

**UNIVERSITY OF NAPLES FEDERICO II
SCHOOL OF MEDICINE
DEPARTMENT OF PHARMACY**



PhD programme in Pharmaceutical Sciences

***Biocomputing for the Discovery of RNA-based therapeutics and Targeting Agents
to Challenge Human Diseases***

**PhD student
Isidora Diakogiannaki**

**Tutor
Prof. Luciana Marinelli**

**Co-Tutor
Prof. Francesco Saverio Di Leva**

**Coordinator
Prof. ROSARIA MELI**

XVIII CYCLE (2022-2025)

PNRR Missione 4, Componente 2, Investimento 1.4

**“Potenziamento strutture di ricerca e
creazione di "campioni nazionali di R&S" su alcune Key Enabling
Technologies”**

Iniziativa finanziata dall'Unione europea — NextGenerationEU.

National Center for Gene Therapy and Drugs based on RNA Technology

Sviluppo di terapia genica e farmaci con tecnologia a RNA

Codice progetto MUR: **CN00000041**

CUP UNINA: **E63C22000940007**

Table of Contents

1. INTRODUCTION	1
1.1. The Role of RNA in Molecular Biology.....	1
1.1.1. RNA functions	1
1.1.2. RNA Structural Diversity	3
1.2. RNA-based Therapies.....	9
1.2.1. RNA as a Therapeutic Target	10
1.2.2. RNA as a Therapeutic Target for Small Molecules.....	13
1.3. Computational Approaches for Studying RNA Structure	14
1.4. Thesis Objectives	15
1.4.1. Targeting EphA2 in Glioblastoma Stem Cells Using RNA Aptamer	15
1.4.2. RNA Force Field (FF) Evaluation	19
1.4.3. Targeting Tubulin Associated Unit (tau) pre-mRNA Using Small Molecules.....	21
2. METHODOLOGIES	27
2.1. RNA Structure Prediction Techniques.....	27
2.1.1. RNA Secondary (2D) Structure Prediction.....	27
2.1.2. RNA tertiary (3D) Structure Prediction.....	31
2.2. Molecular Docking	34
2.2.1. Ligand and Target Preparation.....	35
2.2.2. Generation of Binding Poses	36

2.2.3.	Scoring	37
2.2.4.	Challenges and Recent Advances	42
2.2.5.	Applications of Molecular Docking.....	43
2.3.	Molecular Dynamics (MD) Simulations.....	44
2.3.1.	Molecular Simulation Methods Across Scales	44
2.3.2.	The potential Energy Function (force field)	46
2.3.3.	Energy Minimization	55
2.3.4.	MD Formalism.....	58
2.3.5.	Introduction to Statistical Mechanics.....	61
2.3.6.	Thermostats and Barostats	64
2.4.	Artificial Intelligence in Drug Discovery	67
2.4.1.	Introduction to Artificial Intelligence, Machine learning, and Deep Learning	67
2.4.2.	Overview of AI applications in drug discovery	69
2.4.3.	Current Challenges and Limitations	72
3.	RESULTS	84
3.1.	Targeting Glioblastoma Stem Cells (GSCs) via EphA2 using an RNA Aptamer.....	84
3.1.1.	2D Structure Prediction of A40s.....	85
3.1.2.	In Silico 3D Structure Prediction of A40s	88
3.1.3.	Molecular Dynamics of A40s in Solution	92
3.1.4.	Validation of MD-predicted A40s Structure using In-Line Probing Experimental Data	99

3.1.5.	EphA2/A40s Phosphorylation and Interaction	103
3.1.5.1.	EphA2/A40s Phosphorylation Using Western blot Data.....	103
3.1.5.2.	EphA2/A40s Binding Using Molecular Docking	104
3.1.5.3.	Elucidating EphA2/A40s and EphA2/ephrinA1 Interaction Using MD Simulations	107
3.2.	RNA Force Field Evaluation	114
3.2.1.	Selection of FFs	115
3.2.2.	Selection of RNA motifs.....	117
3.2.3.	MD Simulation Protocols	121
3.2.4.	MD Simulation Results and Analysis	122
3.3.	Targeting tau pre-mRNA with Small Molecules	149
3.3.1.	AI-based Similarity Screening of Ligand Library	150
3.3.2.	1D ¹ H NMR On Selected 15-ligand Library.....	153
3.3.3.	WaterLOGSY On Selected 15-ligand Library.....	157
3.3.4.	Biological Screening on 6 Ligands Derived from NMR Experiments	162
3.3.5.	In-silico Analysis of Tau-pre-mRNA/IP025.....	164
4.	DISCUSSION.....	170
5.	CONCLUSIONS	184

ABSTRACT

RNA plays a central role in molecular biology, participating in a wide range of cellular functions. Owing to its structural versatility, RNA can adopt diverse motifs, enabling specific molecular recognition and interactions with proteins, DNA, small molecules, and ions. As a large fraction of the human genome is transcribed into non-coding RNAs, RNA represents a promising target for therapeutic intervention.

In this thesis, RNA-based therapeutics were investigated from three complementary perspectives, highlighting the potential of RNA in drug discovery. First, RNA was studied as a ligand for the EphA2 receptor, a target implicated in glioblastoma multiforme stem cells. The structure of the A40s RNA aptamer and its interaction with EphA2 were characterized, and the resulting model provides key insights for the rational optimization of A40s variants targeting EphA2-positive glioblastoma. In the second project, we evaluated the state-of-the-art force fields OL3, OL3 cP-ghbfix21, Shaw, and DES-AMBER for Molecular Dynamics simulations across a diverse set of RNA motifs. This analysis revealed strengths and limitations in their ability to accurately capture RNA structure and dynamics, providing guidance for the development of improved RNA force fields. Finally, the tau RNA exon 10–intron hairpin was targeted with small molecules to restore its structural stability and modulate alternative splicing. Through AI-based virtual screening, NMR spectroscopy, and biological assays, a small molecule was identified that binds tau RNA and modulates the 3R:4R tau isoform ratio.

Overall, this work demonstrates the potential of RNA-focused drug discovery through the synergistic integration of computational and experimental approaches.

1. INTRODUCTION

1.1. The Role of RNA in Molecular Biology

1.1.1. RNA functions

Ribonucleic acid (RNA) is one of the most important and versatile chemical species in molecular biology, participating in several biological processes (**Figure 1.1**). Indeed, messenger RNA (mRNA) is an intermediary carrier of genetic information, and transfer RNA (tRNA) is the molecular adaptor used to translate this information into a protein sequence.^{1,2} The discovery of ribozymes has demonstrated that RNA can catalyze specific chemical reactions, achieving accelerations comparable to those of many protein enzyme.³ Ribosomes, RNA (rRNA)/protein complexes that have been characterized as an unusual large ribozyme, catalyze the translation process.⁴ Furthermore, small nucleolar RNA (snoRNA) process rRNA molecules, RNA switches (riboswitches) detect and respond to environmental or metabolic cues affecting gene expression accordingly.⁵ The discovery of a multitude of catalytic RNAs revolutionized modern biology, demonstrating the unique capacity of RNA to simultaneously store genetic information (as DNA genome) and catalyze chemical reactions (as protein enzymes). In eukaryotic and particularly multicellular organisms, RNA is involved in numerous processes essential for the maintenance, regulation, and processing of genetic information, such as RNA silencing.⁶ Therefore, RNA can be classified into protein-coding and noncoding RNA (ncRNA).⁷ Only 1-2% of the human genome encodes for proteins, corresponding to approximately 21,000 protein-coding genes,^{8,9} while the vast majority of RNA transcripts in human cells

(98%) is made up of non-coding regions, which include regulatory sequences, repetitive DNA sequences, and other functional elements.^{10,11} Consequently, because RNA plays multiple roles in cellular regulatory processes, understanding its function has significant biological impact. Notably, mutations in non-coding RNAs are involved in cancer and other diseases, highlighting the growing importance of RNA-focused research.

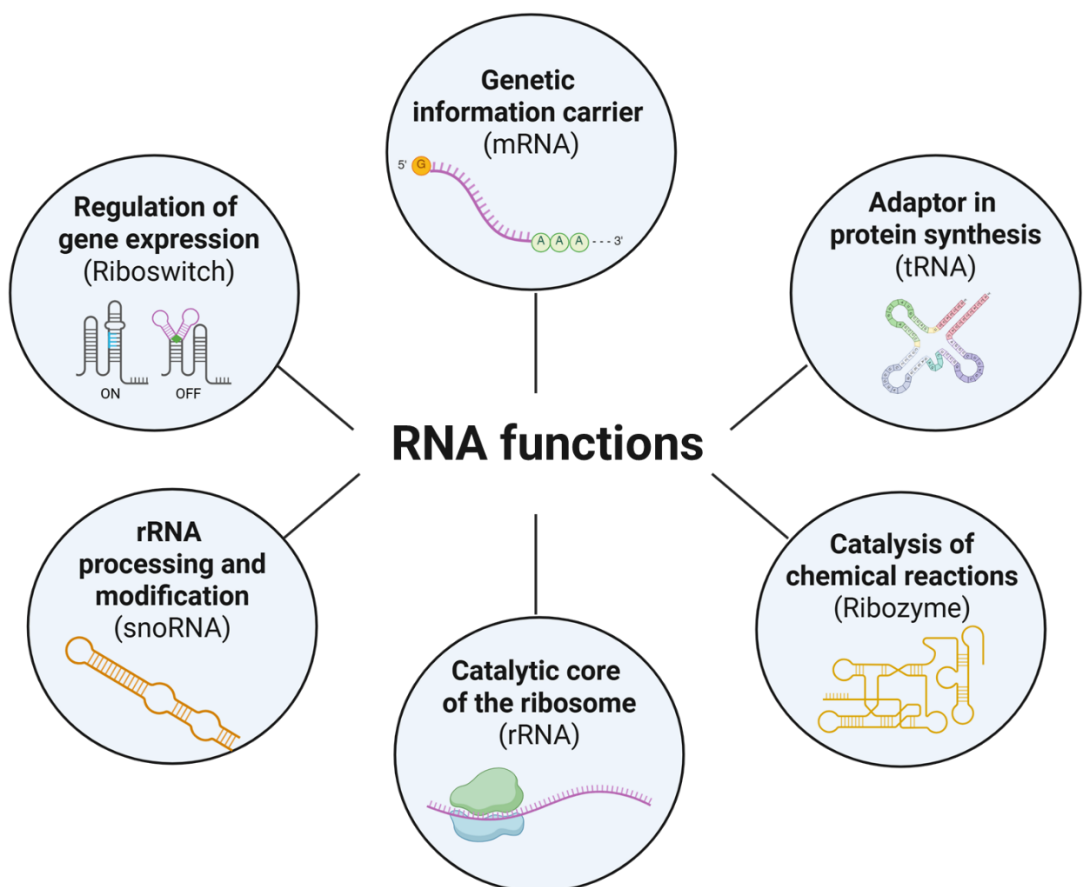


Figure 1.1. Scheme of RNA multiple functions.

1.1.2. RNA Structural Diversity

RNA is a linear polymer composed of four nucleobase (nt) monomers: adenine (A), uracil (U), guanine (G), and cytosine (C), each one consisting of a planar aromatic nucleobase attached to a ribose sugar, which is linked through a 5'-phosphate group (**Figure 1.2**). RNA strands are formed by phosphodiester bonds connecting the 3' carbon of one ribose to the 5' carbon of the next, creating a chemically polarized chain with free 5' and 3' termini. A key feature distinguishing RNA from DNA is the presence of the ribose 2'-hydroxyl (2'-OH) group. This substituent makes RNA less chemically stable than DNA by facilitating self-cleavage reactions. The 2'-OH group of the sugar acts both as bond donor and acceptor, thereby stabilizing non-canonical interactions and enabling the formation of complex conformational architectures that are central to RNA function.¹²

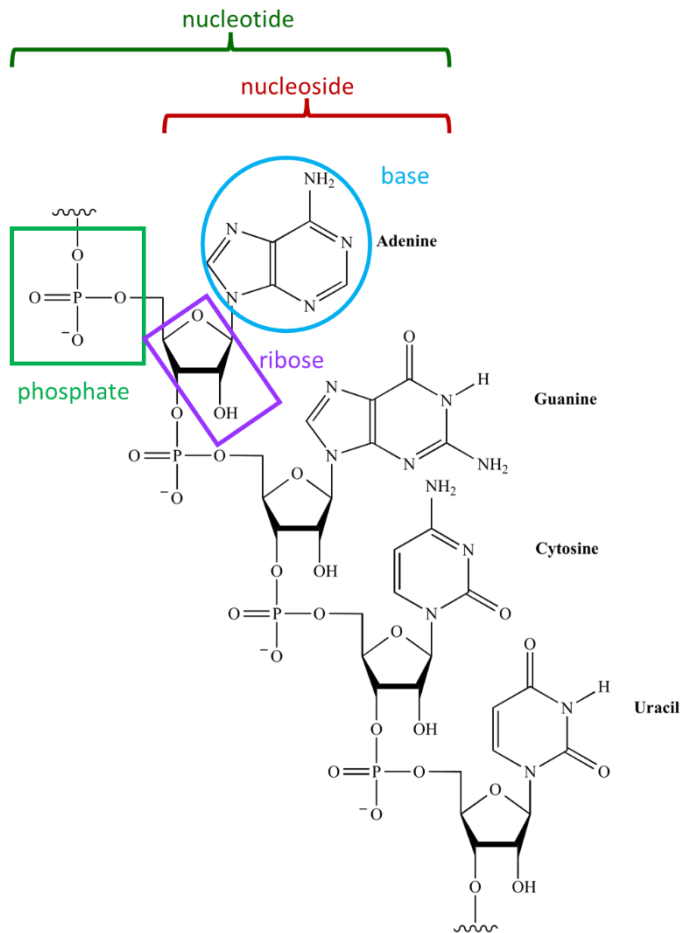


Figure 1.2. Schematic representation of RNA. The phosphate, ribose sugar, and base are marked by green, purple, and cyan squares, respectively. A nucleoside is referred to a base linked to a sugar, while a nucleotide is a nucleoside with a phosphate group attached.

Primary and Secondary Structure

RNA molecules can be described in a hierarchical manner by analyzing their primary, secondary (2D), and tertiary (3D) structures. Primary structure simply defines an RNA sequence in terms of its constituent nucleotides. 2D

structure is defined by the pattern of Watson–Crick (WC) canonical base pairs (A-U and C-G), which contribute approximately by 1-3 kcal/mol per base pair and form the basis of helical regions. In addition to the canonical base pairs, RNA exhibits a wide range of non-canonical base pairs. These interactions can be organized into 12 basic families according to the Leontis–Westhof classification,¹³ based on the three nucleoside edges participating in the formation (sugar edge, WC edge, and Hoogsteen edge) and assuming cis/trans orientations of the glycosidic bonds (**Figure 1.3, Table 1.1**).

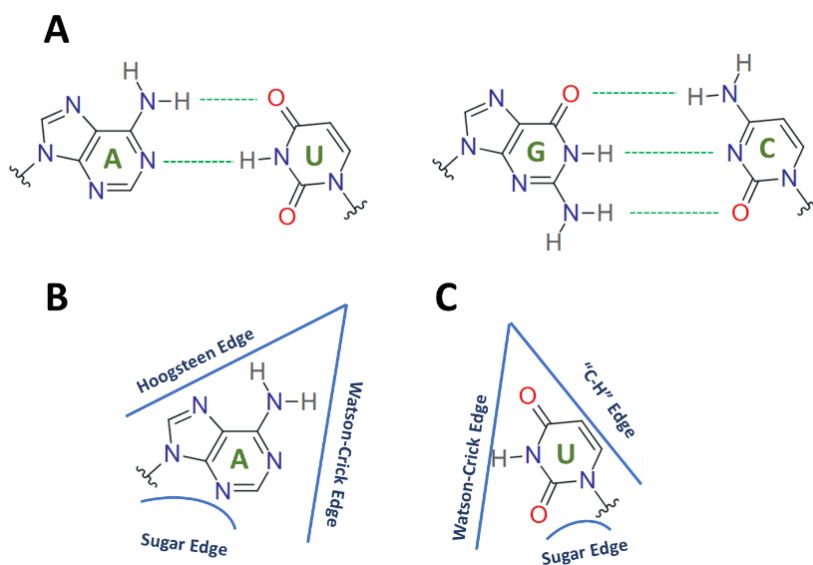


Figure 1.3. A) canonical base pairs (A-U, G-C), B, C) Interacting edges of nucleobases.

Table 1.1. The 12 basic families of non-canonical interactions according to Leontis-Westhof classification.¹²

Watson–Crick (W) / Watson–Crick (W)
1. cWW: cis Watson–Crick / Watson–Crick (canonical A–U, G–C)
2. tWW: trans Watson–Crick / Watson–Crick
Watson–Crick (W) / Hoogsteen (H)
3. cWH: cis Watson–Crick / Hoogsteen
4. tWH: trans Watson–Crick / Hoogsteen
Watson–Crick (W) / Sugar (S)
5. cWS: cis Watson–Crick / Sugar
6. tWS: trans Watson–Crick / Sugar
Hoogsteen (H) / Hoogsteen (H)
7. cHH: cis Hoogsteen / Hoogsteen
8. tHH: trans Hoogsteen / Hoogsteen
Hoogsteen (H) / Sugar (S)
9. cHS: cis Hoogsteen / Sugar
10. tHS: trans Hoogsteen / Sugar
Sugar (S) / Sugar (S)
11. cSS: cis Sugar / Sugar
12. tSS: trans Sugar / Sugar

This expanded pairing underlies the formation of a variety of key 2D substructures, including stems (continued stretches of base pairs), bulges (unpaired nt interrupting a stem), internal loops (unpaired segments between paired regions), hairpin loops (terminal loop regions formed by a single strand

folding back on itself), and multi-helix junctions (branching points from which three or more helices emanate). In addition to these canonical nested motifs, RNA can form pseudoknots, that result from canonical base pairing between a hairpin loop (or another 2D substructure) and a single distal complementary strand region (**Figure 1.4**).^{14,15} The 2D structure of RNA is usually represented using dot bracket notation, where paired bases are described by matching pairs of parentheses, and unpaired bases as dots.

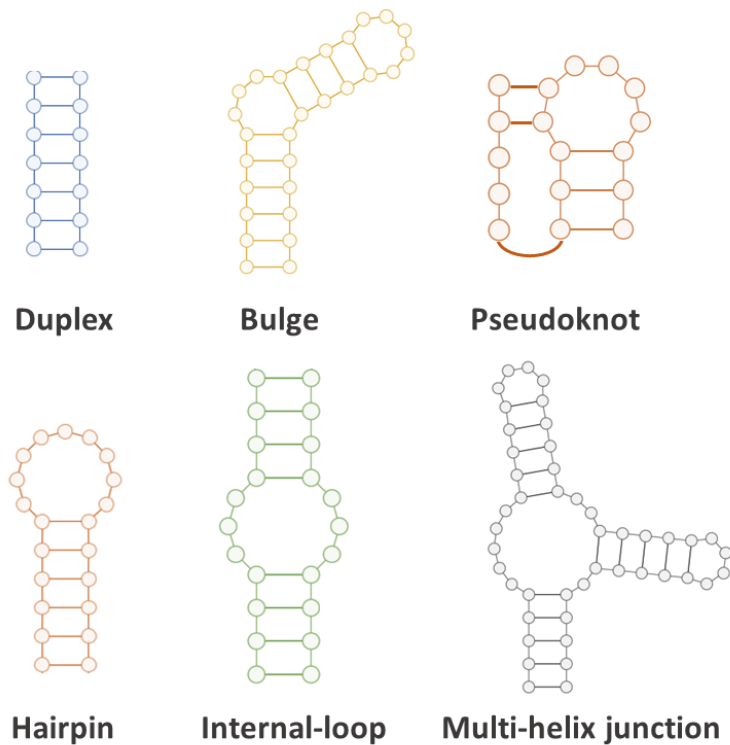


Figure 1.4. RNA 2D key substructures.

Tertiary Structure, Backbone Geometry, and RNA Architectures

The 3D RNA structure is determined by a complex network of canonical and non-canonical interactions coupled with the geometric constraints of its sugar-

phosphate backbone. Although backbone is likely to appear sequence-independent and flexible, it adopts a limited number of preferred conformations. Each phosphodiester linkage is characterized by a “suite” defined by seven backbone torsion angles (α , β , γ , δ , ϵ , ζ) plus the glycosidic χ angle (**Figure 1.5**). Approximately 50 distinct rotameric suites have been identified, each permitting a specific spatial orientation of adjacent nucleotides. Consequently, 3D structure is thus determined by a complex interplay of sequence-dependent internucleotide interactions that are modulated by intrinsically sequence-independent preferences of RNA backbone.¹⁶

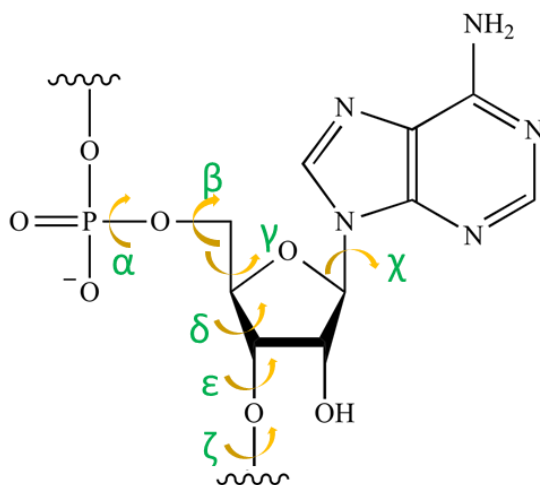


Figure 1.5. RNA torsion angles α , β , γ , δ , ϵ , ζ , χ .

Because of the expanded interaction capability and conformational versatility, RNA can fold into a vast array of structures, including triplex helices, kissing-loops, G-quadruplexes, and i-motifs (**Figure 1.6**).

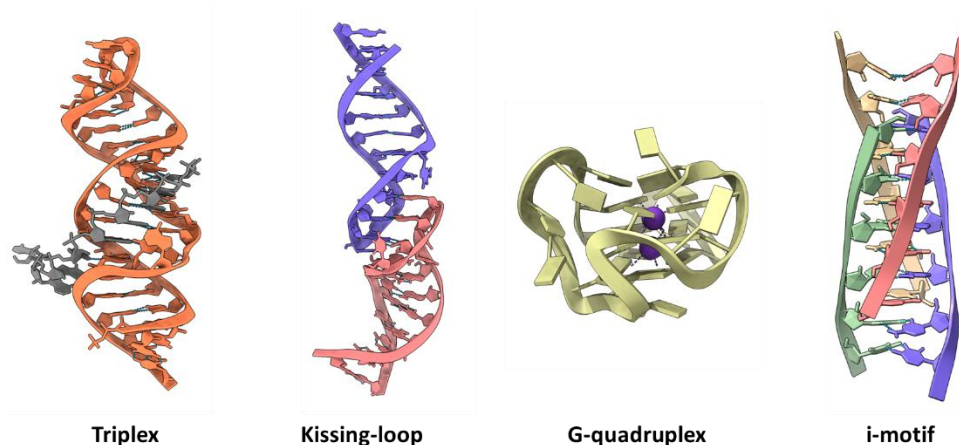


Figure 1.6. Advanced 3D RNA architectures, including triple helices, kissing-loops, G-quadruplexes, and i-motif-like structures.

