



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

**Scienze Biotecnologiche, Biocomputazionali, Farmaceutiche
e Farmacologiche**

Ciclo 38

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

Settore Scientifico Disciplinare: CHIM/08 – CHIMICA FARMACEUTICA

**EXPLORING RNA CATALYSIS AND RNA-LIGAND INTERACTIONS
THROUGH MOLECULAR SIMULATIONS**

Presentata da: *Gianfranco Martino*

Coordinatore Dottorato

Mara Laura Bolognesi

Supervisore

Marco De Vivo

Co-supervisore

Maria Laura Bolognesi

Esame finale anno 2026

Dedicata ad Antonia e Ninny
Alla gioia infinita di rispecchiarmi nei vostri occhi

Table of Contents

ACKNOWLEDGMENTS	9
ABSTRACT	10
1. INTRODUCTION TO RNA BIOLOGY AND ITS THERAPEUTIC POTENTIAL	11
1.1. OVERVIEW OF RNA BIOLOGY	11
1.1. THE CENTRAL ROLE OF ncRNA IN BIOLOGICAL SYSTEMS	12
1.1.1. microRNAs	12
1.1.2. Riboswitches	13
1.1.3. Small nuclear RNAs	14
1.2. CATALYTIC RNA	15
1.2.1. Trans cleaving ribozymes – Ribonuclease P	16
1.2.2. Cis cleaving ribozymes	17
1.2.3. Splicing ribozymes – Group II introns	17
1.3. SPLICESOMAL SPLICING: A CRITICAL REGULATORY HUB	20
1.3.1. Splicing cycle	21
1.3.2. The spliceosome's RNA catalytic core and its mechanism of action	27
1.4. SPLICING MODULATORS AS THERAPEUTIC AGENTS	29
1.4.1. Protein-targeting splicing modulators	30
1.4.2. RNA targeting splicing modulators	30
1.5. RNA IN DRUG DISCOVERY: IMPORTANCE AND THE NEED FOR COMPUTATIONAL ACCELERATION	32
2. THEORY AND METHODS	35
2.1. MOLECULAR DYNAMICS (MD) SIMULATIONS	35
2.1.1. Application of classical mechanics to a molecular system	35
2.1.2. Molecular mechanics	36
2.1.3. The statistical mechanics concept of ensemble	37

2.1.4. MD simulations	38
2.2. METADYNAMICS	40
2.3. FREE ENERGY PERTURBATION (FEP) PRINCIPLES	42
3. ADVANCEMENTS, LIMITATIONS AND USAGE OF RNA FORCE FIELD	46
3.1. ABSTRACT	48
3.2. INTRODUCTION	48
3.3. ADVANCES IN FORCE FIELD PARAMETERS FOR RNA	50
3.3.1. Bonded interactions refinement	52
3.3.2. Water models	55
3.3.3. Other FFs for RNA	56
3.4. CASE STUDIES IN RNA MD SIMULATIONS: MISCELLANEOUS TARGETS AND EMERGING QUESTIONS	57
3.4.1. Influence of ions on the dynamics of RNA	59
3.4.2. RNA folding	65
3.4.3. Integrating simulations and experimental data to unveil RNA dynamics	66
3.5. CONCLUSIONS AND FUTURE PERSPECTIVE	70
4. UNRAVELING RNA CATALYTIC MECHANISMS THROUGH ADVANCED SIMULATIONS	72
4.1. ABSTRACT	73
4.2. INTRODUCTION	73
4.3. RESULTS	77
4.4. DISCUSSION	88
4.5. MATERIAL AND METHODS	94
4.5.1. Structural models for MD simulations	94
4.5.2. Simulations set up	95
4.5.3. Well-tempered Metadynamics simulations	96
4.5.4. MD trajectories analysis	96

5. RELATIVE BINDING FREE ENERGY PREDICTION STRATEGIES FOR RNA-TARGETED DRUG DISCOVERY	97
5.1. ABSTRACT	98
5.2. INTRODUCTION	98
5.3. RESULTS	100
5.4. CONCLUSION	115
6. OVERALL DISCUSSION, CONCLUSIONS AND FUTURE OUTLOOK	117
SUPPORTING MATERIALS FOR CHAPTER 4	120
BIBLIOGRAPHY	131

Table of Figures

Figure 1.1. Ribozymes catalytic mechanisms	16
Figure 1.2. Structural and catalytic details of group II introns	18
Figure 1.3. The spliceosomal splicing cycle	22
Figure 1.4. The spliceosome structure	25
Figure 1.5. Structure and domains of Prp8	26
Figure 2.1. FEP Thermodynamic cycle	44
Figure 3. 1. RNA geometrical features	50
Figure 3. 2. Main Amber RNA FF advancements	53
Figure 3. 3. Amberff12 RNA simulation across 4 water models	56
Figure 3. 4. Mg^{2+} ions mediate ligand binding and conformational transitions	62
Figure 3.5. Controlled K^{+} ion release in group II intron	64
Figure 3.6. Integrating experimental data in MD simulations	68
Figure 4.1. Architecture and evolutionary conservation of key spliceosomal active site elements	76
Figure 4.2. Geometry of the active site of the <i>S. cerevisiae</i> spliceosome from the B ^{ACT} to the C complexes	80
Figure 4.3. Energetics associated with the occupancy of the K1 site and the position of the M2 ion	81
Figure 4.4. Cooperation of Prp11, Cwc24, and Prp8 for the stabilization of the 5'SS	82
Figure 4.5. Yju2 gating dynamics regulate recruitment of potassium ions at the K1 site in the B* complex	85
Figure 4.6. Interaction between the branching point adenosine (BPA) and the 5'SS in the B* and C complex	87
Figure 4.7. Mechanistic model for the dynamic remodeling of the spliceosome active site during the first step of splicing	93
Figure 5. 1 Co-crystal structures of the three case studies	101
Figure 5.2. Case study I: MD equilibration improves FEP predictions on RNA-fragments	104
Figure 5.3 Case study II: benefits of MD sampling on RBFE	109
Figure 5.4. Case study III, series 1: MD equilibration favours RBFE	109
Figure 5.5. Case study III, series 2: MD equilibration improved RBFE prediction accuracy	112

List of Tables

Table 3.1. Summary of the case studies analyzed	58
Table 5.1. Ligands series for case study	102
Table 5.2. Ligands series for case study II	103
Table 5.4. Ligands series 1 for case study III	104
Table 5.3. Ligands series 2 for case study III	104

Acknowledgments

Many people have walked by my side along this journey, contributing to every step.

First of all, I would like to thank my supervisor, Dr. Marco De Vivo for guiding me from the very first day. By wisely dosing the right balance of freedom to make mistakes and correcting me when necessary, deeply contributed to both my personal and scientific progress. Working with your contagious enthusiasm and energy constantly pushed me to do my best. Every suggestion and comment was precious, especially your guidance in writing and in shaping my scientific work, which taught me the importance of precision and clarity.

I am deeply grateful to Dr. Anders Hogner for welcoming me into his laboratory with generosity and kindness, providing not only scientific guidance but also a rare and sincere personal support. Thank you for the encouragement and professional mentorship.

I also wish to warmly thank Dr. Marco Marcia and Dr. Jo La Sala, for sharing with me their extensive expertise through insightful advice and constructive discussions.

A special acknowledgment goes to my former colleague, later mentor, and dear friend Dr. Jacopo Manigrasso. Thank you for your constant teaching and – most importantly – for seeing something in me, helping me to see it too, and bringing out my best. Working together has been a continuous source of growth, and I am deeply grateful to have gained not only knowledge but also a true friend.

Finally, I wish to extend my heartfelt thanks to all those who, in different ways, have contributed to this journey. Sometimes, we do not know exactly what we need, but we simply wander in search of it; and along the path, we find the people who make the journey meaningful

Abstract

Non-coding RNAs (ncRNAs) have emerged as a central player in gene regulation, catalysis and molecular recognition, reshaping the classical protein-centric view of cellular regulation. The ability to adopt complex three-dimensional architectures enables RNA molecules to act as carriers of information, structural scaffolds, and catalysts. These same structural and chemical features make RNA an increasingly attractive therapeutic target. However, the intrinsic flexibility of RNA, its structural and functional dependence on ions and solvent, and the limited availability of high-resolution structural and affinity data continue to challenge both experimental and computational approaches. Indeed, RNA-targeted drug discovery is confined at an earlier stage compared to its protein-focused counterpart. In this thesis, we address these challenges from two complementary perspectives. First, we employ extensive equilibrium and enhanced sampling molecular dynamics (MD) simulations, to reveal the mechanistic principles that govern RNA catalysis during pre-mRNA splicing. Specifically, we describe how conserved positively charged residues and catalytic metal ions dynamically cooperate within the spliceosome to ensure precise and timely activation of the active site. This mechanistic understanding not only advances our knowledge of RNA-based catalysis but also identifies potential regulatory hotspots that could be exploited for therapeutic modulation. In parallel, we test the largely unexplored computational potentiality and limitations of RNA-ligand relative binding free energy (RBFE) predictions by systematically benchmarking FEP+ on different RNA-small molecules systems. By comparing polarizable and fixed-charge force fields, we demonstrate that Free Energy Perturbation calculations can reliably predict RBFE in RNA systems. Altogether, this work bridges mechanistic insight and methodological innovation, contributing to advance both our understanding of RNA catalysis and our ability to computationally model RNA-ligand interactions for the future development of RNA-targeted therapeutics.

1.Introduction to RNA Biology and its Therapeutic Potential

1.1. Overview of RNA biology

The ribonucleic acid (RNA) is a highly dynamic molecule that plays a wide variety of biological roles nearly all forms of life. Chemically, RNA is a single-stranded nucleic acid polymer composed of nucleotide monomers – ribose sugars attached to nitrogenous bases (adenine, guanine, uracil, and cytosine) and phosphate groups – covalently linked by 5'-3' phosphodiester bonds between ribose sugars.

Traditionally, RNA has been viewed as the intermediate that conveys genetic information from DNA to protein in the central dogma of molecular biology¹. This perspective was first solidified with the isolation and synthesis of the messenger RNA (mRNA) in 1961^{2,3}, which highlighted its role in encoding and transmitting instructions from the nucleus to the cytoplasm, where proteins are synthesized. In the early 1960s⁴, the identification of ribosomal and transfer RNA (rRNA and tRNA) further supported RNA's involvement in protein synthesis. However, rRNA and tRNA do not encode for any protein, thus, these findings showed that RNA can be either coding – like mRNA, which carries genetic information – or non-coding, like tRNA and rRNA which play structural and functional roles in translation. Following the functional description of tRNAs and rRNAs, in the late 20th century, transcriptomic analyses have demonstrated that a significant portion of the human genome (76-97%) encodes for RNAs that are not translated into proteins, termed non-coding RNAs (ncRNAs)⁴⁻⁶. A wave of discoveries revealed different classes of ncRNAs, such as microRNAs (miRNAs) that regulate gene expression post-transcriptionally⁷, small nuclear RNAs (snRNAs) that regulate pre-mRNA splicing⁸⁻¹⁰ and long non-coding RNAs (lncRNAs) that are involved in regulating gene expression at various levels^{11,12}, suggesting that ncRNA plays diverse regulatory and structural roles, as explained along this chapter. In parallel, the discovery of catalytic RNAs (Chapter 1.3) and riboswitches further expanded the functional landscape of RNA, showing that some RNAs can act as enzymes or metabolite sensors¹³. This has fundamentally reshaped our understanding of gene regulation and genome function.

Many ncRNAs are now known to regulate chromatin structure, RNA splicing, translation, and even protein localization^{14,15}. Notably, their dysregulation have been implicated in a wide

range of physiological and pathological processes, including cancer progression, neurological disorders, and inflammatory diseases, making them attractive targets for therapeutic intervention^{16,17}. This significantly increased the interest into RNA structure, dynamics, function, and more recently, targeted modulation through small molecules, antisense oligonucleotides, and other RNA-targeted therapeutics (see chapter 1.5 and 1.6).

1.1. The Central Role of ncRNA in Biological Systems

Since their discovery, the biological importance of ncRNAs has become increasingly evident, shifting the classical protein-centric paradigm toward a more RNA-inclusive view of cellular regulation. ncRNAs have emerged as key regulatory and structural elements involved in nearly every layer of gene expression. Unlike mRNAs, which serve as templates for protein synthesis, ncRNAs regulate a wide range of processes including transcriptional and post-transcriptional gene regulation, splicing, translation, and the assembly of large macromolecular complexes.^{15,18} ncRNAs are typically classified based on size, which can vary from ~ 22 nucleotides as for microRNAs (miRNAs), to more than 200 nucleotide as for long non-coding RNAs (lncRNAs), and function.

1.1.1. *microRNAs*

Among the most extensively studied ncRNAs there are miRNAs, which are short (~22 nts), structured ncRNAs that play a central role in post-transcriptional gene silencing. First identified in *Caenorhabditis elegans* about 30 years ago^{19,20}, over 38,000 miRNAs from 271 species, have been annotated in the miRbase (v22.1)²¹. miRNAs are initially transcribed as primary miRNAs, typically longer than 1 kb and containing characteristic stem-loop structures. These hairpins are processed in the nucleus by the Microprocessor complex, which includes the RNase III enzyme Drosha, yielding a ~55–70 nucleotide-long precursor miRNA (pre-miRNA). In the cytoplasm, the pre-miRNA is further cleaved by Dicer, another RNase III enzyme, to produce a short double-stranded RNA duplex. One strand of this duplex is incorporated into an Argonaute protein to form the RNA-induced silencing complex (RISC), while the other strand is degraded. RISC base pairs with complementary sequences of target mRNAs. This interaction leads to translational repression or mRNA degradation²². Given their central roles in gene regulation, their dysregulation is implicated in numerous human diseases, including cancer, neurological disorders. As a result, miRNAs have become prominent targets

in therapeutic development, including small molecule interventions aimed at modulating their expression or maturation²³.

Promising strategies involve targeting miRNA biogenesis at the level of Drosha or Dicer processing. For example, a small-molecule inhibitor with nanomolar affinity was recently identified²⁴, it binds to the Drosha processing site of the oncogenic pre-miR-27a, effectively preventing its maturation. Similarly, miR-21, a highly oncogenic miRNA commonly overexpressed in various cancers, has been targeted using bifunctional molecules that couple neomycin – capable of binding pre-miR-21 – with N-hydroxyimide, a Dicer inhibitor, reducing the levels of mature miR-21 and offering a potential route for therapeutic intervention²⁵.

1.1.2. Riboswitches

Beyond miRNA, other extensively studied structured ncRNAs with distinct regulatory roles include riboswitches. Riboswitches are structured metabolite-sensing regulatory elements typically located within the 5' UTRs of bacterial mRNAs. These structured RNA domains act as genetic switches, directly binding small molecules or ions to control gene expression through conformational changes that affect transcription termination or translation initiation. Over the past two decades, more than 55 distinct classes of natural riboswitches have been identified, and thousands more are predicted to exist²⁶. Each riboswitch generally comprises two functional domains: an aptamer domain, which forms a structured ligand-binding pocket with high specificity, and an expression platform which undergoes conformational changes upon ligand binding to modulate gene expression. Riboswitches exhibit several features that make them particularly attractive as therapeutic targets. Predominantly found in bacteria, where they regulate essential metabolic pathways and homeostasis, they are absent in humans, allowing for selective antimicrobial targeting. Additionally, their aptamer domains readily accommodate drug-like molecules, making them highly suitable for structure-based drug design²⁷.

Notably, riboswitches that respond to antibiotic-like ligands offer promising avenues to fight antibiotic resistance – a major global health threat²⁸. Some riboswitches, such as the thiamine pyrophosphate (TPP) and flavin mononucleotide (FMN) riboswitches, are conserved across numerous bacterial species and may serve as broad-spectrum antibacterial targets. Others, such as prequeuosine 1 (PreQ1) and purine riboswitches, are more restricted in distribution and could be leveraged for species-specific antimicrobial strategies. For instance, the FMN riboswitch, located in the *ribD* operon and involved in regulating riboflavin biosynthesis and transport, is a validated target in *C. difficile*, *E. coli*, and *K. pneumoniae*²⁹. Synthetic ligands

such as 5FDQD and ribocil³⁰ have shown potent antibacterial activity by binding the FMN riboswitch with high affinity and low risk of resistance development. Similarly, a recent structure-based drug design effort targeting the PreQ1 riboswitch led to the development of high-affinity ligands, aided by crystallographic data that revealed key binding interactions within the RNA³¹. To date, the only clinically approved small-molecule antibiotics known to act through direct RNA binding – linezolid³², tedizolid³³ and risdiplam – target bacterial ribosomal RNA or pre-mRNA. However, the growing body of work on riboswitch-targeting compounds has expanded the landscape of RNA-targeted therapeutics³⁴, underscoring the potential of non-ribosomal RNA elements as drug targets and contributing to the increasing interest in RNA-targeted drug discovery.

1.1.3. Small nuclear RNAs

While most known riboswitches have been identified exclusively in bacteria, other classes of non-coding ncRNAs, such as small nuclear RNAs (snRNAs), are abundantly present in all eukaryotic cells. A major subclass of snRNA is the uridine-rich snRNAs (U-snRNAs), which comprise approximately 1% of total mammalian cellular RNA. Typically around 150 nucleotides in length, U-snRNAs characteristically associate specific nuclear proteins to form small nuclear ribonucleoproteins (snRNPs)³⁵. snRNPs act cooperatively to facilitate pre-mRNA splicing - the removal of non-coding introns from precursor transcripts to generate mature mRNAs or functional ncRNAs (see Chapter 1.4). Specifically, the U1, U2, U4, U5, and U6 snRNAs are core components of the spliceosome, a massive and dynamic ribonucleoprotein complex responsible for catalyzing pre-mRNA splicing in eukaryotes.³⁶ These five U-snRNAs, participate in several RNA-RNA and RNA-protein interactions in the assembly and function of canonical spliceosomes: U1 recognizes the 5'-end splice site of the pre-mRNA, and U2 with U6 create the RNA active site, U4 and U5 play key roles in active site assembly and structural rearrangements. Collectively, they coordinate the splicing reaction and help maintain fidelity in exon-intron recognition. Given their essential role in regulating splicing, it is not surprising that dysregulation of snRNAs or their associated snRNPs has been implicated in several human diseases, including cancer and Spinal muscular atrophy. For instance, recurrent mutations in U1 snRNA have been identified in multiple cancers, where they contribute to aberrant splicing and dysregulated gene expression, highlighting U-snRNAs as potential therapeutic targets^{36,37} (see chapter 1.5.2).

Together, these advances illustrate that targeting snRNA-RNA interactions represent a

viable strategy for therapeutically modulating gene expression, offering a way to influence cellular pathways without altering the genome or targeting traditionally 'undruggable' proteins.

This role of snRNAs in gene regulation and disease underscores their relevance, both biologically and therapeutically, and sets the stage for deeper exploration in the following chapters on splicing (Chapter 2.3 and 2.4) and splicing-targeted drug discovery (Chapter 2.5).

1.2. Catalytic RNA

Some RNAs, despite being composed of only four chemically similar nucleotides, can fold into distinct three-dimensional architectures that enable a broad spectrum of biological functions. Beyond constituting scaffolds for protein binding, some RNAs can act as a catalyst, a property that changed the long-standing protein-centric view of enzymatic activity. In 1982, Tom Cech and colleagues demonstrated that an intron from the ribosomal RNA precursor of *Tetrahymena thermophila* could self-excise in vitro, without the aid of proteins or external energy sources.³⁸ Soon after, Sidney Altman's group showed that the RNA component of ribonuclease P (RNase P), M1 RNA from *Escherichia coli*, could independently process precursor tRNAs.³⁹ These landmark discoveries earned Cech and Altman the Nobel Prize in Chemistry in 1989 and expanded the knowledge of catalytic RNA, supporting the RNA world hypothesis, which proposes that ancient RNA molecules may have covered both genetic and catalytic functions in early life.

Ribozymes represent a key class of naturally occurring functional RNAs, which can be broadly divided into two categories, based on their function: cleaving ribozymes and splicing ribozymes (group I and II introns)^{13,40}. Cleaving ribozymes can be further divided into trans-cleaving ribonuclease P (RNase P)⁴¹, and small self-cleaving (nucleolytic) ribozymes – which includes hairpin⁴², hammerhead^{43,44}, hepatitis delta virus (HDV), HDV-like motifs^{45,46}, glucosamine-6-phosphate synthase⁴⁷, *Neurospora* VS⁴⁸, twister, hatchet, and pistol motifs^{49,50}. A schematization of the catalytic actions performed by these three RNA categories is summarized in **Figure1.1**.

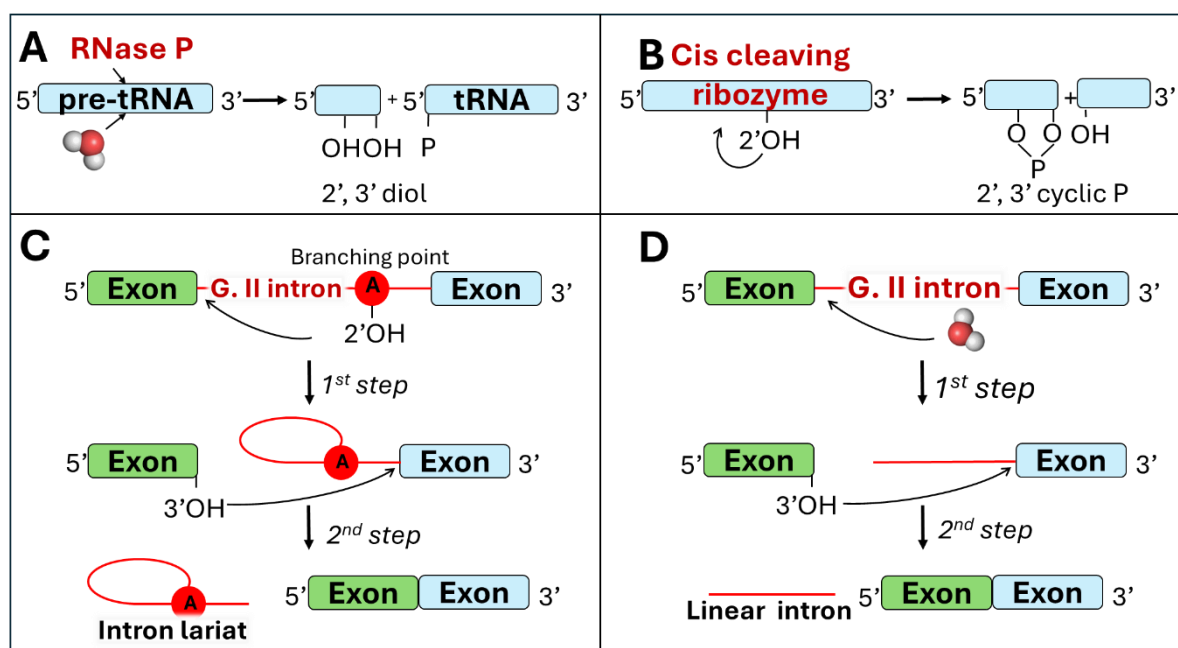


Figure 1.1. **Ribozymes catalytic mechanisms.** A) RNase P cleavage of the tRNA. A water molecule serves as a nucleophile, and the reaction yields 2',3'- diol and a 5'P termini. B) The typical reaction of self-cleaving ribozymes, carried by 2'OH and yielding 2',3'-cyclic phosphate and 5'OH termini. C and D represent the self splicing by group II introns via the branching and hydrolytic pathway, respectively.

1.2.1. Trans cleaving ribozymes – Ribonuclease P

Ribonuclease P (RNase P) is a ubiquitous and essential trans-acting endonuclease that processes precursor tRNAs by removing their 5' leader sequences. It utilizes two Mg^{2+} ions: one activates a hydroxide nucleophile for an SN2 attack, while the other stabilizes the transition state and the leaving group. The reaction yields a mature tRNA with a 5'-phosphate and a 5' leader with a 2',3'-diol (**Figure 1.1A**)^{39,51}. The RNA core catalytic function of RNase P is conserved across the species, while its RNA subunit (RPR) has evolved across species, showing remarkable diversity in structure and complexity. In some bacteria, eukaryotes and archaea, parts of the RPR have been functionally replaced or complemented by proteins, and other eukaryotic species have even transitioned to a protein-only variant^{52,53}. The uniformity in the conserved catalytic core and differences in the peripheral elements of the RPR, underscore strong convergent evolution and makes the ribonucleoprotein form of RNase P an ideal drug target⁵⁴. Due to its therapeutic potential in the field of infectious disease and gene-silencing therapies, several studies attempted to target bacterial RNase P with small molecules, but flexibility, promiscuity, and protein-mediated aggregation considerations have ruled out many candidates^{55,56}. However, current studies are expanding the chemical space for RNase P modulators, opening up possibilities for designing RNA-targeted antibiotics or microbiome

modulators^{55,57}.

1.2.2. *Cis cleaving ribozymes*

Unlike RNase P, most self-cleaving ribozymes perform acid-base catalysis in *cis*⁵⁸. The site-specific phosphodiester bond cleavage involves a nucleophilic attack, with S_N2-type mechanism, often supported by a two-metal-ion catalytic strategy. One Mg²⁺ ion activates the 2'-hydroxyl nucleophile, and another stabilizes the leaving group and transition state. Mechanistically, a 2' deprotonated hydroxyl of the nucleophile attacks the adjacent phosphodiester bond with the 5-O leaving group. The transesterification yields a 2'-3' cyclic phosphate and a 5'-hydroxyl product, in a single step (**Figure 1.1B**)^{40,59}. Nevertheless, these compact catalytic RNAs show structural differences that define unique active sites and reaction mechanisms, which for most of them are still debated⁶⁰. Four main catalytic strategies have been distinguished from nucleolytic ribozymes: in-line attack, stabilization of the negative charge on the non-bridging phosphate oxygen atoms, proton abstraction from the 2-hydroxyl and stabilization of the leaving group⁶¹. Used singularly or in combination, these strategies promote RNA scission. In this framework, computational studies, mainly involving QM/MM, pKa calculations and equilibrium molecular dynamics (MD), in cooperation with experimental works have been critical for elucidating the atomistic nuances of these intricate mechanisms^{61–63}. For instance, MD simulations of the pistol ribozyme revealed three Mg²⁺ ions near the catalytic center. Two of them showed interchangeable roles in supporting catalysis and conformational stability, indicating a fail-safe architecture for catalysis, different from the other ribozymes⁶⁴. Self-cleaving ribozymes are tools able to control RNA functions with different strategies, thus, understanding their exact mechanisms is appealing for future therapeutic uses and to inspire new biotechnological techniques, such as enabling pharmacological selection of precise gene editing events in vivo, overcoming cell-type specificity issues in the expression of guide RNAs in the CRISPR/Cas9 systems⁶⁵.

1.2.3. *Splicing ribozymes – Group II introns*

In contrast to small self-cleaving ribozymes, splicing ribozymes catalyze two sequential transesterification reactions. Among them, group II introns are the most extensively studied. These large catalytic RNAs can excise themselves from precursor transcripts and ligate flanking exons in a fully autocatalytic process⁶⁶. Found in bacteria, archaea, mitochondria, and chloroplasts of fungi and plants, group II introns play essential roles in RNA maturation and

the regulation of gene expression. They typically range from 400 to 800 nucleotides and fold into a conserved six-domain architecture (DI–DVI, **Figure 1.2A**)⁶⁷.

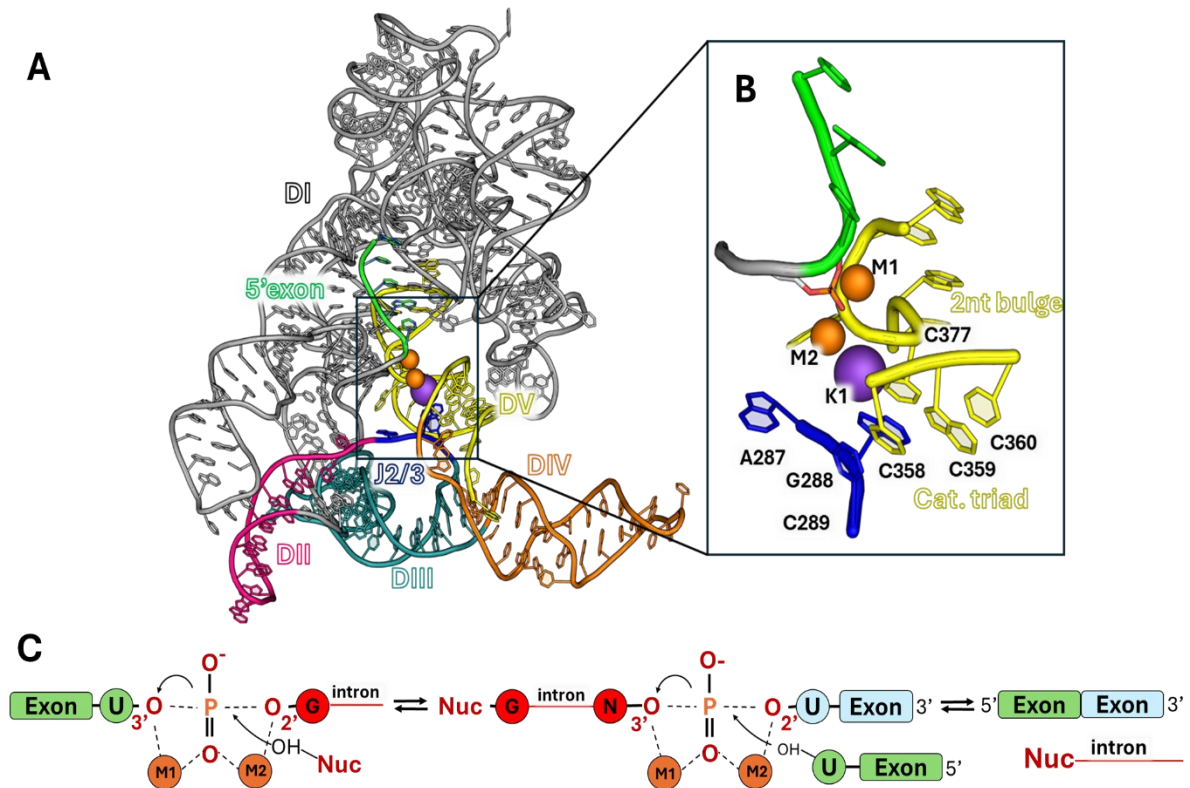


Figure 1.2. Structural and catalytic details for group II introns. (A) Cartoon representation of D1-to-D5 3D architecture of *Oceanobacillus Hiyensis* group II intron (PDB ID 4FAQ). 3' end of 5' exon in green, Domain 1 in grey, Domain 2 in magenta, junction between domains 2 and 3 in blue, Domain 3 in cyan, domain 4 in ssnd, Domain 5 in yellow. (B) Zoom in the catalytic site. The catalytic triplex is formed by the 2 nucleotide bulge (yellow), the catalytic triad (yellow) and the A, G, C nucleotides of the J2/3. The two Magnesium ions (orange spheres) coordinate the scissile phosphate (in stick), while K⁺ ion in K1 site preserve the site structure. Such active site geometry is ready for the first step of splicing, schematized on the left side in (C). The nucleophile (Nuc-OH) can be either a water molecule or the BPA, according to the path followed. Between the I and II steps, the J2/3 toggling controls the site reorganization, allowing the release of I step products and the recruitment of II step reactants. Importantly, each transesterification is reversible: from left to the right, it represent the forward splicing, viceversa is the reverse splicing

DI is the largest domain, functioning as a structural scaffold and contributing to exon recognition. DII and DIII serve structural roles, while DIV often encodes an intron-encoded protein (IEP), a multifunctional factor required for folding, splicing, and mobility. DV contains the catalytic core and coordinates two essential Mg²⁺ ions. DVI includes the branch-point adenosine that initiates the first splicing reaction. Although distant in the linear sequence, these domains come together through long-range tertiary interactions, forming a compact and catalytically competent structure⁶⁸. Splicing proceeds via a two-step mechanism (**Figure 1.1C and D**). In the first step, called branching, the 2'-OH of the bulged adenosine in DVI attacks the 5'-splice site, yielding a lariat intermediate and a 3' exon. In the second step, the 3'-OH of

the cleaved 5' exon attacks the 3'-splice site, resulting in exon ligation and release of the intron as a lariat RNA^{69,70}(**Figure 1.1C**). Some group II introns can undergo hydrolytic splicing via water-mediated nucleophilic attack⁷¹ yielding linear introns (**Figure 1.1D**), or, as group IIA introns, circularization, which lies outside the scope of this thesis.

The first high-resolution structure of a group II intron – solved from the *Oceanobacillus iheyensis* group IIC intron – captured the RNA after the first step of splicing (**Figure 1.2A**) and revealed a catalytically competent active site⁴¹ (**Figure 1.2B**). Central to this is a triple-helix structure formed by three conserved elements: a catalytic triad (C358-G359-C360; all numbering from the crystallized form of the *O. iheyensis* intron, i.e. PDB id: 4FAQ) and a two-nucleotide bulge (A376-C377) in DV, and a segment from the DII-DIII junction (A287-G288-C289). These RNA motifs coordinate a metal ion cluster composed of two catalytic Mg²⁺ ions (M1 and M2) and two potassium ions (K1 and K2), with structural role and reactant-positioning role. Subsequently, numerous structures described different steps of the splicing cycle at medium to high resolution^{72–77}. Those, in combination with further biochemical and computational investigations revealed details and nuances of the splicing process^{78–81}. During the first step, M1 and M2 coordinate the scissile phosphate, stabilizing the 5'exon-intron junction conformation with a kink of ~50°, ensuring the scissile phosphate positioning within the site. M2 coordinate the 2'OH nucleophile too, while K1 reinforces the triple helix. Upon the coordination of the reactants is established, the first step transesterification occurs generating the lariat intermediate. At this stage, the 5' exon is positioned to attack the 3' splice site, but steric hindrance by the 5' end of the intron prevents progression. A rearrangement of the active site is therefore required to transition to the second step. Although no crystal structure captures this transition, recent work⁸¹ from our combining mutagenesis, enzymatic assays, crystallography, and MD clarified the mechanism. After the first transesterification, the proton released by the nucleophile protonates C358 of the catalytic triad, triggering K1 release and disruption of the J2/3 junction, which toggles out and disassembles the metal cluster. This rearrangement positions the catalytic triple helix to recruit the second-step reactants. Upon DVI recruitment, J2/3 toggles back, restoring the metal cluster and active geometry. In the second step, M1 activates the 5'-exon to attack the 3'-splice site, yielding exon ligation and intron release (**Figure 1.2C**).

Remarkably, the excised lariat intron retains catalytic activity and can reinsert into cognate or non-cognate genomic DNA through the processes known as retrohoming or retrotransposition, respectively^{82,83}. The lariat intron RNA inserts into ligated exons guided by

the same exon-intron and base pairing interactions between the intron and flanking exon sequences used for RNA splicing. This mobility relies on the IEP, which contains four domains: a reverse transcriptase (RT), a maturase domain (X/thumb), a DNA-binding domain (D), and a DNA endonuclease (En). The RT and maturase domains bind and stabilize the intron RNA and initiate cDNA synthesis. The En domain, a Mg^{2+} -dependent nuclease nicks the target DNA, creating a primer for reverse transcription⁸⁴. Multiple structural and bioinformatic studies have established strong evolutionary links between group II introns and the eukaryotic spliceosome. The spliceosomal protein Prp8 shares homology with the IEP's RT domain and acts as a scaffold for the RNA catalytic center^{85,86} (Chapter 1.4.2). Moreover, the catalytic triad and bulged nucleotides of group II introns are conserved in the U6 snRNA, where they fulfil equivalent catalytic roles (**Figure 1.3B, C**). Both systems operate via metal-ion-dependent RNA catalysis, resulting in an intron lariat structure and require active site rearrangement between steps, sharing identical stereochemistry⁷⁹. This evidence supports the widely accepted model that group II introns exploited their mobility to invade the ancestral eukaryotic genome, duplicating themselves from the nucleus to several genomic sites and, over time, lost autonomous splicing activity. To prevent genomic instability, these mobile elements fragmented into snRNAs that, together with protein cofactors, evolved into the modern spliceosome⁸⁷. Because group II introns are sub-megadalton ribozymes that can be reconstituted and studied in vitro, they serve as a powerful model system to dissect splicing mechanisms relevant to nuclear pre-mRNA processing. Studies on their structure, dynamics, and catalytic strategies have deepened our understanding of RNA-mediated catalysis and informed the development of RNA-targeting therapeutics⁸⁸. Furthermore, efforts to modulate group II introns with small molecules have provided valuable insights into the druggability of splicing-related machineries^{89,90} – topics that will be explored in the following chapter on the spliceosome.

1.3. Splicesomal splicing: a critical regulatory hub

Eukaryotic genes contain non-coding sequences (introns) that must be accurately removed from pre-mRNA to produce mature transcripts. This task is performed by the spliceosome, a dynamic ribonucleoprotein complex that assembles in a stepwise and highly regulated manner on each intron (**Figure 1.3A, 1.4**). Unlike pre-formed enzymes, the spliceosome forms *de novo* around the pre-mRNA, involving five small nuclear RNAs (U1, U2, U4, U5, U6) and

approximately 100 proteins (**Figure 1.4**). Each snRNA forms a small nuclear ribonucleoprotein (snRNP) by associating with specific proteins. In addition, eight ATP-dependent RNA helicases further drive the conformational rearrangements required for splicing^{36,91} (**Figure 1.3A**). Splicing fidelity is essential, as mis-splicing can lead to frameshifts and defective proteins. In fact, about 35% of genetic diseases are associated with splicing abnormalities⁹². Therefore, a detailed mechanistic understanding of the pre-mRNA splicing is essential for developing strategies to modulate this process and create effective therapeutic strategies. In this context, over the past decade significant advances in cryo-electron microscopy (cryo-EM), have revealed structures of the spliceosome from yeast and humans at various stages of the splicing cycle^{93–97}. These structural insights, combined with biochemical and genetic data from prior experiments, have led to a comprehensive understanding of the splicing process.

1.3.1. Splicing cycle

The spliceosome catalyzes two transesterification reactions – branching and exon ligation – via a single RNA-based active site formed by U2 and U6 snRNAs and supported by U5 snRNA. Spliceosome progression is defined by distinct intermediates: E, A, pre-B, B, B^{ACT}, B*, C, C*, P, and ILS (**Figure 1.3A**), where 4 states of assembly, activation, catalysis, and disassembly are crossed^{36,98–100}. Assembly and catalysis are tightly coordinated to prevent premature reactions and ensure precise substrate positioning.

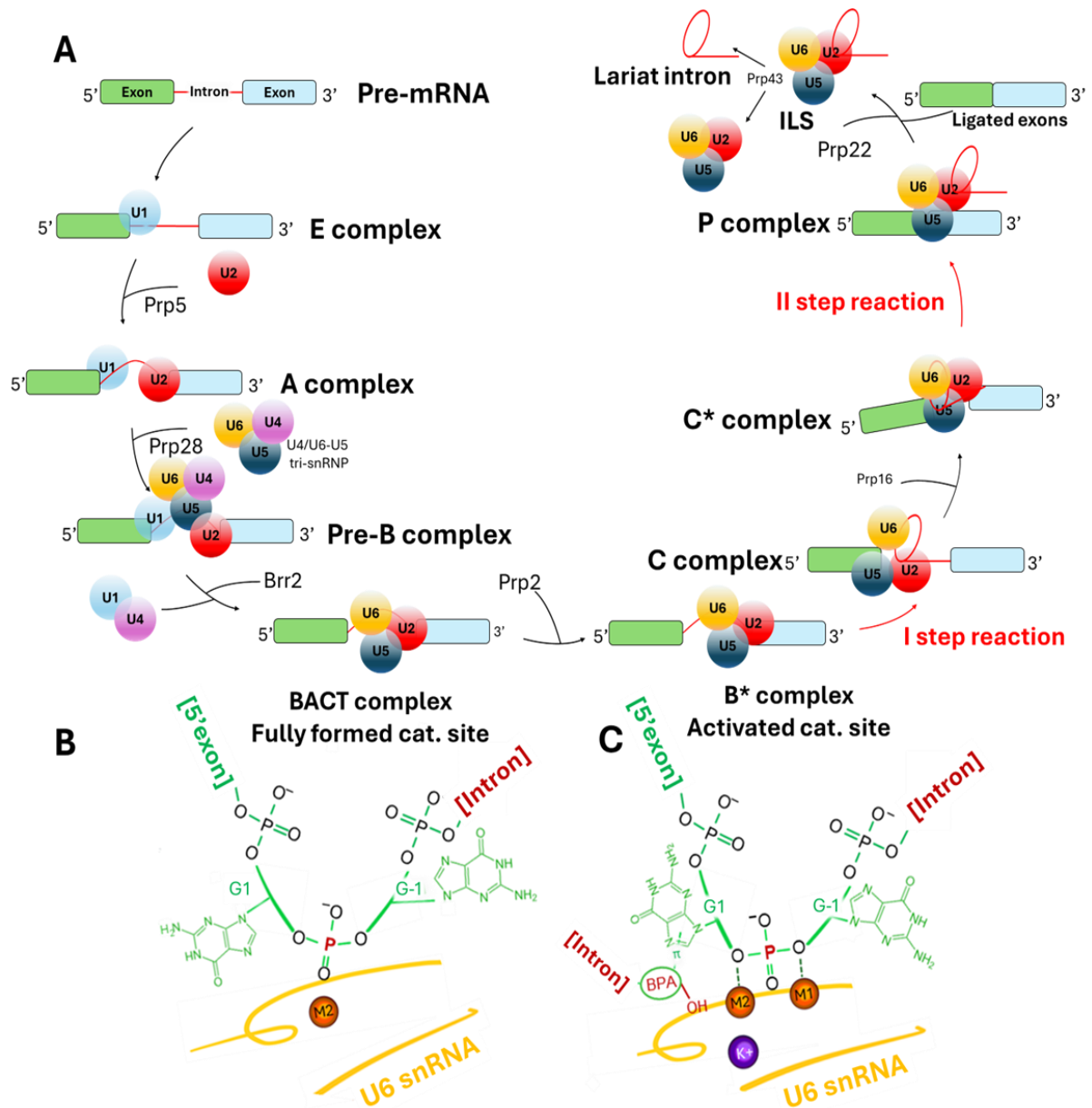


Figure 1.1. **The spliceosomal splicing cycle.** (A) Pre-mRNA splicing cycle. The pre-mRNA is represented by the green and cyan boxes (5' and 3' exon, respectively), separated by the red line (intron). The SnRNPs are simplified as circles. For clarity, only ATP-dependent RNA helicases and SnRNPs movements are shown, avoiding the other proteins. During the cycle, the status of the active site is declared, and schematized in (B) for B^{ACT} and in (C) for B^* complexes. (B) is based on the PDB structure 5GM6 and 7DCO, while (C) is a model, since no B^* structures are available.

E complex (assembly): at this initial stage, the spliceosome assembly begins with U1 snRNP to the 5' splice site (5'SS) via base pairing with U1 snRNA.

A complex (assembly): Prp5 helicase promotes U2 snRNP recruitment to the branching point (BP). U2 snRNA base-pairs with the BP forming the branch helix, while the BPA is involved in interactions with Hsh155 (SF3B1 in human), remaining inaccessible for catalysis. The formation of the A complex corresponds to the pre spliceosome.

Pre-B complex (assembly): At this stage, the pre-assembled U4/U6•U5 tri-snRNP is

recruited to the spliceosome, joining the pre-mRNA-bound U1 and U2 snRNPs to form the fully assembled but catalytically inactive pre-B complex. This tri-snRNP brings all the essential RNA components required for catalysis. U6 snRNA will later form the heart of the catalytic core, while U5 helps align the exons for ligation. However, in the pre-B complex, U6 remains base-paired with U4, holding it in an inactive conformation that prevents premature folding of the active site. Simultaneously, the 5' splice site (5'SS) remains engaged with U1 snRNP, further ensuring that splicing does not initiate before proper assembly

B complex (activation): The ATP-dependent helicase Prp28 starts the transition to the B complex. It promotes the release of the 5'SS from U1 snRNP, enabling the 5'SS to engage with the ACAGAGA box of U6 and U5 snRNA loop1 – key interactions for the active site formation. These molecular rearrangements also reposition the Prp8 RNaseH-like domain exposing the U4 snRNA and the site of binding for Brr2. This, triggers a series of rearrangements in the pre-B complex, ultimately leading to the release of U1, absent in the B complex. While this mechanism is well-supported by cryo-EM structures in human systems, in yeast, Prp28 has not yet been directly visualized in early complexes, leaving certain mechanistic aspects unresolved in this organism.

B^{ACT} complex (activation): Following Brr2 loading onto U4 snRNA, the U4/U6 duplex separates, leaving U6 free to pair with U2 snRNA and form the catalytic core B^{ACT} represents the first spliceosome intermediate with a near-complete active site architecture (see chapter 1.4.3) (**Figure 1.3B**), reminiscent of group II introns (**Figure 1.2B**). However, the branch point adenosine (BPA) – the nucleophile for the first transesterification – is still sequestered ~50 Å from the catalytic center within the Hsh155 protein¹⁰¹. To prevent hydrolytic reaction using water as a nucleophile, protective proteins such as Prp11 (SF3A2 in humans) and Cwc24 (RNF113A) shield the 5'SS and occupy the active site (Chapter 4).

B* complex (catalysis): The transition to B* is driven by the helicase Prp2⁹⁵, which interrupts the interactions of the RNA with SF3b complex. SF3a and SF3b were bound to U2 snRNA, which now has some free space to accept the incoming branch helix, that enters the active site displacing Cwc24. Stabilization of the BPA within the active site requires the recruitment of branching factors Yju2, Cwc25, and Ysy1, which together establish a catalytically competent B* complex. Structural studies of B* intermediates lacking Cwc25, reveal that without this protein the branch helix adopts a misaligned conformation incompatible with catalysis^{94,95}. Upon Prp2 dissociation, the RNA catalytic core becomes fully assembled. The BPA's 2'OH is now precisely aligned for the first transesterification, attacking the phosphate at the 5'SS, with all reactants coordinated by the metal ion cluster (**Figure 1.3C**).

C complex (catalysis I): It represents the spliceosome immediately after the branching reaction. At this stage, the active site hosts the reaction products: a cleaved 5' exon and a lariat intron-3' exon intermediate. Structurally, the spliceosome closely resembles the B* complex, with minimal additional rearrangements beyond the chemistry of the first step.

C* complex (activation and catalysis II): As with group II introns, the same catalytic site is reused for both steps of splicing. However, a remodeling step is required to release the cleaved 5' exon and position the 3' exon for ligation. This transition is driven by the ATP-dependent helicase Prp16, which promotes the dissociation of the branching factors Yju2 and Cwc25 and facilitates a rotation of the branch helix, clearing space within the active site. Structural stabilization of the repositioned branch helix is provided by the exon ligation factor Prp17 and the conformational shift of Prp8. These changes are accompanied by rearrangements in the U2 snRNP and associated NTC components. Additionally, subtle adjustments in the pairing between U6 snRNA and the 5'SS allow the 3'SS to dock, setting the stage for exon ligation, which occurs after C* formation.

P complex. At this point, the two exons have been joined to produce the mature mRNA. The ATP-dependent helicase Prp22 facilitates the release of the ligated exons, along with dissociation of exon ligation factors, allowing the spliceosome to transition to a disassembly-ready state.

ILS complex (disassembly): The intron lariat-spliceosome (ILS) complex retains the snRNAs and the excised lariat intron. Disassembly of this complex is mediated by the helicase Prp43, with support from the NTC protein Syf1. This process releases the U2, U5, and U6 snRNPs, which are then recycled for subsequent rounds of splicing.

