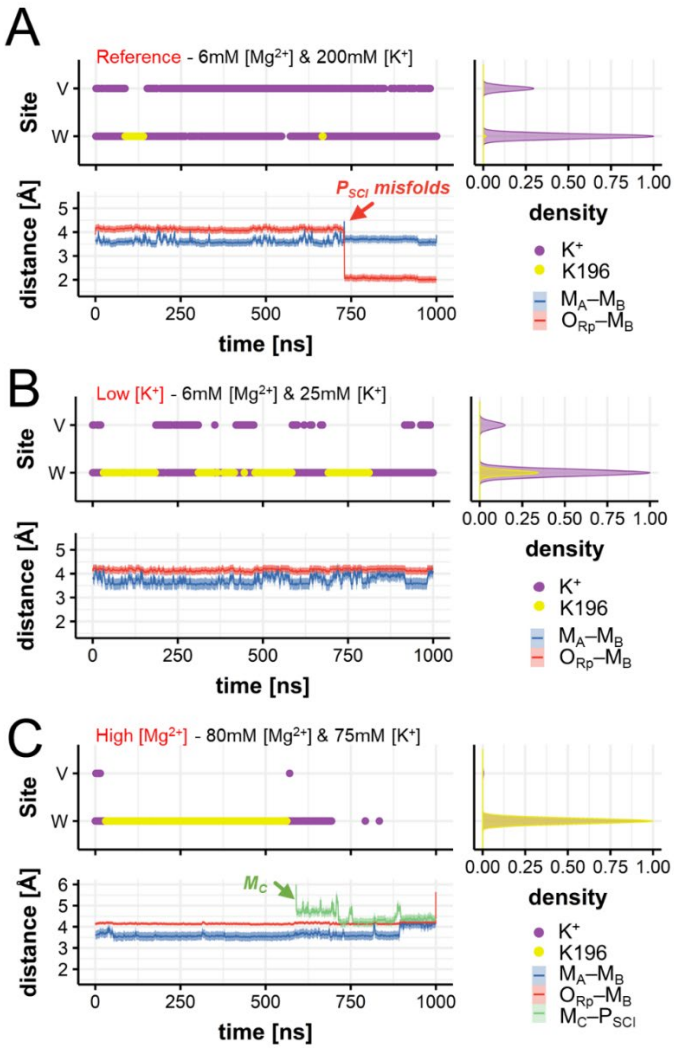


Figure 4.3: Occupancy of positive charges (viz., K^+ ions and K196) and conformational dynamics of the products. ***(A)*** *Occupancy of the W and the V sites (upper panel) by K^+ ions (purple) and by the K196 side chain (yellow); and distances (lower panel) between the catalytic ions M_A–M_B (blue trace), and between the pro-Rp oxygen of the scissile phosphate and M_B (O_{Rp}–M_B, red trace) during MD at optimal concentrations. Normalized densities are on the right.* ***(B-C)*** *Same descriptors as in (A) for MD simulations at low [K^+] (B) and at high [Mg^{2+}] (C). At low [K^+], the interaction between K196 and the scissile phosphate is more frequent (Figure 4B). At high [Mg^{2+}], M_C approaches the scissile phosphate (green trace of M_C–P_{SCI}), while the M_A–M_B distance increases, preluding the release of the products (Figure 4C).*

In these simulations, K_W stably coordinates the scissile phosphate P_{SCI} by locating at the "W" site (Figure 4.3A, upper panel and B.9A). Moreover, one additional K^+ ion transiently locates at the "V" site (i.e., in the vicinity of the catalytic site, but in a distinct position compared to the "W" and "U" sites). In this position, K_V interacts with P_{SCI} (Figure 4.4A), together with K_W. Importantly, this transient K_V ion superimposes well with the Rb^+ ion observed in the X-ray data (PDBid: 6DOX).[72] Interestingly, the simulations show that the K196 side chain can also reversibly access the "W" site, when empty.

Three additional ~1μs-long MD runs considered a lower [K^+] (*i.e.,* 6mM [Mg^{2+}] and 25mM of [K^+]). Here, a more frequent interchange between K^+ and K196 was observed (Figure 4.3B, upper panel). At lower [K^+], the "W" and "V" sites are often not occupied by K^+, compared to their occupancy at higher [K^+]. Consequently, K196 is more often found to interact directly with the phosphate leaving group (Figure 4.4B). This results in a stable configuration of the P_{SCI} (Figure 4.3B, lower panel). On the other hand, when K196 more rarely accesses the "W" site, a partial misfold of the P_{SCI} is observed (Figure 4.3A, lower panel). These findings, which are well reproduced in the simulation replicas (Figure B.9), suggest that K196 could be quite relevant for the precise accommodation and preliminary departure of the solvent-exposed leaving nucleotides. Indeed, K196 can, at times, interchange with K_W and K_V.

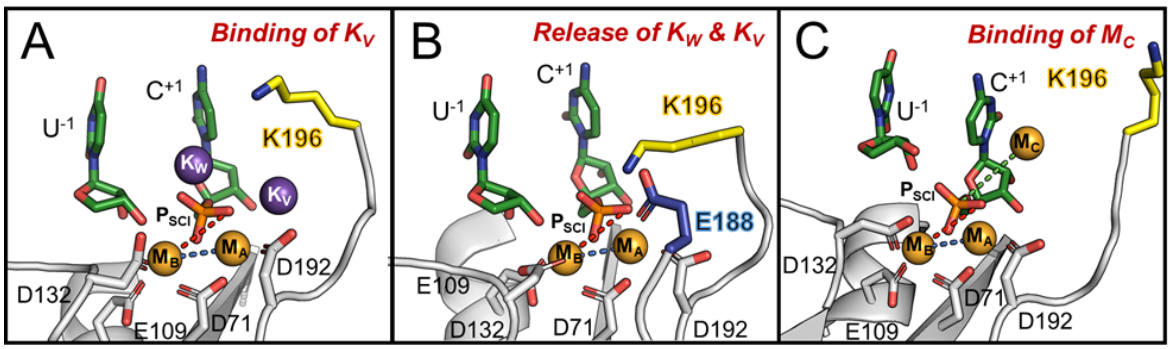


***Figure 4.4: Snapshots of the RNase H1 active site upon catalysis**. **(A)** In the product state, K_W is stably coordinated at the active site, while K_V is only transiently observed. **(B)** K_W and K_V can exchange their position with K196, which directly coordinates the scissile phosphate preventing its misfolding. **(C)** The third additional M_C binds the scissile phosphate and promotes the increase of the distance between the two catalytic M_A–M_B ions for product release.*

As a result, the correct positioning of the phosphate of the leaving group for departure is maintained (as observed in the PDBid: 6DOG, captured right before leaving group departure).[72] Notably, the positively-charged side chain of K196 locates at the W-site (at ~2.7 Å from P_{SCI}) to H-bond the leaving group only if both K_W and K_V ions have departed from the catalytic site (see Figure B.9, where the only presence of K_W still preserves the hydrolyzed leaving group in place, at the two-metal-ion center).

Molecular simulations of the products have also been performed at increased [Mg^{2+}] (*i.e.,* 80mM of Mg^{2+} and 75mM of K^+, five MD runs of ~1μs each) in line with the experimental conditions. These simulations show that K196 can be replaced by a transient third Mg^{2+} ion, which locates at the "C" site (*viz*., M_C, Figures 4.4C). The approach of M_C consistently occurred in MD replicas, conducted using three different non-bonded models for Mg^{2+} (Figures 4.3C, lower panel and B.9C-D). This finding matches the structural data obtained at high divalent metals concentration (≥ 16mM), which identified such M_C in a very similar position, close to the leaving group (PDBid: 6DPD).[72] Moreover, upon binding of M_C, the internuclear M_A–M_B distance is increased (from 3.72 ± 0.11 Å to ~4.11 ± 0.07 Å, Figures 4.3C, lower panel), reflecting the destabilization of the active site's architecture to prelude the products' exit. Thus, such a transient third ion seems recruited in the product state to favor the leaving group departure, as seen in other nucleic acid processing enzymes.[76,210,211]

It is important to recall that MD of the reactants at high [Mg^{2+}] have also shown that a third Mg^{2+} ion can locate in the vicinity of the catalytic site (Figures S5 and B.8E). In that case, however, the third Mg^{2+} ion locates close to the "W" site and binds E188, hampering the dynamic recruitment of K_W and preventing the formation of the K^+ metal cluster for catalysis. Hence, taken together, the simulations at high [Mg^{2+}] support the idea that a third M_C ion could be recruited in the product state to favor the leaving group departure.[76,80,211]

These results are also in line with the structure of the K196A mutant (Figure B.2B; PDBid: 6DPM).[72] Here, M_C is missing, suggesting that the K196A mutation

would somehow impair product stabilization, disfavoring M_C recruitment for leaving group departure. To test this hypothesis, we carried out two additional MD runs of the K196A mutant, at both low K^+ (i.e., 6mM [Mg^{2+}] and 25mM [K^+]) and high Mg^{2+} (i.e., 80mM [Mg^{2+}] and 75mM [K^+]) concentrations (~300 ns and ~1 μs, respectively). These simulations confirm that the K196A mutant does not lead to M_C recruitment, further sustaining the crucial role of K196 for product stabilization and M_C binding (Figure B.11).

In summary, molecular simulations of the product state indicate that a controlled metal trafficking at the catalytic site plays a critical role in RNA hydrolysis and product release. Upon RNA hydrolysis, the potassium metal cluster breaks, while K196 contributes to product stabilization before the third divalent M_C ion can further favor product release (*vide infra*).

4.5 Third Mg^{2+} Trafficking for Product Release.

In the products, upon binding of M_C, the internuclear M_A–M_B distance increased (Figure 4.3C), suggesting that the active site might be prone to release the reaction products, as observed in other two-metal-ion enzymes.[76,80,210] To further investigate how the unbinding of the third M_C may facilitate product exit, we performed well-tempered metadynamics simulations of the product state, which comprehends an extended metal cluster formed by M_A-M_B, M_C, and K_W. The gaussian-shaped potential was deposited using two collective variables (CV_S): *(i)* the coordination number of the M_C ion, which traced the number of water molecules in its first-shell; and *(ii)* the distance between M_C and the scissile phosphate (P_{SCI}–M_C).

Metadynamics identifies two main metastable states (*viz.,* A and B, shown in Figure 4.5, upper panel), which correspond to the two main minima of the computed free-energy surface (FES, Figure 4.5, lower panel).

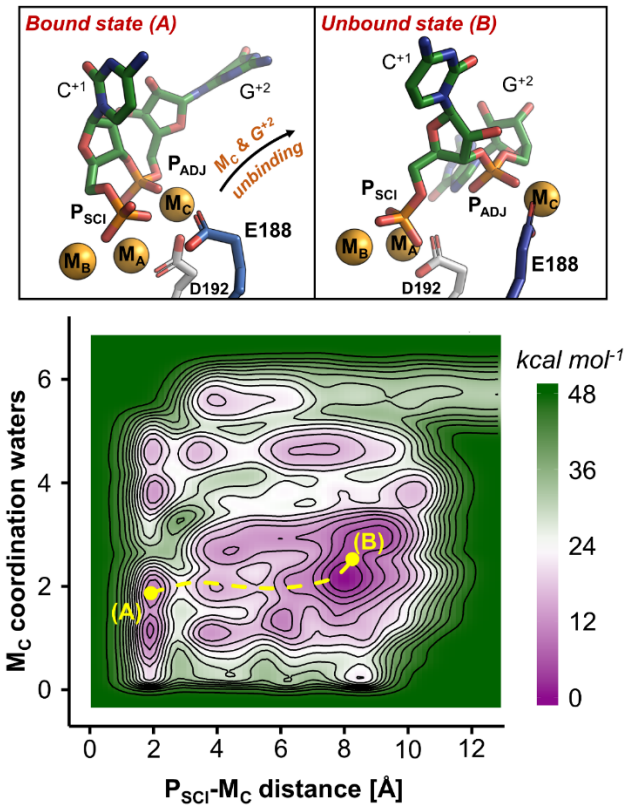


Figure 4.5: Role of M_C and E188 in the product release. *Upper panel: intermediates states (A and B) of the RNase H1 after phosphodiester bond cleavage, identified via free-energy simulations. In state A, M_C is coordinated by E188. Upon rotation of this residue (state B), M_C is released with the departure of the adjacent phosphate (P_{ADJ}). Lower panel: free-energy surface describing the product release mechanism, indicating the two intermediate A and B. The energy scale is in kcal mol^{-1}, contours are traced every 3 kcal mol^{-1}.*

In state A, M_C and bridges the scissile and adjacent phosphates (P_{SCI} and P_{ADJ}), while the catalytic D192 and the second-shell E188 establish a monodentate interaction with M_C. In state B, M_C loses the coordination with D192 and gets bidentate coordination with the E188 carboxylic moiety. At this point, the rotation of E188 shuttles the M_C metal out of the active site, together with P_{ADJ}, now located ~9 Å away from M_A-M_B. Notably, the base G^{+2} and the active site are completely exposed to the bulk, and the distance between the catalytic M_A-M_B, reaches ~4.7 Å, indicating the opening of the active site for product release. The free-energy barrier estimated on the minimum free-energy path for this process is ~15 kcal mol^{-1}, with the unbound state B favored of ~6 kcal mol^{-1} with respect to the bound state A.

We also investigated the role of E188 in the product state via additional metadynamics runs of the E188A mutant (Figure B.2A). Remarkably, in these simulations, M_C preserves the coordination with the adjacent phosphate P_{ADJ} and D192 (state A'; Figure B.12). However, the absence of the E188 carboxylic moiety, which in the wt RNase H1 functions as an anchoring point for the exit of M_C (Figure 4.5), strongly disfavors the release mechanism (state B', Figure B.12). Indeed, in the E188A mutant, M_C needs first to be partially hydrated before being released together with the products, as shown by minimum free energy path connecting the state A' with B' (Figure B.12). Importantly, in this case, the estimated energetic barrier for M_C unbinding and leaving group departure is ~35 kcal mol^{-1} (Figure B.12), more than twice compared to the one obtained for the wt RNase H1. Moreover, in the absence of E188, the bound state A' results favored the unbound state B' of ~18 kcal mol^{-1}, as opposed to what was observed in the wt RNase H1.

These results suggest that M_C acts in concert with the second-shell residue E188 to promote the release of the reaction products.

4.5 Discussion

Controlled cations trafficking and enzymatic strategy in RNase H1.

The RNase H1 enzyme is a fundamental endonuclease involved in DNA replication.[61] Recently, time-resolved X-ray crystallography revealed the transient presence of multiple K^+ and Mg^{2+} ions at the active site of RNase H1, at different stages of catalysis.[72] Here, extensive molecular dynamics (MD) and free-energy simulations (over a sampling of >27 μs) characterized the functional dynamics of this newly determined heteronuclear metal cluster before and upon enzymatic catalysis. Our simulations capture an ordered coordination of motion of transient metal ions and second-shell residues for catalysis. We show that four biding sites in the vicinity of the RNase H1 catalytic core, namely the "U", "W", "V", and the "C" sites (Figure 4.1), are neatly but only intermittently occupied by K^+ and Mg^{2+} metals during catalysis, in agreement with experiments.[72]

We observed that, before RNA hydrolysis (*i.e.,* in the reactant state), a K^+ ions cluster is dynamically formed to favor substrate alignment and, possibly, activation for catalysis. At first, one K^+ ion transiently locates at the "U" site (*viz.,* K_U) and assists in forming the catalytically competent fold of the RNase H1 active site. This K_U is spontaneously released into the bulk only after the proper coordination between the scissile phosphate and the two catalytic M_A-M_B ions (Figure B.4). At this point, the inner ↔ outer conformational swing of the second-shell residue E188 favors the recruitment and the precise positioning of a second K^+ ion at the "W" site (*viz.,* K_W, Figure 4.2), as also observed for enzymes such as human Exo1.[76] In RNase H1, this K_W directly contacts the scissile phosphate to favor the correct positioning of the reactants, prompting catalysis. Such K_W can depolarize the phosphodiester bond, thus promoting the catalytic activation of the reactants.[59,77]

Importantly, we found that this chain of dynamic (un)binding events of K^+ ions is significantly affected by the reaction buffer's ionic strength. In fact, in agreement with *in crystallo* data showing that catalysis is reduced at low concentrations of K^+ (*i.e.,* $[K^+]$),[72] MD simulations performed at low $[K^+]$ show that the E188-mediated

recruitment of K_W occurs less frequently (Figure 4.2D). Accordingly, the inner ↔ outer conformational swing of E188 is energetically favored at low [K^+] with a barrier of ~3.7 kcal mol^{-1} *vs.* ~6.2 kcal mol^{-1} higher [K^+] (Figure B.7). The crucial role observed here for the E188-mediated recruitment of K_W is also sustained by both biochemical and structural evidence. Indeed, alanine mutation of E188 reduces the catalytic efficiency of RNase H1.[72] As well, X-ray structures have shown that such mutation displays no K_W at the active site (Figure B.2A), at odds with the wt RNase H1.[72] In line with this evidence, our MD simulations show that the recruitment of K_W is dramatically hampered when E188 is restrained in its "out" conformation (Figure B.6C), as well as when E188 is mutated into an alanine (Figure 4.2A). These findings support the E188-mediated mechanism for K_W recruitment, which ensures the reactants' proper formation. Notably, these results further corroborate the recruitment role of second-shell residues for two-metal-ion catalysis, such as recently reported for human Exo1, ExoG, Exo-λ and other 5'-exonucleases.[76]

Upon RNA hydrolysis, we observe an additional K^+ ion intermittently located at the so-called "V" site, in the vicinity of the catalytic core. This K_V ion coordinates the scissile phosphate P_{SCI}, together with K_W (Figures 4.3A and 4.4A), as also shown by *in crystallo* data (PDBid: 6DOX).[72] We note that these K_W/K_V ions can interchange their distinct position with the second-shell residue K196, near P_{SCI} (Figure 4.3B and 4.4B). Indeed, such residues favor the stabilization of the P_{SCI}-centered reaction products. The functional role of K196 in products' stabilization is also sustained by the X-ray data of the catalytically impaired K196A mutant of RNase H1 (PDBid: 6DPM),[72] in which the products are misfolded and no third Mg^{2+} ion (M_C) is resolved. Accordingly, MD simulations of this mutant show that the absence of this lysine induces the product's misfolding and hampers the binding of M_C (Figure B.11). On the other hand, in the wt RNase H1, leaving group departure occurs when the K^+ ions and K196 are ultimately displaced by a third Mg^{2+}, which spontaneously sits at the "C" site (*viz.,* M_C) and binds P_{SCI} (Figure 4.4C). At this point, the internuclear M_A–M_B distance is increased (from 3.72 ± 0.11 Å to ~4.11 ±

0.07 Å), as the overall active-site structure evolves towards the products (Figures 4.3C and 4.4C). This is in line with the X-ray structure showing that the distance between the M_A–M_B increases with product release (PDBid: 2G8V), from 3.7 Å to 4.8 Å.[211] From this point, the leaving group departure is expected to occur.

To further investigate the products' release, metadynamics simulations show that M_C is involved in a step-wise product release mechanism. In detail, M_C initially bridges together P_{SCI} and its adjacent phosphate, P_{ADJ} (Figure 4.5, state A). In this conformation, M_C functions as a positively-charged anchor for E188. At this point, the swing of E188 occurs with the simultaneous departure of the movable M_C, together with the reaction products (Figure 4.5, state B), as also observed for other metalloenzymes.[78,79,212–214] This process occurs with an estimated energy barrier of ~15 kcal mol^{-1} (Figure 4.5). Metadynamics simulations of the E188A mutant also confirm such a functional role of E188 for product release. In this case, the estimated energetic barrier for the departure of the leaving group was ~35 kcal mol^{-1} (Figure B.12), which is more than 2-fold higher than that in the wt RNaseH1. In this regard, we note that in the human RNase H1, E188 corresponds to H264. Based on both structural and biochemical analysis,[62,203,211] this histidine residue has been proposed to favor, together with the mobile C-terminal loop, the release of the products by altering the coordination of M_A ion (Figure B.14).

Such a dynamically controlled trafficking of positive charges at RNase H1 active site is finely dependent on the specific ion concentration, as displayed in the X-ray experiments. In fact, MD of the reactant performed at higher [Mg^{2+}] showed a third Mg^{2+} ion that spontaneously locates at the "W" site (*viz.,* M_W differently located than M_C, Figure B.5), closing the accessibility of the "W" site to K^+ ions. In this location, M_W interacts with E188 (Figure B.8E), hampering the E188-mediated recruitment of K_W and preventing the formation of the K^+ metal cluster for catalysis. This mechanistic finding is of particular interest in light of the so-called "attenuation effect".[203,208] Indeed, the precise location of this third Mg^{2+} at the "W" site precludes the formation of a catalytic metal cluster, thereby representing an explanation for the reduced RNA hydrolysis at high [Mg^{2+}].

In summary, within the limits of the employed computational approach, our simulations provide a dynamic characterization of metal ion trafficking and its interplay with surrounding residues prior and upon catalysis. These results qualitatively describe an enzymatic strategy in which an extended and heteronuclear cations cluster (K^+ and Mg^{2+} ions) is dynamically formed and disrupted, with the aid of second-shell residues. This mechanism prompts the processing of RNA:DNA hybrids in RNase H1 (Figure 4.6).

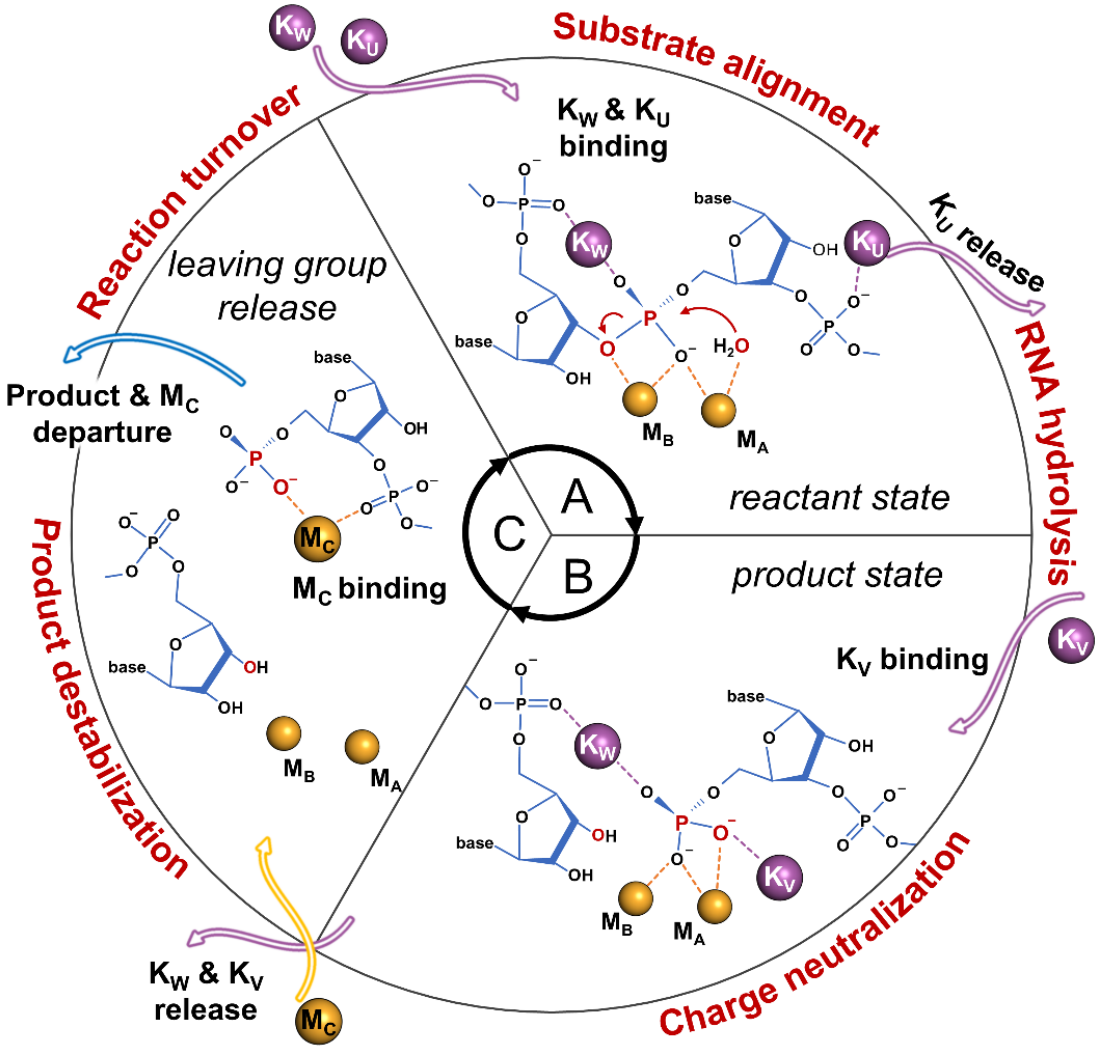


Figure 4.6: Controlled cations trafficking favors RNase H1 catalysis. ***(A)*** *In the reactant state, K_W and K_U prompt the alignment of the substrate. At this point, K_U is released and the RNA hydrolysis can occur.* ***(B)*** *Upon hydrolysis (i.e., product state), the negatively charged product is neutralized by K_W and another K_V ion.* ***(C)*** *The ultimate exchange of these two K^+ ions with a third Mg^{2+} ion (viz. M_C) promotes the destabilization of the products, which are then released together with M_C to induce the reaction turnover and the continuation of catalysis.*

That is, the RNase H1 cation trafficking starts with the binding of K_W and K_U, which prompt the alignment of the substrate (Figure 4.6A). At this point, K_U is released, and the RNA hydrolysis can occur. Upon hydrolysis (*i.e.,* product state, Figure 4.6B), K_W is assisted by another K_V in neutralizing the negatively charged product. Finally, the exchange of these two K^+ ions with a third Mg^{2+} ion (*viz.,* M_C) contributes to the product destabilization, leading to the leaving group's release together with M_C (Figure 4.6C) that precedes the reaction turnover and the continuation of catalysis. Ab-initio simulations will now be needed to evaluate the mechanistic and energetic implications of trafficking and strategic localization of such cations for the chemical step of RNA hydrolysis at the reaction center. Nonetheless, our classical simulations indicate that the controlled trafficking of cations, observed over a multi-microsecond sampling, is crucial to promote the formation of catalytic states that just precede and follow the enzymatic reaction for substrate hydrolysis.

Similarity of binding and trafficking of metals in other nucleic acid-processing metalloenzymes.

It has been recently shown that conserved and positively-charged residues are strategically located at the active core of, *e.g.,* DNA or RNA polymerases and nucleases, at conserved sites, expanding the two-metal-ion functional architecture of these enzymes.[59,73,210] On this basis, we used structural analyses to investigate whether the mechanistic insights herein provided may be shared among other nucleic acid processing enzymes. We found that the binding sites "U", "W", and "V", transiently occupied by K^+ ions in RNase H1, overlap well with second-shell positions occupied by positively-charged residues in other enzymes. In fact, the RNase H1 K_W and K_U/K_V, respectively, well superimpose with, *e.g.,* R92 and K85/G2(NH) in human exonuclease 1 (PDBid: 5V06, Figure B.14A),[192] or with H539 and R557 in HIV-1 reverse transcriptase (PDBid: 6BSH, Figure B.14B),[215] as well as with R100 and K93/G2(NH) in the endonuclease FEN1 (PDBid: 5UM9, Figure B.14C).[216]

Structural analysis of FEN-1 bringing an alanine mutation of R100 (PDBid: 5KSE) revealed that, in the absence of R100, an additional metal ion is captured in the same putative "W" site before the catalysis.[216] Notable similarities also arise from our recent studies investigating the structure and the catalysis of the CRISPR-Cas9 genome editing system. We found that the K970 and R972 residues within the catalytic RuvC domain also transiently engage the so-called "W" and "U/V" binding sites at specific catalytic steps (Figure B.14D).[217,218] Together, these findings support that the trafficking of positive charges at conserved regions of several metalloenzymes is strategically designed and controlled to favor catalysis. In this context, we note that only one positively-charged residue is present in the vicinity of the RNase H1 active site (*i.e.,* K196), contrary to what is found in several other two-metal-ion enzymes.[73,75] This observation may be coupled to the highly solvent-exposed localization of the RNase H1 active site when compared to those mentioned above and second-shell basic residues-rich enzymes (see Figure B.15). This may be why the RNase H1 active site shows such prominent trafficking of multi-cations, in the lack of multiple strategically located positively-charged residues in the surrounding of the enzymatic reaction center. That is, multi-cations binding at the catalytic site may counterbalance the lack of multiple second-shell positively-charged residues. As a matter of fact, this would not represent the first example of its kind. Indeed, a similarly extended, heteronuclear cluster of movable cations has been captured at the catalytic site of the group II introns ribozymes (two K^+, K_1-K_2; and two Mg^{2+} ions, M_1-M_2), RNA-based enzymes that perform RNA hydrolysis in the absence of (positively-charged) aminoacids.[44,138] In group II introns, the position of these ions has been shown to be conserved for catalysis, in analogy with positively-charged residues in proteins.[73,74] Importantly, we have recently demonstrated that such a structured metal cluster needs to be dynamically formed and broken to guarantee the proper formation of the introns' catalytic core and favor the release of the reaction products, which matches well what we uncover herein for RNase H1.[81] Moreover, a recent high-resolution cryo-EM map of the human spliceosome,[219] which is structurally and chemically related to the group II introns,

has revealed the presence of the same conserved heteronuclear metal cluster at its active site, in strike agreement with a model system proposed years before.[73,138] Together, this evidence suggests that the controlled trafficking of cations may indeed be a common enzymatic strategy to trigger and assist phosphoryl transfer reactions in living cells, which has been evolutionary conserved and optimized in proteins. In light of these findings and observations, this study shows that finely regulated and controlled trafficking of multiple cations is an essential feature for the enzymatic-mediated hydrolysis of nucleic acids.