1.2. RNA-based Therapies

The central role of RNA in nearly all biological processes and its ability to form elaborate 2D and 3D structures makes it uniquely suited for specific molecular recognition, engaging countless specific RNA-RNA, RNA-protein, RNA-DNA, RNA-small molecules, and RNA-ion interactions. Because non-coding RNAs comprise the 98% of the genome,^{10,11} it represents an appealing molecule for therapeutic intervention, expanding the potential of therapeutic strategies, which have been so far predominately relied on proteins, using structure-based (SBDD) and ligand-based drug design (LBDD). In fact, around 70-90% of all approved drugs target proteins (including enzymes, ion channels, GPCRs, transporters), a small fraction <10% target lipids, carbohydrates or other biomolecules, and only ca. 1-2% target nucleic acids directly.¹⁷ With the rise of RNA-based therapies, protein expression can be

altered, enabling solutions to one of the most persistent challenges in medicine: target the “undruggable”.¹⁸

Several types of RNA-based therapeutics have been implemented including antisense oligonucleotides (ASOs), small interfering RNA (siRNA), aptamers, and small molecules designed to bind structured RNAs (**Table 1.2**).

1.2.1. RNA as a Therapeutic Target

Antisense Oligonucleotides (ASOs)

ASOs are short single-stranded chains of oligonucleotides (typically consisting of 12-30 nucleotides) complementary to a specific target RNA through Watson-Crick base-pairing, which can alter the target RNA and modify protein expression.¹⁹ Upon binding, ASOs form ASO/RNA heteroduplex that modulates gene expression through RNase H-mediated degradation of the target mRNA, alteration of pre-mRNA splicing, or sterically hindering translation. To date the US Food and Drug Administration (FDA) has approved ten ASO drugs including fomivirsen (1998), mipomersen (2013), nirsinersen (2016), eteplirsen (2016), inotersen (2018), golodirsen (2019), viltolarsen (2020), casimersen (2021), eplontersen (2023), and tofersen (2023). These ASOs are used to treat diseases such as hereditary transthyretin amyloidosis, human cytomegalovirus, homozygous familial hypercholesterolemia, amyotrophic lateral sclerosis, spinal muscular atrophy, duchenne muscular dystrophy, and neuronal ceroid lipofuscinoses. Their clinical success has established ASOs as a mature and versatile therapeutic strategy.²⁰

Small Interfering RNA (siRNA)

Small interfering RNAs (siRNAs) are short double-stranded RNA molecules, typically 20-25 nt in length, that include sequence-specific gene silencing through the RNA interference (RNAi) pathway. Therapeutically, siRNA offers the remarkable ability to silence virtually any gene implicated in disease, making it an attractive modality for conditions driven by toxic gain-of-function proteins, dysregulated pathways, or overexpression. siRNA-based therapies have been developed for the past 20 years and the FDA has approved six siRNA including patisiran (2018), lumasiran (2020), inclisiran (2021), nedosiran (2023), and vutisiran (2022). These drugs treat diseases such as hereditary transthyretin amyloidosis, acute hepatic porphyria, primary hyperoxaluria type 1, and hypercholesterolemia, demonstrating the versatility and therapeutic potential of siRNA technology.^{21,22}

RNA Aptamers

RNA aptamers are single-stranded oligonucleotides that fold into well-defined 3D structures, enabling them to bind molecules with high affinity and specificity. In contrast to linear ASOs and siRNAs, which rely on WC base pairing to recognize complementary RNA sequences, aptamers function similar to small molecules or antibodies. Their binding to the target is achieved through electrostatic interactions (hydrogen bonds, π - π stacking interactions) and shape complementarity, disrupting protein-protein, protein-ligand, or protein-receptor interactions.^{23,24}

Aptamers are typically identified through the combinatorial systematic evolution of ligand by exponential (SELEX) enrichment procedure, a technique used to identify high-affinity nucleic acids that bind specific target molecules through repeated cycles of selection and amplification.^{25,26}

To date, two RNA aptamers have been approved by the FDA, including macugen (2004) for wet age-related macular degeneration, and izervay (2023) for geographic atrophy.²⁷

Table 1.2. Overview of RNA-based therapeutics.

RNA Therapy	Structure	Target	Recognition Mode	Mechanism of Action
ASO	Single-stranded	Pre-mRNA, mRNA	W-C base pairing	RNase H-mediated degradation, splicing modulation, steric blockade of translation
siRNA	Double-stranded	Mature mRNA	W-C base pairing	Gene silencing
Aptamer	Single	Proteins, extracellular receptors, small molecules	Shape complementarity, electrostatic interactions	Antagonism, inhibition

RNA Modifications

Although ASOs, siRNAs, and RNA aptamers are highly promising, their clinical translation is often limited by poor stability, rapid nuclease degradation, and challenges in delivery to the target tissues or cells. To enhance RNA stability and protect them from degradation, several chemically

modified nucleotides have been introduced. Such modifications can occur at the backbone, the sugar moiety, or the nucleobase of the RNA. Common modifications on the backbone include phosphorothioate (PS), and phosphorodiamidate morpholino oligomer (PMO). Sugar modifications contain 2'-fluoro (2'-F), 2'-O-Methyl (2'-OMe), and 2'-O-methoxyethyl (2'-MOE), while a nucleobase can be include 5'-methylcytosine (5'-mC), or abasic sites. These modifications offer advantages primarily concerning nuclease resistance, and improved aqueous solubility. Several FDA-approved RNA drugs incorporate modifications to improve their properties. For example, fomivirsen employs a fully PS-modified backbone, while a combination of modifications like the PS, 2'-O-MOE, and 5-mC are employed in nusinersen.^{20,22,28-30}

1.2.2. RNA as a Therapeutic Target for Small Molecules

The flexibility and the diverse folded structures of RNA enable the formation of grooves, bulges, internal loops that can serve as well-defined ligand binding pockets. This feature makes it attractive for therapeutic target for small molecules. Similar to traditional SBDD used for proteins, small molecule discovery can be applied to RNA targets with the aim to identify compounds that modulate RNA function, or stability.³¹ This strategy has proven successful, prompting the development of databases that contain RNA motifs and their associated small molecule binders. Example of such databases include Infora³² and HARIBOSS.³³

1.3. Computational Approaches for Studying RNA Structure

3D structures of RNA are largely dependent on their sequences; however, they are not static. Usually, RNA has multiple accessible 2D and 3D structures that lie within a narrow range of folding free energies. The time scales over which these alternate structures form and disappear can range from microseconds (for simple base pairing changes) to seconds or longer (for complex 2D refolding events). The experimental characterization of static RNA structural data remains challenging, and eventually, even more challenging remains understanding its structural dynamics.

To address this challenge, there have been intense efforts to complement the available experimental techniques with bioinformatics and advanced Molecular Dynamics (MD) simulation approaches. The main goals of atomistic MD simulations are i) to simulate the structural dynamics of RNA molecules explicitly and to support the interpretation and planning of experimental measurements of such dynamics; ii) to provide reliable and experimentally testable predictions; and iii) to gain insights that are not obtainable by current experimental methods, which will help explain RNA's multitudinous functional roles. Although MD simulations are useful in supporting or guiding experiments, it incorporates approximations that do not allow full prediction or reproduction of experiments. The main factors limiting accuracy of MD simulations is i) reaching the proper time scales, which can be addressed using enhanced sampling methods, and ii) the molecular mechanical force field (FF) that defines the relationship between the simulated molecule's exact geometry and its potential energy. Given the central role of force fields in determining simulation accuracy, their careful evaluation is an

essential step toward reliable structure-based RNA research and therapeutic design.

1.4. Thesis Objectives

In this thesis we focused on studying RNA-based therapeutics from three complementary perspectives that together illustrate the potential of RNA in drug discovery. First, we treat RNA as a ligand and we elucidate the structure and interaction of an RNA aptamer with its target. Second, we evaluate the performance of current state-of-the-art force fields across nine RNA motifs, and finally, we target a pre-mRNA with small molecules to regulate the inclusion of a particular exon.

1.4.1. Targeting EphA2 in Glioblastoma Stem Cells Using RNA Aptamer

Glioblastoma multiforme (GBM) is the most common and aggressive brain tumor in humans with a median survival of approximately 14.6 months and a dismal prognosis. Current treatment approaches typically involve tumor surgical resection, whenever feasible, followed by concomitant chemoradiotherapy with temozolomide.³⁴ Despite initial improvements in survival with this treatment regimen, tumor progression and relapse are frequent, primarily due to the therapy resistance exhibited by a subpopulation of cells known as glioblastoma stem cells (GSCs), which survive longer than

ordinary cells and demonstrate an enhanced ability to withstand conventional therapies, contributing to poor treatment outcomes.^{35–37}

Several key proteins are implicated in the pathogenesis of GBM and are associated with tumor proliferation, invasion, and angiogenesis, including Epidermal Growth Factor Receptor (EGFR),^{38,39} Vascular Endothelial Growth Factor (VEGF),^{40,41} Platelet-Derived Growth Factor Receptor (PDGFR),³⁸ and Erythropoietin-producing hepatocellular carcinoma A2 (EphA2).⁴² The latter is an elongated transmembrane protein that belongs to the Receptor Tyrosine Kinase (RTK) family. The extracellular region (ectodomain) consists of an N-terminal ligand-binding domain (LBD), an adjacent cysteine-rich domain (CRD), followed by two Fibronectin Type III domains (nFN3, cFN3). The EphA2 ectodomain is connected by a single transmembrane α -helix, which is extended intracellularly to a juxtamembrane (JM) region that tethers a tyrosine kinase domain (**Figure 1.7A**).

EphA2 plays a significant role in various physiological processes, such as cell development, migration, and tissue organization via activation of the so-called canonical pathway.⁴³ In this pathway, the ligand ephrinA1 binds to the EphA2 LBD, resulting in the oligomerization of EphA2/ephrinA1 complexes and in the following autophosphorylation of Tyr588, Tyr594, Tyr772 residues within the intracellular domain of EphA2 receptor. This autophosphorylation facilitates signaling through the receptor (forward signaling) and through the ligand (reverse signaling). In contrast, in absence of ephrinA1, EphA2 interacts with other proteins such as E-Cadherin and integrins, leading to the phosphorylation of Ser897 in EphA2's intracellular domain, which activates non-canonical signaling.^{43–45} Notably, this non-canonical signaling is associated with tumorigenesis and resistance to therapy (**Figure 1.7B**).^{43,44,46–}

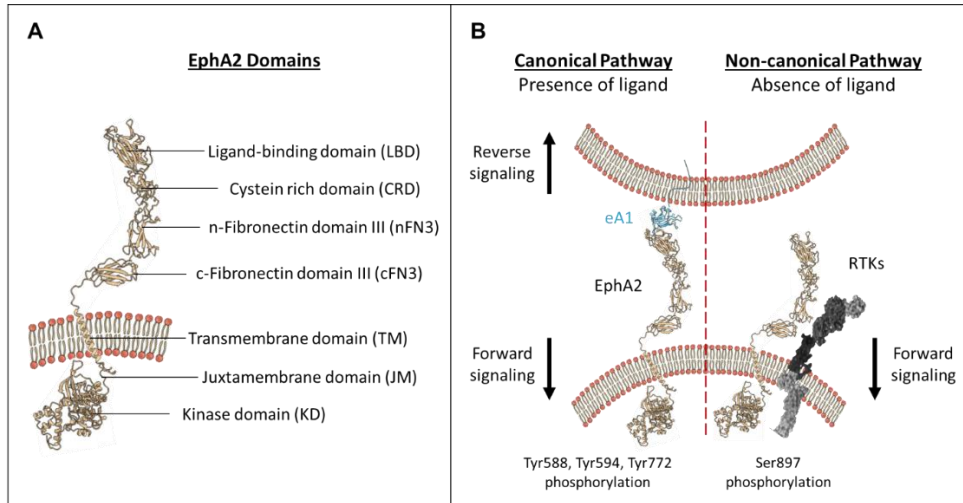


Figure 1.7. A) Structural organization of EphA2. B) Canonical and non-canonical pathways of EphA2. During the canonical pathway eA1 binds to EphA2 enabling signaling through the receptor EphA2 (forward signaling) and the ligand eA1 (reverse signaling). Upon activation by eA1, EphA2 oligomerizes and provokes autophosphorylation of tyrosine residues in the intracellular domain. In contrast, in the non-canonical pathway, EphA2 heterodimerizes with other RTKs, leading to phosphorylation of Ser897.

Importantly, EphA2 is overexpressed not only in various human malignancies, including melanoma, lung cancer, breast cancer, ovarian cancer, and GBM^{45,50–52}, but also in GSCs, allowing these cells to invade the brain more effectively and maintain their stem cell-like properties.^{50,51,53} The overexpression in GSCs contributes to the aggressive behavior of GBM and promotes resistance to conventional therapies. Consequently, EphA2 has emerged as a promising therapeutic target to overcome chemoresistance and improve treatment outcomes in GBM patients. Several strategies have been developed to inhibit this protein, aiming to suppress tumor progression and

enhance the efficacy of existing therapies. For instance, restoring the signaling balance by treating GSCs with EphA2-targeting siRNA⁵⁴ and ephrinA1-Fc (a soluble ephrinA1 dimer fused to Fc),⁵¹ induces astroglial differentiation, limiting the stem-like properties of the cells, and significantly suppressing the growth of xenograft tumors in vivo. Small molecules such as dasatinib have shown promise in pancreatic cancer by inhibiting the EphA2 receptor, disrupting its phosphorylation and key cellular processes.⁵⁵ More importantly, in GBM, dasatinib inhibits critical pathways that drive tumor growth and invasion: preclinical studies demonstrated its strong potential in slowing GBM progression, making it a promising therapeutic option for this aggressive cancer.⁵⁶ Tandem CAR-T cell therapy (TanCAR) destroy glioblastoma tumor cells by recognizing EphA2 and IL-13 receptor $\alpha 2$ in vitro.⁵⁷ However, these strategies face significant limitations due to lack of selectivity and consequent toxicity, delivery issues including blood brain barrier (BBB) membrane permeability or susceptibility to nucleases.^{58–60}

Given these limitations, we adopted an alternative approach to target EphA2 utilizing RNA aptamers due to their potential advantages.⁶¹ RNA aptamers can fold into a plethora of heterogeneous three-dimensional shapes allowing them to form high-affinity, high-specificity binding pockets for their targets. Moreover, by incorporating 2'-fluoro-modified ribonucleotides (2'-F-RNA), it is possible to enhance nuclease resistance and improve stability in biological environments.^{62,63} Prior studies, particularly in short RNAs, have shown that 2'-F-RNA preserve local geometry and stacking interactions with only minor perturbations with respect to their non-fluorinated analogues. Finally, aptamers are easy to synthesize and modify.²³

Using SELEX, our collaborators (Prof. Condorelli's research group, University of Naples Federico II, Italy) recently identified the 2'-fluorinated

Eph2-targeting RNA aptamer A40L and its truncated 30-mer derivative A40s (5'-ccuGuuGuucGAcAGGAGGcucAcAAcAGG-3'; lowercase letters indicate 2'-fluorinated ribonucleotides).⁶⁴ A40s was demonstrated to interact with EphA2 with high affinity ($K_D = 0.76 \pm 0.26$ nM) as well as to bind GSCs by EphA2 recognition and inhibit their growth, stemness and migration. It has exhibited stability in serum in *in vitro* experiments as well as the ability to cross the BBB in vivo after its systemic injection in healthy mice.⁶⁵

In this study, for the first time, we aim to unveil the three-dimensional structure of A40s and its complex with EphA2, by integrating advanced computational methods with experimental data. The model will allow us to identify key interaction points critical for binding, laying the foundation for further aptamer optimization aimed to enhance its binding affinity and therapeutic potential.

1.4.2. RNA Force Field (FF) Evaluation

Computational approaches for exploring RNA structure–dynamics–function relationships play an increasingly important role in RNA biology and medicine. As mentioned above, one critical challenge of RNA MD simulations is accessing the relevant time scales to accurately characterize the structural dynamics, thermodynamics, and kinetics of RNA structural rearrangements. However, enhanced sampling methods and coarse-grained strategies are largely applied to challenge this issue. A further critical challenge concerns the accurate treatment of ions in RNA systems. RNA is a highly negatively charged polymer whose structure and dynamics strongly depend on both monovalent and divalent ions, often occupying specific and functionally

relevant binding sites. In MD simulations, ions may require long equilibration times to localize correctly around RNA, and their behavior is further complicated by limitations in ion parametrization and their compatibility with different FFs. For these reasons, accurately capturing ion–RNA interactions remains a non-trivial task. A detailed analysis of the role of ions in MD simulations lies beyond the scope of this thesis, consequently, we focus here on the remaining critical bottleneck in RNA simulations, namely the accuracy of atomistic RNA force fields.

As will be extensively discussed in Chapter 2, a FF defines the potential energy surface of a biomolecule by specifying bonded terms (bonds, angle, dihedrals) and non-bonded terms (van der Waals, electrostatics). These terms rely on a basic approximation, which is subsequently tuned *via* specific parametrizations to obtain the best possible performance across representative training systems. However, the FF performance across a set of systems will not be uniform and will depend on its parameterization and on the extent to which compensating errors balance one another in a particular structural context. The applicability of a given FF version may vary between simulated biomolecules, and even between different parts of a single simulated molecule. For example, in G-quadruplexes the single-stranded loops are described less accurately than the G-stems.⁶⁶

Existing FFs not always capture the complexity of RNA structure such as diverse non-canonical base pairing interactions and complex backbone torsionals. Consequently, they are not yet sufficient for blind RNA structure prediction. However, in many applications simulated methods provide unique results that, when properly interpreted, enrich our insights into the structural dynamics and evolution of RNA molecules, which are often difficult to obtain experimentally.

This project aims to systematically evaluate the performance of state-of-the-art-RNA FFs across representative structural motifs. By analyzing deviations from experimental observables and comparing the behavior of canonical, non-canonical, and dihedral interactions, this study will identify strengths and limitations of current FFs and provide guidance for future refinement. These insights are critical to continue the successful expansion of MD-based methods within the RNA field and advance the reliability of computational studies.

1.4.3. Targeting Tubulin Associated Unit (tau) pre-mRNA Using Small Molecules

The tubulin associated unit (tau) is an intrinsically disordered protein, lacking a well-defined 3D structure under physiological conditions. In neurons, tau binds to microtubules, a fundamental component of the cytoskeleton, and promotes their assembly and stability (**Figure 1.8**). Through this interaction, tau plays a vital role in preserving the neuronal architecture and supports the organization of the microtubular network. Because microtubules serve as the major tracks for intracellular trafficking, tau indirectly contributes to the transport of essential cargoes and to effective communication between neurons and their target cells. Therefore, proper tau function is essential for maintaining integrity and homeostasis.⁶⁷

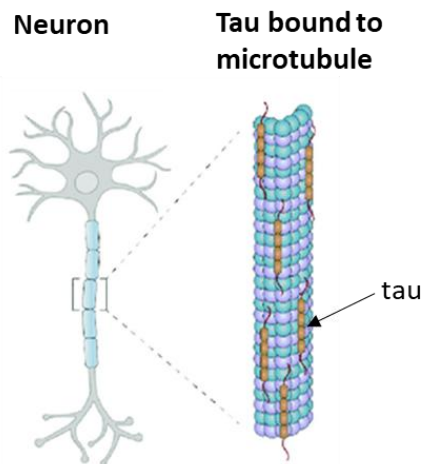


Figure 1.8. Tubulin associated unit (tau) protein binds to microtubule, leading to the proper shape of neurons.

The biosynthesis of tau protein isoforms is regulated by alternative splicing of the microtubule-associated protein tau (MAPT) pre-mRNA. Specifically, the MAPT gene consists of 16 exons: exons 0 and 1 encode the 5' untranslated sequence of MAPT mRNA, exons 2 and 3 encode 29-amino acid residue inserts near the amino-terminus, exons 4a, 6, and 8 are only transcribed in peripheral tissues, exons 9-12 encode microtubule binding domain repeats near the carboxyl-terminus, and exon 14 encodes part of the 3' untranslated region.

Alternative splicing of exons 2, 3, and 10 generates six tau isoforms. Inclusion or exclusion of exon 10 produces isoforms with either four (4R) or three (3R) microtubule-binding repeats, respectively (**Figure 1. 9**). In the healthy adult cerebral cortex, 3R and 4R isoforms are normally expressed at approximately equal levels, maintaining a 1:1 ratio critical for proper microtubule function.

Structurally, tau consists of four major domains: the N-terminal projection domain (NTD, residues 1–150); a proline-rich region (PRR, residues 166–242); the microtubule-binding domain (MTBD, residues 243–368), which contains either three (3R) or four (4R) imperfect repeats; and the C-terminal domain (residues 369–399), which includes a weakly homologous pseudo-repeat R' followed by the final C-terminal tail fragment (400–441).^{67,68}

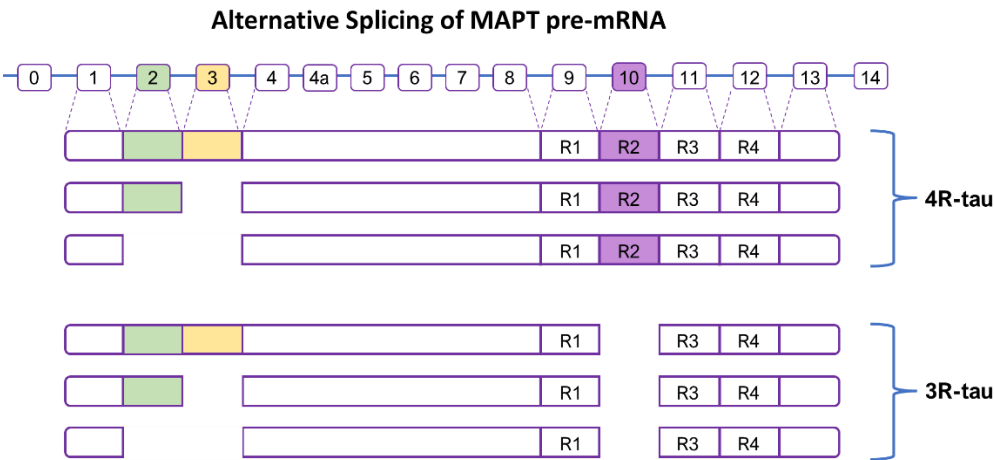


Figure 1. 9. The alternative splicing of MAPT pre-mRNA, which leads to 4R-tau and 4R tau isoforms.