1.2.1. Proteins are crucial components of the spliceosome

The spliceosome undergoes extensive protein regulated remodeling during each splicing cycle, to fold its RNA active site containing the metal ions required for catalysis, proving that it is a protein-directed metalloribozyme. Although the several dynamic changes during each splicing cycle, it retains approximately 20 stably associated RNA and protein components whose spatial organization and remain consistent across all high-resolution spliceosomal structures available for the yeast B^{ACT}, C, C*, and ILS complexes^{97,101–104} (**Figure 1.4**).

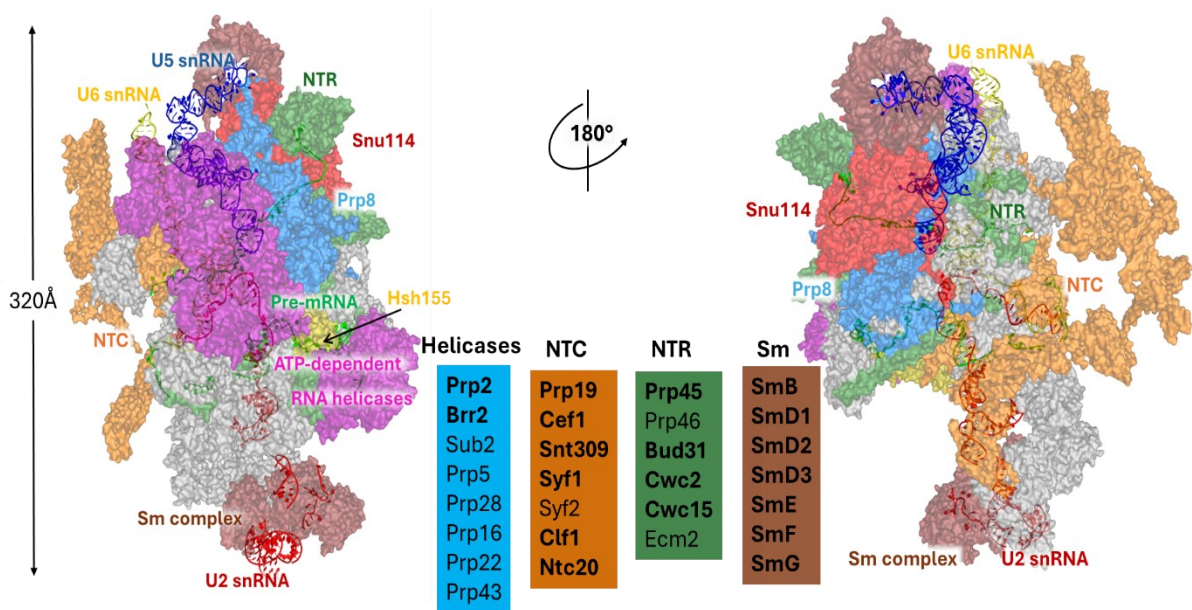


Figure 1.2. **Overview of the spliceosome.** Proteins are shown as surfaces and RNA strands as cartoons in the yeast B^{ACT} complex (PDB ID 7DCO). Distinct colors indicate groups of proteins conserved throughout the splicing cycle, while proteins unique to the B^{ACT} complex are shown in grey. In the legend, the top label of each column indicates the protein group; proteins in bold are members of that group present in the B^{ACT}

The protein scaffold supporting this RNA-based machinery is composed of conserved elements such as Prp8, Snu114, the Sm ring of U5 snRNP, the NineTeen Complex (NTC), and a set of NTC-related proteins (NTR) (**Figure 1.4**). While additional proteins transiently participate at specific stages, these core elements form a stable architectural and functional platform for catalysis.

Prp8 (**Figure 1.5**), the most conserved and structurally complex spliceosomal protein, serves as a dynamic central scaffold throughout the splicing cycle¹⁰⁵.

Located within the U5 snRNP, Prp8 forms a positively charged cavity that accommodates the RNA active site. It adopts two primary conformations: an open state that facilitates U2/U6 catalytic core formation and a closed conformation that stabilizes the active site from B^{ACT} onward. Prp8 comprises seven structured domains: the N-terminal domain, RT fingers/palm, thumb/X, linker, endonuclease-like, RNase H-like, and Jab1/MPN. The central core – comprising RT fingers/palm, thumb/X, linker, and endonuclease-like domains – remains largely invariant across splicing intermediates. In contrast, domains such as the RNase H-like and N-terminal regions reposition dynamically to support catalytic activation and substrate positioning. The Jab1/MPN domain is highly mobile and plays a key regulatory role by modulating the activity of the Brr2 helicase¹⁰⁶.

Additionally, Prp8 hosts three key regulatory elements: i) The switch loop, which stabilizes the bound 5' exon through interaction with Cwc21; ii) The 1585 loop, which interacts with the

U2/U6 duplex and with proteins like Prp11 and Cwc24, aiding 5'SS recognition; iii) The β -finger, which contacts Prp17 and helps facilitate exon ligation⁸⁶.

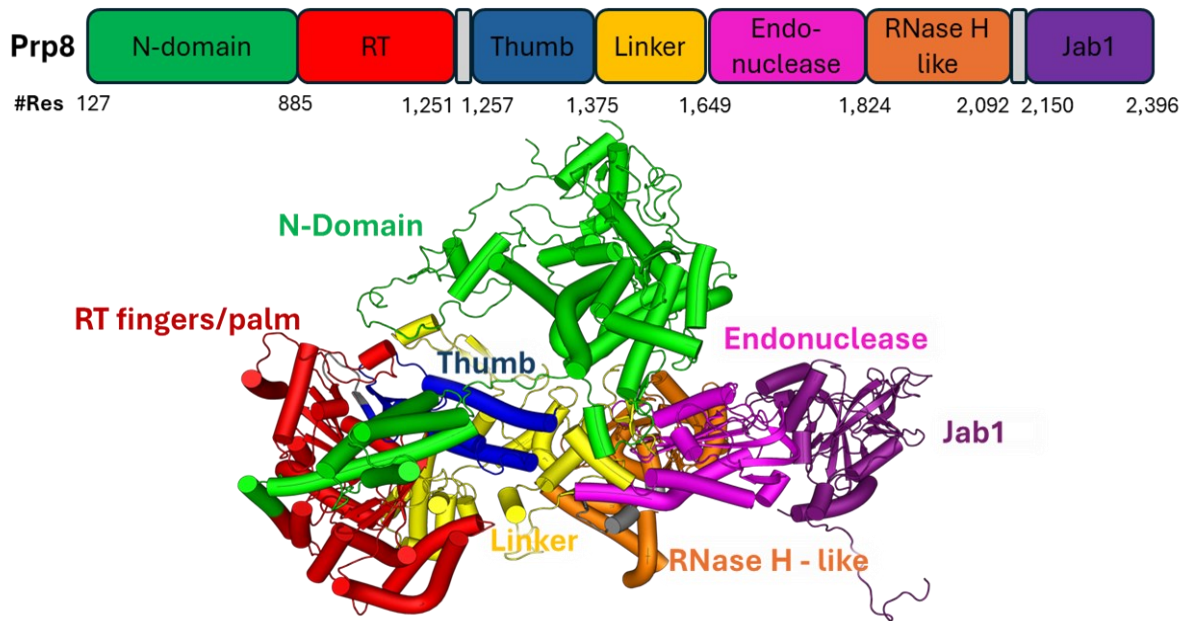


Figure 1.3. **Structure and domains of Prp8.** Prp8 consist of 7 domains. N-domain (green), RT fingers and palm (ref), Thumb (blue), Linker (yellow), Endonuclease (magenta), RNase-H like (orange) and Jab1 (purple) The protein structure was extracted from B^{ACT} complex of the yeast spliceosome (PDB ID 7DCO)

Snu114, a GTP-binding protein in U5 snRNP, modulates the activity of the Brr2 helicase, primarily facilitating the U4/U6 unwinding necessary for active site formation during the Bact transition. It also influences the structural transitions of Prp8, particularly during the B* to C complex transition, enabling 5'SS docking^{107,108}.

The **Sm complex** is a seven-membered ring (composed of SNRPB/B', D1, D2, D3, E, F, and G) that binds the conserved Sm site on U1, U2, U4, and U5 snRNAs. While not catalytically active, the Sm ring is essential for snRNA stabilization, nuclear import, and proper snRNP assembly. Its architectural role is critical for maintaining snRNA function throughout the splicing cycle¹⁰⁹.

The **NineTeen Complex (NTC)**, centered around Prp19, plays a pivotal role in coordinating RNA and protein rearrangements. It is composed of eight core proteins (Prp19, Cef1, Clf1, Syf1, Syf2, Snt309, Ntc20), which remain bound throughout the splicing cycle. Prp19 uses its coiled-coil domain to interact with Cef1 and its WD40 domain to bind NTC-associated proteins like Cwc2, allowing direct regulation of U6 snRNA and pre-mRNA positioning¹¹⁰.

NTC-related proteins (NTR) are a broader group of more dynamic components, which often participate transiently in specific splicing stages to support structural transitions, particularly in conjunction with ATP-dependent helicases activity. For instance, During B to

B* remodeling, the helicase Prp2, supported by the action of the NTRs Spp2 and Cwc22 displaces SF3a/b from U2 snRNP, liberating the 5'SS and enabling recruitment of other two NTRs – Yju2 and Cwc25 – which are required for catalysis initiation. Similarly, Cwc2 (NTC protein) and Isy1 (NTR protein) help rearrange U2/U6 duplex during Prp16 helicase driven exon ligation, as well as the spliceosome disassembly mediated by Prp22 and Prp43 helicases the action of NTC and NTR proteins like Syf1^{111,112}.

ATP-dependent RNA helicases: In total, eight ATP-dependent RNA helicases orchestrate the dynamic rearrangements required for accurate splicing³⁶. These belong to the DEAD-box and DEAH-box families and possess conserved RecA-like domains that mediate RNA binding and ATP hydrolysis. Their roles fall into two main categories: i) RNA and RNP remodeling, mediating essential RNA-RNA, RNA-protein, and protein-protein rearrangements that guide the assembly and function of the spliceosome; ii) kinetic proofreading, acting like quality control checkpoints, rejecting incorrectly assembled intermediates and thus maintaining splicing fidelity. Helicases involved in major transitions of the spliceosome (**Figure 1.3A**) include all the DEAH-box and Brr2¹¹³. Specifically, Prp2 promotes the displacement of the SF3a/b complexes to allow docking of the branch helix into the active site, Prp16 reorganizes the active site to prepare for exon ligation, Prp22 facilitates the release of the ligated exons once splicing is complete, while Prp43 catalyzes disassembly of the intron lariat spliceosome (ILS) complex. Brr2 instead is responsible for unwinding the U4/U6 duplex enabling U6 to form the catalytic core. The DEAD-box helicases act in earlier stages of assembly. Sub2 and Prp5 facilitate recruitment of U2 snRNA and promote its base-pairing with the branch point sequence. Prp28 triggers the displacement of U1 snRNP from the 5' splice site, enabling the handoff of the 5'SS to U6 snRNA and U5 snRNA loop 1. These helicases operate either locally, such as by unwinding RNA duplexes, or remotely, by pulling on RNA to disrupt specific RNA–protein interactions¹¹⁴.

1.3.2. The spliceosome's RNA catalytic core and its mechanism of action

The cooperation of over 150 proteins during the splicing cycle, is essential to ensure timely and accurate formation of the active site. Despite the complexity of protein coordination, catalysis is ultimately carried out by an RNA-based active site formed by the U2 and U6 snRNAs, underscoring its evolutionary connection to group II introns. Recent advances in high-resolution cryo-EM have provided near-complete reconstructions of spliceosomal active site at various catalytic stages^{93–95,97,101–104}.

Upon transition to the B^{ACT} complex (**Figure 1.4**) the U2/U6 helix stably dock onto the surface of Prp8, which subsequently undergoes a conformational shift to a closed state. A conserved catalytic configuration emerges at this stage: a triple helix connects nucleotides C58, A59, and G60 (the catalytic triad) with A79 and U80 (the two-nucleotide bulge), coordinating two Mg²⁺ ions – referred to as M1 and M2 – that are critical for reactant alignment and activation. Adjacent to these is the K1 site, which binds a monovalent ion (typically K⁺) to help stabilize the active site geometry (**Figure 1.3 B and C**). While M1 has been consistently visualized as the only catalytical ion present in yeast B^{ACT} structures^{95,101}, M2 has recently been detected in this intermediate, in humans¹¹⁵. Nevertheless, the presence of M2 does not prematurely activate the B^{ACT} site, first because the nucleophile resides more than 20 Å from the active site (as in the post-B^{ACT} structure, PDB ID: 8I0V), and secondly for the absence of K⁺ at the K1 site that hinder catalytic activities. To date no spliceosomal structure captures the B* intermediate at the moment of the first catalytic step. However, cryo-EM snapshots of the spliceosome before and after branching^{94,96} (PDB IDs: 6J6Q and 7B9V, respectively), along with a group II intron structure at the equivalent catalytic stage¹¹⁶ (PDB ID: 4FAQ) (**Figure 1.2B**), offer important insights into the geometry required for catalysis. As in group II introns, proper alignment of the divalent metal ions is crucial. M1 and M2 must directly coordinate the scissile phosphate, U80, and the catalytic triad, maintaining an approximate 4 Å spacing. The K⁺ ion at the K1 site further stabilizes this arrangement, lying ~4 Å from M2 and ~3 Å from G52 (**Figure 1.3C**). This configuration allows the first transesterification reaction to occur in the B* complex, with the two Mg²⁺ ions coordinating the transition state and properly aligning the reactants.

To explain the chemistry occurring within this active site, the general two-metal-ion mechanism proposed by Steitz and Steitz in 1993 for RNA and DNA catalysis has long served as a foundational model¹¹⁷. According to this scheme, one Mg²⁺ ion activates the nucleophile, the other stabilizes the leaving group, and both support the transition state. However, quantum mechanics/molecular mechanics (QM/MM) and MD studies on the first catalytic step in the group II intron from *Oceanobacillus iheyensis* have challenged this view. That system follows a dissociative mechanism, in which the nucleophile transfers its proton to the bulk solvent during the nucleophilic attack. Here, the Mg²⁺ ion activates the phosphate (the electrophile), yielding a transition state in which bond cleavage precedes bond formation¹¹⁸. Computational studies then extended to the spliceosome, simulating a model of B* based on a yeast C complex. QM/MM and MD calculations confirmed a dissociative mechanism for intron-lariat formation, wherein the 2'-OH of the protonates the 3'OH of the 5' exon with an energy barrier of 21.5

kcal/mol, compatible with the experimental evidence¹¹⁹.

Upon the 5'exon cleavage, at C complex, a protein-mediated structural remodeling releases the BPA and guide the recruitment of 3'exon, forming the C* complex. Within the same active site used for branching, the exon ligation can take place. The chemistry of the second step of splicing has been investigated computationally through QM/MM and MD simulations, revealing an associative mechanism consistent with the classical two-metal-ion model¹²⁰. In this case, a scissile-phosphate-mediated proton transfer from the nucleophile to the leaving group occurs via a metastable phosphorane-like intermediate. Thus, while the first step of splicing in both systems follows a dissociative pathway, the second adheres to the associative mechanism proposed by Steitz and Steitz.

Cryo-EM structures have proven indispensable for mapping the spliceosome's catalytic cycle. These static snapshots reveal the spatial organization of snRNAs and proteins and serve as starting points for dynamic simulations, which – combined with experimental data – allow mechanistic insights at atomic resolution. However, some intermediate states – such as the B* - remain uncharacterized, leaving gaps in our understanding of key transitions. Notably, clarifying these dynamic rearrangements has significant biomedical relevance. Splicing fidelity is vital for maintaining genome stability, and its disruption has been linked to over 200 human diseases, including cancer and neurodegenerative disorders¹²¹. A deeper mechanistic understanding of spliceosome function is thus not only a fundamental goal of molecular biology, but also a critical step toward the development of RNA-targeted therapeutic strategies.

1.4. Splicing Modulators as Therapeutic Agents

Splicing represents an attractive intervention point for therapeutic development, as it occurs early in gene expression and does not involve permanent genomic alteration. This makes it particularly appealing for targeting diseases driven by so-called "undruggable" proteins. Moreover, approximately one-third of disease-causing mutations affect splicing signals or regulatory elements, resulting in aberrant mRNA isoforms that underlie numerous genetic disorders – including spinal muscular atrophy, Duchenne muscular dystrophy, and Huntington's disease – as well as many cancers^{92,121,122}. In oncology, splicing errors are pervasive, with recurrent mutations in core spliceosomal components such as SF3B1, U2AF1, and SRSF2 frequently impairing splicing fidelity and contributing to tumorigenesis, therapy resistance, and disease progression^{123,124}.

1.4.1. Protein-targeting splicing modulators

The first identified splicing inhibitors – derivatives of pladienolide, herboxidienes, spliceostatins and related compounds – belong to the protein-targeting class. These compounds demonstrated potent cytotoxicity in the low to mid-nanomolar range against human tumor cell lines^{125,126}. Mechanistically, they act early in splicing assembly by binding to the SF3B1 subunit within the U2 snRNP, locking it in an open conformation. This prevents the conformational transition needed for stable U2 – pre-mRNA pairing during A complex formation, thereby disrupting early splicing steps. E7107, a semisynthetic derivative of pladienolide, was the first of this class to enter clinical trials. Although it showed initial promise, development was discontinued due to toxicity and limited therapeutic benefits¹²⁷. A derivative currently under investigation, H3B-8800¹²⁸, is orally bioavailable and shows selectivity for cancer cells with mutations on SF3B1, U2AF and SRSF2 splicing factors, common in leukemias. This compound represents a promising next-generation strategy for protein-targeted splicing modulation¹²⁹.

1.4.2. RNA targeting splicing modulators

Promising strategies for therapeutic intervention involve the direct targeting of RNA to modulate splicing^{121,123,130}. This approach is exemplified by the development of treatments for spinal muscular atrophy (SMA), a severe autosomal recessive neuromuscular disorder that affects approximately 1 in 11,000 live births. SMA is caused by the loss-of-function mutation in the SMN1 gene, which encodes the survival motor neuron (SMN) protein – an essential factor for RNA processing and cellular viability across multiple tissues, including motor neurons. The resulting deficiency of functional SMN protein leads to degeneration of spinal motor neurons, culminating in progressive muscle weakness and atrophy¹³¹. Humans possess a paralog gene, SMN2, but a silent C to T mutation at the 5' end of exon 6 produces a weak splice site, leading to exon 7 exclusion. Consequently, the resultant mRNA isoform without exon 7 encodes for an unstable and inactive SMN. Several RNA targeting therapeutic strategies have been developed to correct this splicing defect by either promoting exon 7 inclusion or restoring full-length SMN expression¹³².

The first approved medication, nusinersen, is an anti-sense oligonucleotide (ASO). Antisense therapy is a form of treatment relying on antisense oligonucleotides to target highly specific regions either in the pre-mRNA or mature mRNA to mask enhancer or silencer

sequences near a specific 5'SS, instead of targeting the splice site itself. In the context of splicing, nusinersen acts by displacing the hnRNP A1/A2 splicing repressor from intron 7 of the SMN2 gene, enhancing exon 7 inclusion¹³³. This antisense therapy improved neuromuscular functions in infants with SMA patients and has been approved in 2016¹³⁴. However, due to its inability to cross the blood-brain barrier, it requires repeated intrathecal administration – a typical drawback for ASOs – which is costly, risky and invasive. A second treatment, onasemnogene abeparvovec (marketed as *Zolgensma*), is a gene therapy that delivers a functional copy of SMN1. It is currently approved for infants under two years of age and offers a one-time systemic administration alternative to ASO-based therapy^{135,136}.

In parallel, orally bioavailable and more easily accessible small molecules have been developed to modulate SMN2 splicing by targeting RNA structures and spliceosome interactions^{137,138}. The first attempts used scaffolds containing coumarins and iso coumarins, but they raised safety concerns upon in vitro and in vivo tests¹³⁹. High-throughput phenotypic screening suggested pyridazine as a scaffold for new hits. Multiparameter lead optimization on this scaffolds led to branaplam¹⁴⁰, which was the first oral small-molecule splicing modulator tested in SMA Type I patients. Branaplam acts by stabilizing the interaction between SMN2 pre-mRNA and U1 snRNP, converting a weak 5' splice site into a more efficiently recognized structure. Despite promising clinical performance, it was ultimately discontinued in phase II/III trials due to safety concerns^{141,142}. This setback led to the development of safer compounds based on alternative scaffolds. One such compound, risdiplam¹⁴³, has benzamide core and is a selective SMN2 splicing modifier that binds SMN2 specifically to the RNA helix formed by the 5'SS of exon 7 when in complex with U1 snRNA. Risdiplam stabilizes a bulged adenosine at position –1 of the exon, thereby reinforcing U1 snRNP binding and promoting correct exon 7 inclusion, consequently restoring functional SMN protein levels¹²⁹. The drug was FDA-approved in 2020 for the treatment of SMA across all age groups and disease stages¹⁴⁴. Notably, due to the similar presence of this bulge in other exons of pre-mRNAs involved in several disease causing mutations, this mechanism is thought to be replicable for the design of new RNA-targeting splicing modulators¹⁴⁵.

Importantly, the therapeutic potential of splicing modulation through RNA targeting small molecules extends beyond human genetic disorders. Recent studies have shown that small molecules targeting group II introns in pathogenic fungi – such as *Candida parapsilosis* – can inhibit self-splicing events that are essential for their viability⁸⁹. Because RNA metabolism in fungi differs substantially from that in mammals, these compounds offer a promising path for developing pathogen-specific antifungal agents with reduced host toxicity.

Overall, the success of splicing modulators in treating SMA, combined with emerging applications in cancer, regenerative medicine, and infectious diseases, underscores the clinical value of targeting the splicing machinery^{130,146}. This highlights the need for a continuous increasing of our understanding of the splicing mechanism aimed to provide new generations of therapeutics.

1.5. RNA in Drug Discovery: importance and the need for computational acceleration

The drug discovery process broadly involves three stages: *i*) the discovery phase, where relevant molecular targets and active compounds (hits) are identified; *ii*) the development phase, including preclinical and clinical testing; and *iii*) the registration phase, which allows market approval and clinical use. The costs escalate significantly across these stages, with recent estimates placing the preclinical phase at \$3.4 million, and clinical phases II and III at \$8.6 million and \$21.4 million, respectively¹⁴⁷. Computational approaches have become a cornerstone in modern drug discovery, particularly in the early discovery phase, where they help identify and optimize small-molecule candidates¹⁴⁸. Computer-aided drug design (CADD) have already played a significant role in the investigation of molecules that are currently in clinical use, for example, imatinib¹⁴⁹, zanamivir¹⁵⁰, and nelfinavir¹⁵¹. CADD reduces the attrition rate by quickly assessing the drug likeness of compounds without experimental testing, and filtering the molecules for desirable moieties or ADMET-predicted properties. In particular, *in-silico* Structure-Based Drug Design (SBDD), which relies on the 3D structure of biomolecular targets to identify potential ligands, has been extensively used for protein targets^{148,152}. However, RNA has only recently gained recognition as a viable pharmaceutical target, despite its central role in biology and disease.

The rationale for targeting RNA therapeutically is compelling. First, RNA molecules often offer higher selectivity than structurally conserved proteins, reducing off-target toxicity¹⁵³. Second, RNA allows for indirect modulation of undruggable proteins by interfering at the mRNA level¹⁵⁴. Third, pathogenic RNAs – such as bacterial riboswitches or fungal group II introns – often lack human homologs, providing pathogen-selective drug targets¹⁵⁵. Additionally, ncRNAs – including microRNAs, lncRNAs, and circRNAs – are increasingly implicated in cancer and neurodegenerative diseases, expanding the landscape of RNA-based drug targets (see Chapter 1). While RNA drug discovery still heavily relies on experimental

methods – especially phenotypical screening – the computational acceleration observed for proteins is still far away for RNA, due to the unique challenges that it poses to computational modeling. Compared to proteins, RNA molecules are highly flexible, adopting multiple conformations that fluctuate with their environment. Their structural plasticity is critical for biological function but complicates rational ligand design. Furthermore, RNA's limited chemical diversity, high charge density, and reliance on metal ions and water for structural integrity make suitable binding pockets harder to define and predict. These factors limit the applicability of conventional SBDD tools. The structural gap is evident in public databases. As of June 2025, the Protein Data Bank contains over 205,000 protein-only entries but fewer than 2,100 RNA-only structures, with fewer than 1,000 RNA-ligand co-crystal structures, severely constraining the development of docking, scoring, and free energy prediction tools tailored for RNA.

Several efforts are underway to bridge the structural gap in RNA-targeted drug discovery. Advances in high-resolution structural techniques such as cryo-EM, X-ray crystallography, and NMR spectroscopy have significantly increased the number of experimentally resolved RNA structures in recent years. Additionally, experimental methods like SHAPE, DMS mapping, and cross-linking-based techniques provide secondary and tertiary structure information. These data can serve as structural constraints that enhance the accuracy of computational 3D modeling¹⁵⁶. To complement experimental data, *de novo* RNA structure prediction methods can generate 3D models directly from sequence information. These approaches fall into two main categories: physics-based and fragment-based. Physics-based methods use parameterized energy functions within MD or Monte Carlo frameworks to identify the most stable (lowest free-energy) conformation of an RNA sequence. Available physics-based tools include iFoldRNA¹⁵⁷, NAST¹⁵⁸, and SimRNA¹⁵⁹. These programs have been successfully tested on RNAs up to ~150 nucleotides, reproducing near-native 3D conformations with reasonable accuracy¹⁵⁸. Fragment-based methods, like FARFAR, exploit RNA modularity, assembling full structures from short subsequences matched to known 3D motifs¹⁶⁰.

Nevertheless, modeling a single static structure is often insufficient. RNA flexibility requires some methods able to capture its intrinsic dynamics. In this context, MD simulations are often used to generate conformational ensembles, which can be used to dissect the dynamic behaviour of RNA targets, and RNA-small molecules complexes, or for example applied in ensemble-based virtual screening (EBVS)¹⁵³. In EBVS, compounds are docked across multiple conformations, increasing the likelihood of identifying viable binders. This approach was successfully applied in a structure-based campaign against the HIV-1 TAR element used MD,

docking, and NMR to identify six small-molecule inhibitors of viral replication¹⁶¹. While EBVS offers improvements, it is computationally demanding and requires reliable structural input. Ligand-based methods can offer alternatives. For example, Informa¹⁶² – a template-based platform – matches RNA secondary structure motifs (e.g., bulges, loops) to known ligand interactions to prioritize “druggable” motifs. It has been used to identify ligands targeting oncogenic miRNA precursors, Parkinson’s-associated SNCA mRNA, and diabetes-related miRNAs.

Beyond virtual screening, accurate prediction of binding affinities is critical to guide lead optimization. Among the available tools, Free Energy Perturbation¹⁶³ (FEP) is the most accurate for estimating relative binding free energies. FEP methods simulate the alchemical transformation of one ligand into another in the binding site, typically using MD or Monte Carlo simulations (See chapter 2.3). While extensively used for protein-ligand systems, FEP has rarely been applied to RNA^{90,164}, largely due to the scarcity of structural data and tested force fields for RNA-small molecules description (see chapter 2 and 3). Moreover, computational tools are increasingly used not only to predict ligand binding but also to clarify mechanisms of action. For instance, in our lab, an integrated study combining crystallography, MD, enzymology, and alchemical calculations elucidated the binding mode of intronistatB, a compound targeting group II introns. In this case⁹⁰, Thermodynamic Integration (TI), a variant of FEP, was used to test candidate binding modes by comparing computed and experimental $\Delta\Delta G$ values, guiding the selection of the correct binding pose.

In conclusion, RNA represents a largely untapped reservoir of druggable targets. However, its unique structural complexity and flexibility require the integration of innovative computational approaches with experimental techniques. Advancements in MD, docking and free energy calculations will be essential to unlock RNA’s therapeutic potential and expand the current landscape of small-molecule drug discovery.

2. Theory and Methods

2.1. Molecular Dynamics (MD) Simulations

2.1.1. *Application of classical mechanics to a molecular system*

A single particle system evolves according to the time dependent Schrödinger equation:

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V \right\} \Psi(\mathbf{r}, t) = i\hbar \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} \quad (2.1)$$

Which describes a single particle of mass m which is moving through space given by a position vector $\mathbf{r} = (x, y, z)$ and time t under the influence of an external field V . $\hbar$ is the Plank's constant divided by 2π , and i is the imaginary unit. $\Psi(\mathbf{r}, t)$ is the wavefunction which characterize the particle's motion. If the potential V is time-independent, the wavefunction can be factorized into a spatial part and a temporal part, yielding the time-independent Schrödinger equation:

$$\mathcal{H}\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \quad (2.2)$$

Where $\mathcal{H}$ is the Hamiltonian operator $\mathcal{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r})$ encoding the total energy of the system, as the sum of kinetic and potential contributions.

For a molecular system of I nuclei and i electrons, the Hamiltonian has the following form

$$\mathcal{H} = -\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{i < j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I < J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} \quad (2.3)$$

Where the first two terms are the nuclear and electronic kinetic energies, followed by nucleus-nucleus attraction, electron-electron repulsion and nucleus-nucleus repulsion

Schrödinger equation can be solved for the simplest molecular systems, such as H_2^+ , using quantum mechanics with the Born-Oppenheimer approximation¹⁶⁵. Because nuclei are much heavier than electrons, the latter adjust almost instantaneously to nuclear motion. The Born-Oppenheimer approximation exploits this separation of timescales, allowing the total molecular wavefunction to be factorized into the electronic part and the nuclei part. With this simplification, the electronic and the nuclei problems can be solved separately. Notably, the electronic wavefunction depends only on the nuclear positions $\mathbf{R}$, not their momenta, and the Schrödinger equation can be written as:

$$\mathcal{H}_e(\mathbf{r}; \mathbf{R}) \psi_e(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R}) \psi_e(\mathbf{r}; \mathbf{R}) \quad (2.4)$$

Where H_e is the electronic Hamiltonian, ψ_e , the electronic wavefunction, and E_e the potential energy surface (PES) as a function of nuclear coordinates only. For each fixed nuclear arrangement, the electronic Schrödinger equation can be solved, and quantum mechanical methods (e.g. ab initio or density functional theory) use this formalism to calculate the electronic structure of molecules.

Under Born-Oppenheimer, the nuclear Hamiltonian is:

$$\mathcal{H}_n = \sum_I \frac{1}{2M_I} \nabla_I^2 + E_e(\mathbf{R}) \quad (2.5)$$

Which describes the nuclear motion on the electronic PES. Moreover, as system size increases, quantum effects on nuclear motion become less dominant, allowing a classical approximation. Thus, biologically relevant molecules can be described with molecular mechanics (MM), where the Born-Oppenheimer approximation is retained, but the quantum PES $E_e(\mathbf{R})$ is replaced by an empirical energy function described by a Force Field (FF)

2.1.2. *Molecular mechanics*

FFs are used to describe the energy of our system, treated with a classical approximation where atoms are treated as rigid beads connected by springs, oscillating around equilibrium values. The change in energy depends on the physicochemical properties of the particles, and their bonds (e.g. hybridization, radius, charge, bond length, angle and torsion) which are determined either experimentally or through QM calculations. Using these parameters, the FF describe how the energy changes as bonds are rotated, stretched, or according to the interactions between non bonded parts of the system. In practice, the potential Energy of the system is calculated by solving the force field equation:

$$\begin{aligned} V(\mathbf{r}^N) = & \sum_{bonds} k_r (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi [1 + \cos(n\phi + \phi_0)] \\ & + \sum_i \sum_{j \neq i}^4 \epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] + \sum_i \sum_{j \neq i}^4 \epsilon_{i,j} \frac{q_i q_j}{\epsilon_0 r_{i,j}} \end{aligned} \quad (2.6)$$

$V(\mathbf{r}^N)$ denotes the potential energy, which is a function of the positions ($\mathbf{r}$) of N particles. the first three terms are the potential energy for bond stretching, angle bending and

proper/improper torsions, respectively. While the last two terms refer to the potential energy for Van-der-Waals interactions, derived from the Lennard-Jones 6-12 potential, and electrostatic interactions, as the Coulomb's law. Importantly, the FF should be considered as a single entity, the energy of its individual components taken separately has not real physical meaning due to the parametrization process.

2.1.3. *The statistical mechanics concept of ensemble*

MD simulations use the FF potential for sampling the behaviour of molecular systems¹⁶⁶, based on statistical mechanics principles, which provides the link between microscopic variables (r, p) and macroscopic thermodynamic observables (N, V, P, T, μ). For a system of N particles, there are $6N$ microscopic degrees of freedom, corresponding to the $3N$ positions and $3N$ momenta. The space defined by these variables is the phase space, and each point in it, labelled Γ , represents a microscopic state of the system, which express any property $A(\Gamma)$ of that configuration. To calculate a macroscopic property from the microscopic knowledge of A in one state Γ , we sample different conformations, creating a collection of points Γ in the phase space. For each state sampled in the phases space, $A(\Gamma)$ can be computed, and the value of this property A is obtained through the ensemble average

$$\langle A \rangle_{ens} = \sum_{\Gamma} A(\Gamma) \rho(\Gamma) \quad (2.7)$$

Where $\rho(\Gamma)$ is the probability distribution associated to the chosen ensemble, that describes how the system is disposed. In molecular simulations, different thermodynamic conditions correspond to different ensembles. For instance, in the microcanonical ensemble (NVE), energy, volume, and particle number are conserved, whereas in the canonical (NVT) or isothermal–isobaric (NPT) ensembles, the system exchanges energy and/or volume with a reservoir. The canonical ensemble is most commonly employed, where the probability density is given by:

$$\rho_{NVT} = \frac{e^{-\frac{H(\Gamma)}{k_b T}}}{Q_{NPT}} \quad (2.8)$$

Here, $H(\Gamma)$ is the Hamiltonian and Q_{NPT} is the canonical partition function:

$$Q_{NVT} = C \int dp dr e^{-\beta H(\Gamma)} = C \int dp dr e^{-\beta [K(p) + U(r)]} \quad (2.9)$$

The partition function formally encodes all equilibrium thermodynamic properties of the system, since it contains the statistical weights of all microstates. Since $H(\Gamma)=K(p)+U(r)$, where $K(p)$ is the kinetic energy and $U(r)$ the potential, QNVT can be factorized into two integrals: an integral over the momenta and an integral over the positions. The former is dependent on temperature and particles mass, has a Gaussian form, hence, can be solved analytically. It is an ideal gas-like contribution, taking into account only the kinetic energy of the particles in the system, considered as non-interacting units. In contrast, the second, dubbed configurational integral (Z), it requires integrating over the $3N$ -dimensional coordinate space of the system. For biomolecular systems, this space is too large and the potential energy function $U(r)$ is highly complex, with many local minima. To overcome this, molecular simulations such as Molecular Dynamics or Monte Carlo are employed to generate representative ensembles of configurations distributed according to the Boltzmann weight $e^{-\beta U(r)}$. In this way, thermodynamic averages are obtained through statistical sampling rather than exact integration.

2.1.4. *MD simulations*

Depending on the choice of temperature and pressure control, different statistical ensembles can be sampled. In all cases, MD simulates the time evolution of a molecular system composed of N particles, generating a trajectory of points in the phase space. Following the time evolution of the system means that the observed value of A (A_{obs}) will be the time average of its values along the trajectory, rather than the ensemble average $\langle A \rangle_{ens}$, as defined in statistical mechanics. The connection between the two is the ergodic hypothesis, which assumes that for infinitely long trajectories, the time average converges to the ensemble average

$$A_{obs} = \lim_{n \rightarrow \infty} \frac{1}{t_0} \int_0^{t_0} dt A(\Gamma(t)) = \langle A \rangle_{ens} \quad (2.10)$$

As the simulation time increases, time averages provide a progressively better approximation to ensemble averages.

To predict the system's time-dependent behavior, MD resolves the motion of each atom with femtosecond resolution by solving Newton's equations of motion for each particle.

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i \quad (2.11)$$

Where F_i is the force acting on an atom i of mass m_i and position r_i . Notably, $F_i = -\nabla_i U$; U is the potential Energy evaluated through the FF. Since position is a vector, for a N particle system, $3N$ second order differential equations should be solved. On the other hand, the problem can be rephrased, starting from

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m} \quad (2.12)$$

So that:

$$\frac{d\mathbf{p}_i}{dt} = m_i \frac{d\mathbf{v}_i}{dt} = \mathbf{F}_i = -\nabla_i U \quad (2.13)$$

Where $\mathbf{v}$ is the velocity of the particle i and $-\nabla_i U$ is the gradient of the potential energy. This give $6N$ first order differential equations, that can be solved numerically by employing the finite difference method, in which from the position and its derivatives at time t , we can estimate them at $t+\delta t$, with δt as a small finite time increment. One of the most common variations of this method is the leap frog algorithm¹⁶⁷. It starts from these two equations:

$$\begin{cases} \mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{1}{2}\delta t)\delta t \\ \mathbf{v}(t + \frac{1}{2}\delta t) = \mathbf{v}(t - \frac{1}{2}\delta t) + \mathbf{a}(t)\delta t \end{cases} \quad (2.14)$$

It updates positions and velocities in a staggered manner, such that velocities are calculated at half time steps and positions at full time steps.

In conclusion, given a model of the system of interest, an MD simulation is initiated by assigning initial coordinates and velocities to all atoms, together with a FF whose parameters are appropriate for the system under investigation. At each time step, the force field computes the forces acting on every atom, which the integration algorithm uses to update positions and velocities at $t+\delta t$. If required, thermostats and barostats are applied to maintain the desired temperature and pressure. Repeated updates generate a trajectory of atomic positions and velocities over time, along which properties of interest can be evaluated. These properties are obtained as time averages, which, under the ergodic hypothesis, approximate the corresponding ensemble averages. In practical terms, this means that values obtained from MD can be directly compared with macroscopic observables measured experimentally.

The choice of the FF is a crucial factor when performing MD simulations. In the case of proteins, continuous refinement over decades has produced highly robust and transferable force fields that reproduce experimental data with good accuracy. In contrast, force field

development for RNA remains an active area of research, and although significant advances have been achieved in recent years, further optimization is still required¹⁶⁸. In this thesis, which focuses on the spliceosome – a very large macromolecular complex with an intricate mechanism of action – RNA and protein simulations have been extensively employed to uncover new mechanistic insights into splicing. For this reason, a dedicated discussion on the reliability and limitations of force fields in these two contexts, together with the influence of ions on RNA dynamics is provided in Chapter 3.

Another central issue concerns the timescales accessible to MD. If we simulate long enough, we should cover the whole phase space, but it is not so straightforward. In fact, some systems may not have enough energy to overcome some barriers, remaining trapped in energy minima. In these cases, the system is still ergodic, but such barriers require too much time to climb, making impossible to exploit ergodicity with affordable computational cost. This is particularly problematic for the sampling of rarely populated conformations, or rare events, which may remain inaccessible under standard MD conditions. To overcome this limitation, a family of enhanced sampling methods has been developed. These approaches introduce an external bias to the energy landscape, thereby accelerating the exploration of conformational states while keeping the computational requirements within reasonable limits. One prominent example of such techniques is Metadynamics¹⁶⁹, which will be described in the next paragraph

2.2. Metadynamics

Metadynamics (MetaD) is a molecular simulation technique designed to accelerate the sampling of rare events and reconstruct the free energy surface (FES). It works by introducing a history-dependent biasing potential, built as a sum of adaptive Gaussian functions deposited along a set of selected collective variables (CVs). Each CV, defined as a function $s(\mathbf{x})$ of the atomic coordinates, reduces the dimensionality of the system, and simplifies the description of the phase space. The system, originally governed by the potential energy $U(\mathbf{x})$, is gradually modified by the added bias, which fills already visited energy wells and forces the system to explore new regions. At the end of the simulation, the cumulative Gaussians provide a direct estimate of the biasing potential applied, which can be used to reconstruct the complementary underlying free-energy surface F .

$$F = -\frac{1}{\beta} \ln \int e^{-\beta U(\mathbf{x})} \delta[s - s(\mathbf{x})] d\mathbf{x} \quad (2.15)$$

Where $\beta=(k_B T)^{-1}$ where k_B is the Boltzmann constant, T is the temperature, and the delta function restraints the free energy evaluation on the chosen CV space.

The total biasing potential added during the simulation is time-dependent, since it changes every n steps. It has the shape:

$$U_g(s(\mathbf{x}, t)) = \omega \sum_{n=1}^{\infty} \exp \left\{ -\frac{[s(\mathbf{x}) - s(n\tau_g)]^2}{2\sigma_s^2} \right\} \quad (2.16)$$

Where τ_g is deposition stride, σ_s is the gaussian width, and ω is a constant energy rate defined as W/τ_g , with W being the Gaussian height. If the CVs are chosen correctly, and the sampling is adequate, the biasing potential satisfies the heuristic condition:

$$\lim_{t \rightarrow \infty} U_g(s, t) \cong -F(x) \quad (2.17)$$

In this ideal case, the FES can be accurately estimated from $U_g(s, t)$, and the probability distribution of the CVs becomes approximately flat, allowing the system to diffuse freely across barriers.

However, practical limitations exist. First, the identification of suitable CVs is not trivial: they must (i) clearly distinguish between relevant states (e.g., folded vs. unfolded protein), (ii) capture all slow or rare processes of interest, and (iii) be limited in number (ideally < 3) to make FES reconstruction feasible. Second, since the FES is unknown a priori, the optimal choice of Gaussian parameters is difficult. Looking at the equation 2.2.2, choosing large ω and σ_s and small τ_g lead to rapid filling of the wells, but results in poor reconstruction of the FES, since the statistical error scales as:

$$Err(F) \propto \frac{W}{\tau_g} \quad (2.18)$$

This correlates to a third issue. Convergence in MetaD is typically defined as the complete filling of the CV space with Gaussians. In practice, however, determining when to stop a simulation is challenging. With insufficient sampling, the heuristic condition for convergence is not fulfilled, preventing the biasing potential from converging to the true FES. Conversely, oversampling while continuing to deposit Gaussians may push the system into regions of phase space that are not physically meaningful. In this context, employing well-tempered MetaD provides a solution¹⁷⁰. In this variant, the height of the Gaussians (W) decreases overtime

according to an exponential scaling of the accumulated bias:

$$W = \omega_0 \tau_g e^{-\frac{U_g(S,t)}{k_b \Delta T}} \quad (2.19)$$

Where ΔT is an artificial bias temperature that controls the rate of decay. The resulting bias potential is expressed as:

$$U_g(s, t) = \Delta T \ln \left(1 + \frac{\omega N(s, t)}{\Delta T} \right) \quad (2.20)$$

Where $N(S,t)$ represents the number of accumulated number of Gaussians at point s in the CV space up to time t . In Well-Tempered MetaD, the bias grows more slowly than in standard MetaD, preventing overfilling of energy wells and avoiding exploration of non-physical regions. At long times, the bias converges to a finite value proportional to the free energy:

$$F(S, t) = -\frac{\Delta T}{T + \Delta T} U_g(S, t) \quad (2.21)$$

Here, ΔT regulates the balance between exploration and accuracy. Large ΔT produces slower decay of Gaussian height, faster exploration but lower accuracy; small ΔT faster decay, slower exploration but more accurate free-energy reconstruction. The two limiting cases are $\Delta T=0$, which corresponds to standard molecular dynamics without bias, and $\Delta T \rightarrow \infty$, which recovers standard MetaD. Intermediate values allow the CVs to be sampled at an effective temperature ($T+\Delta T$), leading the system toward a quasi-equilibrium state in long simulations, where the bias potential serves as an exact estimator of the free-energy surface.

2.3. Free Energy Perturbation (FEP) principles

FEP belongs to the family of alchemical free-energy methods¹⁶³. In these approaches, the main focus lies on estimating the overall free-energy difference associated with a process, rather than explicitly modeling the chemical mechanism itself. Since free energy is a state function, its change depends only on the initial and final states, and not on the physical pathway connecting them. Alchemical methods exploit this property by introducing artificial “non-physical” pathways, which allow for more computationally efficient determination of free-energy differences between end states.

Free energy is inherently an ensemble-dependent property. Within the canonical ensemble (NVT), the relevant thermodynamic potential is the Helmholtz free energy (A), whereas in the isothermal–isobaric ensemble (NPT) it is the Gibbs free energy (G). Both share the same general statistical-mechanical definition, differing only in the form of the partition function Q :

$$A = -k_B T \ln Q_{NVT} = -k_B T \ln \int d\Gamma e^{-\beta \mathcal{H}(p,r)} \quad (2.22)$$

Where $\mathcal{H}(p,r)$ is the Hamiltonian, with contributions from both the momenta p and the coordinates r . Thus, the partition function represents the sum over all possible microstates, each weighted by its Boltzmann factor. The Helmholtz free energy difference between two states A and B is:

$$\Delta A = A_A - A_B = -k_B T \ln \left(\frac{Q_B}{Q_A} \right) = -k_B T \ln \left(\frac{Z_B}{Z_A} \right) \quad (2.23)$$

Here, the separation of the Hamiltonian into momentum and positional terms (as discussed in section 2.1.3 is key. The integration on the momenta can be solved analytically and it depends only on temperature and masses. Since these do not change between states, their contribution cancel when taking the ratio Q_A/Q_B , leaving only the configurational integral Z . This justifies the final equality and explains why, in practice, free energy differences are computed from the configurational space alone. The configurational integral is defined as:

$$Z_{NVT} = \int dr e^{-\beta U(r)} \quad (2.23)$$

Where $U(r)$ is the potential energy. Explicing Z and substituting U_B in eq. 2.3.2, the free energy difference is:

$$\Delta A = -k_B T \ln \langle e^{\beta \Delta U(r)} \rangle_A \quad (2.24)$$

Where $\langle \cdot \rangle_A$ is the ensemble average over configurations sampled from the reference state A . This is the exponential average or Zwanzig equation. Although rigorous, the exponential form is numerically unstable: the average is dominated by rare configurations with strongly negative ΔU , which may not be all captured during finite sampling. If ΔU follows a gaussian distribution – i.e., if there is good phase-space overlap between the two states – a second order cumulant expansion can attenuate such numerical instability:

$$\Delta A = \langle \Delta U \rangle_A - \frac{\beta}{2} (\langle \Delta U^2 \rangle_A \langle \Delta U \rangle_A^2) \quad (2.24)$$

The cumulant expansion rewrites Zwanzig's exponential average in terms of the statistical distribution of ΔU . If ΔU is Gaussian, the expansion becomes exact when truncated at the second order, since higher order cumulants vanish. In practice, this makes the estimate less noisy and numerically more stable than evaluating the exponential average directly. However, both approaches (Zwanzig's and its cumulant expansion) converge only if sampling is sufficient and ΔU is close to Gaussian. Thus, a good overlap between the phase spaces of the reference and target states is essential. To facilitate this, the transformation is often divided into smaller steps using an order parameter λ , with the potential energy as:

$$U(\lambda) = (1 - \lambda)U_A + \lambda U_B \quad (2.25)$$

where U_A and U_B are the potential functions for the two states. Increasing the number of λ -values improves the overlap between neighbouring windows, at the cost of increased computational effort. The parameter λ drives the alchemical transformation, smoothly morphing the FF parameters from those of state A to those of state B. When the two states have different topologies, two strategies are possible:

Single topology: The two states are described with only one topology with dummy atoms that are gradually morphed along the alchemical path. Atoms that exist in state A and not in B are gradually decoupled, while atoms that exist in A and not in B are introduced.

Double topology: Both molecules are present in the simulation box simultaneously. During the transformation, state A's atoms are coupled to the system, while state B's atoms are non-interacting ("ghosted"), and then swapped. This avoids the need for one-to-one atom mapping, and is more suitable for chemically diverse transformations, at the expense of higher computational cost.