4.5 Conclusions

In conclusion, we have used extended atomistic molecular dynamics and free-energy simulations to investigate the cations trafficking recently observed *in crystallo* reaction intermediates during catalysis of the RNA ribonuclease H1 (RNase H1) enzyme.[72] Our results illustrate a finely regulated trafficking of multiple and diverse cations with functional implications at the reaction center of RNase H1. In agreement with the recent mechanistic hypotheses and experimental data,[72] our findings corroborate and sensibly expand the recent notion of an extended two-metal-ion architecture for nucleic acid-processing enzymes.[73] Such controlled metal ion trafficking in the vicinity of the catalytic reaction center seems critically designed to aid catalysis in RNase H1, and possibly other similar nucleic-acid processing enzymes. Taken together, these findings may encourage further investigations related to enzyme engineering and drug discovery.

4.5 Methods

Here, the catalytic intermediates of the RNase H1 captured by time-resolved X-ray have been object of extensive molecular dynamics (MD) and enhanced sampling free-energy simulations, with the goal of characterizing the dynamics of cation trafficking prior to and upon RNA hydrolysis. Cation trafficking has also been investigated considering the second shell E188A and K196A mutations and three different ionic strengths of the solution. Overall, we collected a total of >27 μs of MD multi-replica simulations.

Structural models.

To perform MD simulations, six systems were modeled based on the recently time-resolved X-ray structures of *Bacillus Halodurans* RNase H1 in complex with an RNA/DNA hybrid construct.[72] First, the reactant state of the wild type (wt) RNase H1, as obtained from the PDBid: 6DOG (occupancy C), was considered (Figure 4.1B). Here, only the crystallized metals M_A-M_B were included in the starting configuration to verify spontaneous binding events of K^+ ions at the "U", and "W" site. Second, the reactant state of the wt RNase H1 included the ions M_A-M_B and K_U, as in PDBid: 6DMV, to verify the stability of the K_U bound state. Third, the reactant state of the E188A mutant included M_A-M_B, as in PDBid: 6DPO, to verify the effect of the mutation for the recruitment of K_W. Fourth, we considered the product state of the wt RNAse H1, modeled upon the PDBid: 6DOG (occupancy B), including the crystalized ions M_A, M_B, and K_W (Figure 4.1C). Fifth, the product state of the K196A mutant, built on the PDBid: 6DPM, included the crystallized metals M_A, M_B, and K_W (Figure B.2B). Sixth, the product state of the E188A mutant, as obtained from the PDBid: 6DPO, included the crystallized metals M_A, M_B, and M_C (Figure B.2A). All the systems were solvated with a layer of water molecules of 15Å. According to crystallization experiments, each system was simulated at multiple ionic strengths to characterize the effect of the concentrations of the ions on their dynamics. Specifically, Mg^{2+}, K^+ and Cl^- ions were added into the

simulation boxes to match three different ionic strengths: *(i)* reference, optimal concentrations, 6mM [Mg^{2+}] and 200mM [K^{+}]; *(ii)* low K^{+} concentrations, 6mM [Mg^{2+}] and 25mM [K^{+}]; *(iii)* high Mg^{2+} concentration, 80mM [Mg^{2+}] and 75mM [K^{+}].

Molecular simulations.

The RNase H1 protein was parametrized with the AMBER force-field ff14SB,[220] while the RNA and the DNA of the hybrid were parametrized with AMBER force-field RNA.OL3[97,99,221,222] and DNA.OL15.[223] For the K^{+} metal ions, the parameters of Joung and Cheatham were used.[172,224] The Mg^{2+} ions were systematically modelled according to both 12-6 (Aqvist[225] and Allnér[226]) and 12-6-4 (Panteva[227]) non-bonded fixed point charge models, to verify that the observed results were independent from the chosen set of parameters.[209,228] Last, the two catalytic Mg^{2+} ions M_A-M_B were modelled following the atom-in-molecule charges' partitioning scheme,[174] to consider also charge transfer between the ions and the first-shell coordination residues. The TIP3P model was used for water molecule.[169] AMBER[229,230] was used to perform Langevin[231] MD simulations, using an integration time step of 2 fs. The temperature was set at 300K using a collision frequency $\gamma = 1$ per picosecond, while the constant 1 atm pressure was controlled through Berendsen barostat[232] with a relaxation time of 2 ps.

We used the same simulation protocol for all systems. First, we performed an energy minimization to relax the water molecules. At this point, positional restraints of 300 kcal mol·Å^2 were imposed to all the heavy atoms of the systems, including the metals. Then, we used a series of NVT and NPT simulations to smoothly thermalize the systems and remove the positional restraints. First, the systems were heated up with one NVT simulations of ~600 ps, using the same positional restraints as used during the energy minimization. Subsequently, the positional restraints were progressively halved with a series of two NVT simulations of ~200 ps each. Then, two additional simulations in the isothermal-isobaric ensemble (NPT) of ~200 ps were performed by further halving the positional restraints, while a third NPT run of ~2 ns was performed without any positional restraints to relax the density of the

systems to ~1.01 g cm^{-3}. Last, we performed multiple-replicas production runs to collect overall ~20 μs of simulation time. Specifically, we collected: ~9 μs for the pre-reactive state of the wild type, 9 replicas; ~9 μs for the post-reactive state of the wild type, 9 replicas; ~1.3 μs for the post-reactive state of the K196A mutants, 2 replicas.

Metadynamics.

Two independent metadynamics simulations were performed to characterize the conformational change of E188 in the pre-reactive state. To ensure the convergence of the calculations, we used the well-tempered variant,[113] setting the temperature at 300K and the bias factor at 8. The gaussian-shaped potential (height = 1.2 kJ mol^{-1}, width = 0.35 *rad.*) was added with a frequency of 0.5 ps on the torsional angle E188-θ (between the E188 atoms N-Cα-Cβ-Cδ). Overall, we collected ~400 ns at both high and low K^+ concentrations.

Additional well-tempered metadynamics simulations were performed to elucidate the release of the M_C metal and the reaction products. The post-reactive state of both the wild type and the E188A mutant was used as starting configurations for the two independent simulations. Here, the temperature was set to 300K and the bias factor was set to 15. The coordination number of the M_C, which defines the number of water molecules in the first coordination shell, and the distance between the scissile phosphate and the M_C were used as collective variables (CV1 and CV2, respectively). The gaussian-shaped potential was added with a frequency of 0.2 ps on both the CV1 (height = 0.3 kJ mol^{-1}, width = 0.1) and CV2 (height = 0.3 kJ mol^{-1}, width = 0.5 Å).

Chapter 5. Finding the Ion in the RNA-stack: Can Computational Models Accurately Predict Key Functional Elements in Large Macromolecular Complexes?

5.1 Abstract

This viewpoint discusses the predictive power and impact of computational analyses and simulations to gain prospective, experimentally-supported mechanistic insights into complex biological systems. Remarkably, two newly-resolved cryoEM structures have confirmed the previous, and independent, prediction of the precise localization and dynamics of key catalytic ions in megadalton-large spliceosomal complexes. This outstanding outcome endorses a prominent synergy of computational and experimental methods in the prospective exploration of such large multicomponent biosystems.

5.2 Discussion

The challenge of computationally predicting and refining the 3D structure of biological macromolecules has been highly appealing over the last decades. Significantly, such activity can accelerate impactful discoveries in life, environmental and pharmacological sciences. As a matter of fact, structure predictions by new artificial intelligence-driven algorithms have now achieved unprecedented accuracy, at least for single-subunit proteins. However, predicting 3D structures is often not sufficient to provide mechanistic insights for dynamic biological systems. Computational tools like molecular dynamics (MD) simulations are therefore powerfully used to interpret experimental data, and generate integrative models of biological structures or investigate their complex function, dynamics, and even chemical reactions.[69,233]

On the other hand, the power of prospective mechanistic insights from computational studies is still often underestimated. Indeed, it is particularly challenging to detail the functional mechanism of very large macromolecular complexes, such as transcriptional, translational, splicing, or protein/RNA degradation machineries. At a time when technological advances make these multi-subunit protein and RNA-protein complexes experimentally tractable, reliably predicting their structures and dynamics can be crucial in rationally guiding and accelerating their characterization and, ultimately, their modulation. In this context, what is the best approach for reliable mechanistic predictions into such large macromolecular systems? How to generate such predictions, and ensure they are valued and exploited by experimentalists? Here, we address these questions with a recent example that shows how the integration of computational and experimental data helped provide key predictive structural and mechanistic insights into vital, ubiquitous and medically-relevant splicing machineries.

Splicing is a two-step biological reaction whereby introns are excised from precursor RNA molecules and exons are ligated into mature functional protein-

coding or non-coding transcripts. In detail, splicing chemistry consists of two sequential scissions of phosphodiester bonds at the 5'- and 3'-intron/exon junctions, respectively. Both reactions, which are S_N2-like nucleophilic additions, occur within an active site comprising two divalent metal ions that coordinate and activate the reacting residues (Figure 5.1A).[37,234] This two-metal-ion reaction chemistry is identical to that of other nucleic-acid-processing protein enzymes, such as endo/exonucleases and polymerases. All these complex enzymes catalyze the scission or synthesis of phosphodiester bonds in DNA/RNA, respectively. The ubiquitous nature of such metal-aided structural architecture of the catalytic site is corroborated by the large therapeutic spectrum of drugs that target two-metal-ion enzymes and are thus broadly used to treat cancers and viral infections.[71]

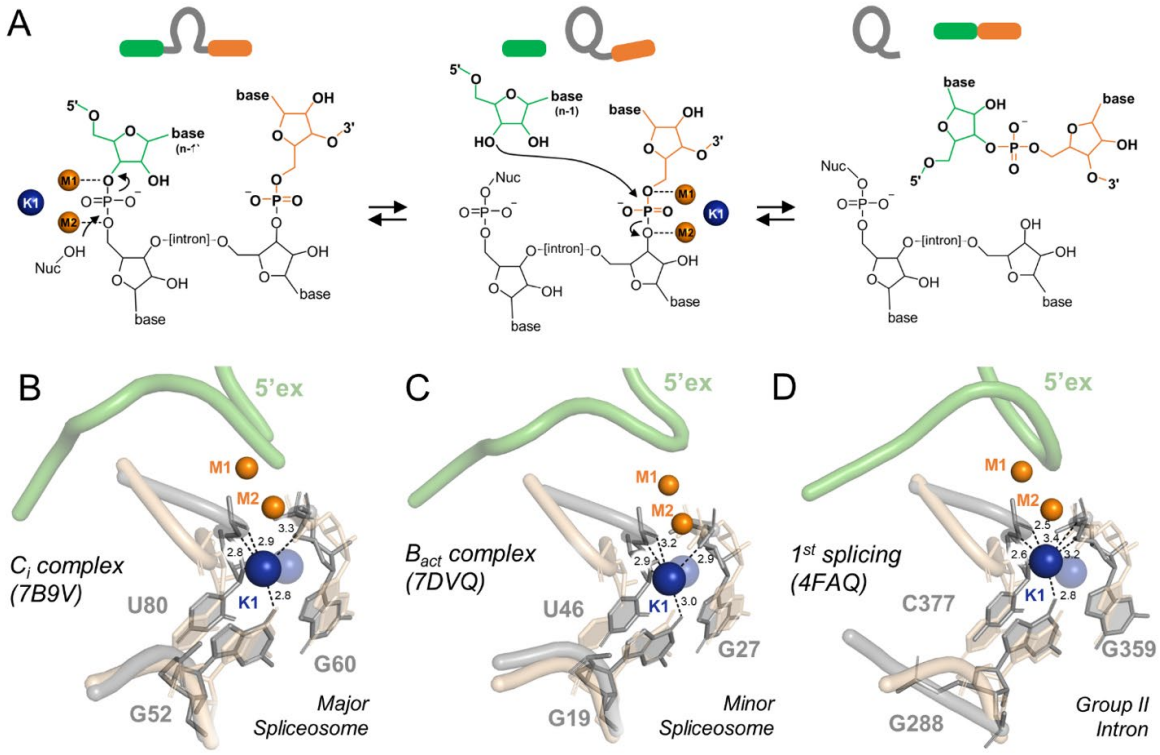


Figure 5.1: K1 is required for both steps of forward splicing. ***A)*** *Schematic representation of the two steps of the splicing reactions. The 5'- and the 3'-exon are highlighted in green and orange, respectively. Catalytic ions M1-M2 (orange) and K1 (blue) are shown as spheres. Black arrows indicate nucleophilic attacks, while "Nuc" indicates the reaction nucleophile.* ***B-D)*** *Structural superposition of the K1 binding model over the major spliceosome (C_i complex, panel* ***B****), the minor spliceosome (B_{act} complex, panel* ***C****), and the group II intron (panel* ***D****). Catalytic ions M1-M2 (orange) and K1 (blue) are shown as spheres. The nucleotides coordinating K1 (grey) are represented as sticks, while the 5'-exon (green) is shown as cartoon. The predicted K1 binding model (beige) is depicted in semitransparent representation. The predicted K1 ion (blue) is depicted as a semitransparent blue sphere. Black dotted lines represent K1 coordination distances in angstrom.*

In eukaryotes, splicing is mostly catalyzed by megadalton-large ribonucleoproteins called the spliceosomes, which exist in two isoforms: The major spliceosome, which processes 99% of all transcripts in humans, is formed by the U1/U2/U3/U4/U5/U6 snRNPs; and the minor spliceosome, which processes the remaining 1% of human transcripts, is formed by the U11/U12/U4atac/U5/U6atac snRNPs.[235]

Until recently, mechanistic insights into spliceosomal complexes were mainly based on phylogenetic analysis, chemical-mapping, and enzymatic and biochemical assays.[236] Limitations in obtaining more detailed mechanistic insights were primarily due to the large dimensions, heterogeneous biopolymeric composition (6 large RNAs and hundreds of protein subunits), and dynamic assembly of the spliceosome along the catalytic cycle. In fact, high-resolution structural insights have been initially obtained only indirectly from crystallographic work on the so-called group II self-splicing introns,[34,35,37,130] which are the evolutionary ancestors of the spliceosomes. Nonetheless, with 38 new cryoEM 3D structures produced in the last five years, we have now reached an increasingly good understanding of the dynamic assembly and remodeling of spliceosomal protein and RNA subunits throughout the splicing cycle.[236] Despite this progress, so far, the resolution of the available structures (>3.0 Å, most structures at >4.0 Å) had remained a limiting factor in defining the catalytic site's atomic details.

Among other properties, the dependence of the spliceosome on potassium ions, which had been functionally reported already since the early 1980s,[157] had remained unexplained at the molecular level. A nearly 40-year-long research effort to explain this enzymatic observation was crucially informed and guided by structure-function studies on the group II introns and by closely-related computational analyses of various classes of nucleic-acid-processing protein and RNA enzymes (see below). These analyses had led to the prediction that a specific potassium ion (named K1) could be localized near the catalytic site of the spliceosome and thus contribute to catalysis through precise structural and functional interactions (Figure 5.2).[73]

K1 was first identified in the active site of the group IIC intron from the bacterium *Oceanobacillus iheyensis* in 2012 by crystallizing this ribozyme in the presence of different mono- and divalent metal ions and by performing anomalous diffraction X-ray studies.[35,36,133,134] In the *O. iheyensis* intron structures, one of which was solved at 2.7 Å resolution, K1 is coordinated by active site residues G288, G359, and C377 (Figure 5.1D).[35–37] At that time, it was questioned whether K1 was a conserved active site element or idiosyncratic only to the *O. iheyensis* intron, and as a consequence this ion was not modeled in other lower-resolution structures of homologous group II introns.[130]

Subsequent systematic integration of evolutionary and structural alignments, electrostatic potential calculations, and MD simulations made it possible to appreciate the ubiquitous presence of positively-charged residues surrounding the active sites of several and diverse nucleic-acid-processing enzymes.[73] These basic residues were structurally and functionally analogous to the group II intron K1, suggesting evident mechanistic similarities in enzymes where K1, or K1-like residues, were likely to provide a key functional contribution for nucleic acid processing. Indeed, in these enzymes, K1-like residues act in synergy with the previously-recognized two-divalent-metal-ion core and specifically contribute to shape the electrostatics of the active site, modulating the orientation and dynamics of key reacting residues for catalysis. For instance, microsecond-long equilibrium MD simulations have shown that the absence of K1-like residues in polymerase-η induces a distortion in the reaction substrates and disrupts the Michaelis-Menten complex, thus hampering catalysis.[73] Such comparative analysis of group II introns, exo/endonucleases, and polymerases offered solid and accurate bases to predict the presence, identity and exact location of K1 also in the spliceosome (Figure 5.1B-D).[73] In more detail, through our analysis and simulations, we predicted K1 to be located at a site coordinated by G52, G60 and U80, which are the evolutionarily and structurally homologous residues to the group II intron G288, G359, and C377, in the structure of the major spliceosomal C complex. At *that* time, this was the most reliable structure to model the location of K1 (PDB id: 5LJ3; Figure 5.1B-D).

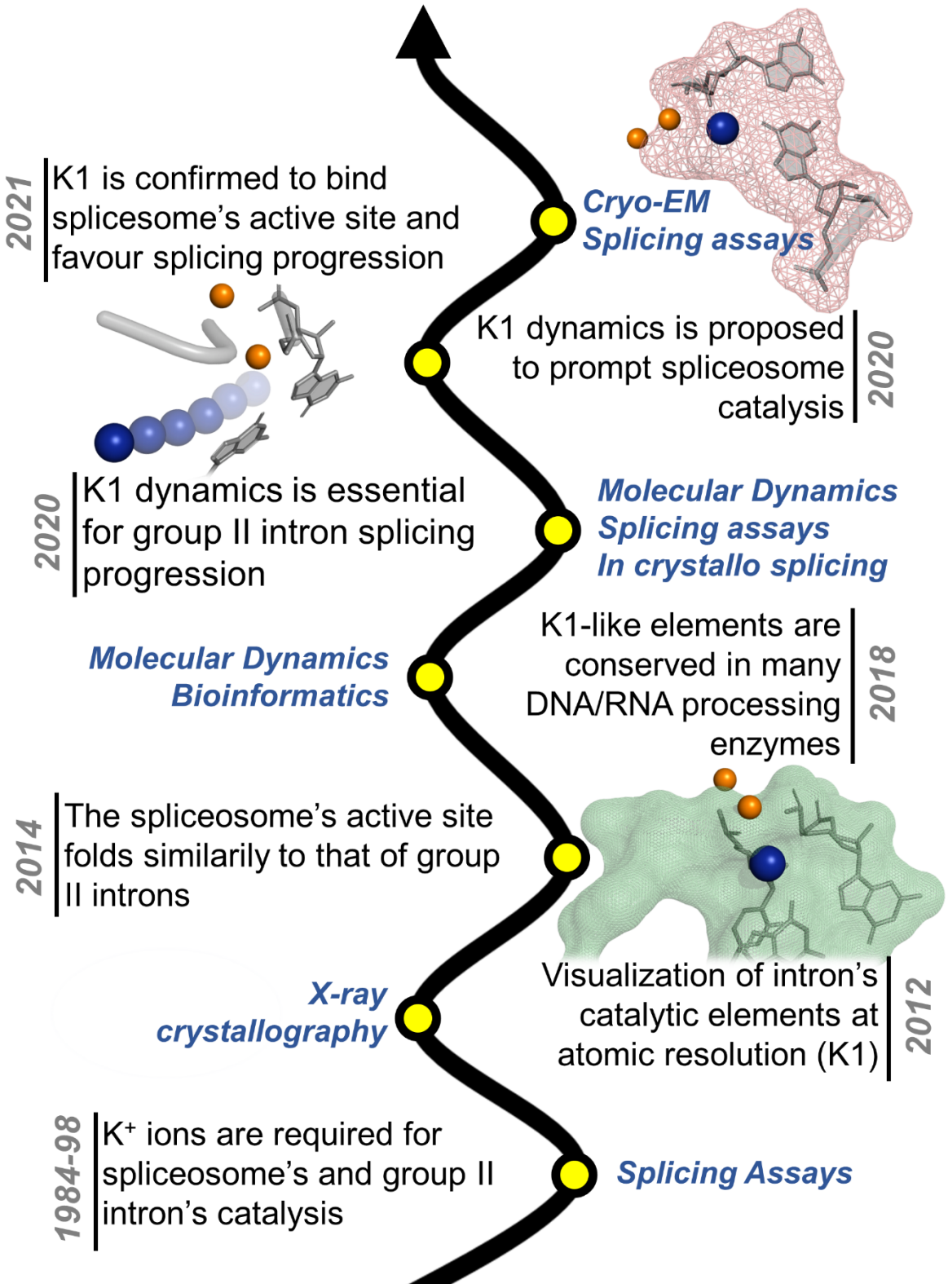


Figure 5.2: Computational and experimental milestones that marked the progressive discovery of the role of K1. *Functional studies initially revealed the importance of potassium for group II intron and spliceosomal splicing[157] but only three decades later was the K1 potassium ion identified in intron active site by crystallography.[35] Subsequent generalization of the K1 importance in multiple classes of nucleic-acid processing enzymes by means of structural/computational analyses,[73] and the elucidation of the functional role and dynamics of K1 by structural, enzymatic, and MD analyses[81] led to the prediction that a similar ion would also bind within the spliceosomal active site. This ion and its dynamics have now been successfully identified in the major and minor spliceosomes.[219,237]*

Remarkably, in the last few weeks, two new structures below 3.0 Å resolution of the major spliceosomal C_i complex[219] and the minor spliceosomal B_{act} complex[237] provided the necessary information to experimentally identify and localize specific structural and functional elements in and around the spliceosomal active site. Outstandingly, K1 was identified in the exact same position as predicted by our analyses and simulations, back in 2018 (Figure 5.1B-D).[73] In the major spliceosome K1 is coordinated by G52, G60, and U80 (U6 snRNA) (Figure 5.1B), and in the minor spliceosome K1 is coordinated by G19, G27, and U46 (U6atac snRNA) (Figure 5.1C – see also coordination distances). The overlap of our predictive model with these two new cryoEM structures returned an RMSD of ~0.6 Å (Figure 5.2), calculated using the first shell coordination of K1 – which is the exact K1 coordination shell we had predicted.[73] In both structures, K1 appears, indeed, crucial for catalysis, and specifically engaged both in the first (in the minor spliceosome B_{act}) and the second (in the major spliceosome C_i) steps of splicing.[219,237] The functional engagement of K1 was also confirmed by splicing assays, monitored using a stalled major spliceosomal C complex.[219]

But K1 is not just a static structural component of the intron and spliceosomal active site. Extensive computational studies performed since 2012 on group II intron X-ray structures at multiple stages of catalysis had contributed to explain the complex functional role and dynamics of K1 throughout the splicing cycle.[81] Indeed, multi-microsecond equilibrium MD simulations and free-energy calculations (metadynamics) have shown that K1 is dynamically bound and released to and from the intron active site favoring functionally important conformational rearrangements of the intron's active site, which are required for exchanging the reaction products and substrates in between the first and second steps of splicing.[81] Notably, according to our results, this sequence of events is directly triggered by the protonation of one catalytic residue just after the first splicing reaction, as indicated by quantum-mechanics/molecular-mechanics (QM/MM) simulations coupled with enzymatic assays, mutagenesis and structural characterization.[81]

Importantly, the K1 dynamics observed in the intron are analogous to dynamics of K1-like residues in protein enzymes. In polymerase-η, the altered dynamics of K1-like second-shell basic residues was shown, via equilibrium MD and enhanced sampling simulations, to disfavor the formation of the Michaelis-Menten complex, ultimately impairing the catalysis.[73,175] Similar positively-charged second-shell residues have been recently shown via MD simulations to play an active role for catalysis in other nucleic-acid processing metalloenzymes, as in the case of λ-exonuclease, dUTPases, RNase H, and human exonuclease 1 enzymes.[74–76,82] As a result of these studies, and in light of the extended similarities between the intron and protein enzymes, it seemed likely that similar dynamic events are ubiquitous and necessary for nucleic acid processing, and would thus also occur in the spliceosome.[81] Remarkably, by comparing the new spliceosomal cryoEM structures with previous ones obtained at various steps throughout the splicing cycle, K1 appears to be indeed dynamic and transient, i.e., bound to the active site for the catalytic steps, but released during conformational rearrangements.[219]

Taken together, these results illustrate how comparative structural, functional, evolutionary studies on the group II intron coupled with extensive molecular simulations and free energy calculations of both such a challenging system and other convergently-evolved nucleic-acid-processing enzymes enabled accurate predictions into the intricate catalytic core of the spliceosomes. Notably, experimental characterization of these complex systems has inevitably lagged behind due to the complexity of these megadalton-large ribonucleoproteins.

The recent experimental confirmation of the accuracy of such structural and mechanistic predictions offers increased confidence in the predictive power of such computational simulations, when appropriately integrated with evolutionary analysis and experimental data. Predicting such precise structure/functional correlations between the catalytic heteronuclear metal clusters of the group II intron and the spliceosomes can guide the mechanistic interpretation of high-resolution structural data, facilitating the functional dissection of these vital splicing machineries. This example thus shows that a computationally-informed approach

can deliver valuable insights into complex protein-nucleic acid systems, particularly those relevant to human diseases.

This is a time when RNA-targeted drug development is emerging as a viable strategy for developing new therapeutics, including against the spliceosome itself, as shown by the recent FDA approval of risdiplam for the treatment of spinal muscular atrophy.[58] The precise modeling and design of organic and inorganic compounds, located at binding pockets in and around catalytic centers of complex ribonucleoprotein machineries, can be of great value for fostering structure-based drug design even before high-resolution structures are experimentally determined.[238] In this scenario, the confirmed accuracy of mechanistic predictions based on integrated MD simulations and structural data, about the functional role of K1,[73] reinforces the notion that positively-charged second-shell residues are indeed essential regulators of nucleic-acid-processing chemistry together with the previously-recognized two divalent metal ions.[73] This observation provides precise grounds for understanding the mechanism, modulating function, and targeting many medically-relevant enzymes beyond splicing complexes with small molecules.[238] For instance, our predictive structure-function insights can rationally support the biotechnological engineering of CRISPR-Cas systems, which are now emerging as potentially powerful "drug machines" for gene or RNA editing and for personalized gene therapies. More broadly, integrated and computationally-driven approaches, as the one we discuss here, can serve as a useful reference (and confidence boost, too) for future modeling and prospective mechanistic interpretation of many other large macromolecular machineries that are essential for life and critically involved in diseases, but difficult to experimentally characterize at high resolution. Critical example of such complex machineries include membrane-embedded supercomplexes or ribonucleoproteins formed by highly-structured long non-coding RNAs.[11] Prospective applications of molecular simulations for structure refinement and mechanistic insights will undoubtedly play a prominent role in the incessant exploration of such complex biological multicomponent systems.

Chapter 6. Computer-Aided design of small molecules targeting RNA

6.1 Abstract

Structured RNAs are a drug-discovery breakthrough. These molecules regulate many vital cell functions, and are involved in the development of several human pathologies, from cancer to viral and bacterial infections. Among them, group II introns regulate gene expression in plants, fungi, and bacteria. Moreover, owing to their structural and functional properties, group II introns are considered ancestors of the human spliceosome, whose function is altered in severe human diseases. Recently, it has been shown that small molecules can be designed to interact with and modulate the biological function of structured RNAs. Here we discuss the possibility of targeting group II intron ribozymes to develop novel antimicrobial agents as well as to design splicing inhibitors with potential application in the rescue of disease-relevant altered splicing functions in humans.