The exon 10–intron junction, located within the MTBD-coding region, forms a stem-loop (hairpin) structure, to which U1 small nuclear ribonucleoprotein (snRNP) splicing factor can access, thereby regulating splicing. Perturbations on the exon 10-intron hairpin can promote U1 snRNP access to the splicing site, increasing exon 10 inclusion, and finally promoting 4R-tau production. 4R-tau aggregation promotes loss of structural integrity, cell dysfunction and ultimately neuronal death. Such dysregulation is associated with

neurodegenerative disorders, referred as tauopathies. Tauopathies are related to more than 20 disorders, including Alzheimer's Disease, frontotemporal dementia, chronic traumatic encephalopathy, and progressive supranuclear palsy.⁶⁷⁻⁶⁹

Several factors can perturb the tau exon 10-intron hairpin. Pathogenic MAPT mutations, often located in the intron following exon 10, can destabilize the hairpin. Specifically, four intronic mutations at positions +3, +10, +13, +14, and +16 (numbering with the first nucleotide of the splice-donor site as +1) lie within the hairpin that partially masks the splice donor (**Figure 1.10**). A characteristic example is the +14 C→U mutation in frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). Studies show that these mutations decrease the thermodynamic stability of the RNA hairpin in vitro, thereby enhancing exon 10 inclusion and increasing 4R-tau production. Conversely, mutations that stabilize the hairpin reduce exon 10 inclusion, demonstrating the bidirectional control of splicing by RNA structure.⁷⁰

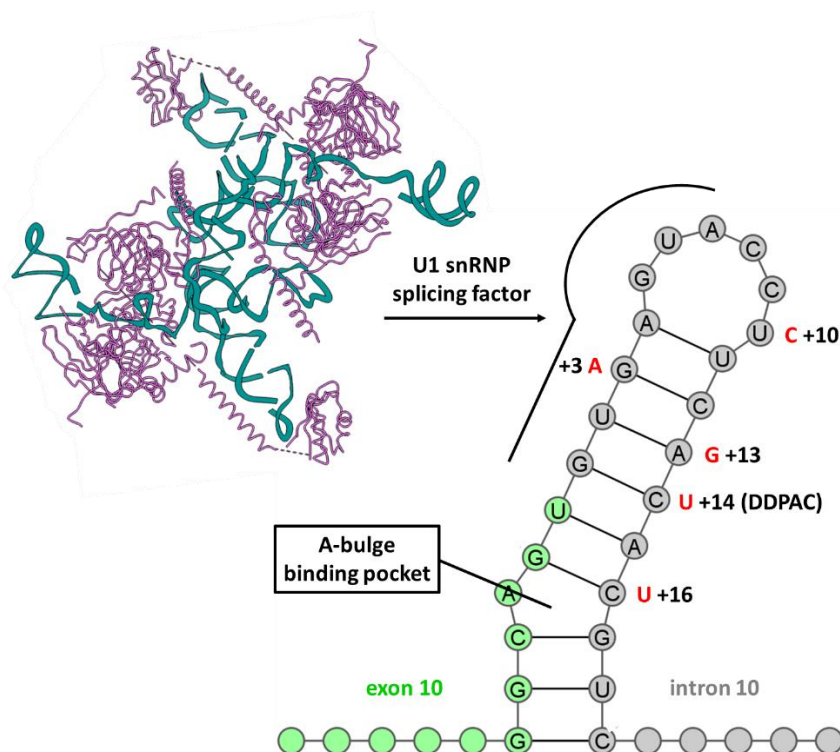


Figure 1.10. Hairpin structure formed by the exon 10–intron junction, which regulates access of the U1 snRNP splicing factor and thereby controls alternative splicing. Mutations at positions +3, +10, +13, +14, and +16 (numbered relative to the first nucleotide of the splice-donor site, +1) destabilize the hairpin, promoting U1 snRNP binding and increased production of 4R-tau. Targeting the A-bulge binding pocket of the hairpin with small molecules represents a promising strategy to stabilize the structure and restore physiological splicing.

These experimental findings suggest that the exon 10-intron hairpin represents a promising therapeutic target. Inspecting the hairpin structure, an A-bulge is identified on the stem, which can serve as a binding pocket for small molecules. Therefore, development of small molecules capable of stabilizing

the hairpin could restore proper 4R:3R tau ratios, preventing tau-mediated neurodegeneration.

Indeed, previous studies identified a reference molecule, GM118,⁷¹ a carbazole analogue known to bind the A-bulge of the tau pre-mRNA hairpin, stabilizing it primarily through stacking interactions. Nevertheless, due to inherent challenges associated with small molecules, in particular issues of selectivity and specificity, a need to discover novel, optimized compounds with improved binding properties remains.

The aim of this project is to identify small molecules capable of binding and stabilizing the tau harpin and modulate tau splicing. To achieve this, we explored a guanylylhydrazone (GH) library inspired by GM118 and employed artificial intelligence (AI)-based tools to analyze 2D and 3D similarity with the reference molecule. The resulting narrowed library was screened using 1D ¹H and WaterLOGSY NMR experiments to identify candidate binders. The most promising compounds were subsequently evaluated by our collaborators to assess their activity on tau isoforms. Finally, computational studies were performed on the top candidate to investigate its interactions with the tau hairpin and provide guidance for rational optimization.

This integrated approach, combining computational and experimental techniques, enables the discovery of small molecules that bind the tau exon 10-intron hairpin, provides insights into their binding mechanisms, and establishes a rational pathway for the future optimization of RNA-targeting compounds for tau-related neurodegeneration.

2. METHODOLOGIES

2.1. RNA Structure Prediction Techniques

RNA structure plays a central role in determining RNA function. To uncover how RNA molecules participate in diverse cellular processes, it is crucial to accurately predict both their 2D and 3D conformations, which govern molecular recognition, gene expression, catalytic activity, and interaction with proteins, metabolites or other RNAs. However, obtaining RNA structures experimentally using NMR techniques, X-ray crystallography or cryo-EM remains challenging, expensive and time consuming. As a result, computational RNA structure prediction has become an essential tool for connecting RNA sequence to structure and function.

2.1.1. RNA Secondary (2D) Structure Prediction

The secondary structure is defined as the pattern of base-pairing interactions, typically involving canonical base pairs (A-U, G-C) and wobble G-U pairs, arranged into structural elements such as stems, bulges, loops, and hairpins. Accurately predicting this 2D architecture is essential because it forms the foundation for understanding RNA folding, function, and tertiary structure formation. Several computational and experimental approaches have been developed to model RNA secondary structures, each relying on different principles and offering complementary advantages (**Figure 2.1**).

Main Approaches for RNA 2D Structure Predictions

- *Thermodynamic (nearest-neighbour) approaches:* The thermodynamic methods are based on nearest neighbour rules that predict the stability of helices and loop motifs from experimentally determined free-energy parameters. The structure prediction is then accomplished by finding the lowest free energy structure, which is the single most probable structure in a folding ensemble.⁷² Software such as RNAstructure⁷³, RNAfold⁷⁴, and NUPACK⁷⁵ implement the minimum free energy approach. Alternatively, these algorithms compute the Boltzmann ensemble allowing sampling of alternative conformations and identification of centroid structures that best represent the ensemble instead of obtaining a single optimal structure.
- *Probabilistic and statistical learning approaches:* A widely used method for RNA 2D structure prediction includes probabilistic approaches that use statistical models to provide probabilities for base-pairing interactions. Specifically, instead of taking into account a unique solution with the highest probability, these approaches employ an estimator, which takes into account the ensemble of all possible solutions. Representative examples are: i) the centroid estimator (e.g. Fold),⁷⁶ ii) the maximum expected accuracy estimator (CONTRAFold),⁷⁷ or iii) a generalized centroid estimator,⁷⁸ called a γ -centroid estimator (CENTROIDFOLD).⁷⁹
- *Kinetic folding approaches:* Complementary to the thermodynamic approaches, kinetic folding approaches have also been employed to predict RNA secondary structures. Kinetic models consider the dynamic folding process, the energy landscape (including local minima and folding barriers), and the possibility that an RNA may become trapped in a metastable conformation rather than reaching the global minimum. For example, the Kinefold web server⁸⁰ performs

online stochastic folding simulations of nucleic acids using the folding dynamics algorithm⁸¹ and the exactly clustered stochastic simulation algorithm⁸², enabling the capture of alternative base-pairing patterns or folding pathways that pure thermodynamic methods might miss.⁸³

- *Coarse-grained, physics-based models*: These approaches have been successfully used to improve the representation of RNA conformational entropy. In particular, such models employ virtual bond-based coarse-graining of RNA backbone conformation to generate structural models through Monte Carlo sampling. Using then a coarse-grained statistical potential, the entropy and free energy parameters for the hairpin, bulge, internal loops, and multiway junctions can be obtained, enabling accurate prediction of secondary structures.^{84,85}
- *Hybrid computational-experimental approaches*: In addition to the principal methods, experimental data can be integrated to guide predictions. Software such as RNAstructure⁷³ and Vfold2D⁸⁵ incorporate experimental constraints (i.e. chemical probing data) to obtain more accurate RNA 2D structure predictions.
- *Machine-Learning (ML) approaches*: In ML approaches, scoring parameters are trained directly from reference RNA structures rather than relying solely on experimentally derived thermodynamic data. (e.g. CONTRAfold uses statistical learning).⁷⁷ ML can also be integrated with thermodynamic data. For example, MXfold⁸⁶ combines thermodynamic parameters with richly parameterized weight terms learned from training data. This hybrid strategy helps prevent overfitting by reverting to thermodynamic parameters when encountering substructures not represented in the training set. Deep neural networks have also been employed for RNA 2D structure

prediction. Such algorithms formulate RNA secondary structure prediction as multiple binary classification problems that predict whether each pair of nucleotides forms a base pair or not, using a deep neural network trained with a large amount of training data. Although these models can capture complex patterns from large datasets, their highly flexible parameterization raises concerns about overfitting.⁸⁷

- *Pseudoknot-prediction approaches:* Pseudoknots are RNA 2D structures in which base pairs cross the usual nested pattern, typically forming when unpaired bases in a loop pair with bases outside the loop. Standard algorithms cannot predict them because they assume non-crossing (nested) structures. To predict pseudoknots, dynamic programming algorithms (e.g. NUPACK⁷⁵), integer linear programming²⁴ and heuristic sampling methods (e.g. IPknot^{15,88}) have been developed, all of which allow crossing base pairs.

Despite the development of diverse approaches for RNA 2D structure prediction, including thermodynamic, probabilistic, statistical, coarse-grained, and machine-learning methods, significant limitations remain. Insufficient high-quality experimental data hinder accurate predictions, particularly for RNA families lacking structural information. Machine-learning models are also susceptible to overfitting, which can lead to unreliable results when applied beyond their training distribution. In addition, incomplete parametrization, especially for modified nucleotides, continues to constrain the accurate modeling of RNA secondary structure.

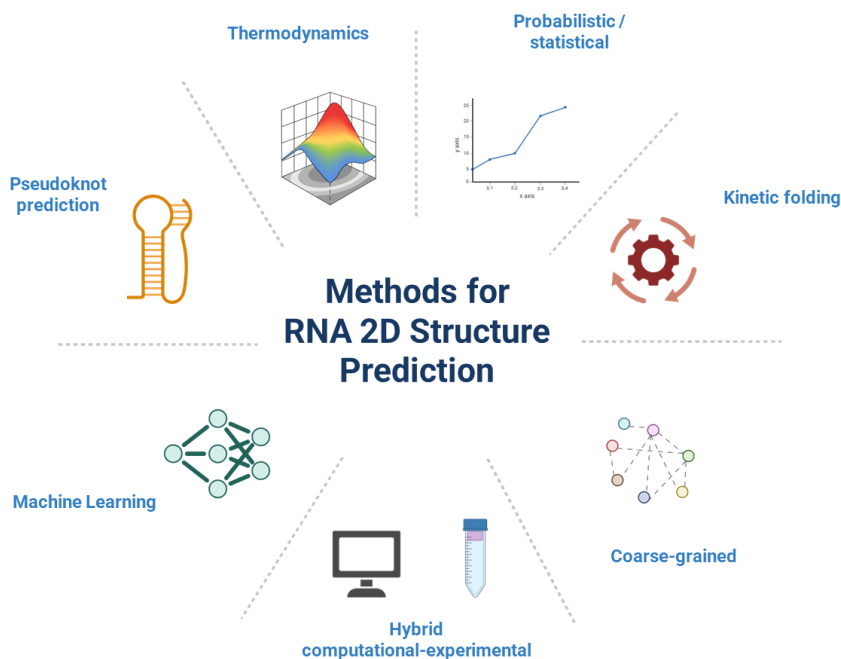


Figure 2.1. Methods for RNA 2D structure prediction.

2.1.2. RNA tertiary (3D) Structure Prediction

Similar to the RNA 2D structure, the tertiary (3D) conformation of RNAs is essential for understanding its biological function. Experimental determination, as mentioned previously, remains laborious and technically challenging, making the computer-aided structure prediction from sequences an effective way to address the challenge. To this aim, three main methods have been developed.

- The first strategy employs the *motif template assembly* method. In the modeling process, the input secondary structure is first decomposed into fragments, which are then matched to 3D elements using libraries, such as RNA FRABASE⁸⁹, that contain experimental elements. The

selected 3D elements are subsequently merged to form the complete 3D structure, followed by refinement through energy minimization (**Figure 2.2**). Prominent tools using this approach include RNAcomposer⁹⁰, 3dRNA⁹¹, and Vfold3D⁸⁴.

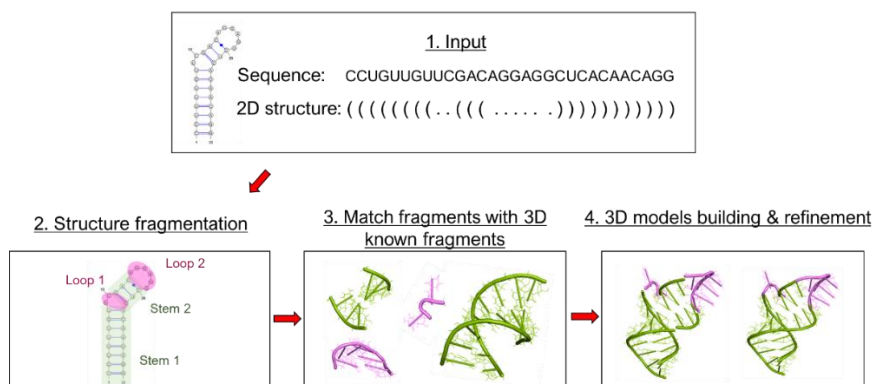


Figure 2.2. Fragment-based approach for RNA 3D prediction.

- Another computational method for RNA folding simulations involves the *coarse-grained approach* implementation. In this method, the nucleotide chain is first represented with a simplified, coarse-grained model to reduce computational cost while preserving key structural features. The conformational space is then explored, typically using a Monte Carlo sampling, and a knowledge-based energy function is employed to produce the most energetically favorable 3D conformations (**Figure 2.3**). A widely used webserver implementing this technique is SimRNA.⁹²

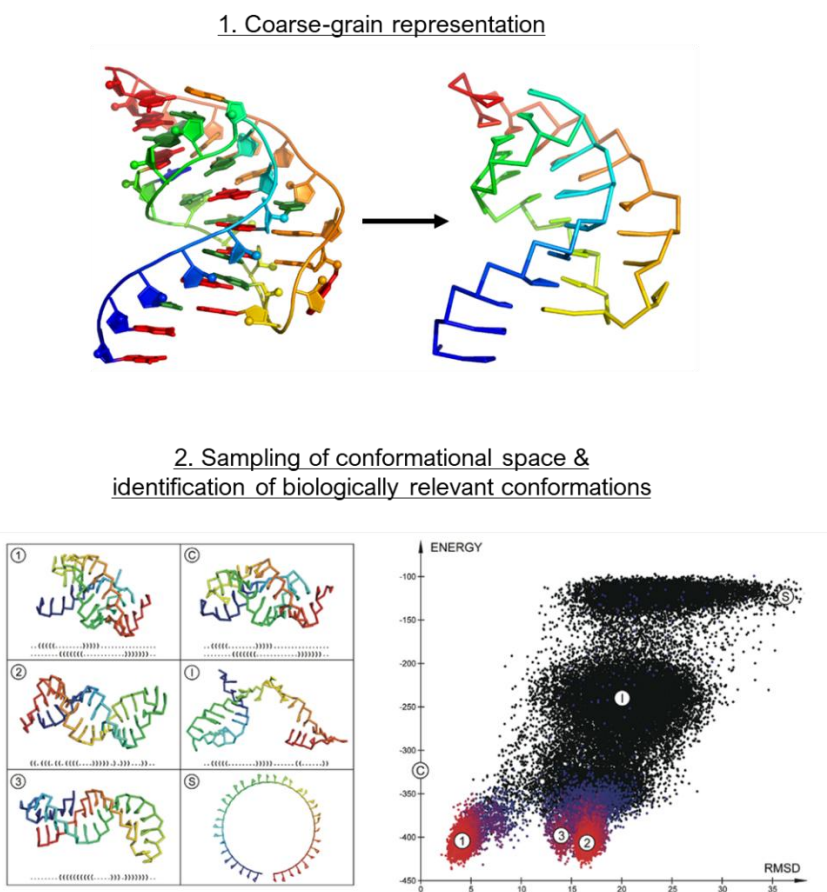


Figure 2.3. Coarse-grained approach for RNA 3D prediction. Adapted from Boniecki et al.⁹²

- Finally, machine learning (ML), and in particular deep learning (DL) techniques have been developed for the accurate prediction of RNA 3D structure from sequences. Several frameworks, including DeepFoldRNA⁹³, trRosettaRNA⁹⁴, E2Efold-3D⁹⁵, and AlphaFold3⁹⁶, use multiple sequence alignments (MSAs) and transformer architectures to predict inter-nucleotide geometries or directly generate 3D coordinates. These methods provide state-of-the-art accuracy and generally outperform traditional RNA 3D structural prediction

methods. For example, trRosettaRNA outperformed RNAComposer and SimRNA in 86.7% and 96.7% of the 30 benchmark cases, respectively, while DeepFoldRNA surpassed the best models submitted in the RNA-Puzzles challenges. AlphaFold3 also outperforms previous blind docking methods, such as Vina and RoseTTAFold All-Atom.^{93,94,96}

Despite the promising advantages, ML-based techniques remain limited by the availability and quality of experimental RNA structure data, which are far less abundant than for proteins. Consequently, overfitting and poor generalization to rare or poorly characterized RNA motifs can lead to misleading predictions, highlighting the need for continued integration of experimental data and improved datasets.

2.2. Molecular Docking

Molecular docking is an established *in silico* structure-based method widely used in drug discovery to explore the binding of a molecule to a target protein. When molecular docking was discovered in mid-1970s, it was limited to predicting interactions between a ligand, typically a small molecule, and the binding pocket of a macromolecule of biological interest, such as a protein.^{97,98} However, advances in computational methods and structural biology have since broadened its scope to include more complex biomolecular systems, such as protein-protein, protein-nucleic acid, protein-membrane, nucleic acid-small molecule, and cofactor/ion interactions.^{99–102} The identification of the most likely binding conformations typically requires two steps: (i) exploration of the conformational space to generate candidate binding poses; (ii) estimation

of the interaction energy associated with each of the predicted binding conformations (scoring). In practice, however, docking accuracy is often limited by the incomplete sampling of ligand and the neglect of the target conformational flexibility and explicit solvent, as well as the use of approximated scoring functions, which can provide results that do not correlate with the experimental binding affinities.^{103,104}

2.2.1. *Ligand and Target Preparation*

Docking begins with ligand and target preparation, a critical step that defines the accuracy of all subsequent predictions. Depending on the nature of the system, whether it involves small molecule, peptide, protein, nucleic acid, or another biomolecule, appropriate structural refinement is essential. When only a two-dimensional representation of the ligand is available, a three-dimensional conformation must be generated, with careful consideration of tautomeric, ionization states, and stereochemistry. Prior to docking, energy minimization of both ligand and target structures is recommended to avoid steric clashes.¹⁰⁵

The structural state of the target also has an impact on docking performance. Comparative studies have demonstrated that using the *holo* form of a receptor yields higher enrichment rates (around 70%) compared to the *apo* (around 20%) or modelled forms (around 10%) of the binding site. However, the *holo* form can be over-influenced by the ligand in *holo* complex, if the active site has undergone conformational collapse around the co-crystallized ligand. Overall, the *holo* form will give the best enrichment, if appropriate structural preparation and validation are performed.^{106,107}

2.2.2. *Generation of Binding Poses*

Several techniques have been implemented to explore the conformational space and generate potential binding poses. The most widely used docking tools treat the ligand as flexible and the receptor rigid. They use different search methods, which can be divided in three categories: systematic methods, random/stochastic methods, and simulation methods.¹⁰⁸

Systematic methods: To generate all possible conformations of a molecule, systematic algorithms are used to explore all the degrees of freedom. The most common systematic approach is the *exhaustive search*, in which each rotatable bond of the ligand is incrementally rotated at predefined intervals to sample all possible orientations. Although conceptually straightforward, this approach is limited by the exponential increase in computational cost with the number of rotatable bonds. To overcome this, geometrical or chemical constraints are applied during the first steps of docking to provide an initial binding pose, followed by hierarchical optimization of the most promising conformations.¹⁰⁸ In other cases, the ligand is divided in rigid core fragments and in flexible parts. The rigid fragments are first docked into the binding site, and the flexible parts of the ligand are added incrementally. This *fragment-based* strategy has been adopted by several docking software such as DOCK4.0,¹⁰⁹ HammerHead,¹¹⁰ and FlexX.¹¹¹

Random/Stochastic methods: In this approach, random conformational sampling is applied by probabilistic algorithms rather than systematic enumeration. Random perturbations are applied to a ligand or a population of ligands to explore conformational space within the binding site. The most

widely used methods are Monte Carlo (e.g. MCDOCK¹¹²), tabu search and genetic algorithms.^{113,114} After each round of docking and scoring, the highest ranked conformations are used to seed the iteration, progressively optimizing the search. The key consideration is to ensure adequate iterations to sample the active site properly. Overall, these methods provide a good balance between predictive performance and computational efficiency.

Simulation methods: The advantage of simulation-based docking approaches is that both the ligand and the target are treated as flexible molecules. Molecular dynamic programs commonly used for simulations include AMBER,^{115,116} CHARMM,¹¹⁷ and GROMACS¹¹⁸ software. Nevertheless, molecular dynamic simulations are computationally demanding, and the system can often become trapped in a local energy minima. In addition, inadequate sampling can prevent the system from overcoming energy barriers and exploring alternative binding modes.

Docking methods that combine elements of systematic, stochastic, and simulation-based protocols have been developed to improve efficiency. Glide (Schrödinger, LLC)^{119,120} is one of the most commonly used tools. In this approach, a grid is used to map the physicochemical properties of the active site, followed by generation of low energy conformations of the ligand using Monte Carlo sampling. Subsequently, these poses are refined by energy minimization within the binding site, and the top three to six low energy poses are further optimized through additional Monte Carlo simulations.

2.2.3. *Scoring*

Once a set of binding poses has been generated, each pose must be evaluated by a *scoring function*. The aim of scoring the poses is firstly to discriminate incorrect binding modes, and secondly, to distinguish the active compounds from the inactive. Ideally, a scoring function should be able to accurately calculate binding free energies (ΔG_b). Robust computational methods such as free energy calculations¹²¹ could be used to quantitatively predict the ΔG_b , however they are computationally expensive and unpractical in the docking of a set of molecules. Therefore, scoring functions are based on approximations and do not fully account for solvation and entropic effects. In fact, they finally estimate the receptor-ligand complementary rather than the binding free energy using a functional such as:

$$\Delta G_{bind} = \Delta G_{solv} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib} \quad (2.1),$$

where ΔG_{solv} is the contribution of the solvation free energy. ΔG_{conf} is the contribution of conformational changes in the ligand and in the protein. ΔG_{int} is the free energy, due to ligand-protein interactions. ΔG_{rot} (entropic contribution) is the free energy loss arising from freezing internal rotations of protein and the ligand. $\Delta G_{t/r}$ is the penalty for the loss of translational and rotational free energy caused by the association of two-bodies (ligand and receptor) to create a single body (complex). ΔG_{vib} is the free energy of the system due to changes in vibrational modes.^{120,122,123}

Currently, two main classes of scoring functions are available: classical scoring functions, which include molecular mechanics force field-based, empirical and knowledge-based scoring functions, and machine learning (ML) scoring functions (**Figure 2.4**). The first class scoring functions are distinguished by the type of structural descriptors employed. However, from a mathematical perspective, all classical scoring functions assume an additive functional form. In contrast, nonparametric machine-learning scoring

functions do not make assumptions about the form of the functional, but the functional form is inferred from training data in an unbiased manner, instead. As a result, classical and machine-learning scoring functions behave very differently in practice.