To accelerate the convergence of free energy calculations, relative binding free energies between ligands (RBFEE, $\Delta\Delta G_{\text{bind}}$) can be calculated instead of absolute binding free energies

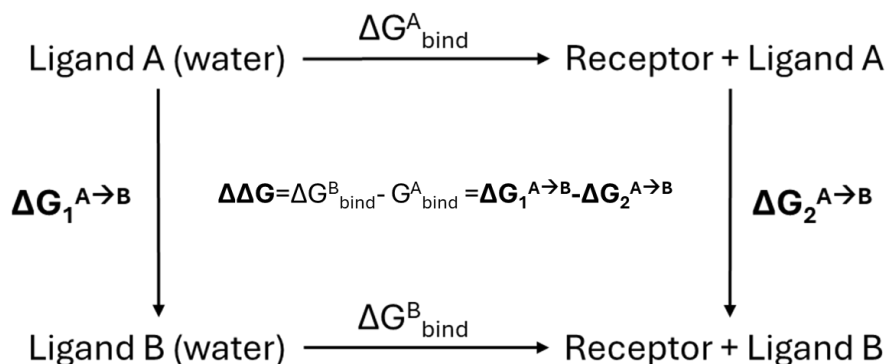


Figure 2.1. *FEP Thermodynamic cycle*

(ABFE, ΔG_{bind}). RBFE exploits the fact that two chemically similar ligands bound to the same receptor typically have much greater phase-space overlap than a ligand in the bound vs. unbound state. RBFE is obtained from a thermodynamic cycle, in which the alchemical transformation of one ligand into another is performed both in aqueous solution (unbound state) and in the receptor–ligand complex (bound state). Since free energy is a state function, the RBFE can be computed as the difference between the two transformation free energies

Considering that all the transformations are reversible, the cycle can be run in both directions, and the free energy of interest can be calculated from either forward or backward transformations. If the same order parameter, λ is used for the forward and the reverse transformations, the changes in the free energy with should be reversible, and the sum of the forward and reverse ΔG should be zero. Deviations from this ideal behaviour define the hysteresis, which constitutes a measure of the error in the free energy calculation. If the hysteresis is markedly larger than the estimated statistical errors, it is usually indicative of ergodicity issues during the transformations, or sampling problems.

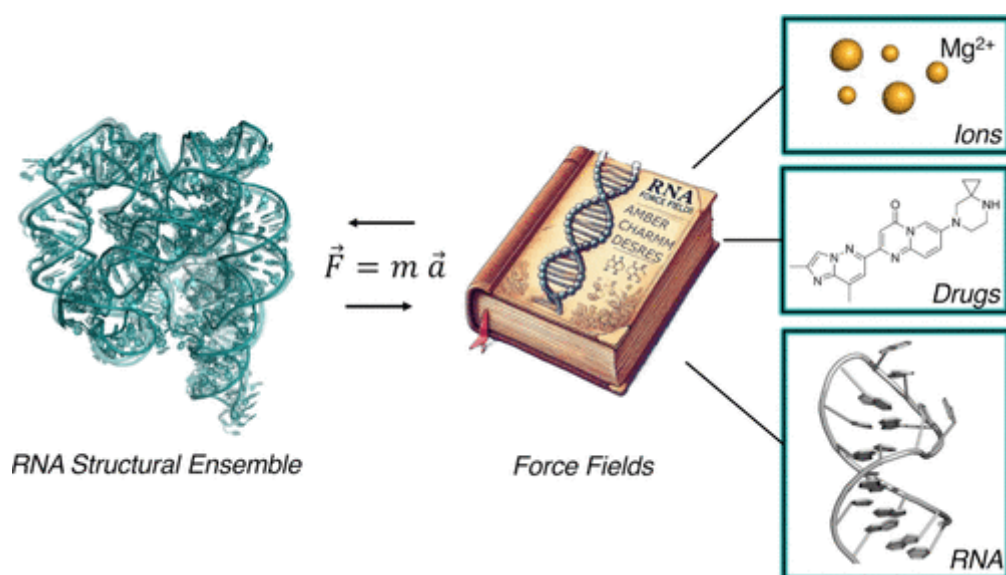
Another way to assess reliability is to use convergence cycles, in which a ligand is transformed through a series of others and back to itself. The cycle-closure hysteresis estimates the statistical error due to incomplete sampling, independent of force field inaccuracies. For a three-ligand cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, the closure free energy is:

$$\Delta G_{\text{closure}} = \Delta \Delta G_{1 \rightarrow 2} + \Delta \Delta G_{2 \rightarrow 3} + \Delta \Delta G_{3 \rightarrow 1} \quad (2.26)$$

Which ideally should be zero.

FEP has been widely and successfully applied to protein–ligand systems, where force fields are mature and benchmark datasets are abundant. In contrast, its application to RNA–ligand complexes is more recent and presents unique challenges due to the greater flexibility and electrostatic complexity of RNA. In this thesis, at chapter 5, we systematically assess the accuracy and reliability of FEP for RNA–ligand systems, testing its performance and identifying the specific limitations and potential of this approach in the context of RNA-targeted drug discovery

3. Advancements, Limitations and usage of RNA Force Field



Reference paper: Stefano Muscat, Gianfranco Martino, Jacopo Manigrasso, Marco Marcia and Marco De Vivo, *On the Power and Challenges of Atomistic Molecular Dynamics to investigate RNA Molecules*, J. Chem. Theory Comput., 2024



An open book spans time. On the left, an aged, yellowed page represents the past of RNA forcefield. The right modern page symbolizes its future. This image was created with the assistance of DALL·E.

This perspective article on RNA simulations was on the cover of the *Journal of Chemical Theory and Computation*, August 27, 2024, Volume 20, Issue 16.

3.1. Abstract

RNA molecules play a vital role in biological processes within the cell, with significant implications for science and medicine. Notably, the biological functions exerted by specific RNA molecules are often linked to the RNA conformational ensemble. However, the experimental characterization of such three-dimensional RNA structures is challenged by the structural heterogeneity of RNA and by its multiple dynamic interactions with binding partners like small molecules, proteins, and metal ions. Consequently, our current understanding of the structure-function relationship of RNA molecules is still limited. In this context, we highlight molecular dynamics (MD) simulations as a powerful tool to complement experimental efforts on RNAs. Despite the recognized limitations of current force fields for RNA MD simulations, examining the dynamics of selected RNAs has provided valuable functional insights into their structures.

3.2. Introduction

Over the years, the resolution of numerous three-dimensional (3D) protein structures has laid the groundwork for studying protein dynamics. These experimental structures serve as crucial starting points for deepening our knowledge of how proteins function at a molecular level. In this framework, theoretical and computational techniques, particularly molecular dynamics (MD) simulations, powerfully complement experimental techniques for unraveling macromolecule dynamics and their functional interactions with other biological partners^{171,172}. MD simulations act as a computational microscope, comprehensively characterizing macromolecule conformational ensembles, thus enhancing our understanding of the intricate relationship between structure, dynamics, and biological function.

Conversely, the scenario for RNA is quite different, as RNA structures have been less frequently resolved, i.e., there are currently only 7'731 deposited RNA structures in the PDB compared to 215'930 protein structures. Moreover, identifying well-resolved ion binding sites in RNA structures remains relatively scarce^{173,174}, partly due to the resolution limits of traditional structural biology techniques such as X-ray crystallography, NMR, or cryo-electron microscopy (cryo-EM) when applied to RNA molecules. In this context, deciphering the possible relationship between RNA structure, its dynamics, and its precise biological functions is still challenging^{175,176}.

However, with the rapid improvements in RNA structure determination, there is a pressing need to enhance RNA MD methodologies. This advancement will ensure that as new RNA structures are experimentally resolved, we can promptly explore their dynamics through MD, which is vital for comprehending their roles in physiological and pathological contexts. RNA plays a crucial role in essential cellular processes, functioning as a carrier of genetic information, a catalyst for biological reactions such as splicing and protein synthesis, and a regulator of transcription, translation, and metabolic protein activity^{88,177–180}. Furthermore, RNA molecules are emerging as an appealing and novel class of drug targets to fight several human diseases, with a few RNA-based or RNA-targeted medications having already received approval from the FDA and a few more undergoing pre-clinical trials^{156,181}.

As a matter of fact, the biological function of RNA can be intrinsically associated with the 3D structure and dynamical molecular interactions, calling for help from MD simulations to dissect the complex RNA structure-function relationship^{182,183}. For instance, catalytic RNAs, such as small nuclear RNAs and ribozymes, need to properly fold in a catalytically competent state to process their reaction substrates. Also, stable 3D non-coding RNA (ncRNA) elements, such as 5'- and 3'-untranslated regions (UTRs) or internal ribosome entry sites (IRES), directly control the recruitment and activation of the translation machinery acting on mRNA molecules, ultimately favoring or disfavoring the synthesis of corresponding proteins. Similarly, riboswitches specifically recognize small molecule ligands, undergo profound conformational rearrangements, and thus control translation and the activation of specific metabolic pathways that guarantee the correct response of the cell to environmental stress^{180,184}. Finally, in the context of mRNA-based vaccines, the structural dynamics of mRNA contribute significantly to the molecular organization of lipid nanoparticles and their physicochemical properties, thereby playing a crucial role in the formulation and efficacy of these vaccine vectors¹⁸⁵.

Given these premises, MD is often elected as a powerful tool for retrospective analyses of the experimental evidence and for generating prospective mechanistic hypotheses of such complex RNA molecules. However, accurate sampling through MD simulations of the RNA conformational space and dynamics requires the correct definition of the potential energy function implemented for MD simulations. Here, we will highlight the limitations encountered over the years in studying RNA dynamics and how these limitations have been tackled by appropriately tailoring *ad-hoc* force field parameters developed to investigate RNA dynamics. Through a series of selected case studies, we will focus on how MD simulations of RNA have been successfully used to elucidate complex molecular phenomena, synergizing with experimental data. Finally, we will highlight some of the most promising advances to overcome

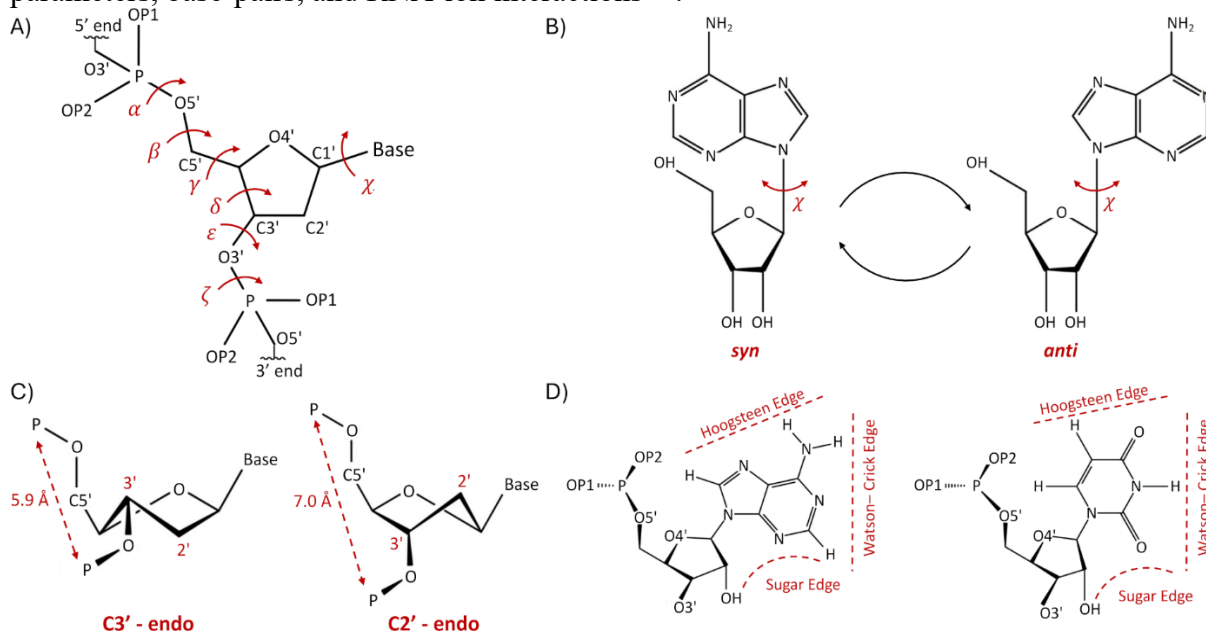
recognized limitations of MD simulations of RNA¹⁸⁶.

3.3. Advances in force field parameters for RNA

Classical atomistic MD stands as a pivotal physics-based method for examining the interactions and movements of atoms and molecules, guided by the principles of Newtonian physics. These simulations are based on the use of a force field, which defines the energy of a system depending on the chemical nature and exact coordinates of its particles. The force field, therefore, contains a set of atomic parameters and *bonded* and *non-bonded* interatomic potentials to define the overall potential energy surface of the system under investigation. Then, the forces acting on every atom are derived as a gradient of such potential energy with respect to the particle coordinates

$$f_i = -\frac{\partial}{\partial r_i} V(r^N) = -\frac{\partial}{\partial r_i} (E_{bonded} + E_{non\ bonded}) \quad (3.1)$$

It is, therefore, evident how accurate force fields are essential for yielding physically meaningful results for the model system under investigation. However, the defining parameters for RNA are uniquely complex compared to other biological systems, making it harder to accurately capture the dynamics of RNA molecules. These critical parameters include backbone torsional angles (α , β , γ , δ , ϵ , ζ), glycosidic torsion angle (χ), sugar pucker parameters, base-pairs, and RNA-ion interactions¹⁸⁷.



In more detail, the six torsion angles of the RNA backbone (**Figure 4.1A**), describing the phosphate-sugar backbone conformation, confer high flexibility. This contrasts with proteins where the backbone is sterically more restricted, and only two torsional angles thus describe it. Also, the glycosidic torsional angle in RNA defines the relative orientation between the base and the sugar moiety. In this case, two are the primary possible rearrangements: the *anti* and the *syn* conformation (**Figure 3.1B**). The predominant conformation is the *anti*, where the base is rotated away from the sugar. However, the less common *syn* conformation is crucial in mediating tertiary interactions at functional sites within RNA molecules¹⁸⁸. In addition, the C3'-endo (A-form) and C2'-endo (B-form) sugar puckers, dictating the ribose sugar conformation, set the inter-phosphate distances, contributing to the distinctive compact nature of RNA structures (**Figure 3.1C**). These conformations are often essential for protein recognition, nucleic acid polymerization, and catalytic reactions^{189–192}.

Furthermore, RNA bases contribute to the stabilization and specificity of folding through two main types of interactions. Stacking interactions occur between consecutive bases, providing sequence-dependent stabilization crucial for maintaining the RNA structure. On the other hand, hydrogen bonds form across three distinct edges—Watson-Crick, Hoogsteen, and Sugar, as shown in **Figure 3.1D** – of which Watson-Crick pairing is particularly integral to conferring the characteristic A-form double helix structure to RNA. However, most bases are involved in non-Watson-Crick (i.e., non-canonical pairings), facilitating specific RNA–RNA self-assembly processes and RNA–protein recognition. Each edge of a base can interact with any of the three edges of another base in either a *cis* or *trans* orientation concerning the glycosidic bonds. This potential for interaction results in twelve distinct edge-to-edge pairing geometries¹⁹³.

Other RNA structural peculiarities include the 2' hydroxyl group (2'-OH) on the sugar edge, which endows RNA base pairing with additional versatility. The 2'-OH abilities of both H-bonds acceptor and donor are often exploited in the stabilization of non-canonical base pairs, also through bifurcated H-bonds¹⁹⁴. Finally, we mention RNA-counterions interactions, which can also critically influence RNA structure and function. For example, positively charged ions, diffused around the entire RNA structure due to its large negative charge, screen the Coulomb repulsion between the overall negative charge of the RNA backbone, allowing tertiary contacts. Additionally, ions can sometimes bind in specific functional sites and stabilize active RNA conformations that, for example, enable catalytic reactions, as occurs in ribozymes¹⁹⁵.

Over the years, significant progress has been made for RNA force fields. Three major force

fields currently used to investigate RNA are AMBER, CHARMM, and DESRES^{194,196,197}. Indeed, AMBER is a widely utilized force field by computational researchers worldwide. Its extensive development and continuous refinements since the 1990s also covered the development of *ad-hoc* parameters for RNA molecules. The advancements of the AMBER force field for RNA cover crucial refinements in handling RNA-RNA interactions through *bonded* and *non-bonded* interactions. Furthermore, these improvements were tested over different water and ion models to represent the interactions more accurately between RNA and its surrounding solvent^{198,199}.

3.3.1. *Bonded interactions refinement*

Since the introduction of the initial AMBER99 version²⁰⁰, numerous variants have been developed, and various parameters have been integrated to address and overcome the initial limitations (**Figure3.2**). In fact, initially, 50 ns of unbiased MD simulations using the AMBER99 force field revealed severe distortions in the DNA double helix, primarily caused by unrealistic transitions in the α/γ dihedral angles²⁰¹. One of the most notable advancements is the refinement of AMBER99 explicitly aimed at enhancing the α/γ torsion terms within the AMBER parameters bsc0 (parmbsc0). This crucial development significantly improved the precision in modeling nucleic acid conformations, creating the AMBER ff99-BSC0 forcefield²⁰².

Subsequently, emphasis has been put on the glycosidic dihedral angle χ , after observing that MD simulations overestimated the population of *syn* conformations of the base orientation and of C2'-endo sugar puckering of the pyrimidines, when compared to NMR data. Among various χ_{OL} reparameterizations, the χ_{OL3} for the AMBER99 RNA ff corrected such a destabilization issue, improving the *syn-anti* balance in RNA and preventing spurious ladder-like distortions^{203,204}. Optimized for use with AMBER ff99-BSC0, it is based on high-level quantum mechanics (QM) calculations. It incorporates crucial solvation-related effects, although it is not recommended for canonical B-DNA structures, leading to the creation of AMBERff12 force field.

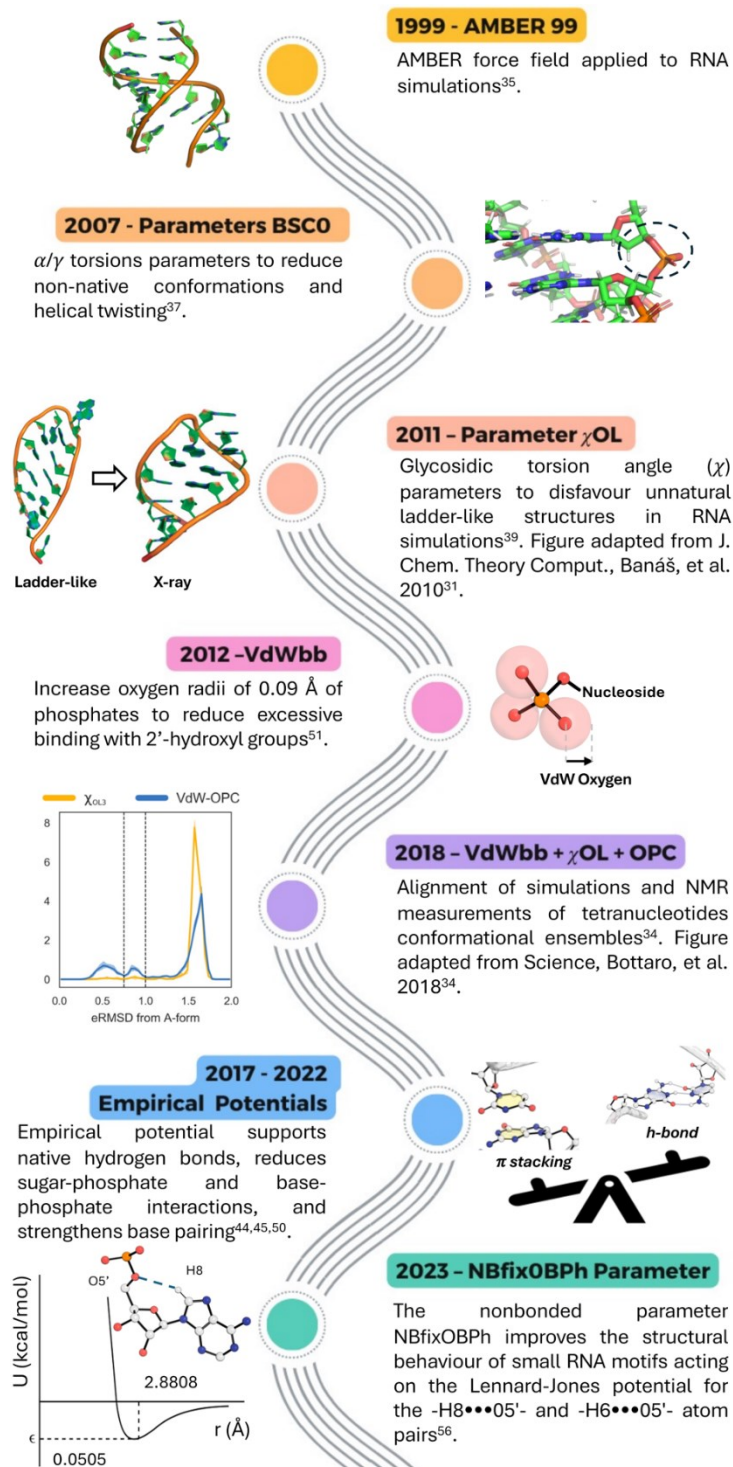


Figure 3. 2. **Main Amber RNA FF advancements**

Nonbonded interactions refinement

Over the years, several issues were highlighted on non-bonded interactions, based on the underestimation of base pairing and overestimation of the base stacking²⁰⁵, and artificially strong charge-charge interactions leading to unrealistic aggregation of RNA-protein and DNA-DNA biomolecules^{206,207}. Bottaro et al. found criticisms that included an imbalanced

distribution of RNA torsional in AMBER99SB with parmbsc0 and χ_{OL3} corrections, with failure in identifying the most stable structures for 5' untranslated region tetraloops in the SARS-CoV-2 genome, and an overestimation of base-stacking¹⁹⁹. Additionally, recent insights from NMR studies have highlighted specific inaccuracies in force fields, particularly in predicting stable hydrogen bond formations in certain intercalated conformations¹⁹⁹. Further efforts have been made to refine nonbond interactions in RNA simulations evaluating intramolecular hydrogen bonding, base stacking, protein-nucleic acid interactions, and various thermodynamic properties^{207–211} (**Figure 3.2**). In detail, Steinbrecher et al. focused on increasing the van der Waals radii of phosphate oxygens²¹² (vdW_{bb}), aiming to reduce excessive binding between 2'-hydroxyl groups and phosphates. An increase in oxygen radii of 0.09 Å was beneficial for the biologically relevant negative charge states. Notably, double negative compounds necessitated tailored adjustments to this oxygen radii. The adjustment successfully reproduced the experimentally available solvation free energies and maintained thermodynamic consistency, as validated using the thermodynamic Pearson cycle.²¹² This was later supported by replica exchange MD (REMD) simulations of tetranucleotides¹⁹⁸. These have highlighted the enhanced accuracy achieved by AMBER ff12 through adjustments to the vdW_{bb} ^{212,213}, particularly when used in tandem with the optimal 3-charge 4-point (OPC) water model²¹⁴. This advancement has been further validated by the closer alignment between computational predictions and NMR observations, underscoring the superiority of the $vdW_{bb} + \chi_{OL3}$ with the OPC water model over the χ_{OL3} parameter alone¹⁹⁹. To refine the balance between stacking, hydrogen bonding, and solvation, another improvement has been reparameterizing VdW terms for nucleobases and applying distinct empirically fine-tuned VdW combination rules for base – water interactions²⁰⁹. Primarily aimed at reducing stacking overstabilization, the modification has the side effect of reinforcing base pairing. However, it reproduced the folded structure of 8-mer RNA tetraloops via T-REMD, starting from an unfolded configuration.²⁰⁹ Notably, Šponer and colleagues have proposed fine-tuning non-bonded interactions by introducing a local auxiliary potential (HBfix, gHBfix19, and gHBfix21) that supports native hydrogen bonds, reduces sugar-phosphate and base-phosphate interactions, and strengthens base pairing²⁰⁸. These potentials are designed to be integrated with χ_{OL3} corrections²⁰³ complemented by modified phosphate parameters and OPC water model. Subsequently, the gHBfix term was combined with another term called tHBfix, specifically designed to address the hydrogen bond interactions of terminal nucleotides. This innovative approach fine-tuned the simulations of RNA tetranucleotides, ensuring precision without impacting the simulation fidelity of other RNA structures²¹⁵. Finally, Mlýnský et al. proposed

to solve the minor intranucleotide clashes in the *anti* nucleobase conformation. They identified that the root cause of these steric clashes was a weak hydrogen bond between the O5' phosphate oxygen and the -C8H8 and -C6H6 groups of purines and pyrimidines, respectively. This issue was mitigated by adjusting the nonbonded parameters (NBfix_{0BPh}), specifically targeting the Lennard-Jones potential for the -H8...O5'- and -H6...O5'- atom pairs.^{212,216}

3.3.2. *Water models*

Alongside the adjustments made in non-bonded interactions, the importance of the water model in RNA simulations has also come to the forefront. Water molecules influence RNA dynamics, helping stabilize their structure and acting as a buffer in mediating RNA-ion interactions. Moreover, the parameterization of ions also plays a crucial role in performing reliable MD simulations. The importance of accurately parameterized ions will be further discussed in the context of the case studies detailed in the next section.

Recent studies have highlighted that the outcomes of RNA simulations are significantly influenced by the choice of the explicit water model, underscoring its substantial impact²¹⁷. While changing water models alone may not fully address all the accuracy challenges in RNA simulations, the role of water in these simulations cannot be overlooked. In this context, Bergonzo et al. compared 4 different water models, TIP3P²¹⁸, SPC/E²¹⁹, TIP4PEw²¹⁹ and OPC²¹⁴, with two force fields ff12 and ff12 + vdW_{bb}²¹² running multidimensional replica exchange MD (M-REMD) simulations. The water models were tested on over 192 MD replicas on two tetranucleotides, of which the conformational NMR ensemble is known¹⁹⁸. The results suggested that the specific combination of water model and nucleic acid force field can drastically influence the final conformational ensemble (**Figure 3.3**). The authors concluded that the OPC water model with the ff12 + vdW_{bb}²¹² RNA force field modifications was the optimal combination to represent the RNA tetranucleotides structural ensemble. Additionally, addressing the imbalance between water–water, and water–solute interactions, the TIP4P-D water model was introduced, revisioning the AMBER ff14 RNA force field²²⁰. Furthermore, this model plays a crucial role in alleviating the issue of overly stable stacking interactions. The revised force field demonstrated substantial improvements, including reversible

hybridization of RNA duplexes and unfolding of long ssRNAs, tetraloops, and riboswitches²¹¹.

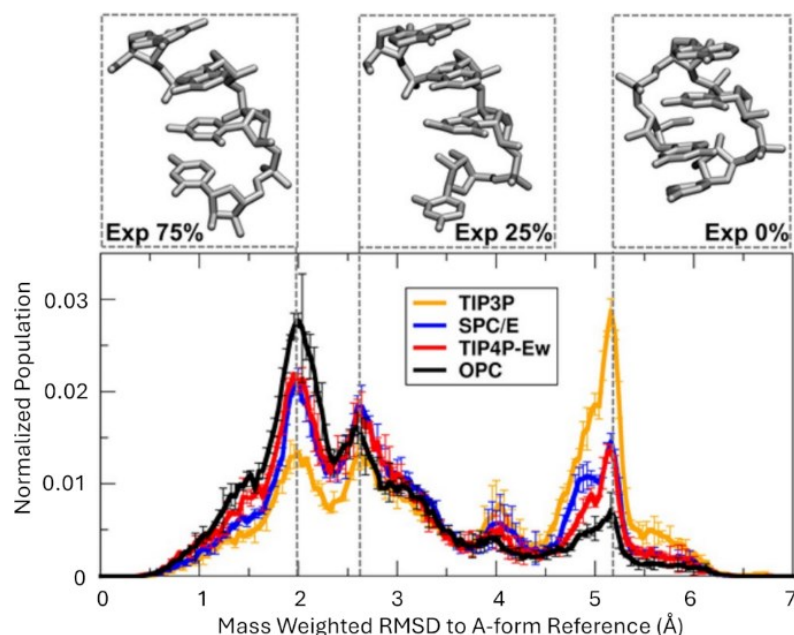


Figure 3. 3 *Amberff12 RNA simulation across 4 water models.* Mass-weighted RMSD histograms depicting the structural distribution as a function of deviation from the A-form reference, for simulations using AMBERff12 with vdWbb across four water models. Figure adapted with permission from *J. Chem. Theory Comput.*, Bergonzo, *et al.* 2015³². Copyright 2015 American Chemical Society.

3.3.3. Other FFs for RNA

Parallel to the AMBER force field development, significant contributions have been made by the DESRES and CHARMM communities. The DESRES force field, developed by De Shaw and colleagues, represents a revised version of AMBER ff14²¹¹. It aims to rebalance stacking and base-pairing interactions by recalibrating nucleobase charge parameters, van der Waals interaction coefficients, and fine-tuning χ , γ and ζ torsional parameters. DESRES was validated using $\sim 100 \mu\text{s}$ MD simulations of various RNA structures, demonstrating effective capture of disordered ssRNA properties, even if with a tendency to over-stabilize double-helical conformations²¹¹. Despite its successes, recent tests revealed limitations in correctly folding RNA kink-turns and the ribosomal L1-stalk RNA segment²²¹. This led to the development of DES-Amber, a new version that addresses these issues by transferring non-bonded RNA parameters and rescaling dihedral terms.

Likewise, the CHARMM force field has undergone several iterations, each addressing specific aspects of nucleic acid behavior. CHARMM22 established the foundation of CHARMM RNA force field, deriving bonded parameters from both experimental data and quantum mechanical calculations. Its nonbonded parameters were designed to reproduce

quantum mechanical interactions between water and model compounds such as nucleic acid bases and dimethyl phosphate, as well as experimental gas-phase properties of base pairing²²². CHARMM27 further refined these parameters, with a particular focus on improving sugar puckering description, aiming to accurately represent both A-form and B-form conformations. This version showed promising results in simulating the B to A form transition under varying solvent conditions. However, longer simulations revealed limitations, particularly in representing BI/BII states, defined by the difference between the ϵ and ζ torsion angles (BI: $\epsilon - \zeta < 0$, BII: $\epsilon - \zeta > 0$). While CHARMM36 improved upon its predecessor, particularly in BI/BII balance, challenges persist in representing local distortions and twisting of the DNA double helix²²³. Notably, the CHARMM family also includes Drude-2017²²⁴, the first polarizable force field for nucleic acids, which incorporates charge transfer mechanisms through a classical Drude oscillator model²²⁴, crucial for RNA-metal interactions. However, Drude-2017 has shown discrepancies in replicating experimentally observed RNA conformations, often favoring hydrogen-bonding and base-pairing interactions over π - π stacking²²⁵. Also, such force field is costly compared to classical force fields, thus the concrete applicability of the polarizable force fields is currently limited for small-to-medium RNA systems where fast convergence of simulations can be achieved. Importantly, both DESRES and CHARMM force fields continue to evolve, therefore contributing to our understanding of nucleic acid dynamics and interactions.

3.4. Case studies in RNA MD simulations: miscellaneous targets and emerging questions

In recent years, RNA force field accuracy has been remarkably improved thanks to the dedicated efforts of the scientific community, unlocking new frontiers in the understanding of the RNA structure-function relationship and RNA structural dynamics. For example, these advancements set the stage for unraveling the intricacies of ion interactions with RNA assemblies^{226,227}, which play a pivotal role in stabilizing RNA folding, structural remodeling, and mediating interactions with proteins^{228,229} and small molecules^{230,231}. Moreover, state-of-the-art RNA force fields have been reliably used to elucidate the RNA-small molecule recognition mechanism, showing how small ligands can regulate the activity of intricate RNA molecular machinery^{90,161,230,232,233}. In the following sections, we will delve into some of the

most significant case studies (**Table 3.1**) that exemplify these topics and their investigation via MD simulations and experiments, impacting our understanding of RNA biology.

Table 3.1: Summary of analyzed case studies. The diversity of examined RNA systems is highlighted and the rational selection of simulation parameters. Each row details the type of study, the RNA system studied, the force field used, the water model employed, molecular dynamics (MD) specifics, and corresponding references

Case study	RNA System	Force fields	Water Model	MD	Ref.
RNA-Ions interactions	Varkud satellite ribozyme	AMBER99 + parmbsc0 + χ_{OL3} Joung and Cheatham parameters ¹⁰⁵ (monovalent ions) Li ¹¹⁰ , Panteva ⁹⁶ parameters (divalent ions)	TIP4P	Classical	226
RNA-Ions interactions	A-form RNA double helix, Telomerase RNA pseudoknot, Ribozyme junction, riboswitches.	AMBER99, AMBER99 + parmbsc0 + χ_{OL3} Allnér Parameters ¹⁰⁰ (divalent ions)	TIP3P	Classical	173
RNA-Ions interactions	Tetranucleotide RNA duplex	AMBER99 + parmbsc0 + χ_{OL3} Allnér Parameters ¹⁰⁰ (divalent ions)	-	Classical, Be-MetaD	234
RNA-Drug-Ions interactions	Env9b SAM/SAH riboswitch	AMBER99 + parmbsc0 + χ_{OL3} Li and Merz Parameters ¹¹⁰ (divalent ions)	TIP3P	Classical	235
RNA-DNA-Protein-Ions interactions	RNA ribonuclease H1	ff14SB ¹⁵⁶ + χ_{OL3} Joung and Cheatham parameters ¹⁰⁵ (monovalent ions) Åqvist ¹⁰⁶ , Allnér ¹⁰⁰ , and Panteva ⁹⁶ (divalent ions)	TIP3P	Classical, MetaD	116
RNA-Ions interactions	Group II intron ribozymes	AMBER99 + parmbsc0 + χ_{OL3} Joung and Cheatham parameters ¹⁰⁵ (monovalent ions) Li and Merz parameters ¹¹⁰ (divalent ions)	TIP3P	Classical, MetaD	81
RNA folding	RNA stem loops in the 5' untranslated	AMBER99SB + parmbsc0 + χ_{OL3} Joung and Cheatham	OPC	Classical, Partial tempering	208

	region of SARS-CoV-2	parameters ¹⁰⁵ (monovalent ions)			
RNA-Ions interactions, RNA folding	RNA hairpin motif	AMBER14	TIP4P-D	Classical	
RNA folding	RNA hairpin motif	AMBER14	TIP4P-D	DPMD	[NO_PRINTED_FORM]
RNA-Drug interactions	Riboswitches, HIV1- TAR, HCV-IRES-IIa, IAV promoter	AMBER99SB-ILDN + parmb0 + χ_{OL3} Allnér, Joung and Cheatham parameters ^{100,105}	OPC	MetaD	Fare clic o toccare qui per immettere il testo.
RNA-Drug Interactions	Group II intron ribozymes	AMBER99SB + parmb0 + χ_{OL3} Joung and Cheatham parameters ¹⁰⁵	TIP3P	Classical, HREX	82
RNA-Drug Interactions	(TEP)-binding aptamer RNA	AMBER99 + χ_{OL3}	TIP3P	GaMD	121
RNA-Ions Interactions	GTPase-associated center	AMBER99SB + parmb0 + χ_{OL3} Allnér, Joung and Cheatham parameters ^{100,105}	OPC	Classical, HREX, MaxEnt	126
RNA folding	RNA tetranucleotides	AMBER99 + parmb0 + χ_{OL3} , AMBER99 + parmb0 + χ_{OL3} + VdW Oxygen correction	TIP3P, OPC	REMD, MaxEnt	199

3.4.1. Influence of ions on the dynamics of RNA

Modeling ions in the context of RNA structural dynamics is a hurdle when using classical force fields²³⁶. This is primarily due to the role of the polarizability of charged groups and ions, as well as charge transfer phenomena that are inherently overlooked in classical MD¹⁵⁶. These phenomena are crucial in RNA-ion interactions, which are pivotal for stabilizing unique RNA structures^{237,238}, promoting folding and conformational transitions, aiding the recognition of binding partners^{198,238,239}, and facilitating catalysis²⁴⁰. Within these limitations, Mg^{2+} cations are RNA structures' most prevalent metal ions. Mg^{2+} is well-known to control both RNA folding and RNA-mediated catalysis²⁴¹. However, Mg^{2+} ions are notoriously difficult to detect in standard X-ray crystallography because they scatter X-rays similarly to water molecules and Na^+ ions²⁴¹, and they do not significantly alter spectroscopic readings, making it hard to identify their specific binding sites through spectroscopic signatures²³⁶. Furthermore, RNA catalysis investigations frequently employ metals such as Ca^{2+} that inhibit catalytic activity, preventing

accurate characterization of Mg^{2+} binding sites[NO_PRINTED_FORM]. In this context, multiple Mg^{2+} parameters have been developed to monitor the diverse effects of the cation on RNA structures²³⁶. The following case studies will demonstrate the adaptability and utility of MD simulations in the elucidation of intricate phenomena involving RNA-cation interactions^{81,116,173,236,240,242}.

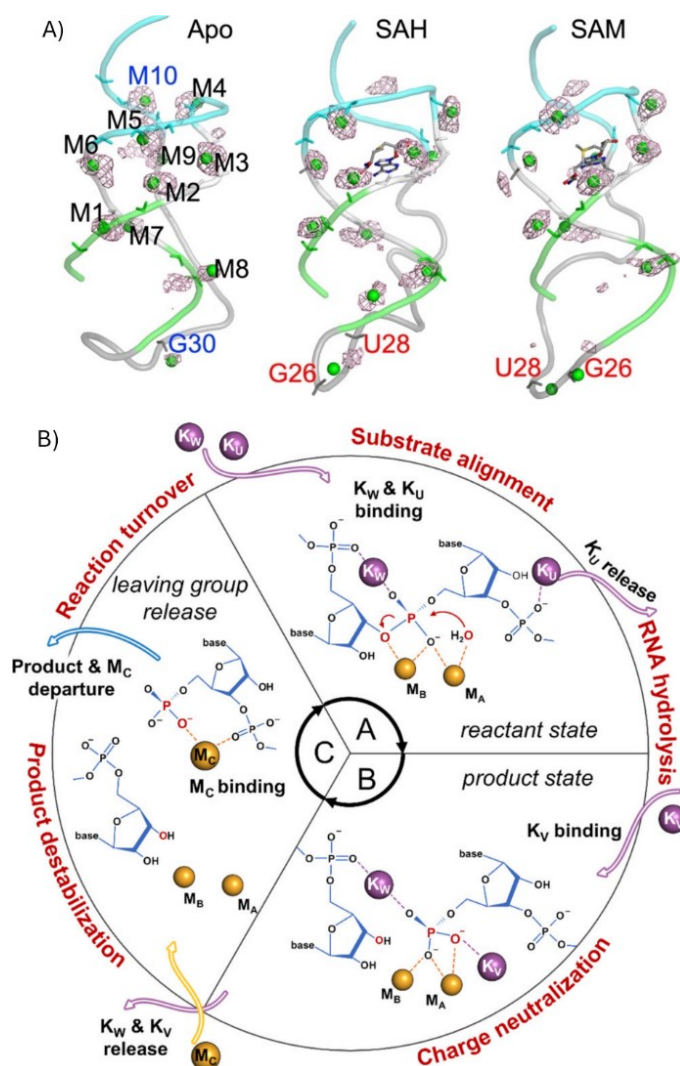
For instance, a combination of computational and experimental data converged on a configuration of interactions within the active site of the Varkud satellite (VS) ribozyme, which is the largest known self-cleaving ribozyme and among the largest endonucleolytic ribozymes responsible for site-specific scission of multimeric VS RNAs into linear monomers²²⁶. Classical MD simulations create a dynamic model of VS catalysis that integrates and expands the existing experimental data, converging on a functionally inactive state of the ribozyme in the absence of Mg^{2+} . MD simulations and molecular solvation theory calculations identified a Mg^{2+} binding site within the active site. This site interacts with the pro-SP oxygen of the scissile phosphate, aiding in the organization of the active site. Notably, no divalent metal ions were detected in the VS ribozyme crystal structures, likely due to the high concentrations of ammonium ions during crystallization. Overall, molecular dynamics simulations employing the AMBER99 force field, combined with parmbsc0 and χ_{OL3} corrections, TIP4P water model, and RNA-tailored divalent cations parameters¹¹⁹, elucidate the configuration of the catalytic site and reveal conformational changes in the transition states, thereby enhancing our understanding of VS ribozyme catalysis.

Fischer et al. elucidated the distinct structural role of mono- and divalent metal ions over 24 RNA structures through equilibrium MD simulations¹⁷³. The dataset comprised twelve helical and twelve complex folded RNA structures. Helical structures were defined as those forming a single double helix consisting of one or two nucleotide strands emblematic of simpler structural motifs. In contrast, the complex folded RNA structures, characterized by their more globular shapes, often encompass multi-helical configurations and include categories such as ribozymes and pseudoknots. Initially employing the AMBER99²⁴³ force field and TIP3P water model, they considered four different ion concentrations, conducting three replicas for each condition. This rigorous approach resulted in 288 simulations, cumulatively spanning 14.4 microseconds of simulation time. However, the authors employed the parmbsc0²⁰² + χ_{OL3} ²⁰³ corrections for a subset of these structures, which demonstrated fewer structural changes and lower RMSD values than structures simulated under identical ion conditions with the original AMBER99. The predicted ion binding sites in RNA structures simulated with Mg^{2+} ions closely matched the experimental data, achieving accuracies up to 1-2 Å. Also, their results

implementing Mg^{2+} ions matched the findings of X-ray crystallography better when compared to simulations conducted with Na^+ ions. Their research revealed that both Mg^{2+} and Na^+ ions exhibited increased direct interactions with RNA at lower ion concentrations. However, divalent ions were more effective in stabilizing native RNA base-base contacts for helical structures than for complex folded structures. This was primarily due to water-mediated interactions at sites identified by experimental studies and in accordance with the ‘ion atmosphere’ theory. Although the binding positions of the ions were accurately predicted, the Gibbs activation energies for second shell dissociation/binding of Mg^{2+} ions in contact with backbone phosphate oxygens are underestimated by approximately 24.1 – 26.7 kJ mol⁻¹. Nonetheless, Cunha et al. have shown that MD simulations have reproduced binding affinities of Mg^{2+} for a flexible tetranucleotide RNA duplex, which represents the most frequent motif within ribosomal RNA²⁴⁴. Employing a combination of equilibrium MD and a modified variant of bias-exchange metadynamics (Be-MetaD) using AMBER99 with parmbsc0 and χ_{OL3} corrections, the authors dissected the influence of ion competition, RNA flexibility, and the effects of RNA single-strand compared to double-strand on Mg^{2+} binding dynamics. In detail, the authors shed light on the intricate interplay between RNA conformational dynamics and Mg^{2+} interactions, highlighting that the Mg^{2+} -phosphate interactions drive the affinity. For both direct and water-mediated binding, ion competition between K^+ and Mg^{2+} reduces the overall binding affinity of Mg^{2+} . Furthermore, the study delved into RNA flexibility by comparing scenarios where RNA was either free to move or conformationally restrained, thereby minimizing changes in RNA conformational entropy. The RNA- Mg^{2+} binding affinity decreased under high RNA flexibility while increasing upon transitioning to an RNA double-stranded state, indicating a Mg^{2+} preference for highly structured and stable regions. Notably, the authors used enhanced sampling methods in combination with the Mg^{2+} parameters developed by Allnér et al.²⁴⁵ due to their improved accuracy in modeling the kinetics of Mg^{2+} water-phosphate exchange. This selection of methodology and parameters underscores the importance of specifically tailored parameters to accurately reflect the intricate dynamics of RNA- Mg^{2+} interactions.

In another study, extensive equilibrium MD simulations elucidated the Mg^{2+} essential role in mediating the ligand binding and conformational transition on the env9b SAM/SAH riboswitch²⁴⁶. Mechanistically, the apo-form of the riboswitch engages with the ribosome, initiating the mRNA translation. Conversely, the binding of SAM or SAH induces the conformational remodeling of the riboswitch, preventing the interaction with the ribosome and eventually inhibiting the translation process. Experimental investigations confirmed that the

ligands (un)binding is mediated by Mg^{2+} ions²³⁷, although their precise detection within a resolved structure remains challenging²⁴⁷, requiring high-resolution structures²⁴¹. MD simulations showed that six inner-shell Mg^{2+} coordinate with the RNA backbone surrounding the ligand binding pocket, ultimately expanding the ligand entry pathway in the apo form of the riboswitch and increasing the accessibility of the binding pocket (**Figure 3.4A**).



Upon ligand binding, these six Mg^{2+} contribute to shaping and stabilizing the outer-shell
Figure 3. 4 Mg^{2+} ions mediate ligand binding and conformational transitions. (A) Spatial distribution of inner-shell Mg^{2+} ions from MD simulations of the apo, SAH and SAM liganded conformations, depicted by mesh densities and overlaid on a representative MD snapshot. Green spheres represent the inner-shell Mg^{2+} ions, while the coordinating phosphate groups are displayed as stick. Figures adapted from Panei, et al. 2023¹¹⁹. (B) Mechanism of RNase H1 catalysis facilitated by cation exchange. K_W and K_U cations align the substrate, followed by K_U dissociation, enabling RNA cleavage. Product state: The negatively charged cleavage product is stabilized by K_W and a new K_V cation. The replacement of K^+ ions with a Mg^{2+} ion facilitates product destabilization, and catalytic turnover. Figure reprinted with permission from Manigrasso, et al. 2021. Copyright 2021 American Chemical Society

coordination with the ligand and additionally reinforcing the stability of an RNA-ligand hydrogen bond²⁴⁵. Another notable example is the use of MD simulations to explore cation

dynamics at the edge of the RNA-protein interaction interphase of the RNA ribonuclease H1 (RNase H1), a metal-dependent endonuclease¹¹⁶. The integration of reaction intermediates obtained by time-resolved X-ray crystallography with extensive equilibrium and non-equilibrium MD simulations, have shed light on the dynamics and functions of second-shell metal ions, i.e., K^+ and Mg^{2+} , in the catalytic cycle of RNase H1 (**Figure 3.4B**). MD simulations, using Joung and Cheatham parameters^{248,249} for monovalent ions, accurately replicated K^+ recruitment and binding at the active site of the enzyme. Further *in silico* investigations, including protein point mutations and variation of the buffer ionic strength, confirmed the role of specific RNase H1 residues for cations recruitment and the influence of ions concentration on K^+ binding events. When modeling divalent cations, three distinct parameter sets, namely Åqvist²⁵⁰, Allnér²⁴⁵, and Panteva¹¹⁹, specifically designed to accurately model RNA-cation interactions, were employed to corroborate the findings and mitigate biases of any particular parameterization. This computational approach provided an atomistic explanation of the so-called 'attenuation effect', where high Mg^{2+} concentration leads to the recruitment of divalent cations in one specific binding site, ultimately preventing the K^+ ion from binding which is essential for catalysis. Overall, these simulations complemented well experimental data²⁵¹, enriching the mechanistic rationale for such complex biochemical phenomena.

Notably, MD simulations have been combined with X-ray crystallography and enzymatic assays to characterize the splicing progression of group II intron ribozymes²⁴², which are large RNA ancestors of the human spliceosome. In this study, the use of equilibrium MD and MetaD simulations was pivotal to characterize the binding and controlled release of one K^+ ion from within the catalytic site (**Figure 3.5A**). These metal binding and unbinding events, occurring at different stages of splicing, were shown to be crucial to guarantee the correct catalytic activation and structural remodeling of the active site for splicing progression (**Figure 3.5B**). Importantly, the synergy between *in silico* and experimental data supported one unified splicing mechanism for both the group II intron and the spliceosome, which are splicing machines that share a highly similar catalytic core architecture. This mechanism has been confirmed by biochemical investigation and high-resolution cryo-EM structures of the spliceosome independently published afterward, as well as by additional *a posteriori* computational studies⁸¹, confirming the predictive power of MD simulations of RNA systems²⁵².