6.2 Small Molecules Can Target Functionally Important Structured RNAs

RNA molecules can control vital cellular processes, such as translation, splicing, gene expression, immunity, and metabolic proteins regulation. These RNA molecules often adopt complex tertiary structures essential to their function.[239] For instance, thanks to their well-defined 3D structures, rRNAs, and tRNAs cooperatively catalyze protein synthesis. Their activity is modulated by interactions with structured regions of mRNA, such as 5'- and 3'-UTRs or IRES motifs.[240] Self-splicing group I/II introns and snRNA subunits of the spliceosome also require a precise 3D architecture to recognize splicing junctions with high fidelity and efficiency, and produce in-frame mRNA.[81] In some instances, these precise 3D structures can be targeted by organic compounds, particularly at protein-recognition sites.[135] In other instances, structured RNAs are tuned by small organic molecules at the functional level. For example, riboswitches are regulatory motifs generally located within 5'-UTRs of certain mRNAs that can be targeted by a variety of endogenous metabolites including vitamins and ions.[241]

Importantly, structured RNAs are potential targets to treat many serious diseases, from cancer to viral and bacterial infections.[6,7] Indeed, RNA druggability has been demonstrated by studies on bacterial riboswitches. For instance, a new class of potent PreQ1 riboswitch binders was recently identified in a study that boosted the potency of the initial PreQ1-targeting binders by crystallizing PreQ1 in complex with a ligand, highlighting the impact of structure-based approaches to RNA drug design.[242] Micro-RNA (miRNA) and their precursors are also promising targets. For example, the small molecule Targaprimir-515 was recently reported to selectively target miR-515 hairpin precursor to sensitize HER2-negative cancer cells to Herceptin.[243] Notably, Targaprimir-515 was identified with help from Inforna, a hybrid experimental-computational tool[244] for mining and collecting unique RNA-small molecule interaction patterns. The success of this sequence-based approach further demonstrates that folded RNA motifs are druggable with selective small

molecules.[244] Furthermore, structured functional motifs in the genome of RNA viruses are attractive targets for drug discovery. For instance, one C6 phenylacetylene substituted amiloride derivative was recently identified as a potent binder of an internal ribosome entry site (IRES) in the enterovirus 71.[245] Also, the HIV-1 TAR motif was successfully targeted via structure-based virtual screening strategies that combined molecular dynamics (MD) simulations, docking, and NMR experiments, leading to the identification of 6 small molecule inhibitors of HIV replication.[120]

6.3 Targeting Group II Intron Ribozymes with Small Molecules

Group II introns are structured RNAs that control gene expression in many microorganisms, thus representing a target to develop novel antimicrobial agents. Indeed, small molecules have also been successfully targeted at group II self-splicing introns. For example, the small molecule Intronistat B inhibits group II introns of pathogenetic fungi. Intronistat B was discovered via fluorescence-based and radioanalytical screening assays.[28] Moreover, group II introns are evolutionarily-related ancestors of the human spliceosome and, thus, can be potentially exploited to develop small molecules modulating human splicing functions involved in the development of severe diseases. Indeed, this approach builds upon the novel evidence of small molecules targeting the spliceosomal RNA, as confirmed by the recent FDA approval of Risdiplam as a treatment for spinal muscular atrophy (SMA).[246]

In *Chapter 3*, we have shown that the conformational dynamics of the group II introns is fundamental to guarantee the progression of catalysis.[81] Importantly, with the use of X-ray crystallography and molecular dynamics simulations, we reached an unprecedented atomistic resolution of such intron's structural rearrangement during its splicing. As a natural follow-up study, we built upon these pieces of knowledge to set up a hybrid computational-experimental pipeline to develop novel small molecule splicing inhibitors.

Despite, at the current stage, the promising results of this drug discovery campaign cannot be disclosed in this thesis, here I report a brief summary of this research project. We have identified functional introns' metastable states involved in splicing progression, as characterized with atomistic MD simulations. These high-resolution catalytic intermediates are used as structural models for a consensus docking protocol, aiming to identify novel small molecule binders through *in silico* screening of compounds' libraries. Notably, in order to enhance the hit-selection rate, this library is initially filtered to confine the small molecules' chemical space to that of known RNA binders, according to the available literature. That is, we have collected the physicochemical properties of known RNA binders from publicly available databases,[247] and we have used them to filter out unlikely RNA-binding molecules from the in-house compounds' library. Overall, this computational protocol, as combined with compounds' activity evaluation via splicing assays and medicinal chemistry hit expansion, has driven to the identification of one class of splicing inhibitors, which are currently investigated with X-ray crystallography in order to gain more insights into their binding mode.

Chapter 7. Conclusions and Future Perspectives

Both coding and non-coding RNA molecules play a crucial role in cellular homeostasis, and the altered expression of certain RNAs has been recently linked to the development of severe human pathologies.[3,5] Accordingly, understanding the enzymatic processing of RNA molecules is fundamental to characterize the biological regulation of their expression and to develop novel therapeutic strategies and tools.[6–8]

In this context, this research project aimed to investigate the metal-aided processing of RNA as catalyzed by two two-metal-ion enzymes, namely the group II intron ribozymes and the RNase H1. We used state-of-the-art molecular dynamics simulation and free energy calculation, complemented with X-ray crystallography, biochemical assays, and comprehensive structural analysis to show that an extended two-metal-ion architecture is required to properly support enzymatic catalysis. That is, our results suggest that a heteronuclear cations cluster is formed and disrupted at the active site of RNA-processing enzymes to support their catalytic cycle. Specifically, the initial proper positioning and the following timely release of a second-shell K^+ ion favors group II introns' catalysis by regulating the structural dynamics of their active site.[81] Indeed, such controlled cations trafficking guarantees for, at first, reactants alignment and, then, product release and splicing progression.[81] Similarly, multiple and diverse cations, i.e., Mg^{2+} and K^+, are strategically recruited at and opportunely shuttled out from the active site of the RNase H1 to support the formation of the catalytic Michaelis-Menten complex and, upon catalysis, the removal of the leaving group from the active site for reaction turnover.[82]

Importantly, our structural analysis revealed that an active site architecture similar to that of the abovementioned metalloenzymes is shared among a plethora of nucleic acid processing enzymes, including CRISPR-Cas systems.[82] In fact, either second-shell basic residues or cations are found at the active site of these

enzymes to occupy regions well overlapping with those found transiently filled by metal ions in both group II introns and RNase H.[82] Together, these pieces of knowledge suggest that these metalloenzymes may adopt a similar catalytic strategy that has convergently evolved to guarantee the functional and accurate processing of RNA molecules.[82]

Building upon this evidence, and the structural and functional similarity between group II introns and the human spliceosome, we proposed that these two splicing machines adopt the same mechanism to process RNA molecules.[81] Importantly, our prediction of the dynamic formation and controlled disruption of a heteronuclear metal ion cluster for spliceosome catalysis has been recently confirmed by two new high-resolution Cryo-EM maps of both the major[219] and the minor[237] spliceosome, capturing the predicted heteronuclear cations' cluster at their active site.[49] Furthermore, biochemical data confirmed our MD-driven findings of the transient formation of this cluster during catalysis, further supporting the reliability of the predicted catalytic mechanism.[49]

In conclusion, in this research project, we used MD simulations to characterize, at atomistic resolution, the catalytic cycle of disease-relevant RNA-processing metalloenzymes and ribozymes. Such high-resolution structural and functional insights certainly corroborate and sensibly expand the current understanding of metal-aided nucleic acids' catalysis. Most remarkably, in the era of RNA-targeted drug discovery,[246] our results will support the rational design of small molecules targeting the active site of pharmaceutically relevant ribozymes, such as group II introns as well as the human spliceosome.

Appendix A. Supporting Information for *Chapter 3*

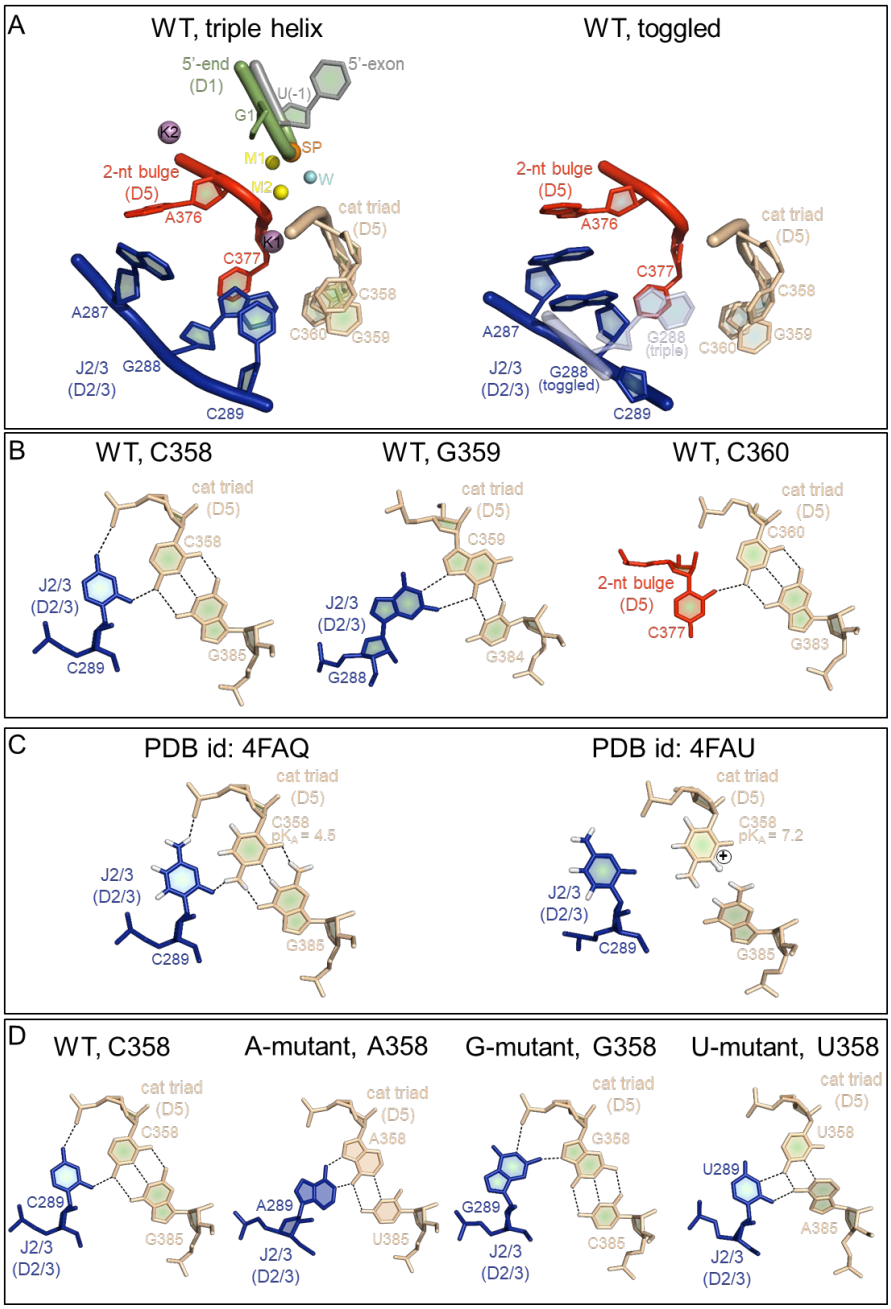


Figure A.1: Triple helix and toggled conformations in group II intron. ***A)*** *The active site of the pre-hydrolytic intron (PDB id.: 4FAQ, left) and of the toggled intron (PDB id.: 4FAX, right). J2/3 junction is in blue, two-nucleotide bulge in red, catalytic triad in brown, intron 5'-end in green, and 5' exon in gray. Catalytic magnesium ions M1-M2 are depicted as yellow spheres, potassium ions K1-K2 as violet spheres, the nucleophilic water molecule is in cyan, and the scissile phosphate (SP) in orange.* ***B)*** *Triple interactions for the three nucleotides of the catalytic triad (from left to right: C358, G359, and C360) in wild type.* ***C)*** *Left: C358 in the pre-hydrolytic state forms a triple interaction, as observed in PDB id.: 4FAQ. In this state, the pK_A of C358 is 4.5, so that its nucleobase is not protonated. Right: In the toggled structure, the triple interaction is disrupted, as observed in PDB id.: 4FAX. In this state, the pK_A of C358 is 7.2, so that its nucleobase can be protonated on the N3 atom. Hydrogen bonds are depicted as black dotted lines, and hydrogen atoms are in white.* ***D)*** *From left to right: triple interactions of residue 358 in wild type (same as panel C, left), A-mutant, G-mutant, and U-mutant.*

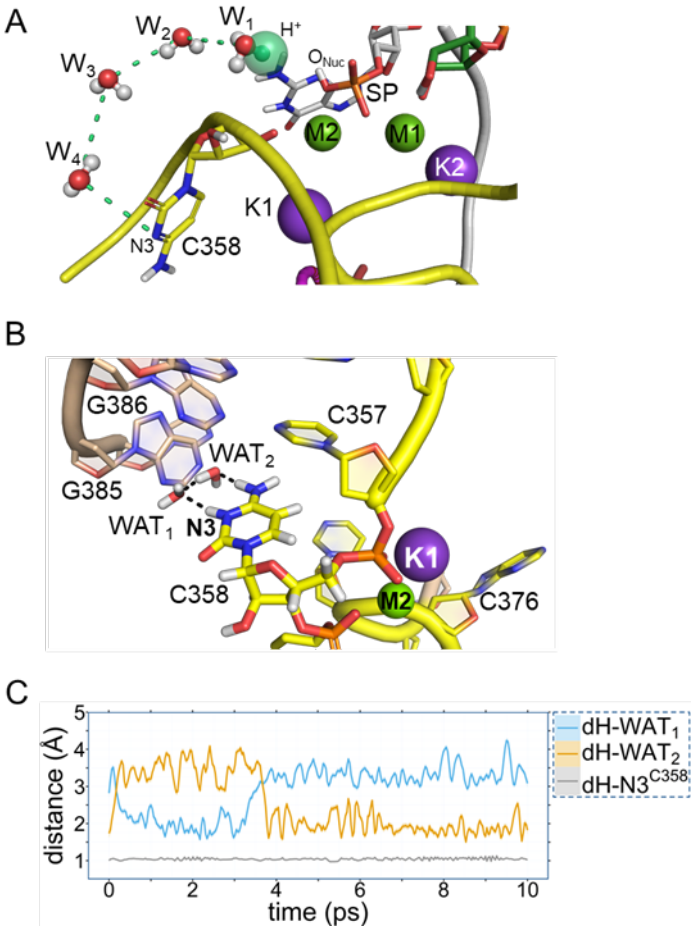


Figure A.2: N3C358 can be protonated after the first step of splicing. A) *Active site representation of the solvated structural model of the cleaved intron. The nucleophilic water (O_{Nuc}) has attacked the scissile phosphate (SP) and released a proton (H^+, semi-transparent dark green sphere) to a neighboring water molecule in the bulk solvent (W_1). A proton shuttling pathway composed of 3 water molecules at H-bond distances from one other (W_2-W_4) can conduct this proton to $N3^{C358}$ (green dashed lines). Other proton transfer pathways would also be compatible with our structures and can count up to 5 water molecules, based on previously reported QM/MM calculations.[44] Alternatively, proton transfer can also occur by unspecific diffusion through the bulk solvent.* ***B)*** *A representative snapshot of intron active site during hybrid quantum (DFT/BLYP)/classical simulations of $N3^{C358}$ protonated in the cleaved state. Dotted lines highlight H-bonds between C358 and nearby water molecules. The intron backbone is represented as a cartoon (catalytic triad and two-nucleotide bulge in yellow, triple helix partner nucleotides in brown). M2 and K1 are represented as green and purple spheres, respectively. Atoms in bold sticks are treated at the quantum level, while the rest of the atoms are treated at the classical level.* ***C)*** *Hybrid quantum/classical simulations of $N3^{C358}$ protonated in the cleaved state. Time series of selected distances during the last 10 ps of the simulations (two water molecules, in the quantum region, exchange their position relative to $N3^{C358}$ at ~3.5 ps): $d_{H\text{-}N3C358}$ (grey trace), $d_{H\text{-}WAT2}$ (orange trace), $d_{H\text{-}WAT1}$ (light-blue trace). Notably, $N3^{C358}$ remained protonated throughout the simulation time (~15 ps), never exchanging its proton with the nearby water molecules.*

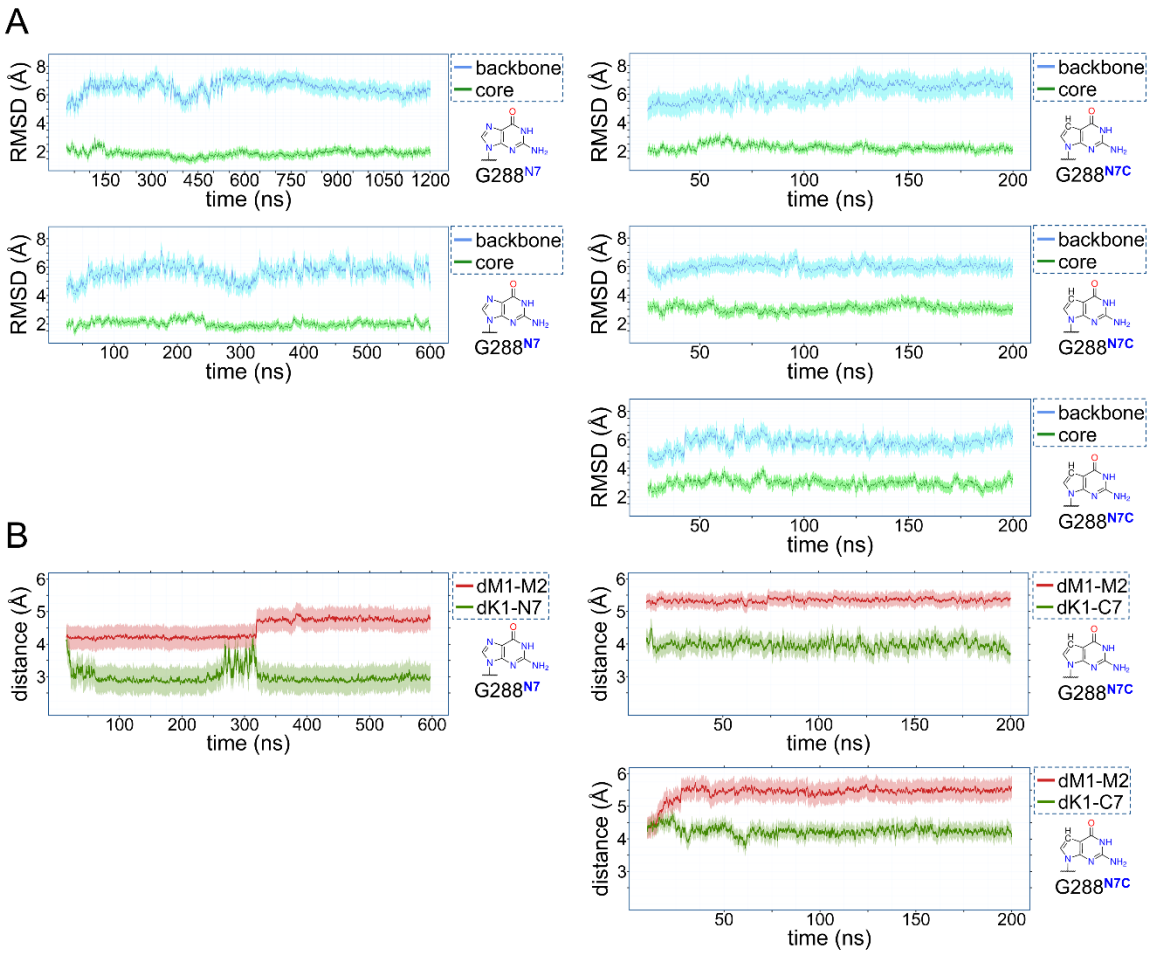


Figure A.3: MD simulations of the pre-reactive state (PDB id: 4FAQ). A) *Changes in RMSD for intron backbone (blue trace) and core (defined as the backbone and side chains of all nucleotides in D5 plus nucleotides 287-289 of the J2/3 junction, green trace) during the two simulations of the wild type pre-reactive state (left) and during the three simulations of the* $N7^{G288}$*-deaza mutant pre-reactive state (right).* ***B)*** *Changes in* $d_{K1\text{-}N7G288}$ *(green trace),* $d_{M1\text{-}M2}$ *(red trace), during the 600-ns long MD simulation of the pre-reactive state (left; 1200-ns long simulation reported in* ***Figure 4****). In this simulation, K1 transiently drifted away from the active site after ~280 ns, causing a shift of G288 and M1-M2 (*$d_{M1\text{-}M2}$ *= 4.77 ± 0.16 Å). However, in contrast with the simulations of the cleaved state (****Figure 5****), here K1 was unable to leave the active site spontaneously. The concerted shift of K1, M1-M2, and G288 in this simulation confirms that the reciprocal structural position of these residues is intimately interconnected. On the right, changes in* $d_{K1\text{-}N7G288}$ *(green trace),* $d_{M1\text{-}M2}$ *(red trace), during two additional 200-ns-long replicas of the N7-deaza mutant pre-reactive state.*

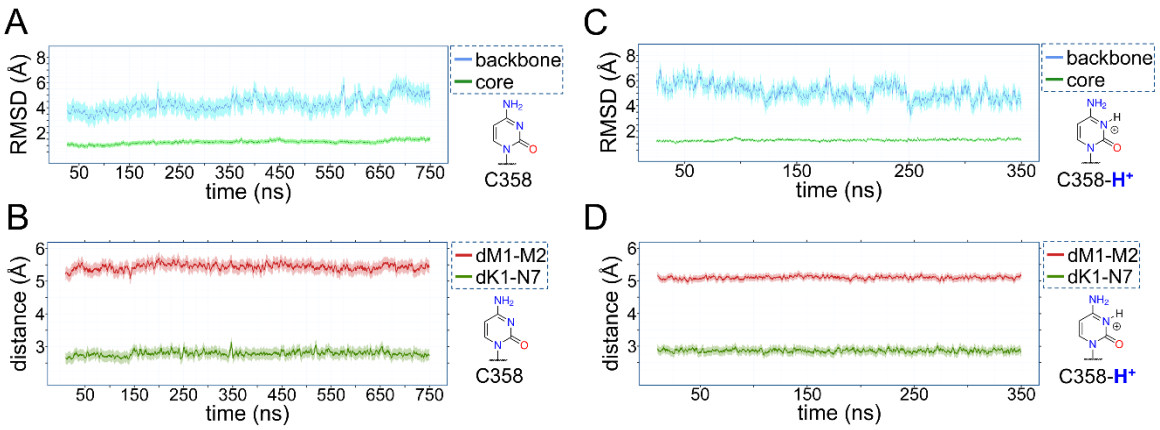


Figure A.4: MD simulations of the post-hydrolytic state (PDB id: 4FAR). A) *Changes in RMSD during the simulations of the post-hydrolytic state for the backbone of all intron nucleotides (blue trace) and for the intron "core" (i.e. backbone and side chains of all nucleotides in D5 plus the three nucleotides of the J2/3 junction; green trace).* ***B)*** *Evolution of* $d_{K1\text{-}N7G288}$ *(green trace) and* $d_{M1\text{-}M2}$ *(red trace) during the simulations of the post-hydrolytic state.* ***C)*** *Changes in RMSD during the simulations of the post-hydrolytic* H^+ *state for the intron backbone (blue trace) and the intron core (green trace).* ***D)*** *Evolution of* $d_{K1\text{-}N7G288}$ *(green trace) and* $d_{M1\text{-}M2}$ *(red trace) during the simulations of the post-hydrolytic* H^+ *state.*

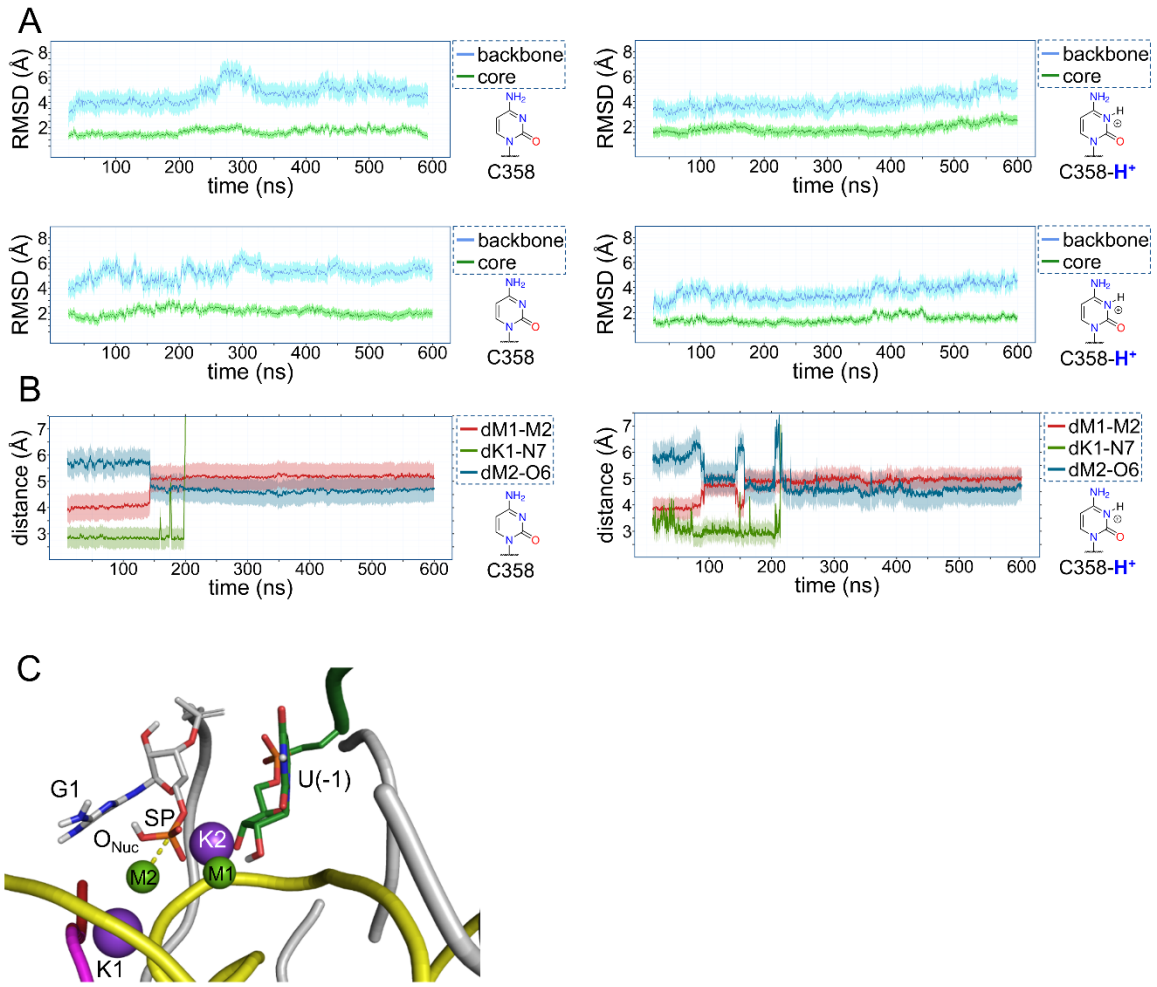


Figure A.5: MD simulations of the cleaved state of wild-type (modeled from PDB id: 4FAQ). A) *Changes in RMSD during the two simulations of the cleaved state of the non-protonated (left) and protonated (right) wild type intron for the backbone (blue trace) and the core (green trace, defined as in* ***Figure S3****).* ***B)*** *Changes in* $d_{K1\text{-}N7G288}$ *(green trace),* $d_{M1\text{-}M2}$ *(red trace), and* $d_{M2\text{-}O6}$ *(blue trace) during the second MD simulation of the cleaved state of the non-protonated (left) and protonated (right) wild type intron.* ***C)*** *The active site of the pre-hydrolytic intron (PDB id: 4FAQ) modified to cleave the 5'-splice junction. The intron backbone is represented as a cartoon representation (catalytic triad and two-nucleotide bulge in yellow, J2/3 junction in purple, 5'-exon in green). M1 and M2 are represented as green spheres. The scissile phosphate (SP) is represented as sticks in the cleaved configuration. Oxygen atoms are red (*O_{Nuc} *is the oxygen atom contributed by the nucleophile), hydrogen atoms white, and the phosphorus atom is orange. The dashed yellow line indicates the distance between the scissile phosphate (SP) and M2 (*$d_{SP\text{-}M2}$*).*