Molecular mechanics force field-based scoring functions: This kind of scoring functions treat energetics of a ligand-protein complex as a sum of the interaction energy between the ligand and the receptor and the internal energy of the ligand itself, while neglecting the internal energy of the protein. The binding energy is estimated by a sum of non-bonded interactions, including electrostatic forces (described by Coulomb's law) and van der Waals interactions (commonly modeled using the Lennard-Jones potential). Standard force field-based functions face several limitations. These include the omission of solvation and entropic contributions, dependence of the score on ligand size (larger ligands tend to yield artificially better scores), and inaccuracies arising from the use of cut-off distances in the treatment of non-bonded interactions.¹²⁴

Empirical scoring functions: These scoring functions are parametrized by fitting to experimental data, such as binding affinity, steric, hydrogen-bonding and lipophilic interactions, as well as solvation and entropic effects. Their major advantage lies in their simplicity in evaluating individual energy terms. However, a key limitation is their dependence on the quality and diversity of the training set, which can reduce their ability to accurately predict the binding affinities of ligands that structurally differ from the training set.^{125–127}

Knowledge-based scoring functions: In these approaches, ligand-protein complexes are modeled using relatively simple atomic interaction-pair potentials derived from statistical analyses or experimentally determined

structures. They are mainly designed to reproduce experimental structures rather than binding energies. The main advantages of knowledge-based scoring functions include their computational efficiency, which enables the rapid screening of large libraries, their ability to account for uncommon interaction types, such as sulphur-aromatic or cation- π interactions, which are generally poorly handled in empirical approaches, and their transferability, as they can be applied on systems that were not used for the setup of the training set. A drawback of this method is that their parameterization is essentially based on information implicitly encoded in limited sets of ligand-protein complex structures, which can lead to insufficient differentiation among atom types.^{128–130} Examples of popular knowledge-based scoring functions include PMF¹²⁸ and DrugScore.¹³¹

Machine learning (ML) scoring functions: This type of scoring functions for binding affinity predictions are trained on complexes with continuous binding data (i.e. binding constants), binary binding data or a combination of continuous and binding data. A scoring function can be made to be generic by training on diverse complexes or family specific by restricting to complexes within the considered protein family. Each complex is then represented as a set of features and a binding measurement paired with its experimental binding measurement. The data are divided into a training set, which is used for model training and selection, and a test set that validates the trained scoring function. Finally, the predictions are compared with known measurements that the algorithm has never seen in the training or test set to evaluate the performance of the scoring function. The most widely used ML methods for scoring functions are the tree-based methods, such as random forests and gradient boosting machines, due to their robustness to heterogeneous features and easy interpretation of the method (i.e. RF-Score). However, deep learning and neural networks show great promising (e.g. Docking Score ML¹³²) but their

use is still limited due to the large amount of data required for the training set.^{133,134}

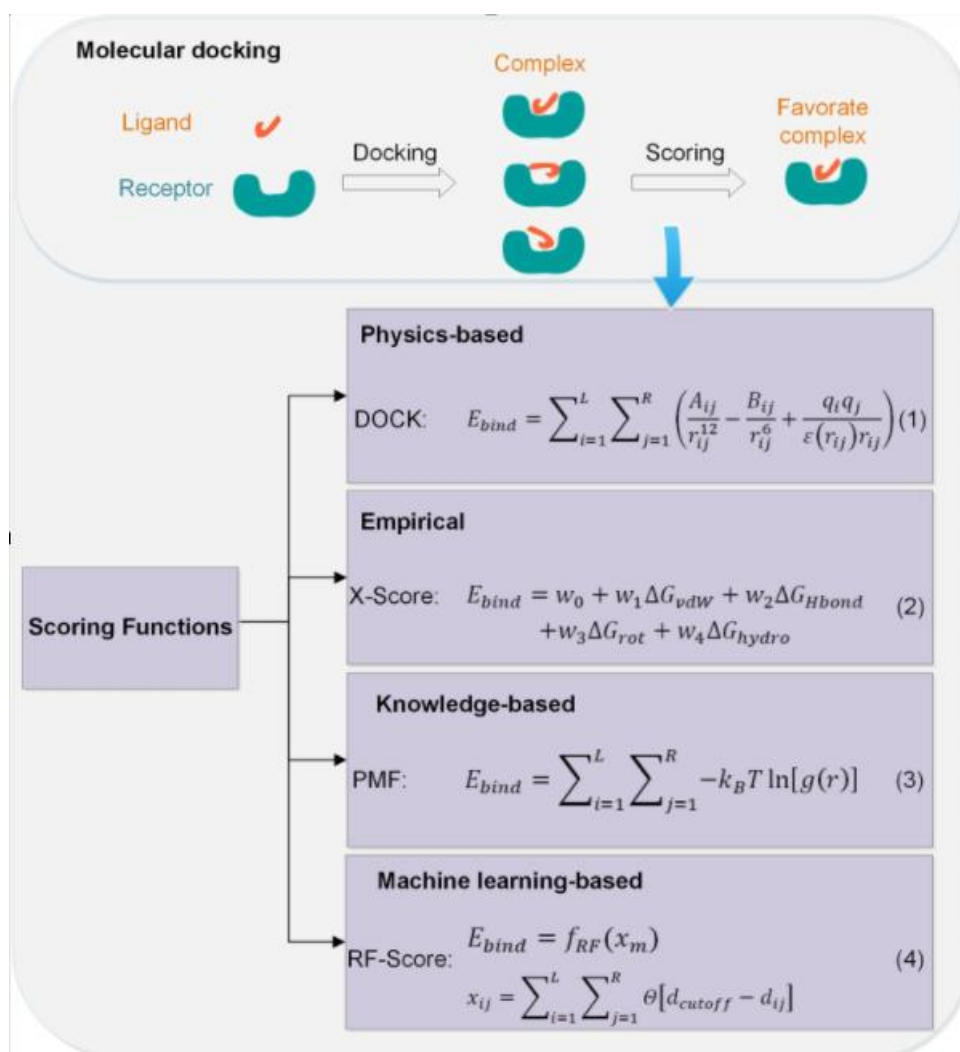


Figure 2.4. Four categories of the scoring functions for receptor–ligand docking.

Adapted from Jin et al.¹³⁴

2.2.4. *Challenges and Recent Advances*

Although the algorithms for binding pose generation appear to be successful at predicting ligand-target binding modes, major challenges remain. In particular, considering protein flexibility and the explicit role of water molecules during docking continues to be a significant limitation. Recently developed methods are beginning to address these issues into docking workflows.

Protein flexibility: Allowing both ligand and target to be flexible by sampling all the conformational degrees of freedom would avoid prediction of incorrect ligand binding modes or poor docking scores and would improve accurate docking. However, increasing degrees of flexibility enhances the computational cost. The most common strategy to overcome this limitation is modeling the flexibility of the binding site exclusively using molecular dynamics or Monte Carlo calculations or rotamer libraries. Another strategy includes the use of an ensemble of protein conformations rather than a single one. Also, a method called *soft* docking, in which the effect of the repulsion term in the scoring function is reduced by permitting some steric clashes between the ligand and the protein. Recent docking workflows incorporate ensemble docking and induced-fit¹³⁵ (change of the shape of the receptor's active site explaining enzyme specificity and catalysis) docking protocols, sometimes coupled with molecular dynamics or enhanced sampling methods, to better capture receptor flexibility and conformational changes upon ligand binding.^{107,136}

Explicit solvation: The presence of water molecules has a significant impact on determining the accuracy of ligand-target docking. Some docking software treat water either implicitly or explicitly. In the former approach, water is

represented as a continuous dielectric medium, which simplifies calculations and reduces computational cost, whereas in the latter approach, water molecules are included in the vicinity of the binding site where they are capable of forming hydrogen bonds. Although computationally more expensive, accounting for water molecules in the vicinity of the ligand and the target, particularly after energy minimization of their orientation, has shown to statistically improve the reliability of docking accuracy.^{136,137,138}

2.2.5. *Applications of Molecular Docking*

Molecular docking was originally developed to rationalize ligands activity towards a target of interest and to perform structure-based virtual screening for the discovery of new potential drugs.

Besides these applications, docking has found valuable uses in other drug discovery tasks. Among these is target fishing and profiling, where docking helps identify alternative targets that exhibit good complementarity to known ligands, including potential off-targets responsible for adverse drug reactions. In addition, docking supports polypharmacology studies, by enabling the identification of ligands that simultaneously bind to a pool of selected targets of interest, and contributes to drug repurposing, by identifying novel uses for chemical compounds with already optimized safety profiles.¹³⁹

Ultimately, molecular docking has shown a promising role in overcoming the challenges related to drug formulation and development. This can be applied to predict stable drug–excipient inclusion complexes, analyze molecular interactions governing solubility, dissolution, and bioavailability, and assist in

the rational design of dosage forms that optimize drug performance and accelerate development.¹⁴⁰

2.3. Molecular Dynamics (MD) Simulations

2.3.1. *Molecular Simulation Methods Across Scales*

In this section we will discuss the scope of MD, how MD simulations are performed, and how macroscopic properties can be computed through statistical mechanics. The study of molecular systems' properties is achieved with the aid of computer models. Depending on the size of the system and the time scales in which the most interesting phenomena occur varies and thus different techniques are employed to model them (**Figure 2.5**).

For instance, Quantum Mechanics (QM) describe the electrons of a system and are used when chemical reactions or bond rearrangements must be captured.¹⁴¹ These methods operate at electronic resolution, typically handling tens to a few hundred atoms and accessing time scales up to a few picoseconds due to their high computational cost. Coarse-Grained (CG) simulations simplify groups of atoms into single interaction sites, dramatically increasing accessible time and length scales, and are particularly valuable for studying mesoscale biological processes such as membrane dynamics, protein folding, and self-assembly over microsecond to millisecond time scales, while Brownian Dynamics (BD), are used to simulate the diffusive motion of mesoscopic particles rather than atoms in Cartesian space by integrating the Langevin equation and can efficiently explore microsecond to millisecond time scales.¹⁴²

MD uses an empirical potential to calculate the energy of a system (usually simulating systems of 10^3 - 10^6 atoms over nanosecond to microsecond time scales) based on the nuclear positions only by numerical integration of Newton's equations of motion in order to describe the dynamics of the system and to generate information about the system on the microscopic level. Subsequently, the microscopic information is linked to the macroscopic observables through statistical mechanics.

To bridge different spatial and temporal scales, hybrid approaches have been developed. QM/MM (Quantum Mechanics/Molecular Mechanics) combines quantum and classical mechanics by treating the chemically active region of a system at the QM level while modeling the surrounding environment using MM, thereby capturing electronic effects with reduced computational cost. MM/CG methods integrate atomistic molecular mechanics with coarse-grained (CG) models to extend simulations to larger systems while retaining key atomistic details in regions of interest.^{143,144}

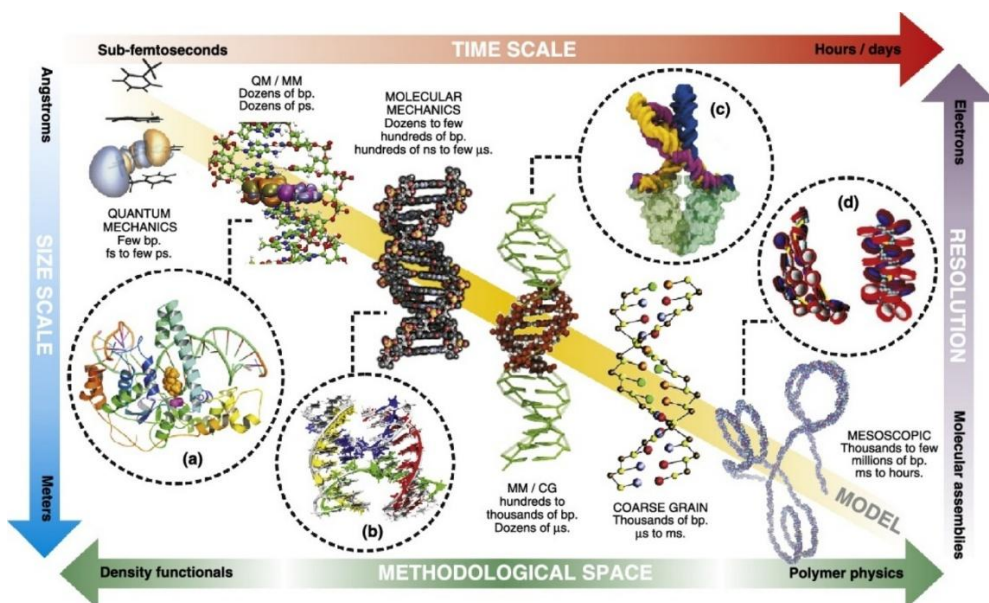


Figure 2.5. Schematic representation of multiscale simulation methods in computational chemistry, illustrating the relationship between system size, spatial resolution, and accessible time scales. Adapted from Dans et al.¹⁴⁴

MD is based on different approximations, such as the Born-Oppenheimer approximation and the potential energy function. According to Born-Oppenheimer approximation, electron motion in molecules is considerably faster than nuclear motion and therefore can be considered negligible. Moreover, the empirical potential energy function describes the intra- and inter-molecular interactions that occur between the nuclei and it is further discussed in the next section.^{145,146}

2.3.2. *The potential Energy Function (force field)*

The energy of a system is represented by the Hamiltonian:

$$H = T + V \quad (2.2)$$

Where T is the Kinetic Energy and V is the Dynamic (potential) Energy. In classical mechanics, the kinetic energy of a system can be described by:

$$T = \frac{1}{2}mv^2 \quad (2.3)$$

However, the description of the potential energy term, V , is more complicated, due to the intermolecular interactions need to be taken into consideration. Therefore, the potential energy, $V(r)$, can be defined as the non-bonded and bonded energy terms of the system.¹⁴⁷

$$V(r) = V_{bonded} + V_{non-bonded} \quad (2.4)$$

The bonded terms include the simple covalent binding in addition to the complex hybridization and π -orbital effects, while the non-bonded interactions are described by the intermolecular van der Waals and electrostatic interactions. In order to simulate the bonded and non-bonded energy terms of a molecular system in a classical mechanic's framework, an empirical potential energy function that describes these interactions is introduced with a set of equations, also known as force field.¹⁴⁷ The different equations for bonded and non-bonded interactions that comprise the potential energy function are presented below.

The V_{bonded} term is the summation of the terms for bonds, angles, proper dihedrals and improper dihedrals (**Figure 2.6**):

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{impropers} \quad (2.5).$$

Non bonded interactions involve two types of interactions: the van der Waals interaction energy and the electrostatic interaction energy:

$$V_{non-bonded} = V_{vdW} + V_{electrostatics} \quad (2.6)$$

These interactions are the most computationally demanding, due to the fact that they include long-range interactions between the atoms in the system.

Bonded Interactions

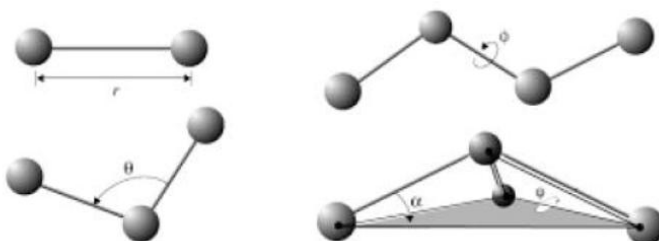


Figure 2.6. Symbolic representation of the bonded interactions: bond stretching r (upper left), bond angle bending θ (bottom left), dihedral ϕ (upper right) and improper dihedral ψ (bottom right) and the small out-of-plane angle α . Adapted from Adcock and McCammon.¹⁴⁶

Bond stretching

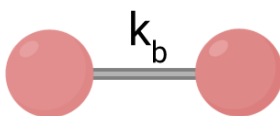


Figure 2. 7. Definition of bond stretching between two atoms.

Atoms in molecules are connected with covalent bonds, which vibrate (**Figure 2. 7**). The bond vibrations are described by a harmonic potential according to Hooke's law:

$$F = -kx \quad (2.7)$$

Taking into account that the force can be written as the gradient of the potential energy of the system, we conclude to this formula:

$$F = -kx = -\nabla V \quad (2.8),$$

which leads to the potential of bond stretching:

$$V = -kx^2 \quad (2.9),$$

which describes the force constant k_b of the vibration of the spring and the distance between the two atoms, b , from the average distance, b_0 , of the two bonded atoms:

$$V_{bonds} = k_b(b - b_0)^2 \quad (2.10)$$

Bond angle bending

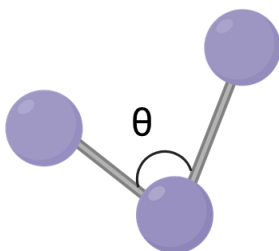


Figure 2. 8. Definition of angle bending between three atoms.

The angle bending terms describe the force originating from the deformation of the valence angles between three covalently bonded atoms (**Figure 2. 8**). This conformational change can be described again using Hooke's law (eq. 2.10) ensuing into an angle bending term, which is expressed as:

$$V_{angles} = k_{\theta}(\theta - \theta_0)^2 \quad (2.11)$$

where k_θ is a force constant, θ is the angle between three atoms and θ_0 is the reference angle.

Dihedral torsion term

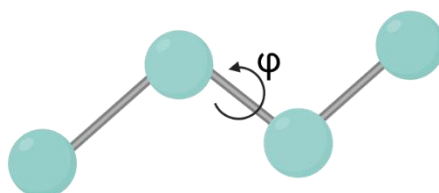


Figure 2. 9. Definition of dihedral torsion between four atoms.

Torsional terms describe the barriers to rotations existing between four covalently bonded atoms. They are weaker than the bond stretching and angle bending terms, however, the need for the proper torsional potential arises, as bond-stretching or angle-bending potentials cannot describe the energetics of rotatable bonds, since they appear in four covalently bonded atoms (**Figure 2. 9**)

The rotation along a covalent bond is periodical, therefore the torsional potentials can be outlined by a cosine function as presented below:

$$V_{dihedrals} = k_\varphi(1 - \cos(n\varphi - \delta)) \quad (2.12)$$

where k_φ is a force constant, n is the periodicity, φ is the angle between the planes formed by the first and the last three of the four atoms, and δ sets the minimum energy angle.

Improper dihedral term

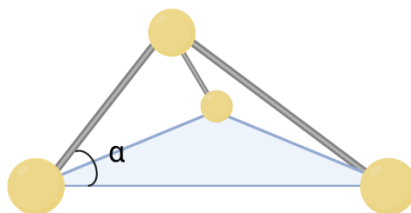


Figure 2.10. The improper dihedral.

The improper dihedral term is designed both to maintain chirality about a tetrahedral heavy atom and planarity of certain atoms. The improper angle is defined as the angle between the plane formed by the central atom and two peripheral atoms and the plane formed by the peripheral atoms (**Figure 2.10**). The improper potential is expressed by a harmonic function:

$$V_{impropers} = k_{\omega}(\omega - \omega_0)^2 \quad (2.13)$$

where ω is the angle between the planes as previously described.

Non-bonded Interactions

Van der Waals interactions

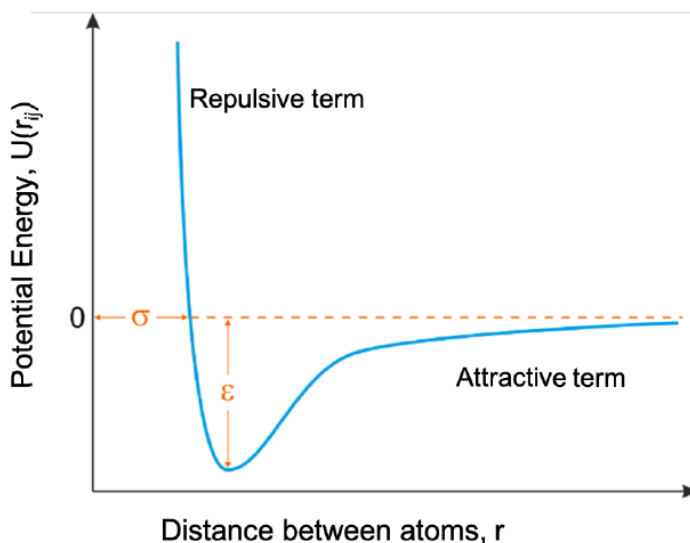


Figure 2.11. The Lennard-Jones potential, where σ is the collision parameter and ϵ the well depth. Adapted from Periodni Glossary.¹⁴⁸

The van der Waals (vdW) interaction occurs on atoms of close proximity. It is strongly repulsive at short range and attractive when there are vacillations in the charge dispersion in the electron clouds (**Figure 2.11**). It is described by a Lennard-Jones potential:

$$V_{vdW} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.14)$$

Where r_{ij} represents the distance between two atoms. It is parametrized by σ , which is the collision parameter (the separation for which the energy is zero) and the depth of the potential well, ϵ .

Electrostatic interactions

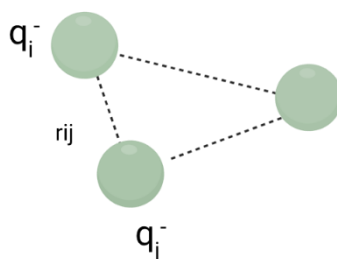


Figure 2.12. Schematic representation of electrostatic interactions between three charged atoms q_i , q_k , q_j .

Two partial charged non-bonded atoms can interact electrostatically with each other (**Figure 2.12**). This long-distance electrostatic interaction between two atoms is described by Coulomb's law:

$$V_{electrostatics} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.15)$$

Potential energy function

Finally, the equation of the potential energy function that represents the force field is:

$$\begin{aligned} V(r) = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{dihedrals} k_\varphi(1 - \cos(n\varphi - \delta)) + \sum_{impropers} k_\omega(\omega - \omega_0)^2 \\ & + \sum_{vdW} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{i\xi}} \right)^6 \right] + \sum_{electrostatics} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \end{aligned} \quad (2.16)$$

In MD simulations, long-range interactions are among the most time-consuming ones. For example, in a simulation box consisting of 500 molecules, the range of these forces often exceeds half the box length. The most commonly used method in MD simulations to correctly account of the electrostatic interactions, is the Ewald sum, which includes the interaction of an ion or molecule with all its periodic images.¹⁴⁹

Simulation systems with periodic boundary conditions (PBCs) provide a good approximation of a small subsystem embedded in a larger bulk phase. During the simulation it is assumed that when a molecule exits the central box, its periodic image simultaneously enters the box from the opposite side (**Figure 2.13**) ensuring continuity without physical walls or surface molecules at the boundaries. To prevent double-counting interactions (a single particle interacting with the same particle multiple times), a cut-off distance is applied to non-bonded interactions. The cut-off is typically set to less than half the box length in any dimension. In this way, a natural lower limit to the size of a periodic simulation box is applied, as the box must be sufficiently large to capture all the most significant non-bonded interactions.¹⁵⁰

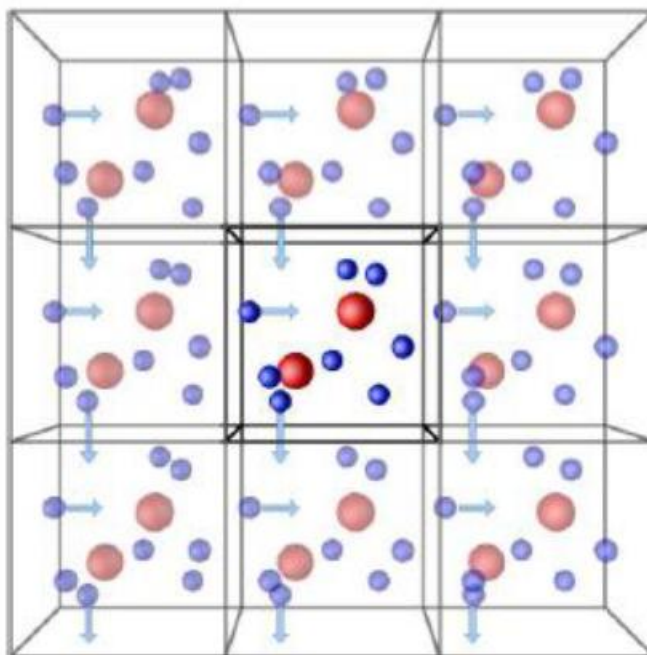


Figure 2.13. A periodic system illustrating the periodic boundary conditions used in MD simulations. Adapted from Isaac's Physics Notes.¹⁵¹

2.3.3. *Energy Minimization*

The energy landscape of a biomolecule is characterized by a large number of local minima. Prior to an MD simulation, the system should undergo energy minimization to reach a local energy minima (**Figure 2.14**). This process alleviates steric clashes between atoms, corrects distorted bond lengths or angles, and removes any non-physical van der Waals contacts that may exist, thereby ensuring a realistic starting structure. Without proper minimization, these unfavorable interactions can provoke collapse of the simulation, and destabilization of the system leading to unreliable trajectories.¹⁵²

Given a function f , which depends on the variables $x_1, x_2, \dots, x_n$, a minimum or maximum of f is defined as a point where the first derivative of the function with respect to each of the variables is zero and the second derivative is positive (Figure 2.14):

$$\frac{\partial f}{\partial x_i} = 0 \quad (2.17)$$

$$\frac{\partial^2 f}{\partial x_i^2} > 0 \quad (2.18)$$

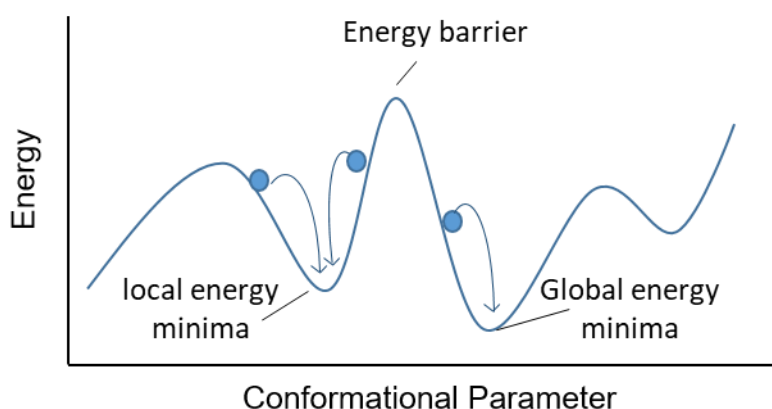


Figure 2.14. Representation of one-dimensional potential energy surface. Minimization methods move downhill to the nearest minimum.