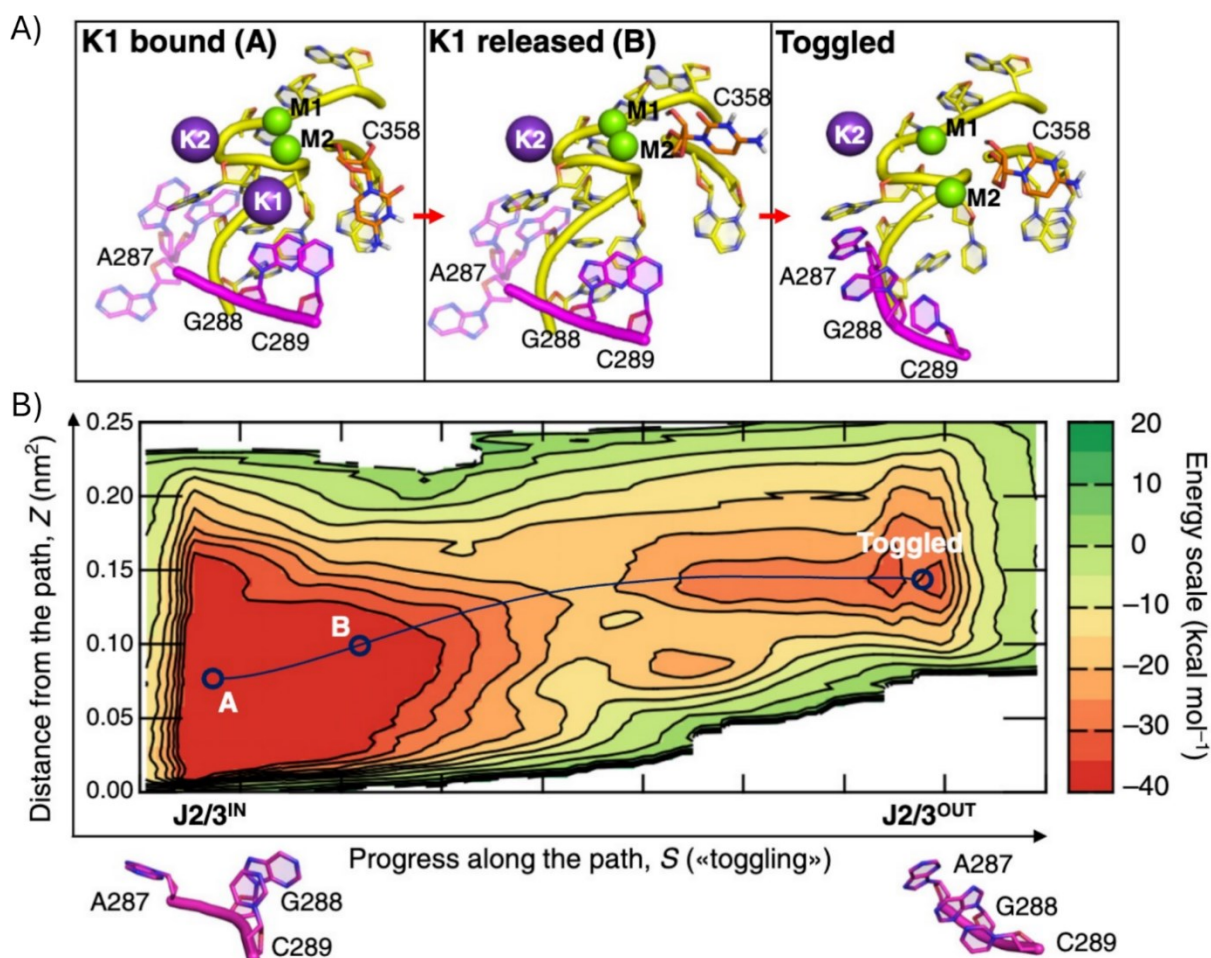


Figure 3.5. **Controlled K^+ ion release in group II intron.** (A) Structural conformation of the active site in “cleaved” (left), intermediate (center), and toggled (right) states as revealed by MetaD simulation. (B) Free energy profile for the cleaved and protonated state, with key intermediates highlighted along the conformational path (dark line). Figure reprinted from Manigrasso, et al. 2020²⁴².

In summary, despite the functional form of MD force field parameters for ions neglects the ion polarizability, various parameterizations exist that accurately reproduce specific experimental properties, as demonstrated from the case studies. Åqvist²⁵⁰ and Allnér²⁴⁵ parameterizations, despite their tendency to overestimate the contact between Mg^{2+} and the oxygen on the bases, were tuned to compare experimental and computed hydration free energy and more accurately reflect the experimental ion distributions associated with RNA. Recently, Li & Merz²⁵³ added an r^{-4} term to the Lennard-Jones model for non-bonded interactions, which allowed better reproducibility of experimental hydration free energy, ion-oxygen distances in the first solvation shell, and coordination numbers. While Åqvist²⁵⁰ and Li & Merz²⁵³ seem superior in capturing the intricacies of highly coordinated RNA sites, Allnér²⁴⁵ and Panteva¹¹⁹ models describe well the diversity and hydration of Mg^{2+} binding sites, including RNA-cations interactions not involving the phosphate backbone¹¹⁹. A careful parameter analysis in the

context of RNA simulations is therefore advisable.

3.4.2. *RNA folding*

Small RNA structures like 5'- or 3' UTR motifs can be dynamic entities, and often, their biological function is exerted by rarely populated conformation²⁵⁴. On the other hand, large RNA structures such as group II intron ribozymes can show an overall compact architecture, where the folding of functional motifs is preserved. In general, regardless of the length of the ncRNA, many of these molecules need to fold to exert their function. However, experiments and MD simulations showed certain limitations when studying RNA folding. The primary *in silico* constraint lies in the size of the system under examination. Classical MD simulations are adept at exploring RNA flexibility but are limited by the accessible timescale, which typically extends to 1–100 μ s. Modern force fields provide accurate descriptions on a microsecond timescale for small and well-investigated RNA molecules as tetranucleotides, especially when simulations commence from established high-resolution experimental structures and remain within the native state basin^{208,255,256}. This level of accuracy partly stems from these molecular systems frequently serving as benchmarks for force field parameterization. Furthermore, a timescale of microseconds is generally insufficient to capture more extensive structural rearrangements, and employing enhanced sampling techniques or coupling experimental data with simulations becomes necessary.²⁵⁶ Furthermore, observing extensive conformational changes in larger RNA systems within accessible simulation times also requires ensuring the energy profile is accurately depicted in the folded, misfolded, and unfolded states. Despite these challenges, MD simulations have been shown to fairly characterize the structural dynamics of different types of RNA molecules¹⁹⁴.

For instance, Bottaro et al. employed the combination of partial tempering and classical MD to reconstruct the conformational ensemble of five structured stem loops within the 5'UTR of the SARS-CoV-2 genome, which vary in size from 17 to 44 nucleotides¹⁹⁹. The computational protocol employed the Amberff99SB force field with parmbsc0¹⁹³ and χ_{OL3} ¹⁹⁴ corrections, alongside the OPC water model²⁰⁹. The simulations began with extended, single-strand A-form RNA structures. A subsequent pulling phase created a reference structure modeled on an ideal A-form double-stranded RNA template for regions with known secondary structures. Then, each molecular system was investigated by partial tempering simulation involving 20 replicas ranging from 300 K to 1000 K of temperature, with two independent runs per system to guarantee thorough exploration. This meticulous approach, amassing 0.5 ms of cumulative

simulation time, generated *in-silico* structures very close to the experimental reference, having an eRMSD below 0.5 Å from NMR data.¹⁹⁹

Furthermore, unbiased MD simulations have been instrumental in quantifying the impact of Mg^{2+} , K^{+} and Ca^{2+} on the folding of the L4 loop within the magnesium-sensing M-box riboswitch²⁵⁷, which undergoes conformational changes that regulate gene expression by terminating transcription in response to Mg^{2+} concentration variations. This study observed several folding/unfolding events of the L4 loop employing AMBER14 force field²⁰³ in conjunction with the TIP4P-D water model²⁵⁸. The study highlighted the crucial role of Mg^{2+} ions, which, through site-specific binding guided by the spatial arrangement of electronegative nucleobase atoms, play a pivotal role in folding the RNA hairpin motifs. The researchers proposed that without the structured octahedral geometry of water molecules in the metal ion first coordination shell, RNA cannot fold efficiently into its native state²⁵⁷.

In addition, probabilistic Bayesian neural networks were employed to enhance MD simulations through a new method called deep boosted molecular dynamics (DBMD)[NO_PRINTED_FORM]. This approach was applied to study the folding process of three different hairpin RNAs, each containing a distinct tetraloop (GCAA, GAAA, or UUCG). Simulations were executed using AMBER14 force field²¹¹, with initial configurations set to unfolded states. The simulations structures reached up to 1.1 Å of RMSD compared to experimental data and revealed a progressive decrease in the frequency of folding and unfolding events from the GCAA to GAAA to UUCG system, highlighting the importance of base stacking in the tetraloop for RNA folding. The observed folding mechanism was similar for all three variants, beginning with extended conformations and proceeding through the formation of Watson-Crick base pairs to the final base stacking in the tetraloop.

In summary, while the simulation of RNA folding poses significant challenges for large RNA systems due to the limitations of the accessible time scales MD simulations, it does not imply the inadequacy of RNA force fields. Indeed, MD simulations employing RNA force fields for smaller RNA systems have demonstrated remarkable success, ultimately retrieving conformations closely matching those found by experiments. This underscores the capability of current RNA force fields to describe the dynamics of many RNA systems effectively, mainly when the simulation scale is within the temporal reach of MD techniques.

3.4.3. Integrating simulations and experimental data to unveil RNA dynamics

A diverse array of experimental methods has been devised to refine biomolecular structural

models, including X-ray crystallography, NMR spectroscopy, cryo-EM, small-angle X-ray scattering (SAXS), and small-angle neutron scattering (SANS). The abovementioned methods provide insights into the RNA molecular size, shape, atomic and residue proximity, local chemical environment, and global structure. From an ensemble refinement perspective, it appears to be critical to effectively incorporate experimental structural information into computational models²⁵⁹. The integration encompasses two primary aspects: first, a forward model that computes an experimental observable based on a given conformation, and second, an integration approach, such as applying restraints on-the-fly during the simulations or reweighting the results *a posteriori*²⁶⁰ leveraging maximum entropy (MaxEnt) or Bayesian inference^{261,262}. The underlying idea of these methods is to minimally perturb an MD simulation ensemble to align with experimental data (**Figure 3.6A**). Furthermore, accounting for experimental uncertainties is essential for constructing a properly weighted configurational ensemble²⁶¹. In this context, advancements in MaxEnt approach inspired methodologies enabling the on-the-fly estimation of biases from experimental data during simulations^{262–264}. For instance, metainference recently stood out as a noteworthy method^{265–267}. Rooted in Bayesian inference and the MaxEnt principle, metainference adeptly combines prior system information with experimental data while managing the inherent errors in such data. It operates by constructing a noise model dependent on uncertainty parameters, which are inferred alongside the system model. Incorporating these methodologies into several open-source software such as IMP²⁶⁸, ROSETTA²⁶⁹, and PLUMED-ISDB²⁷⁰ has spurred their widespread use. Such strategies significantly improve the exploration of relevant conformational spaces, enhancing the overall efficacy and precision of the simulation processes. For example, SAXS and SANS data, generally referred to SAS, have proven to be highly accurate and predictive approaches to characterize the structures of biomolecules in solution. These techniques are precious for determining the size, shape, interactions stoichiometry, and dynamics of biomolecules under concentrations closely mimicking physiological environments and eliminating the need for labeling agents^{271–274}. Despite their relatively low resolution, the ease and lower cost of obtaining such structural data make SAS appealing compared to higher-resolution methods like NMR and cryo-EM. As a matter of fact, SAS approaches have been applied to a wide range of RNA systems, being particularly useful for characterizing RNA structural dynamics^{272–274}. these data can be integrated with computational models by techniques such as reweighting or on-the-fly enhanced sampling with good forward models^{260,272,275}. Over the years, increasingly accurate forward models have been developed, among the most interesting of which are those that also consider the contribution of the

solvation shell, allowing for significantly improved matching with experimental data^{276,277}.

A noteworthy study investigated the ion-dependent conformational dynamics of the GTPase-associated center (GAC) of the 23S ribosomal subunit, a 57-nucleotide RNA molecule critical for protein translation²⁶⁰. Initial microsecond-long unbiased MD simulations showed that GAC retained its compact crystal structure across various ionic conditions, suggesting limitations in exploring the full spectrum of experimental conformations through unbiased MD. The study employed enhanced sampling methods to overcome these constraints, combining HREX with MetaD. The results confirmed that Mg^{2+} ions stabilize the folded state, whereas K^+ ions lead to less compact and more extended conformations. The generated structural ensemble was subsequently reweighted to align with reference SAXS experimental data under various ionic conditions, utilizing accurate explicit-solvent forward models (**Figure 3.6B**). The reweighting procedure, based on the MaxEnt approach, underscored the study's ability to capture the relationship between ionic environment and RNA structure, offering a refined understanding of GAC conformational ensembles.

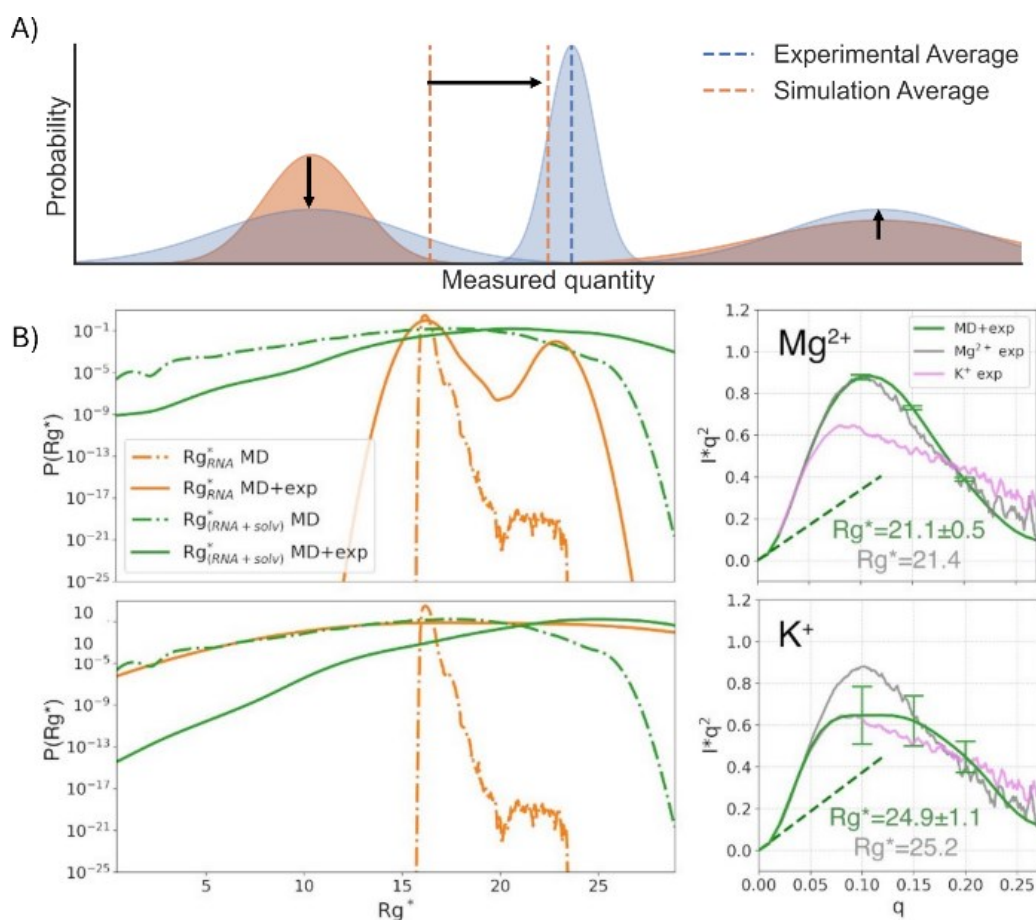


Figure 3.6. Integrating experimental data in MD simulations. (A) In the MaxEnt approach, the computational ensemble undergoes minimal modification that brings the calculated averages to match the experimental data, resulting in an optimal combination of simulations and experiments. (B) Distribution of radius of gyration (Rg^*)

and SAXS spectra for ensembles reweighted according to Mg^{2+} (top panels) and K^+ (bottom panels) experimental data. The left panels display R_g^* values derived from a Guinier fit to SAXS spectra of both the initial (dashed lines) and adjusted (solid lines) ensembles, illustrating data for the solute alone (orange) and the entire system (green). In the right panels, reweighted ensemble SAXS spectra are presented in Kratky plots, with corresponding R_g^* values and their statistical errors (in green) highlighted. Experimental R_g^* values, derived from SAXS measurements, are shown in light gray and violet for Mg^{2+} and K^+ , respectively, and are included for comparison in gray. Figure reprinted from Bernetti, et al. 2021²⁶⁰.

The Integration of SAS data helps to overcome the limitations inherent in MD force field sampling capabilities, ensuring a comprehensive exploration of experimentally relevant conformations. However, interpreting SAS data, especially in the context of RNA molecules, poses unique challenges. Contributions from the hydration layer and ion cloud surrounding RNA molecules can significantly complicate data analysis. The intricate interplay between RNA solutes and their solvent environment can profoundly influence the scattering profile, making applying to nucleic acids particularly daunting²⁷⁸. It therefore becomes evident that integrating SAS data with MD simulations, where the latter detailed insights into atomic interactions and solvent effects can clarify the contributions of hydration layers and ion clouds, thereby enhancing the overall interpretation of SAS data and leading to a more accurate understanding of RNA structures in their physiological environments.

In addition, NMR is another suitable method to unravel the RNA structures in solution. However, it faces challenges with large systems where the complexity and size can hinder the resolution and clarity of structural insights. Integrating atomistic molecular dynamics simulations with NMR spectroscopy data allowed to overcome the limitations of current simulation methods and challenges in interpreting ensemble- and time-averaged signals. The strength of integrating NMR data with MD simulations is exemplified by a study investigating the thermodynamic conformational ensembles of four RNA tetranucleotides¹⁹⁹, explored performing REMD using two force fields: *i*) AMBER99 enhanced with parmbsc0 and χ_{OL3} parameters alongside the TIP3P water model^{202,203}, and *ii*) χ_{OL3} modified with van der Waals oxygen radii corrections and the OPC water model (VdW-OPC)^{216,218}. Applying the MaxEnt/Bayesian approach facilitated the minimal adjustment of the simulated structural ensemble to align with experimental NMR configurations, particularly for the r(CCCC) tetranucleotides. The reweighted ensemble demonstrated a significant shift in conformational population towards the A-form, and well-matched the NMR conformation with an eRMSD below 0.75.¹⁹⁹ The selection of the VdW-OPC force field notably resulted in a closer alignment with the experimental NMR conformation both before and after reweighting, as evidenced by the improved eRMSD for all tetranucleotides systems. Most remarkably, the integration of

NMR observations shed light on specific force field inaccuracies. A key finding was the observation of distance distributions between OP2 in C3 and the hydrogen at the 5' terminus in C1, suggesting stable hydrogen bonds in intercalated conformations that were significantly adjusted following the reweighting process. Additionally, this detailed analysis led to the discovery of shifts in the populations of α and γ torsion angles from gauche⁺ to gauche⁻, indicating a nuanced understanding of RNA structure dynamics. This careful integration of NMR data with computational models transcends mere structural alignment, paving the way for substantial improvements in force field accuracy. By meticulously reweighting based on NMR insights, the study achieved a more reliable reconstruction of the conformational ensemble and highlighted specific force field parameters needing refinement.

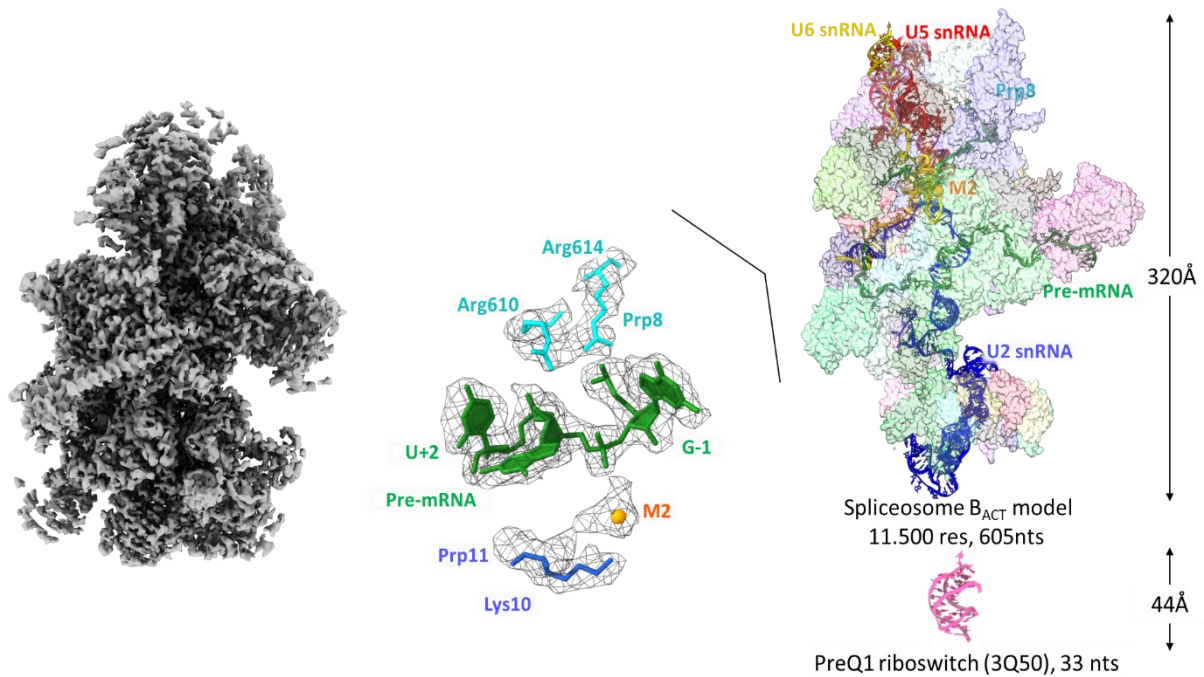
Ensemble-based virtual screening (EBVS), informed by experimental data like NMR¹⁶¹, was illustrated in the case of targeting the transactivation response element (TAR) RNA from human immunodeficiency virus type-1 (HIV-1)¹⁶¹. NMR data were instrumental in refining the MD-simulated ensemble of TAR RNA, characterized by its considerable flexibility, with the pairwise RMSD among conformations exceeding 1.9 Å and averaging 5.9 Å. The effectiveness of this screening approach was underscored by impressive metrics, including Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) values ranging between 0.91 and 0.94 of the active compounds. The results demonstrate that experimentally determined RNA ensembles significantly enrich libraries with actual hits, ensuring a more accurate and relevant virtual screening process for RNA-targeted drug development.

3.5. Conclusions and future perspective

MD simulations of RNA pose significant challenges concerning peculiar and poorly characterized experimentally inter- and intra- molecular interactions, especially in complex systems where RNA is in complex with proteins and multiple metal ions. The flexibility of the RNA backbone, the π - π stacking among aromatic bases, and the hydration of RNA functional groups are areas for improvements of the current force fields^{186,194,279–281}. Nonetheless, force fields show higher reliability and accuracy in structured systems due to error compensation²⁸² that involves mutual offsetting of inaccuracies among various energy terms. For instance, the lack of electronic polarization in many force fields leads to underestimating direct ion–ligand attractions and inter-ligand repulsions. However, this shortfall is mitigated by the concurrent

underestimation of the opposing forces, effectively reducing the discrepancy in the energy landscape between hexa-hydrated Mg^{2+} ions and those partially dehydrated ions coordinated with RNA. Such error compensation supports the overall simulation quality, which often remains robust across different system sizes and complexities^{283,284}. Furthermore, attaining atomistic accuracy in simulations involving the overall dynamics of large and complex structures like the ribosome may not be feasible.^{194,284,285–155} In such cases, the uncertainties inherent in the experimental starting structures could overshadow the errors in the force fields, suggesting a more pragmatic approach in relation to our expectations for applications of MD simulations in extremely complex molecular systems. It is worth noting that current RNA force fields are parameterized based on data from controlled experimental conditions. While highly effective in many contexts, these force fields may face additional challenges in describing the complex cellular environment, potentially opening avenues for future developments. Despite the limitations of RNA force fields, which hinder their generalizability, they have proven capable of providing significant quantitative and qualitative insights into molecular phenomena. Consequently, the meticulous selection of the appropriate force field, tailored to the specific molecular system under study, is essential. In addition, integrating experimental data into computational models is a promising approach to control the limitations of RNA force fields. In this context, as molecular mechanics force fields advance, maximum entropy/Bayesian approach is expected to require less experimental data for reliable determination of structural ensembles, offering a precise tool for exploring RNA conformational dynamics and identifying force field inaccuracies¹⁹⁹. The substantial increase in computational efficiency is also expected to generate more insights from QM calculations and evaluate the force fields over longer timescales and in more complex scenarios than was possible in the early stages of computational modeling, ultimately improving the reliability of computational models.

4. Unraveling RNA Catalytic Mechanisms through Advanced Simulations



Reference paper: Gianfranco Martino, Jacopo Manigrasso, Giuseppina La Sala, Marco Marcia and Marco De Vivo, *Controlled dynamic remodeling of the spliceosome active site enables the first step of splicing*, PNAS, 2026.

4.1. Abstract

RNA splicing removes non-coding regions from precursor mRNAs, a process essential to produce functional transcripts in all living organisms. In eukaryotes, this reaction is catalyzed by the spliceosome, a dynamic RNA–protein complex that assembles and disassembles through tightly regulated steps. The spliceosome is a key therapeutic target, and indeed its early assembly steps can be modulated using FDA-approved drugs like risdiplam to treat congenital disorders. Recent cryo-electron microscopy data have revealed significant conformational rearrangements of the megadalton-large spliceosome structure during splicing, an essential catalytic process for maturing most human transcripts. However, the molecular trigger of these structural rearrangements for splicing catalysis remains unclear at the atomic level. Here, by analysing a consistent dataset of multicomponent spliceosome structures, we identified a minimal set of positively-charged residues whose dynamic action remodels the spliceosome active site during the first step of splicing. Through equilibrium and enhanced sampling molecular dynamics simulations of multiple splicing intermediates (>2M atoms), we uncover how these residues dynamically operate in coordination with the transient, temporally ordered binding of key spliceosomal proteins (Prp11, Prp8, and Yju2) at the spliceosome core. Our findings reveal a molecular mechanism that ensures precise and timely spliceosome activation, opening new directions to probe splicing regulation and potentially target its dysfunction in diseases, with broad implications for drug discovery and medicine.

4.2. Introduction

RNA splicing is the molecular process of removing non-functional regions from precursor transcripts to produce functional protein-coding or non-coding RNAs. This process is thus tightly regulated as it controls the expression of essentially all human genes⁹⁸, leading to severe diseases when it malfunctions. Indeed, the spliceosome – the megadalton-large ribonucleoprotein complex that catalyzes splicing in eukaryotes – is a validated drug target in cancer and neurological disorders^{98,286,287}.

Over the past 30 years, extensive genetic and biochemical studies on the spliceosome have revealed that more than a hundred proteins are involved in the maturation of the spliceosome through four sequential phases: assembly, activation, catalysis, and disassembly. However, due to the molecular complexity of the megadalton-large eukaryotic spliceosome, only recent cryo-

electron microscopy (cryo-EM) efforts have enabled the visualization of crucial splicing intermediates at near-atomic resolutions^{288–290}. Specifically, these intermediates are defined as: *i*) B^{ACT} (activated but not catalytically primed, **Figure 4.1A**)^{95,101}; *ii*) B* (primed for branching before the first step of splicing, **Figure 4.1B**) and its related substructures B*a1, B*a2, B*b1, and B*b2;⁹⁴ and *iii*) C (after the first step of splicing, **Figure 4.1C**)^{96,97}.

These structural data have revealed the molecular architecture of the spliceosomal active site, where the U6 small nuclear RNA (snRNA) folds into a triple helix, connecting nucleotides C58, A59, and G60 from the so-called catalytic triad motif with nucleotides A79 and U80 from the so-called two-nucleotide bulge motif (**Figure 4.1D-F**, nucleotide numbering from yeast spliceosome, PDB ID: 7DCO). These evolutionary and structurally conserved nucleotides coordinate two catalytic Mg²⁺ ions, commonly called M1 and M2, required to align and activate all reactants (**Figure 4.1D-F**). Adjacent to the M1-M2 site, the so-called K1 site, which is occupied either by the side chain of Lys10 of spliceosomal protein Prp11 in the B^{ACT} complex⁹⁵ or by potassium ions in the C complex^{96,291} (**Figure 4.1D-F**), further stabilizes the active site and contributes to splicing progression^{95,288,290}.

Interestingly, a similar active site architecture is found in bacterial and organellar self-splicing ribozymes, i.e., the group II introns, which are evolutionarily related ancestors of the small-nuclear RNA (snRNA) components of the eukaryotic spliceosome's active site and follow the same splicing cycle^{94,101}. During the first step, the 5'-splice site (5'SS) undergoes a transesterification reaction with the 2'-OH group of the intronic branch-point adenosine (BPA) to produce an intron-3'exon intermediate. In the second step of splicing, the 5'-exon attacks the 3'-splice site (3'SS), releasing the spliced exons and the lariat intron⁹⁸. However, the dynamics and functional assembly of spliceosomal intermediates that control the splicing progression remain challenging to understand at the atomic level.

Besides conserved nucleotides and metal cofactors, several protein subunits are located transiently in the second shell of the spliceosome active site, at different stages of the catalytic cycle. Specifically, proteins Prp11 and Cwc24 interact with the 5'SS in the B^{ACT} intermediate (**Figure 4.1A, D**)¹⁰¹. Progression from the B^{ACT} to the B* splicing intermediate shows the disassembly of Prp11 and Cwc24, followed by the recruitment of the protein Yju2, near the active site (**Figure 4.1B, E**)^{94,292}. Then, binding of the protein Cwc25 is observed during the first step of splicing and during the transition from the B* to the C intermediate (**Figure 4.1C, F**)^{94,293}. Additionally, Prp8 is consistently bound to the spliceosome throughout the splicing cycle. While these proteins therefore engage at the spliceosome's core at different stages of the multi-step splicing cycle, it remains unclear how they dynamically act to regulate the required

structural remodeling of the active site architecture during spliceosome activation and catalysis, ensuring reaction efficiency.

Here, we have employed extensive all-atom molecular dynamics simulations (~26 μ s of multiple >2M atoms ribonucleoprotein systems) coupled with free energy calculations and comprehensive structural analyses to elucidate the structure-function relationship of the spliceosome during the first step of splicing. Our findings reveal the crucial role of a minimal set of second-shell residues during the first step of splicing, particularly in modulating the transition from B^{ACT} to B* and C complexes. Mechanistically, we found that these residues – which are highly conserved in evolution and mutated in human diseases – control the occupancy of the K1 site and the timely positioning of catalytic metals together with the 5'SS. These orchestrated dynamics of key active site elements ultimately govern the remodeling of the spliceosome core for catalytic activation, which ensures splicing efficiency and fidelity.

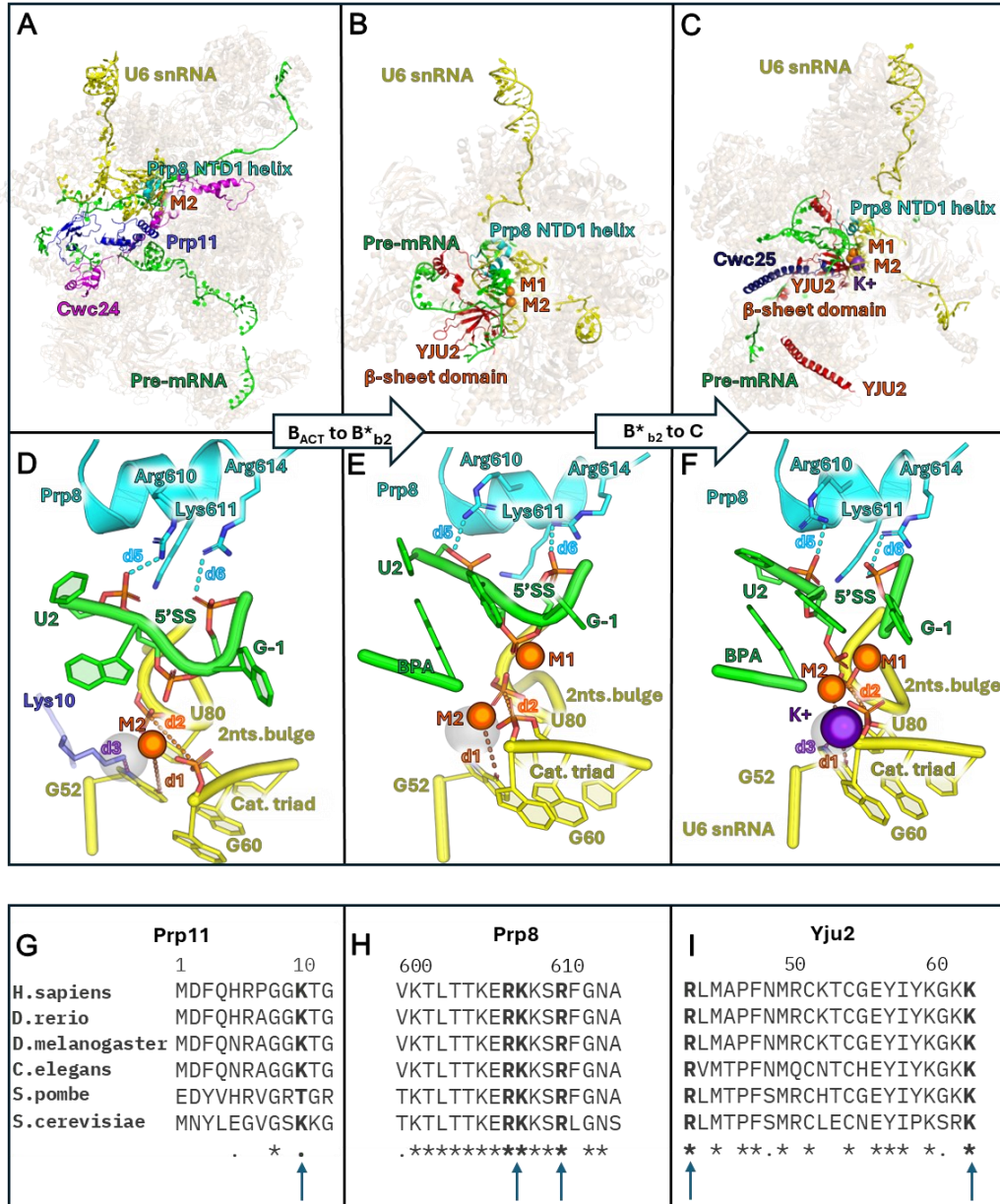


Figure 4. 1. Architecture and evolutionary conservation of key spliceosomal active site elements. A–C) Overview of the spliceosome (cartoon, light brown) highlighting key components: Prp11 (light blue), Prp8 (cyan), Cwc24 (magenta), Yju2 (red), Cwc25 (blue), U6 snRNA (yellow), and pre-mRNA (green). Catalytic magnesium ions (M1, M2, orange) and potassium ions (K⁺, purple) are shown in their binding sites. During the B^{ACT} to B^{*b2} transition, Prp11 and Cwc24 are released while Yju2 is recruited, and the pre-mRNA undergoes conformational changes in preparation for catalysis. Cwc25 joins in the transition to the C intermediate. D–F) Close-up of the active site showing U6 snRNA and the 5'SS (green) forming the catalytic core, anchored by the Prp8 triad (cyan). In B^{ACT} , Prp11's Lys10 (light blue) occupies K1 (purple), and M2 (orange) is in place. In B^{*b2} , the branch point adenosine (BPA) is aligned for attack, M2 shifts toward K1, and the scissile phosphate is uncoordinated. In the C intermediate, M1, M2, and K1 are fully coordinated, with the cleaved 5' exon and lariat intron remaining at the site. G–I) Multiple sequence alignments (MSAs) of selected regions from spliceosomal proteins surrounding the active site. Asterisks (*) denote fully conserved residues, colons (:) indicate positions with strong similarity (e.g., in charge or polarity), and periods (.) mark weaker physico-chemical similarity. Arrows point to conserved residues analyzed in this study. The alignment includes sequences from six species: *H. sapiens* (*Homo sapiens*), *D. rerio* (*Danio rerio*), *D. melanogaster* (*Drosophila melanogaster*), *C. elegans* (*Caenorhabditis elegans*), *S. pombe* (*Schizosaccharomyces pombe*), and *S. cerevisiae* (*Saccharomyces cerevisiae*).

4.3. Results

Positioning of a minimal set of positive residues correlates with the K1 site occupancy

Our first intriguing observation was the varying occupancy of the K1 binding site, whose occupancy changes depending on the catalytic stage and intermediate structure. To examine how such occupancy is likely regulated during catalysis, we examined different first-step-of-splicing intermediates of the *S. cerevisiae* spliceosome, which represents a consistent structural dataset, as it covers the B^{ACT}, B*, and C complexes (**Figure 4.1A-C**). In line with previous studies⁸¹, we specifically analyzed interatomic distances between the catalytic metal M2 and residues that are present at the K1 site in the various structures, namely: *i*) the distance $d1$ between M2 and the O6 atom of G52 and *ii*) the distance $d2$ between the phosphate of G60 and that of U80.

At the B^{ACT} stage (PDB ID 7DCO), the K1 site is occupied by the positively charged side chain of Lys10 from the protein Prp11 (**Figure 4.1D**). Despite the B^{ACT} stage being catalytically inactive, the catalytic M2 ion is found in its proper position, located at 7.3 (± 0.4 Å averaged over the PDB structures of B^{ACT}, 7DCO and 5GM6) from the O6 atom of G52 (**Figure 4.2A**). In all B^{ACT} structures, the U6 snRNA motif folds into the characteristic triplex helix architecture ($d2 = 6.4 \pm 0.3$ Å, **Figure 4.2A**). These findings suggest that Prp11 has adjusted its position to ensure that the Lys10 side chain could favorably locate at the K1 site upon its recruitment at the spliceosome core.

At the B* stage (PDB ID 6J6Q), both Cwc24 and Prp11 are absent⁹⁴. As a result, Lys10 is no longer located at the K1 site, which is empty (**Figure 4.1B, E**). Instead, the protein Yju2 is recruited. We noted that a β -sheet domain of Yju2 (His38 to Arg114) – together with nucleotides A69 and A70 of the pre-mRNA, and C58 of the U6 snRNA – shapes a channel connecting the K1 site to the bulk. Such a channel seems likely gated by Yju2 invariant residues Arg42 and Lys63 (**Figure S1**). Additionally, in B*, M2 is placed out of its catalytic position, partially occupying the K1 site ($d1 = 4.8 \pm 0.2$ Å, **Figure 4.2A**), while the triple-helix active site is tightened ($d2 = 4.9 \pm 0.2$ Å, **Figure 4.2A**). Such a positioning of M2 is found in all B* substates, namely B*a1, B*a2, B*b1, and B*b2 (PDB ID 6J6H, 6J6G, 6J6N, 6J6Q, **Figure 4.2A**)⁹⁴.

At C stage (PDB ID 7B9V), the first step of splicing is facilitated by the binding of Cwc25, which coincides with a positional shift of the Yju2 β -sheet domain¹⁰¹. Importantly, the K1 site

is now filled with one K⁺ ion, while both metals M1-M2 are properly placed in their catalytic position ($d1 = 6.9 \text{ \AA}$ and $d2 = 4.8 \text{ \AA}$, **Figure 4.2A**). Altogether, these unprecedented structural observations suggest that the spliceosome employs a protein-mediated dynamic mechanism to control the occupancy of the K1 site, with either Lys10 of Prp11 or a transient potassium ion during the first step of splicing. Notably, the occupancy of the K1 site is correlated with the placement of M2 in its catalytic position.

Then, we noted that the 5'SS (**Figure 4.1D-F**) – one of the two substrates of the first step of splicing – makes varying contacts with a triad of evolutionarily conserved residues of Prp8, namely Arg610, Lys611, and Arg614 (hereafter referred to as “Prp8 triad”). In the B^{ACT} complex, the Prp8 triad is positioned close to the 5'SS, behind the phosphates of the nucleotides G-1, G1, and U2. At this stage, the 5'SS adopts a kinked conformation, characterized by a 5'SS α angle of $55.5 \pm 2.3^\circ$ (averaged across PDB ID 7DCO and 5GM6, **Figure 4.2B**), defined as the angle formed by the phosphate atoms of G-1, G1, and U2. Arg610 and Arg614 interact with pre-mRNA nucleotides U2 (downstream of the splice junction) and G-1 (upstream of the splice junction), respectively (**Figure 4.2C**), while Lys611 occupies the conserved position K2, as previously described²⁹⁴ (**Figure 4.1D**). Moreover, Cwc24 residues Tyr155 and Phe161, and Prp11 residue Tyr3, interact with the 5' SS (**Figure S2**), ensuring pre-mRNA recognition and positioning. This suggests that Prp8, together with Cwc24, may support the formation of a 5'SS kinked conformation.

In the B* complex, following the Prp2-mediated remodeling and Cwc24 disassembly,^{96,101} the spliceosome progresses to the B*a1 complex (PDB ID 6J6H). Here, the Prp8 triad no longer interacts with the 5'SS, which is captured in a flattened, non-catalytic conformation (5'SS $\alpha = 161.1^\circ$, PDB ID = 6J6H, **Figure 4.2B**). The structural analysis of subsequent B* intermediates shows that the kinked conformation of the 5'SS and the Prp8-5'SS interactions are progressively restored. Indeed, Arg614 and Arg610 alternatively contact the 5'SS from B*a2 to B*b1 complex (**Figure 4.2B**). Only at the B*b2 stage, the Prp8 triad re-engages with the 5'SS, which is found in a kinked conformation (5'SS $\alpha = 52.7^\circ$, PDB ID 6J6Q, **Figure 4.2B**). Thus, the interaction of Prp8 triad residues Arg610, Lys611, and Arg614 with the 5'SS seems associated with the structural arrangement of the 5'SS in a kinked, catalytically productive conformation, even in the absence of Cwc24.

Overall, these observations on the structural evolution of the K1 binding site and 5' splice motif during the first step of splicing suggest a previously uncharacterized functional role of a specific set of second-shell residues (Lys10 of Prp11; Arg42 and Lys63 of Yju2; and Arg610, Lys611, Arg614 of Prp8) distributed across different spliceosomal proteins. The occupancy of

the K1 site and the exact position of M2 seem codependent during the first step of splicing. Furthermore, residues in the Prp8 triad transiently interact with the 5'SS at various splicing intermediates, supporting their dynamic functional role in stabilizing the first-step reactants.

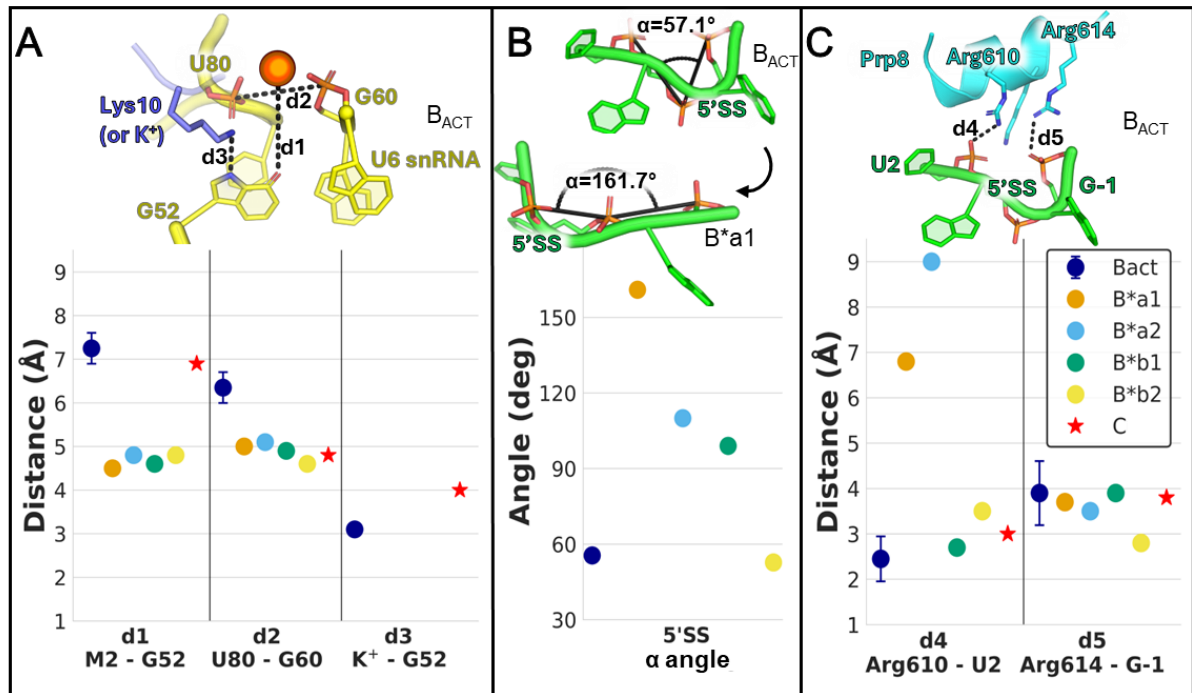


Figure 4.2. Geometry of the active site of the *S. cerevisiae* spliceosome from the B^{ACT} to the C complexes. **A)** Key distances at the spliceosome core: d1 and d3 represent the positions of catalytic metals M1 and K⁺, respectively; d2 indicates the distance between the 2-nt bulge and the U6 snRNA catalytic triad. **B)** The 5'SS angle is defined by the phosphate groups of G-1, G1, and U2 of the pre-mRNA. This angle is flattened in the B^{*1a} complex and progressively restored to a catalytically competent conformation by B^{*b2}. **C)** Distance between Prp8 residues Arg610/Arg614 and the 5'SS nucleotides (NH1/NH2 of Arg to OP1/OP2 of 5'SS). Productive interactions are fully formed in the B^{ACT} and B^{*b2} complexes, lost during the B^{*a1} to B^{*b2} transition, and gradually re-established as the reactive kinked 5'SS conformation reforms.

The energetically favorable conformational switch of Lys of Prp11 controls the positioning of M2 at the B^{ACT} stage

Having identified a minimal set of conserved and medically relevant^{295,296} second-shell residues potentially controlling the remodeling of the spliceosome core during the first step of splicing, we performed multiple molecular dynamics (MD) simulations of each intermediate state of the first splicing reaction to investigate their functional dynamics (B^{ACT}, 12 replicas; B^{*}, 8 replicas; C, 6 replicas; totaling 26 μ s).

First, we observed that Lys10 of Prp11 consistently occupied the K1 site throughout two independent trajectories of the B^{ACT} *wild type* systems (2 replicas, 1 μ s each). Lys10 was stably located at 3.1 ± 0.3 Å from the N7 atom of G52. Meanwhile, M2 remained in its catalytically competent site ($d1 = 6.9 \pm 0.2$ Å, **Figure S3A-C**).