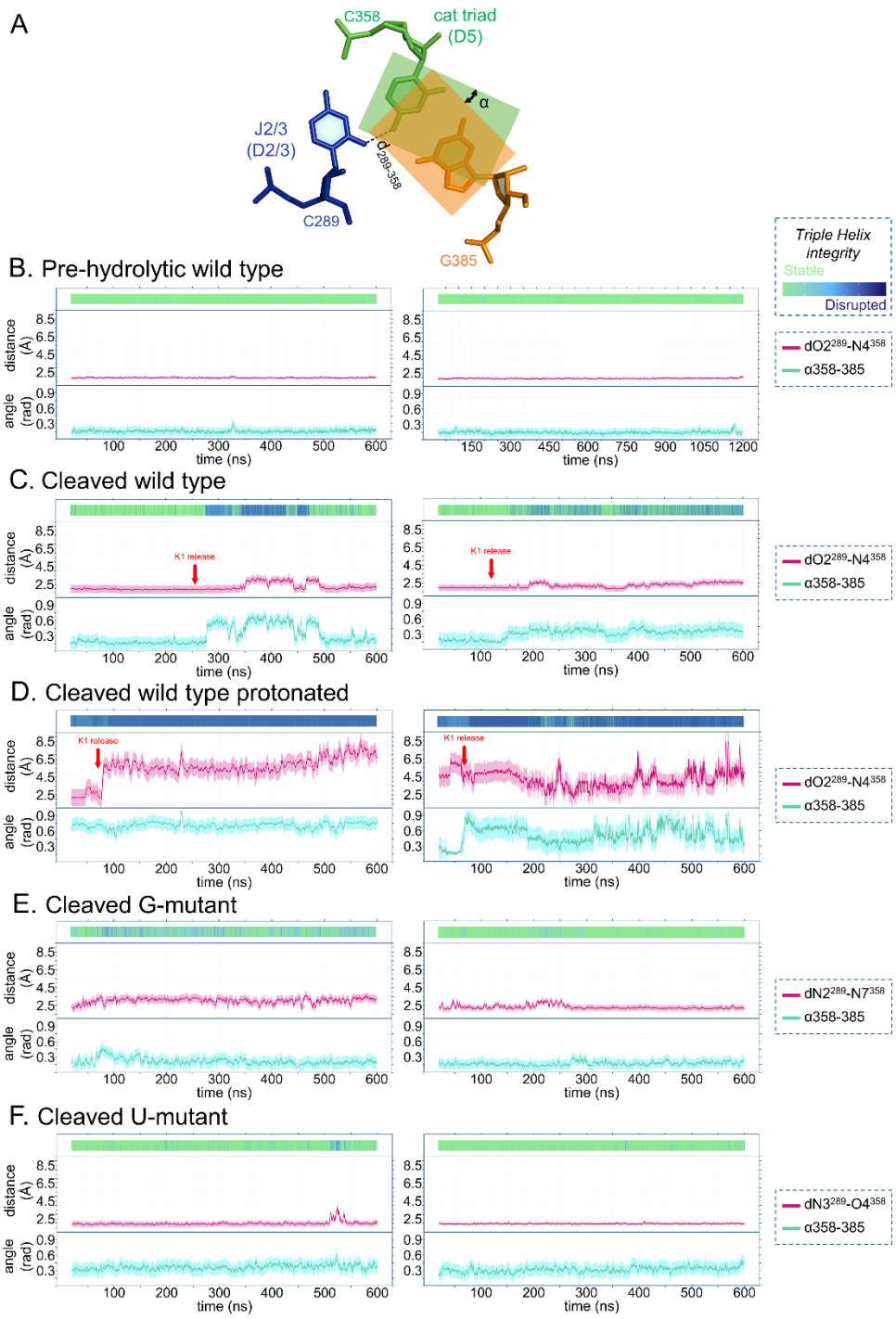


Figure A.6: K1 release induces the destabilization of the triple helix active site. A) *Definition of the two parameters used to monitor the geometry of the intron triple helix at position 289: $d_{289\text{-}358}$ corresponds to the distance between $C289^{O2}$ and $C358^{N4}$ (black dotted line); α corresponds to the dihedral angle between the planes of the nucleobases of residues 358 and 385 (here represented as semi-transparent green and orange planes, respectively).* ***B)*** *Changes of $d_{289\text{-}358}$ and α during the two simulations of the pre-hydrolytic state of wild type intron.* ***C-D)*** *Changes of $d_{289\text{-}358}$ and α during the two simulations of the non-protonated cleaved state (C) and protonated cleaved state (D) of wild type intron.* ***E-F)*** *Changes of $d_{289\text{-}358}$ and α during the two simulations of the cleaved states of the G- (E) and U-mutants (F). In each plot, the colored barcode plotted on top of every MD trajectory monitors the integrity of the triple helix. When the triple helix is formed ($d_{289\text{-}358}$ < 3 Å and α < 0.35 rad) the bar is green, when the triple helix is disrupted ($d_{289\text{-}358}$ > 3 Å or α > 0.35 rad) the bar is dark blue. Release of K1 in the protonated (D) and non-protonated (C) cleaved states of wild type intron (red arrows) favors the destabilization of the triple helix. The strongest destabilization occurs in the protonated cleaved state of wild type (D). Instead, in the cleaved states of the mutants (E and F), K1 is not released into the bulk water, and the triple helix remains stable.*

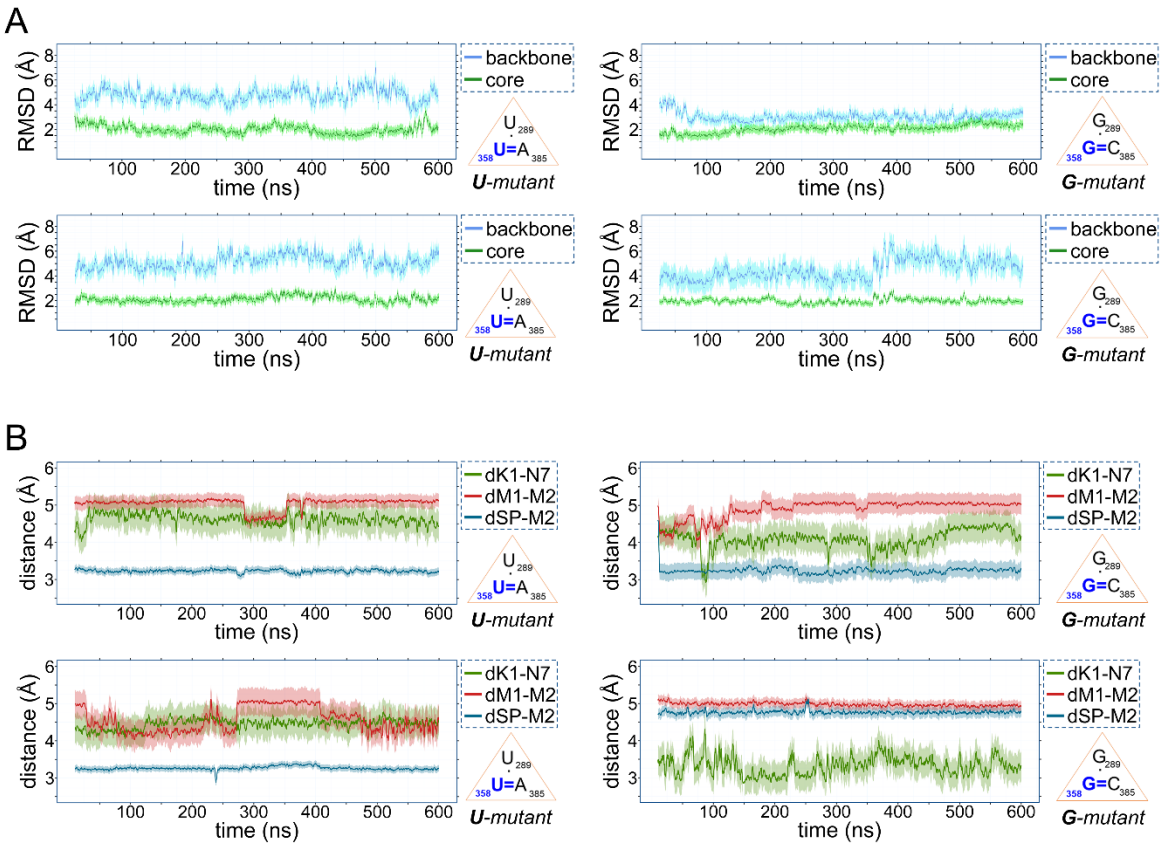


Figure A.7: MD simulations of the post-hydrolytic state of the mutants (PDB id: 6T3K and 6T3R). A) *Changes in RMSD during the two simulations of the post-hydrolytic states of the U- (left) and G- (right) mutants for intron backbone (blue trace) and core (green trace, defined as in* ***Figure S3****).* ***B)*** *Changes in $d_{K1\text{-}N7G288}$ (green trace), $d_{M1\text{-}M2}$ (red trace), and $d_{SP\text{-}M2}$ (blue trace) during the 600-ns-long MD simulation of the post-hydrolytic states of the U- (left) and G- (right) mutants. The SP becomes coordinated by the catalytic metal cluster ($d_{SP\text{-}M2}$ = 3.63 ± 0.18 Å) while K1-N7^{G288} interaction is only transiently established and K1 is not released into the bulk water ($d_{M1\text{-}M2}$ = 4.87 ± 0.24 Å).*

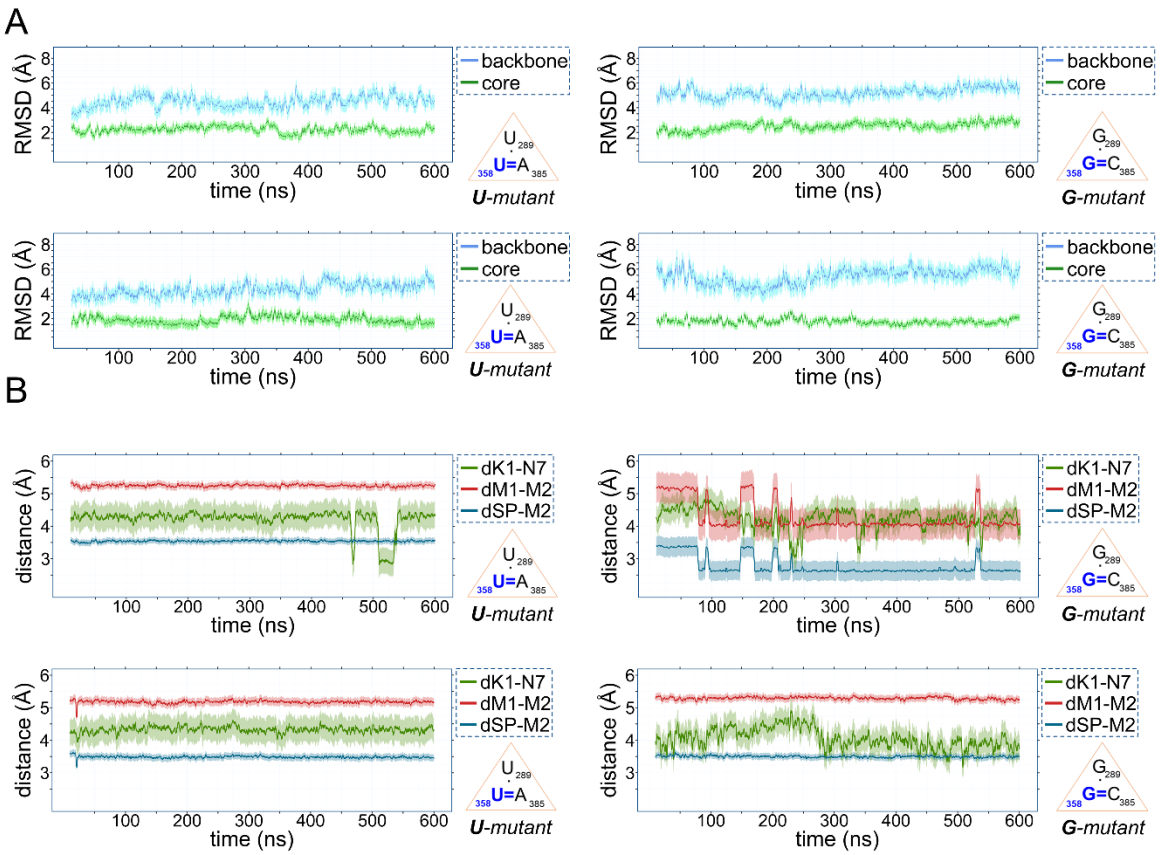


Figure A.8: MD simulations of the cleaved state of the mutants (modeled from PDB id: 6T3K and 6T3R). A) *Changes in RMSD during the two simulations of the cleaved states of the U- (left) and G- (right) mutants for intron backbone (blue trace) and core (green trace, defined as in* ***Figure S3****).* ***B)*** *Changes in $d_{K1\text{-}N7G288}$ (green trace), $d_{M1\text{-}M2}$ (red trace), and $d_{SP\text{-}M2}$ (blue trace) during the 600-ns-long MD simulation of the cleaved states of the U- (left) and G- (right) mutants. The SP is firmly coordinated by the catalytic metal cluster ($d_{SP\text{-}M2}$ = 3.33 ± 0.15 Å) while the K1-N7^{G288} interaction is only transiently established and K1 is not released into the bulk water ($d_{M1\text{-}M2}$ = 4.21 ± 0.40 Å).*

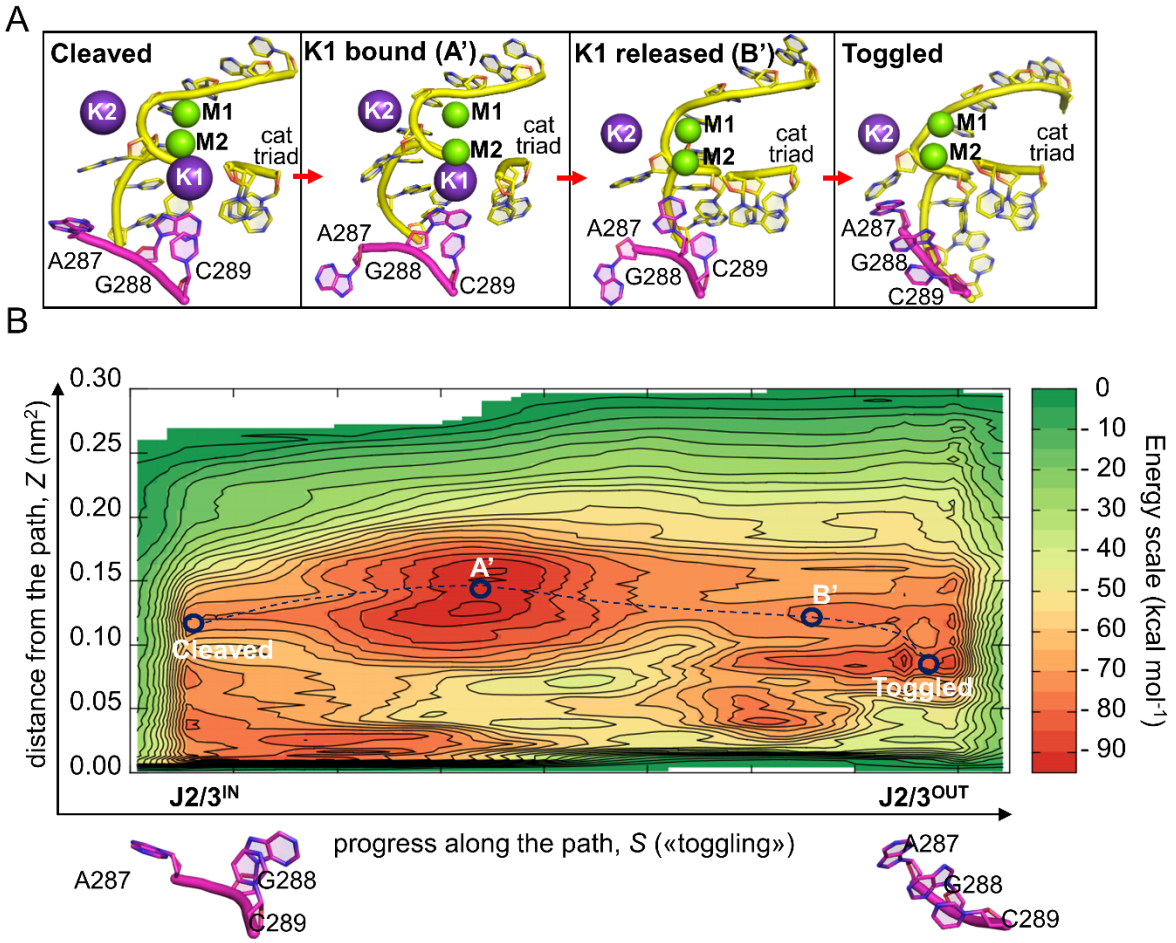


Figure A.9: Energetics associated with intron toggling in the non-protonated state. A) *Structural architecture of the active site for each state identified by the MtD simulation reported in panel B.* ***B)*** *Path metadynamics (MtD) free energy landscape of the cleaved state. The cleaved, A', B', and toggled states are indicated along the MtD trajectory (dark blue dotted line). The energy scale is indicated in kcal mol^{-1} on the right. The conformations of the J2/3 junction in state A' and in the toggled state are represented at the bottom of the figure.*

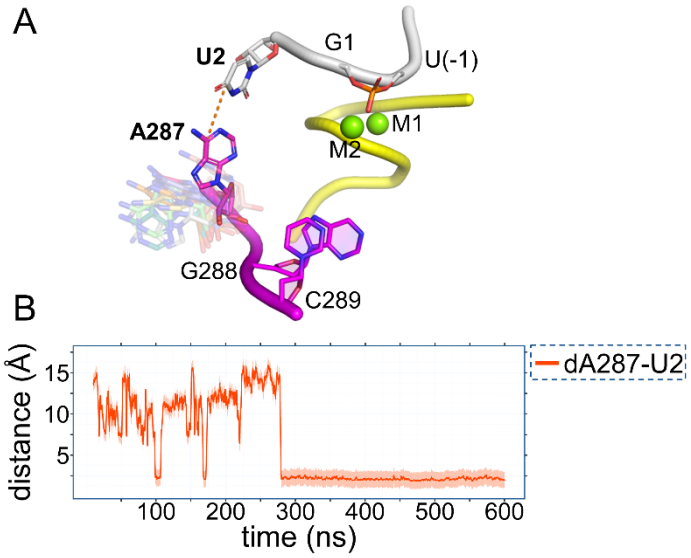


Figure *A.10*: *A287-U2 interaction.* *A)* *Graphical representation of the stabilized base pair between A287 with U2 (red dotted line indicated $d_{U2\text{-}A287}$).* ***B)*** *Evolution of $d_{U2\text{-}A287}$ during the MD simulation of the pre-hydrolytic state.*

***Table A.1: pK_A values of adenine and cytosine residues in the group II intron active site**, as empirically estimated using continuum electrostatics non-linear Poisson-Boltzmann calculations. We note that these calculations provide only qualitative approximations of the tendency of the reported nucleotides to shift pK_A in different conformational states.*

Residue	Precatalytic state (PDB id.: 4FAQ)	Toggled state (PDB id.: 4FAX)
J23 junction		
A287	6.6	7.3
C289	7.0	8.5
cat triad		
C358	4.5	7.2
C360	3.2	3.7
2nt-bulge		
A376	11.9	9.7
C377	10.3	0.7

***Table A.2: Kinetic rate constants of O. iheyensis wild type intron and A-, G-, and U-mutants**. Standard errors of the mean were calculated from n = 3 independent experiments. The wild type and A-mutant show a double exponential decay of precursor, so the kinetic rate constants of both the fast and slow populations are reported for these two constructs. The relative sizes of each population are indicated in the right column.*

Construct	k_1 (min^{-1})	k_2 (min^{-1})	population size
WT	fast: 0.133 ± 0.014	fast: 0.096 ± 0.015	fast: 64 %
	slow: 0.007 ± 0.003	slow: 0.006 ± 0.004	slow: 36 %
A-mutant	fast: 0.075 ± 0.009	fast: 0.045 ± 0.008	fast: 73 %
	slow: 0.008 ± 0.006	slow: < 0.006	slow: 27 %
G-mutant	0.011 ± 0.001	0.002 ± 0.001	
U-mutant	0.019 ± 0.001	0.012 ± 0.001	

Table A.3: Data collection and refinement statistics (molecular replacement).

	G-mutant (Mg^{2+}-K^{+})	U-mutant (Mg^{2+}-K^{+})	G-mutant (Mg^{2+}-Na^{+})	U-mutant (Mg^{2+}-Na^{+})
PDB code	6T3K	6T3R	6T3N	6T3S
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
unit cell dimensions				
a (Å)	88.7	89.5	89.9	89.9
b (Å)	95.4	95.1	95.3	95.2
c (Å)	225.0	222.4	217.8	227.9
α=β=γ (°)	90	90	90	90
energy (eV)	12,658	12,662	12,662	12,658
resolution (Å)[1]	48.5-3.44 (3.63-3.44)	44.5-3.57 (3.76-3.57)	32.1-3.22 (3.39-3.22)	49.5-3.28 (3.46-3.28)
R_{merge} (%)[1]	5.3 (150.8)	13.0 (79.8)	6.9 (54.2)	7.4 (99.5)
R_{meas} (%)[1]	8.1 (159.3)	15.7 (92.1)	8.9 (69.5)	8.9 (119.7)
R_{pim} (%)[1]	4.4 (86.2)	6.8 (38.5)	3.9 (30.2)	3.5 (46.7)
I/σ(I)[1]	17.0 (1.6)	9.4 (2.5)	14.5 (2.4)	15.8 (2.5)
completeness (%)[1]	94.1 (95.0)	98.7 (99.3)	98.6 (99.7)	99.9 (99.8)
multiplicity[1]	3.3 (3.2)	4.7 (4.8)	4.8 (5.0)	6.5 (6.5)
R_{work} (%)	21.2	21.2	18.5	19.8
R_{free} (%)[2]	25.7	25.0	24.2	22.9
RMSD				
bond length (Å)	0.009	0.002	0.009	0.008
bond angle (°)	1.89	0.601	1.96	1.792

[1]: highest resolution shell shown in parenthesis
[2]: R_{free} calculated using 5 % of the data

Appendix B. Supporting Information for *Chapter 4*

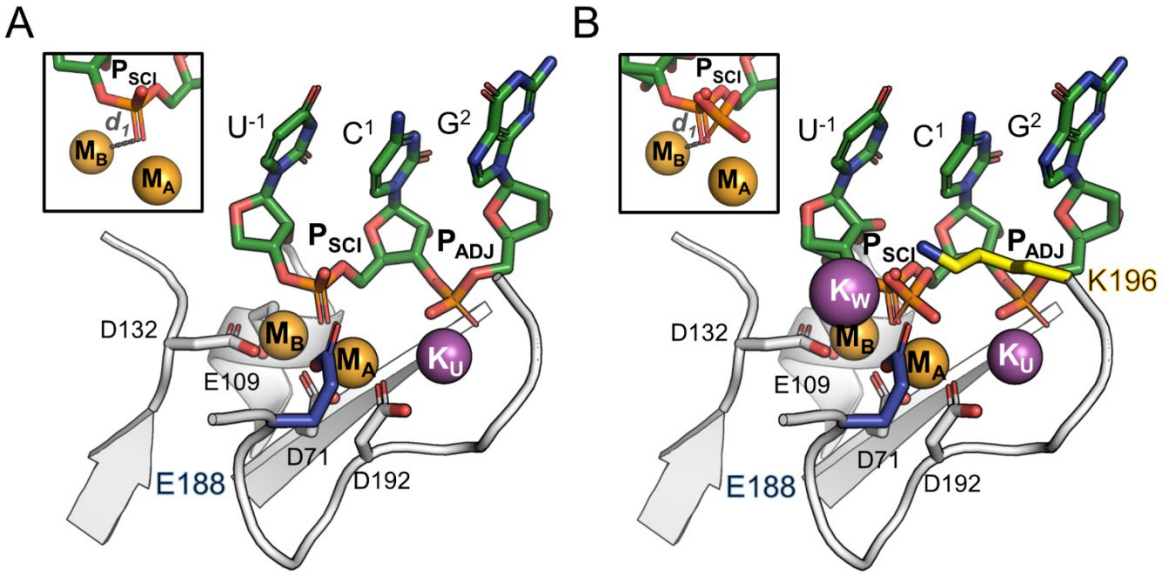


Figure B.1: Reaction intermediates of wild type RNase H1 as captured by X-ray crystallography. ***(A)*** *The reactant state of RNase H1 as obtained after 40s from the incubation (PDBid: 6DMV). The catalytic residues forming the DEDD motif (viz. D71, E109, D132, and D192; white) and the residue E188 (blue) are represented as sticks. The RNA nucleotides (green) including the scissile phosphate (P_{SCI}) and its adjacent phosphate (P_{ADJ}) are also shown as sticks. The two catalytic magnesium ions M_A-M_B (orange) and the ion K_U (purple) are shown as spheres. Notably, at this stage, K_W is not found at the active site, the P_{SCI} is not properly aligned onto the catalytic M_A-M_B (i.e., d1=2.6Å, left corner), and the products are not formed.* ***(B)*** *The reactant state of RNase H1 as obtained after 120s from the incubation (PDBid: 6DO9). Here, K196 (yellow) is represented as sticks, while the additional K_W (purple) is shown as a sphere. Importantly, at this stage, the Michaelis-Menten complex is properly formed (i.e., d1=2.2Å, left corner), and the formation of the products is observed.*

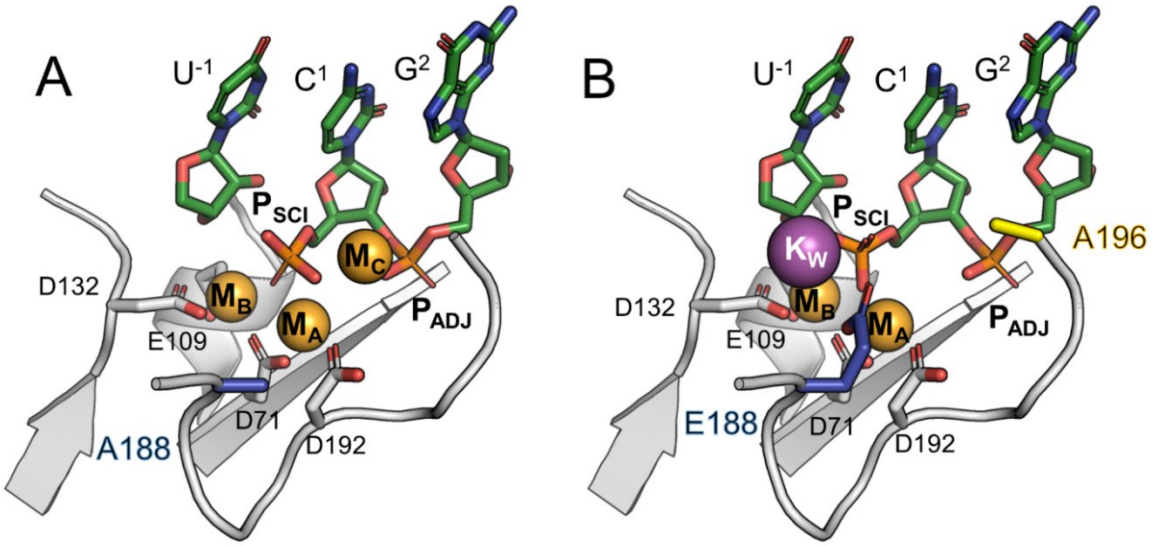


***Figure B.2: Reaction intermediates of RNase H1 mutants as captured by X-ray crystallography**. **(A)** The mutant E188A in the post-reactive state (PDBid: 6DPO), as obtained upon alanine mutation of the second-shell residue E188. The catalytic residues forming the DEDD motif (viz. D71, E109, D132, and D192; white) and the residue A188 (blue) are represented as sticks. The RNA nucleotides (green) including the scissile phosphate (P_{SCI}) and its adjacent phosphate (P_{ADJ}) are also shown as sticks. The two catalytic magnesium ions M_A-M_B (orange) and the additional third M_C ion (orange) are shown as spheres. Notably, in absence of the residue E188, K_W is not present in the vicinity of the active site, as opposite to what found in the wild type RNase H1. **(B)** The mutant K196A in the post-reactive state, as obtained upon alanine mutation of the second-shell residue K196 (PDBid: 6DPM). Here, the residues E188 (blue) and A196 (yellow) are represented as sticks, while the additional K_W (purple) is shown as a sphere. Notably, in absence of K196, the configuration of the scissile phosphate is rotated and the additional metal ion M_C does not locate at the active site.*