Most common minimization algorithms use derivatives of the energy with respect to the coordinates to predict the location of the closest minimum.¹⁵³ The three main algorithms that are used are: Steepest Descent, Conjugate Gradient and Newton-Raphson algorithms. The appropriate selection of the method is based on the quantity of iterations essential to converge, as well as on the quantity of function assessments necessary per iteration.¹⁵²

Steepest Descent Algorithm

The basis for the Steepest Descent method, also known as the gradient descent method, is the simple observation that a continuous function should decrease, at least initially, if one takes a step along the direction of the negative gradient. The search direction is given by:

$$F(r) = -\frac{dV(r)}{dr} \quad (2.19)$$

In the steepest descent algorithm, the step length is determined through a line search, which utilizes the direction vector obtained from the first derivative of the potential function. This process identifies the optimal step size that minimizes the energy in that direction. Once the local minimum along this search direction is located, the step is applied to update the coordinates. To find the best possible step size, the line search needs an enormous number of function assessments. However, the two main computational advantages of the steepest descent algorithm are the simplicity with which a computer algorithm can be implemented and the low storage requirements necessary.^{153,154}

Conjugate gradient and Newton-Raphson

The Conjugate Gradient and Newton-Raphson algorithms offer more efficient approaches to energy minimization compared to steepest descent method. The first, utilizes information from previous first directions and first derivatives to find the best direction for a line search, avoiding redundant searches and accelerating converge. In contrast, the Newton-Raphson method makes use of both the first and second derivatives to precisely estimate the curvature of the potential energy surface. By considering this curvature, it predicts where the

gradient will change direction, leading to a more accurate and often faster convergence to the nearest minimum.¹⁵³

$$V(r) = V(r_0) + V'(r_0)(r - r_0) + \frac{V''(r_0)(r - r_0)^2}{2} + \dots$$

$$\frac{dV}{dr} = 0; \frac{d^2V}{dr^2} > 0 \quad (2.20)$$

2.3.4. *MD Formalism*

As previously mentioned, MD simulations are based on Newton's second law of motion:

$$F_i = m_i \alpha_i = m_i \frac{\theta^2 r_i}{\theta t^2}, i = 1, \dots, N \quad (2.21)$$

where m_i is the mass of each atom i and α_i is the acceleration on atom i .

Forces can be expressed as the negative derivatives of the potential energy function $V(r_1, r_2, r_3, \dots, r_N)$:

$$F_i = -\frac{\theta V}{\theta r_i} \quad (2.22)$$

By combining the two equations:

$$-\frac{\theta V}{\theta r_i} = m_i \frac{\theta^2 r_i}{\theta t^2} \quad (2.23)$$

and by solving this differential equation (eq. 2.23) a trajectory of the system is obtained.

To start an MD simulation, an initial structure of the system is required. Most often, the initial configuration is an X-ray crystallography, a cryo-EM or NMR

structure, which can be easily obtained by the Protein Data Bank (PDB). The initial coordinates are defined from these experimental structures, while velocities can be obtained from a distribution, e.g. Maxwell – Boltzmann distribution.¹⁵⁵

The equations are solved synchronously in short time steps, while the parameters such as the temperature and pressure remain constant at the desired values throughout the simulation. The different coordinates, related to time, represent the trajectories of the system. MD simulations evolved into a mature technique that can be used effectively to understand macromolecular structure-to-function relationships and along with the continuous evolution of technology that provides faster and cheaper computers the time and the cost for performing a MD simulation has been significantly decreased. The time scale of solvated protein simulations is up to nanoseconds, however, simulations up to microseconds have been also recorded.¹⁵⁶

Methods for integrating Newton's equations of motion

The potential energy function, which was analyzed in 2.3.2, does not allow for an analytical solution to solve the equations of motion, therefore we conclude to treat the classical equations numerically. To this end, several algorithms have been developed to integrate Newton's equations of motion, and some of those algorithms are presented below.¹⁵⁷

Overall, all the integration algorithms predict the positions, velocities and accelerations of the atoms using the Taylor series expansion:

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots$$

$$\begin{aligned}
v(t + \delta t) &= r(t) + \alpha(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \dots \\
\alpha(t + \delta t) &= \alpha(t) + b(t)\delta t + \dots
\end{aligned}
\tag{2.24}$$

where r is the position, v is the velocity (also can be expressed as the first derivative with respect to time) and α is the acceleration (second derivative with respect to time). δt is a short time interval and is defined by the fastest vibration of the system, typically the C-H bond ($\delta t = 1 \text{ fs} = 10^{-15} \text{ s}$).

The total force remains constant during each time step and is equal to the vector sum of its interactions with other particles at time t . From the force, the acceleration, positions and velocities of the particles are calculated, and the available information at time t are used to calculate the system's positions and velocities at time $t + \delta t$.

Verlet Algorithm

The most common integration algorithm is the Verlet algorithm, which does not require the velocities, but requires a single force calculation per cycle and uses the following two equations:

$$\begin{aligned}
r(t + \delta t) &= r(t) + v(t)\delta t + \frac{1}{2}\alpha(t)\delta t^2 + \dots \\
r(t - \delta t) &= r(t) - v(t)\delta t + \frac{1}{2}\alpha(t)\delta t^2 + \dots
\end{aligned}
\tag{2.25}$$

The sum of these two is the basic formalism of this algorithm:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \alpha(t)\delta t^2 \tag{2.26}$$

The Verlet algorithm utilizes the positions (r) and accelerations (a) at time t , as the updated positions at time $t + \delta t$ are calculated from the positions at time $t - \delta t$. The velocities can be calculated from the equation below:

$$V(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} \quad (2.27)$$

The main advantage of Verlet algorithm is that it requires small amount of storage memory; however, it may generate a lot of errors.

Leap-Frog Algorithm

The Leap-Frog algorithm, which is a variation of the Verlet algorithm, uses these two equations:

$$\begin{aligned} r(t + \delta t) &= r(t) + v\left(t + \frac{1}{2}\delta t\right) \\ v\left(t + \frac{1}{2}\delta t\right) &= v\left(t - \frac{1}{2}\delta t\right) + a(t)\delta t \end{aligned} \quad (2.28)$$

Specifically, it calculates the velocities at time $t + 1/2\delta t$, and these velocities are used to calculate the positions at time $t + \delta t$. This algorithm calculates the velocities directly, but not at the same time with the positions. The advantage of this algorithm lies on the fact that it generates less errors than the Verlet algorithm.¹⁵⁷

2.3.5. *Introduction to Statistical Mechanics*

MD simulations provide a way to relate the microscopic properties of atoms to macroscopic thermodynamic observables using statistical ensembles *via*

statistical mechanics. A statistical ensemble is a set of representative points in the $6N$ -dimensional phase space ($3N$ for the spatial coordinates, r , and $3N$ for the momentum, p).¹⁵⁸ The “ergodic hypothesis”, which derives from the greek word $\epsilon\rho\gamma\omega\nu$ (energy) and $\omicron\delta\acute{o}\varsigma$ (path), states that over an infinitely long simulation, the time average of a property along a trajectory is equivalent to its ensemble average. With this assumption, ensemble properties (e.g., energy, pressure, temperature) can be computed using a distribution function that gives the probability of the system adopting a particular configuration at a given energy under specified thermodynamic conditions. As a result, by calculating the values of the property at different times t , the macroscopic average can be obtained.¹⁵⁹

The partition function

To calculate the distribution function of an ensemble, the type of the ensemble should be specified. The canonical ensemble (NVT) is the set of the possible positions and momenta for all particles such that the number of particles, N , the volume, V , and the temperature, T , are fixed and it specifies variation of energy.

The probability function, $P(\Gamma)$ in the NVT ensemble is described by the Gibbs distribution:

$$P_i(\Gamma) = \frac{e^{-\beta H_i(r^N, p^N)}}{\sum_i e^{-\beta H_i(r^N, p^N)}} \quad (2.29)$$

where r^N, p^N are a specific set of positions and momentum respectively, H_i is the Hamiltonian function that describes the total energy in state i , $\beta = 1/k_B T$, T is the the temperature, and k_B is the Boltzmann constant ($k_B = 1.38 \times 10^{-23} J K^{-1}$).¹⁶⁰

The dominator of Equation 2.29 is the partition function, Q_{NVT} .

$$Q_{NVT} = \sum_i e^{-\beta H_i(r^N, p^N)} \quad (2.30)$$

The partition function computes the number of the microstates accessible to the system at a specific temperature.¹⁶¹

As the microstates form a continuous set, the partition function is expressed as an integral:

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int e^{-\beta H(r^N, p^N)} dr^N dp^N \quad (2.31)$$

where h is the Plank's constant ($h = 6.62 \times 10^{-34}$), $1/h^{3N}$ is introduced to make the partition function dimensionless, and the factor $1/N!$ accounts for the indistinguishability of the N particles.

Under the assumption that the kinetic (T) and potential energy (V) terms of the Hamiltonian are separable, Equation 2.31 becomes:

$$H(r, p) = T + V$$

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int e^{-\beta V(r^N)} dr^N e^{-\beta T(p^N)} dp^N = Z_{NVT} Q_P \quad (2.32)$$

where Z_{NVT} is the configuration integral:

$$Z_{NVT} = \int e^{-\beta V(r^N)} dr^N \quad (2.33)$$

and Q_P is the momentum integral, which can be analytically obtained using the following formula:

$$Q_P = \frac{V^N}{N!} \left(\frac{\sqrt{2\pi m k_B T}}{h^2} \right)^{3N} \quad (2.34)$$

where m is the molecular mass.

The partition function Q_{NVT} plays significant role in statistical mechanics, due to the fact that it is associated with the Helmholtz free energy:

$$A_{NVT} = -k_B T \ln(Q_{NVT}) \quad (2.35)$$

The Gibbs free energy can be computed using another ensemble, the isothermal-isobaric ensemble (NPT). In this ensemble, the number of particles, N , the pressure, P and the temperature, T , are constant. The partition function can be written as:

$$\mathcal{E}_{NPT} = \iiint e^{-\beta H(r^N, p^N)} e^{\beta p V} V^N dr^N dp^N dV \quad (2.36)$$

Then the Gibbs free energy is:

$$G = -k_B T \ln(\mathcal{E}_{NPT}) \quad (2.37)$$

Consequently, the Gibbs free energy can be computed from Z_{NVT} , which can be calculated from the potential energy of the system. Since MD simulations directly sample the potential energy surface under constant T and P , ensemble averages obtained from NPT simulations can be used to compute Gibbs free-energy differences.^{159–161}

2.3.6. *Thermostats and Barostats*

In the following section, we introduce the purpose of thermostats and barostats in MD simulations and explain why they are essential for controlling temperature and pressure, respectively. We also analyze the theoretical principles underlying these algorithms, providing the foundation needed to understand how they regulate the system and ensure correct sampling of the desired statistical ensemble.

Thermostats

When performing MD simulations, it is desirable to emulate the laboratory experiments conditions. to achieve this, sampling from the canonical ensemble is demanded, where the temperature is constant. Normally, if the temperature of the system must be maintained during the simulation, some thermostat algorithm will be employed.

During an MD simulation the temperature is measured from the kinetic energy using the equation:

$$\frac{3}{2}Nk_bT = \langle \sum_{i=1}^N \frac{1}{2}m_i v_i^2 \rangle \quad (2.38)$$

Where the angled brackets indicate that the temperature is defined as a time-averaged quantity. However, if equation 2.42 is used to calculate the temperature for a single snapshot in time of an MD simulation instead of time-averaging, this quantity is referred to as the instantaneous temperature. The instantaneous temperature will undergo fluctuations around the target temperature.

There are several thermostat algorithms that can be implemented to control temperature. Some thermostats rescale velocities outside of the molecular dynamics' equations of motion, while others include stochastic collisions between the system and an implicit bath of particles, or they explicitly include additional degrees of freedom in the equations of motion giving the effect of an external heat bath.¹⁶² Some popular thermostats are the Langevin,¹⁶³ Andersen,¹⁶⁴ Gaussian,¹⁶⁵ Simple velocity rescaling¹⁶⁶ and Berendsen.^{164,167}

Barostats

The thermodynamic properties, which are studied in a laboratory, are measured under open-air conditions, meaning that they are measured at

essentially constant pressure and temperature (isothermal - isobaric ensemble). In order to obtain a non-atmospheric pressure, some device, would be needed to control the pressure and volume of the system. As is the case with thermostats, if the pressure must be maintained in a simulation, a barostat algorithm will be employed to sample this ensemble.

Barostat algorithms control pressure alone, so if the target ensemble is isothermal-isobaric, they must be applied along with a thermostat. The pressure of an MD simulation is commonly measured using the Clausius Virial theorem (an expectation value relating to positions and forces) and when pairwise interactions and PBC are considered, different approaches are utilized. These formulas give pressure as a time-averaged quantity, so if we use these formulas to compute the pressure for a single snapshot, this quantity, as in the case with temperature, is referred to as instantaneous pressure, which will undergo under fluctuations around the target pressure.

Considering a hypothetical system being compressed by a fictitious piston, which has a mass acting from all directions, it can be considered as applying a uniform compression. The mass of the piston can be adapted such that it will change the frequency that the particles in the system will interact with the system enclosure. These impacts from the particles on the “enclosure” will transfer a stress on the system box from the surroundings and serve as a type of barostat. Some widely used barostats include the Simple volume rescaling, Berendsen,¹⁶⁷ Andersen,¹⁶⁴ Parrinello-Rahman^{168,169} and Monte Carlo.^{170,171}

2.4. Artificial Intelligence in Drug Discovery

2.4.1. *Introduction to Artificial Intelligence, Machine learning, and Deep Learning*

Artificial intelligence (AI) is a technology-based system involving various advanced tools and networks that can mimic human intelligence. Over the past few years it has garnered widespread attention in various sectors of society, including the pharmaceutical industry. AI utilizes systems and software that can interpret and learn from input data to make independent decisions for accomplishing specific objectives.¹⁷²

Machine learning (ML) is a branch of AI that focuses on utilizing statistical approaches and trained algorithms to generate classifications or predictions that reveal critical insights in data mining. The method uses subsets of data to generate an algorithm that may use novel or different combinations of features and weights than can be derived from first principles. ML is divided into four categories based on the learning methods: supervised, unsupervised, semi-supervised, and reinforcement learning.¹⁷³

Supervised ML: In this approach, a labeled dataset is given as an input to train a model to predict an output (target) from a set of input (features). Datasets are generally split into training, validation, and testing subsets. During training, the algorithm relates the input features (x) to the target variable (Y), so that future predictions can be associated $Y=f(x)$.¹⁷⁴ The model changes the parameters based on how accurately it predicts the known targets in the training data, a process “supervised” by the presence of correct output labels. The validation dataset is then validating how well the learned mapping

generalizes beyond training set, and fine-tunes the model avoiding overfitting. The performance of the algorithm is evaluated on the test subset, which includes data that the algorithm has never seen before.

Unsupervised ML: This method uses unlabeled data that are not connected to any predefined outputs or target variables. The aim of unsupervised ML is to detect hidden patterns or relationships within a dataset based on the structure of the data itself. Unsupervised learning algorithms group or categorize according to the similarities and differences of the samples. The most common unsupervised methods include clustering algorithms, such as k -means and hierarchical clustering, which group data points into clusters with similar characteristics; and dimensionality reduction approaches, such as Principal Component Analysis, which turns data into simpler representations while preserving essential structural information.

Semi-supervised ML: This approach falls between supervised and unsupervised learning. During the learning process, it utilizes datasets that contain a small set of labeled data together with a much larger set of unlabeled data. This choice takes advantage of the abundant unlabeled information to improve model performance while minimizing the time and cost required for manual annotation. First, the model implements unsupervised methods to identify hidden patterns within the unlabeled data. Subsequently, these patterns are used to propagate labels from the small labeled subset to the unlabeled samples. Finally, the combined dataset is used to train a predictive model in a semi-supervised way.

Reinforcement Learning: In this method, the algorithm learns to make a sequence of decisions by interacting with an environment and receiving feedback. Based on the desired result, the algorithm will be “rewarded” or will receive a “penalty” and will adjust its actions to maximize the cumulative

reward. It is the closest attempt at modeling the human learning experience, as the algorithm improve its behavior progressively through trial and error rather than using solely data.¹⁷⁵⁻¹⁷⁷

Deep learning (DL) is a subfield of ML that uses artificial neural networks (ANNs) composed of multiple layers to automatically learn complex representations of data. ANNs are inspired by the structure of human brain that communicates and transmit information to several parts of the body through neurons. In a similar way, a neural network processes input data by combining features, assigning weights, adding bias terms, and applying activation functions to produce an output. After iterative training, the weights and biases are adjusted to minimize prediction errors, enabling the ANNs to identify, classify, or generate patterns directly from raw data.^{178,179}

2.4.2. Overview of AI applications in drug discovery

AI is an innovative approach that play a crucial role in modern drug discovery. With its robust data processing and learning capabilities, it can extract meaningful patterns from vast volumes of chemical and biological data. AI-based drug discovery promise to accelerate and enhance efficiency, accuracy, and predictive power of drug discovery process.¹⁸⁰

AI approaches is being successfully applied in structure-based drug design. Specifically, DL models are accelerating molecular docking by training algorithms to predict ligand-protein complexes. Protein information can be represented at different levels of complexity. Protein sequence as strings, tertiary conformation through graphs and voxels, and surface using point clouds and meshes. Each representation is suited to distinct neural network

architectures. Similarly, ligands can be encoded as strings, molecular graphs or surface representations. These representations allow models to detect potential binding sites, predict binding affinities, and evaluate ligand–target complementarity with high precision.¹⁸¹

In addition, data quality and availability play major role in the reliability of SBDD. For instance, inaccurate or incomplete protein structures can lead to unreliable docking predictions and misleading conclusions. To overcome this challenge, AI-driven tools have been developed to enhance protein structural information. In particular, the release of AlphaFold2,¹⁸² an AI-based tool used in this study, developed by DeepMind, was a groundbreaking discovery by accurately predicting three-dimensional (3D) protein structures from amino acid sequences with atomic-level accuracy. The successful release of AlphaFold2 was followed by its successor AlphaFold3,¹⁸³ which further extended its capabilities by predicting the joint structure of complexes that include proteins, nucleic acids, small molecules, ions, and modified residues, offering new opportunities for integrative modeling. Although AlphaFold has only recently been introduced, numerous studies across biology and medicine have already demonstrated its great potential in drug discovery, protein function prediction, and biomolecular design. In the current study, AlphaFold3 was used to predict the secondary structure of an RNA aptamer, as well as its interaction with the EphA2 receptor.

Another recent DL-based tool, namely MolProphet,¹⁸⁴ has been developed for accelerating lead optimization in the early stages of small molecule drug development. The platform integrates multiple AI-driven modules covering key aspects of both structure- and ligand-based drug design. These include binding site detection using DL, LBDD using a graph neural network for 2D and 3D similarity analysis, AI docking based on geometric DL (e.g. a small

molecule to a flexible receptor), and molecule optimization using scaffold hopping, fragment and R-group exploration. Additionally, MolProphet provides a suite of analysis tools including oral bioavailability assessment and structural alerts, binding mode assessment, ADMET prediction, as well as visualization of structural alerts for toxicity or metabolic instability. In this study, we leveraged MolProphet to assess 2D and 3D molecular similarity between a library of designed molecules and a reference compound. The approach trains a neural network with the vector search Milvus algorithm¹⁸⁵, a purpose-built data management system to efficiently manage large-scale vector data, on a large dataset of ligands and their known properties, then predicts how closely new molecules resemble a lead compound in terms of structural features and physicochemical behavior.

Furthermore, ML models have been developed able to re-rank docking poses more accurately compared to classical scoring functions, while DL models more accurately predict binding free energies after training on experimental data. This is achieved either by optimizing an existing scoring function (e.g., refining the weights of an empirical function) or by developing a new scoring function taking the three-dimensional structure of a complex as input. Importantly, these AI-based methods can process and evaluate millions of compounds *in silico*, making it possible to explore chemical libraries at a scale previously unattainable with conventional computational approaches.¹⁸⁶

Among the most challenging tasks in computer-assisted drug discovery is the generation of novel chemical entities possessing desired properties (*de novo* design). Generative DL approaches for *de novo* design are usually applied to produce molecules in the form of molecular graphs or strings such as SMILES.¹⁸¹

Several studies have focused on using various ML and DL algorithms, to analyze large data sets of drug and toxicity information to identify the potential toxic effects of drugs during development. These models have the potential to accelerate drug development significantly by identifying toxicities early in the development process, thereby avoiding the need for expensive and time-consuming animal testing. Additionally, AI-based toxicity prediction models can be used to prioritize compounds for further testing and identify new drug targets and mechanisms of toxicity.¹⁸⁷

2.4.3. Current Challenges and Limitations

Although many advantages and applications might be available using AI, there are still many challenges and limitations that need to be addressed. The drug discovery pipeline is inherently complex and lengthy, involving multiple stages such as target identification, hit discovery, lead optimization, preclinical testing, and clinical trials. Each stage introduces its biological variability, experimental uncertainty, and data heterogeneity, making the comprehensive control of all these variable factors through AI highly challenging.