To further characterize the interconnection between the occupancies of K1 and M2 binding sites, we performed enhanced sampling free energy simulations via well-tempered

metadynamics using as collective variables the distance $d1$, and the distance $d3$ (defined between the NZ atom of residue Lys10 in Prp11 and the N7 atom of G52 in U6 snRNA, **Figure 4.2A**). In this way, we could identify three metastable states (state1, 2 and 3, **Figure 4.3A**). At the free energy minimum (state 1), M2 was positioned in its designated site, while Lys10 was stabilized at the K1 site, consistent with the cryo-EM B^{ACT} structure (**Figure 4.3B**, PDB ID: 7DCO). Another free-energy minimum (state 2, **Figure 4.3A**) was large as M2 oscillated between its designated position and a location near the K1 site. When M2 was in its original site, a transient K⁺ ion sat at K1, thus compensating for the absence of the positive charge of Lys10 (**Figure 4.3C**). In the additional metastable minimum (state 3, **Figure 4.3A, D**), M2 was stabilized directly at the K1 site, preventing the access of both Lys10 or one K⁺ ion (**Figure 4.3D**). However, states 2 and 3 were higher in energy compared to state 1 ($\Delta G_{1-2} = 12$ kcal/mol and $\Delta G_{1-3} = 22$ kcal/mol, respectively), with energy barriers of $\Delta G^*_{1-2} = 16$ kcal/mol and $\Delta G^*_{1-3} = 28$ kcal/mol, respectively. Taken together, these free energy data suggest that the transitory occupancy of K1 with a positively charged residue directly affects the precise accommodation of the catalytic M2 cation in the B^{ACT} complex.

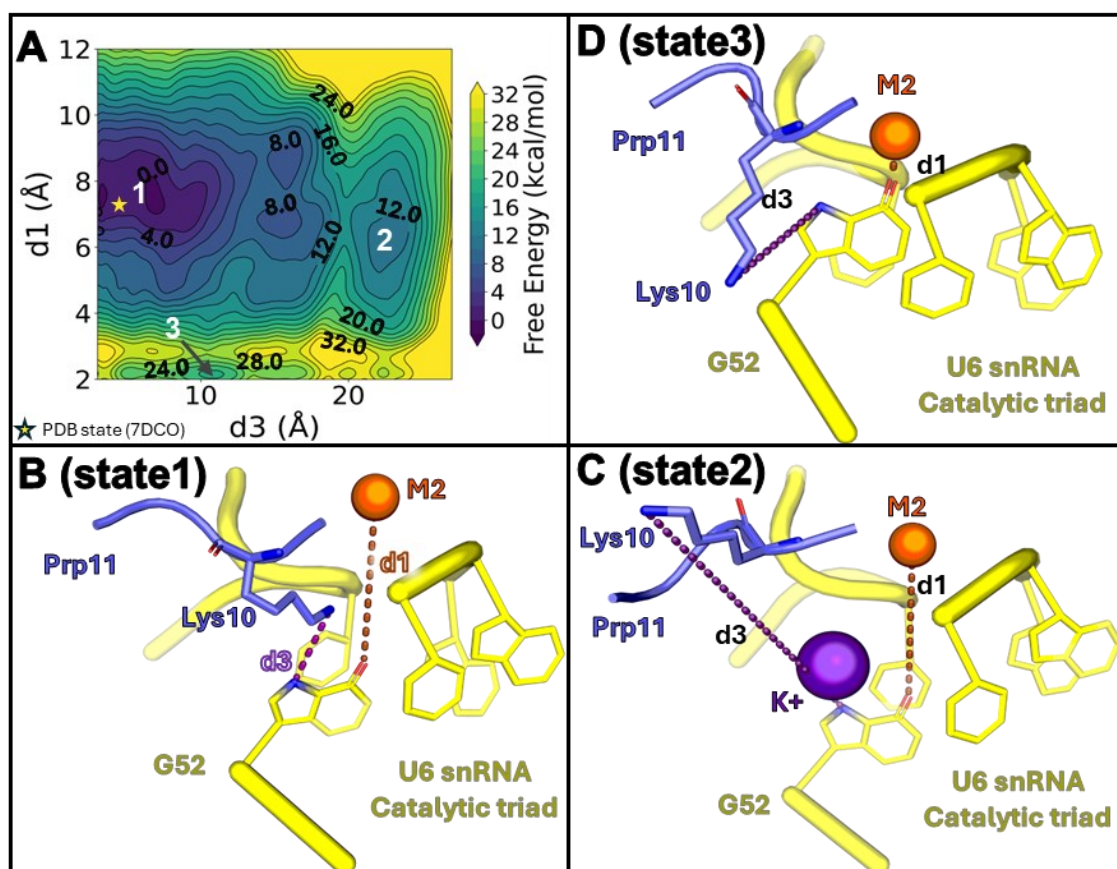


Figure 4.3. Energetics associated with the occupancy of the K1 site and the position of the M2 ion. **A)** Free energy surface of the concerted motions of Prp11 Lys10 and M2, with the initial PDB conformation marked by a yellow star. **B-D)** Representative structures of three metastable states from metadynamics simulations. State 1 (**B**) represents a PDB-like arrangement where Lys10 (slate blue sticks) remains at the K1 site, and M2 (yellow sphere)

is properly positioned at the U6 snRNA triple-helix (yellow cartoon). State 2 (C): Lys10 leaves the K1 site, allowing a K^+ ion to occupy it and stabilize M2 in place. State 3 (D): M2 directly coordinates G52, losing its catalytic position and partially occupying the K1 site

We then simulated the single-point mutation Lys10Ala in the same B^{ACT} model (2 replicas of 1 μ s each), thereby leaving the K1 site empty in all replicas. We designed this mutant to mimic a physiopathological condition, as Lys10Ala mutation is linked to reduced human cell fitness²⁹⁶ and alternative splicing modulation²⁹⁵. In this mutant, the absence of the N-butylamine side chain of Lys10 led to the repositioning of M2, which shifted out of its catalytic position during the first 25 ns ($d1 = 6.4 \pm 0.4$ Å, **Figure S3D-F**). Nonetheless, M2 transiently returned to its catalytic site (~at 25 μ s), due to the spontaneous recruitment of one K^+ ion from the bulk to the K1 site, similarly to what was observed during well-tempered metadynamics simulations (**Figure S3E**). These results support that the K1 site is designed to accommodate a positive charge, thus favoring the correct placement of M2 in its catalytic position at the B^{ACT} stage.

Arg610 and Arg614 of Prp8 cooperate to align and fold the 5'SS in the B^{ACT} complex

Established that the position of M2 is controlled by Lys10 of Prp11 near the 5'SS, we then further looked at the dynamics of this site in our MD simulations of the B^{ACT} complex.

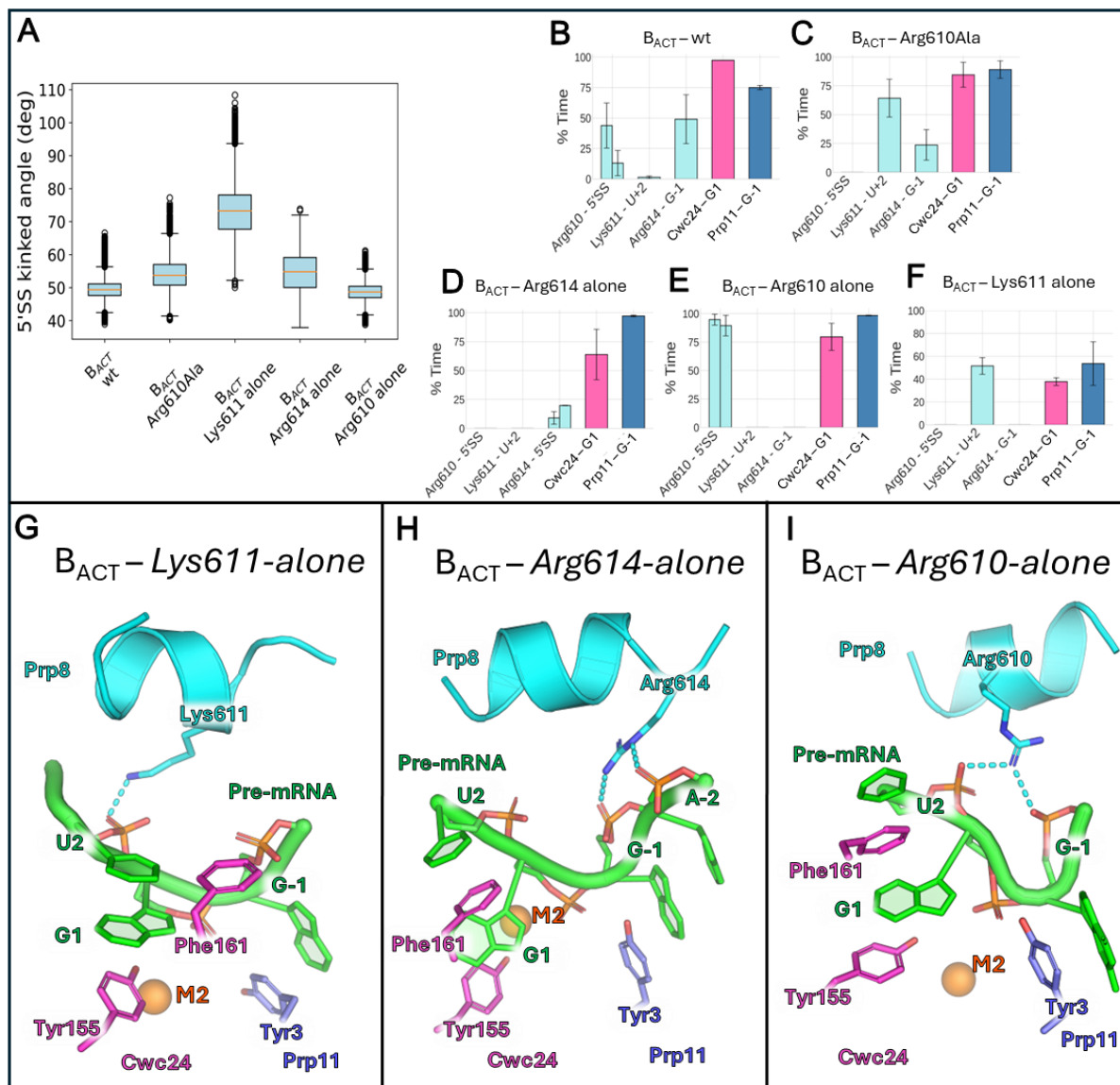


Figure 4.4. Cooperation of Prp11, Cwc24, and Prp8 for the stabilization of the 5'SS. **A)** Boxplot showing 5'SS angle values across simulated B^{ACT} models. **B-E)** Bar plots of interactions between Prp8 residues Arg610, Lys611, Arg614 and 5'SS nucleotides (light blue), π -stacking by Tyr155@Cwc24 with G1 and Phe161@Cwc24 with U2 (magenta), and Tyr3@Prp11 with G-1 (blue). Two bars are shown for Arg610 due to its bifurcated binding mode. Data and errors are from independent replicas. Productive 5'SS stacking requires at least one interaction from Arg610 or Arg614, supporting cooperative stabilization by Cwc24 and Prp11. The Lys611-only mutant shows the weakest stacking and a highly flexible, unproductive 5'SS angle. **G-I)** Active site rearrangements in Prp8 (cyan) mutants. **G)** In the Lys611-only mutant, 5'SS (green) fails to stack with Cwc24 (Tyr155, Phe161, magenta) and Prp11 (Tyr3, slate blue). **H, I)** In Arg614-only and Arg610-only mutants, 5'SS stacking is preserved through retained stabilizing interactions.

During the simulations of the *wild type* B^{ACT} (2 replicas of 1 μ s each), the 5'SS angle (5'SS α) was stabilized at $49.4 \pm 2.7^\circ$, in agreement with cryo-EM data (5'SS $\alpha = 57.1^\circ$, PDB ID = 7DCO, **Figure 4.4A, S4, S5D**). Indeed, the π -stacking interactions of Phe161, Tyr155 of Cwc24, and Tyr3 of Prp11, with U+1, G+1, and G-1 nucleotides of the pre-mRNA were preserved, together with the contacts between Arg610 and Arg614 of Prp8 and G-1 and U+2 phosphates (**Figure 4.4B**). In these simulations, however, we noted that Arg610 dynamically

rearranged to establish a new bifurcated interaction, bridging both ends of the 5'SS, G-1, and U+2 (**Figure 4.4B**). To further characterize the contribution of this residue in the stabilization of the 5'SS, we inserted the Arg610Ala mutation. We performed additional simulations (2 replica of 1 μ s each), observing that the interactions between the 5'SS and both Cwc24 and Prp11 were preserved, thus even in the absence of Arg610. The 5'SS angle was still kinked as in the cryo-EM structure ($54.0 \pm 4.5^\circ$ in the simulations *vs* 57.1° in PDB ID 7DCO, **Figure 4.4C**). This was due to the contribution of Prp8 residues Lys611 and Arg614, which partially compensated for the Arg610Ala mutation by dynamically readjusting their position to interact with U+2 and G-1 (**Figure S6**). This suggests that Arg610, Lys611 and Arg614 act independently to efficiently stabilize the fold of the 5'SS, avoiding dysregulated splicing functions induced by the 5'SS destabilization. Thus, to further determine the contribution of each of the conserved Prp8 residues Arg610, Lys611 and Arg614, we ran MD simulations on three double mutant models (6 μ s in total), namely: *i*) *611-alone* (i.e., Lys611 is the only residue that is not mutated) and in analogy; *ii*) *610-alone*; and *iii*) *614-alone*.

Models of the double mutant *614-alone* and *610-alone* showed that these residues could stabilize the 5'SS kinked conformation with similar magnitude. In model *614-alone*, Arg614 compensated for the mutations and stabilized the 5'SS angle at $54.6 \pm 5.4^\circ$ (**Figure 4.4D, S5A**). Remarkably, also simulations of model *610-alone* showed that Arg610 is enough to stabilize the 5'SS in a B*-like kinked conformation ($5'SS \alpha = 48.6 \pm 2.6^\circ$ in the simulations *vs* 52.7° in PDB ID = 6J6Q, **Figure 4.4E, S5B**), further underscoring the importance of this residue for the first step of splicing. Notably, in both these models, the interactions between the 5'SS and surrounding splicing factors Cwc24 and Prp11 were preserved (**Figure 4.4E-F**). On the other hand, the simultaneous mutation of Arg610 and Arg614 (model *611-alone*) led to a great increase in the 5'SS flexibility (RMSD= 14.3 ± 4.9 Å), as Lys611 alone is not sufficient to stabilize the 5'SS. Importantly, we also noted that due to the missing stabilizing effect of Arg610 and Arg614 on the 5'SS, the interactions of the 5'SS with the proteins Cwc24 and Prp11 were also weakened, resulting in the 5'SS flattening ($5'SS \alpha = 73.2 \pm 7.9^\circ$, **Figure 4.4F**). A partially kinked conformation of the 5'SS was temporarily restored only when the 5'SS interactions with Cwc24 and Prp11 were simultaneously and transitorily re-established (**Figure S5C**).

These simulations suggest that Arg610 and Arg614 of Prp8 favor folding the 5'SS motif into a catalytically competent, kinked conformation. Also, interatomic contacts between these residues and the 5'SS motif reinforced the interaction network of 5'SS with the first-step splicing ribosomal proteins Cwc24 and Prp11.

Yju2 and M2 dynamically control the accessibility of the K1 site in the B* complex

Having captured the dynamics of the structural determinants that control the active site architecture of the B^{ACT} complex, we then moved on to monitor the same elements in the B* complex. Our structural analysis of B* intermediates showed that the catalytic triad and the two-nucleotide bulge are found at 4.9 ± 0.2 Å, and M2 is placed out of its catalytic position, being closer to the K1 site (**Figure 4.1E, 4.2A**), which is surrounded by the β -sheet domain of Yju2 (**Figure 4.1B**). Cryo-EM experiments have captured several B* substates, namely B*a1, B*a2, B*b1, and B*b2. To perform the MD simulations, we used the coordinates of the B*b2 intermediate, as it captured the spliceosome just before the branching reaction.⁹⁴ Initially performed MD simulations of the B*b2 *wild type* intermediate (2 replicas of 1 μ s each). Here, M2 steadily located near the K1 site ($d1 = 5.5 \pm 0.2$ Å) and never occupied its catalytic position – in agreement with cryo-EM data. Nonetheless, despite the K1 site being empty, our simulations showed that K⁺ ions (i.e., K⁺_a and K⁺_b) remained in the surroundings of this position at 3.8 ± 1.0 Å from N7 of G52 (**Figure 4.5A, D**). Further analyses of these simulations showed that K⁺ ions get recruited at the spliceosome core through a solvation channel directly connecting the bulk with the K1 site. This solvation channel is formed by the negatively charged backbone of pre-mRNA nucleotides A69 and A70, U6-snRNA C58, and Yju2 conserved residues Lys63 and Arg42 (**Figure 4.1I**), which act as gating residues (**Figure S1**). Indeed, at the beginning of the simulations (~ 20 μ s), Lys63 was stabilized in an “open” position, sitting at 5.8 ± 0.8 Å from the OP1 atom of A70 (**Figure 4.5A, D**). In this conformation, one potassium ion (K⁺_a) could spontaneously enter the gate, inducing the structural rearrangement of Lys63. K⁺_a could diffuse towards the K1 site, at 3.8 ± 1.0 Å from N7 of G52 (**Figure 4.5B, D**). After about 300 ns, one additional K⁺ ion (K⁺_b) entered the channel, sitting at the core of the spliceosome together with the previously recruited K⁺_a ion, although in a slightly different position. This second K⁺_b ion acted similarly to Lys63, further stabilizing the first K⁺_a ion close to the K1 site. At this point, Lys63 rearranged in a “closed” conformation ($d_{\text{NZ-OP1@A70}} = 4.5 \pm 1.3$ Å, **Figure 4.5C-D**), thus closing the gate and blocking the two K⁺ cations near the K1 site. Together, these simulations show that Lys63 controlled the gating of the spliceosome active site at the B* stage, ensuring that multiple K⁺ ions were recruited and stabilized in the surroundings of the K1 site (**Figure 4.5 and Figure S7**).

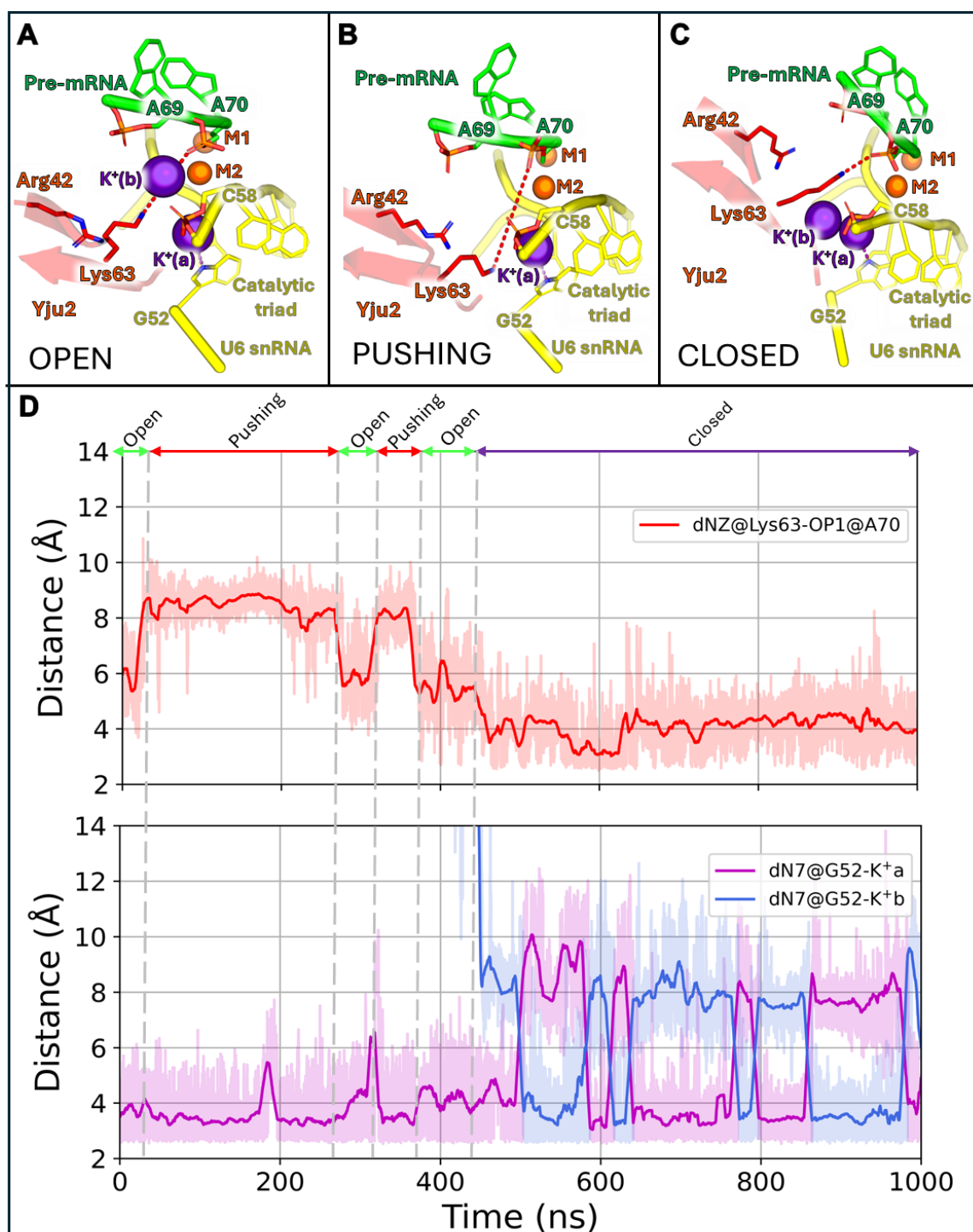


Figure 4.5. **Yju2 gating dynamics regulate recruitment of potassium ions at the K1 site in the B* complex.** **A-C)** Structural view of the cation channel formed by Arg42 and Lys63 (red sticks) of Yju2 (red cartoon), A69 and A70 of the pre-mRNA (green), and C55 of U6 snRNA (yellow). **A)** Open state: Lys63-A70 distance (red dashed line) permits free access of K^+ ions (purple sphere) to the active site. **B)** Pushing state: Lys63 drives a K^+ ion (K^+_a) toward G52 of U6, near the K1 site. M2 (orange sphere) partially occupies K1, repelling K^+_a . **C)** Closed state: An additional K^+ (K^+_b) is recruited, displacing K^+_a toward K1. Lys63 moves closer to A70, closing the gate. **D)** Distance plots from molecular dynamics simulations show transitions between gate states, tracking Lys63-A70 (red), G52- K^+_a (purple), and G52- K^+_b (blue).

Surprisingly, despite two potassium ions sitting at the core of the spliceosome during these simulations, these ions were never steadily bound at the K1 site, which was partially filled by

the catalytic metal M2 (**Figure 4.5A-C**). Indeed, we observed that these potassium ions remain confined at the spliceosome's core while not being able to displace M2 from its position close to the K1 site, at odds with simulations of the B^{ACT} complex (**Figure 4.3**). To check this evidence, we performed two simulations (1 μ s each) starting from the B*b2 complex (PDB ID 6J6Q) in which we modeled M2 in its catalytic site. As expected, in these simulations, M2 was stable in its catalytic position ($d1 = 6.9 \pm 0.3$ Å, **Figure S5**), leaving the K1 site vacant, confirming that the triple helix in B* can hinder M2 mobility through this structure ($d2 = 4.6 \pm 0.1$ Å). Notably, Lys63 also acts as a gatekeeper of the spliceosome core during these simulations. Upon the recruitment of one K⁺ ion and the closure of the channel gate by Lys63, one K⁺ ion spontaneously bound and was stabilized at the K1 site (3.0 ± 0.3 Å from N7@G52, **Figure S8**). Overall, these simulations suggest that the conformation of the triple helix core at the B* complex (PDB ID 6J6Q) ensures the stabilization of M2 close to the K1 site, preventing the recruitment of K⁺ ions at the same site, thus avoiding the premature catalytic activation of the spliceosome. K⁺ access K1 site only upon the recruitment of Cwc25, in the C complex (PDB ID 7B9V), which triggers local rearrangements of the active site.⁹⁶ Indeed, in additional simulations of the C complex (2 μ s), one K⁺ ion remained steadily bound at its original K1 site (i.e., the K1 site, 2.9 ± 0.3 Å from N7@G52) stabilized by the conformation of the active site triple helix ($d2 = 4.4 \pm 0.1$ Å). In the same simulations, both metals M1 and M2 also remained in their positions ($dM2-O6@G52 = 6.9$ Å; $dM1-O6@G62 = 8.8$ Å, **Figure S9**), confirming the functional role of the triple helix conformation in facilitating the timely and accurate positioning of catalytic components.

Taken together, our simulations showed that the controlled dynamics of Lys63 of spliceosomal protein Yju2 gates the core of the spliceosome for the recruitment of potassium ions near the K1 site, which is temporarily filled by the catalytic M2 to prevent premature catalytic activation during the B* stages.

RNA-RNA contacts for reactant alignment and product stabilization support the B* to the C transition

We further extended our trajectories to characterize the dynamic role of Arg610, Lys611, and Arg614 of Prp8 in the transition from B* to C complexes.

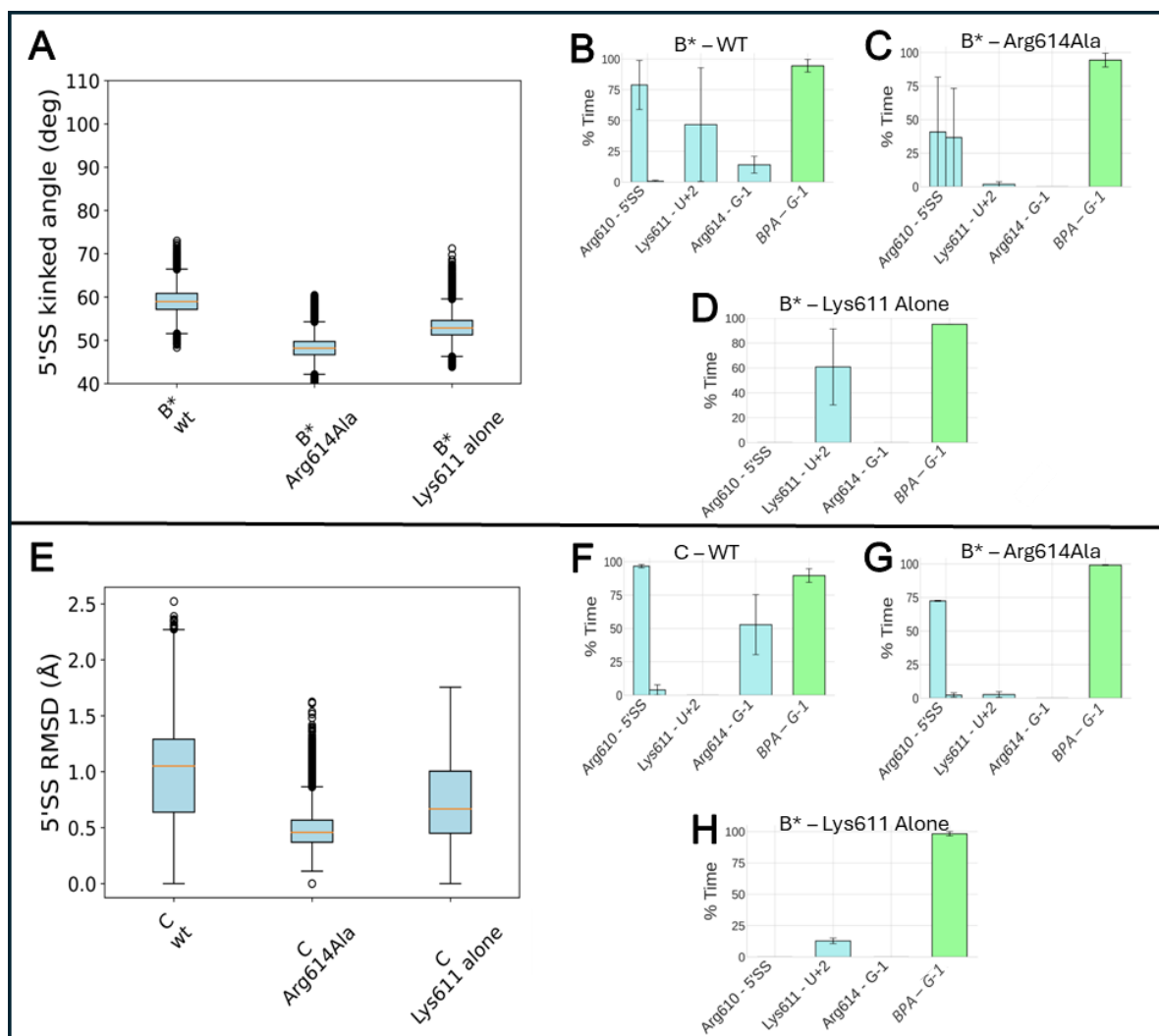


Figure 4.6. Interaction between the branching point adenosine (BPA) and the 5'SS in the B* and C complex. **A)** Boxplot showing the 5'SS angle across simulated B* models. **B-D)** Bar plots of interactions between Prp8 triad residues and 5'SS nucleotides (light blue), and stability of BPA-G1 π -stacking (light green). Data and errors are from independent replicas. The 5'SS kinked angle remains stable across mutations due to consistent stacking with BPA. **E)** Boxplot of 5'SS stability in C models, monitored by RMSD of G-1 and G+1 nucleotides (the 5'SS angle is absent post-cleavage of the G-1 and G+1 phosphodiester bond). **F-I)** Bar plots showing Prp8-5'SS interactions (light blue) and BPA-G1 stacking stability (light green). As in B*, the 5'SS remains stable in C models due to preserved BPA-5'SS interaction network.

During our simulations of the B*b2 structure (PDB ID 6J6Q) – which captured the spliceosome just before the first step of splicing – these 3 residues interacted with the 5'SS, stabilizing it in a kinked conformation (5'SS $\alpha = 52.7^\circ$). This angle was also steadily maintained throughout our multiple trajectories (2 μ s, 5'SS $\alpha = 59.2 \pm 2.8^\circ$, **Figures 4.6A and S10**) together with the contacts between the Prp8 triad and 5'SS (**Figure 4.6B**). Thus, we modeled the same mutants that showed the most significant destabilization of the 5'SS dynamics in the B_{ACT}, namely Arg610Ala, and the double mutant model *611-alone*. Remarkably, in all these models, such mutations did not alter the structural properties of the 5'SS, which was stabilized in its original kinked conformation (Arg610Ala 5'SS $\alpha = 48.2 \pm$

2.4°, and *611-only* 5'SS $\alpha = 53.2 \pm 2.8^\circ$, **Figure 4.6A, C-D**), suggesting that the role of the Prp8 triad in stabilizing the fold of the 5'SS is less critical at B* late stages as compared to the B^{ACT} stage. Our simulations suggest that this structural resilience may arise from the branch-point adenosine (BPA) and its ordered RNA-RNA interaction network with the 5'SS, made by π stacking interactions with G1 and Watson-Crick interactions with U2. These contacts were preserved in all simulations (6 μ s, **Figure 4.6B-D**), ultimately guaranteeing the proper positioning of the reactants just before the first step of splicing.

In the C complex, on the contrary, the 5'SS is no longer found in the so-called kinked conformation, as it is cleaved to release the first step splicing products, i.e., the 5'-exon and the 3'-exon-intron intermediate. Indeed, our simulations of the *wild type* C complex showed that this system is stable, as Arg610 and Arg614 of Prp8 steadily interacted with the just-cleaved 5'exon and the 5'-end of the intron (**Figure 4.6E-F**). Following the same approach used for the B* complex, we modelled the Arg610Ala and *611-alone* mutant models in the C complex. In both cases, the mutations did not alter the conformation of the cleaved products, as evidenced by RMSD values consistently below 2.5 Å and comparable to those of the *wild type* throughout the trajectories (6 μ s, **Figure 4.6E, S11**). As observed for the B* complex, the increased stability of the 5'SS in the C complex can be attributed to the presence of BPA. Once positioned, the BPA steadily remained in its original position, interacting with the 5'SS (**Figure 4.6F-H**), whose stabilization is independent of the contacts with Prp8 residues. In conclusion, these simulations of multiple splicing systems suggest that RNA-RNA contacts between the BPA and the 5'SS in the B* and C intermediates are sufficient to stabilize the 5'SS in productive conformations.

4.4. Discussion

RNA splicing is crucial for eukaryotic cells and requires the precise and timely assembly of functional spliceosome complexes to orchestrate mRNA maturation. Through comprehensive structural analyses, extensive equilibrium molecular simulations of a set of multimillion-atom systems (totaling 26 μ s) and free energy calculations of *wild type* and medically-relevant mutant spliceosome complexes, we characterized the structural remodeling required for the progression through the major intermediates of the first step of splicing, i.e., B^{ACT}, B*, and C (**Figure 4.7**), leveraging recent high-resolution cryo-EM structures^{94-96,101}.

At the B^{ACT} stage, we observed that Lys10 of Prp11 - a residue conserved across Prp11 orthologues (**Figure 4.1G**) and associated with reduced human cell fitness²⁹⁷ and alternative

splicing modulation²⁹⁸ - controls the positioning of the catalytic metal M2. Our free energy calculations show that when Lys10 disengages from the K1 site, M2 spontaneously moves out of its position, destabilizing the active site (**Figure 4.3**). MD simulations of the Lys10Ala mutant confirmed this mechanistic finding, explaining the splicing defects associated with this mutation (**Figure S3**). Thus, the transient occupancy of the K1 site with positively-charged residues, either by Lys10 or trafficking K⁺ ions, promotes early active site folding. Intriguingly, this mechanistic strategy parallels catalytic mechanisms observed in nucleic acid-cleaving enzymes, where catalytic positions similar to the K1 site are occupied by either positively charged residues or cations²⁹⁴. For instance, in BamHI (PDB ID: 2BAM), Lys161 fulfills a site analogous to the spliceosomal K1 site^{116,299}, resembling Prp11's Lys10. Similarly, Lys131 of exonuclease λ (modeled from PDB IDs 4WUZ and 3SM4) occupies a comparable position^{300,301}, as does Lys231 of DNA polymerase η (Pol- η , PDB ID: 4ECS)³⁰² and Lys974 of the RuvC domain of SpyCas9.³⁰³ Notably, single-point mutations of these lysine residues abolish or significantly reduce enzymatic activity, suggesting a fundamental role for these "K1 site" residues in catalytic processing of nucleic acids.^{301,304–306} Moreover, recent biophysical and computational evidence has shown that the catalytic cycle of these enzymes relies on the timely binding and removal of positively charged residues and transitory ions at K1 site-like positions.^{294,307,308} For instance, similarly to what was observed at the K1 site of the spliceosome, the exchange between cations and Lys196 is found at the Kw site of RNase H1 during catalysis to favor reactant alignment, activation, and product release.^{307,309} This parallel extends to further functional evidence: just as the RNase H1 is altered by Lys196 mutations and benefits from high K⁺ concentration, spliceosome function is impaired by mutations of Prp11 Lys10 and shows similar K⁺ ion dependence.^{96,310–312} This suggests that a mechanism of catalytic control relying on both dynamic second-shell positively charged residues and transitory bound cations is shared among numerous RNA processing enzymes, likely indicating an evolutionarily convergent catalytic strategy.

Another layer of splicing control in B^{ACT} is provided by a triad of positively charged and evolutionarily conserved residues in Prp8: Arg610, Lys611, and Arg614. We show that this triad ensures catalytically productive folding of the 5'SS through a series of redundant interactions. Indeed, our simulations show that even in the case of a severe two-out-of-three alanine mutation of the Prp8 triad, the remaining residue (Arg610 or Arg614, **Figure 4.4D-E**) is still able to stabilize the 5'SS, suggesting the evolutionarily selected functional redundancy of this triad. That is, multiple residues with similar physicochemical properties are positioned within a narrow structural space to ensure critical contacts with the 5'SS, stabilizing its folding

and ultimately favoring alignment of catalytic elements at the spliceosome core even under mutational stress.

Remarkably, the functional role of the Prp8 triad mirrors that of group II intron maturase proteins,³¹³ which share structural similarity with Prp8 and assist bacterial and organellar intron self-splicing by directly interacting with the splice site. Thus, our findings further expand upon the evolutionary connections between group II intron splicing and the spliceosome, elucidating a specific and extended network of interdependent interactions for splicing. Through the stabilization of the 5'SS, the Prp8 triad enables a broader spliceosomal interaction network involving the 5'SS and two splicing factors crucial for assembly of the spliceosome: Cwc24, and Prp11. Indeed, the simultaneous mutation of Arg610 and Arg614 in the Prp8 triad (*611-alone* model) destabilizes the complex interaction network between Cwc24, Prp11, and the 5'SS, altering the catalytic architecture of the splice site (**Figure 4.4F** and **Figure S5C**). These interdependent interactions highlight that the spliceosome's catalytic core is stabilized and regulated through the allosteric actions of multiple proteins.

Following the proper formation of the catalytic triple helix active site – ensured by Lys10 of Prp11 at K1 and 5'SS alignment by the Prp8 triad, Cwc24, and Prp11 – the spliceosome moves from B^{ACT} to B* complex. Prp11 and Cwc24 dissociate at this stage while Yju2 assembles onto the spliceosome, positioning a beta-sheet domain in front of the K1 site.¹⁰¹ Just before the first catalytic step (B*b2 complex), the K1 site is partially occupied by M2, which is displaced from its catalytic position (**Figure 4.2A**). Our simulations also reveal a dynamic flow of K⁺ ions near the K1 site, recruited *via* a previously uncharacterized channel gated by the invariant residue Lys63 of Yju2 (**Figure S1**). Transiently recruited ions bounce between M2 and Lys63 but do not settle at the K1 site (**Figure 4.5**), potentially explaining their absence in cryo-EM structures.⁹⁴ Nonetheless, the excess of transitory K⁺ ions near K1 just before catalysis suggests a controlled catalytic activation mechanism. Indeed, upon Cwc25 recruitment, Yju2 undergoes a positional shift⁹⁶, enabling M2 to transition into its catalytic position (**Figure 4.2A**), followed by K⁺ ion occupancy at the K1 site, which completes the catalytic transition from B* to C. Again, this mechanism is reminiscent of the group II intron self-splicing ribozyme, where potassium ions transiently occupy the K1 site to regulate catalytic activation¹¹⁹ and first-to-second step transition⁸¹, further reinforcing the evolutionary link between the complex eukaryotic spliceosome machine and the group II introns.

Such a tight regulation of the spliceosome catalytic activation at B*b2 is essential, as both reactants – the 5'SS and the branch point adenosine (BPA) – are positioned very close to each other (**Figure 4.1E**) and may prematurely react, forming non-functional products. While the

Prp8 triad stabilizes the 5'SS in B*b2, as observed in the B^{ACT}, our simulations indicate that its role is less critical at this stage. Even the most disruptive Prp8 triad mutations tested in B^{ACT} do not significantly alter the 5'SS conformation in B*b2 (**Figure 4.6A**), suggesting a structural compensation mechanism. We identify the BPA and its structured interaction network with the 5'SS as the primary stabilizing factor. Once BPA is recruited, stable base-pairing and stacking interactions (**Figure S12**) form between BPA and the 5'SS, providing self-assisted stabilization of the reactants (**Figure 4.6B-D**). Remarkably, a similar π -stacking and pairing interaction network is observed in group II introns (**Figure S12**). For instance, the structure of the *Agathobacter rectalis* intron obtained before the branching (PDB ID: 8T2S) – i.e., corresponding to the B*b2 spliceosomal complex – shows that the branching A632 stacks with the bases G1 and U2, and contacts the ribose of C630, positioned two nucleotides downstream of the intronic BPA A632 (**Figure S12**).³¹⁴ This specific interaction is also observed in the pre-branching structure of the *Thermosynechococcus vestitus* intron (PDB ID: 8FLI), where the sugar moiety of C858 interacts with the BPA A860, which in turn stacks with U1.³¹⁵ Analogous interactions are also captured at the B*b2 spliceosomal stage (PDB ID: 6J6Q), where the BPA A70 and nucleotides G1 and U2 interact with each other¹⁰¹. Notably, current structures of B* do not capture the interaction between the BPA A70 and the nucleotide located at two positions downstream (U68), as observed in group II introns. However, the A70-U68 interaction is observed and consistently maintained at the C complex, even in the case of severe mutation of the Prp8 triad (**Figure 4.6 E-H**), suggesting a role in the stabilization of reactants during catalysis. Altogether, these observations show that specific and dynamically-engaged RNA-RNA interactions occurring at the core of the spliceosome between the 5'SS and the BPA represent a strategy for the stabilization of reaction intermediates and products conserved among splicing machines.

In summary, our work constitutes an atomistic-resolution and dynamic view of the complex structural remodeling of the spliceosome catalytic architecture required for the first step of splicing (**Figure 4.7**). Our findings reveal the crucial role and functional dynamics of a precise set of positively-charged residues that are evolutionarily conserved across species from human to yeast and are associated with diseases if mutated. Mechanistically, we found that these residues' dynamics control the occupancy of the K1 site and the timely positioning of catalytic metals together with the 5'SS, during the transition from B^{ACT} to B* and C complexes. These coordinated dynamics of specific key active site elements, distributed across distinct spliceosomal proteins that only transiently localize near the catalytic core – i.e., Prp11, Prp8, and Yju2 – ultimately governs the spliceosome structural remodeling and ensures splicing

efficiency and fidelity. Notably, our mechanistic findings parallel those of group II introns and other nucleic acid-processing enzymes, suggesting that the regulatory mechanisms we describe may represent fundamental principles of catalysis that have been selected and refined throughout evolution, in both protein and RNA enzymes.

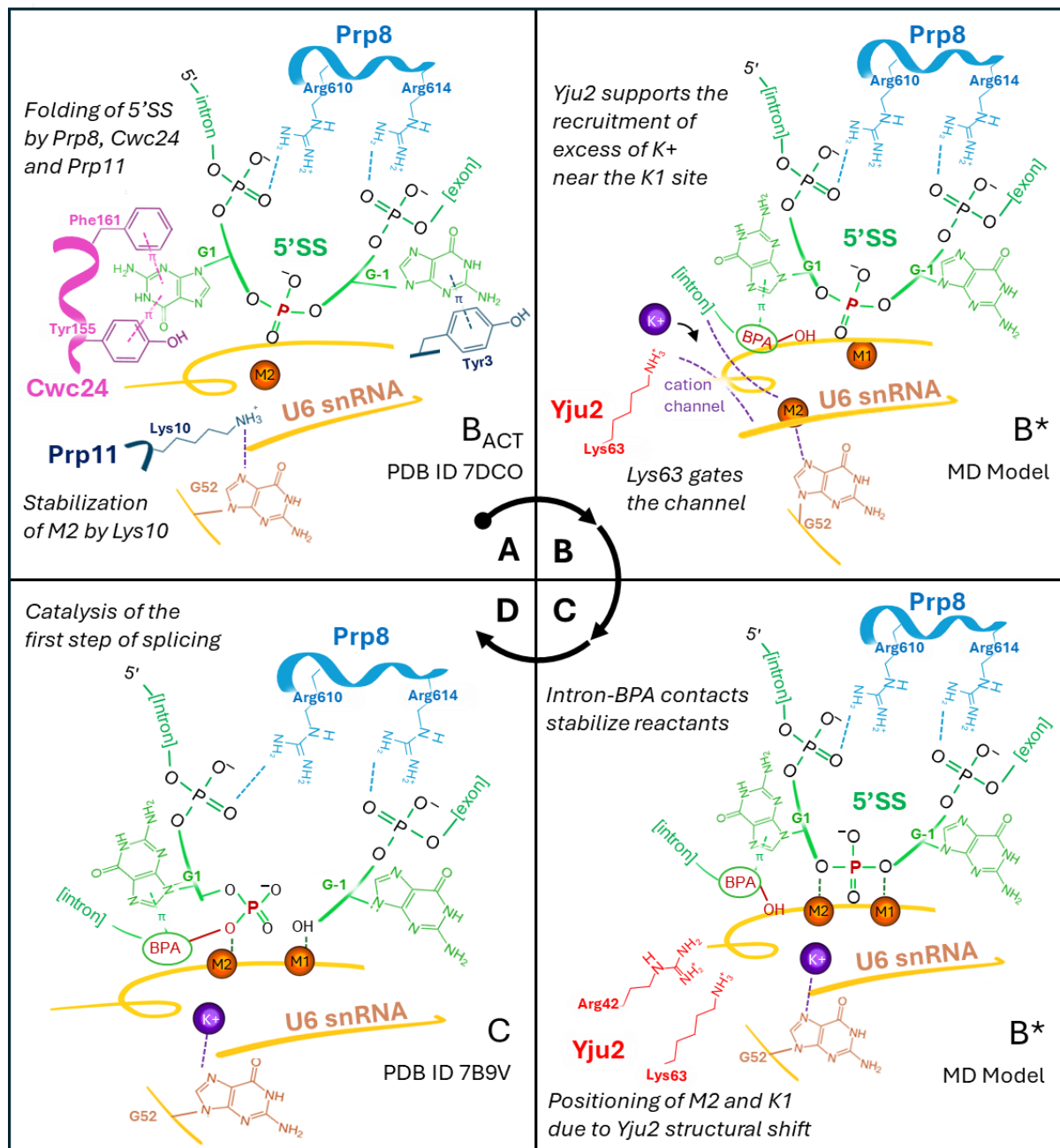


Figure 4.7. Mechanistic model for the dynamic remodeling of the spliceosome active site during the first step of splicing. In the B^{ACT} complex (A), M2 is stabilized at its position by the Prp11's Lys10, which sits at the K1 site. The 5'SS adopts a kinked conformation that positions the scissile phosphate in the active site, supported by π -stacking with Cwc24 (Phe161, Tyr155) and Prp11 (Tyr3). This interaction network is allosterically reinforced by Prp8 (Arg610, Arg614). During remodeling to the B^* complex (B), Cwc24 and Prp11 dissociate, while M1, the branch point adenosine (BPA), and Yju2 are recruited. Though not yet fully positioned, the BPA stabilizes the 5'SS via RNA-RNA π -stacking. M2 shifts toward G52, partially occupying K1, while Yju2 (Arg42, Lys63) helps form an ion channel for K⁺ entry. However, M2 still blocks full access, preventing premature catalysis. In the C complex (C), Cwc25 promotes K⁺ binding at K1, allowing M2 to adopt its catalytic position and directly coordinate the scissile phosphate. The BPA also shifts into place, completing active site assembly. Upon catalysis (D), Yju2 dissociates, and the reaction products are stabilized by Prp8 and the BPA, priming the spliceosome for the second step of splicing.

4.5. Material and Methods

4.5.1. Structural models for MD simulations

The initial coordinates for molecular dynamics (MD) simulations of the B^{ACT},³¹⁷ B*,³¹⁸ and C³¹⁹ spliceosomal complexes (*Saccharomyces cerevisiae*) were derived from cryo-EM structures available in the Protein Data Bank (PDB IDs: 7DCO, 6J6Q, and 7B9V), with average resolutions of 2.50 Å, 3.70 Å, and 2.80 Å, respectively. All missing protein regions were modelled using Crosslink Protein tool³²⁰ by Maestro (v12.9.123 from the Schrodinger release 2021.3, Schrödinger, LLC, New York, NY, 2021), unless they were structurally disconnected from the spliceosomal core to contribute meaningfully to the system's stability – similarly to what has been done by parallel investigation.^{321–323} RNA gaps were manually modeled on Maestro. All structural models were subsequently subjected to energy minimization and MD equilibration with the restraint schemes detailed below.

For the B^{ACT} complex, we included all proteins from the 7DCO PDB entry except Cdc40, Cwc21, Pre-mRNA processing factor 19, Pre-mRNA-splicing factor Snt309, Small nuclear ribonucleoprotein Sm D3, Spp2, Small nuclear ribonucleoprotein E. In the B* complex, all proteins were modelled except Prp19, Snt309, Syf1, SNRNP-E, SNRNP-F, SmD1, SmD2, SmD3, SNRNP-G, Prp17, SNRNP-B, and SNRNP-A. In the C complex, all proteins were modelled except Brr2, SmD1, SmD2, SNRNP-F, SNRNP-E, SNRNP-G, SmD3, Syf1, Ntc20, Cdc40, Prp18, Hlj1, Prp19, Snt309, SNRNP-B, and Prp16.