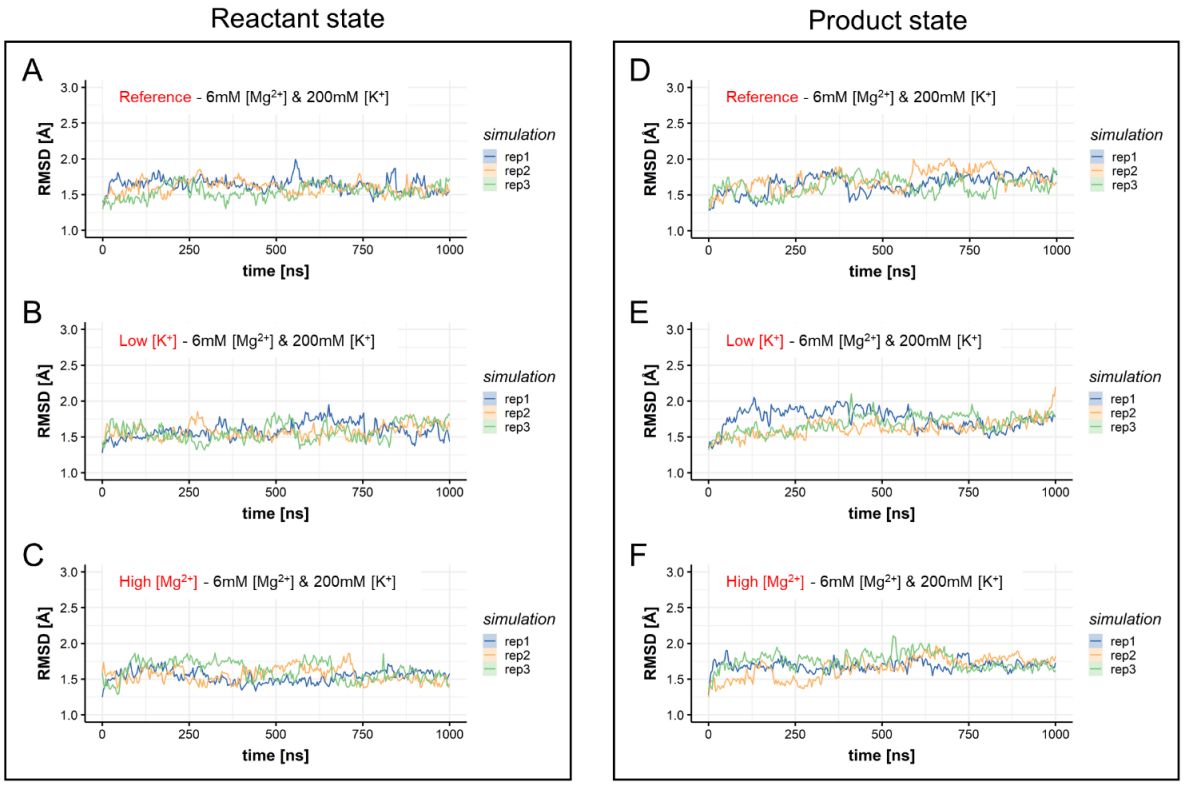


Figure B.3: Root mean square deviation (RMSD) for the backbone atoms of RNase H1 during molecular dynamics (MD) simulations. *On the right are shown the simulation replicas (blue, yellow, and green lines) of the RNase H1 in the reactant state at* ***A)*** *reference ionic strength;* ***B)*** *low K^+ concentrations;* ***C)*** *high Mg^{2+} concentrations. On the left are reported the simulation replicas (blue, yellow, and green lines) of the RNase H1 in the product state at* ***D)*** *reference ionic strength;* ***E)*** *low K^+ concentrations;* ***F)*** *high Mg^{2+} concentrations. Three simulation replicas were performed at each condition to verify the robustness of our observations.*

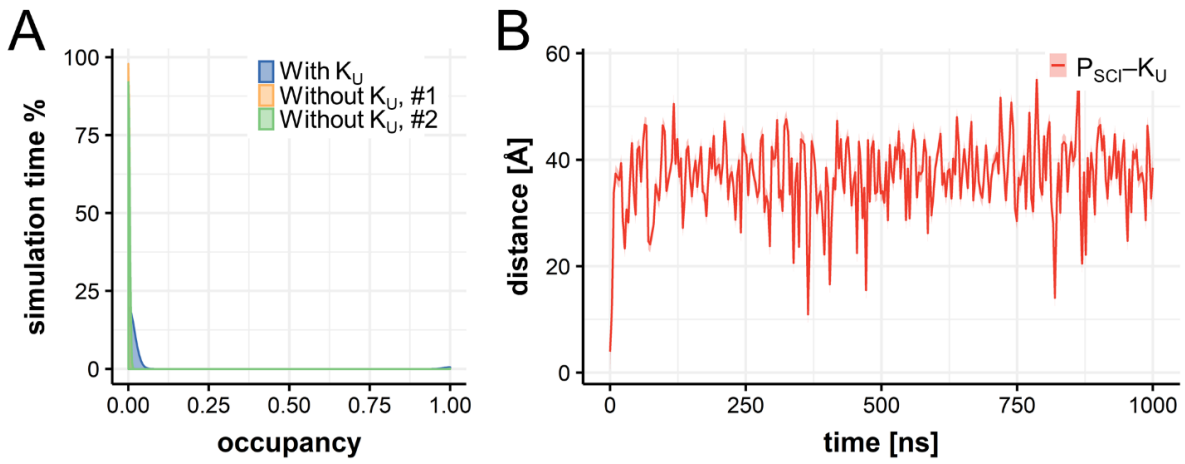


Figure B.4: Mobility of K_U in the reactant state. ***(A)*** *Occupancy of the U-site averaged along three ~1 μs-long MD simulations of RNase H1 in the reactant state (at optimal ionic strength, 6 mM [Mg^{2+}] and 200 mM [K^+]) including K_U (blue trace) and without K_U (yellow and green traces). The almost null occupancy of K^+ ion at the U-site indicates that K_U is not present during these simulations.* ***(B)*** *Distance between the scissile phosphate (P_{SCI}) and K_U (P_{SCI}–K_U) during ~1 μs MD simulations of RNase H1 in the reactant state. Here, K_U was included in the starting conformation. However, the ion is immediately released into the bulk.*

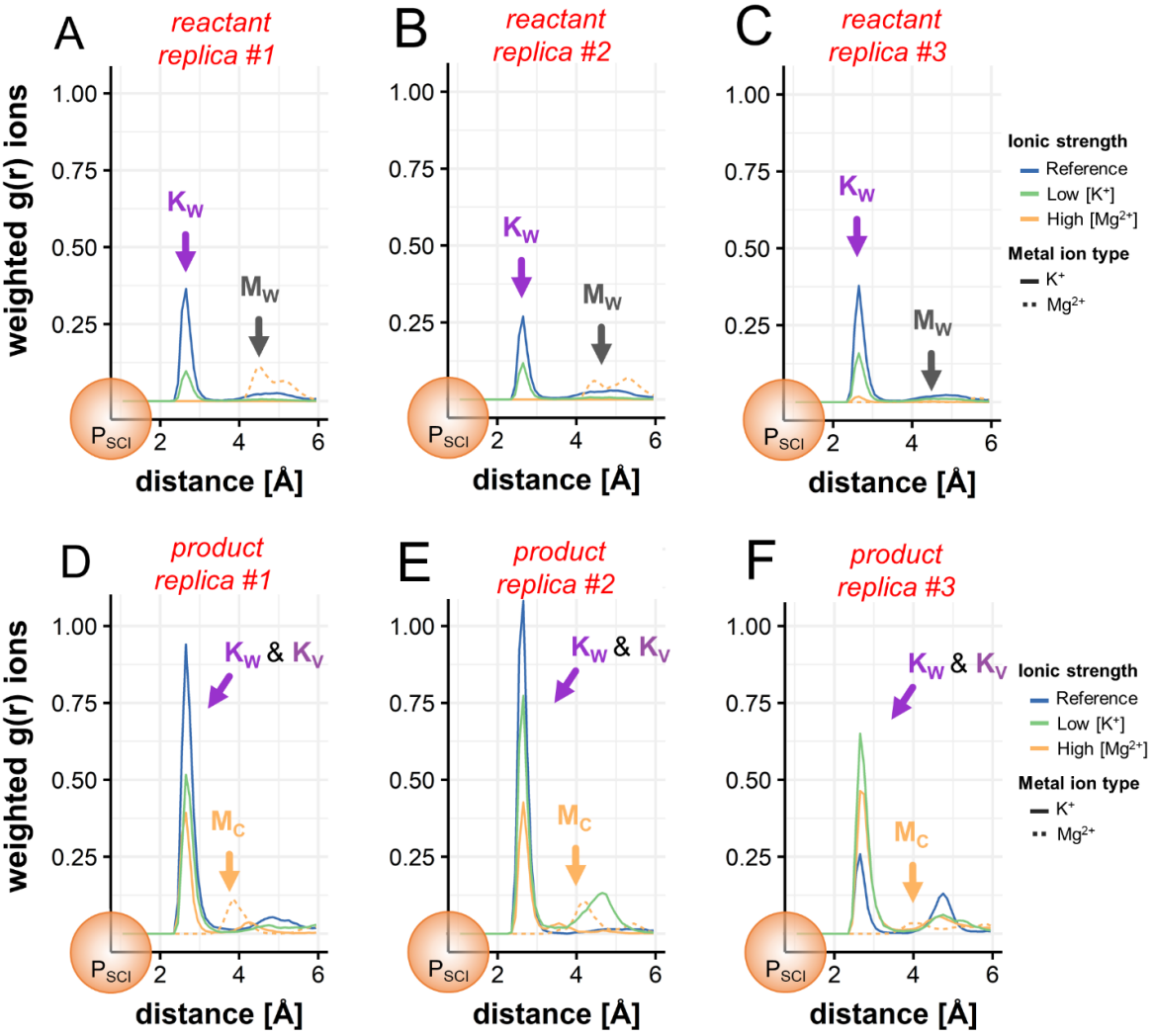


Figure B.5: Distribution of the ions at the RNase H1 active site before and after RNA hydrolysis. (A-C) *Weighted radius of gyration (g(r)) for K^+ (solid line) and Mg^{2+} (dashed line) ions around the scissile phosphate in the reactant state (viz. before hydrolysis), along three MD simulation runs of ~1 μs each, performed at optimal (blue trace, 6mM [Mg^{2+}] and 200mM [K^+]), high Mg^{2+} (yellow trace, 85 mM [Mg^{2+}] and 75 mM [K^+]) and low K^+ (green trace, 6mM [Mg^{2+}] and 25mM [K^+]) concentrations. Data are shown for three sets of simulation replicas. K_W (purple arrow) is transiently captured at the active site at both high and low K^+, but not at high Mg^{2+} concentrations, where an additional M_W (grey arrow) binds with E188 (Figure B.7D). Importantly, while panels (A-B) report data performed using the Aqvist model for Mg^{2+} ions, the Allnér force field has been employed to generate data reported in panel (C). The latter reveals a more flexible binding between M_W and E188 (Figure B.7C), resulting in a less pronounced M_W weighted distribution peak.* ***(D-F)*** *The same descriptors as in panels (A-C) are reported for three sets of ~1 μs-long simulation replicas of RNase H1 in the product state (viz. after hydrolysis). Here, regardless of the ionic strengths, K_W (purple arrow) is stably captured at the active site, while transient binding of K_V is also observed. At high Mg^{2+} concentration, the third M_C is bound to the scissile phosphate. This binding event is mediated by first-shell coordination waters of M_C. Data reported in panel (F) have been collected employing the Allnér force field for Mg^{2+} ions, showing a weakened interaction between M_C and the scissile phosphate, reducing the magnitude of M_C peak.*

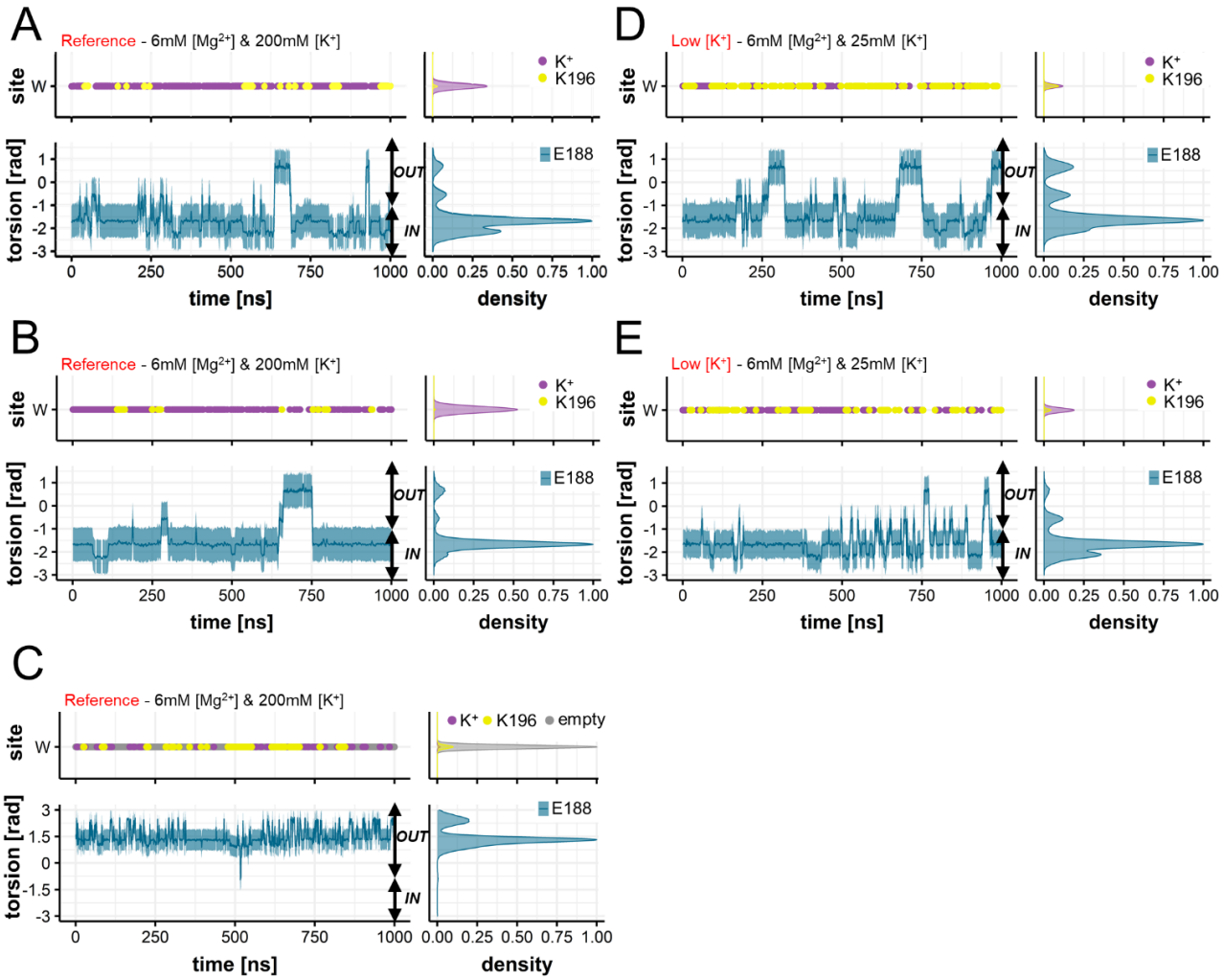


Figure B.6: MD simulations replica of the RNase H1 in the reactant state. (A-C) *Occupancy of the W-site (upper panel) by K^+ ions (purple) and by the K196 residue (yellow); and variation of the torsional angle E188-θ (lower panel, computed between the N-Cα-Cβ-Cδ atoms) along three additional MD replicas at optimal ionic strength (6mM [Mg^{2+}] and 200mM [K^+]). Normalized densities are shown on the right. High concentrations of K^+ favor the binding of K_W and the stabilization of $E188_{IN}$ conformation. An harmonic restraint has also been applied to E188 torsion (N-Cα-Cβ-Cδ) during the replica reported in the panel (C). Here, as E188 cannot explore "in" conformations, the binding of K_W is dramatically reduced.* ***(D-E)*** *The same descriptors reported as in panel (A), for two additional MD replicas at low K^+ concentration. These MD replicas corroborate the synergistic interaction between K_W and E188 for active site stabilization, as shown in Figure 4.2.*

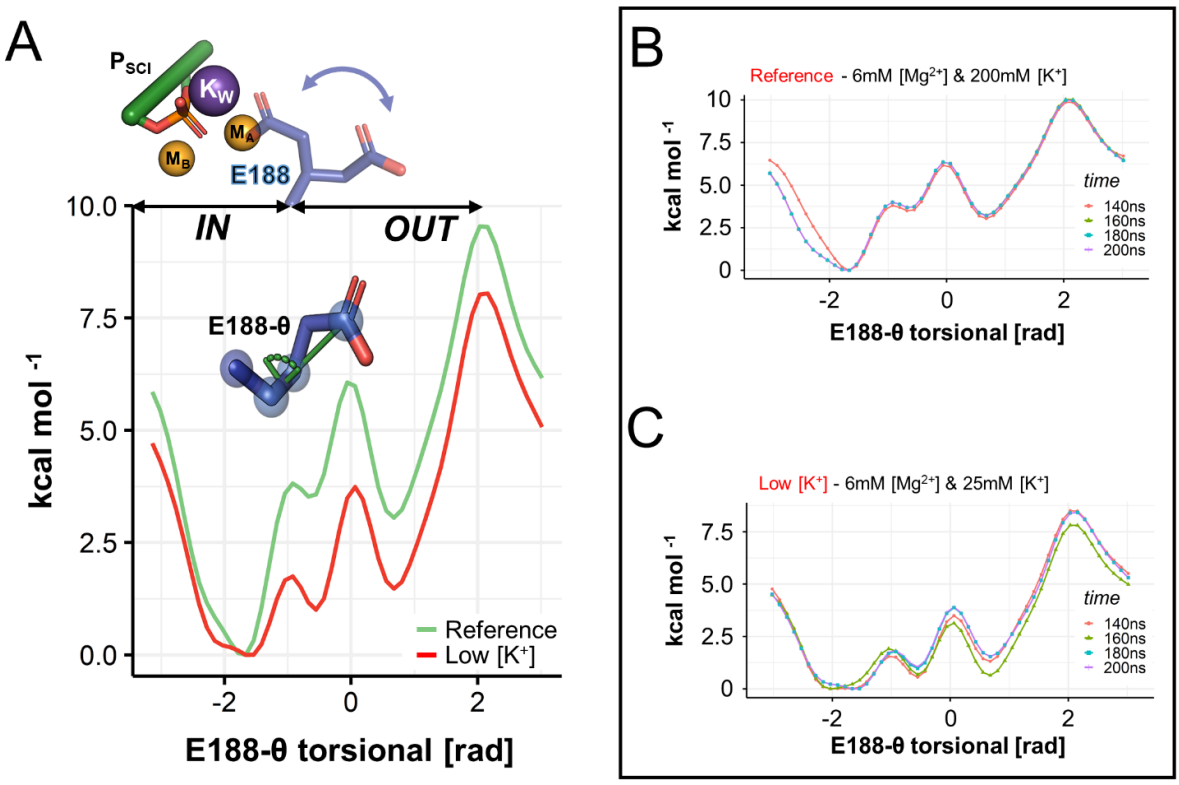


Figure B.7: Metadynamics simulations for the E188 inner ↔ outer conformational swing. ***(A)*** *Top: the conformational change of E188 is shown (blue arrows); the residue E188 (blue) and the scissile phosphate (P_{SCI}, green) are shown as sticks, while the ions K_W (purple), M_A-M_B (orange) are shown as spheres. Bottom: Free-energy profiles computed along the E188-θ torsional angle, using well-tempered metadynamics simulations. The inner ↔ outer conformational swing of E188 is ~2.3 kcal mol^{-1} less favored at reference (green) than at low (red) K^+ concentration.* ***(B)*** *The free energy for the inner ↔ outer conformational of E188 is estimated as a function of time over non-overlapping windows of ~20 ns during the last ~60 ns of metadynamics (viz. from ~140 to ~200 ns) performed at optimal ionic strength (6mM [Mg^{2+}] and 200mM [K^+]).* ***(C)*** *The free energy is estimated as in panel (A) along metadynamics simulations at low K^+ concentrations (6mM [Mg^{2+}] and 25mM [K^+]). In both cases, after ~160 ns no significant changes in the free energy profiles are detected, suggesting that both the metadynamics simulations are converged.*

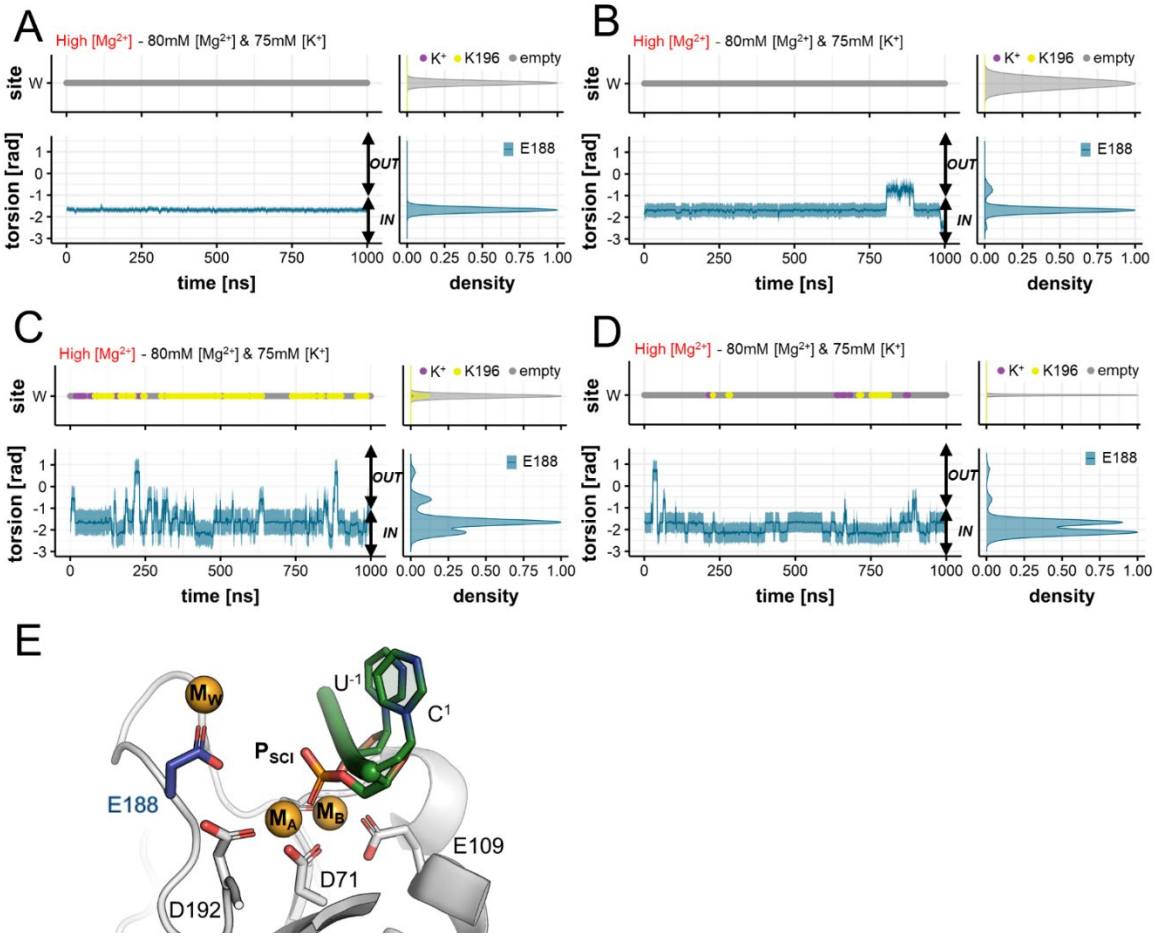


Figure B.8: MD simulations of RNase H1 in the reactant state at high Mg^{2+} concentrations. (A-D) *The occupancy of the W-site by K^+ ions and by the K196 residue together with the variation of the torsional angle E188-θ (computed between the N-Cα-Cβ-Cδ atoms) are reported for three MD replicas at high Mg^{2+} concentration (80mM [Mg^{2+}] and 75mM [K^+]), following the same scheme as in Figure B.5. Despite $E188_{IN}$ conformation is stabilized, no K^+ ions are captured at W-site because an additional Mg^{2+} metal (viz. M_W) binds with E188 closing the accessibility of the W-site to K^+ ions. Importantly, panels (A-C) report data performed using 12-6 model for Mg^{2+} ions (A-B, Aqvist; C, Allnér), while in panel (D) the Mg^{2+} ions were modeled according to 12-6-4 model from Panteva. Both the simulations performed with Allnér and Panteva force field parameters showed a weakened interaction between M_W and the E188. However, the interaction between M_W and E188 still occurs, preventing the binding of K^+ at the W-site, which is only transiently occupied by K196.* ***(E)*** *Structural representation of M_W binding preventing the E188-mediated recruitment of K_W at W-site. The active site of RNase H1 is represented according to Figure B.1. Together, these MD simulations corroborate the fundamental role of E188 in the recruitment of K_W – as shown in Figure 4.2 – and provide a rationale for the attenuation of RNase H1 catalysis at high Mg^{2+} concentrations.*

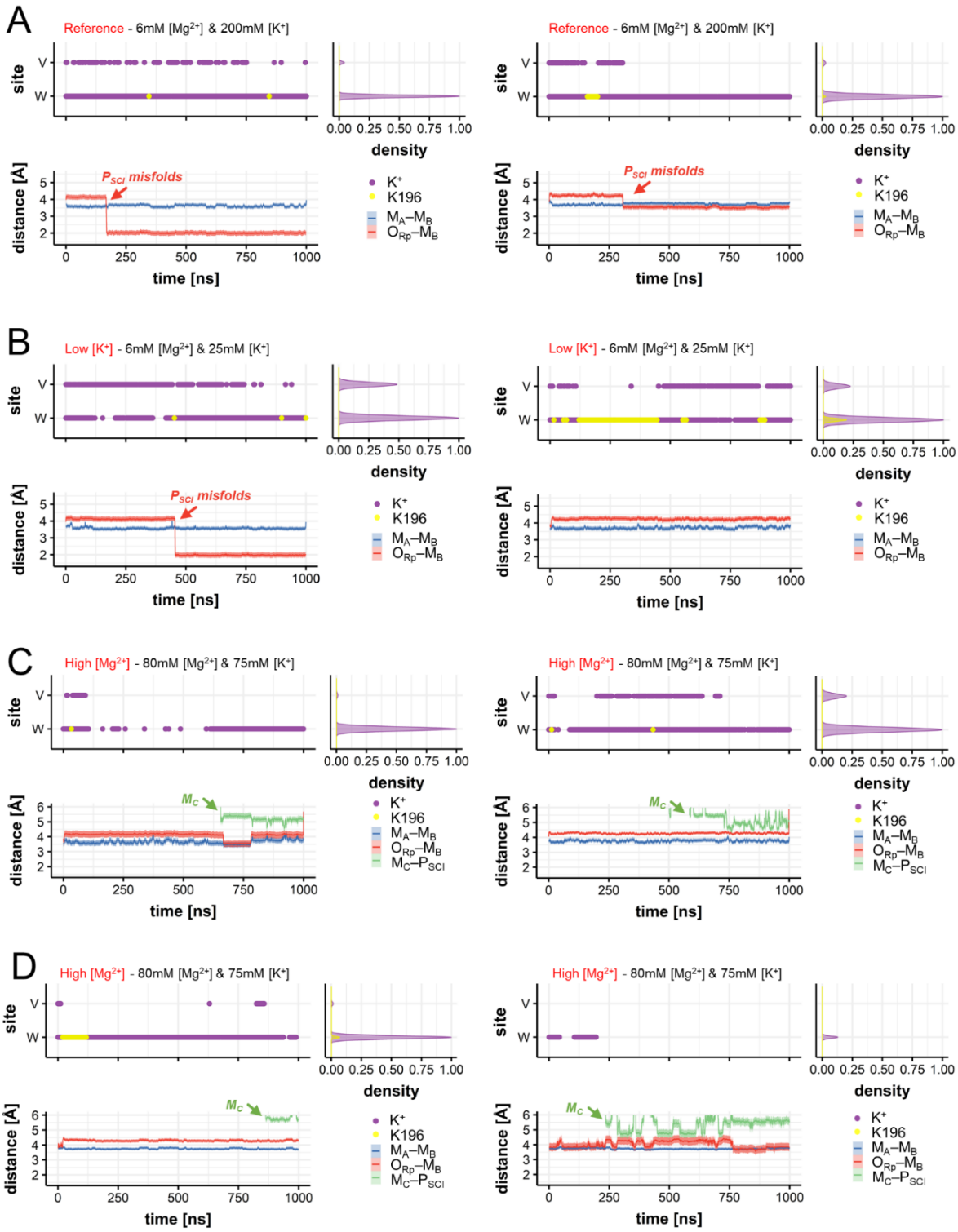


Figure B.9: MD simulations of RNase H1 in the product state. (A) *Occupancy of the W- and the V-site (upper panel) by K^+ ions (purple) and K196 side chain (yellow); and the distances (lower panel) between the catalytic ions M_A-M_B (M_A–M_B, blue trace) and between the pro-R oxygen (O_{Rp}) of the scissile phosphate and the metal M_B (O_{Rp}–M_B, red trace) during two MD replicas at optimal ionic strength (right and left panels). Normalized densities are shown on the right of each panel. After the RNA hydrolysis both K_W, K_V and K196 concurrently interact with the products as a consequence of their binding at the W- or the V-site. Despite the scissile phosphate (P_{SCI}) get misfolded when only K_W locates at the active site, this event does not favor the release of the products as the two catalytic ions M_A-M_B are stabilized at their position.* ***(B)*** *The same descriptors as in panel (A) are shown for two MD replicas at low K^+ concentrations. At this ionic strength, the binding of K^+ ions at the W- or V- site is less recurrent, and thus these ion exchange more frequently its position with the residue K196. This seems to favor the product stabilization.* ***(C)*** *The same descriptors as in panel (A)*

are shown for two MD replicas at high Mg^{2+} concentrations. Additional distance between M_C and scissile phosphate is reported as green trace (M_C–SP). Despite two different Mg^{2+} models have been used in the two MD simulations, Aqvist (left) and Allnér (right), the additional M_C gets recruited at the active site to coordinate the scissile phosphate in both the simulations. As a consequence of this binding event the distance M_A–M_B is increased, suggesting that the active site is now prone to release the reaction products. ***(D)*** *The same descriptors as in panel (A) are shown for two MD replicas at high Mg^{2+} concentrations, performed using the Panteva 12-6-4 model for Mg^{2+} ions. These simulations are in good agreement with those reported in panel (C). Together, these MD replicas sustain the role of K196, K_W, and K_V in the stabilization of the scissile phosphate for the recruitment of M_C and product release, in agreement with the findings showed in Figure 4.3.*