Another challenge is that the success of AI approaches is heavily dependent on the availability of large, high-quality, well-structured, and comprehensive datasets. However, many datasets suffer from limited accessibility, poor quality, and inconsistency, all of which compromise their reliability and therefore the reliability of the AI outcomes.¹⁸⁸

AI models rely on complex algorithms and are frequently described as “black boxes” due to the difficulty in understanding how they obtain their decisions,

as mentioned earlier. This lack of transparency poses significant challenges in securing regulatory approval for AI-driven drug discovery tools, as it becomes difficult to prove the accuracy and reliability of the model's predictions and to ensure that decisions are not influenced by unrelated correlations between input and output values within the training dataset. Furthermore, this opacity can lead to distrust in the model's predictions, especially when they contradict the expectations of clinicians or researchers.¹⁷⁶

2.5. Nuclear Magnetic Resonance (NMR) Used to Investigate RNA-Ligand Interaction/Binding

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful analytical tools and has been widely used to analyze in quantitatively and qualitatively way compositional and structural information of various molecules. This method was discovered in mid 1940s by two independent groups, Edward Mills Purcell and Felix Bloch, who were also awarded with the Nobel prize in 1952 in recognition of their discovery of NMR.^{189,190} It has played an important role in structure elucidation for organic and inorganic compounds and has been widely implemented to provide structural, kinetic and dynamic information for biological macromolecules, such as proteins, DNA, and RNA.^{191–194}

To understand the principles behind NMR, some basic concepts, including the nuclear spin and spin distribution, chemical shift, and spin-spin coupling, have been briefly introduced below.

Nuclear spin and spin distribution: NMR spectroscopy relies in the fact that atomic nuclei have spin I . To calculate the spin of a nucleus, the following rules are used:

- If the number of neutrons and protons are even, then the nucleus does not have spin
- If the number of neutrons and protons are odd, then the nucleus has an integer spin
- If the sum of the number of neutrons and protons is odd, then the nucleus has a half-integer spin.

A nucleus of spin I has $2I + 1$ degenerate states. When there is no external magnetic field, the degenerate states are at the same energy level, however, when an external magnetic field is applied, the degeneracy will be broken, and the nucleus will interact with the external magnetic field giving rise to different energy levels. The splitting between nuclear spin levels is called the nuclear Zeeman splitting and the different energy levels can be characterized by the magnetic quantum number m . The value of m is limited to $-I$ and I in integer steps. For instance, ^1H has a half-integer spin, $I = 1/2$, which will give two different energy levels ($2 \times (1/2) + 1 = 2$), state α with $m = 1/2$ and state β with $m = -1/2$, according to Zeeman splitting after applying an external magnetic field. The spin population at different energy levels are distributed according to Boltzmann distribution:

$$\frac{N_{\beta}}{N_{\alpha}} = e^{-\frac{\Delta E}{kT}} \quad (2.39)$$

where N_{α} is the numbers of spins in the ground “spin up” state and N_{β} is the numbers of spins in the excited “spin down” state, ΔE is the energy difference between two states, k is the Boltzmann’s constant ($1.381 \times 10^{-23} \text{ J/K}$), and T is the temperature.

The lower energy spin can jump to a higher energy state with the use of radiation. The energy difference between two states is:

$$\Delta E = h\nu_0 \quad (2.40)$$

where h is the Planck's constant (6.625×10^{-34} Js), and ν_0 is the Larmor frequency.

In an external magnetic field, nuclear spins precess around the field axis with a characteristic motion known as Larmor precession. The corresponding precession (Larmor) frequency is given by:

$$\nu_0 = \frac{\gamma B_0}{2\pi} \quad (2.41)$$

where B_0 is the strength of the external magnetic field and γ is the gyromagnetic ratio of the nucleus.

The precession of the net magnetic moment generates a time-dependent transverse magnetization component, which forms the basis of NMR detection. A radiofrequency coil positioned in the xy -plane, perpendicular to the main magnetic field along the z -axis, detects this oscillating transverse magnetization as an induced voltage. The resulting signal, known as the free induction decay (FID), reflects the sum of all precessing nuclear dipoles in the sample. Fourier transformation of the FID converts this time-domain signal into the frequency domain, yielding the conventional NMR spectrum in which chemical shifts and peak intensities are analysed.

Chemical shift: When an external magnetic field is applied, nuclear spins split into discrete energy levels and resonate at characteristic frequencies. As described above, the resonance frequency ν_0 , depends on the magnetic field strength B_0 and the gyromagnetic ratio γ of the nucleus. However, electrons

surrounding the nucleus also generate local magnetic fields, which oppose the applied field. As a result, the effective magnetic field experienced by the nucleus is reduced compared to the externally applied field.

Because chemically distinct nuclei reside in different electronic environments, the degree of this electronic shielding varies from one site to another. Consequently, nuclei in different chemical environments resonate at slightly different frequencies. This phenomenon is known as the chemical shift (δ) (Figure 2.15).

The absolute resonance frequency of a nucleus depends on the strength of the external magnetic field, meaning that identical chemical environments would appear at different frequencies on spectrometers operating at different field strengths. Reporting absolute frequencies is therefore impractical. To overcome this, chemical shifts are expressed relative to a standard reference compound. For ^1H and ^{13}C NMR, tetramethylsilane (TMS) is commonly used as the reference, whose chemical shift is defined as 0 ppm.

Chemical shifts are reported in parts per million (ppm) to provide a field-independent measure of the relative frequency difference between a nucleus in the sample and the reference. The chemical shift is calculated as:

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{ref}}}{\nu_{\text{ref}}} \times 10^6 \quad (2.42)$$

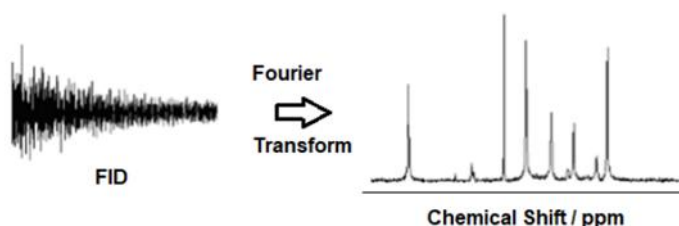


Figure 2.15. NMR spectrum in frequency-domain obtained after the Fourier Transform of the detected signal FID in time-domain.

Spin-spin coupling: The effective magnetic field experienced by a nucleus is also influenced by nearby nuclear spins within the same molecule. Depending on their spin state and the orientation of their magnetic moments, these neighboring nuclei can either increase or decrease the local magnetic field at the observed nucleus. This phenomenon, known as spin–spin coupling, leads to the splitting of NMR signals into characteristic multiplets.

Spin–spin coupling is an indirect interaction transmitted through the electrons of the chemical bonds connecting the two nuclei. The magnitude of the splitting is quantified by the spin–spin coupling constant, J . Notably, the J coupling constant is independent of the external magnetic field strength, in contrast to the resonance frequency itself.

The multiplicity of a signal is determined by the number of chemically bonded neighboring nuclei that couple to the observed nucleus. For simple first-order systems, the splitting pattern follows the familiar $(n + 1)$ rule, where n is the number of equivalent vicinal nuclei. The magnitude of the coupling depends on the number of bonds separating the interacting spins. Typically, couplings transmitted through one to three bonds are readily observed, but longer-range couplings can appear when electrons are delocalized, such as in conjugated systems.

Together, chemical shift and spin–spin coupling provide essential information in NMR spectroscopy, allowing detailed insight into the electronic environment and connectivity of nuclei within a molecule.

Understanding the principles of how a target interacts with a ligand allows the development of additional more potent compounds that can serve as drug

candidates.¹⁹⁵ The process of identifying and optimizing such drug candidates is usually a time consuming task and consists of several steps. One major step comprises the identification of molecules with binding activity for the target and requires the screening of a compound library. Classic NMR methods such as 1D-¹H NMR and Water-LOGSY have emerged as powerful techniques to understand the binding process at a molecular level in solution and have proven to be useful for the identification of new bioactive substances. The first is a macromolecule-based approach and focuses on the NMR signals of the target, typically a protein or a nucleic acid. The second is a ligand-based approach that monitors signals and utilizes the nuclear Overhauser effects (NOE) between the target and the ligand.

In the present study, these approaches are applied specifically to RNA-small molecule systems, where both methods provide complementary insights about binding. As will become clear in the following, the two approaches are complementary and have different advantages. To fully understand the results of such experiments, it is necessary to introduce some fundamental NMR concepts that underpin these techniques, including NOEs, chemical shift perturbations, and line broadening.

2.5.1. Fundamental NMR Concepts Relevant to This Work

The Nuclear Overhauser Effect (NOE)

The nuclear Overhauser effect or enhancement (NOE) is one of the most extensively and intensively used NMR parameters to determine molecular 3D structure. In contrast to chemical shift and J-coupling, which are effects of the electron shell, NOE is a relaxation process caused by a direct dipole–dipole

interaction between nuclei. When two protons are close in space, fluctuations of their dipole–dipole interaction modulate the relaxation pathways of each nucleus. As a result, perturbing the magnetization of one spin can induce changes in the steady-state population of the other. This leads to an enhancement (or reduction) of the observed NMR signal intensity, a process referred to collectively as the NOE.¹⁹⁶

The magnitude of the NOE has an r^{-6} dependence on the spatial separation between the nuclei. Because of this steep dependence, only nuclei separated by roughly 5 Å or less contribute significantly to NOE signals making the NOE one of the most powerful sources of short-range structural information in solution-state NMR.¹⁹⁷

In addition to its structural utility, the NOE is highly sensitive to molecular tumbling. Small, fast-tumbling molecules typically display negative NOEs, whereas large, slowly tumbling macromolecules exhibit positive NOEs. This contrast forms the basis of many ligand-observed NMR approaches. For example, in RNA–ligand systems, bound ligands generally show positive NOEs characteristic of the RNA–ligand complex, whereas unbound ligands maintain negative NOEs. Thus, NOEs provide a sensitive and unambiguous probe for detecting binding events.

Beyond binding detection, NOE provide quantitative structural restraints using the intensity of an NOE cross-peak, which are used in structure determination and model validation. In the context of this thesis, NOE-derived distances will be used to evaluate and validate molecular simulations of RNA molecules performed using different force fields, to evaluate whether the simulated ensemble is consistent with experimentally accessible internuclear proximities.^{196,198}

Chemical Shift Perturbations

In the bound state, the protons of a ligand alter local electronic and magnetic environments exhibiting chemical shift perturbations (CSPs). In RNA-ligand complexes, these perturbations are driven by π - π stacking interactions, proximity to imino or amino protons in helical elements, changes in solvation upon burial within pockets, and electrostatic interactions with the phosphodiester backbone.¹⁹⁸

Line Broadening

Upon binding two effects give rise to line broadening in NMR spectra. The first involves the increased rotational correlation time when the ligand transiently associates with much larger RNA and the second concerns chemical exchange contributions when the bound and free states interconvert on timescales comparable to the transverse relaxation process. Specifically, structural motifs generally impose slower tumbling and therefore stronger broadening, while flexible regions may produce weaker effects. Overall, strong broadening of specific ligand resonances usually reflects localized interactions with structured RNA elements.¹⁹⁸

2.5.2. *Macromolecule-based Approach (1D-¹H)*

1D¹H experiments provide crucial information regarding the binding of a ligand to an RNA molecule. In this experiment, all proton spins in the sample are excited simultaneously with a broadband radiofrequency pulse, and the resulting free induction decay (FID) is recorded and transformed into a frequency-domain spectrum. Each resonance corresponds to a distinct proton environment, with its chemical shift, line width, and intensity reporting on local structural and dynamic properties.

In RNA-ligand interaction studies, 1D ^1H can be used as a macromolecule-observed method. Upon binding, RNA perturbs the ligand's local magnetic environment and its dynamic properties, giving rise to CSPs changes or induce line broadening. These effects are detectable at very low RNA concentrations, allowing fast screening of binding events and can suggest which regions participate in the interaction.

Although 1D ^1H NMR is valuable for rapid detection of ligand binding, the technique has several limitations. Proton chemical shift dispersion can be limited, particularly in RNA molecules, where ribose and base protons are in narrow spectral regions. In this way, overlapping peaks can overlap small CSPs or make it difficult to assign affected residues. 1D spectra provide only indirect information about the binding site but they don't define the precise interaction mode or geometry. In addition, weak binders can be ignored due to non-detectable CSPs, while aggregation or degradation in RNA samples can broaden resonances and diminish the ability to detect ligand-induced changes. For these reasons, 1D ^1H NMR is often used as a first-line screening method to detect the presence of binding, while complementary techniques such as Water-LOGSY are used to elucidate the interaction in more detail.

2.5.3. *Ligand-based Approach (Water-LOGSY)*

The water-LOGSY (water-ligand observed via gradient spectroscopy) method is a ligand-based NMR experiment used to detect weak to moderate RNA-ligand interactions (K_D in the high μM to low mM) in solution. The technique relies on magnetization transfer from bulk water to small molecules and from the macromolecule to bound ligands via NOEs in selective manner (**Figure**

2.16). In this experiment, the resonances of non-binding compounds appear with opposite NOE signs and tend to be weaker than those of the interacting ligands. In contrast, bound ligands acquire magnetization primarily from the macromolecule through intermolecular NOEs, resulting in positive signals. The idea for this technique was based on many experimental observations and the sign inversion between bound and unbound ligands provides an effective way for identifying interacting molecules, even in complex mixtures.^{189,198,199}

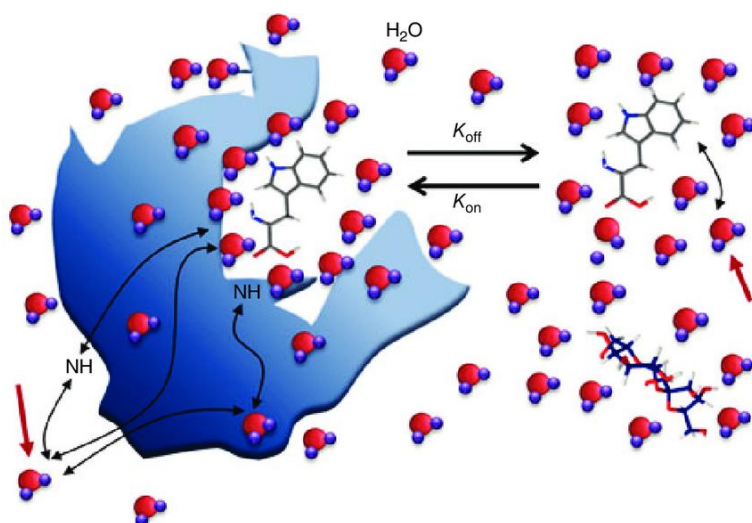


Figure 2.16. Water-LOGSY principles. The target is shown with buried cavities and the active binding site. The ligand is shown in the bound and free states. Water resonances are selectively irradiated. The resulting water magnetization can be transferred to ligands interacting with the receptor and to free molecules. The different transfer pathways are illustrated by arrows. The tumbling properties of the molecules receiving the water magnetization modulate the sign of the ligand signals observed in the final Water-LOGSY spectra. Receptor-bound ligands, behaving as part of a large molecular complex, yield Water-LOGSY signals in the same phase, whereas opposite signals arise from fast-tumbling free molecules. Adapted from Gairi et al.²⁰⁰

Despite the advantages of water-LOGSY technique, it faces several limitations that must be considered. LOGSY signal intensities are primarily qualitative and thus they do not correlate with binding affinity. Moreover, aggregated ligands can display false positive signals even in the absence of specific binding, because aggregates mimic slow-tumbling macromolecules. In addition, some hydrophobic ligands could appear positive due to nonspecific interactions with RNA surfaces or solvent-exposed regions. Overall, Water-LOGSY is best used as a qualitative screening method, ideally in combination with 1D ^1H NMR.¹⁹⁸

Table 2. 1. Comparison Table: 1D ^1H NMR vs Water-LOGSY

Feature	1D ^1H NMR	Water-LOGSY
Detection mode	Macromolecule-based	Ligand-based
Readout	CSPs, line broadening in RNA signals	Positive/negative sign of ligand signals
Binding sensitivity	Moderate-to-strong binders	Weak-to-moderate binders
Best application	Binding confirmation Identification of affected RNA regions	Rapid screening of candidate ligands
Limitations	Overlap, low resolution, exchange broadening	False positives, lack of quantification, exchange effects
Complementarity	Identifies <i>where</i> binding affects the RNA	Identifies <i>if</i> binding occurs

3.RESULTS

3.1. Targeting Glioblastoma Stem Cells (GSCs) via EphA2 using an RNA Aptamer

In this chapter, we will explore the RNA aptamer A40s as a ligand and we will study its interaction with the protein receptor EphA2. EphA2 is overexpressed in GBM, particularly within GSCs, where it contributes to tumor growth and resistance to therapy. This makes EphA2 a potential target for GBM treatment. The RNA aptamer A40s is a high affinity ligand of EphA2 that inhibits GSC growth stemness and migration, thus representing a promising RNA-based candidate for therapeutic development.

First, we unveil the 2D structure of A40s using several state-of-the-art RNA 2D prediction algorithms, providing a view of its structural organization. Subsequently, we predict the 3D structure of A40s and its complex with EphA2, by integrating advanced computational modeling, MD simulations, and experimental data.

Our results suggest that A40s interacts with the LBD of EphA2 through specific contact points. The predicted binding interface highlights structural features of the aptamer that are critical for recognition and affinity, thereby providing a mechanistic basis for its biological activity. These insights establish a framework for rational optimization of A40s with the aim of enhancing binding strength and therapeutic potential.

3.1.1. 2D Structure Prediction of A40s

Methods

The primary sequence of A40s was used to predict its secondary (2D) structure. To this aim, eleven different online tools were employed: RNAstructure 6.5,⁷³ RNAfold 2.6.3,⁷⁴ CentroidFold 0.0.16,⁷⁹ NUPACK 4.0,⁷⁵ Vfold2D 2.5,⁸⁵ Ipknott++ 2.2.1,¹⁵ MC-Fold 8,²⁰¹ KineFold 3,⁸⁰ pKiss 2.3.0,²⁰² sFold 2.2,⁷⁶ and AlphaFold 3.¹⁸³ These servers are based on at least one of the following methods: 1) Minimum Free Energy, which predicts the most probable structure within a folding ensemble (RNAstructure, RNAfold, NUPACK); 2) probabilistic approaches, which use statistical models to provide probabilities for base-pairing interactions (CentroidFold, sFold); 3) experimental data integration, which incorporates experimental data to guide predictions (RNAstructure, Vfold2D); 4) coarse-grained approach that simplifies the RNA structure for computational efficiency (Vfold2D); 5) dynamic folding and kinetics, which simulate the RNA folding process over time, offering insights into folding pathways and kinetics (Kinefold); 6) handling pseudoknots, which are specifically designed to predict RNA structures with pseudoknots (IPknot++, McFold, KineFold, pKiss); and 7) machine learning-based probabilistic approaches, which apply deep learning algorithms trained on large datasets to predict complex structures with high accuracy (AlphaFold 3). All available parameter settings provided by each tool were considered for 2D structure prediction. For AlphaFold 3, which allowed for both isolated RNA and RNA-protein complex predictions, we included 2D structures of A40s predicted alone as well as those predicted in complex with EphA2. In cases where multiple structures with varying free energies were obtained, we selected the top three solutions. These servers

allowed us to capture a broad spectrum of 2D predictions generating a total of 46 models, which were then clustered based on their structural similarity using in-house Python scripts that employed affinity propagation.

Results

After generating 46 structures using eleven servers and performing clustering, we proceeded with the analysis of the resulting models. The most populated cluster (13 out of 46 predictions) represented a hairpin structure, consisting of an 8-base-paired stem and a 14-nucleotide loop (dot-bracket notation: “(((((((.....)))))))))”) or a 9-base paired stem and a 12-nucleotide loop (dot-bracket notation: “(((((((.....)))))))))”), while the second most populated cluster (10 out of 46 structures) represented a left-bulged hairpin model (bulged), including an 8-base paired stem, a 2-nucleotide bulge, two base pairs between G11-C22 and A12-U21, and an 8-nucleotide loop (dot-bracket notation: “(((((((..((.....)))))))))”) (**Figure 3.1**). **Figure 3.2** lists the employed tools and their corresponding solutions. Notably, employed software such as Ipknnot++ and KineFold, which account for pseudoknots, did not yield any pseudoknot 2D structure of our aptamer in the top three ranked solutions.

MC-Fold 8, KineFold 3, pKiss 2.3.0, sFold 2.2, and AlphaFold 3. All available parameter settings provided by each tool were considered for 2D structure prediction. For AlphaFold 3, which allowed for both isolated RNA and RNA-protein complex predictions, we included 2D structures of A40s predicted alone as well as those predicted in complex with EphA2. In cases where multiple structures with varying free energies were obtained, we selected the top three solutions, resulting in a total of 46 predicted 2D structures.

3.1.2. *In Silico 3D Structure Prediction of A40s*

Methods

For the tertiary structure prediction of the two selected 2D configurations - hairpin (dot bracket “((((((((.....)))))))))”) and bulged (dot-bracket “((((((((..(.....)))))))))”), the aptamer’s sequence and dot-bracket notation were used as input. Structure predictions were performed using three well-established programs for RNA 3D predictions: RNAComposer,⁹⁰ 3dRNA,⁹¹ and Vfold3D⁸⁴, all with default parameters. The accuracy of these methods has been validated in previous studies on five aptamers, with all of them showing high accuracies for the 3D structure prediction of short aptamers (<40 nucleotides).²⁰⁴ The use of multiple tools allowed us exploring a variety of conformations and evaluating the convergence.

To analyze structural diversity, the conformations from the two models were clustered separately using two complementary strategies: i) clustering based on backbone root mean square deviation (RMSD), applying the GROMOS²⁰⁵ algorithm with a 4.5 Å cutoff, and ii) clustering based on the ensemble RMSD (eRMSD) metric through the Barnaba software.²⁰⁶ Unlike RMSD, eRMSD

focuses solely on the relative positions and orientations of nucleobases, making it particularly sensitive to base-pairing and base-stacking interactions. This makes eRMSD particularly suited for capturing RNA conformational variability and for distinguishing structurally meaningful substates within the ensemble. Consequently, it offers a complementary perspective to our backbone-oriented RMSD-based clustering. For the eRMSD-based clustering, G-vectors were first computed from the aptamer conformations to capture base-level structural features. These vectors were then projected onto the first two principal components derived from a principal component analysis (PCA), and clustering was performed using the DBSCAN algorithm by setting *min_samples* = 2 and *eps* = 0.9. Setting the minimum number of samples to 2 was essential due to the limited size of the dataset. Notably, smaller *eps* values led Barnaba to fail in detecting any clusters, justifying the chosen parameters.

Results

For each of the two predicted 2D structures, namely hairpin and bulged, the top three scored conformations from RNAComposer and 3dRNA were selected. Additionally, a single conformation generated by Vfold3D was used for the hairpin structure, while two conformations, being the only ones produced by Vfold3D, were used for the bulged structure. A total of seven hairpin and eight bulged 3D conformations were identified.

Regarding the hairpin model, clustering based on RMSD grouped all conformations into a single, cohesive cluster characterized by substantial overlap in both the stem and loop regions. In contrast, the bulged model yielded three distinct clusters, with the most populated cluster accounting for 50% of the sampled conformations. While the conformations within this

dominant cluster exhibited consistent alignment in the stem region, variations were observed in the orientation of the unpaired bases, reflecting structural heterogeneity in the bulge region (**Figure 3.3**).

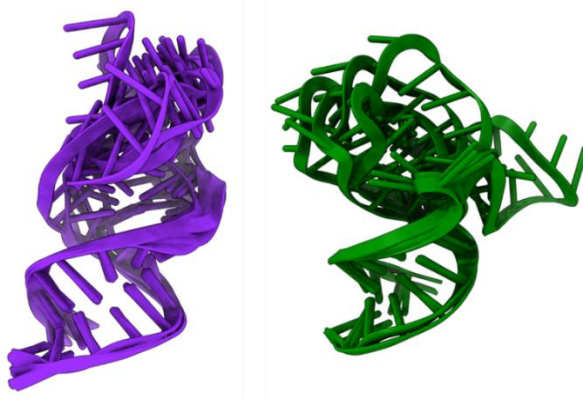


Figure 3.3. Ensemble of 3D conformations of hairpin structure (left, violet cartoon) and bulged structure (right, dark green cartoon) obtained using RNAcomposer, 3dRNA, and Vfold3D.