In total, we prepared 13 systems for MD simulations: (1) *Wild type* B^{ACT} complex modelled on PDB ID: 7DCO; (2) B^{ACT} complex with the of Prp11-Lys10Ala mutation; (3) B^{ACT} complex with Prp8-Arg610Ala mutation; (4) B^{ACT} complex with Prp8-Arg610Ala and Lys611Ala mutations; (5) B^{ACT} complex with Prp8-Arg610Ala and Arg614Ala mutation; (6) B^{ACT} complex with Prp8-Arg611Ala and Arg614Ala mutation; (7) *Wild type* B* complex modelled on PDB ID: 6J6Q; (8) B* complex with Prp8-Arg614Ala mutation; (9) B* complex with Prp8-Arg610Ala and Arg614Ala mutation; (10) *Wild type* B* complex with M2 repositioned from the native to the catalytical position; (11) *Wild type* C complex modelled from PDB ID: 7B9V; (12) C complex with the Prp8-Arg614Ala mutation of Prp8; (13) C complex with Prp8-Arg610Ala Arg614Ala mutation. Alanine mutations were performed using PyMOL. All systems were solvated in a 10 Å layer of TIP3P water³²⁴ and neutralized with K⁺ ions. Final system sizes were as follows: B^{ACT} models enclosed in a box of ~286·227·297 Å³, containing ~621,000 water molecules, resulting in a total of ~2,080,000 atoms; B* models within a box

of $\sim 193 \cdot 202 \cdot 185 \text{ \AA}^3$, containing $\sim 227,000$ water molecules and $\sim 77,500$ atoms in total; C models within a box of $\sim 198 \cdot 179 \cdot 178 \text{ \AA}^3$ filled with $\sim 200,000$ water molecules, resulting in a system of $\sim 700,000$ atoms.

4.5.2. *Simulations set up*

RNA was parameterized using AMBER ff12SB (ff99+bsc0+ χ OL3),³²⁵ proteins with AMBER ff14SB,³²⁶ monovalent ions with Joung–Cheatham parameters,³²⁷ and divalent ions following Allner's parametrization.³²⁸

All systems were subjected to the same simulation protocol. First, an energy minimization was performed using the steepest descent algorithm implemented in GROMACS 2023.3,³²⁹ with harmonic positional restraints of $5,000 \text{ kJ/mol nm}^2$ applied to all heavy atoms and catalytic ions (M1, M2, K^+). Next, each system was smoothly thermalized to 310 K over a 5 ns NVT equilibration using a 1 fs timestep. Temperature coupling was performed using the velocity-rescale thermostat³³⁰ with a relaxation time $\tau = 0.1 \text{ ps}$. The same positional restraints used during minimization were maintained throughout the NVT phase. This was followed by a 5 ns NPT equilibration using a 1 fs timestep. Pressure control was applied using the Parrinello–Rahman barostat³³¹ with a relaxation time $\tau = 2 \text{ ps}$. Positional restraints on heavy atoms and catalytic ions remained unchanged during this step. To allow system relaxation while preserving the conformation of critical features – particularly at the catalytic core – four additional NPT equilibration steps of 5 ns each were performed, with gradually decreasing positional restraints. The system was divided into two regions: (i) the catalytic core, consisting of heavy atoms from the 5'SS, the two-nucleotide bulge, the U6 snRNA catalytic triad, U6-G52, and the catalytic metal ions; and (ii) the bulk, comprising all other heavy atoms in the system. The restraint scheme was as follows: 1) $2,000 \text{ kJ/mol}\cdot\text{nm}^2$ on the core and $1,000 \text{ kJ/mol}\cdot\text{nm}^2$ on the bulk; 2) $1,000 \text{ kJ/mol}\cdot\text{nm}^2$ on the core and $500 \text{ kJ/mol}\cdot\text{nm}^2$ on the bulk; 3) $500 \text{ kJ/mol}\cdot\text{nm}^2$ on the core and $250 \text{ kJ/mol}\cdot\text{nm}^2$ on the bulk; 4) $250 \text{ kJ/mol}\cdot\text{nm}^2$ on the core and unrestrained bulk.

Production simulations were then performed in the NPT ensemble using a 2 fs integration timestep. Bond lengths were constrained with the LINCS algorithm.³³² Periodic boundary conditions were applied in all directions, and long-range electrostatics were treated using the particle mesh Ewald method³³³ with a Fourier grid spacing of $1.6 \text{ \AA}$. A temperature of 310 K was maintained using the velocity-rescale thermostat ($\tau = 1 \text{ ps}$), and pressure was regulated at 1 atm using the Parrinello–Rahman barostat ($\tau = 2 \text{ ps}$).

Over 26 μ s of MD simulations were collected in total, two replicas of ~ 1 μ s for each of the 13 systems described above.

4.5.3. *Well-tempered Metadynamics simulations*

Well-tempered metadynamics³³⁴ simulations were run on the *wild type* B^{ACT} complex, using PLUMED 2.9³³⁵ patched into GROMACS 2023.3.³²⁹ Two collective variables (CVs) were defined: CV1 corresponding to the distance between Lys10 of Prp11 protein and N7 of G52 belonging to U6 snRNA, while CV2 as the distance between O6 of G52 belonging to U6 snRNA and M2.

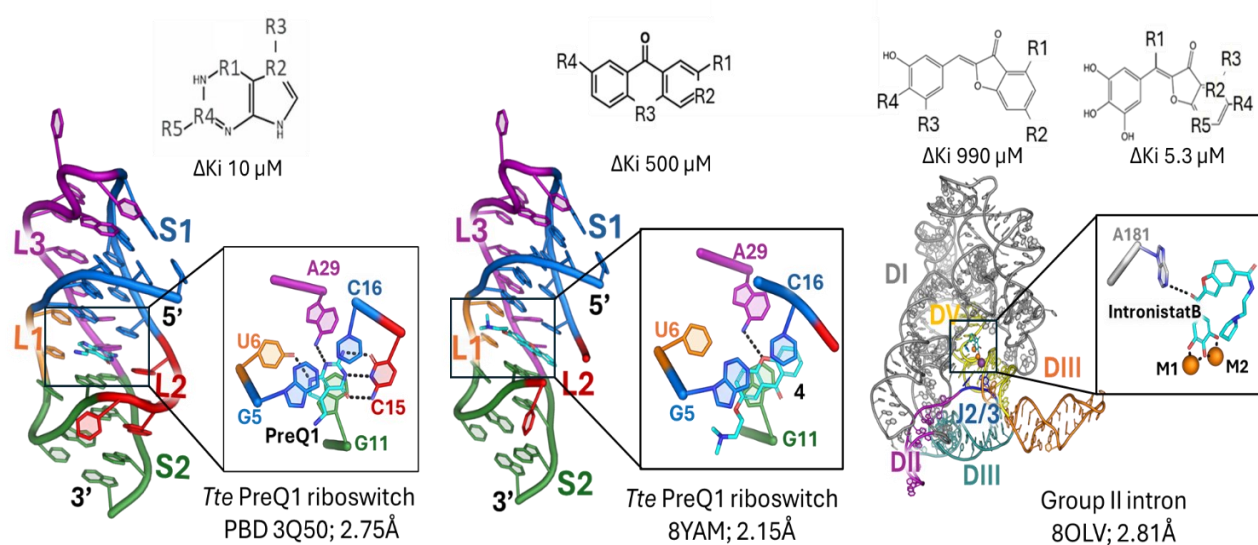
G52 position was restrained with a harmonic potential of 500 kJ/mol·nm² to avoid unphysical distortions and preserve the structural architecture of this part of the active site. Additionally, we imposed a lower wall at 0.35 nm to the CV1 and an upper wall at 0.85 nm to the CV2, both with a constant of 40 kJ/mol·nm². The metadynamics bias was applied on both CVs using a Gaussian potential deposited every 500 integration steps. The initial Gaussian height was 0.5 kJ/mol, and widths (sigma) of 0.025 nm and 0.047 nm were set on the standard deviations of G52 and Lys10 obtained from the unbiased simulation of the same system.

Free energy surfaces (FES) were reconstructed using PLUMED's `sum_hills` utility.

4.5.4. *MD trajectories analysis*

MD trajectories were analyzed using VMD³³⁶. To assess the preservation of native π stacking interactions, we monitored the RMSD of π -stacked bases with a threshold of 1 Å, considering them preserved when RMSD remained below this threshold throughout the trajectory. Hydrogen bonds were defined as present when the donor–acceptor distance was less than 3.5 Å. The 5'SS angle α was defined using the phosphorus atoms of G-1, G1, and U2 in the pre-mRNA sequence.

5. Relative Binding Free Energy Prediction Strategies for RNA-Targeted Drug Discovery



Reference paper: Gianfranco Martino, Jacopo Manigrasso, Giuseppina La Sala, Anders Hogner and Marco De Vivo, *Benefits of Polarizable Force Fields for RNA-Ligand Relative Binding Free Energy Estimations*, In preparation

5.1. Abstract

Relative binding free energy (RBFE) calculations of RNA-small molecule systems remain largely unexplored and challenging despite the intense efforts towards RNA-targeted small molecule therapeutics. In this context, we present a systematic comparison of polarizable OPLS5 and fixed-charge OPLS4 force fields across three representative RNA-small molecules complexes: i) The PreQ1 riboswitch bound to native ligand-like fragments; ii) The partially resolved complex of PreQ1 riboswitch with synthetic ligands; and iii) The group II intron in complex with metal-coordinating inhibitors. We demonstrate that preparatory molecular dynamics simulations are crucial for accurate predictions through RBFE calculations, with equilibration-induced structural relaxation that improves fragment ranking. We also show that polarizable force fields perform better by promoting spontaneous recovery of native-like conformations, particularly in systems with crystal structures that are only partially resolved. Additionally, polarizable force fields consistently deliver superior RBFE accuracy compared to fixed-charge approaches for ligand/metal coordination systems, as charge-changing transformations at metal coordination sites remain difficult. Ultimately, explicit polarizability was essential for structural modeling and computation involving flexible RNA regions and metal-ligand interactions. These findings suggest context-dependent benefits of polarizability, helping determine when this is justified despite its additional cost, and highlighting current challenges of alchemical free energy methods for RNA targets.

5.2. Introduction

Regulatory RNAs fold into complex three-dimensional structures essential for their biological functions, including regulating gene expression, viral replication, and splicing. These structured RNA elements often form small molecule-druggable pockets, making them compelling therapeutic targets across various diseases, such as cancer, neurodegenerative disorders, and viral or bacterial infections^{153,156}.

Multiple drug discovery campaigns illustrate the therapeutic potential of RNA targets. For instance, the FDA-approved small molecule Risdiplam³⁰ is proposed to target a specific bulge structure within the duplex formed by the SMN2 pre-mRNA and the U1 small nuclear RNA. Another class of small molecule splicing modulators acts on self-splicing group II intron ribozymes by directly coordinating the two catalytic Mg^{2+} ions, preventing essential

conformational changes required for splicing⁸⁸. Beyond splicing targets (further details in chapter 1.5), riboswitches represent another promising class of structured regulatory RNAs that recognize specific metabolites or metal ions with high affinity²⁶. Several drug discovery campaigns have targeted these regulatory RNAs^{336,337}, as exemplified by the development of Ribocil³⁰, which inhibits the FMN riboswitch to disrupt bacterial growth.

Building on the promise of RNA as a therapeutic target, computational tools for RNA-targeted drug discovery represent an emerging field, with relative binding free energies (RBFE) calculations that represent a powerful methodology routinely used to target proteins, while still challenging when applied to RNA targets. Recent successful examples include the characterization of antifungal agents targeting group II intron ribozymes^{89,90}, which demonstrated that protein-based protocols can be effectively adapted to predict RNA-ligand binding affinities, with similar success reported in two independent benchmarking studies.

Despite these promising initial benchmarks, significant challenges in applying standard RBFE estimation methods to RNA-ligand complexes emerge^{338–340}. Unlike proteins, RNA molecules display inherent conformational flexibility, existing as structural ensembles rather than unique states. This complicates the experimental identification of exact binding modes and functionally relevant conformations for RBFE calculations^{254,341}. Molecular dynamics (MD) simulations can address this challenge, as a powerful tool to link static experimental structures with the dynamic conformational landscapes that exist in solution. MD has proved particularly crucial to characterize RNA-ligand conformational ensembles and recognition modes. This has been successfully demonstrated for ligand targeting, e.g., the HIV-1 TAR²⁵⁷ RNA element and the Tte PreQ₁ riboswitch³⁴². However, the optimal extent of preliminary MD sampling required to perform reliable RBFE predictions remains undefined, and the systematic assessment of the impact of preliminary MD on RNA-ligand RBFE calculations remains largely unexplored.

In this context, another challenge involves the role of electrostatic modeling in RNA-ligand RBFE calculations. RNA's limited chemical diversity and pronounced charge density create an electrostatic environment that complicates the design of RNA binders with drug-like properties. The degree to which this characteristic electrostatic environment affects computational predictions remains challenging to investigate. In this regard, recent developments in RNA polarizable FFs represent a potential solution with significantly reduced computational cost, but similar accuracy, compared to computationally expensive, quantum-based approaches^{343,344}. These advanced force fields have demonstrated success across diverse molecular systems³⁴⁵, enhancing hydrogen-bond network stability and enabling better

characterization of metastable RNA conformations³⁴⁶. Nevertheless, whether the additional computational complexity of polarizable models translates to meaningful improvements in RNA-ligand binding predictions has not been systematically evaluated, yet.

To assess these methodological challenges in RNA-ligand RBFE calculations, we systematically benchmark the performance of polarizable OPLS5³⁴⁴ versus fixed-charge OPLS4³⁴⁷ force fields across three representative RNA-targeted drug discovery scenarios. In particular, we show the role and impact of preparatory MD simulations for prediction accuracy, while evaluating when explicit polarizability provides tangible benefits over standard approaches. Our comprehensive assessment offers practical guidelines for selecting appropriate computational protocols based on system complexity and project requirements in RNA-targeted drug discovery.

5.3. Results

System Selection and dataset design: from fragments to metal-mediated RNA-ligand interactions

We designed a benchmark dataset to evaluate relative binding free energy (RBFE) prediction performance using fixed-charge versus polarizable force fields across the spectrum of challenges in RNA-targeted drug discovery. While the PDB contains over 205,000 protein-including structures, only ~2,000 RNA-only structures exist, with fewer than 1,000 RNA-ligand co-crystal structures available^{348,349}. Working within these constraints, we selected three benchmark systems that progress from well-resolved RNA-fragment complexes to increasingly challenging scenarios involving structural gaps and essential metal-mediated RNA-ligand interactions (**Figure 5.1 A-C**).

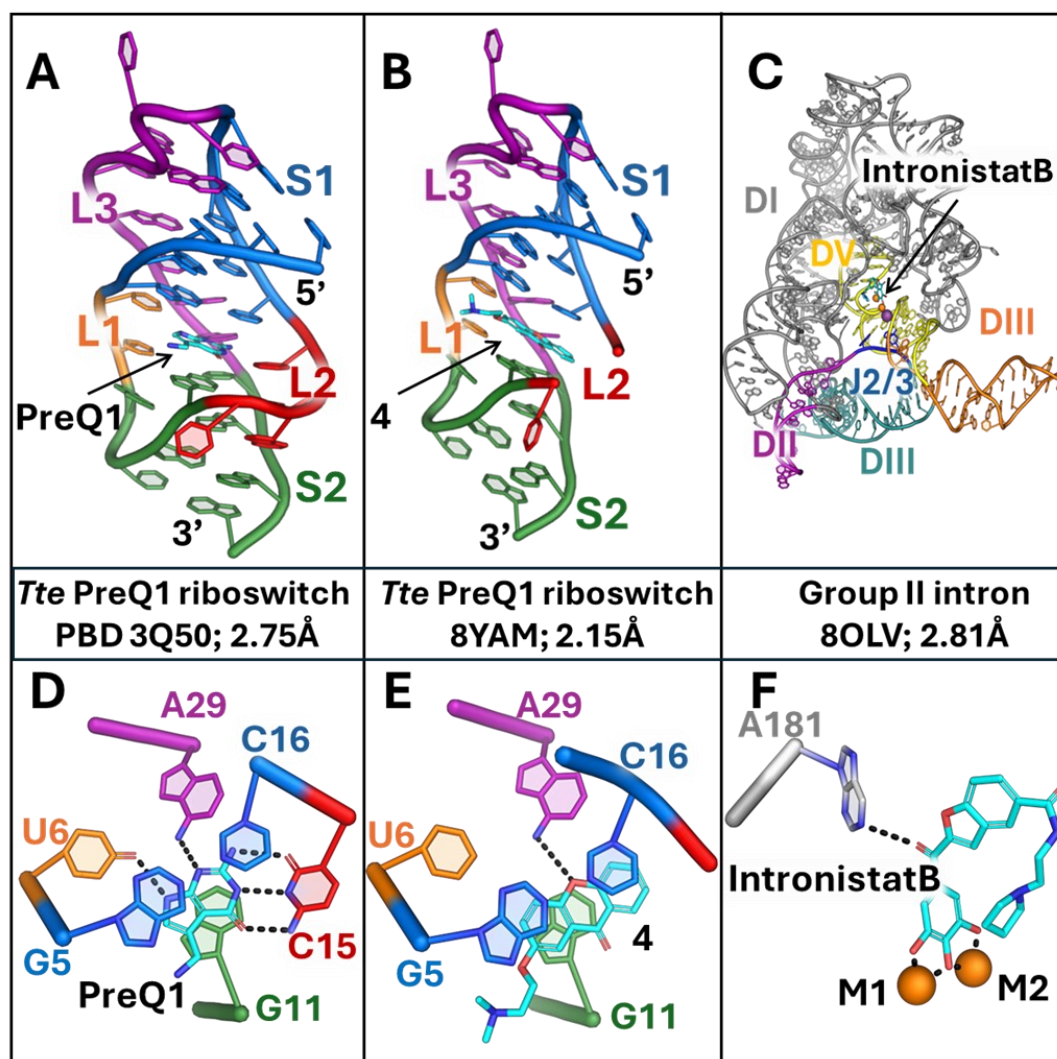
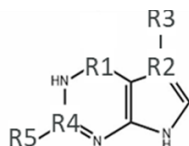


Figure 5.1 Co-crystal structures of the three case studies. (A) Case study I: *Thermoanaerobacter tengcongensis* (Tte) PreQ₁ riboswitch bound to its natural ligand PreQ₁ (PDB ID: 3Q50, 2.75 Å). The receptor is shown in cartoon representation, with domains color-coded as follows: S1 (marine blue), S2 (forest green), L1 (orange), L2 (red), and L3 (purple). The PreQ₁ ligand is shown in cyan licorice. All residues are fully resolved, revealing the native base orientations within the binding pocket (D). (B) Co-crystal structure of the same Tte PreQ₁ riboswitch (PDB ID: 8YAM, 2.15 Å) in complex with a small-molecule analog. The same color scheme as in (A) is maintained to emphasize sequence and structural conservation. The gap within the active site, and the abasic bases are visible at domain L2. Zoom in the active site (D) shows the binding mode of the small molecule (dubbed 4) of the series. (C) Case study III: *Oceanobacillus iheyensis* group II intron bound to intronistat B (PDB ID: 8OLV, 2.81 Å). The receptor is shown in cartoon form with domains color-coded as follows: DI (grey), DII (purple), DII–DIII junction (blue), DIII (cyan), DIV (orange), and DV (yellow). The ligand is represented in cyan licorice. Panel (F) provides a close-up of the active site, highlighting the coordination of the pyrogallol moiety with the two catalytic Mg²⁺ ions (orange spheres).

Case studies I and II focus on the *Thermoanaerobacter tengcongensis* (Tte) PreQ₁ riboswitch in complex with (i) natural ligand-like fragments^{350,351} (Table 5.1), and (ii) synthetic small molecule ligands³³⁷ (Figure 5.1A–B, Table 5.2), respectively. In both cases, ligands were connected with a multi-edge perturbation map, calculating ligand similarity with the Canvas MCS maximum common substructure algorithm implemented in FEP+. These systems allow us to probe RBE sensitivity across ligand size, paralleling protein-ligand systems where

fragment-series RBE calculations are often highly predictive due to reduced conformational complexity^{352,353}. Specifically, case study I examines a congeneric series of the natural PreQ₁ ligand, including eight small guanine derivatives in which up to three heavy atoms are replaced in a single transformation. K_i values were determined with in-line probing analysis, ranging over ~10 μ M³⁵⁰ (**Table 5.1**). The coordinates of the fully resolved ligand-bound riboswitch were extracted from the PDBid 3Q50³⁵¹ (2.75 Å, **Figure 5.1A**).

Table 5.1. Ligands series for case study I. The ligands are guanine derivatives fragments with $\Delta K_i=10\mu$ M



Ligand	R1	R2	R3	R4	R5	K _i (μ M)
1	CO	C	CONH ₂	C	NH ₂	0.03
2	CO	C	CH ₃ NH ₂	N	NH ₂	0.02
3	CO	N	/	C	NH ₂	0.4
4	CO	C	HCN	C	NH ₂	0.1
5	CO	C	H	C	NH ₂	1.6
6	CO	C	H	C	CH ₃ NH ₂	1.6
7	CH	N	/	C	NH ₂	10
8	CO	N	CH ₃	C	NH ₂	5

Case study II examines an additional set of eight recently discovered PreQ₁ riboswitch binders structurally unrelated to the natural ligand³³⁷ (**Table 5.2**). Compounds 1-5 share a xanthone scaffold, with compounds 2 and 3 differing by an oxygen-to-sulfur substitution within the heterocycle. Compounds 7 and 8 contain a distinct seven-membered heterocyclic scaffold, while compound 9 features a harmol core. K_i values were determined using intrinsic ligand fluorescence and MST measurements, spanning ~500 μ M³³⁷ (**Table 5.2**). Notably, the ligand-bound structure used in case study II (PDBid 8YAM, 2.15 Å) contains three abasic nucleotides (**Figure 5.1B**) and is only partially resolved at the active site. For this reason, we could evaluate RBE calculations, combined with structural modeling approaches and equilibrium molecular dynamics, as opposed to a more conventional use on RBE calculations on high-quality models of the target.

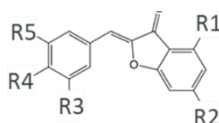
Table 5.2. **Ligands series for case study II.** The ligands are small molecules with $\Delta K_i=500\mu M$

Ligand	R1	R2	R3	R4	K _i (μ M)
1	C ₁₀ H ₁₃ O ₂ N ₂	C	O	H	8.0 $\pm$ 9
2	OCH ₂ CONH ₂	C	S	H	1.6 $\pm$ 1.6
3	OCH ₂ CH ₂ (NH ₃) ₂	N	S	H	58 $\pm$ 30
4	OCH ₂ CH ₂ (NH ₃) ₂	C	O	H	16 $\pm$ 21
5	OCH ₂ CONH ₂	C	O	H	>500
6	NHCH ₃	N	O	OH	>500
7					18.3 $\pm$ 7
8					1.9 $\pm$ 0.2
9					104.4 $\pm$ 20

Case study III features two novel series of antifungal agents targeting group II introns where direct metal-ligand coordination is essential for binding. These ligands were also connected via a multi-edge perturbation map, and K_i values were obtained from radioanalytical self-splicing assays⁸⁹. The ligands can be subdivided into three structural regions (A, B, and C). In series 1 (**Table 5.3**), the chemical space of region A is explored, with the pyrogallol moiety being essential for activity, while its replacement with a catechol or resorcinol markedly reduces activity (K_i difference of $\sim 249\mu M$ between compounds 1 and 3). In the same series, region B features a conserved benzofuran-3-one scaffold with varied hydroxylation, while region C is consistently represented by a hydrogen. In series 2 (**Table 5.4**), region A remains a pyrogallol, whereas region B alternates between benzofuran-3-one and benzofuran, with the linker between regions A and B being either a carbonyl or an α -unsaturated ketone, producing the largest affinity change ($\Delta K_i \sim 5.4\mu M$ between compounds 8 and 13). Region C consists of chemically diverse tails, including aromatic groups and tertiary amines. Importantly, the binding mode of these compounds has been recently characterized (PDBid 8OLV, 2.81Å), showing that this system represents a particularly challenging test for RBE calculations (**Figure 5.1C, F**). Indeed, the group II intron binding pocket is solvent-exposed, and ligand

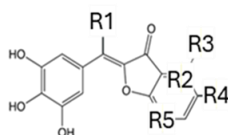
binding involves direct metal coordination⁹⁰, creating a critical benchmark for evaluating polarizable force field models.

Table 5.3. Ligands series 1 for case study III. The ligands are small molecules, all coordinating the Mg^{2+} ions with $\Delta Ki=990\mu M$



Ligand	R1	R2	R3	R4	R5	Ki(μM)
1	OH	OH	OH	OH	OH	2.3 $\pm$ 0.4
2	OH	OH	H	OH	OH	>100
3	OH	OH	OH	H	OH	255 $\pm$ 37
4	OH	OH	H	COOH	COOH	>1000
6	OH	H	OH	OH	OH	8 $\pm$ 2
7	H	H	OH	OH	OH	7 $\pm$ 1

Table 5.4. Ligands series 2 for case study III. The ligands are small molecules, all coordinating the Mg^{2+} ions with $\Delta Ki=5.3\mu M$



Ligand	R1	R2	R3	R4	R5	Ki(μM)
6	H	C	OH	H	H	8 $\pm$ 2
7	H	C	H	H	H	7 $\pm$ 1
8	O	C	H	Cl	Cl	33 $\pm$ 3
9	O	C	H	H	H	3.7 $\pm$ 0.6
12	O	C	H	C ₅ H ₃ NCH ₃	H	6 $\pm$ 8
13	O	C	H	PhCH ₂ OH	H	0.9 $\pm$ 0.1
14	O	C	H	PhCH ₃ N(CH ₃ CH ₂) ₂	H	2.1 $\pm$ 0.4
15	O	C	H	PhCON(CH ₃ CH ₂) ₂	H	2.4 $\pm$ 0.4
16	O	C	H	PhSO ₂ NH(C ₄ H ₉)	H	9.9 $\pm$ 0.7
17	O	N	H	C ₅ H ₃ NCF ₃	H	2.8 $\pm$ 0.2
18	O	C	H	PhCOCH ₃	H	1.5 $\pm$ 0.1
19						3.8 $\pm$ 0.5

The aforementioned benchmark was then used to assess the following RBEF performance metrics, using both polarizable (OPLS5) and fixed-charge (OPLS4) force fields: 1) compounds ranking with Kendall's τ and Pearson coefficient (R^2), as well as pairwise mean unsigned error (MUE) and 2) the pairwise root-mean-square error (RMSE) between predicted and

experimental $\Delta\Delta G$. Additionally, predicted ΔG is computed from the $\Delta\Delta G$ by FEP+ standard workflow³⁵⁴, and ligand-dependent constants based on the experimental ΔG available. We estimated the error on the predicted ΔG by analyzing cycle closure of the transformations (CCC error and CCC convergence) and the cycle closure hysteresis.

Collectively, these three case studies provide a comprehensive evaluation framework spanning key RNA drug discovery challenges – from ligand size effects and incomplete structural data to metal coordination – enabling systematic assessment of polarizability effects on both structural stability and binding predictions.

Case study I: explicit polarizability improves structural sampling with limited impact on fragment ranking accuracy

We began our RBFE benchmark by examining the PreQ₁ riboswitch bound with native ligand-like fragments. The X-ray crystal structure³⁵¹ (PDB 3Q50, 2.75 Å) captures the riboswitch bound to its natural ligand, PreQ₁, which forms a well-characterized^{351,355} hydrogen-bond network with riboswitch bases C15, U6, and A29 while being sandwiched between two base-pairs (G5-C16 and G11-A14) (**Figure 5.1D**).

To assess the structural stability of this complex, we first performed an MD-based equilibration protocol with two rounds of NVT and three of NPT followed by 50 ns of MD simulations using OPLS5 force field. Notably, the equilibration induced a subtle but important change in the PreQ₁ binding position, enabling an interaction between the atom N9 of the ligand and nucleotide C7, not observed in the crystal structure (**Figure 5.2A**). This configuration showed overall stability also during equilibrium MD simulations, as demonstrated by steady root mean square fluctuation (RMSF) and root mean square deviation (RMSD) values, which are below 1Å for both the ligand and the receptor throughout the simulations, respectively.

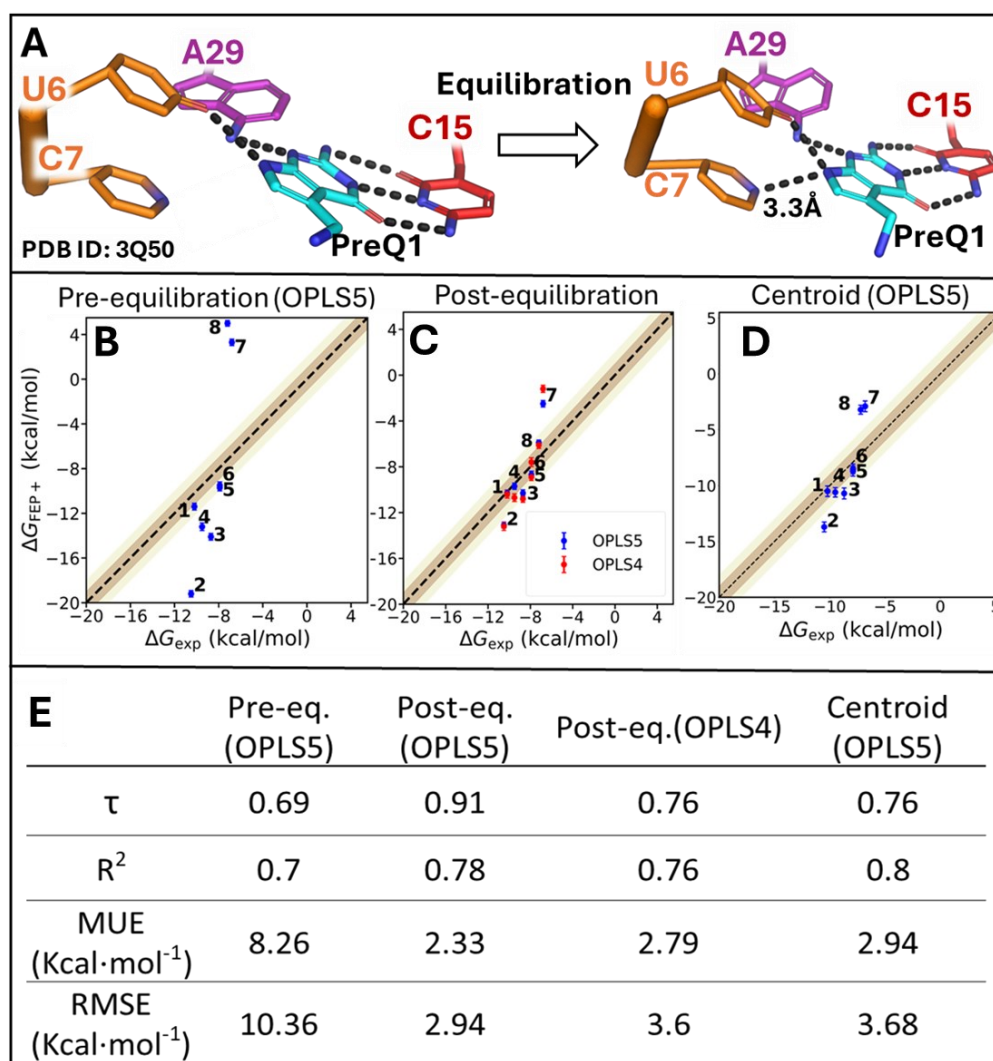


Figure 5.2. Case study I: MD equilibration improves FEP predictions on RNA-fragments. (A) Structural rearrangement within the riboswitch active site before (left) and after (right) MD equilibration. The receptor is shown in cartoon representation, with nucleotide C7 (orange) forming a new interaction with the ligand (cyan licorice) upon equilibration. (B–D) Predicted versus experimental ΔG values for the pre-equilibrated, post-equilibrated, and centroid starting structures, respectively. In (C), red points indicate results obtained from FEP calculations performed with the OPLS4 force field on a structure equilibrated using the same FF. (E) Summary of evaluation metrics (Kendall's τ , R^2 , MUE, and RMSE) for all RBEF predictions in this case study, highlighting the improved agreement with experimental data after equilibration.

Given this conformational rearrangement observed during MD, we next quantified the impact of our structural relaxation protocol on RBEF accuracy, using three different starting coordinates: the original X-ray crystal structure (i.e., the pre-equilibration structure), the equilibrated structure (i.e., the post-equilibration structure), and the centroid of the most populated MD cluster (i.e., the MD centroid structure). We employed 12, 16, and 24 lambda windows of 5 ns each for default, core hopping, and charge transformations, respectively, following the FEP+ standard protocol. For the best-performing systems, we tested alternative simulation protocols including doubled lambda window counts, with and without modified ionic atmospheres. We report here the optimal conditions (20 ns per lambda window with

standard lambda spacing).

The exact starting structure employed for the calculations proved to be critical for prediction accuracy. The poorest predictions were obtained when starting from the pre-equilibration structure (Kendall's $\tau = 0.69$, $R^2=0.7$ MUE = 8.3 kcal·mol⁻¹, RMSE=10.37 kcal·mol⁻¹, **figure 5.2B**), while the best performance was obtained with the post-equilibration structure ($\tau = 0.91$, $R^2 = 0.77$, MUE = 2.34 kcal·mol⁻¹, RMSE=2.97, **figure 5.2C**). This improvement stems directly from the equilibration-induced formation of the ligand-C7 interaction (**Figure 5.2A**), which particularly benefited predictions for 7-methylguanine – a compound that was poorly ranked without this equilibration-induced structural rearrangement. Visual inspection of MD trajectories confirmed that this ligand consistently establishes an additional contact with nucleotide C7. This nucleotide underwent further rearrangement during alchemical simulations to stabilize this interaction over the competing contact with U6, providing a structural rationale for the improved predictive accuracy (**Figure 5.2A**). Notably, RBE calculations starting from the MD centroid structure did not further improve predictions ($\tau = 0.76$, $R^2 = 0.80$, MUE = 2.95 kcal·mol⁻¹, RMSE = 3.39 kcal·mol⁻¹; **Figure 5.2D, E**). Additional protocol modifications, such as extended simulation times, increased lambda windows, or modified ionic atmospheres, also failed to enhance performance (Data not shown). This suggests that proper MD-based system equilibration and standard FEP+ protocol are sufficient for accurate fragment ranking in well-resolved RNA-ligand complexes.

To isolate the effects of explicit polarizability from benefits coming from the MD equilibration, we repeated the structural relaxation protocol using the OPLS4 force field and compared RBE performance with the best-performing OPLS5-based model, i.e., the post-equilibration structure (**Figure 5.2C**). Remarkably, OPLS4 failed to sample the crucial ligand-C7 interaction that emerged with OPLS5, yet both force fields achieved comparable RBE accuracy (**Figure 5.2C, E**). This suggests that, for small ligands in structurally stable systems, explicit polarizability had minimal impact on binding affinity predictions. However, the enhanced conformational sampling observed with OPLS5 can be beneficial for identifying novel interaction hotspots that could guide the identification of growing vectors and strategies for rational ligand optimization.

Case study II: MD-Driven recovery of native-like riboswitch-ligand conformations

enables accurate RNA-ligand affinity predictions

Next, we focused on the PreQ₁ riboswitch in a complex of synthetic small molecule ligands that are structurally unrelated to the natural PreQ₁ substrate. X-ray crystal structures of the *T. tengcongensis* (Tte) PreQ₁ riboswitch bound to ligands 4 (xanthone core) and 9 (harmol core) revealed that these chemically distinct scaffolds bind at the same site with similar modes³³⁷ (PDB IDs: 8YAM, 8YAN, resolved at 2.15 and 2.25 Å, respectively). In the complex with compound 4 (PDB 8YAM), the xanthone core stacks against conserved nucleotides G11, G5, and C16, and forms a hydrogen bond with A29 (**Figure 5.1B, E**) – key interactions also observed for the natural ligand PreQ₁ (**Figure 5.1D**). Remarkably, these complexes presented a significant structural challenge as they contained three abasic nucleotides at the binding cavity (A13–C15 missing the nucleobase and only made by the backbone in the crystallization construct), with A14 being completely unresolved.

To address these structural gaps and replicate the experimental construct used for Ki-determination, we modelled the missing bases and unresolved nucleotides, and performed 20 ns of partially restrained MD simulations to adapt the modelled bases to the ligand pose. Then, we carried out 30 ns of unrestrained equilibrium MD simulations using the polarizable OPLS5 force field. The resulting trajectories confirmed the overall system stability, with compound 4 maintaining an H-bond with the A29. Most remarkably, the modelled bases converged toward a native-like orientation resembling the PreQ₁-bound structure during the equilibrium MD simulations (**Figure 5.3A**). Indeed, the RMSD of the modelled bases, and their neighbouring nucleotides relative to their counterparts in the fully-resolved structure bound to the natural ligand PreQ₁ (3Q50), decreased from ~4 Å to ~2 Å during the equilibrium MD simulations (**Figure 5.3A**). This spontaneous structural recovery restored critical features of the preQ₁ riboswitch structure, such as the G11-A14 base pairing, which stabilized the small molecule binding cavity in a native-like architecture (**Figure 5.3A**). Control simulations of the abasic receptor confirmed its unsuitability for RBFE calculations, as the missing bases rendered the active site highly flexible and overly solvent-exposed, leading to ligand dissociation within nanoseconds of MD simulations.

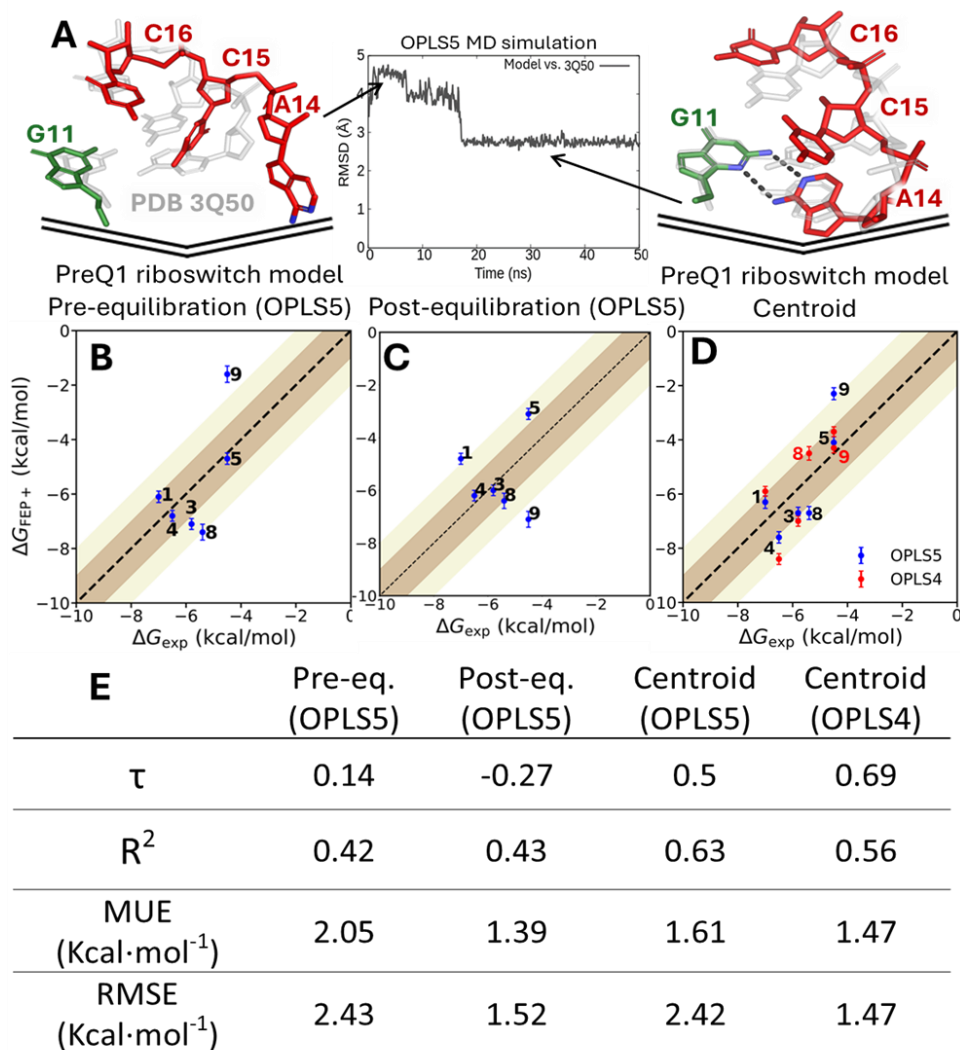


Figure 5.3 Case study II: benefits of MD sampling on RBEF prediction. ((A) Molecular dynamics (MD) relaxation of the modelled PreQ₁ riboswitch active site leads to recovery of a native-like conformation. On the left, modelled bases from PDB 8YAM are shown in red licorice, and G11 in green, overlaid with the native conformation from PDB 3Q50 (transparent grey licorice). After ~20 ns, the A14–G11 base pairing is re-established (right panel), restoring a native-like active site geometry. The central panel shows the RMSD of residues A14, C15, C16, and G11 with respect to 3Q50 over the 50-ns trajectory, confirming structural stabilization. (B–D) FEP results for the pre-equilibrated, post-equilibrated, and centroid structures, respectively, with the corresponding performance metrics summarized in (E). (B) and (C) reports only the data with OPLS5, while the best performing system (D) was repeated with OPLS4 for comparison. In (D), the data obtained with the OPLS5 force field are shown in blue, while results from OPLS4 are shown in red.

Building on this structural recovery of native-like riboswitch-ligand complex using the polarizable force field, we extracted three representative structures from the MD relaxation protocol to perform RBEF calculations (pre- and post-equilibration models, and the native-like centroid of the most populated MD cluster). Similar to what was observed for case study I, the exact starting structure again proved to be critical for prediction accuracy, here showing even more pronounced effects. Calculations using pre-equilibrated ($\tau = 0.14$, $R^2 = 0.42$, MUE = 2.08 kcal·mol⁻¹, RMSE = 2.45 kcal·mol⁻¹; **Figure 5.3B**) and post-equilibrated structures ($\tau = -0.27$,

$R^2 = 0.01$, MUE = 2.11 kcal·mol⁻¹, RMSE = 2.42 kcal·mol⁻¹, **Figure 5.3C, E**) yielded poor correlations, likely reflecting the impact of misoriented bases on binding predictions. In contrast, using the MD-sampled native-like structure achieved substantially improved agreement with experimental data ($\tau = 0.50$, $R^2 = 0.63$, MUE = 1.61 kcal·mol⁻¹, RMSE = 1.92 kcal·mol⁻¹; **Figure 5.3D**). This performance improvement directly resulted from the spontaneous recovery of native-like RNA-RNA and RNA-ligand contacts during the MD simulations. For instance, in both pre- and post-equilibration systems, C15 formed a hydrogen bond with ligand 9, likely over-stabilizing this binding pose. Instead, in the native-like MD-centroid structure, the restored C15-A14-G11 interaction network prevented this H-bond contact (**Figure 5.3A, D, E**).

Despite these improvements, we also identified three outliers based on calculations starting from the MD centroid structure. In detail, compounds 7 and 8 consistently appeared as outliers across all starting structures (data not shown), with their flexible seven-membered rings frequently adopting conformations that disrupted essential π -stacking interactions and led to binding poorly stable binding poses during simulations. However, given the absence of structural data for these compounds, we can only argue that their experimental affinities may reflect their binding to alternative sites. Compound 2 was also an outlier, despite retaining stable and native-like contacts during the simulations. RBE estimates for these outliers showed no improvement when varying perturbation maps, doubling lambda windows, or modifying ionic conditions. We therefore speculate that this may reflect different binding orientations that have not been observed in this study.

To evaluate whether explicit polarizability was essential for the observed structural recovery, we performed the same workflow using the fixed-charge OPLS4 force field. Importantly, MD-based structural modeling of the unresolved gap failed in the absence of explicit polarizability. OPLS4-based equilibrium simulations could only retrieve the RNA-ligand interactions between the ligand 4 and A29, but A14 remained bulged, preventing recovery of a native-like conformation. Despite this structural limitation, OPLS4-based RBE predictions showed a similar trend compared to best-performing OPLS5-based calculations (MD centroid, $\tau = 0.69$, $R^2 = 0.56$, MUE = 1.47 kcal·mol⁻¹, RMSE = 4.27 kcal·mol⁻¹; **Figure 5.3D, E**), with the compounds 2, 7, and 8 appearing as outliers – arguably for same reasons as for OPLS5-based models. This suggests that explicit polarizability is essential for reliable structural sampling of flexible RNA regions during MD simulations, while having less impact on RBE calculations, where error cancellation may still provide a beneficial self-correcting factor even with suboptimal force fields.

Taken together, these results demonstrate that MD sampling is essential for recovering native-like RNA-RNA and RNA-ligand interactions from computationally modeled structures, enabling RBE calculations to capture critical binding determinants and achieve meaningful predictive accuracy even when starting from partially unresolved crystallographic data.

Case Study III: ligand-metal coordination and non-zero net charge transformations exposes critical boundaries of current alchemical methods

Case study III addresses one of the most challenging scenarios: modeling RNA-ligand interactions mediated by direct metal coordination, where metal-driven polarization effects may critically influence ligand binding properties. To exemplify such a case, we selected the *Oceanobacillus iheyensis* group II intron, a structurally ordered RNA molecule comprising approximately 450 nucleotides, using the highest resolution available structure⁹⁰ (PDB 8OLV, 2.81 Å) bound to a dibrominated derivative of intronistat B. In this complex, the ligand occupies the two-metal-ion active site, directly coordinating both catalytic Mg²⁺ ions while engaging nucleotide A181 through its carbonyl linker (**Figure 5.1C, F**).

As for the case study I, we applied a structural relaxation protocol after computationally converting the dibrominated derivative to Intronistat B by replacing bromine atoms with hydrogens. Following the MD-based equilibration and 50 ns of equilibrium MD using OPLS5, the system maintained structural stability with preserved metal coordination as the primary RNA-ligand anchor. The Mg-Mg and pyrogallol-Mg distances remained intact, though the A181 hydrogen bond was replaced by transient contacts with nucleotide G1 at the intron 5' terminus, consistent with previous observations. We extracted three starting configurations (pre-equilibration, post-equilibration, and MD centroid structures) for RBE calculations across both ligand series.

Series I comprised ligands containing the α -unsaturated ketone linker and varying at both regions A and B. Compounds 2, 3, and 4 differed solely at region A through pyrogallol modifications, while compounds 1, 6, and 7 shared the pyrogallol core but featured different R-group substitutions on the benzofuran scaffold of region B. Although all models achieved similar ranking performance (Figure 5.4A-D), the post-equilibration system delivered the best estimates with the lowest prediction error (Kendall $\tau = 0.67$, R²=0.82, MUE=0.59, RMSE=0.69, **Figure 5.4B**).