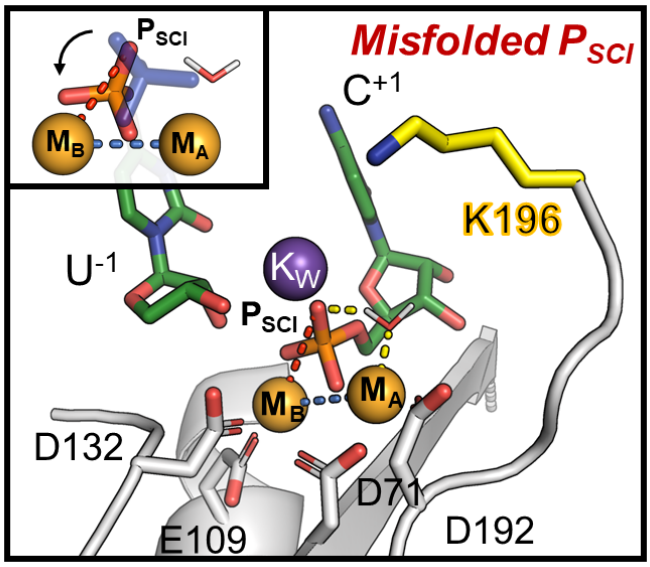


Figure B.10: Structural representation of the scissile phosphate misfolding. *MD simulations of the RNase H1 in the product state shown that the scissile phosphate (P_{SCI}) can, at times, rotate around the phosphodiester bond. Upon rotation, one bulky water molecule replaces the phosphate oxygen to coordinate the catalytic M_A. This event occurs when only K_W is bound at the active site, while K_V and K196 are unbound (main panel). Notably, the misfolded conformation observed in our MD simulations perfectly matches that captured in the K196 mutant by X-ray crystallography (top corner, orange sticks; PDBid: 6DPM). The conformation of the scissile phosphate as crystallized in the wild type RNase H1 is shown in the top corner (blue semi-transparent sticks, PDBid: 6DOG). All the representations are color-coded as in Figure 4.1.*

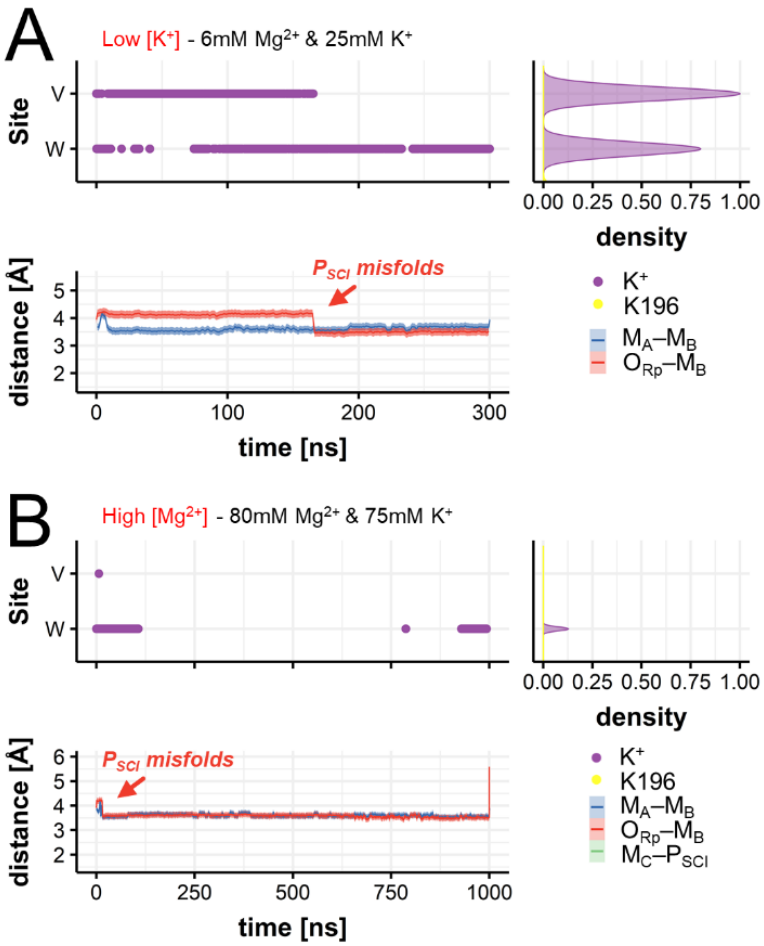


Figure B.11: MD simulations of RNase H1 mutant K196A in the product state. (A) *Occupancy of the W- and the V-site (upper panel) by K^+ ions (purple) and for the K196 (yellow); and the distances (lower panel) between the catalytic ions M_A-M_B (M_A–M_B, blue trace) and between the pro-R oxygen (O_{Rp}) of the scissile phosphate and the M_B ion (O_{Rp}–M_B, red trace) during two MD replicas at low K^+ concentration. Normalized densities are shown on the right. The absence of K196 induces the fast misfolding of the scissile phosphate (P_{SCI}).* ***(B)*** *The same descriptors as in panel (A) are reported for MD simulations at high Mg^{2+} concentration. Here, the scissile phosphate is rotated and M_C is not recruited at the active site, as opposite to what observed for the wild type RNase H1 MD simulations performed at the same high Mg^{2+} concentration (Figure 4.3C and B.8C).*

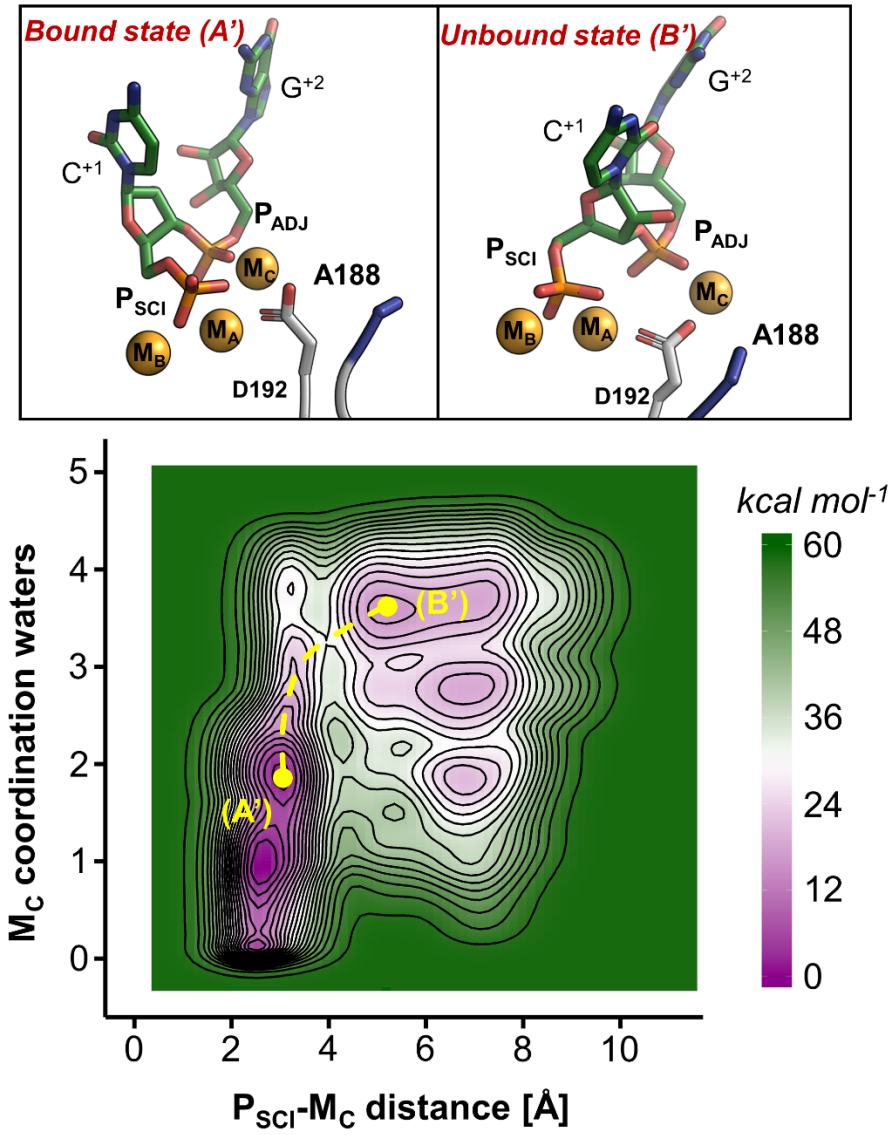


Figure B.12:. Metadynamics simulation of the RNase H1 mutant E188A in the product state. *Structural representation of the metastable state A' and B' (upper panel) corresponding to the two main minima highlighted on the free energy surface (bottom panel). The energy scale is indicated in kcal mol^{-1}, while contour traces are reported every 3 kcal mol^{-1}. In absence of the residue E188, the release of the third Mc is guided by D192, which is not as efficient as E188 in shuttling the Mc out of the active site together with the products. The free energy barrier associated with the A' → B' transition (bound → unbound state) is estimated at ~35 kcal mol^{-1}, while the state A' is ~18 kcal mol^{-1} more stable than the state B'.*

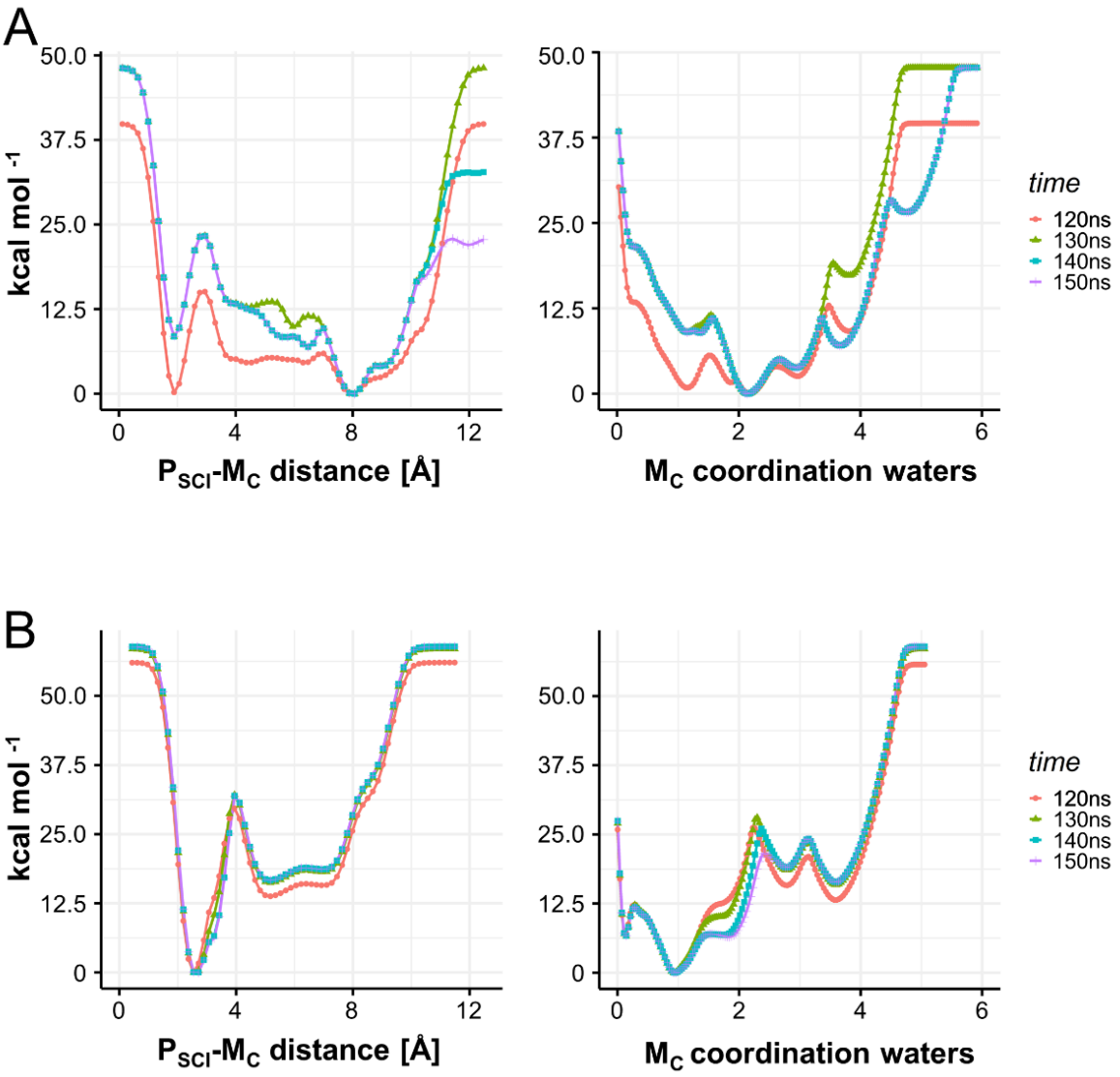


Figure B.13: The convergence of the free energy profiles for the unbinding of M_C metal. (A) *Free energy profiles projected on the distance between the scissile phosphate (P_{SCI}) and the third M_C ion (P_{SCI}–M_C distance, left panel) and on the number of M_C coordination water (M_C coordination waters, right panel) as obtained from metadynamics simulations in the wild type RNase H1. The free energy profiles are computed over non-overlapping windows of ~10 ns, considering the last ~30 ns of simulations (viz. from ~120 to 150 ns).* ***(B)*** *The free energy profiles are estimated as in panel (A) for the metadynamics simulations of the E188A mutant. In both the simulations, after ~120 ns no significant changes in the free energy profiles are detected, suggesting that both the metadynamics calculations are converged.*

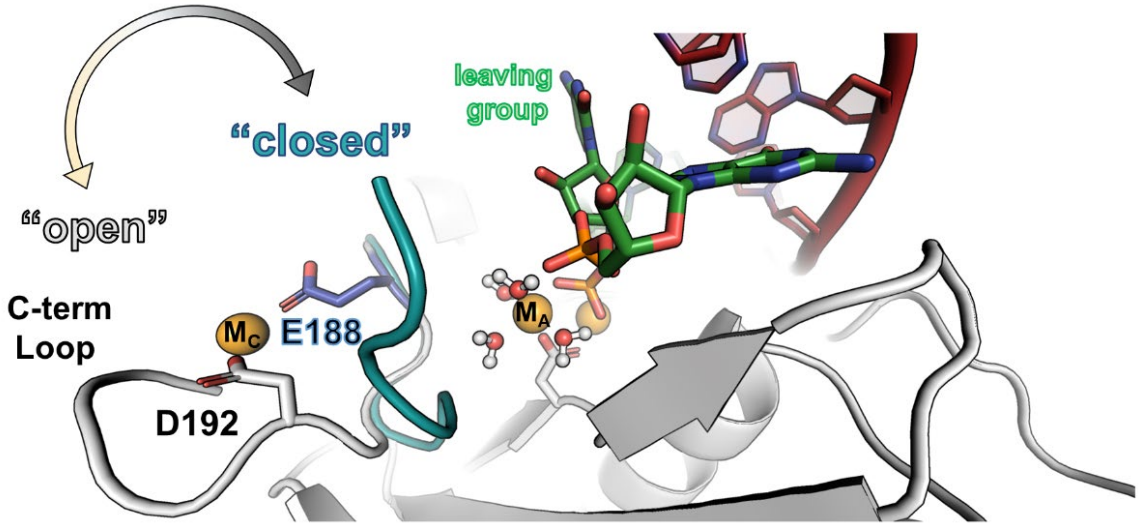


Figure B.14: Release of M_C and hydration of the active site for products' release. *RNase H1 (white) and DNA (red) are shown as cartoons, while RNA leaving group (green), the second-shell residue E188 (blue), and the catalytic residue D192 (white) are represented as sticks. The catalytic M_A-M_B ions and the third M_C metals are shown as orange spheres. As observed in the metadynamics simulations of wild type RNase H1, M_C gets coordinated by both the second-shell residue E188 and the catalytic D192. Here, the mobile C-terminal loop is in the "closed" state (blue cartoon). Upon the rotation of E188, both M_C and the adjacent phosphate are shuttled out of the active site, while D192 loses the coordination with M_A. As a result of this conformational change, the C-terminal loop rearranges in the "open" state (white cartoon), favoring the hydration of the catalytic M_A and, eventually, the release of the reaction products.*

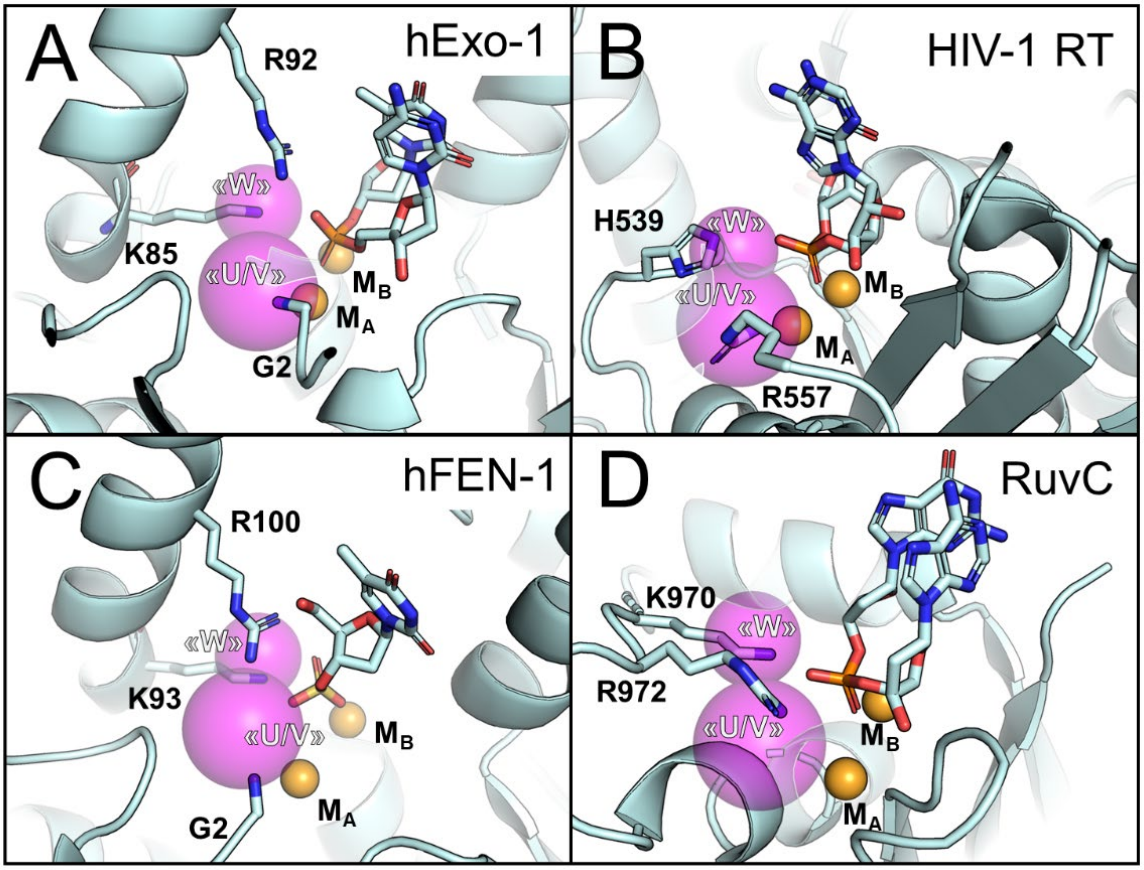


Figure B.15: Structural overlap of positive charges at the active site of RNase H1, hExo-1, HIV-1 RT, and hFEN-1. *The binding site "W", "U", and "V" (purple semi-transparent spheres) at the active site of RNase H1 are superimposed on the reaction center of* ***A)*** *of hExo-1,* ***B)*** *HIV-1 reverse transcriptase (RT),* ***C)*** *hFEN-1, and* ***D)*** *RuvC catalytic domain of CRISPR-Cas9 system (cyan cartoon). The catalytic ions M_A-M_B (orange) are shown as spheres, while the substrate/product (cyan) is represented as sticks. Remarkably, the position of structured residues (cyan sticks) located at the core of either hExo-1 (G2, K85, R95), or HIV-1 RT (R539, R557), or hFEN-1 (G2, K93, R100), or RuvC (K970, R972) well superimpose with the additional metals' binding sites "W", "U", and "V" at the active site of RNase H1.*

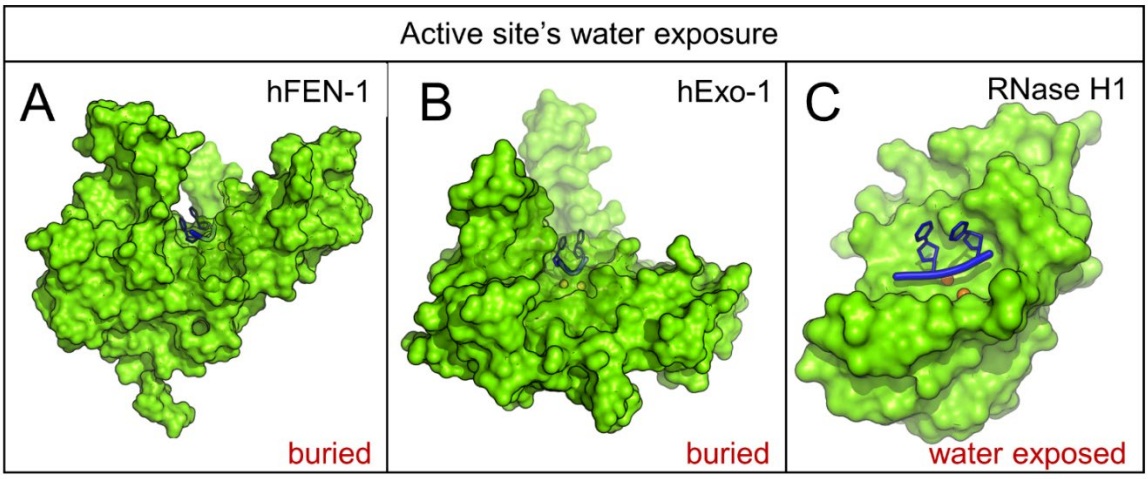


Figure B.16: The active site of RNase H1 is highly hydrated as compared to other nucleases. *The surface (green), the catalytic metals (orange spheres), and the reaction substrate (blue cartoon) at the catalytic center are shown for* ***(A)*** *hExo-1,* ***(B)*** *hFEN-1, and* ***(C)*** *for RNase H1. In the first two nucleases, the active site is buried inside the protein (transparent green surface), while in RNase H1 the active site is completely exposed to the bulk solvent and ions.*

Bibliography

(1) Crick, F. H. On Protein Synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163.

(2) Ramakrishnan, V. Ribosome Structure and the Mechanism of Translation. *Cell* **2002**, *108* (4), 557–572. https://doi.org/10.1016/s0092-8674(02)00619-0.

(3) Mattick, J. S.; Makunin, I. V. Non-Coding RNA. *Hum. Mol. Genet.* **2006**, *15 Spec No* (1), 17–29. https://doi.org/10.1093/hmg/ddl046.

(4) Clamp, M.; Fry, B.; Kamal, M.; Xie, X.; Cuff, J.; Lin, M. F.; Kellis, M.; Lindblad-Toh, K.; Lander, E. S. Distinguishing Protein-Coding and Noncoding Genes in the Human Genome. *Proc. Natl. Acad. Sci.* **2007**, *104* (49), 19428–19433. https://doi.org/10.1073/pnas.0709013104.

(5) Morris, K. V.; Mattick, J. S. The Rise of Regulatory RNA. *Nat. Rev. Genet.* **2014**, *15* (6), 423–437. https://doi.org/10.1038/nrg3722.

(6) Falese, J. P.; Donlic, A.; Hargrove, A. E. Targeting RNA with Small Molecules: From Fundamental Principles towards the Clinic. *Chem. Soc. Rev.* **2021**, *50* (4), 2224–2243. https://doi.org/10.1039/D0CS01261K.

(7) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-like Small Molecules. *Nat. Rev. Drug Discov.* **2018**, *17* (8), 547–558. https://doi.org/10.1038/nrd.2018.93.

(8) Rizvi, N. F.; Smith, G. F. RNA as a Small Molecule Druggable Target. *Bioorganic Med. Chem. Lett.* **2017**, *27* (23), 5083–5088. https://doi.org/10.1016/j.bmcl.2017.10.052.

(9) Donlic, A.; Hargrove, A. E. Targeting RNA in Mammalian Systems with Small Molecules. *Wiley Interdiscip. Rev. RNA* **2018**, *9* (4), 1–21. https://doi.org/10.1002/wrna.1477.

(10) Hombach, S.; Kretz, M. *Non-Coding RNAs: Classification, Biology and Functioning*; Slaby, O., Calin, G. A., Eds.; Advances in Experimental Medicine

and Biology; Springer International Publishing: Cham, 2016; Vol. 937. https://doi.org/10.1007/978-3-319-42059-2.

(11) Chillón, I.; Marcia, M. The Molecular Structure of Long Non-Coding RNAs: Emerging Patterns and Functional Implications. *Crit. Rev. Biochem. Mol. Biol.* **2020**, *55* (6), 662–690. https://doi.org/10.1080/10409238.2020.1828259.

(12) Mercer, T. R.; Dinger, M. E.; Mattick, J. S. Long Non-Coding RNAs: Insights into Functions. *Nat. Rev. Genet.* **2009**, *10* (3), 155–159. https://doi.org/10.1038/nrg2521.

(13) Wahlestedt, C. Targeting Long Non-Coding RNA to Therapeutically Upregulate Gene Expression. *Nature Reviews Drug Discovery*. 2013, pp 433–446. https://doi.org/10.1038/nrd4018.

(14) Arun, G.; Diermeier, S. D.; Spector, D. L. Therapeutic Targeting of Long Non-Coding RNAs in Cancer. *Trends in Molecular Medicine*. Elsevier Ltd March 1, 2018, pp 257–277. https://doi.org/10.1016/j.molmed.2018.01.001.

(15) Breaker, R. R. Prospects for Riboswitch Discovery and Analysis. *Mol. Cell* **2011**, *43* (6), 867–879. https://doi.org/10.1016/j.molcel.2011.08.024.

(16) Lünse, C. E.; Schmidt, M. S.; Wittmann, V.; Mayer, G. Carba-Sugars Activate the GlmS-Riboswitch of Staphylococcus Aureus. *ACS Chem. Biol.* **2011**, *6* (7), 675–678. https://doi.org/10.1021/cb200016d.

(17) Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; De Jesus, R.; Pan, W.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R.; Sheth, P. R.; Gill, C. J.; Tang, H.; Roemer, T. Selective Small-Molecule Inhibition of an RNA Structural Element. *Nature* **2015**, *526* (7575), 672–677. https://doi.org/10.1038/nature15542.