To further validate these observations, we performed an independent cluster analysis using the nucleic acid-specific eRMSD metric.²⁰⁷ Interestingly, the centroids identified through eRMSD closely matched those obtained with RMSD clustering. However, eRMSD classified a larger proportion of structures as outliers (**Figure 3.4A**), likely due to its heightened sensitivity to local base orientation and the relatively small sample size.

Moreover, while RMSD analysis suggested higher structural conservation within the hairpin ensemble, eRMSD-based clustering revealed reduced variability among bulged structures (**Figure 3.4B**). This apparent discrepancy can be attributed to the distinct structural features emphasized by the two metrics: the hairpins contain fewer base pairs and a more flexible loop, which

allows diverse base orientations across predictions, leading to greater variation in eRMSD. In contrast, bulged structures, although more variable in backbone geometry, tend to maintain more consistent local base orientation due to two additional base pairs involving nucleotides G11-C22 and A12-U21.

This interpretation is further supported by cross-eRMSD matrices computed for all hairpin and bulged structures, excluding the loop region (**Figure 3.4C**). When this loop is excluded, the analysis aligns more closely with the RMSD-based results (**Figure 3.4 D and E**), suggesting that loop flexibility is a major contributor to the divergence among 3D predictions, an aspect that warrants further investigation in a dynamic context.

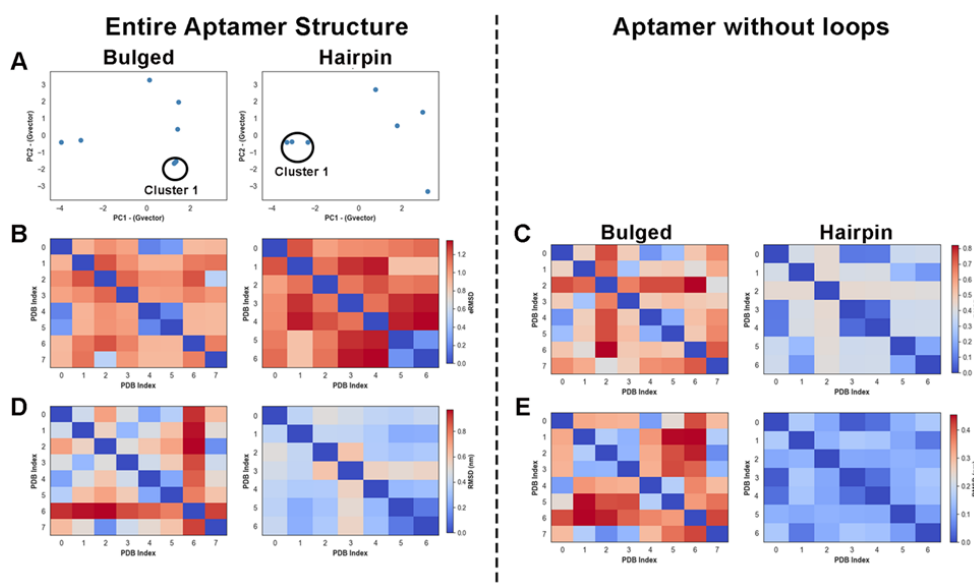


Figure 3.4. eRMSD-based analysis of 3D A40s structure predictions, alternatively including (A, B, D) or excluding (C, E) the aptamer's loop (nucleotides 9-22). A) Projection onto PC1 and PC2 of G-vectors computed from the PDBs provided by the 3D-conformation predictor tools for bulged (left) and hairpin (right) conformations.

B) and C) Cross-eRMSD matrix between the PDBs generated by the 3D-conformation predictor tools for bulged (left) and hairpin (right) conformations. D) and E) Cross-RMSD matrix between the PDBs generated by the 3D-conformation predictor tools for bulged (left) and hairpin (right) conformations, based on the GROMOS algorithm.

3.1.3. *Molecular Dynamics of A40s in Solution*

Methods

The stability and energetics of the A40s hairpin and bulged predicted models were then investigated through 3 μ s full-atom MD simulations in three replicas using GROMACS 2020.6²⁰⁸ with the OL3²⁰⁹ force-field. For the replica simulations all conditions were identical apart from different initial velocities. All simulations were performed on the non-fluorinated RNA sequence of A40s. In fact, the 2'-fluoro-modifications are not expected to significantly affect the aptamer's secondary or tertiary structure, as supported by prior literature.^{62,63}

The leap program available in AmberTools²¹⁰ was used to prepare each system for the MD simulations. The aptamer was described using the OL3²⁰⁹ force field and then solvated in a TIP3P²¹¹ water cubic box with a 12.0 Å edge distance under PBC. 150 mM K⁺ and Cl⁻ ions, modelled with Joung and Cheatham parameters^{211,212}, were randomly placed to neutralize the total charge of the system and mimic physiological conditions. Finally, coordinates and topology files for the whole systems were obtained.

The Verlet cut-off scheme was used for non-bonded interactions neighbor search, whereas the long-range electrostatic interactions were treated using the particle-mesh Ewald summation method (PME)²¹³. Short-range coulombic and van der Waals interactions were treated with a cutoff distance of 12 Å.

Prior to MD simulations, all structures were relaxed with energy minimization followed by subsequent NVT and NPT runs. The energy minimization consisted of two steps performed by using the steepest descent algorithm: i) 20,000 steps, with harmonic restraints of 5,000 kJ mol⁻¹ nm⁻² applied to the aptamer heavy atoms, so that only the solvent was unconstrained; and ii) 50,000 steps during which the entire system was allowed to relax.

To integrate Newton's equations of motion, the leap-frog²¹⁴ algorithm was employed and a time step of 2 fs was chosen, while the LINCS²¹⁵ algorithm was used to constrain bonds involving hydrogen atoms. During the MD equilibration procedure, the system was gradually heated by increasing the temperature with subsequent MD runs in the canonical ensemble (NVT) using the weak-coupling Berendsen¹⁶⁷ scheme. Specifically, three 500 ps NVT steps were performed by gradually increasing the temperature from 100 K up to 300 K. At each step, harmonic restraints were applied to all heavy atoms of the aptamer and were gradually decreased from 1,000 to 500 and finally to 250 kJ mol⁻¹ nm⁻². Subsequently, two NPT equilibration runs of 1 and 5 ns, respectively, were performed to adjust the box volume, using the Berendsen algorithm for pressure coupling. In the first NPT step, restraints of 100 kJ mol⁻¹ nm⁻² were applied on heavy atoms, while no restraints were used in the last equilibration run.

During the 3 µs production runs, the temperature was kept constant at 300 K using the v-rescale¹⁶⁶ thermostat with a time constant of 0.5 ps. The pressure

was isotropically maintained at 1 atm using the Parrinello-Rahman¹⁶⁸ barostat with a time constant of 5 ps and a compressibility of 4.5×10^{-5} bar.

For each aptamer configuration, the RMSD and root mean square fluctuations (RMSF) were analyzed, along with the stability and formation of hydrogen bonds throughout the trajectory. For the hairpin configuration, which showed the highest agreement with experimental data, all frames from the three replicas were clustered using the GROMOS algorithm with a 4.5 Å cutoff. A representative conformation was then selected from the most populated cluster.

Results

At the end of the MD simulations, RMSD calculation of the aptamer's heavy atoms (**Figure 3.5A**) and eRMSD (**Figure 3.5C**) with respect to the initial frame confirmed the general stability of the hairpin conformation (averaged $\text{RMSD}_{\text{hairpin}} = 2.8 \pm 0.4$ Å, averaged eRMSD = 0.85). Subsequent RMSF analysis of the hairpin structure indicated flexibility in the terminal nucleotides, nucleotides 13 and 15–19 across all replicas, and nucleotide 9 only in the first replica, with RMSF values reaching 2.5–3.0 Å (**Figure 3.5E**). Notably, all of these nucleotides are part of the hairpin loop.

Monitoring hydrogen bonds between the bases of the hairpin model revealed the noteworthy formation of a new base pair between nucleotides G11 and C20, which was consistently present in all three replicas (93.1%, 96.0%, and 91.6%, respectively). Additionally, the formation of a noncanonical base pair between nucleotides A12 and G19 was observed in all three replicas, with

frequencies of occurrence of 52.6%, 51.5%, and 49.7%, respectively, which indicates transitioning between paired and unpaired bases (**Figure 3.8A**).

Compared to the hairpin model, the bulged model demonstrated reduced stability across all simulations, as reflected by the RMSD value of the aptamer's heavy atoms, measured at 4.6 ± 1.2 Å (**Figure 3.5B**), and an average eRMSD of 1.12 (**Figure 3.5D**) across the three replicas. In line with this observation, the RMSF values for the terminal and unpaired nucleotides (10 and 16–18) ranged from 4.5 to 7.0 Å, while the expected paired nucleotides (11–12 and 20–22) exhibited lower values between 4.0 and 5.0 Å (**Figure 3.5F**). This structural instability is further emphasized by pronounced changes in the aptamer's 2D arrangement compared to its initial configuration. As shown in **Figure 3.8A**, the three simulations failed to converge to a single conformation, underscoring the low thermodynamic stability of the initial bulged structure. For the sake of thoroughness, we also evaluated the potential overlap in conformational space between the two sets of RNA-only simulations. For each simulated replica, we computed the cross-eRMSD relative to the hairpin structure (for simulations initialized from the bulged conformation) and the bulged structure (for those initialized from the hairpin conformation). As shown in **Figure 3.7**, the resulting plots indicate that no interconversion occurs between the two conformations in any of the replicas over the 3 μs simulation time.

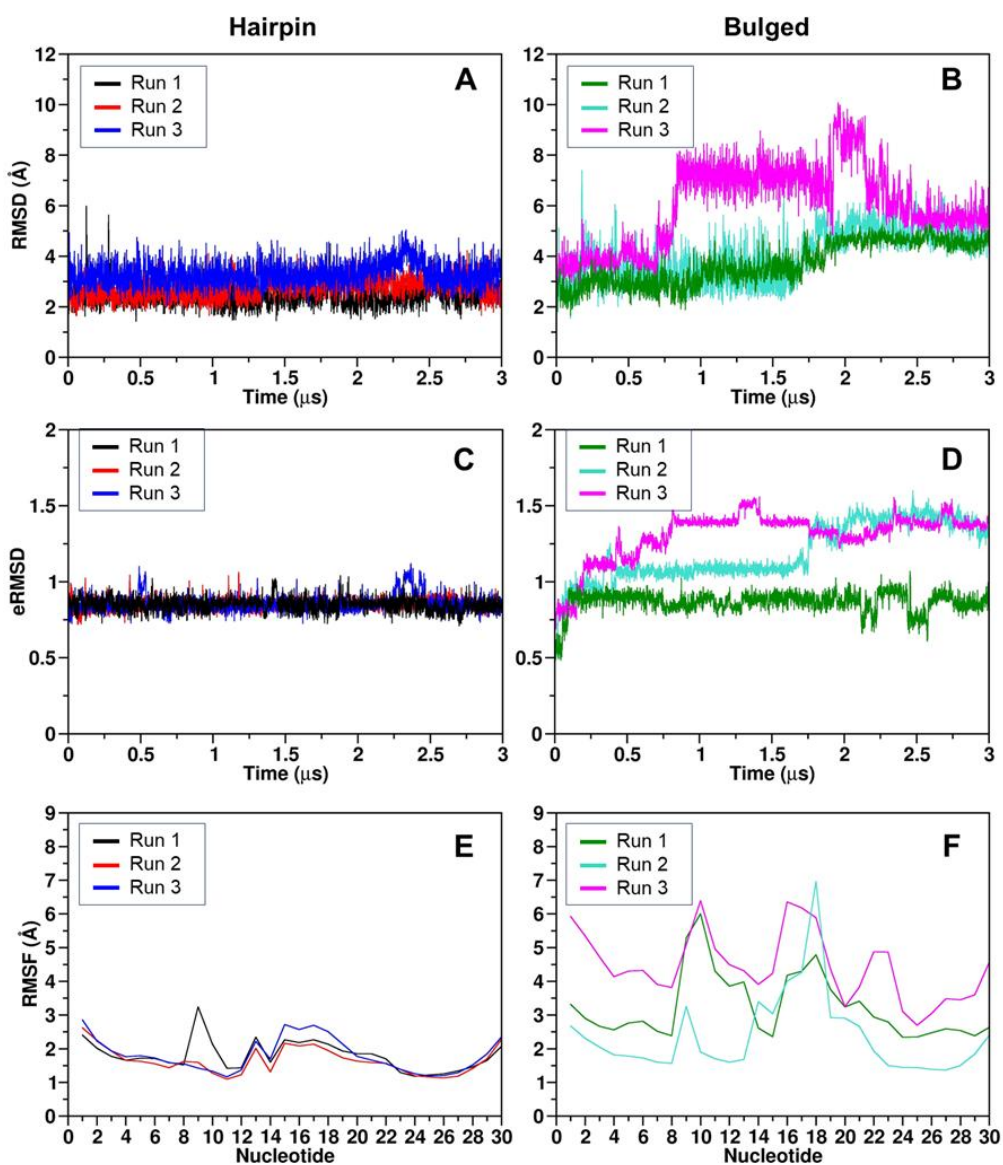


Figure 3.5. RMSD (top panels), eRMSD (central panels), and RMSF (bottom panels) plots of the hairpin (A, C, E) and bulged (B, D, F) models. The values are computed on the A40s's heavy atoms with respect to the initial frame. For the RMSD calculations, the terminal nucleotides were excluded.

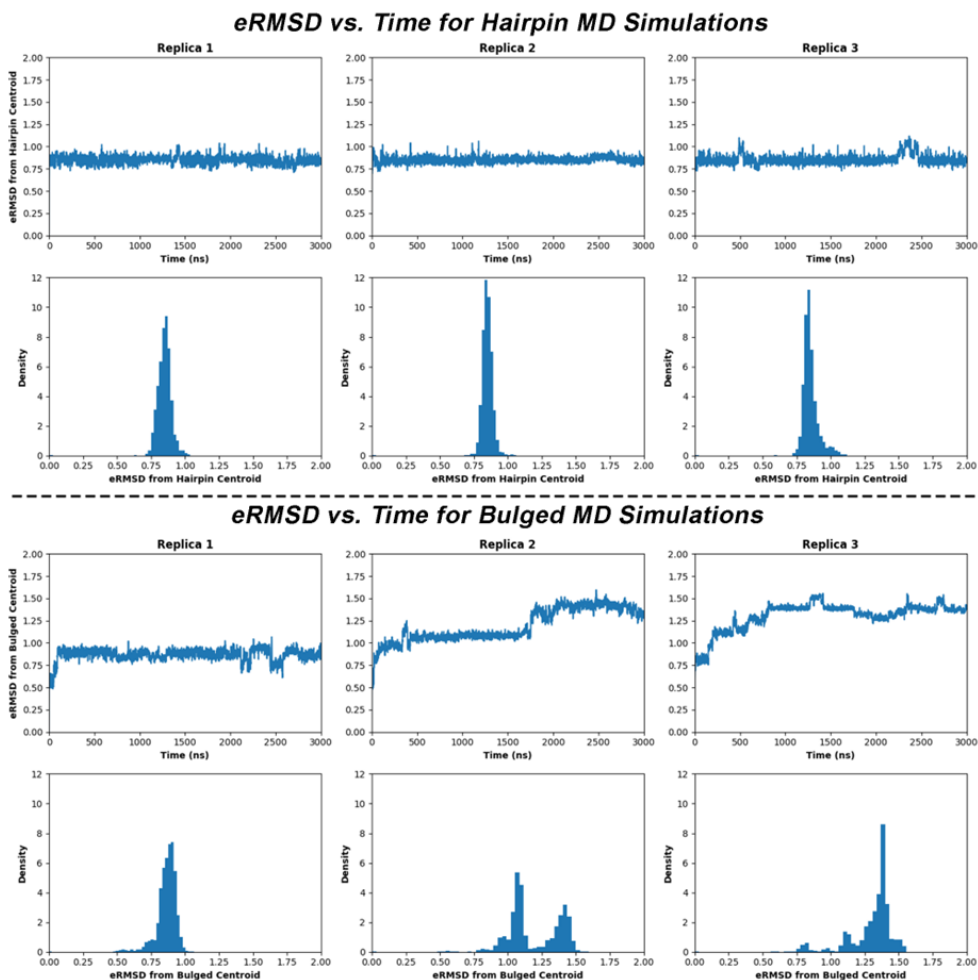


Figure 3.6. Time evolution of the eRMSD¹² and corresponding density histograms computed over three independent RNA-only MD simulations of the A40s aptamer, initiated from either the hairpin (upper panels) or bulged (lower panels) conformations. The initial offset in eRMSD values reflects the use of the respective cluster centroids (from hairpin or bulged ensembles) as reference structures for eRMSD calculation. These centroids undergo slight rearrangements during the initial equilibration phase, which is not included in the plotted trajectories.

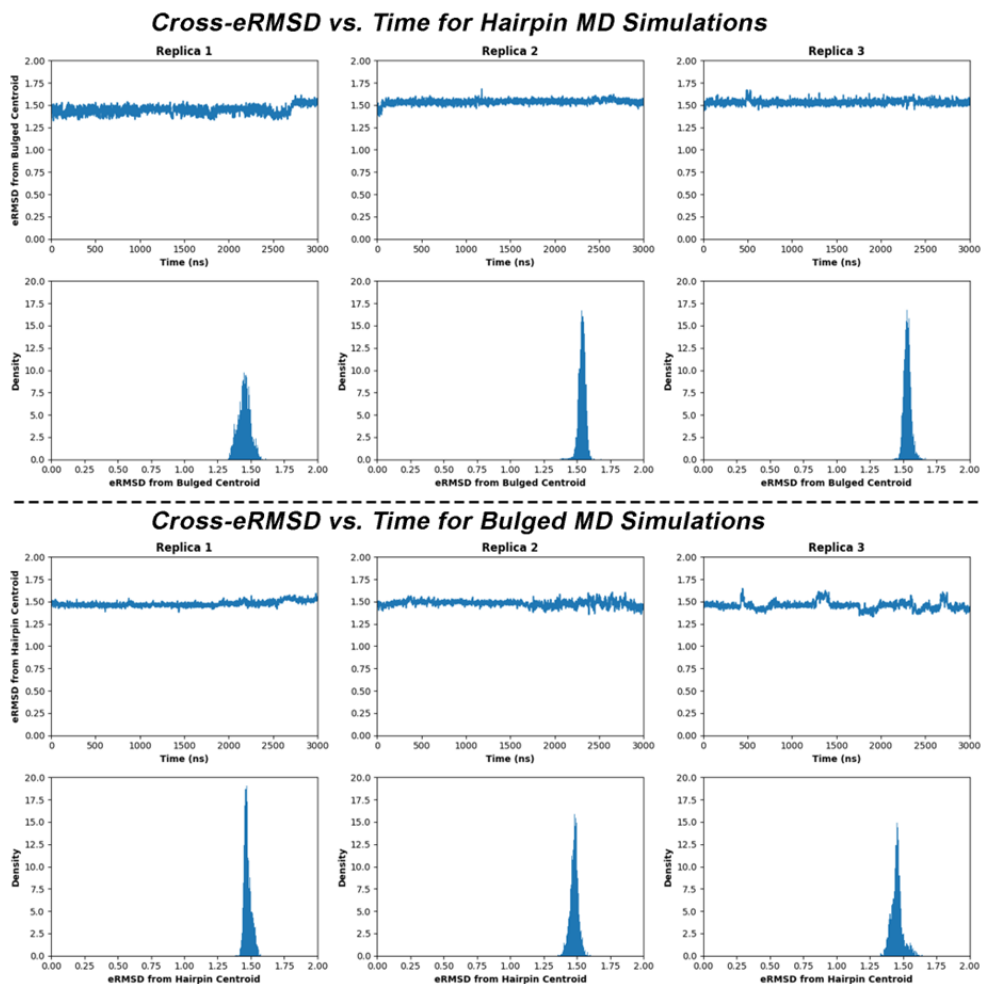


Figure 3.7. Cross-eRMSD evolution (upper row) and density histograms (lower row) computed over the A40s aptamer in three independent replicas of RNA-only simulations initialized from hairpin (upper panels) and bulged (lower panels) conformations. The reference conformations for eRMSD calculations are the centroids of the hairpin structure (when the simulation was initialized from the bulged conformation) and the bulged structure (when initialized from the hairpin conformation), both obtained through clustering.

This observation suggests that the transition between these structural states likely occurs on time scales longer than those sampled here, implying a lower bound for the interconversion rate under equilibrium conditions. Notably, replicas initiated from the same conformation tend to converge toward similar cross-eRMSD values, further supporting consistent sampling within each system.

3.1.4. Validation of MD-predicted A40s Structure using In-Line Probing Experimental Data

To validate the computational predictions, our collaborators (Prof. Mario Mörl, University of Leipzig, Germany) conducted In-Line Probing (ILP) and RNase T1 cleavage experiments according to Regulski and Breaker. ILP results in a preferential cleavage of single-stranded positions and thus offers valuable insights into the 2D structure by revealing band signals at the cleaved sites. While the ILP reaction targets unpaired positions without base specificity, RNase T1 specifically cleaves at guanine residues, with a strong preference for single-stranded regions. To enhance band clarity and simplify the interpretation of the probing results, the longer version of the aptamer was used, A40L, which provides additional cleavage sites. The extended sequence not only minimizes the risk of ambiguous results but also improves RNA stability during handling, streamlining the experimental workflow.

Figure 3.8B illustrates the results of the probing experiments, highlighting signals that indicate unpaired nucleotides. ILP data suggest that nucleotides 1–7, 11–12, and 19–30 are paired, while 8–10 and 13–18 are unpaired; RNase T1 cleavage suggests that nucleotides G15, G16, G18, and G19 are unpaired,

while G4, G7, and G11 are paired. Altogether, these observations support the hairpin MD-predicted A40s conformation. Nonetheless, nucleotide G19 is classified as paired in the ILP lane but seems to be unpaired in the RNase T1 cleavage reaction (**Figure 3.8B**, **Figure 3.9**). This discrepancy can be explained by the fact that unpaired bases can form stacking interactions with the neighbouring bases. If such stacking limits the conformational freedom of G19, in-line-induced cleavage might be strongly reduced. Furthermore, the computational data indicate that G19 transiently forms a noncanonical base pair with A10, with a frequency of occurrence of 50% in the three simulations. Hence, it is possible that such a transient pairing contributes to a partially paired character of G19, as indicated in the ILP analysis.

On the other hand, there is a mismatch between the computational predictions and the experimental data for nucleotides 21–23. Indeed, the computational model predicted these nucleotides to be unpaired, whereas the ILP results showed either very weak signals or no signal at all. Visualization of the A40s conformation revealed potential base-stacking interactions in this region, which could reduce its accessibility for cleavage, resulting in the absence of a signal (**Figure 3.8C**). It is worth noting that minor discrepancies between RNase T1 and ILP reactivity, such as at G19 and nucleotides 21–23, likely reflect differences in strand break mechanisms and local RNA dynamics, as previously observed for other structured RNAs.