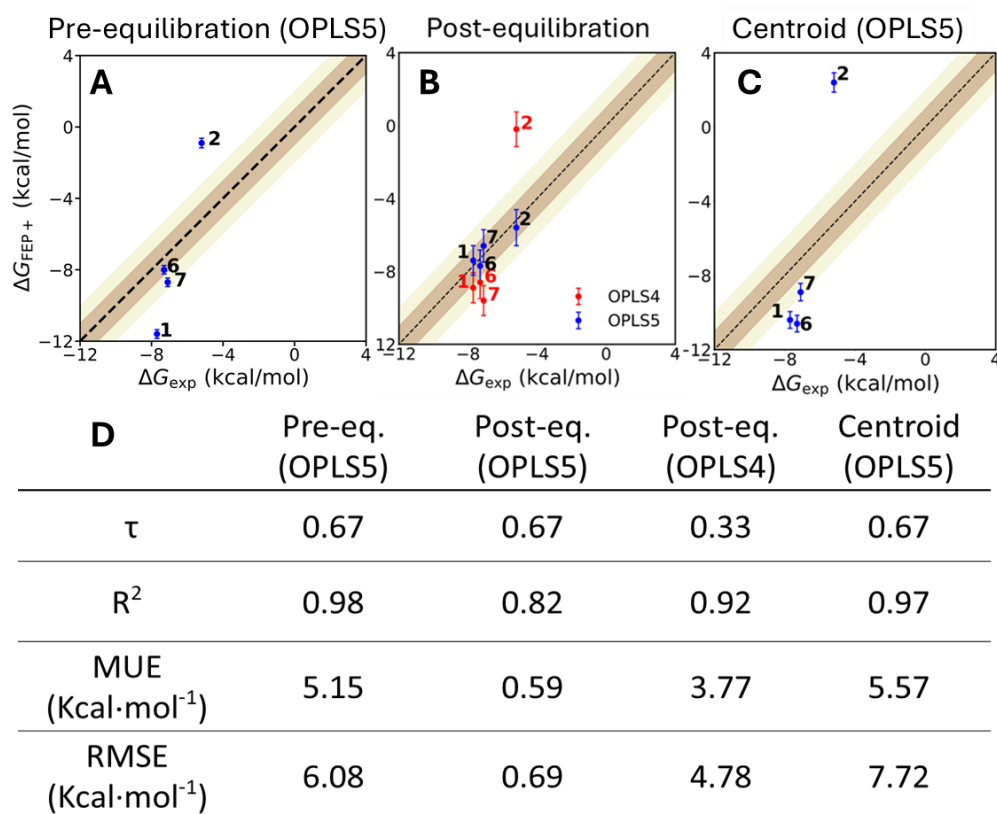


Figure 5.4. **Case study III, series 1: MD equilibration favours RBFE prediction.** (A–C) Predicted versus experimental ΔG values for the pre-equilibrated, post-equilibrated, and centroid starting structures, respectively. In all panels, blue points correspond to data obtained with the OPLS5 force field, while in (B) red points indicate results reproduced using OPLS4 on the same equilibrated structure. (E) Summary of the evaluation metrics (Kendall's τ , R^2 , MUE, and RMSE) for all RBFE predictions, highlighting that MD equilibration leads to improved correlation with experimental affinities compared to both pre-equilibrated and centroid structures.

However, non-zero net charge transformations revealed fundamental limitations of current alchemical methods for metal-coordinated systems. All perturbations involving compounds 3 and 4 – the only non-zero net charge transformations – produced poorly predicted affinities, regardless of cycle closure quality. Indeed, while compound 3 was correctly identified as inactive, its predicted binding affinity drastically deviated from experimental values. Analysis of the MD trajectory revealed that compound 3 spontaneously dissociated from the binding site, losing all metal coordination during the simulations. This behavior stemmed from the critical para-OH to para-H substitution relative to intronistat B. The para-OH group of intronistat B is indeed deprotonated and essential for simultaneous coordination of both catalytic Mg^{2+} ions, satisfying the geometric and electronic requirements of this two-metal-ion center. Replacing this group with hydrogen created steric clashes and electronic incompatibility, destabilizing the coordination complex and causing ligand dissociation during simulation, suggesting that RBFE estimation inaccuracies for compound 3 were linked to structural instability of the complex. Compound 4 presented a different but equally problematic

scenario. Despite remaining bound throughout the MD trajectories, it was incorrectly predicted as active (experimental $K_i > 1000 \mu\text{M}$, **Table 5.3**). As the only ligand introducing substantial changes to the pyrogallol chelating moiety, we hypothesized that further enhanced sampling simulations might improve the energetic predictions. However, doubling the lambda windows unexpectedly triggered severe cycle closure failures for all transformations involving compounds 3 and 4. This outcome demonstrates that even extended sampling combined with explicit polarizability could not adequately handle alchemical transformations occurring at metal-ligand coordination centers, particularly those involving net charge changes. However, fixed-charge OPLS4 calculations showed increased fundamental limitations (**Figure 5.4B, D**), leading to RBFE models with the worst compound ranking accuracy, indicating that polarizable force fields still offer an advantage in metal-coordinated systems.

Building on the challenges observed with series I, series II provided an opportunity to evaluate systems without structural variations at the metal-coordination center. Indeed, this series included only pyrogallol-containing ligands, characterized by variations at region C and modifications to the linker between regions A and B.

For zero net-charge transformations, the post-equilibration model demonstrated superior performance across all statistical measures with OPLS5 ($\tau = 0.43$, $R^2 = 0.31$, $\text{MUE} = 0.54 \text{ kcal}\cdot\text{mol}^{-1}$, $\text{RMSE} = 0.72 \text{ kcal}\cdot\text{mol}^{-1}$, **Figure 5B, D**). Conversely, RBFE results from both pre-equilibrated and MD-centroid models remained unreliable due to high hysteresis in transformations involving compounds 6, 7 and 8 (**Figure 5.5A, C, D**). This pattern mirrors case study I, where optimal performance emerged from post-equilibration models, reinforcing that MD-based equilibration protocols sufficiently enhance RBFE reliability when starting from well-characterized experimental structures. However, the fundamental challenge of charge-altering transformations persisted. Despite the more conservative small molecules structural framework, charged ligands 14 and 19 were poorly predicted across all models and tested conditions, confirming that charge-changing perturbations represent a systematic limitation in RNA-ligand RBFE calculations, similar to challenges observed in protein-ligand systems.

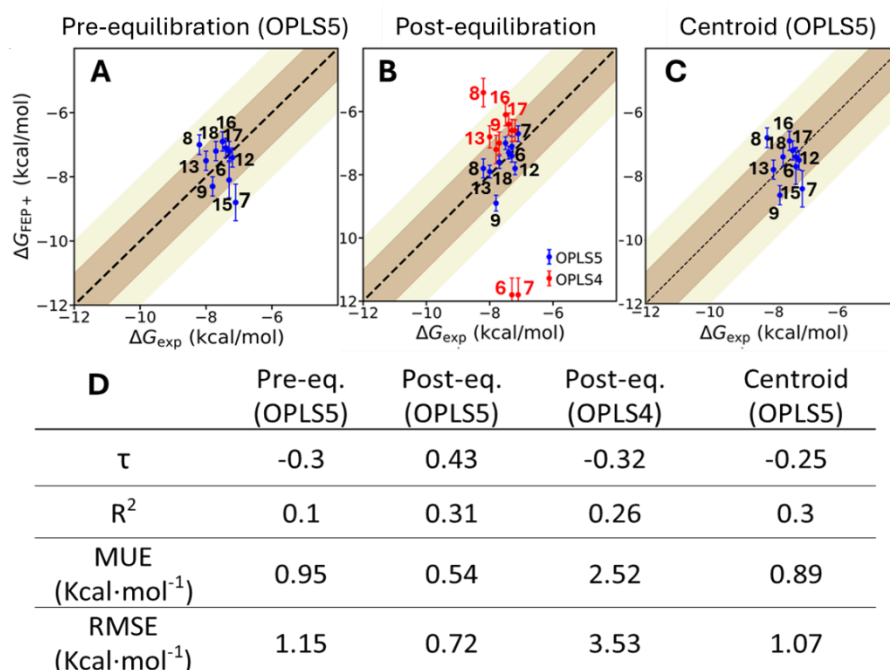


Figure 5.5 **Case study III, series 2: MD equilibration improved RBE prediction accuracy.** (A–C) Predicted versus experimental ΔG values for the pre-equilibrated, post-equilibrated, and centroid starting structures, respectively. In all panels, blue points correspond to data obtained with the OPLS5 force field, while in (B) red points indicate results reproduced using OPLS4 on the same equilibrated structure. (E) Summary of the evaluation metrics (Kendall's τ , R^2 , MUE, and RMSE) for all RBE predictions, highlighting that MD equilibration leads to improved correlation with experimental affinities compared to both pre-equilibrated and centroid structures.

The comparison between polarizable and fixed-charge force fields revealed additional limitations that extended beyond charge-changing transformations. We run the best-performing OPLS5-based calculations (post-equilibration) using OPLS4, achieving consistently poor performance 4 (**Figure 5.5B, D**). Cycle closure criteria were never satisfied for charged compounds 14 and 19. Moreover, OPLS4 exhibited poor cycle closure with neutral compounds that had been successfully handled by OPLS5, such as compounds 6, 7, 8, and 9, suggesting additional inadequacies in modeling the metal-coordinated environment. This clustering indicates that scaffold hopping transformations are inadequately sampled with default OPLS4 protocols, particularly evident in the challenging α -unsaturated ketone versus carbonyl linker transformation that distinguishes compounds 6 and 7 from other series members. Additionally, transformations involving substantial differences in molecular size proved problematic – compounds 8 and 9, being the smallest carbonyl-containing ligands in the series, showed higher cycle closure errors when connected to larger compounds through alchemical perturbations. These systematic failures suggest that fixed-charge force fields struggle to adequately sample the complex conformational and electrostatic changes required for reliable alchemical transformations in metal-coordinated RNA systems. Such limitations could potentially be

addressed through enhanced sampling protocols or redesigned perturbation maps incorporating intermediate transformation steps to decompose large structural changes into smaller alchemical perturbations not explored in this study.

Taken together, these results emphasize that non-zero net charge transformations involving metal-coordinating regions represent fundamental challenges for FEP prediction accuracy, regardless of force field. While explicit polarizability reduces estimation uncertainties and enables more reliable convergence, fixed-charge models can achieve similar compound ranking performance for small, structurally conservative transformations that avoid both charge changes and significant scaffold modifications.

5.4. Conclusion

This study provides one of the first systematic evaluations of RBFE calculations applied to RNA – small molecule complexes, integrating equilibrium MD simulations with both polarizable (OPLS5) and fixed-charge (OPLS4) force fields. Across three representative case studies, ranging from fragment-like riboswitch binders to metal-coordinating inhibitors of group II introns, we established a practical framework for evaluating the predictive reliability of FEP in RNA-targeted drug discovery. Our results highlight the trade-offs associated with polarizable vs. fixed-charge FFs, the benefits of MD equilibration and sampling in improving the final RBFE prediction, and the high dependency of the predictive success on the quality of the starting structure.

Our results highlight the following key conclusions and take-home messages:

- 1) **MD equilibration is essential for accurate RBFE predictions.** Preliminary MD relaxation consistently improved ranking performance by correcting local geometries, recovering native-like conformations, and stabilizing key ligand–RNA interactions that were absent in the starting crystal or modeled structures.
- 2) **Explicit polarizability enhances structural realism and sampling efficiency.** Polarizable force fields, such as OPLS5, enabled spontaneous recovery of native-like RNA base pairing and hydrogen-bond networks, particularly in partially resolved or flexible systems. These effects were less pronounced for small, rigid ligands but remain critical when RNA conformational adaptability or metal coordination governed binding.
- 3) **Charge-changing transformations and metal coordination remain fundamental challenges.** Even with polarizable models, free energy estimation struggles with charge-changing perturbations and transformations involving metal-chelating groups. Likely,

these limitations stem from electronic and structural complexities not fully captured by current force fields and sampling schemes.

- 4) **Context determines the benefit of polarizability.** Polarizable force fields provide a measurable advantage in systems with high electrostatic heterogeneity or partially resolved structures but may not justify their higher computational cost in simpler, well-defined RNA – ligand complexes.

Collectively, these findings underscore that the success of RBFE predictions for RNA targets depends strongly on system preparation, force field choice, and the nature of the ligand – RNA interactions. While polarizable models improve accuracy and sampling in challenging contexts, achieving quantitative precision for metal-mediated or charge-varying systems will require further methodological development, such as refined perturbation mapping or enhanced sampling approaches. This work establishes a benchmark and provides methodological insights for future computational studies on RNA – ligand binding. The manuscript presenting these results is currently in preparation, and further discussion, detailed statistical analysis, and expanded methodological considerations will be provided in the final publication.

6. Overall Discussion, Conclusions and Future Outlook

Pre-mRNA splicing is a fundamental step in eukaryotic gene expression, during which coding exons are joined and non-coding introns are removed. This reaction is catalysed by the spliceosome, a highly dynamic ribonucleoprotein machinery that timely catalyses two consecutive transesterification reactions through precise protein-regulated RNA active site rearrangements⁹⁸. Being directly involved in shaping the cellular transcriptome, mutations in the splicing machinery and defects in the splicing mechanisms result in a wide range of human diseases and disorders^{122,123,146}. As a consequence, multidisciplinary efforts have focused on characterizing the spliceosomal splicing cycle to support the rational design of new therapeutics (**Chapter 1.4**). However, an atomistic level understanding of how these molecular machines reach precise catalytic control is not yet addressed. Achieving this aim requires high-resolution structural data combined with dynamic information. In this context, computational approaches can bridge the gap, providing atomistic insights into the splicing mechanism^{81,90,118–120,137,146,156,242,268}.

In this thesis, after a comprehensive structural analysis, we employed state-of-the-art MD simulations and Metadynamics on three spliceosomal complexes^{94–96} – B^{ACT}, B* and C. We elucidated how conserved positively charged residues of second shell proteins coordinate with catalytic ions and RNA elements, to remodel the active site during the first step of splicing. Our results revealed a precise set of conserved residues, such as Lys10 and Tyr3 of Prp11, Tyr155 and Phe161 of Cwc24, Arg610 and Arg614 of Prp8, Lys63 and Arg42 of Yju2, which dynamically cooperate to i) regulate the timely metal cluster formation and its placement and ii) maintain the overall site geometry while preventing premature reactions at early splicing stages. This multi-layered regulation underscores a fine evolutionary tuning of the active site, where structural remodelling are synchronized with metal exchange and ion-channel formation to ensure precise timing of catalysis. Interestingly, this mechanistic strategy mirrors those observed in other nucleic-acid-processing enzymes – such as RNase H1²⁴², BamHI²⁹⁹, or DNA polymerases³⁵⁶ – where lysine plays a similar role at the analogous of the K1 site to regulate the catalytic activity, underlying a shared evolutionary principle of charge-assisted catalysis. Moreover, further structural and functional parallels were found with group II introns. On top of the well-known conserved active site architecture and mechanism^{79,87}, we have individuated

further RNA – RNA contacts between the BPA and the 5'SS that preserve the catalytic geometry in the spliceosome as in the group II introns. Furthermore, this parallel extends beyond structural biology and pure mechanistic understanding: group II introns themselves can be pharmacologically targeted directly at the active site, as shown by the binding mechanism of intronstatB, disclosed by our group⁹⁰. This suggests that also the spliceosome harbours potential druggable hot spots within its conserved RNA catalytical site. Building upon this evidence, our mechanistic findings could open the way for rational design of small molecules to modulate splicing activity, with therapeutic effects in the broad spectrum of diseases that involve the spliceosome.

Besides the regulation of the gene expression through the splicing process, both coding and non-coding RNAs are central regulators of metabolism, cellular homeostasis and stress responses, such that their malfunctioning is nowadays linked to multiple human diseases^{18,25,92}. The biological activity of RNA arises from its three-dimensional structure, which defines both its function and its druggability, making RNA molecules compelling drug targets. Nonetheless, the repetitive sequence composition and intrinsic flexibility of RNA make it difficult to identify stable and specific binding sites. Additionally, the scarcity of high-resolution RNA-ligand complex structures limits the training and validation of predictive computational models. As a result, RNA-targeted drug discovery still lags behind its protein-based counterpart³⁵⁷, where advanced computational tools already complement experimental efforts and efficiently reduce the attrition rate.

To help bridge this gap, we systematically evaluated the application of FEP methods to RNA – small molecule systems. Through three case studies, ranging from riboswitch fragment – complexes to metal coordinating inhibitors of group II introns, we demonstrated that MD-based equilibration significantly improves the predictive power of FEP, while polarizable force fields further improve the modeling of flexible and electrostatically complex RNA environments. These results contribute to defining practical strategies for applying FEP to RNA targets and provide a benchmark for future developments in RNA-ligand affinity prediction.

In conclusion, this thesis demonstrates how computational methods can complement experimental data to investigate complex biological mechanisms and tend to maximize the largely unexplored therapeutic potential of RNA molecules. By resolving the atomistic details of RNA catalysis and improving the reliability of RNA – ligand binding predictions, this work contributes to advancing our understanding of RNA as both a fundamental biological catalyst and a target in modern drug discovery. In an era where RNA biology stands at the intersection of structural biochemistry, molecular medicine, and computational innovation, such integrative

approaches will be key to transforming mechanistic insights into therapeutic avenues.

Supporting materials for chapter 4

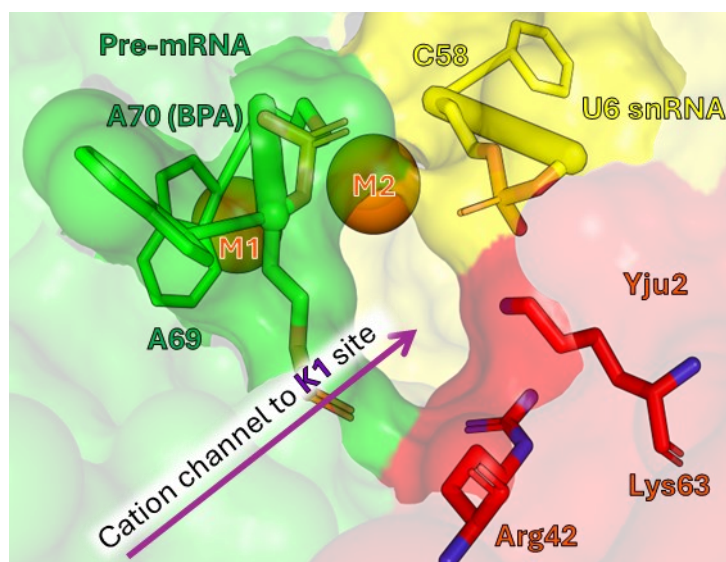


Figure S1. Cation channel connecting the K1 site and the bulk solvent in the B*b2 stage. Graphical representation of Yju2 residues Lys63 and Arg42 (red licorice), U6 snRNA nucleotide C58 (yellow stick), and pre-mRNA nucleotides A69 and A70 (green sticks), which collectively regulate the opening of a cation channel. The channel, depicted in its open conformation as an opaque surface, enables K⁺ entry into the spliceosomal active site. Two orange spheres mark the catalytic Mg²⁺ ions (M1 and M2), illustrating the direct access provided to the catalytic core.

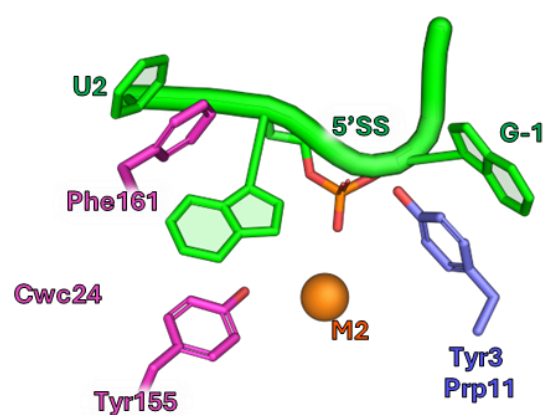


Figure S2. Interactions between Prp11, Cwc24, and the 5' splice site (5'SS) in the B_{ACT} complex. π -Stacking interactions observed in the B_{ACT} cryo-EM structure stabilize the kinked conformation of the 5'SS. Phe161 and Tyr155 of Cwc24 (magenta licorice) stack with U2 and G1 of the pre-mRNA (green licorice), while Tyr3 of Prp11 (blue licorice) stacks with G-1. These interactions position the scissile phosphate for coordination by the catalytic M2 ion.

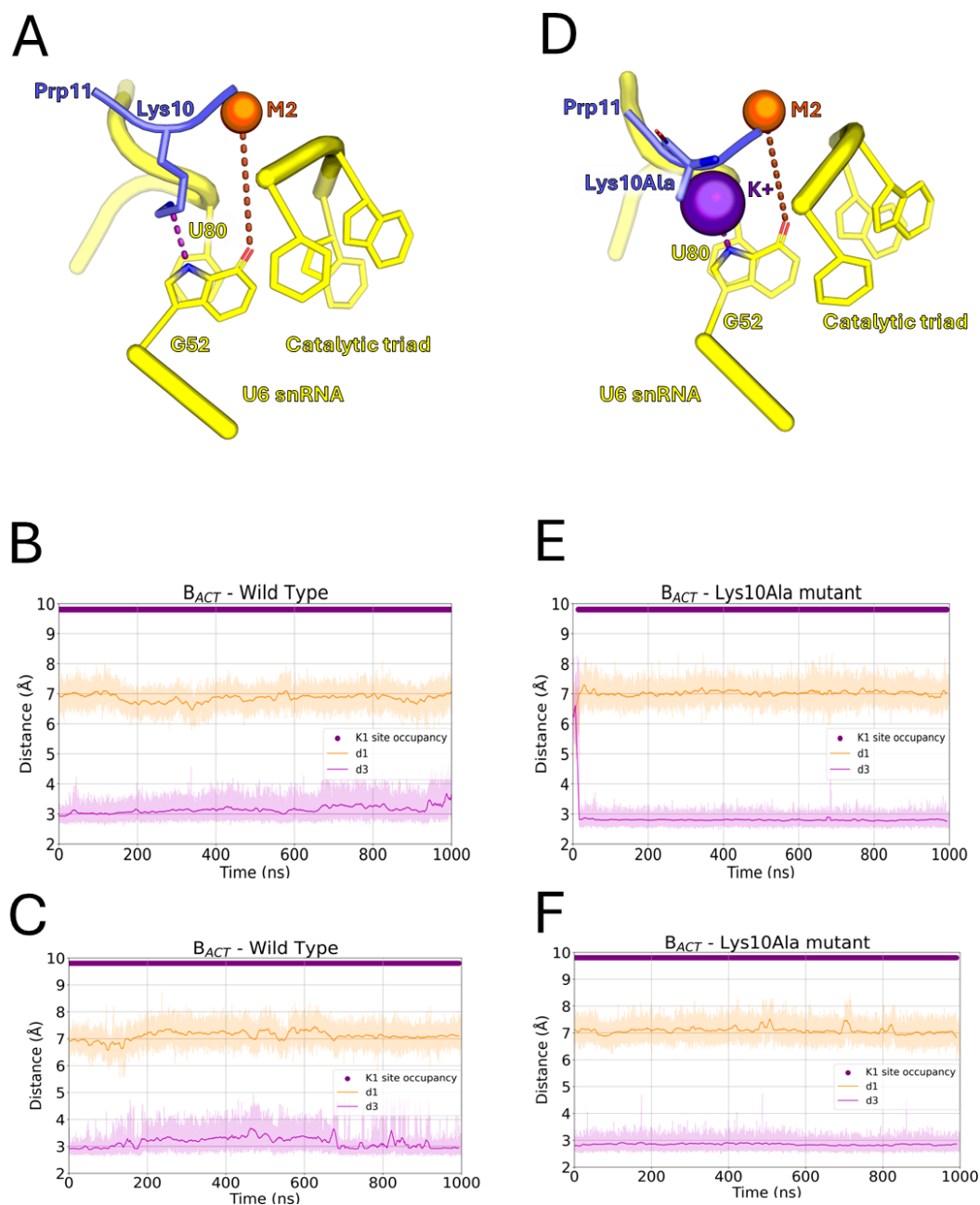


Figure S3. Molecular dynamics (MD) simulations of the B_{ACT} wild type and Prp11-Lys10Ala mutant. **A)** Definition of reference distances used to monitor the positions of Lys10 and M2 in the wild type. **B, C)** MD snapshots showing stable coordination of Lys10 at the K1 site, maintaining M2 positioning. **D)** Equivalent view for the Lys10Ala mutant as from panel A. **E, F)** Time evolution of D1 and D3 in two independent mutant simulations. Purple dots mark K1 site occupancy (distance < 3.5 Å). In F, K⁺ is recruited early and stabilizes M2. In E, K⁺ enters after ~25 ns, restoring M2 position following initial displacement.

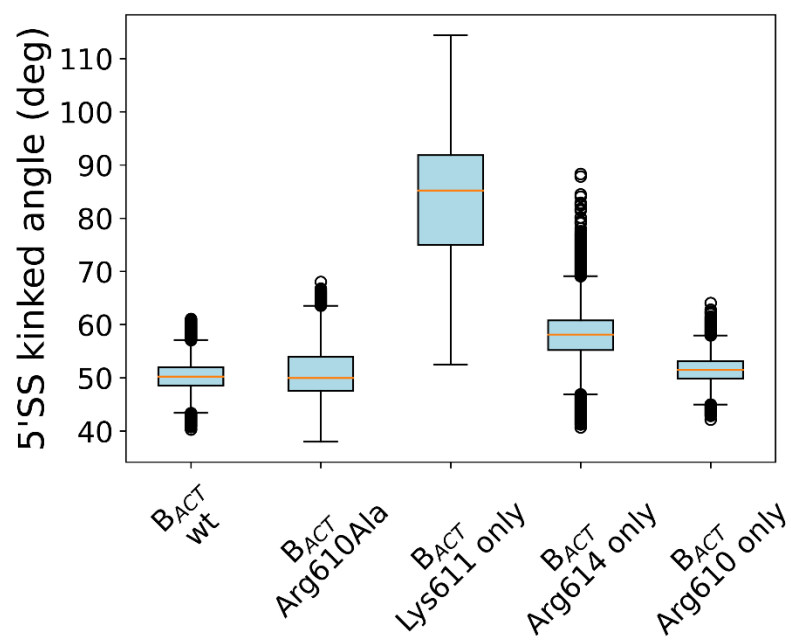


Figure S4. 5'SS angle variations and stability in B_{ACT} wild type and mutants. Boxplot showing the distribution of 5'SS angle values from the second replicate of each simulated B_{ACT} system. The angle measured in the wild type serves as a reference for the stable, kinked conformation of the 5'SS.

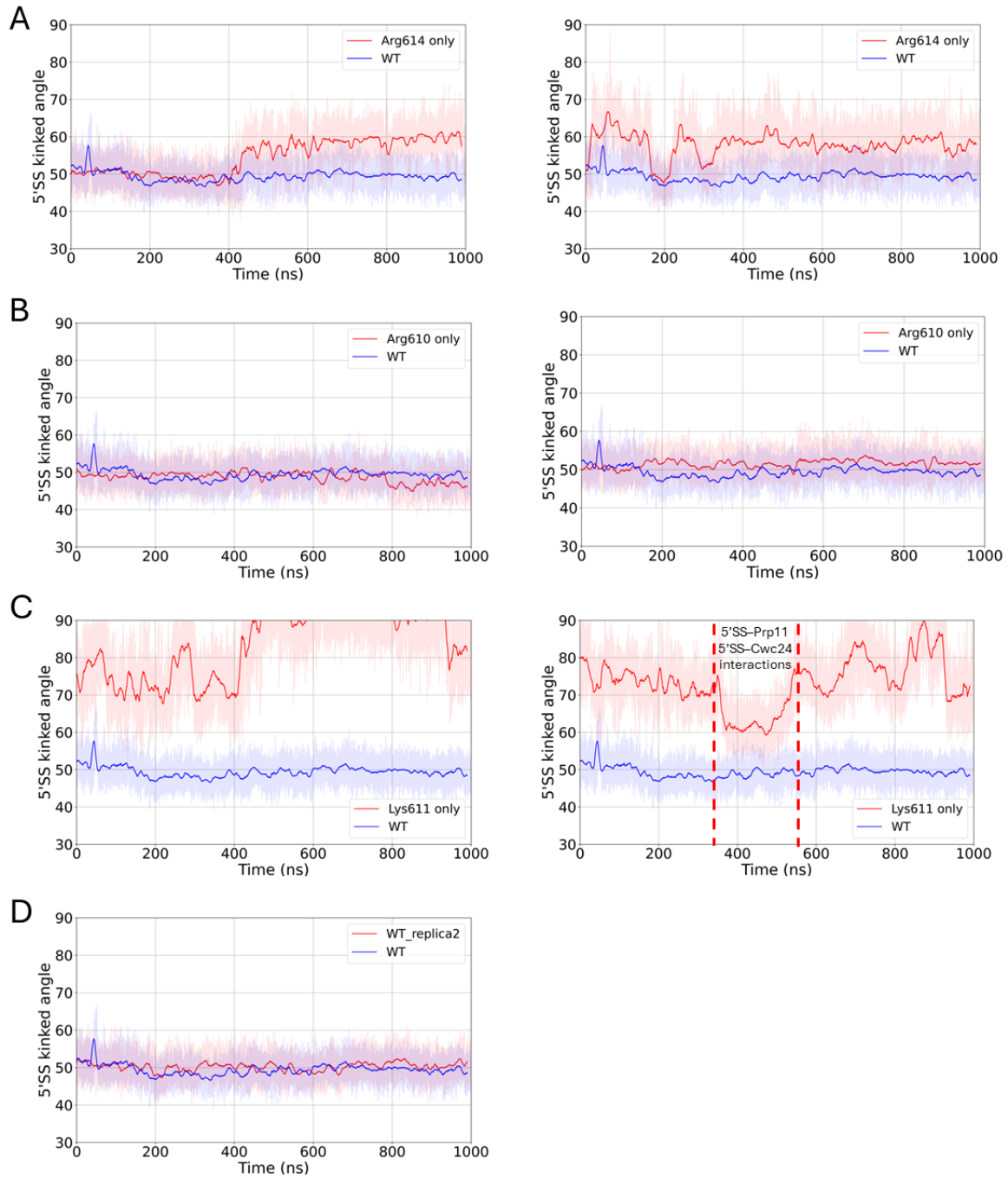


Figure S5. 5'SS angle fluctuations in *B_{ACT}* wild type and *Prp8* double mutants. Each plot shows two lines: the red line represents the 5'SS angle values of the mutant system throughout the trajectory, while the blue line corresponds to the first replicate of the wild type, used as a reference for a stable kinked conformation. **A)** Arg610Ala / Lys611Ala double mutant: The kinked conformation is largely preserved, with a slightly elevated average angle ($54.6 \pm 5.4^\circ$) compared to wild type ($49.4 \pm 2.7^\circ$). **B)** Lys611Ala / Arg614Ala double mutant: Arg610 alone is sufficient to maintain a wild-type-like angle in both replicas. **C)** Arg610Ala / Arg614Ala double mutant: The kinked conformation is destabilized, though a transient re-stabilization between 380–580 ns (red dashed lines) coincides with temporary reformation of Prp11, Cwc24, and 5'SS interactions. **D)** Wild type: Both trajectories behave similarly, confirming the stable kinked conformation observed in the model.

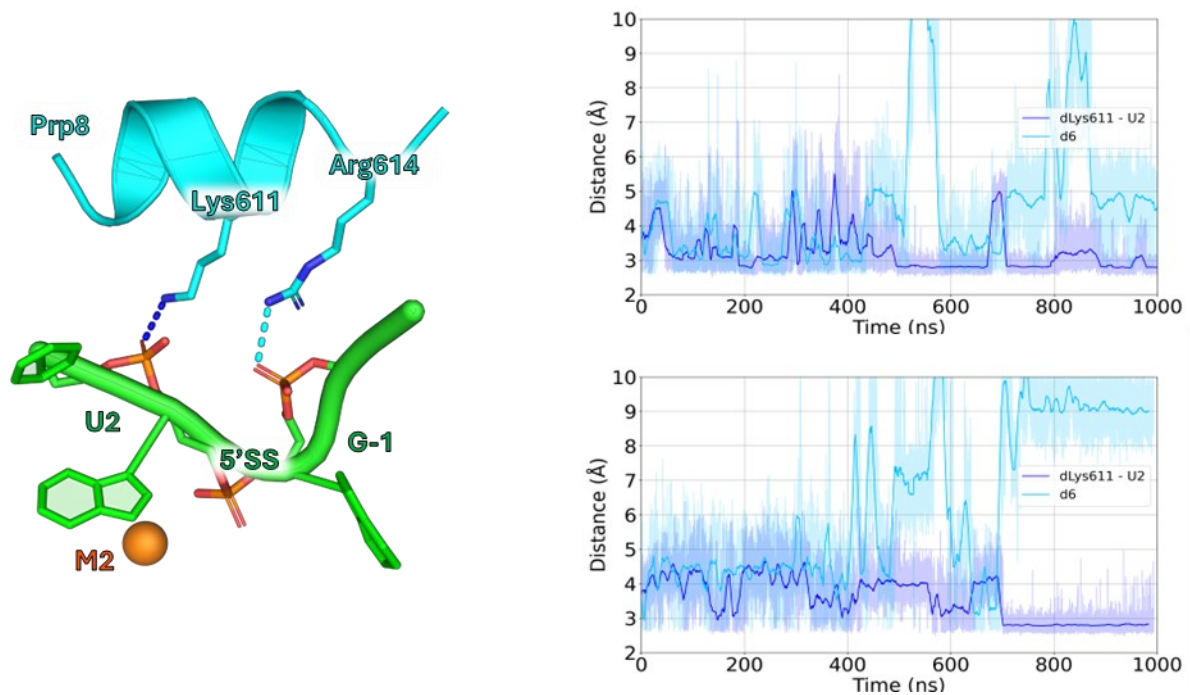


Figure S6. Arg614 and Lys611 partially compensate for the Arg610Ala mutation by stabilizing the 5'SS at the B_{ACT} stage.

Left: Stabilizing interactions between Lys611 of Prp8 (cyan licorice) and U2 of the pre-mRNA (green licorice), and between Arg614 (cyan licorice) and G-1 (green licorice). Donor-acceptor distances—blue dashed for Lys611–U2 and cyan dashed for Arg614–G-1 (distance d6)—were monitored during simulations. **B:** Plots show the time evolution of these distances in two 1- μ s replicates. Interactions are considered maintained when distances remain below 3.5 Å.

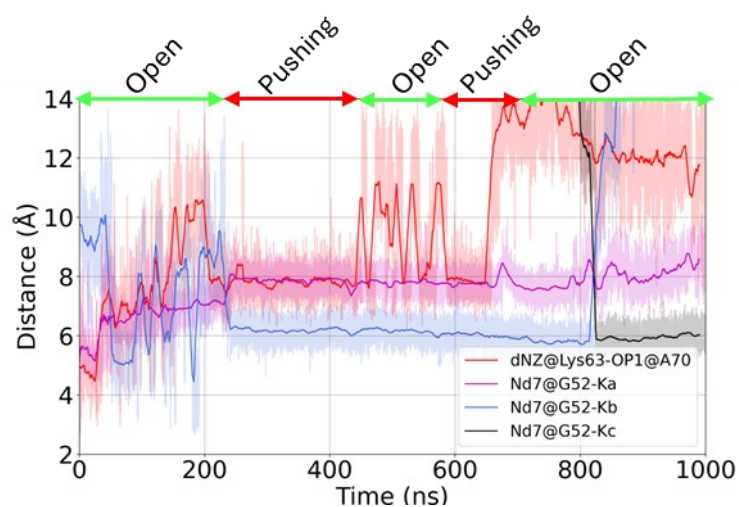


Figure S7. Yju2 gating dynamics and recruitment of potassium ions at the K1 site in the B* complex. Plot showing the distances over time in the second replicate of metadynamics simulations between Lys63-A70 (red), G52-K⁺_a (purple), and G52-K⁺_b (blue). K⁺_a, K⁺_b, and K⁺_c are potassium ions recruited through the ion channel depicted in Figure S1. The plot captures transitions between different gate states.

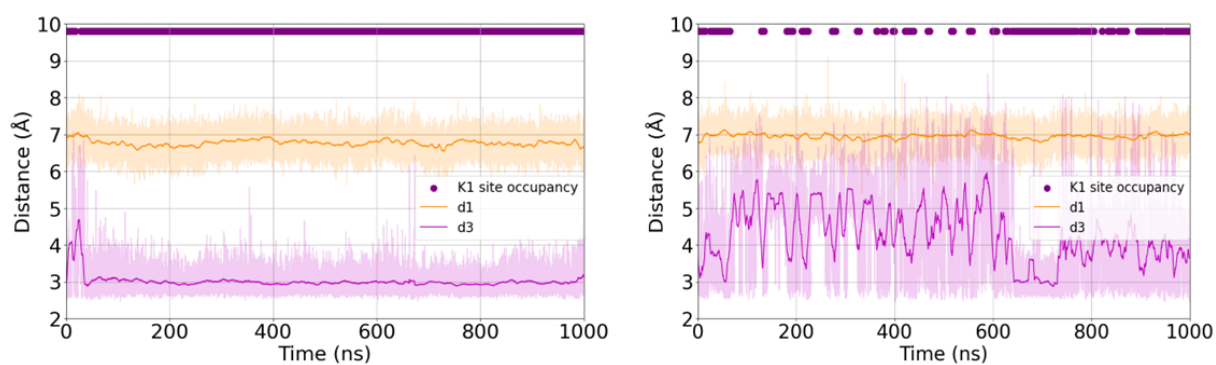


Figure S8. MD simulations of B^* with M2 modeled in its catalytic position. Changes in distances d1 (orange) and d3 (purple) are tracked throughout the trajectory. Purple dots indicate K^+ occupancy at the K1 site ($d_{K^+-N7 \text{ of } G52} < 3.5 \text{ \AA}$). Compared to the B^* active site with M2 in the cryo-EM-like position, K^+ transiently approaches (right panel) or stably occupies (left panel) the K1 site.

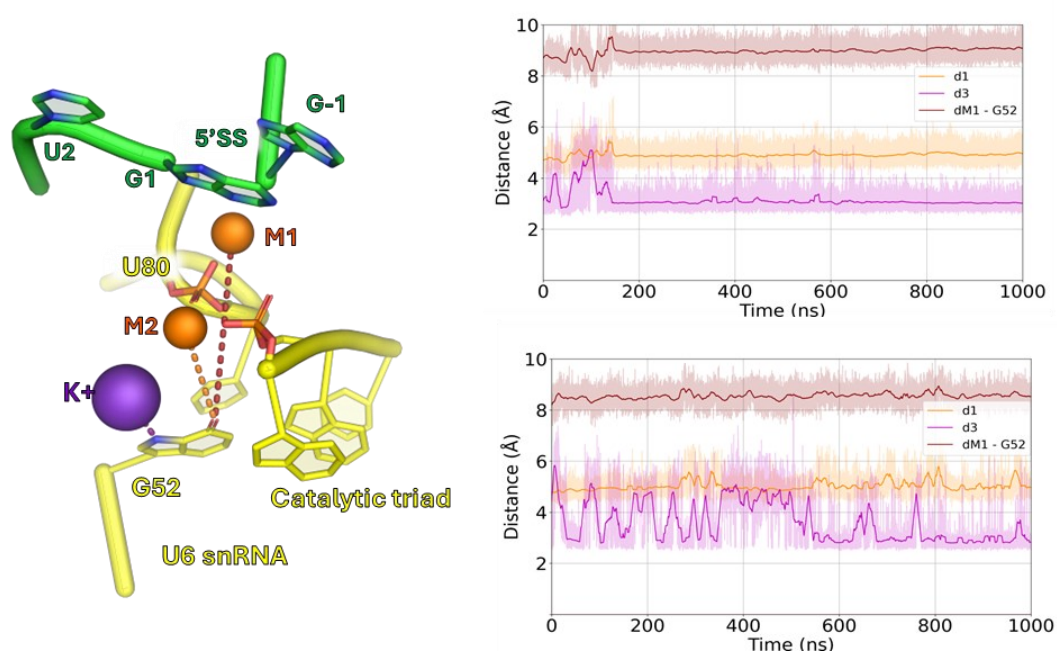


Figure S9. MD simulations of the C complex. Left: Key distances monitored during simulations are shown – yellow dashed line for M1 to O6 of U6-G52 (d1), orange dashed line for M2 to O6 of G52, and purple dashed line for K⁺ to N7 of G52 (d3). Right: Time evolution of these distances over two independent 1- μ s replicates. All critical structural and catalytic elements within the active site remain stable throughout the trajectories.

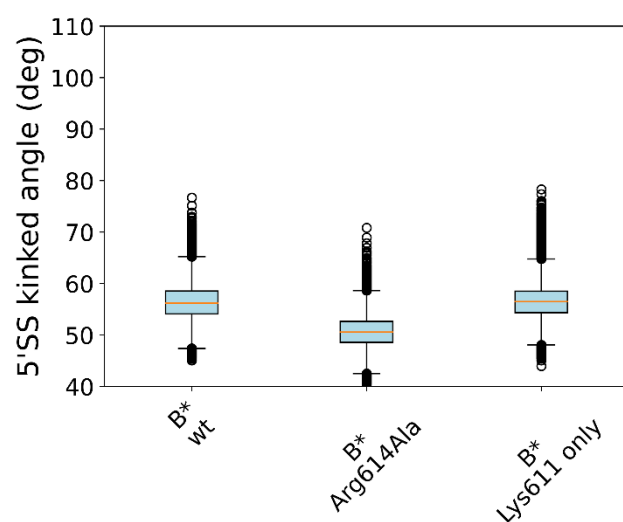


Figure S10. 5'SS angle variations and stability in the B* complex of wild type and mutants. Boxplot displaying the distribution of 5'SS angle values from the second replicate of each simulated B* system. The wild-type angle serves as a reference for the stable kinked conformation.

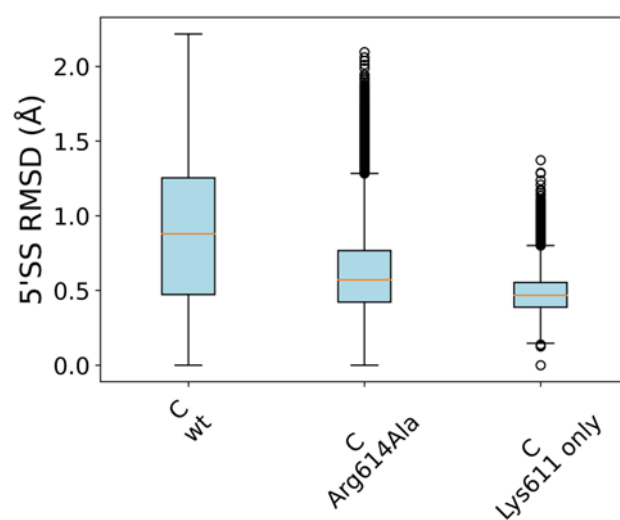


Figure S11. 5'SS stability in the C complex of wild type and mutants. Boxplot showing the distribution of 5'SS RMSD values from the second replicate of each simulated C complex system, reflecting 5'SS stability. RMSD was calculated over the three nucleotides forming the 5'SS throughout the trajectory. Wild-type RMSD serves as a reference for the stable kinked conformation.

Bibliography

1. On Protein Synthesis - Francis Crick - Profiles in Science.
<https://profiles.nlm.nih.gov/spotlight/sc/catalog/nlm:nlmuid-101584582X404-doc>.
2. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190, 576–581 (1961).
3. Gros, F. *et al.* Unstable ribonucleic acid revealed by pulse labelling of *Escherichia coli*. *Nature* 190, 581–585 (1961).
4. Nomura, M., Hall, B., Biology, S. S.-J. of M. & 1960, undefined. Characterization of RNA synthesized in *Escherichia coli* after bacteriophage T2 infection. *ElsevierM Nomura, BD Hall, S SpiegelmanJournal of Molecular Biology*, 1960•Elsevier
<https://www.sciencedirect.com/science/article/pii/S0022283660800277>.
5. Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* 8, 413–423 (2007).
6. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799 (2007).
7. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854 (1993).
8. Zieve, G., Cell, S. P.- & 1976, undefined. Small RNA species of the HeLa cell: metabolism and subcellular localization. *cell.comG Zieve, S PenmanCell*, 1976•*cell.com*
[https://www.cell.com/cell/pdf/0092-8674\(76\)90181-1.pdf](https://www.cell.com/cell/pdf/0092-8674(76)90181-1.pdf).
9. Benecke, B., Cell, S. P.- & 1977, undefined. A new class of small nuclear RNA molecules synthesized by a type I RNA polymerase in HeLa cells. *cell.comBJ Benecke, S PenmanCell*, 1977•*cell.com* [https://www.cell.com/cell/pdf/0092-8674\(77\)90158-1.pdf](https://www.cell.com/cell/pdf/0092-8674(77)90158-1.pdf).
10. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. Are snRNPs involved in splicing? [26]. *Nature* 283, 220–224 (1980).
11. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the H19 gene may function as an RNA. *Mol Cell Biol* 10, 28 (1990).
12. Y., O., M., F. & T., K. FANTOM CONSORTIUM; RIKEN GENOME EXPLORATION RESEARCH GROUP PHASE I & II TEAM. ANALYSIS OF THE MOUSE TRANSCRIPTOME BASED ON FUNCTIONAL ANNOTATION OF 60,770 FULL-LENGTH CDNAS. 420, 563–573 (2002).
13. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* 8, 776 (2007).
14. Mou, X., Liew, S. W. & Kwok, C. K. Identification and targeting of G-quadruplex structures in MALAT1 long non-coding RNA. *Nucleic Acids Res* 50, 397–410 (2022).
15. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nature Reviews Genetics* 2014 15:6 15, 423–437 (2014).
16. Goodall, G. J. & Wickramasinghe, V. O. RNA in cancer. *Nat Rev Cancer* 21, 22–36 (2021).

17. Wang, W., Van Niekerk, E., Willis, D. E. & Twiss, J. L. RNA transport and localized protein synthesis in neurological disorders and neural repair. *Dev Neurobiol* 67, 1166–1182 (2007).
18. Fu, X. D. Non-coding RNA: A new frontier in regulatory biology. *Natl Sci Rev* 1, 190–204 (2014).
19. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854 (1993).
20. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862 (1993).
21. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. MiRBase: From microRNA sequences to function. *Nucleic Acids Res* 47, D155–D162 (2019).
22. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology* 2014 15:8 15, 509–524 (2014).
23. Im, H. I. & Kenny, P. J. MicroRNAs in neuronal function and dysfunction. *Trends Neurosci* 35, 325–334 (2012).
24. Benhamou, R. I. *et al.* DNA-encoded library versus RNA-encoded library selection enables design of an oncogenic noncoding RNA inhibitor. *Proc Natl Acad Sci U S A* 119, e2114971119 (2022).
25. Tadesse, K. & Benhamou, R. I. Targeting MicroRNAs with Small Molecules. *Noncoding RNA* 10, 17 (2024).
26. Kavita, K. & Breaker, R. R. Discovering riboswitches: the past and the future. *Trends Biochem Sci* 48, 119 (2022).
27. Ellinger, E. *et al.* Riboswitches as therapeutic targets: Promise of a new era of antibiotics. *Expert Opin Ther Targets* 27, 433 (2023).
28. Mulhbach, J. *et al.* Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog* 6, 1–11 (2010).
29. Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotechnol* 24, 1558–1564 (2006).
30. Howe, J. A. *et al.* Selective small-molecule inhibition of an RNA structural element. *Nature* 2015 526:7575 526, 672–677 (2015).
31. Connelly, C. M. *et al.* Synthetic ligands for PreQ1 riboswitches provide structural and mechanistic insights into targeting RNA tertiary structure. *Nature Communications* 2019 10:1 10, 1–12 (2019).
32. Ford, C. W., Zurenko, G. E. & Barbachyn, M. R. The discovery of linezolid, the first oxazolidinone antibacterial agent. *Curr Drug Targets Infect Disord* 1, 181–199 (2001).
33. Rybak, J. M. & Roberts, K. Tedizolid Phosphate: a Next-Generation Oxazolidinone. *Infect Dis Ther* 4, 1–14 (2015).