(18) Blount, K. F.; Megyola, C.; Plummer, M.; Osterman, D.; O'Connell, T.; Aristoff,

P.; Quinn, C.; Chrusciel, R. A.; Poel, T. J.; Schostarez, H. J.; Stewart, C. A.; Walker, D. P.; Wuts, P. G. M.; Breaker, R. R. Novel Riboswitch-Binding Flavin Analog That Protects Mice against Clostridium Difficile Infection without Inhibiting Cecal Flora. *Antimicrob. Agents Chemother.* **2015**, *59* (9), 5736–5746. https://doi.org/10.1128/AAC.01282-15.

(19) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. In Vitro Splicing of the Ribosomal RNA Precursor of Tetrahymena: Involvement of a Guanosine Nucleotide in the Excision of the Intervening Sequence. *Cell* **1981**, *27* (3), 487–496. https://doi.org/10.1016/0092-8674(81)90390-1.

(20) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell* **1983**, *35* (3), 849–857. https://doi.org/10.1016/0092-8674(83)90117-4.

(21) Thomas, J. M.; Perrin, D. M. Probing General Acid Catalysis in the Hammerhead Ribozyme. *J. Am. Chem. Soc.* **2009**, *131* (3), 1135–1143. https://doi.org/10.1021/ja807790e.

(22) Wilson, T. J.; Liu, Y.; Li, N.-S.; Dai, Q.; Piccirilli, J. A.; Lilley, D. M. J. Comparison of the Structures and Mechanisms of the Pistol and Hammerhead Ribozymes. *J. Am. Chem. Soc.* **2019**, *141* (19), 7865–7875. https://doi.org/10.1021/jacs.9b02141.

(23) Lee, T. S.; López, C. S.; Giambaşu, G. M.; Martick, M.; Scott, W. G.; York, D. M. Role of Mg2+ in Hammerhead Ribozyme Catalysis from Molecular Simulation. *J. Am. Chem. Soc.* **2008**, *130* (10), 3053–3064. https://doi.org/10.1021/ja076529e.

(24) Pyle, A. M. Group II Intron Self-Splicing. *Annu. Rev. Biophys.* **2016**, *45* (1), 183–205. https://doi.org/10.1146/annurev-biophys-062215-011149.

(25) Galej, W. P.; Toor, N.; Newman, A. J.; Nagai, K. Molecular Mechanism and Evolution of Nuclear Pre-MRNA and Group II Intron Splicing: Insights from Cryo-Electron Microscopy Structures. *Chemical Reviews*. 2018, pp 4156–4176. https://doi.org/10.1021/acs.chemrev.7b00499.

(26) Lambowitz, A. M.; Zimmerly, S. Mobile Group II Introns. *Annu. Rev. Genet.* **2004**, *38* (1), 1–35. https://doi.org/10.1146/annurev.genet.38.072902.091600.

(27) Lambowitz, A. M.; Belfort, M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. In *Mobile DNA III*; ASM Press: Washington, DC, USA, 2015; pp 1209–1236. https://doi.org/10.1128/9781555819217.ch52.

(28) Fedorova, O.; Jagdmann, G. E.; Adams, R. L.; Yuan, L.; Van Zandt, M. C.; Pyle, A. M. Small Molecules That Target Group II Introns Are Potent Antifungal Agents. *Nat. Chem. Biol.* **2018**, *14* (12), 1073–1078. https://doi.org/10.1038/s41589-018-0142-0.

(29) Augustin, S.; Müller, M. W.; Schweyen, R. J. Reverse Self-Splicing of Group II Intron RNAs in Vitro. *Nature* **1990**, *343* (6256), 383–386. https://doi.org/10.1038/343383a0.

(30) Cousineau, B.; Smith, D.; Lawrence-Cavanagh, S.; Mueller, J. E.; Yang, J.; Mills, D.; Manias, D.; Dunny, G.; Lambowitz, A. M.; Belfort, M. Retrohoming of a Bacterial Group II Intron. *Cell* **1998**, *94* (4), 451–462. https://doi.org/10.1016/S0092-8674(00)81586-X.

(31) Nisa-Martinez, R.; Jimenez-Zurdo, J. I.; Martinez-Abarca, F.; Munoz-Adelantado, E.; Toro, N. Dispersion of the RmInt1 Group II Intron in the Sinorhizobium Meliloti Genome upon Acquisition by Conjugative Transfer. *Nucleic Acids Res.* **2006**, *35* (1), 214–222. https://doi.org/10.1093/nar/gkl1072.

(32) Lambowitz, A. M.; Zimmerly, S. Group II Introns: Mobile Ribozymes That Invade DNA. *Cold Spring Harb. Perspect. Biol.* **2011**, *3* (8), a003616–a003616. https://doi.org/10.1101/cshperspect.a003616.

(33) Guo, H. Group II Introns Designed to Insert into Therapeutically Relevant DNA Target Sites in Human Cells. *Science (80-.).* **2000**, *289* (5478), 452–457. https://doi.org/10.1126/science.289.5478.452.

(34) Toor, N.; Keating, K. S.; Taylor, S. D.; Pyle, A. M. Crystal Structure of a Self-Spliced Group II Intron. *Science (80-.).* **2008**, *320* (5872), 77–82.

https://doi.org/10.1126/science.1153803.

(35) Marcia, M.; Pyle, A. M. Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* **2012**, *151* (3), 497–507. https://doi.org/10.1016/j.cell.2012.09.033.

(36) Marcia, M.; Pyle, A. M. Principles of Ion Recognition in RNA: Insights from the Group II Intron Structures. *RNA* **2014**, *20* (4), 516–527. https://doi.org/10.1261/rna.043414.113.

(37) Marcia, M.; Somarowthu, S.; Pyle, A. Now on Display: A Gallery of Group II Intron Structures at Different Stages of Catalysis. *Mob. DNA* **2013**, *4* (1), 14. https://doi.org/10.1186/1759-8753-4-14.

(38) Toor, N.; Keating, K. S.; Fedorova, O.; Rajashankar, K.; Wang, J.; Pyle, A. M. Tertiary Architecture of the Oceanobacillus Iheyensis Group II Intron. *Rna* **2010**, *16* (1), 57–69. https://doi.org/10.1261/rna.1844010.

(39) Zimmerly, S.; Guo, H.; Eskest, R.; Yang, J.; Perlman, P. S.; Lambowitz, A. M. A Group II Intron RNA Is a Catalytic Component of a DNA Endonuclease Involved in Intron Mobility. *Cell* **1995**, *83* (4), 529–538. https://doi.org/10.1016/0092-8674(95)90092-6.

(40) Jacquier, A.; Michel, F. Multiple Exon-Binding Sites in Class II Self-Splicing Introns. *Cell* **1987**, *50* (1), 17–29. https://doi.org/10.1016/0092-8674(87)90658-1.

(41) Fedorova, O.; Mitros, T.; Pyle, A. M. Domains 2 and 3 Interact to Form Critical Elements of the Group II Intron Active Site. *J. Mol. Biol.* **2003**, *330* (2), 197–209. https://doi.org/10.1016/S0022-2836(03)00594-1.

(42) Costa, M.; Dème, E.; Jacquier, A.; Michel, F. Multiple Tertiary Interactions Involving Domain II of Group II Self-Splicing Introns 1 1Edited by M. Yaniv. *J. Mol. Biol.* **1997**, *267* (3), 520–536. https://doi.org/10.1006/jmbi.1996.0882.

(43) Fedorova, O.; Pyle, A. M. Linking the Group II Intron Catalytic Domains: Tertiary Contacts and Structural Features of Domain 3. *EMBO J.* **2005**, *24* (22),

3906–3916. https://doi.org/10.1038/sj.emboj.7600852.

(44) 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.* **2016**, *138* (33), 10374–10377. https://doi.org/10.1021/jacs.6b01363.

(45) de Lencastre, A.; Hamill, S.; Pyle, A. M. A Single Active-Site Region for a Group II Intron. *Nat. Struct. Mol. Biol.* **2005**, *12* (7), 626–627. https://doi.org/10.1038/nsmb957.

(46) Pyle, A. M. The Tertiary Structure of Group II Introns: Implications for Biological Function and Evolution. *Crit. Rev. Biochem. Mol. Biol.* **2010**, *45* (3), 215–232. https://doi.org/10.3109/10409231003796523.

(47) Peebles, C. L.; Perlman, P. S.; Mecklenburg, K. L.; Petrillo, M. L.; Tabor, J. H.; Jarrell, K. A.; Cheng, H.-L. A Self-Splicing RNA Excises an Intron Lariat. *Cell* **1986**, *44* (2), 213–223. https://doi.org/10.1016/0092-8674(86)90755-5.

(48) Jarrell, K. A.; Peebles, C. L.; Dietrich, R. C.; Romiti, S. L.; Perlman, P. S. Group II Intron Self-Splicing. Alternative Reaction Conditions Yield Novel Products. *J. Biol. Chem.* **1988**, *263* (7), 3432–3439. https://doi.org/10.1016/S0021-9258(18)69089-1.

(49) Marcia, M.; Manigrasso, J.; De Vivo, M. Finding the Ion in the RNA-Stack: Can Computational Models Accurately Predict Key Functional Elements in Large Macromolecular Complexes? *J. Chem. Inf. Model.* **2021**, *61* (6), 2511–2515. https://doi.org/10.1021/acs.jcim.1c00572.

(50) Muntoni, F.; Torelli, S.; Ferlini, A. Dystrophin and Mutations: One Gene, Several Proteins, Multiple Phenotypes. *Lancet Neurol.* **2003**, *2* (12), 731–740. https://doi.org/10.1016/S1474-4422(03)00585-4.

(51) Disset, A.; Bourgeois, C. F.; Benmalek, N.; Claustres, M.; Stevenin, J.; Tuffery-Giraud, S. An Exon Skipping-Associated Nonsense Mutation in the Dystrophin Gene Uncovers a Complex Interplay between Multiple Antagonistic Splicing

Elements. *Hum. Mol. Genet.* **2006**, *15* (6), 999–1013. https://doi.org/10.1093/hmg/ddl015.

(52) Lefebvre, S.; Bürglen, L.; Reboullet, S.; Clermont, O.; Burlet, P.; Viollet, L.; Benichou, B.; Cruaud, C.; Millasseau, P.; Zeviani, M.; Le Paslier, D.; Frézal, J.; Cohen, D.; Weissenbach, J.; Munnich, A.; Melki, J. Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene. *Cell* **1995**, *80* (1), 155–165. https://doi.org/10.1016/0092-8674(95)90460-3.

(53) Lorson, C. L.; Hahnen, E.; Androphy, E. J.; Wirth, B. A Single Nucleotide in the SMN Gene Regulates Splicing and Is Responsible for Spinal Muscular Atrophy. *Proc. Natl. Acad. Sci.* **1999**, *96* (11), 6307–6311. https://doi.org/10.1073/pnas.96.11.6307.

(54) Scotti, M. M.; Swanson, M. S. RNA Mis-Splicing in Disease. *Nat. Rev. Genet.* **2016**, *17* (1), 19–32. https://doi.org/10.1038/nrg.2015.3.

(55) Stein, C. A.; Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Mol. Ther.* **2017**, *25* (5), 1069–1075. https://doi.org/10.1016/j.ymthe.2017.03.023.

(56) Crooke, S. T.; Baker, B. F.; Crooke, R. M.; Liang, X. Antisense Technology: An Overview and Prospectus. *Nat. Rev. Drug Discov.* **2021**, *20* (6), 427–453. https://doi.org/10.1038/s41573-021-00162-z.

(57) Ratni, H.; Ebeling, M.; Baird, J.; Bendels, S.; Bylund, J.; Chen, K. S.; Denk, N.; Feng, Z.; Green, L.; Guerard, M.; Jablonski, P.; Jacobsen, B.; Khwaja, O.; Kletzl, H.; Ko, C. P.; Kustermann, S.; Marquet, A.; Metzger, F.; Mueller, B.; Naryshkin, N. A.; Paushkin, S. V.; Pinard, E.; Poirier, A.; Reutlinger, M.; Weetall, M.; Zeller, A.; Zhao, X.; Mueller, L. Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular Atrophy (SMA). *J. Med. Chem.* **2018**, *61* (15), 6501–6517. https://doi.org/10.1021/acs.jmedchem.8b00741.

(58) Sheridan, C. First Small-Molecule Drug Targeting RNA Gains Momentum. *Nat. Biotechnol.* **2021**, *39* (1), 6–8. https://doi.org/10.1038/s41587-020-00788-1.

(59) Genna, V.; Donati, E.; De Vivo, M. The Catalytic Mechanism of DNA and RNA Polymerases. *ACS Catal.* **2018**, *8* (12), 11103–11118. https://doi.org/10.1021/acscatal.8b03363.

(60) Nowotny, M. Retroviral Integrase Superfamily: The Structural Perspective. *EMBO Rep.* **2009**, *10* (2), 144–151. https://doi.org/10.1038/embor.2008.256.

(61) Broccoli, S.; Rallu, F.; Sanscartler, P.; Cerritelli, S. M.; Crouch, R. J.; Drolet, M. Effects of RNA Polymerase Modifications on Transcription-Induced Negative Supercoiling and Associated R-Loop Formation. *Mol. Microbiol.* **2004**, *52* (6), 1769–1779. https://doi.org/10.1111/j.1365-2958.2004.04092.x.

(62) Yang, W.; Hendrickson, W. A.; Crouch, R. J.; Satow, Y. Structure of Ribonuclease H Phased at 2 Å Resolution by MAD Analysis of the Selenomethionyl Protein. *Science (80-.).* **1990**, *249* (4975), 1398–1405. https://doi.org/10.1126/science.2169648.

(63) Katayanagi, K.; Miyagawa, M.; Matsushima, M.; Ishikawa, M.; Kanaya, S.; Ikehara, M.; Matsuzaki, T.; Morikawa, K. Three-Dimensional Structure of Ribonuclease H from E. Coli. *Nature* **1990**, *347* (6290), 306–309. https://doi.org/10.1038/347306a0.

(64) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester Cleavage in Ribonuclease H Occurs via an Associative Two-Metal-Aided Catalytic Mechanism. *J. Am. Chem. Soc.* **2008**, *130* (33), 10955–10962. https://doi.org/10.1021/ja8005786.

(65) Ganguly, A.; Thaplyal, P.; Rosta, E.; Bevilacqua, P. C.; Hammes-Schiffer, S. Quantum Mechanical/Molecular Mechanical Free Energy Simulations of the Self-Cleavage Reaction in the Hepatitis Delta Virus Ribozyme. *J. Am. Chem. Soc.* **2014**, *136* (4), 1483–1496. https://doi.org/10.1021/ja4104217.

(66) Ho, M. H.; De Vivo, M.; Dal Peraro, M.; Klein, M. L. Understanding the Effect of Magnesium Ion Concentration on the Catalytic Activity of Ribonuclease H through Computation: Does a Third Metal Binding Site Modulate Endonuclease Catalysis? *J. Am. Chem. Soc.* **2010**, *132* (39), 13702–13712.

https://doi.org/10.1021/ja102933y.

(67) Steitz, T. A.; Steitz, J. A. A General Two-Metal-Ion Mechanism for Catalytic RNA. *Proc. Natl. Acad. Sci.* **1993**, *90* (14), 6498–6502. https://doi.org/10.1073/pnas.90.14.6498.

(68) Yang, W. *Nucleases: Diversity of Structure, Function and Mechanism*; 2011; Vol. 44. https://doi.org/10.1017/S0033583510000181.

(69) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. *Acc Chem Res* **2015**, *48* (2), 220–228. https://doi.org/10.1021/ar500314j.

(70) Schmidt, B. H.; Burgin, A. B.; Deweese, J. E.; Osheroff, N.; Berger, J. M. A Novel and Unified Two-Metal Mechanism for DNA Cleavage by Type II and IA Topoisomerases. *Nature* **2010**, *465* (7298), 641–644. https://doi.org/10.1038/nature08974.

(71) Riccardi, L.; Genna, V.; De Vivo, M. Metal–Ligand Interactions in Drug Design. *Nat. Rev. Chem.* **2018**, *2* (7), 100–112. https://doi.org/10.1038/s41570-018-0018-6.

(72) Samara, N. L.; Yang, W. Cation Trafficking Propels RNA Hydrolysis. *Nat. Struct. Mol. Biol.* **2018**, *25* (8), 715–721. https://doi.org/10.1038/s41594-018-0099-4.

(73) Genna, V.; Colombo, M.; De Vivo, M.; Marcia, M. Second-Shell Basic Residues Expand the Two-Metal-Ion Architecture of DNA and RNA Processing Enzymes. *Structure* **2018**, *26* (1), 40-50.e2. https://doi.org/10.1016/j.str.2017.11.008.

(74) Genna, V.; Marcia, M.; De Vivo, M. A Transient and Flexible Cation−π Interaction Promotes Hydrolysis of Nucleic Acids in DNA and RNA Nucleases. *J. Am. Chem. Soc.* **2019**, *141* (27), 10770–10776. https://doi.org/10.1021/jacs.9b03663.

(75) Vidossich, P.; Castañeda Moreno, L. E.; Mota, C.; de Sanctis, D.; Miscione, G.

Pietro; De Vivo, M. Functional Implications of Second-Shell Basic Residues for DUTPase DR2231 Enzymatic Specificity. *ACS Catal.* **2020**, 13825–13833. https://doi.org/10.1021/acscatal.0c04148.

(76) Donati, E.; Genna, V.; De Vivo, M. Recruiting Mechanism and Functional Role of a Third Metal Ion in the Enzymatic Activity of 5′ Structure-Specific Nucleases. *J. Am. Chem. Soc.* **2020**, *142* (6), 2823–2834. https://doi.org/10.1021/jacs.9b10656.

(77) Gao, Y.; Yang, W. Capture of a Third Mg2+ Is Essential for Catalyzing DNA Synthesis. *Science (80-.).* **2016**, *352* (6291), 1334–1337. https://doi.org/10.1126/science.aad9633.

(78) Perera, L.; Freudenthal, B. D.; Beard, W. A.; Shock, D. D.; Pedersen, L. G.; Wilson, S. H.; Broyde, S. Requirement for Transient Metal Ions Revealed through Computational Analysis for DNA Polymerase Going in Reverse. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (38), E5228–E5236. https://doi.org/10.1073/pnas.1511207112.

(79) Stevens, D. R.; Hammes-Schiffer, S. Exploring the Role of the Third Active Site Metal Ion in DNA Polymerase η with QM/MM Free Energy Simulations. *J. Am. Chem. Soc.* **2018**, *140* (28), 8965–8969. https://doi.org/10.1021/jacs.8b05177.

(80) Genna, V.; Vidossich, P.; Ippoliti, E.; Carloni, P.; De Vivo, M. A Self-Activated Mechanism for Nucleic Acid Polymerization Catalyzed by DNA/RNA Polymerases. *J. Am. Chem. Soc.* **2016**, *138* (44), 14592–14598. https://doi.org/10.1021/jacs.6b05475.

(81) Manigrasso, J.; Chillón, I.; Genna, V.; Vidossich, P.; Somarowthu, S.; Pyle, A. M.; De Vivo, M.; Marcia, M. Visualizing Group II Intron Dynamics between the First and Second Steps of Splicing. *Nat. Commun.* **2020**, *11* (1). https://doi.org/10.1038/s41467-020-16741-4.

(82) Manigrasso, J.; De Vivo, M.; Palermo, G. Controlled Trafficking of Multiple and Diverse Cations Prompts Nucleic Acid Hydrolysis. *ACS Catal.* **2021**, *11* (14), 8786–8797. https://doi.org/10.1021/acscatal.1c01825.

(83) Oppenheimer, J.; Born, M. On the Quantum Theory of Molecules. *Ann. Phys.* **1927**, *84*, 458.

(84) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci.* **2005**, *102* (19), 6679–6685. https://doi.org/10.1073/pnas.0408930102.

(85) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061. https://doi.org/10.1021/acs.jmedchem.5b01684.

(86) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9* (1), 71. https://doi.org/10.1186/1741-7007-9-71.

(87) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D. Current Status of Protein Force Fields for Molecular Dynamics Simulations BT - Molecular Modeling of Proteins; Kukol, A., Ed.; Springer New York: New York, NY, 2015; pp 47–71. https://doi.org/10.1007/978-1-4939-1465-4_3.

(88) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science (80-.).* **2012**, *338* (6110), 1042–1046. https://doi.org/10.1126/science.1219021.

(89) Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci.* **2014**, *111* (45), 15873–15880. https://doi.org/10.1073/pnas.1411798111.

(90) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. Challenges in Protein-Folding Simulations. *Nat. Phys.* **2010**, *6* (10), 751–758. https://doi.org/10.1038/nphys1713.

(91) Naganathan, A. N.; Muñoz, V. Scaling of Folding Times with Protein Size. *J. Am. Chem. Soc.* **2005**, *127* (2), 480–481. https://doi.org/10.1021/ja044449u.

(92) Best, R. B. A "Slow" Protein Folds Quickly in the End. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (15), 5744–5745. https://doi.org/10.1073/pnas.1303539110.

(93) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.;

Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; 2014; pp 41–53. https://doi.org/10.1109/SC.2014.9.

(94) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science (80-.).* **2011**, *334* (6055), 517–520. https://doi.org/10.1126/science.1208351.

(95) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Lerardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51* (7), 91–97. https://doi.org/10.1145/1364782.1364802.

(96) Pasquali, S.; Derreumaux, P. HiRE-RNA: A High Resolution Coarse-Grained Energy Model for RNA. *J. Phys. Chem. B* **2010**, *114* (37), 11957–11966. https://doi.org/10.1021/jp102497y.

(97) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92* (11), 3817–3829. https://doi.org/10.1529/biophysj.106.097782.

(98) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H.

Reparameterization of RNA χ Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *J. Chem. Theory Comput.* **2010**, *6* (5), 1520–1531. https://doi.org/10.1021/ct900604a.

(99) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886–2902. https://doi.org/10.1021/ct200162x.

(100) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell, A. D. Impact of 2'-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*.

(101) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for RNA Based on the Classical Drude Oscillator. *J. Comput. Chem.* **2018**, *39* (32), 2624–2646. https://doi.org/10.1002/jcc.25709.

(102) Bottaro, S.; Bussi, G.; Kennedy, S. D.; Turner, D. H.; Lindorff-Larsen, K. Conformational Ensembles of RNA Oligonucleotides from Integrating NMR and Molecular Simulations. *Sci. Adv.* **2018**, *4* (5), eaar8521. https://doi.org/10.1126/sciadv.aar8521.

(103) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *Rna* **2015**, *21* (9), 1578–1590. https://doi.org/10.1261/rna.051102.115.

(104) Banáš, P.; Mládek, A.; Otyepka, M.; Zgarbová, M.; Jurečka, P.; Svozil, D.; Lankaš, F.; Šponer, J. Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J. Chem. Theory Comput.* **2012**, *8* (7), 2448–2460. https://doi.org/10.1021/ct3001238.

(105) Dans, P. D.; Walther, J.; Gómez, H.; Orozco, M. Multiscale Simulation of DNA.

Curr. Opin. Struct. Biol. **2016**, *37*, 29–45. https://doi.org/10.1016/j.sbi.2015.11.011.

(106) Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118* (8), 4177–4338. https://doi.org/10.1021/acs.chemrev.7b00427.

(107) Bottaro, S.; Lindorff-Larsen, K. Biophysical Experiments and Biomolecular Simulations: A Perfect Match? *Science*. 2018, pp 355–360. https://doi.org/10.1126/science.aat4010.

(108) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. The Roles of Structural Dynamics in the Cellular Functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (8), 474–489. https://doi.org/10.1038/s41580-019-0136-0.

(109) Al-Hashimi, H. M.; Walter, N. G. RNA Dynamics: It Is about Time. *Curr. Opin. Struct. Biol.* **2008**, *18* (3), 321–329. https://doi.org/10.1016/j.sbi.2008.04.004.

(110) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.

(111) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in Free Energy Space. *J. Chem. Phys.* **2007**, *126* (5). https://doi.org/10.1063/1.2432340.

(112) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nat. Rev. Phys.* **2020**, *2* (4), 200–212. https://doi.org/10.1038/s42254-020-0153-0.

(113) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 1–4. https://doi.org/10.1103/PhysRevLett.100.020603.

(114) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science (80-.).* **1992**, *257* (5066), 76–80.

https://doi.org/10.1126/science.1621097.

(115) Aboul-ela, F.; Karn, J.; Varani, G. Structure of HIV-1 TAR RNA in the Absence of Ligands Reveals a Novel Conformation of the Trinucleotide Bulge. *Nucleic Acids Res.* **1996**, *24* (20), 3974–3981. https://doi.org/10.1093/nar/24.20.3974.

(116) Henriksen, N. M.; Davis, D. R.; Cheatham III, T. E. Molecular Dynamics Re-Refinement of Two Different Small RNA Loop Structures Using the Original NMR Data Suggest a Common Structure. *J. Biomol. NMR* **2012**, *53* (4), 321–339. https://doi.org/10.1007/s10858-012-9642-5.

(117) Do, T. N.; Ippoliti, E.; Carloni, P.; Varani, G.; Parrinello, M. Counterion Redistribution upon Binding of a Tat-Protein Mimic to HIV-1 TAR RNA. *J. Chem. Theory Comput.* **2012**, *8* (2), 688–694. https://doi.org/10.1021/ct2005769.

(118) Frank, A. T.; Stelzer, A. C.; Al-Hashimi, H. M.; Andricioaei, I. Constructing RNA Dynamical Ensembles by Combining MD and Motionally Decoupled NMR RDCs: New Insights into RNA Dynamics and Adaptive Ligand Recognition. *Nucleic Acids Res.* **2009**, *37* (11), 3670–3679. https://doi.org/10.1093/nar/gkp156.

(119) Salmon, L.; Bascom, G.; Andricioaei, I.; Al-Hashimi, H. M. A General Method for Constructing Atomic-Resolution RNA Ensembles Using NMR Residual Dipolar Couplings: The Basis for Interhelical Motions Revealed. *J. Am. Chem. Soc.* **2013**, *135* (14), 5457–5466. https://doi.org/10.1021/ja400920w.

(120) Stelzer, A. C.; Frank, A. T.; Kratz, J. D.; Swanson, M. D.; Gonzalez-Hernandez, M. J.; Lee, J.; Andricioaei, I.; Markovitz, D. M.; Al-Hashimi, H. M. Discovery of Selective Bioactive Small Molecules by Targeting an RNA Dynamic Ensemble. *Nat. Chem. Biol.* **2011**, *7* (8), 553–559. https://doi.org/10.1038/nchembio.596.

(121) Ganser, L. R.; Lee, J.; Rangadurai, A.; Merriman, D. K.; Kelly, M. L.; Kansal, A. D.; Sathyamoorthy, B.; Al-Hashimi, H. M. High-Performance Virtual Screening by Targeting a High-Resolution RNA Dynamic Ensemble. *Nat.*

Struct. Mol. Biol. **2018**, *25* (5), 425–434. https://doi.org/10.1038/s41594-018-0062-4.

(122) Pyle, A. M.; Lambowitz, A. M. Group II Introns: Ribozymes That Splice RNA and Invade DNA. *RNA World (eds. Gesteland, R.F., Cech. T.R. Atkins, J.F.) Cold Spring Harb. Press. Cold Spring Harb.* **2006**, 496–505.

(123) García-Rodríguez, F. M.; Barrientos-Durán, A.; Díaz-Prado, V.; Fernández-López, M.; Toro, N. Use of RmInt1, a Group IIB Intron Lacking the Intron-Encoded Protein Endonuclease Domain, in Gene Targeting. *Appl. Environ. Microbiol.* **2011**, *77* (3), 854–861. https://doi.org/10.1128/AEM.02319-10.

(124) Boudvillaln, M.; De Lencastre, A.; Pyle, A. M. A Tertiary Interaction That Links Active-Site Domains to the 5' Splice Site of a Group II Intron. *Nature* **2000**, *406* (6793), 315–318. https://doi.org/10.1038/35018589.

(125) Chillón, I.; Martínez-Abarca, F.; Toro, N. Splicing of the Sinorhizobium Meliloti RmInt1 Group II Intron Provides Evidence of Retroelement Behavior . *Nucleic Acids Res.* **2011**, *39* (3), 1095–1104. https://doi.org/10.1093/nar/gkq847.

(126) Chillón, I.; Molina-Sánchez, M. D.; Fedorova, O.; García-Rodríguez, F. M.; Martínez-Abarca, F.; Toro, N. In Vitro Characterization of the Splicing Efficiency and Fidelity of the RmInt1 Group II Intron as a Means of Controlling the Dispersion of Its Host Mobile Element. *RNA* **2014**, *20* (12), 2000–2010. https://doi.org/10.1261/rna.047407.114.

(127) Costa, M.; Walbott, H.; Monachello, D.; Westhof, E.; Michel, F. Crystal Structures of a Group II Intron Lariat Primed for Reverse Splicing. *Science (80-.).* **2016**, *354* (6316). https://doi.org/10.1126/science.aaf9258.

(128) Qu, G.; Kaushal, P. S.; Wang, J.; Shigematsu, H.; Piazza, C. L.; Agrawal, R. K.; Belfort, M.; Wang, H.-W. Structure of a Group II Intron in Complex with Its Reverse Transcriptase. *Nat. Struct. Mol. Biol.* **2016**, *23* (6), 549–557. https://doi.org/10.1038/nsmb.3220.

(129) Robart, A. R.; Chan, R. T.; Peters, J. K.; Rajashankar, K. R.; Toor, N. Crystal

Structure of a Eukaryotic Group II Intron Lariat. *Nature* **2014**, *514* (7521), 193–197. https://doi.org/10.1038/nature13790.

(130) Chan, R. T.; Peters, J. K.; Robart, A. R.; Wiryaman, T.; Rajashankar, K. R.; Toor, N. Structural Basis for the Second Step of Group II Intron Splicing. *Nat. Commun.* **2018**, *9* (1), 4676. https://doi.org/10.1038/s41467-018-06678-0.

(131) Chan, R. T.; Robart, A. R.; Rajashankar, K. R.; Pyle, A. M.; Toor, N. Crystal Structure of a Group II Intron in the Pre-Catalytic State. *Nat. Struct. Mol. Biol.* **2012**, *19* (5), 555–557. https://doi.org/10.1038/nsmb.2270.

(132) Zhao, C.; Rajashankar, K. R.; Marcia, M.; Pyle, A. M. Crystal Structure of Group II Intron Domain 1 Reveals a Template for RNA Assembly. *Nat. Chem. Biol.* **2015**, *11* (12), 967–972. https://doi.org/10.1038/nchembio.1949.

(133) Marcia, M. Using Molecular Replacement Phasing to Study the Structure and Function of RNA. In *Methods Mol Biol*; European Molecular Biology Laboratory, Grenoble Outstation, 6, rue Jules Horowitz, Grenoble Cedex 09, 38041, France, mmarcia@embl.fr., 2016; Vol. 1320, pp 233–257. https://doi.org/10.1007/978-1-4939-2763-0_15.

(134) Marcia, M.; Humphris-Narayanan, E.; Keating, K. S.; Somarowthu, S.; Rajashankar, K.; Pyle, A. M. Solving Nucleic Acid Structures by Molecular Replacement: Examples from Group II Intron Studies. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (11), 2174–2185. https://doi.org/10.1107/S0907444913013218.

(135) Marcia, M.; Pyle, A. M. Principles of Ion Recognition in RNA: Insights from the Group II Intron Structures. *Rna* **2014**, *20* (4), 516–527. https://doi.org/10.1261/rna.043414.113.

(136) Mikheeva, S.; Murray, H. L.; Zhou, H.; Turczyk, B. M.; Jarrell, K. A. Deletion of a Conserved Dinucleotide Inhibits the Second Step of Group II Intron Splicing. *RNA* **2000**, *6* (11), 1509–1515. https://doi.org/10.1017/s1355838200000972.

(137) Podar, M.; Perlman, P. S.; Padgett, R. A. Stereochemical Selectivity of Group II Intron Splicing, Reverse Splicing, and Hydrolysis Reactions. *Mol. Cell. Biol.* **1995**, *15* (8), 4466–4478. https://doi.org/10.1128/MCB.15.8.4466.

(138) Palermo, G.; Casalino, L.; Magistrato, A.; Andrew McCammon, J. Understanding the Mechanistic Basis of Non-Coding RNA through Molecular Dynamics Simulations. *J. Struct. Biol.* **2019**, *206* (3), 267–279. https://doi.org/10.1016/j.jsb.2019.03.004.

(139) Dayie, K. T.; Padgett, R. A. A Glimpse into the Active Site of a Group II Intron and Maybe the Spliceosome, Too. *RNA* **2008**, *14* (9), 1697–1703. https://doi.org/10.1261/rna.1154408.

(140) Pechlaner, M.; Donghi, D.; Zelenay, V.; Sigel, R. K. O. Protonation-Dependent Base Flipping at Neutral PH in the Catalytic Triad of a Self-Splicing Bacterial Group II Intron. *Angew. Chemie - Int. Ed.* **2015**, *54* (33), 9687–9690. https://doi.org/10.1002/anie.201504014.

(141) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. General Acid-Base Catalysis in the Mechanism of a Hepatitis Delta Virus Ribozyme. *Science* **2000**, *287* (5457), 1493–1497. https://doi.org/10.1126/science.287.5457.1493.

(142) Huppler, A.; Nikstad, L. J.; Allmann, A. M.; Brow, D. A.; Butcher, S. E. Metal Binding and Base Ionization in the U6 RNA Intramolecular Stem-Loop Structure. *Nat. Struct. Biol.* **2002**, *9* (6), 431–435. https://doi.org/10.1038/nsb800.

(143) Reiter, N. J.; Blad, H.; Abildgaard, F.; Butcher, S. E. Dynamics in the U6 RNA Intramolecular Stem-Loop: A Base Flipping Conformational Change. *Biochemistry* **2004**, *43* (43), 13739–13747. https://doi.org/10.1021/bi048815y.

(144) Keating, K. S.; Toor, N.; Perlman, P. S.; Pyle, A. M. A Structural Analysis of the Group II Intron Active Site and Implications for the Spliceosome. *RNA* **2010**, *16* (1), 1–9. https://doi.org/10.1261/rna.1791310.

(145) Roitzsch, M.; Fedorova, O.; Pyle, A. M. The 2'-OH Group at the Group II Intron

Terminus Acts as a Proton Shuttle. *Nat. Chem. Biol.* **2010**, *6* (3), 218–224. https://doi.org/10.1038/nchembio.312.

(146) Peebles, C. L.; Zhang, M.; Perlman, P. S.; Franzen, J. S. Catalytically Critical Nucleotide in Domain 5 of a Group II Intron. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (10), 4422–4426. https://doi.org/10.1073/pnas.92.10.4422.

(147) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8* (7), 2247–2254. https://doi.org/10.1021/ct3002464.

(148) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3* (2), 107–115. https://doi.org/10.1063/1.1749604.

(149) Evans, M. G.; Polanyi, M. Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Trans. Faraday Soc.* **1935**, *31* (0), 875–894. https://doi.org/10.1039/TF9353100875.

(150) de Lencastre, A.; Pyle, A. M. Three Essential and Conserved Regions of the Group II Intron Are Proximal to the 5'-Splice Site. *RNA* **2008**, *14* (1), 11–24. https://doi.org/10.1261/rna.774008.

(151) Boudvillaln, M.; De Lencastre, A.; Pyle, A. M. A Tertiary Interaction That Links Active-Site Domains to the 5' Splice Site of a Group II Intron. *Nature* **2000**, *406* (6793), 315–318. https://doi.org/10.1038/35018589.

(152) Jacquier, A.; Michel, F. Base-Pairing Interactions Involving the 5′ and 3′-Terminal Nucleotides of Group II Self-Splicing Introns. *J. Mol. Biol.* **1990**, *213* (3), 437–447. https://doi.org/https://doi.org/10.1016/S0022-2836(05)80206-2.

(153) Madhani, H. D.; Guthrie, C. Randomization-Selection Analysis of SnRNAs in Vivo: Evidence for a Tertiary Interaction in the Spliceosome. *Genes Dev.* **1994**, *8* (9), 1071–1086. https://doi.org/10.1101/gad.8.9.1071.

(154) Konarska, M. M.; Vilardell, J.; Query, C. C. Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome. *Mol. Cell* **2006**, *21* (4), 543–553. https://doi.org/10.1016/j.molcel.2006.01.017.

(155) Query, C. C.; Konarska, M. M. Suppression of Multiple Substrate Mutations by Spliceosomal Prp8 Alleles Suggests Functional Correlations with Ribosomal Ambiguity Mutants. *Mol. Cell* **2004**, *14* (3), 343–354. https://doi.org/10.1016/s1097-2765(04)00217-5.

(156) Schwer, B.; Guthrie, C. A Conformational Rearrangement in the Spliceosome Is Dependent on PRP16 and ATP Hydrolysis. *EMBO J.* **1992**, *11* (13), 5033–5039.

(157) Hardy, S. F.; Grabowski, P. J.; Padgett, R. A.; Sharp, P. A. Cofactor Requirements of Splicing of Purified Messenger RNA Precursors. *Nature* **1984**, *308* (5957), 375–377.

(158) Fabrizio, P.; Abelson, J. Two Domains of Yeast U6 Small Nuclear RNA Required for Both Steps of Nuclear Precursor Messenger RNA Splicing. *Science* **1990**, *250* (4979), 404–409. https://doi.org/10.1126/science.2145630.

(159) Ho Faix, P. Conserved Nucleotides in the Joining Segment between Domains 2 and 3 Are Important for Group II Intron Splicing, University of Pittsburgh, Pittsburgh, 1998, Vol. PhD thesis.

(160) Smith, D. J.; Query, C. C.; Konarska, M. M. "Nought May Endure but Mutability": Spliceosome Dynamics and the Regulation of Splicing. *Mol. Cell* **2008**, *30* (6), 657–666. https://doi.org/10.1016/j.molcel.2008.04.013.

(161) Chillón, I.; Marcia, M.; Legiewicz, M.; Liu, F.; Somarowthu, S.; Pyle, A. M. Native Purification and Analysis of Long RNAs. In *Methods in Enzymology*; Academic Press Inc., 2015; Vol. 558, pp 3–37. https://doi.org/10.1016/bs.mie.2015.01.008.

(162) Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Crystallogr.* **1993**, *26* (6), 795–800. https://doi.org/https://doi.org/10.1107/S0021889893005588.

(163) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **1994**, *50* (Pt 5), 760–763. https://doi.org/10.1107/S0907444994003112.

(164) Adams, P. D.; Afonine, P. V; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66* (Pt 2), 213–221. https://doi.org/10.1107/S0907444909052925.

(165) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **2004**, *60* (Pt 12 Pt 1), 2126–2132. https://doi.org/10.1107/S0907444904019158.

(166) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B. 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W375-83. https://doi.org/10.1093/nar/gkm216.

(167) Schrodinger, L. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

(168) Wang, L.; Zhang, M.; Alexov, E. DelPhiPKa Web Server: Predicting PKa of Proteins, RNAs and DNAs. *Bioinformatics* **2016**, *32* (4), 614–615. https://doi.org/10.1093/bioinformatics/btv607.

(169) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(170) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. https://doi.org/https://doi.org/10.1002/jcc.20035.

(171) Besler, B. H.; Merz Jr., K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11* (4), 431–439. https://doi.org/https://doi.org/10.1002/jcc.540110404.

(172) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041. https://doi.org/10.1021/jp8001614.

(173) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. J. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733–2748. https://doi.org/10.1021/ct400146w.

(174) Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L. Modeling the Charge Distribution at Metal Sites in Proteins for Molecular Dynamics Simulations. *J. Struct. Biol.* **2007**, *157* (3), 444–453. https://doi.org/10.1016/j.jsb.2006.10.019.

(175) Genna, V.; Carloni, P.; De Vivo, M. A Strategically Located Arg/Lys Residue Promotes Correct Base Paring During Nucleic Acid Biosynthesis in Polymerases. *J. Am. Chem. Soc.* **2018**, *140* (9), 3312–3321. https://doi.org/10.1021/jacs.7b12446.

(176) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/https://doi.org/10.1016/j.softx.2015.06.001.

(177) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (1), 116–122. https://doi.org/10.1021/ct700200b.

(178) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 14101. https://doi.org/10.1063/1.2408420.

(179) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.

(180) Krebs, W. G.; Gerstein, M. The Morph Server: A Standardized System for Analyzing and Visualizing Macromolecular Motions in a Database Framework. *Nucleic Acids Res.* **2000**, *28* (8), 1665–1675. https://doi.org/10.1093/nar/28.8.1665.

(181) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180* (10), 1961–1972. https://doi.org/https://doi.org/10.1016/j.cpc.2009.05.011.

(182) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. Cp2k: Atomistic Simulations of Condensed Matter Systems. *WIREs Comput. Mol. Sci.* **2014**, *4* (1), 15–25. https://doi.org/https://doi.org/10.1002/wcms.1159.

(183) 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* **1988**, *37* (2), 785–789. https://doi.org/10.1103/PhysRevB.37.785.

(184) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100. https://doi.org/10.1103/PhysRevA.38.3098.

(185) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.

(186) Lippert, G.; Hutter, J.; Parrinello, M. A Hybrid Gaussian and Plane Wave Density Functional Scheme. *Mol. Phys.* **1997**, *92* (3), 477–488. https://doi.org/10.1080/002689797170220.

(187) VandeVondele, J.; Hutter, J. An Efficient Orbital Transformation Method for Electronic Structure Calculations. *J. Chem. Phys.* **2003**, *118* (10), 4365–4369. https://doi.org/10.1063/1.1543154.

(188) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J. Chem. Theory Comput.* **2005**, *1* (6), 1176–1184. https://doi.org/10.1021/ct050123f.

(189) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W. Watching DNA Polymerase η Make a Phosphodiester Bond. *Nature* **2012**, *487* (7406), 196–201. https://doi.org/10.1038/nature11181.

(190) Uson, M. L.; Carl, A.; Goldgur, Y.; Shuman, S. Crystal Structure and Mutational Analysis of Mycobacterium Smegmatis FenA Highlight Active Site Amino Acids and Three Metal Ions Essential for Flap Endonuclease and 5 Exonuclease Activities. *Nucleic Acids Res.* **2018**, *46* (8), 4164–4175. https://doi.org/10.1093/nar/gky238.

(191) Prieto, J.; Redondo, P.; Merino, N.; Villate, M.; Montoya, G.; Blanco, F. J.; Molina, R. Structure of the I-SceI Nuclease Complexed with Its DsDNA Target and Three Catalytic Metal Ions. *Acta Crystallogr. Sect. Struct. Biol. Commun.* **2016**, *72*, 473–479. https://doi.org/10.1107/S2053230X16007512.

(192) Shi, Y.; Hellinga, H. W.; Beese, L. S. Interplay of Catalysis, Fidelity, Threading, and Processivity in the Exo-and Endonucleolytic Reactions of Human Exonuclease I. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (23), 6010–6015. https://doi.org/10.1073/pnas.1704845114.

(193) Nishino, T.; Morikawa, K. Structure and Function of Nucleases in DNA Repair: Shape, Grip and Blade of the DNA Scissors. *Oncogene* **2002**, *21* (58 REV. ISS. 8), 9022–9032. https://doi.org/10.1038/sj.onc.1206135.

(194) Potapov, V.; Fu, X.; Dai, N.; Corrêa, I. R.; Tanner, N. A.; Ong, J. L. Base Modifications Affecting RNA Polymerase and Reverse Transcriptase Fidelity. *Nucleic Acids Res.* **2018**, *46* (11), 5753–5763. https://doi.org/10.1093/nar/gky341.

(195) Pavlov, Y. I.; Shcherbakova, P. V.; Rogozin, I. B. Roles of DNA Polymerases in Replication, Repair, and Recombination in Eukaryotes. *Int. Rev. Cytol.* **2006**, *255* (06), 41–132. https://doi.org/10.1016/S0074-7696(06)55002-8.

(196) Pommier, Y. Drugging Topoisomerases: Lessons and Challenges. *ACS Chem. Biol.* **2013**, *8* (1), 82–95. https://doi.org/10.1021/cb300648v.

(197) Arencibia, J. M.; Brindani, N.; Franco-Ulloa, S.; Nigro, M.; Kuriappan, J. A.; Ottonello, G.; Bertozzi, S. M.; Summa, M.; Girotto, S.; Bertorelli, R.; Armirotti, A.; De Vivo, M. Design, Synthesis, Dynamic Docking, Biochemical Characterization, and in Vivo Pharmacokinetics Studies of Novel Topoisomerase II Poisons with Promising Antiproliferative Activity. *J. Med. Chem.* **2020**, *63* (7), 3508–3521. https://doi.org/10.1021/acs.jmedchem.9b01760.

(198) Zheng, L.; Jia, J.; Finger, L. D.; Guo, Z.; Zer, C.; Shen, B. Functional Regulation of FEN1 Nuclease and Its Link to Cancer. *Nucleic Acids Res.* **2011**, *39* (3), 781–794. https://doi.org/10.1093/nar/gkq884.

(199) Yang, W.; Weng, P. J.; Gao, Y. A New Paradigm of DNA Synthesis: Three-Metal-Ion Catalysis. *Cell Biosci.* **2016**, *6* (1), 1–7. https://doi.org/10.1186/s13578-016-0118-2.

(200) Borišek, J.; Magistrato, A. An Expanded Two-Zn 2+ -Ion Motif Orchestrates Pre-MRNA Maturation in the 3′-End Processing Endonuclease Machinery . *ACS Catal.* **2021**, *11* (7), 4319–4326. https://doi.org/10.1021/acscatal.0c05594.

(201) Cerritelli, S. M.; Crouch, R. J. Ribonuclease H: The Enzymes in Eukaryotes. *FEBS J.* **2009**, *276* (6), 1494–1505. https://doi.org/10.1111/j.1742-4658.2009.06908.x.

(202) Tadokoro, T.; Kanaya, S. Ribonuclease H: Molecular Diversities, Substrate Binding Domains, and Catalytic Mechanism of the Prokaryotic Enzymes. *FEBS J.* **2009**, *276* (6), 1482–1493. https://doi.org/10.1111/j.1742-4658.2009.06907.x.

(203) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. Crystal Structures of RNase H Bound to an RNA/DNA Hybrid: Substrate Specificity and Metal-Dependent Catalysis. *Cell* **2005**, *121* (7), 1005–1016. https://doi.org/10.1016/j.cell.2005.04.024.

(204) Nowotny, M.; Gaidamakov, S. A.; Ghirlando, R.; Cerritelli, S. M.; Crouch, R. J.; Yang, W. Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription. *Mol. Cell* **2007**, *28* (2), 264–276. https://doi.org/10.1016/j.molcel.2007.08.015.

(205) Nowotny, M.; Cerritelli, S. M.; Ghirlando, R.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. Specific Recognition of RNA/DNA Hybrid and Enhancement of Human RNase H1 Activity by HBD. *EMBO J.* **2008**, *27* (7), 1172–1181. https://doi.org/10.1038/emboj.2008.44.

(206) Champoux, J. J.; Schultz, S. J. Ribonuclease H: Properties, Substrate Specificity and Roles in Retroviral Reverse Transcription. *FEBS J.* **2009**, *276* (6), 1506–1516. https://doi.org/10.1111/j.1742-4658.2009.06909.x.

(207) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.* **2011**, *133* (23), 8934–8941. https://doi.org/10.1021/ja200173a.

(208) Keck, J. L.; Goedken, E. R.; Marqusee, S. Activation/Attenuation Model for RNase H. *J. Biol. Chem.* **1998**, *273* (51), 34128–34133. https://doi.org/10.1074/jbc.273.51.34128.

(209) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Comparison of Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg2+ Ion Models Commonly Used in Biomolecular Simulations. *J. Comput. Chem.* **2015**, *36* (13), 970–982. https://doi.org/10.1002/jcc.23881.

(210) Genna, V.; Gaspari, R.; Dal Peraro, M.; De Vivo, M. Cooperative Motion of a Key Positively Charged Residue and Metal Ions for DNA Replication Catalyzed by Human DNA Polymerase-η. *Nucleic Acids Res.* **2016**, *44* (6), 2827–2836. https://doi.org/10.1093/nar/gkw128.

(211) Nowotny, M.; Yang, W. Stepwise Analyses of Metal Ions in RNase H Catalysis from Substrate Destabilization to Product Release. *EMBO J.* **2006**, *25* (9), 1924–1933. https://doi.org/10.1038/sj.emboj.7601076.

(212) Golosov, A. A.; Warren, J. J.; Beese, L. S.; Karplus, M. The Mechanism of the Translocation Step in DNA Replication by DNA Polymerase I: A Computer Simulation Analysis. *Structure* **2010**, *18* (1), 83–93. https://doi.org/10.1016/j.str.2009.10.014.

(213) Da, L. T.; Pardo Avila, F.; Wang, D.; Huang, X. A Two-State Model for the Dynamics of the Pyrophosphate Ion Release in Bacterial RNA Polymerase. *PLoS Comput. Biol.* **2013**, *9* (4). https://doi.org/10.1371/journal.pcbi.1003020.

(214) Yoon, H.; Warshel, A. Simulating the Fidelity and the Three Mg Mechanism of Pol η and Clarifying the Validity of Transition State Theory in Enzyme Catalysis. *Proteins Struct. Funct. Bioinforma.* **2017**, *85* (8), 1446–1453. https://doi.org/10.1002/prot.25305.

(215) Tian, L.; Kim, M. S.; Li, H.; Wang, J.; Yang, W. Structure of HIV-1 Reverse Transcriptase Cleaving RNA in an RNA/DNA Hybrid. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (3), 507–512. https://doi.org/10.1073/pnas.1719746115.

(216) Tsutakawa, S. E.; Thompson, M. J.; Arvai, A. S.; Neil, A. J.; Shaw, S. J.; Algasaier, S. I.; Kim, J. C.; Finger, L. D.; Jardine, E.; Gotham, V. J. B.; Sarker, A. H.; Her, M. Z.; Rashid, F.; Hamdan, S. M.; Mirkin, S. M.; Grasby, J. A.; Tainer, J. A. Phosphate Steering by Flap Endonuclease 1 Promotes 5′-Flap Specificity and Incision to Prevent Genome Instability. *Nat. Commun.* **2017**, *8* (May), 1–14. https://doi.org/10.1038/ncomms15855.

(217) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59* (5), 2394–2406. https://doi.org/10.1021/acs.jcim.8b00988.

(218) 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.* **2020**, *10*, 13596–13605. https://doi.org/10.1021/acscatal.0c03566.

(219) Wilkinson, M. E.; Fica, S. M.; Galej, W. P.; Nagai, K. Structural Basis for Conformational Equilibrium of the Catalytic Spliceosome. *Mol. Cell* **2021**, *81*

(7), 1439-1452.e9. https://doi.org/10.1016/j.molcel.2021.02.021.

(220) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(221) Darian, E.; Gannett, P. M. Application of Molecular Dynamics Simulations to Spin-Labeled Oligonucleotides. *J. Biomol. Struct. Dyn.* **2005**, *22* (5), 579–593. https://doi.org/10.1080/07391102.2005.10507028.

(222) Cieplak, P.; Kollman, P. A.; Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. ...* **2000**, *21* (12), 1049–1074.

(223) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114–4127. https://doi.org/10.1021/acs.jctc.6b00186.

(224) Joung, S.; Cheatham, T. E. Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, *113* (40), 13279–13290. https://doi.org/10.1021/jp902584c.

(225) Åqvist, J. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94* (21), 8021–8024. https://doi.org/10.1021/j100384a009.

(226) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion-Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8* (4), 1493–1502. https://doi.org/10.1021/ct3000734.

(227) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Force Field for Mg2+, Mn2+, Zn2+, and Cd2+ Ions That Have Balanced Interactions with Nucleic Acids. *J.*

Phys. Chem. B **2015**, *119* (50), 15460–15470. https://doi.org/10.1021/acs.jpcb.5b10423.

(228) Casalino, L.; Palermo, G.; Abdurakhmonova, N.; Rothlisberger, U.; Magistrato, A. Development of Site-Specific Mg2+-RNA Force Field Parameters: A Dream or Reality? Guidelines from Combined Molecular Dynamics and Quantum Mechanics Simulations. *J. Chem. Theory Comput.* **2017**, *13* (1), 340–352. https://doi.org/10.1021/acs.jctc.6b00905.

(229) D.A. Case I.Y. Ben-Shalom, S. R. B. D. S. C. T. E. C. I. I. I. V. W. D. C. T. A. D. R. E. D. D. G. M. K. G. H. G. A. W. G. D. G. R. H. N. H. Y. H. S. I. A. K. T. K. T. S. L. S. L. P. L. C. L. J. L. T. L. R. L. D. J. M. K.; Kollman, P. A. AMBER 2018. *Univ. California, San Fr.*

(230) Lee, T. S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; Legrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58* (10), 2043–2050. https://doi.org/10.1021/acs.jcim.8b00462.

(231) Turq, P.; Lantelme, F.; Friedman, H. L. Brownian Dynamics: Its Application to Ionic Solutions. *J. Chem. Phys.* **1976**, *66* (7), 3039–3044. https://doi.org/10.1063/1.434317.

(232) 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.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(233) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061. https://doi.org/10.1021/acs.jmedchem.5b01684.

(234) Fica, S. M.; Mefford, M. A.; Piccirilli, J. A.; Staley, J. P. Evidence for a Group II Intron–like Catalytic Triplex in the Spliceosome. *Nat. Struct. Mol. Biol.* **2014**, *21* (5), 464–471. https://doi.org/10.1038/nsmb.2815.

(235) Turunen, J. J.; Niemelä, E. H.; Verma, B.; Frilander, M. J. The Significant Other: Splicing by the Minor Spliceosome. *Wiley Interdiscip. Rev. RNA* **2013**, *4* (1), 61–76. https://doi.org/10.1002/wrna.1141.

(236) Wilkinson, M. E.; Charenton, C.; Nagai, K. RNA Splicing by the Spliceosome. *Annual Review of Biochemistry*. 2020, pp 359–388. https://doi.org/10.1146/annurev-biochem-091719-064225.

(237) Bai, R.; Wan, R.; Wang, L.; Xu, K.; Zhang, Q.; Lei, J.; Shi, Y. Structure of the Activated Human Minor Spliceosome. *Science (80-.).* **2021**, *1220* (March), eabg0879. https://doi.org/10.1126/science.abg0879.

(238) Manigrasso, J.; Marcia, M.; De Vivo, M. Computer-Aided Design of RNA-Targeted Small Molecules: A Growing Need in Drug Discovery. *Chem* **2021**, 1–24. https://doi.org/10.1016/j.chempr.2021.05.021.

(239) Butcher, S. E.; Pyle, A. M. The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. *Acc. Chem. Res.* **2011**, *44* (12), 1302–1311. https://doi.org/10.1021/ar200098t.

(240) Wilson, D. N.; Cate, J. H. D. The Structure and Function of the Eukaryotic Ribosome. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (5), 5. https://doi.org/10.1101/cshperspect.a011536.

(241) Breaker, R. R. Riboswitches and Translation Control. *Cold Spring Harb. Perspect. Biol.* **2018**, *10* (11). https://doi.org/10.1101/cshperspect.a032797.

(242) Connelly, C. M.; Numata, T.; Boer, R. E.; Moon, M. H.; Sinniah, R. S.; Barchi, J. J.; Ferré-D'Amaré, A. R.; Schneekloth, J. S. Synthetic Ligands for PreQ 1 Riboswitches Provide Structural and Mechanistic Insights into Targeting RNA Tertiary Structure. *Nat. Commun.* **2019**, *10* (1), 1–12. https://doi.org/10.1038/s41467-019-09493-3.

(243) Costales, M. G.; Hoch, D. G.; Abegg, D.; Childs-Disney, J. L.; Velagapudi, S. P.; Adibekian, A.; Disney, M. D. A Designed Small Molecule Inhibitor of a Non-Coding RNA Sensitizes HER2 Negative Cancers to Herceptin. *J. Am. Chem.*

Soc. **2019**, *141* (7), 2960–2974. https://doi.org/10.1021/jacs.8b10558.

(244) Disney, M. D.; Winkelsas, A. M.; Velagapudi, S. P.; Southern, M.; Fallahi, M.; Childs-Disney, J. L. Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem. Biol.* **2016**, *11* (6), 1720–1728. https://doi.org/10.1021/acschembio.6b00001.

(245) Davila-Calderon, J.; Patwardhan, N. N.; Chiu, L.-Y.; Sugarman, A.; Cai, Z.; Penutmutchu, S. R.; Li, M.-L.; Brewer, G.; Hargrove, A. E.; Tolbert, B. S. IRES-Targeting Small Molecule Inhibits Enterovirus 71 Replication via Allosteric Stabilization of a Ternary Complex. *Nat. Commun.* **2020**, *11* (1), 4775. https://doi.org/10.1038/s41467-020-18594-3.

(246) Sheridan, C. First Small-Molecule Drug Targeting RNA Gains Momentum. *Nat. Biotechnol.* **2021**, *39* (1), 6–8. https://doi.org/10.1038/s41587-020-00788-1.

(247) Morgan, B. S.; Forte, J. E.; Culver, R. N.; Zhang, Y.; Hargrove, A. E. Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angew. Chemie - Int. Ed.* **2017**, *56* (43), 13498–13502. https://doi.org/10.1002/anie.201707641.