34. Panchal, V. & Brenk, R. Riboswitches as Drug Targets for Antibiotics. *Antibiotics* 2021, Vol. 10, Page 45 10, 45 (2021).
35. Reddy, R. & Busch, H. Small Nuclear RNAs: RNA Sequences, Structure, and Modifications. *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles* 1–37 (1988) doi:10.1007/978-3-642-73020-7_1.
36. Will, C. L. & Lührmann, R. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* 3, a003707 (2011).
37. Shuai, S. *et al.* The U1 spliceosomal RNA is recurrently mutated in multiple cancers. *Nature* 2019 574:7780 574, 712–716 (2019).
38. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* 31, 147–157 (1982).
39. 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* 35, 849–857 (1983).
40. Micura, R. & Höbartner, C. Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chem Soc Rev* 49, 7331–7353 (2020).
41. Kazantsev, A. V. *et al.* Crystal structure of a bacterial ribonuclease P RNA. *Proc Natl Acad Sci U S A* 102, 13392–13397 (2005).
42. Rupert, P. B. & Ferré-D'Amaré, A. R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* 410, 780–786 (2001).
43. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science* 231, 1577–1580 (1986).
44. Martick, M. & Scott, W. G. Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell* 126, 309–320 (2006).
45. Ke, A., Zhou, K., Ding, F., Cate, J. H. D. & Doudna, J. A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* 429, 201–205 (2004).
46. Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* 395, 567–574 (1998).
47. Klein, D. J. & Ferré-D'Amaré, A. R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* 313, 1752–1756 (2006).
48. Lilley, D. M. J. The Varkud satellite ribozyme. *RNA* 10, 151–158 (2004).
49. Weinberg, Z. *et al.* New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat Chem Biol* 11, 606–610 (2015).
50. Roth, A. *et al.* A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol* 10, 56–60 (2014).
51. Walker, S. C. & Engelke, D. R. Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit Rev Biochem Mol Biol* 41, 77–102 (2006).
52. Torres-Larios, A., Swinger, K. K., Pan, T. & Mondragón, A. Structure of ribonuclease P - a universal ribozyme. *Curr Opin Struct Biol* 16, 327–335 (2006).

53. Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* 418, 222–228 (2002).
54. Bhatta, A. & Hillen, H. S. Structural and mechanistic basis of RNA processing by protein-only ribonuclease P enzymes. *Trends Biochem Sci* 47, 965–977 (2022).
55. Sidharthan, V. *et al.* Use of a small molecule microarray screen to identify inhibitors of the catalytic RNA subunit of *Methanobrevibacter smithii* RNase P. *Nucleic Acids Res* 53, gkae1190 (2024).
56. Schencking, I. *et al.* RNase P Inhibitors Identified as Aggregators. *Antimicrob Agents Chemother* 65, e00300-21 (2021).
57. Chamberlain, A. R., Huynh, L., Huang, W., Taylor, D. J. & Harris, M. E. The specificity landscape of bacterial ribonuclease P. *J Biol Chem* 300, (2024).
58. Janzen, E. *et al.* Promiscuous Ribozymes and Their Proposed Role in Prebiotic Evolution. *Chem Rev* 120, 4879–4897 (2020).
59. Kellerman, D. L., York, D. M., Piccirilli, J. A. & Harris, M. E. Altered (transition) states: Mechanisms of solution and enzyme catalyzed RNA 2'-O-transphosphorylation. *Curr Opin Chem Biol* 21, 96–102 (2014).
60. Ren, A., Micura, R. & Patel, D. J. Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr Opin Chem Biol* 41, 71–83 (2017).
61. Seith, D. D., Bingaman, J. L., Veenis, A. J., Button, A. C. & Bevilacqua, P. C. Elucidation of Catalytic Strategies of Small Nucleolytic Ribozymes From Comparative Analysis of Active Sites. *ACS Catal* 8, 314 (2017).
62. Messina, K. J., Kierzek, R., Tracey, M. A. & Bevilacqua, P. C. Small Molecule Rescue and Glycosidic Conformational Analysis of the Twister Ribozyme. *Biochemistry* 58, 4857–4868 (2019).
63. Zhang, S. *et al.* Role of the Active Site Guanine in the glmS Ribozyme Self-Cleavage Mechanism: Quantum Mechanical/Molecular Mechanical Free Energy Simulations. *J Am Chem Soc* 137, 784–798 (2015).
64. Joseph, N. N., Roy, R. N. & Steitz, T. A. Molecular dynamics analysis of Mg²⁺-dependent cleavage of a pistol ribozyme reveals a fail-safe secondary ion for catalysis. *J Comput Chem* 41, 1345–1352 (2020).
65. Tiyaboonchai, A. *et al.* Self-cleaving guide RNAs enable pharmacological selection of precise gene editing events in vivo. *Nat Commun* 13, 1–12 (2022).
66. Smathers, C. M. & Robart, A. R. The mechanism of splicing as told by group II introns: Ancestors of the spliceosome. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1862, 194390 (2019).
67. Zhao, C. & Pyle, A. M. Structural insights into the mechanism of group II intron splicing. *Trends Biochem Sci* 42, 470 (2017).
68. Boudvillain, 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* 406, 315–318 (2000).

69. Lambowitz, A. M. & Zimmerly, S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* 3, 1–19 (2011).
70. Daniels, D. L., Michels, W. J. & Pyle, A. M. Two competing pathways for self-splicing by group II introns: A quantitative analysis of in vitro reaction rates and products. *J Mol Biol* 256, 31–49 (1996).
71. Podar, M., Chu, V. T., Pyle, A. M. & Perlman, P. S. Group II intron splicing in vivo by first-step hydrolysis. *Nature* 391, 915–918 (1998).
72. Costa, M., Walbott, H., Monachello, D., Westhof, E. & Michel, F. Crystal structures of a group II intron lariat primed for reverse splicing. *Science* 354, (2016).
73. Qu, G. *et al.* Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol* 23, 549–557 (2016).
74. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* 514, 193–197 (2014).
75. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* 320, 77–82 (2008).
76. Chan, R. T. *et al.* Structural basis for the second step of group II intron splicing. *Nat Commun* 9, (2018).
77. 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* 19, 555–557 (2012).
78. Toor, N., Rajashankar, K., Keating, K. S. & Pyle, A. M. Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol* 15, 1221 (2008).
79. 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* 21, 464–471 (2014).
80. Fica, S. M. *et al.* RNA catalyzes nuclear pre-mRNA splicing. *Nature* 503, 229 (2013).
81. Manigrasso, J. *et al.* Visualizing group II intron dynamics between the first and second steps of splicing. *Nat Commun* 11, 1–15 (2020).
82. 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* 6, 1509 (2000).
83. Podar, M., Perlman, P. S. & Padgett, R. A. Stereochemical selectivity of group II intron splicing, reverse splicing, and hydrolysis reactions. *Mol Cell Biol* 15, 4466–4478 (1995).
84. Truong, D. M., Hewitt, F. C., Hanson, J. H., Cui, X. & Lambowitz, A. M. Retrohoming of a Mobile Group II Intron in Human Cells Suggests How Eukaryotes Limit Group II Intron Proliferation. *PLoS Genet* 11, e1005422 (2015).
85. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* 493, 638–643 (2013).
86. Dlakić, M. & Mushegian, A. Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA* 17, 799–808 (2011).

87. Zimmerly, S. & Semper, C. Evolution of group II introns. *Mob DNA* 6, 1–19 (2015).
88. Chillón, I. & Marcia, M. Self-Splicing Group II Introns. *Ribozymes: Volume 1 and Volume 2* 1–2, 143–167 (2021).
89. Fedorova, O. *et al.* Small molecules that target group II introns are potent antifungal agents. *Nat Chem Biol* 14, 1073 (2018).
90. Silvestri, I. *et al.* Targeting the conserved active site of splicing machines with specific and selective small molecule modulators. *Nat Commun* 15, 1–15 (2024).
91. Matera, A. G., Terns, R. M. & Terns, M. P. Non-coding RNAs: Lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 8, 209–220 (2007).
92. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and Disease. *Cell* 136, 777–793 (2009).
93. Bai, R., Wan, R., Yan, C., Lei, J. & Shi, Y. Structures of the fully assembled *Saccharomyces cerevisiae* spliceosome before activation. *Science* (1979) 360, 1423–1429 (2018).
94. Wan, R., Bai, R., Yan, C., Lei, J. & Shi, Y. Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching. *Cell* 177, 339–351.e13 (2019).
95. Bai, R. *et al.* Mechanism of spliceosome remodeling by the ATPase/helicase Prp2 and its coactivator Spp2. *Science* 371, (2021).
96. Wilkinson, M. E., Fica, S. M., Galej, W. P. & Nagai, K. Structural basis for conformational equilibrium of the catalytic spliceosome. *Mol Cell* 81, 1439–1452.e9 (2021).
97. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* (1979) 353, 895–904 (2016).
98. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu Rev Biochem* 89, 359–388 (2020).
99. Wan, R. *et al.* The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* 351, 466–475 (2016).
100. Wachtel, C. & Manley, J. L. Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. *Mol Biosyst* 5, 311–316 (2009).
101. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* (1979) 353, 904–912 (2016).
102. Rauhut, R. *et al.* Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* (1979) 353, 1399–1405 (2016).
103. Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* (1979) 349, 1182–1191 (2015).
104. Fica, S. M. *et al.* Structure of a spliceosome remodelled for exon ligation. *Nature* 542, 377–380 (2017).
105. Grainger, R. J. & Beggs, J. D. Prp8 protein: at the heart of the spliceosome. *RNA* 11, 533–557 (2005).

106. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* 18, 655–670 (2017).
107. Jia, J. *et al.* A Snu114–GTP–Prp8 module forms a relay station for efficient splicing in yeast. *Nucleic Acids Res* 48, 4572 (2020).
108. Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* 530, 298–302 (2016).
109. Urlaub, H., Raker, V. A., Kostka, S. & Lührmann, R. Sm protein–Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J* 20, 187 (2001).
110. de Almeida, R. A. & O’Keefe, R. T. The NineTeen Complex (NTC) and NTC-associated proteins as targets for spliceosomal ATPase action during pre-mRNA splicing. *RNA Biol* 12, 109–114 (2015).
111. Chen, H.-C., Tseng, C.-K., Tsai, R.-T., Chung, C.-S. & Cheng, S.-C. Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. *Mol Cell Biol* 33, 514–525 (2013).
112. Hogg, R., McGrail, J. C. & O’Keefe, R. T. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem Soc Trans* 38, 1110–1115 (2010).
113. Cordin, O. & Beggs, J. D. RNA helicases in splicing. *RNA Biol* 10, 83 (2013).
114. Liu, Y. C. & Cheng, S. C. Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. *J Biomed Sci* 22, 1–9 (2015).
115. Zhan, X., Lu, Y. & Shi, Y. Molecular basis for the activation of human spliceosome. *Nat Commun* 15, (2024).
116. Marcia, M. & Pyle, A. M. Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* 151, 497–507 (2012).
117. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* 90, 6498–6502 (1993).
118. Casalino, L., Palermo, G., Rothlisberger, U. & Magistrato, A. Who Activates the Nucleophile in Ribozyme Catalysis? An Answer from the Splicing Mechanism of Group II Introns. *J Am Chem Soc* 138, 10374–10377 (2016).
119. Aupič, J., Borišek, J., Fica, S. M., Galej, W. P. & Magistrato, A. Monovalent metal ion binding promotes the first transesterification reaction in the spliceosome. *Nat Commun* 14, 8482 (2023).
120. Borišek, J. & Magistrato, A. All-Atom Simulations Decrypt the Molecular Terms of RNA Catalysis in the Exon-Ligation Step of the Spliceosome. *ACS Catal* 10, 5328–5334 (2020).
121. Lee, S. C. W. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat Med* 22, 976–986 (2016).
122. Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med* 18, 472 (2012).

123. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat Rev Genet* 17, 19–32 (2016).
124. Szelest, M. & Giannopoulos, K. Biological relevance of alternative splicing in hematologic malignancies. *Molecular Medicine* 30, 1–14 (2024).
125. Sakai, Y. *et al.* GEX1 Compounds, Novel Antitumor Antibiotics Related to Herboxidiene, Produced by *Streptomyces* sp. II. The Effects on Cell Cycle Progression and Gene Expression. *J Antibiot (Tokyo)* 55, 863–872 (2002).
126. Motoyoshi, H. *et al.* Structure-activity Relationship for FR901464: A Versatile Method for the Conversion and Preparation of Biologically Active Biotinylated Probes. *Biosci Biotechnol Biochem* 68, 2178–2182 (2004).
127. Eskens, F. A. L. M. *et al.* Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors. *Clin Cancer Res* 19, 6296–6304 (2013).
128. Steensma, D. P. *et al.* Results of a Clinical Trial of H3B-8800, a Splicing Modulator, in Patients with Myelodysplastic Syndromes (MDS), Acute Myeloid Leukemia (AML) or Chronic Myelomonocytic Leukemia (CMML). *Blood* 134, 673–673 (2019).
129. Campagne, S. *et al.* Structural basis of a small molecule targeting RNA for a specific splicing correction. *Nat Chem Biol* 15, 1191–1198 (2019).
130. Schneider-Poetsch, T., Chhipi-Shrestha, J. K. & Yoshida, M. Splicing modulators: on the way from nature to clinic. *Journal of Antibiotics* 74, 603–616 (2021).
131. Pearn, J. Classification of spinal muscular atrophies. *Lancet* 1, 919–922 (1980).
132. Aslesh, T. & Yokota, T. Restoring SMN Expression: An Overview of the Therapeutic Developments for the Treatment of Spinal Muscular Atrophy. *Cells* 11, 417 (2022).
133. Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F. & Krainer, A. R. Antisense Masking of an hnRNP A1/A2 Intronic Splicing Silencer Corrects SMN2 Splicing in Transgenic Mice. *Am J Hum Genet* 82, 834–848 (2008).
134. Singh, N. N., Howell, M. D., Androphy, E. J. & Singh, R. N. How the discovery of ISS-N1 led to the first medical therapy for spinal muscular atrophy. *Gene Ther* 24, 520–526 (2017).
135. Mendell, J. R. *et al.* Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *New England Journal of Medicine* 377, 1713–1722 (2017).
136. Chand, D. *et al.* Hepatotoxicity following administration of onasemnogene abeparvovec (AVXS-101) for the treatment of spinal muscular atrophy. *J Hepatol* 74, 560–566 (2021).
137. Naryshkin, N. A. *et al.* Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* 345, 688–693 (2014).
138. Woll, M. G. *et al.* Discovery and Optimization of Small Molecule Splicing Modifiers of Survival Motor Neuron 2 as a Treatment for Spinal Muscular Atrophy. *J Med Chem* 59, 6070–6085 (2016).

139. Ratni, H. *et al.* Specific Correction of Alternative Survival Motor Neuron 2 Splicing by Small Molecules: Discovery of a Potential Novel Medicine To Treat Spinal Muscular Atrophy. *J Med Chem* 59, 6086–6100 (2016).
140. Charnas, L. *et al.* Safety and efficacy findings in the first-in-human trial (FIH) of the oral splice modulator branaplam in type 1 spinal muscular atrophy (SMA): interim results. *Neuromuscular Disorders* 27, S207–S208 (2017).
141. Chaytow, H., Faller, K. M. E., Huang, Y. T. & Gillingwater, T. H. Spinal muscular atrophy: From approved therapies to future therapeutic targets for personalized medicine. *Cell Rep Med* 2, 100346 (2021).
142. White, D. S., Dunyak, B. M., Vaillancourt, F. H. & Hoskins, A. A. A sequential binding mechanism for 5' splice site recognition and modulation for the human U1 snRNP. *Nature Communications* 15, 1–16 (2024).
143. Ratni, H. *et al.* 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* 61, 6501–6517 (2018).
144. Dhillon, S. Risdiplam: First Approval. *Drugs* 80, 1853–1858 (2020).
145. Roca, X. *et al.* Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev* 26, 1098–1109 (2012).
146. Rogalska, M. E., Vivori, C. & Valcárcel, J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat Rev Genet* 24, 251–269 (2023).
147. Martin, L., Hutchens, M., Hawkins, C. & Radnov, A. How much do clinical trials cost? *Nat Rev Drug Discov* 16, 381–382 (2017).
148. Macalino, S. J. Y., Gosu, V., Hong, S. & Choi, S. Role of computer-aided drug design in modern drug discovery. *Arch Pharm Res* 38, 1686–1701 (2015).
149. Lambert, G. K., Duhme-Klair, A. K., Morgan, T. & Ramjee, M. K. The background, discovery and clinical development of BCR-ABL inhibitors. *Drug Discov Today* 18, 992–1000 (2013).
150. Elliott, M. Zanamivir: from drug design to the clinic. *Philosophical Transactions of the Royal Society of London. Series B* 356, 1885 (2001).
151. Subeha, M. R. & Telleria, C. M. The Anti-Cancer Properties of the HIV Protease Inhibitor Nelfinavir. *Cancers* 2020, Vol. 12, Page 3437 12, 3437 (2020).
152. Jorgensen, W. L. The many roles of computation in drug discovery. *Science* 303, 1813–1818 (2004).
153. Childs-Disney, J. L. *et al.* Targeting RNA structures with small molecules. *Nat Rev Drug Discov* 21, 736–762 (2022).
154. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* 17, 547–558 (2018).
155. Aboul-Ela, F. Strategies for the design of RNA-binding small molecules. *Future Med Chem* 2, 93–119 (2010).

156. Manigrasso, J., Marcia, M. & De Vivo, M. Computer-aided design of RNA-targeted small molecules: A growing need in drug discovery. *Chem* 7, 2965–2988 (2021).
157. Ding, F. *et al.* Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *rnajournal.cshlp.org* F Ding, S Sharma, P Chalasani, VV Demidov, NE Broude, NV Dokholyan *Rna*, 2008 • *rnajournal.cshlp.org* <https://doi.org/10.1261/rna.894608> doi:10.1261/rna.894608.
158. Jonikas, M. A. *et al.* Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *rnajournal.cshlp.org* MA Jonikas, RJ Radmer, A Laederach, R Das, S Pearlman, D Herschlag, RB Altman *Rna*, 2009 • *rnajournal.cshlp.org* <https://doi.org/10.1261/rna.1270809> doi:10.1261/rna.1270809.
159. Boniecki, M. J. *et al.* SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *academic.oup.com* MJ Boniecki, G Lach, WK Dawson, K Tomala, P Lukasz, T Soltysinski, KM Rother *Nucleic acids research*, 2016 • *academic.oup.com* 44, 63 (2015).
160. Miao, Z., biophysics, E. W.-A. review of & 2017, undefined. RNA structure: advances and assessment of 3D structure prediction. *annualreviews.org* 46, 483–503 (2017).
161. Stelzer, A., Frank, A., Kratz, J., ... M. S.-N. chemical & 2011, undefined. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *nature.com* AC Stelzer, AT Frank, JD Kratz, MD Swanson, MJ Gonzalez-Hernandez, J Lee, I Andricioaei *Nature chemical biology*, 2011 • *nature.com* <https://www.nature.com/articles/nchembio.596>.
162. Disney, M. D. *et al.* Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem Biol* 11, 1720–1728 (2016).
163. Jorgensen, W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc Chem Res* 22, 184–189 (2002).
164. Kumar, A. & Vashisth, H. Alchemical free energy to probe RNA/ligand interactions. *Biophys J* 122, 217a (2023).
165. Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann Phys* 389, 457–484 (1927).
166. Tuckerman, M. E. & Martyna, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. *Journal of Physical Chemistry B* 104, 159–178 (1999).
167. Mazur, A. K. Common molecular dynamics algorithms revisited: Accuracy and optimal time steps of Störmer-Leapfrog integrators. *J Comput Phys* 136, 354–365 (1997).
168. Muscat, S., Martino, G., Manigrasso, J., Marcia, M. & De Vivo, M. On the Power and Challenges of Atomistic Molecular Dynamics to Investigate RNA Molecules. *J Chem Theory Comput* 16, 18 (2024).
169. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* 99, 12562–12566 (2002).

170. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett* 100, 020603 (2008).
171. Muscat, S., Grasso, G., Scapozza, L. & Danani, A. In silico investigation of cytochrome bc1 molecular inhibition mechanism against *Trypanosoma cruzi*. *PLoS Negl Trop Dis* 17, (2023).
172. De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* 59, 4035–4061 (2016).
173. Fischer, N. M., Polëto, M. D., Steuer, J. & Van Der Spoel, D. Influence of Na⁺ and Mg²⁺ ions on RNA structures studied with molecular dynamics simulations. *Nucleic Acids Res* 46, 4872 (2018).
174. Marcia, M. & Pyle, A. M. Principles of ion recognition in RNA: insights from the group II intron structures. *RNA* 20, 516–527 (2014).
175. Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* vol. 482 322–330 Preprint at <https://doi.org/10.1038/nature10885> (2012).
176. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* vol. 20 474–489 Preprint at <https://doi.org/10.1038/s41580-019-0136-0> (2019).
177. Huarte, M. The emerging role of lncRNAs in cancer. *Nat Med* 21, 1253–1261 (2015).
178. Yamamoto, H., Unbehaun, A. & Spahn, C. M. T. Ribosomal Chamber Music: Toward an Understanding of IRES Mechanisms. *Trends Biochem Sci* 42, 655–668 (2017).
179. Chillón, I. & Marcia, M. The molecular structure of long non-coding RNAs: emerging patterns and functional implications. *Crit Rev Biochem Mol Biol* 55, 662–690 (2020).
180. Breaker, R. R. Riboswitches and Translation Control. *Cold Spring Harb Perspect Biol* 10, (2018).
181. Martin, W. J., Grandi, P. & Marcia, M. Screening strategies for identifying RNA- and ribonucleoprotein-targeted compounds. *Trends Pharmacol Sci* 42, 758–771 (2021).
182. Deriu, M. A. *et al.* Elucidating the role of surface chemistry on cationic phosphorus dendrimer-siRNA complexation. *Nanoscale* 10, 10952–10962 (2018).
183. Grasso, G. *et al.* Free energy landscape of siRNA-polycation complexation: Elucidating the effect of molecular geometry, polymer flexibility, and charge neutralization. *PLoS One* 12, e0186816 (2017).
184. Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Hierarchy of RNA functional dynamics. *Annu Rev Biochem* 83, 441–466 (2014).
185. Trollmann, M. F. W. & Böckmann, R. A. mRNA lipid nanoparticle phase transition. *Biophys J* 121, 3927–3939 (2022).
186. Vangaveti, S., Ranganathan, S. V. & Chen, A. A. Advances in RNA molecular dynamics: a simulator’s guide to RNA force fields. *Wiley Interdiscip Rev RNA* 8, (2017).

187. Richardson, J. S. *et al.* RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* 14, 465–481 (2008).
188. Sokoloski, J. E., Godfrey, S. A., Dombrowski, S. E. & Bevilacqua, P. C. Prevalence of syn nucleobases in the active sites of functional RNAs. *RNA* 17, 1775–1787 (2011).
189. Leulliot, N. & Varani, G. Current Topics in RNA–Protein Recognition: Control of Specificity and Biological Function through Induced Fit and Conformational Capture†. *Biochemistry* 40, 7947–7956 (2001).
190. Huang, M., Giese, T. J., Lee, T. S. & York, D. M. Improvement of DNA and RNA Sugar Pucker Profiles from Semiempirical Quantum Methods. *J Chem Theory Comput* 10, 1538–1545 (2014).
191. Zhang, N., Zhang, S. & Szostak, J. W. Activated Ribonucleotides Undergo a Sugar Pucker Switch upon Binding to a Single-Stranded RNA Template. *J Am Chem Soc* 134, 3691–3694 (2012).
192. Julien, K. R., Sumita, M., Chen, P. H., Laird-Offringa, I. A. & Hoogstraten, C. G. Conformationally restricted nucleotides as a probe of structure–function relationships in RNA. *RNA* 14, 1632–1643 (2008).
193. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* 7, 499–512 (2001).
194. Sponer, J. *et al.* RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem Rev* 118, 4177–4338 (2018).
195. Schnabl, J. & Sigel, R. K. Controlling ribozyme activity by metal ions. *Curr Opin Chem Biol* 14, 269–275 (2010).
196. Gallardo, A., Bogart, B. M. & Dutagaci, B. Protein–Nucleic Acid Interactions for RNA Polymerase II Elongation Factors by Molecular Dynamics Simulations. *J Chem Inf Model* 62, 3079–3089 (2022).
197. Banáš, P. *et al.* Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J Chem Theory Comput* 6, 3836–3849 (2010).
198. Bergonzo, C. & Cheatham, T. E. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J Chem Theory Comput* 11, 3969–3972 (2015).
199. 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* 4, (2018).
200. Cheatham, T. E., Cieplak, P. & Kollman, P. A. A Modified Version of the Cornell *et al.* Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J Biomol Struct Dyn* 16, 845–862 (1999).
201. Várnai, P. & Zakrzewska, K. DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res* 32, 4269–4280 (2004).

202. Pérez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys J* 92, 3817–3829 (2007).
203. Zgarbová, M. *et al.* Refinement of the Cornell *et al.* Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput* 7, 2886–2902 (2011).
204. 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* 6, 1520–1531 (2010).
205. Kührová, P. *et al.* Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J Chem Theory Comput* 12, 4534–4548 (2016).
206. Yoo, J., Winogradoff, D. & Aksimentiev, A. Molecular dynamics simulations of DNA–DNA and DNA–protein interactions. *Curr Opin Struct Biol* 64, 88–96 (2020).
207. Yoo, J. & Aksimentiev, A. Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J Chem Theory Comput* 12, 430–443 (2015).
208. Šponer, J. *et al.* How to understand atomistic molecular dynamics simulations of RNA and protein–RNA complexes? *Wiley Interdiscip Rev RNA* 8, e1405 (2017).
209. Chen, A. A. & García, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc Natl Acad Sci U S A* 110, 16820–16825 (2013).
210. Janeček, M. *et al.* W-RESP: Well-Restrained Electrostatic Potential-Derived Charges. Revisiting the Charge Derivation Model. *J Chem Theory Comput* 17, 3495–3509 (2021).
211. Tan, D., Piana, S., Dirks, R. M. & Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proceedings of the National Academy of Sciences* 115, E1346–E1355 (2018).
212. Steinbrecher, T., Latzer, J. & Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J Chem Theory Comput* 8, 4405–4412 (2012).
213. Gil-Ley, A., Bottaro, S. & Bussi, G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J Chem Theory Comput* 12, 2790–2798 (2016).
214. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different Approach. *Journal of Physical Chemistry Letters* 5, 3863–3871 (2014).
215. Mlynsky, V. *et al.* Fine-Tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides. *J Chem Theory Comput* 16, 3936–3946 (2020).
216. Mlynsky, V. *et al.* Reactive conformation of the active site in the hairpin ribozyme achieved by molecular dynamics simulations with ϵ/ζ Force Field Reparametrizations. *Journal of Physical Chemistry B* 119, 4220–4229 (2015).
217. Kruse, H. *et al.* Quantum Chemical Benchmark Study on 46 RNA Backbone Families Using a Dinucleotide Unit. *J Chem Theory Comput* 11, 4972–4991 (2015).

218. Jorgensen, W. L. & Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences* 102, 6665–6670 (2005).
219. Horn, H. W. *et al.* Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 120, 9665–9678 (2004).
220. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* 115, E4758–E4766 (2018).
221. Kührová, P. *et al.* Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J Chem Theory Comput* 15, 3288–3305 (2019).
222. MacKerell, A. D., Wiórkiewicz-Kuczera, J. & Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J Am Chem Soc* 117, 11946–11975 (2002).
223. Hart, K. *et al.* Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J Chem Theory Comput* 8, 348–362 (2011).
224. Lemkul, J. A., Huang, J., Roux, B. & Mackerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem Rev* 116, 4983–5013 (2016).
225. Dans, P. D., Walther, J., Gómez, H. & Orozco, M. Multiscale simulation of DNA. *Curr Opin Struct Biol* 37, 29–45 (2016).
226. Robertson, M. J., Qian, Y., Robinson, M. C., Tirado-Rives, J. & Jorgensen, W. L. Development and Testing of the OPLS-AA/M Force Field for RNA. *J Chem Theory Comput* 15, 2734–2742 (2019).
227. Ganguly, A. *et al.* Confluence of theory and experiment reveals the catalytic mechanism of the Varkud satellite ribozyme. *Nat Chem* 12, 193–201 (2020).
228. Petrov, A. S. *et al.* RNA–Magnesium–Protein Interactions in Large Ribosomal Subunit. *Journal of Physical Chemistry B* 116, 8113–8120 (2012).
229. Lee, T. S., Radak, B. K., Harris, M. E. & York, D. M. A Two-Metal-Ion-Mediated Conformational Switching Pathway for HDV Ribozyme Activation. *ACS Catal* 6, 1853–1869 (2016).
230. Padroni, G., Patwardhan, N. N., Schapira, M. & Hargrove, A. E. Systematic analysis of the interactions driving small molecule–RNA recognition. *RSC Med Chem* 11, 802–813 (2020).
231. Garmann, R. F. *et al.* Single-particle studies of the effects of RNA–protein interactions on the self-assembly of RNA virus particles. *Proceedings of the National Academy of Sciences* 119, e2206292119 (2022).
232. Ganser, L. R. *et al.* High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. *Nat Struct Mol Biol* 25, 425–434 (2018).

233. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* 441, 1167–1171 (2006).
234. Lipfert, J., Doniach, S., Das, R. & Herschlag, D. Understanding nucleic acid-ion interactions. *Annu Rev Biochem* 83, 813–841 (2014).
235. Allnér, O., Nilsson, L. & Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J Chem Theory Comput* 8, 1493–1502 (2012).
236. 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* 13, 340–352 (2016).
237. Hu, G. & Zhou, H. X. Magnesium ions mediate ligand binding and conformational transition of the SAM/SAH riboswitch. *Commun Biol* 6, 1–10 (2023).
238. Kognole, A. A. & MacKerell, A. D. Mg²⁺ Impacts the Twister Ribozyme through Push-Pull Stabilization of Nonsequential Phosphate Pairs. *Biophys J* 118, 1424–1437 (2020).
239. Xu, X. *et al.* Insights into xanthine riboswitch structure and metal ion-mediated ligand recognition. *Nucleic Acids Res* 49, 7139–7153 (2021).
240. Hu, G. & Zhou, H. X. Binding free energy decomposition and multiple unbinding paths of buried ligands in a PreQ1 riboswitch. *PLoS Comput Biol* 17, e1009603 (2021).
241. Ward, W. L., Plakos, K. & Deroose, V. J. Nucleic Acid Catalysis: Metals, Nucleobases, and Other Cofactors. *Chem Rev* 114, 4318–4342 (2014).
242. Manigrasso, J., De Vivo, M. & Palermo, G. Controlled Trafficking of Multiple and Diverse Cations Prompts Nucleic Acid Hydrolysis. *ACS Catal* 11, 8786–8797 (2021).
243. 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. *Journal of Physical Chemistry B* 119, 15460–15470 (2015).
244. Lipfert, J., Doniach, S., Das, R. & Herschlag, D. Understanding nucleic acid-ion interactions. *Annu Rev Biochem* 83, 813–841 (2014).
245. Cunha, R. A. & Bussi, G. Unraveling Mg²⁺–RNA binding with atomistic molecular dynamics. *RNA* 23, 628–638 (2017).
246. Allnér, O., Nilsson, L. & Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J Chem Theory Comput* 8, 1493–1502 (2012).
247. Weickhmann, A. K. *et al.* The structure of the SAM/SAH-binding riboswitch. *Nucleic Acids Res* 47, 2654–2665 (2019).
248. 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. *Journal of Physical Chemistry B* 113, 13279–13290 (2009).
249. Gonzalez, J. & Tinoco I., J. Identification Characterization of Metal Ion Binding Sites in RNA. *Methods Enzymol* 338, 421–443 (2002).

250. Suk Joung, I. *et al.* Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *Journal of Physical Chemistry B* 112, 9020–9041 (2008).
251. Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *Journal of Physical Chemistry* 94, 8021–8024 (1990).
252. Samara, N. L. & Yang, W. Cation trafficking propels RNA hydrolysis. *Nat Struct Mol Biol* 25, 715–721 (2018).
253. Li, P. & Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J Chem Theory Comput* 10, 289–297 (2013).
254. 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* 20, 474–489 (2019).
255. Mlýnský, V. *et al.* Toward Convergence in Folding Simulations of RNA Tetraloops: Comparison of Enhanced Sampling Techniques and Effects of Force Field Modifications. *J Chem Theory Comput* 18, 2642–2656 (2022).
256. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J Chem Theory Comput* 9, 2733–2748 (2013).
257. 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* 37, 3670–3679 (2009).
258. Halder, A., Kumar, S., Valsson, O. & Reddy, G. Mg²⁺ Sensing by an RNA Fragment: Role of Mg²⁺-Coordinated Water Molecules. *J Chem Theory Comput* 16, 6702–6715 (2020).
259. Miner, J. C., Chen, A. A. & García, A. E. Free-energy landscape of a hyperstable RNA tetraloop. *Proceedings of the National Academy of Sciences* 113, 6665–6670 (2016).
260. Wang, Y., Parmar, S., Schneekloth, J. S. & Tiwary, P. Interrogating RNA–Small Molecule Interactions with Structure Probing and Artificial Intelligence-Augmented Molecular Simulations. *ACS Cent Sci* 8, 741–748 (2022).
261. Hummer, G. & Köfinger, J. Bayesian ensemble refinement by replica simulations and reweighting. *J Chem Phys* 143, (2015).
262. Bernetti, M., Hall, K. B. & Bussi, G. Reweighting of molecular simulations with explicit-solvent SAXS restraints elucidates ion-dependent RNA ensembles. *Nucleic Acids Res* 49, e84–e84 (2021).
263. Cesari, A., Gil-Ley, A. & Bussi, G. Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement. *J Chem Theory Comput* 12, 6192–6200 (2016).
264. Rangan, R. *et al.* Determination of Structural Ensembles of Proteins: Restraining vs Reweighting. *J Chem Theory Comput* 14, 6632–6641 (2018).

265. Hoff, S. E., Thomasen, F. E., Lindorff-Larsen, K. & Bonomi, M. Accurate model and ensemble refinement using cryo-electron microscopy maps and Bayesian inference. *PLoS Comput Biol* 20, e1012180 (2024).
266. Bonomi, M., Camilloni, C., Cavalli, A. & Vendruscolo, M. Metainference: A Bayesian inference method for heterogeneous systems. *Sci Adv* 2, (2016).
267. White, A. D. & Voth, G. A. Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data. *J Chem Theory Comput* 10, 3023–3030 (2014).
268. Posani, E. *et al.* Ensemble Refinement of mismodeled cryo-EM RNA Structures Using All-Atom Simulations. *bioRxiv* 2024.07.24.604258 (2024) doi:10.1101/2024.07.24.604258.
269. Russel, D. *et al.* Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biol* 10, e1001244 (2012).
270. Bonomi, M. & Camilloni, C. Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* 33, 3999–4000 (2017).
271. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun* 185, 604–613 (2014).
272. Cantara, W. A., Olson, E. D. & Musier-Forsyth, K. Analysis of RNA structure using small-angle X-ray scattering. *Methods* 113, 46–55 (2017).
273. Chen, Y. & Pollack, L. SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip Rev RNA* 7, 512–526 (2016).
274. Koch, M. H. J., Vachette, P. & Svergun, D. I. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* 36, 147–227 (2003).
275. Lipfert, J., Chu, V. B., Bai, Y., Herschlag, D. & Doniach, S. Low-resolution models for nucleic acids from small-angle X-ray scattering with applications to electrostatic modeling. *urn:issn:0021-8898* 40, s229–s234 (2007).
276. Ballabio, F. *et al.* Accurate and Efficient SAXS/SANS Implementation Including Solvation Layer Effects Suitable for Molecular Simulations. *J Chem Theory Comput* 19, 8401–8413 (2023).
277. Hub, J. S. Interpreting solution X-ray scattering data using molecular simulations. *Curr Opin Struct Biol* 49, 18–26 (2018).
278. Bernetti, M. & Bussi, G. Comparing state-of-the-art approaches to back-calculate SAXS spectra from atomistic molecular dynamics simulations. *The European Physical Journal B* 2021 94:9 94, 1–8 (2021).
279. Liebl, K. & Zacharias, M. The development of nucleic acids force fields: From an unchallenged past to a competitive future. *Biophys J* 122, 2841–2851 (2023).
280. Rinaldi, S., Moroni, E., Rozza, R. & Magistrato, A. Frontiers and Challenges of Computing ncRNAs Biogenesis, Function and Modulation. *J Chem Theory Comput* 20, 993–1018 (2024).

281. Fischer, M., Coleman, R. G., Fraser, J. S. & Shoichet, B. K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem* 6, 575–583 (2014).
282. Dans, P. D. *et al.* Modeling, Simulations, and Bioinformatics at the Service of RNA Structure. *Chem* 5, 51–73 (2019).
283. Mlynsky, V. *et al.* Reactive conformation of the active site in the hairpin ribozyme achieved by molecular dynamics simulations with ϵ/ζ Force Field Reparametrizations. *Journal of Physical Chemistry B* 119, 4220–4229 (2015).
284. Dans, P. D. *et al.* Modeling, Simulations, and Bioinformatics at the Service of RNA Structure. *Chem* 5, 51–73 (2019).
285. Trabuco, L. G. *et al.* The Role of L1 Stalk–tRNA Interaction in the Ribosome Elongation Cycle. *J Mol Biol* 402, 741–760 (2010).
286. Jiang, W. & Chen, L. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Comput Struct Biotechnol J* 19, 183–195 (2021).
287. Pistoni, M., Ghigna, C. & Gabellini, D. Alternative splicing and muscular dystrophy. *RNA Biol* 7, 441 (2010).
288. 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. *Chem Rev* 118, 4156–4176 (2018).
289. Fica, S. M. & Nagai, K. Cryo-electron microscopy snapshots of the spliceosome: Structural insights into a dynamic ribonucleoprotein machine. *Nat Struct Mol Biol* 24, 791–799 (2017).
290. Zhang, L., Vielle, A., Espinosa, S. & Zhao, R. RNAs in the spliceosome: Insight from cryoEM structures. *Wiley Interdiscip Rev RNA* 10, (2019).
291. 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* 61, 1 (2021).
292. Liu, Y.-C., Chen, H.-C., Wu, N.-Y. & Cheng, S.-C. A Novel Splicing Factor, Yju2, Is Associated with NTC and Acts after Prp2 in Promoting the First Catalytic Reaction of Pre-mRNA Splicing. *Mol Cell Biol* 27, 5403–5413 (2007).
293. Chiu, Y.-F. *et al.* Cwc25 Is a Novel Splicing Factor Required after Prp2 and Yju2 To Facilitate the First Catalytic Reaction. *Mol Cell Biol* 29, 5671 (2009).
294. 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* 26, 40–50.e2 (2018).
295. Huang, Q. *et al.* Ginsenoside Rb2 inhibits p300-mediated SF3A2 acetylation at lysine 10 to promote Fscn1 alternative splicing against myocardial ischemic/reperfusion injury. *J Adv Res* 65, 365–379 (2024).

296. Bao, Y. *et al.* Unbiased interrogation of functional lysine residues in human proteome. *Mol Cell* 83, 4614–4632.e6 (2023).
297. Bao, Y. *et al.* Unbiased interrogation of functional lysine residues in human proteome. *Mol Cell* 83, 4614–4632.e6 (2023).
298. Huang, Q. *et al.* Ginsenoside Rb2 inhibits p300-mediated SF3A2 acetylation at lysine 10 to promote Fscn1 alternative splicing against myocardial ischemic/reperfusion injury. *J Adv Res* <https://doi.org/10.1016/j.jare.2023.12.012> (2023) doi:10.1016/j.jare.2023.12.012.
299. Viadiu, H. & Aggarwal, A. K. The role of metals in catalysis by the restriction endonuclease bamhi. *Nat Struct Biol* 5, 910–916 (1998).
300. Zhang, J., McCabe, K. A. & Bella, C. E. Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. *Proc Natl Acad Sci U S A* 108, 11872–11877 (2011).
301. Zhang, J., Pan, X. & Bell, C. E. Crystal structure of λ exonuclease in complex with DNA and Ca(2+). *Biochemistry* 53, 7415–7425 (2014).
302. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y. J. & Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature* 487, 196–201 (2012).
303. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* (1979) 343, (2014).
304. Sun, J., Viadiu, H., Aggarwal, A. K. & Weinstein, H. Energetic and Structural Considerations for the Mechanism of Protein Sliding along DNA in the Nonspecific BamHI-DNA Complex. *Biophys J* 84, 3317–3325 (2003).
305. Zhang, J., McCabe, K. A. & Bella, C. E. Crystal structures of λ exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. *Proc Natl Acad Sci U S A* 108, 11872–11877 (2011).
306. Johnson, R. E., Trincao, J., Aggarwal, A. K., Prakash, S. & Prakash, L. Deoxynucleotide triphosphate binding mode conserved in Y family DNA polymerases. *Mol Cell Biol* 23, 3008–3012 (2003).
307. Manigrasso, J., De Vivo, M. & Palermo, G. Controlled Trafficking of Multiple and Diverse Cations Prompts Nucleic Acid Hydrolysis. *ACS Catal* 11, 8786–8797 (2021).
308. Marcia, M., Somarowthu, S. & Pyle, A. M. Now on display: A gallery of group II intron structures at different stages of catalysis. *Mobile DNA* vol. 4 Preprint at <https://doi.org/10.1186/1759-8753-4-14> (2013).
309. Samara, N. L. & Yang, W. Cation trafficking propels RNA hydrolysis. *Nat Struct Mol Biol* 25, 715 (2018).
310. Hardy, S. F., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Cofactor requirements of splicing of purified messenger RNA precursors. *Nature* 308, 375–377 (1984).
311. Tseng, C. K. & Cheng, S. C. Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science* (1979) 320, 1782–1784 (2008).

312. Tseng, C. K. & Cheng, S. C. The spliceosome catalyzes debranching in competition with reverse of the first chemical reaction. *RNA* 19, 971–981 (2013).
313. Xu, L., Liu, T., Chung, K. & Pyle, A. M. Structural insights into intron catalysis and dynamics during splicing. *Nature* 2023 624:7992 624, 682–688 (2023).
314. Xu, L., Liu, T., Chung, K. & Pyle, A. M. Structural insights into intron catalysis and dynamics during splicing. *Nature* 624, 682–688 (2023).
315. Haack, D. B., Rudolfs, B., Zhang, C., Lyumkis, D. & Toor, N. Structural basis of branching during RNA splicing. *Nat Struct Mol Biol* 31, 179–189 (2024).
316. Bai, R. *et al.* Mechanism of spliceosome remodeling by the ATPase/helicase Prp2 and its coactivator Spp2. *Science* 371, eabe8863–eabe8863 (2020).
317. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* (1979) 353, 904–912 (2016).
318. Wilkinson, M. E., Fica, S. M., Galej, W. P. & Nagai, K. Structural basis for conformational equilibrium of the catalytic spliceosome. *Mol Cell* 81, 1439–1452.e9 (2021).
319. Zhu, K. *et al.* Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. *Proteins: Structure, Function, and Bioinformatics* 82, 1646–1655 (2014).
320. Casalino, L., Palermo, G., Spinello, A., Rothlisberger, U. & Magistrato, A. All-atom simulations disentangle the functional dynamics underlying gene maturation in the intron lariat spliceosome. *Proc Natl Acad Sci U S A* 115, 6584–6589 (2018).
321. Borišek, J. & Magistrato, A. All-Atom Simulations Decrypt the Molecular Terms of RNA Catalysis in the Exon-Ligation Step of the Spliceosome. *ACS Catal* 10, 5328–5334 (2020).
322. Saltalamacchia, A. *et al.* Decrypting the Information Exchange Pathways across the Spliceosome Machinery. *J Am Chem Soc* 142, 8403–8411 (2020).
323. 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* 79, 926–935 (1983).
324. Pérez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys J* 92, 3817–3829 (2007).
325. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 11, 3696–3713 (2015).
326. 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* 112, 9020 (2008).
327. Allnér, O., Nilsson, L. & Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J Chem Theory Comput* 8, 1493–1502 (2012).
328. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25 (2015).

329. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* 126, (2007).
330. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52, 7182–7190 (1981).
331. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4, 116–122 (2008).
332. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98, 10089–10092 (1993).
333. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* 100, 020603 (2008).
334. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun* 185, 604–613 (2014).
335. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph* 14, 33–38 (1996).
336. Connelly, C. M. *et al.* Synthetic ligands for PreQ 1 riboswitches provide structural and mechanistic insights into targeting RNA tertiary structure. *Nat Commun* 10, 1–12 (2019).
337. Parmar, S. *et al.* Mechanistic analysis of Riboswitch Ligand interactions provides insights into pharmacological control over gene expression. *Nature Communications* 15, 1–13 (2024).
338. Rasouli, A., Pickard, F. C., Sur, S., Grossfield, A. & Işık Bennett, M. Essential Considerations for Free Energy Calculations of RNA–Small Molecule Complexes: Lessons from the Theophylline-Binding RNA Aptamer. *J Chem Inf Model* 65, 223–239 (2024).
339. Abramyan, A. M. *et al.* Accurate Physics-Based Prediction of Binding Affinities of RNA- and DNA-Targeting Ligands. *J Chem Inf Model* 65, 1392–1403 (2025).
340. Chen, W. *et al.* Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J Chem Theory Comput* 14, 6346–6358 (2018).
341. Falese, J. P., Donlic, A. & Hargrove, A. E. Targeting RNA with small molecules: from fundamental principles towards the clinic. *Chem Soc Rev* 50, 2224–2243 (2021).
342. Fullenkamp, C. R. *et al.* Machine learning-augmented molecular dynamics simulations (MD) reveal insights into the disconnect between affinity and activation of ZTP riboswitch ligands. *bioRxiv* <https://doi.org/10.1101/2024.09.13.612887> (2024) doi:10.1101/2024.09.13.612887.
343. Hoang, G. L. *et al.* Refining Ligand Poses in RNA/Ligand Complexes of Pharmaceutical Relevance: A Perspective by QM/MM Simulations and NMR Measurements. *J Phys Chem Lett* 16, 1702–1708 (2025).
344. Damm, W. *et al.* OPLS5: Addition of Polarizability and Improved Treatment of Metals. <https://doi.org/10.26434/CHEMRXIV-2024-2SVR7> (2024) doi:10.26434/CHEMRXIV-2024-2SVR7.

345. Winkler, L. & Cheatham, T. E. Benchmarking the Drude Polarizable Force Field Using the r(GACC) Tetranucleotide. *J Chem Inf Model* 63, 2505–2511 (2023).
346. Sengul, M. Y. & Mackerell, A. D. Accurate modeling of RNA hairpins through the explicit treatment of electronic polarizability with the classical Drude oscillator force field. *Journal of computational biophysics and chemistry* 21, 461 (2022).
347. Lu, C. *et al.* OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J Chem Theory Comput* 17, 4291–4300 (2021).
348. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* 28, 235–242 (2000).
349. Panei, F. P., Torchet, R., Menager, H., Gkeka, P. & Bonomi, M. HARIBOSS: a curated database of RNA-small molecules structures to aid rational drug design. *Bioinformatics* 38, 4185–4193 (2022).
350. Roth, A. *et al.* A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol* 14, 308–317 (2007).
351. Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V. & Wedekind, J. E. Comparison of a PreQ1 Riboswitch Aptamer in Metabolite-bound and Free States with Implications for Gene Regulation. *Journal of Biological Chemistry* 286, 24626–24637 (2011).
352. Steinbrecher, T. B. *et al.* Accurate Binding Free Energy Predictions in Fragment Optimization. *J Chem Inf Model* 55, 2411–2420 (2015).
353. Shivakumar, D. *et al.* Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J Chem Theory Comput* 6, 1509–1519 (2010).
354. Wang, L. *et al.* Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J Am Chem Soc* 137, 2695–2703 (2015).
355. Spitale, R. C., Torelli, A. T., Krucinska, J., Bandarian, V. & Wedekind, J. E. The Structural Basis for Recognition of the PreQ0 Metabolite by an Unusually Small Riboswitch Aptamer Domain. *Journal of Biological Chemistry* 284, 11012–11016 (2009).
356. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y. J. & Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature* 487, 196–201 (2012).
357. Luo, J., Wei, W., Waldispühl, J. & Moitessier, N. Challenges and current status of computational methods for docking small molecules to nucleic acids. *Eur J Med Chem* 168, 414–425 (2019).