In contrast, the different conformations observed at the end of the bulged MD simulations do not align with the experimental data. In particular, the paired G15 in replicas 2 and 3 is inconsistent with the RNase T1 cleavage pattern that exhibits a clear, intense band signal indicative of an unpaired nucleotide, and the paired C13 in replica 3 that was expected to remain unpaired according to ILP data. Altogether, these results provide strong support for the predicted MD

conformation of the aptamer, reinforcing our confidence in the computational hairpin model. It is worth noting that minor discrepancies between RNase T1 and ILP reactivity, such as at G19 and nucleotides 21–23, likely reflect differences in strand break mechanisms and local RNA dynamics, as previously observed for other structured RNAs.²¹⁶

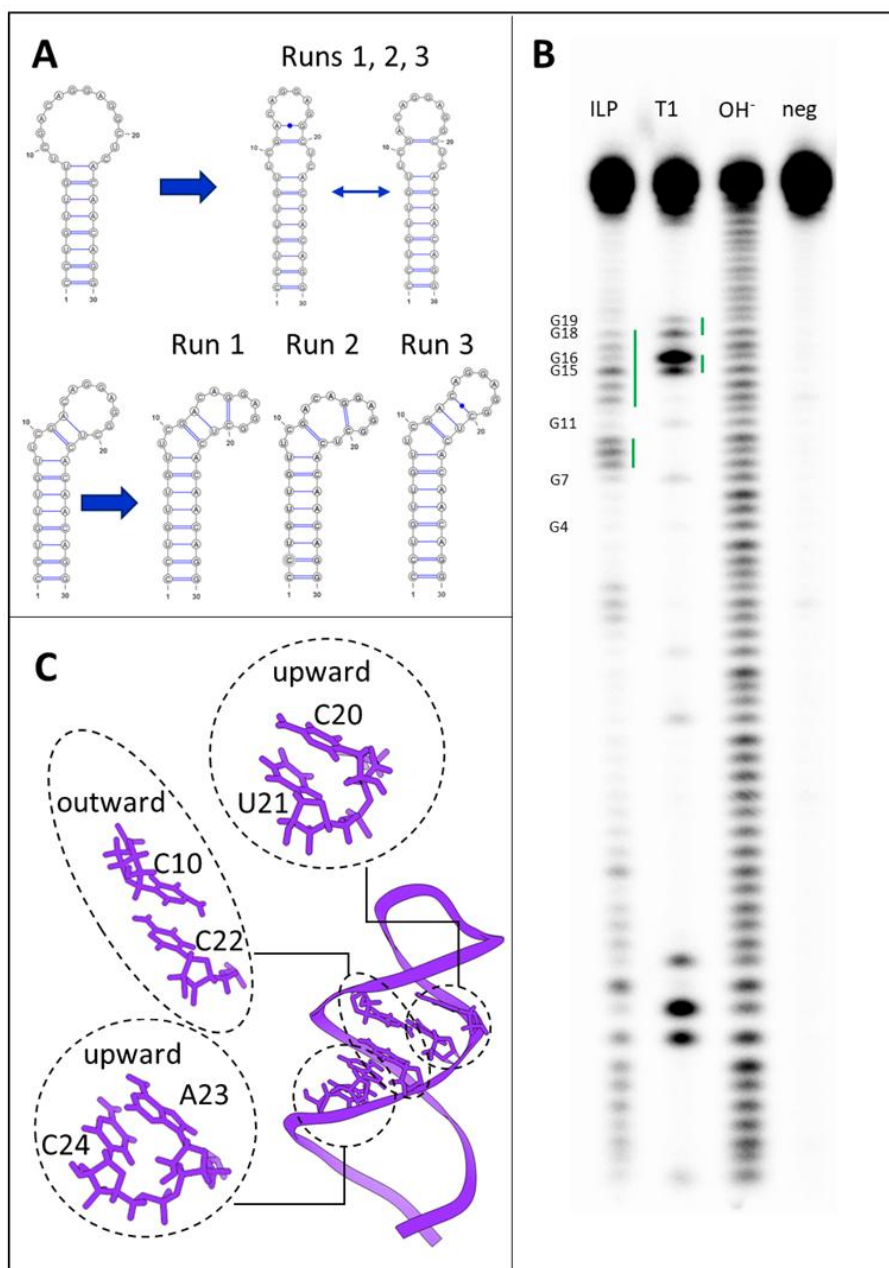


Figure 3.8. A) Changes of the A40s's secondary structures during MD simulations for both the hairpin and the bulged conformation. B) Structure probing results on the full length of the aptamer (A40L), where signals indicate unpaired nucleotides. The unpaired nucleotides of the shorter aptamer A40s are highlighted with green lines. C) Base stacking interactions in the hairpin model.

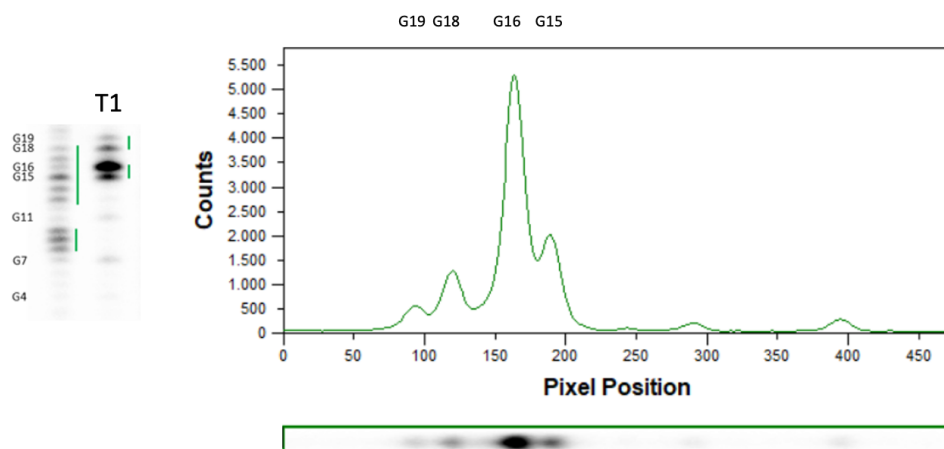


Figure 3.9. Semi-quantitative analysis of RNase T1 cleavage at G19. (Left) Autoradiograph showing cleavage at G residues, with G19 displaying markedly weaker cleavage compared to G15–G16. (Right) Densitometric profile of the gel, where peaks correspond to cleavage events. The weak signal at G19 supports transient unpairing at this position.

3.1.5. *EphA2/A40s Phosphorylation and Interaction*

3.1.5.1. *EphA2/A40s Phosphorylation Using Western blot Data*

To evaluate the ability of A40s to modulate the EphA2 canonical and noncanonical signaling pathways, we used Western blot analysis data on patient-derived GSCs, performed by our collaborators (Prof. Condorelli, University of Naples Federico II, Naples, Italy and Dr. Quintavalle Cristina, Consiglio Nazionale delle Ricerche, Naples, Italy) for the purposes of this project.

The results show that A40s treatment promotes EphA2 phosphorylation at Tyr588, which is associated with canonical EphA2 signaling through its natural ligand ephrinA1 to maintain cellular homeostasis. Concurrently, A40s treatment reduces EphA2 phosphorylation at Ser897, which is instead linked to noncanonical signaling and receptor's oncogenic activity. In addition, A40s treatment also decreased the levels of phosphorylated RSK2 and AKT (pAKT) (**Figure 3.10**), which are both known to activate the oncogenic function of EphA2.

These findings suggest that A40s acts as a functional mimic of the endogenous ligand ephrinA1 and impairs the oncogenic activation of EphA2 receptor, highlighting its potential for GSCs-targeted therapy.

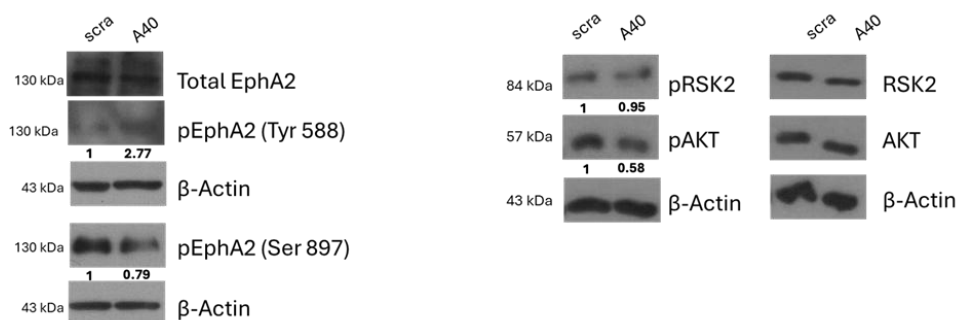


Figure 3.10. Western blot analysis of phosphorylation of EphA2 (left) and associated kinases RSK2 and AKT (right).

3.1.5.2. *EphA2/A40s Binding Using Molecular Docking*

Methods

To elucidate the binding of A40s to EphA2, we initially performed molecular docking using the HADDOCK 2.4²¹⁷ and ZDOCK 3.0.2²¹⁸ servers, both

widely cited in the literature for predicting nucleic acids-protein interactions.^{204,219–225} HADDOCK requires the selection of up to 150 active residues for docking simulations. As the EphA2 protein has 427 surface residues, we focused our docking efforts exclusively on the LBD. In fact, Western blot data suggest that our aptamer mimics ephrinA1 by inducing phosphorylation patterns consistent with canonical signaling. Given that ephrinA1 binds specifically to the LBD of EphA2, we concentrated our docking analysis on this region.

Results

Notably, the lowest-energy poses provided by the two programs converged to the same conformation (RMSD = 0.3 Å), in which A40s adopts an horizontal position on the top of EphA2, with nucleotides 19-22 entering within the surface pocket of the LBD (Figure 7). In addition, we used AlphaFold 3¹⁸³, an advanced machine learning tool designed to predict RNA-protein complexes using only sequence data as input. AlphaFold successfully predicted the binding of A40s to the surface of EphA2, with the loop inserting into the EphA2 structure (**Figure 3.11**). However, it failed to accurately predict the secondary structure of A40s, leading us to exclude this model from further analyses. The predicted complexes of the EphA2/A40s by HADDOCK and ZDOCK were aligned with the natural EphA2/eA1 complex (PDB ID: 3CZU²²⁶), revealing that A40s and eA1 occupy the same binding region, as illustrated in **Figure 3.12**.

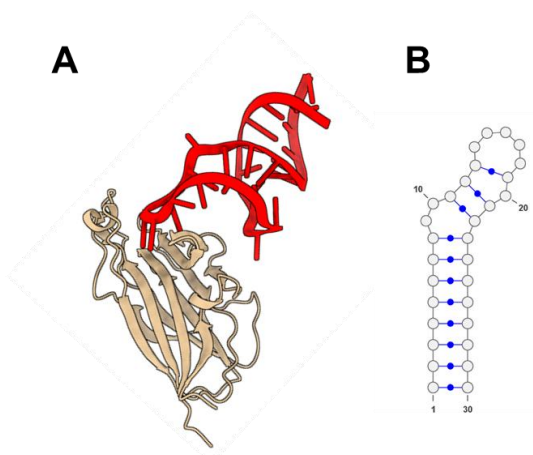


Figure 3.11. A) Prediction of EphA2/A40s complex using AlphaFold 3. B) 2D structure of A40s on EphA2 predicted by AlphaFold 3. The structure was not supported by the experimental data and was therefore excluded from further studies.

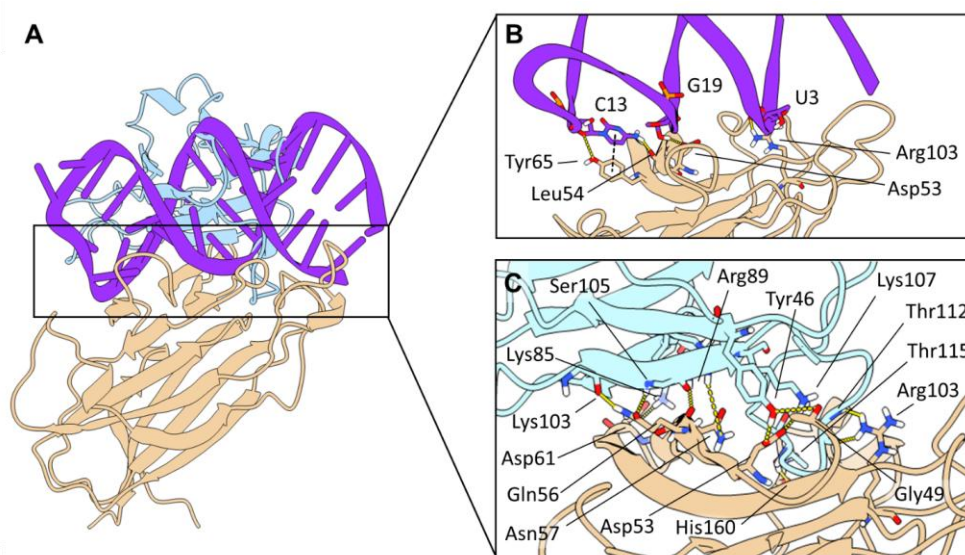


Figure 3.12. A) Superposition of the EphA2/A40s complex predicted by HADDOCK 2.4 and the EphA2/eA1 X-ray complex (PDB code: 3CZU) showing that A40s and the endogenous ligand occupy the same binding surface. Graphical representation of the most important interactions between EphA2 and A40s (B), and between EphA2

and eA1 (C) are shown as insets. The main interacting residues are illustrated as sticks, with specific interactions highlighted by dashed lines. EphA2 is colored in beige, A40s in violet, and eA1 in cyan.

3.1.5.3. Elucidating EphA2/A40s and EphA2/ephrinA1 Interaction Using MD Simulations

After generating the EphA2/A40s complex using molecular docking, we investigated the stability and the energetics of the docking-predicted EphA2/A40s complex by performing 5 μ s all-atom MD simulations using GROMACS 2020.6 software. In parallel, we performed 5 μ s MD simulations of the reference EphA2/ephrinA1 complex to assess whether A40s and ephrinA1 share common key-binding residues. The atomistic model for EphA2/ephrinA1 was retrieved from the PDB ID: 3CZU, which includes the LBD of EphA2 bound to its endogenous ligand, ephrinA1.

Methods

The systems were prepared using the leap program available in AmberTools.²¹⁰ The ff14SB²²⁷ and OL3²⁰⁹ force field were used to describe the proteins and the aptamer, respectively. Each complex was solvated in a TIP3P²¹¹ water cubic box with a 15.0 Å edge distance under PBC. To neutralize the total charge of the EphA2/A40s system and mimic physiological conditions, 33 Na⁺ ions, modelled with Joung and Cheatham parameters,^{211,212} were added. Finally, coordinates and topology files for the two systems were generated. The equilibration and MD production protocols followed were the same as previously described.

Trajectories visualization, RMSD and time series of bonds between nucleotides analyses were performed with the VMD²²⁸ software. The root mean square fluctuation (RMSF) calculations and cluster analysis on the MD trajectory of the aptamer were carried out using specific tools implemented in GROMACS.²⁰⁸

EphA2/A40s

Visualization of the trajectory and plotting of the RMSD of A40s relative to the protein revealed a slight shift of the aptamer at approximately 200 ns of the simulation (**Figure 3.13**) to reach a new arrangement that is conserved for the rest of the simulations. In addition, the overall conformation of A40s itself displayed enhanced stability with a RMSD value of 2.7 ± 0.3 Å along the whole trajectory and monitoring the hydrogen bonds between its base pairs indicated the preservation of its 2D structure throughout the 5 μs (**Figure 3.14**).

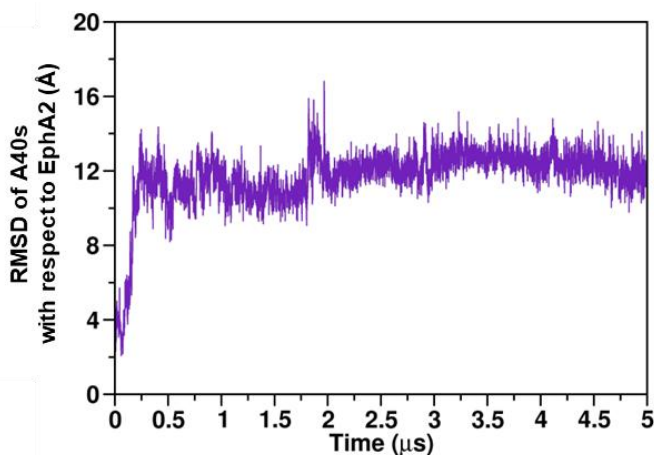


Figure 3.13. RMSD of A40s (calculated on heavy atoms, excluding terminal nucleotides) relative to EphA2 on the EphA2/aptamer complex. The trajectory was aligned on the EphA2 secondary structure C α atoms.

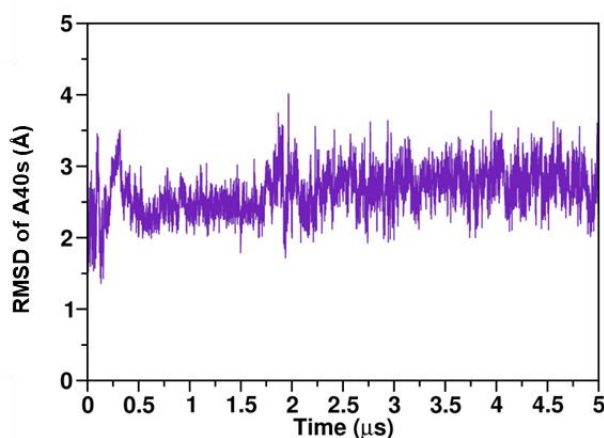


Figure 3.14. RMSD of A40s (calculated on heavy atoms, excluding terminal nucleotides) over the MD simulations on the EphA2/aptamer complex. The trajectory was aligned on the A40s heavy atoms, excluding the terminal nucleotides.

Next, the structural refinement of the A40s/Eph2 interaction throughout the simulation was evaluated by computing the DRPScore.²²⁹ DRPScore is a deep-learning-based scoring tool designed to evaluate the nativity of RNA-protein complexes based on known complexes, taking into account atom positions, residue types, intermolecular distances, and interaction interfaces. As shown in **Figure 3.15**, DRPScore increases from an initial value of 0.85 (corresponding to the docking pose) to 0.99 over the simulation, indicating that the final MD model achieves a high degree of native-likeness according to this metric. These results confirm that while docking provides a reasonable initial model, MD refinement is crucial for capturing a more native-like and stable interaction interface that static docking poses might miss. The DRPScore analysis independently validates the structural refinement achieved, adding extra strength to our MD predictions, as the probability of a native-like structure converges to 99.9% across the trajectory.

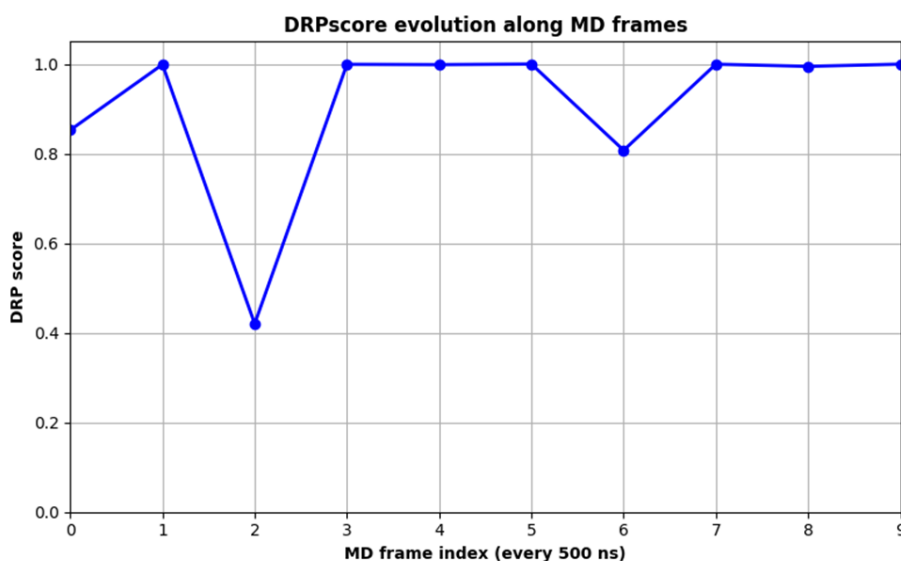


Figure 3.15. DRPScore¹³ computed for the EphA2/A40s complex over MD timescale. Representative frames for DRPScore calculation were extracted each 500 ns.

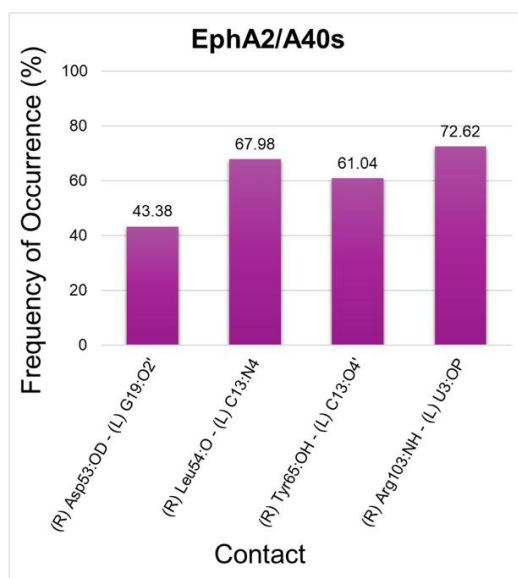
To investigate the structural basis of EphA2 and A40s binding, we utilized PyContact89, a toolkit designed to analyze noncovalent biomolecular interactions in MD trajectories. Our analysis identified the most significant aptamer-protein interactions, focusing on those that persisted for at least 0.4 μ s during the simulation (**Figure 3.16A**). These interactions involve amino acids Asp53, Leu54, Tyr65, and Arg103 of EphA2, as well as nucleotides C2, U3, C13, and G19. Hydrogen bonds are formed between the carboxylic oxygen of Asp53 and the hydroxyl sugar oxygen of G19, the backbone oxygen of Leu54 and the amino nitrogen of C13, the hydroxyl oxygen of Tyr65 and the sugar oxygen of C13. Additionally, a salt bridge is formed between the positively charged guanidinium nitrogen of Arg103 and the negatively charged phosphate oxygen of U3. Lastly, a T-shaped π - π interaction is formed between the aromatic phenolic ring of Tyr65 and the pyrimidine ring of C13 (**Figure 3.12B**).

EphA2/ephrinA1

RMSD analysis of secondary structure showed, as expected, high stability of ephrinA1 (the ligand, L) bound to EphA2 (the receptor, R), with a value of 2.8 ± 0.8 Å (**Figure 3.17**). Intermolecular interaction analysis revealed hydrogen bonds between the backbone oxygen of ^RGly49 and the hydroxyl group of ^LTyr46, and between the latter and the ^RAsp53 carboxyl moiety. Additional H-bonds are formed by the ^LArg89 guanidinium group with both the side nitrogen of ^RAsn57 and the backbone oxygen of ^RGln56, and between the side chain atoms of the latter with both ^LSer105 and ^LLys103 backbone atoms. Finally, H-bonds are established between the ^RArg103 side chain and the ^LThr112 backbone atoms, and between the ^RHis160 backbone atoms and the

^LThr115 hydroxyl. Additionally, salt bridges are formed between the negatively charged ^RAsp53 and the positively charged ^LLys107, as well as between ^RAsp61 and ^LLys85 (**Figure 3.12B**). Notably, in the crystal structure, the interactions between ^RGly49 and ^LAsp53, ^RAsn57 and ^LArg89, ^RAsp61 and ^LLys85, and ^RArg103 and ^LThr112 were absent or characterized by a distance of 5.22 Å, 4.30 Å, 4.62 Å, and 3.85 Å (measured between heavy atoms), respectively, which exceeds the typical range for hydrogen bonding or salt-bridge. During the MD simulations, these interactions are formed and stabilized, suggesting that dynamic conformational changes or flexibility facilitated their establishment. In contrast, only a single interaction between ^RArg103 and ^LGlu119, observed in the crystal structure, is disrupted or formed for less than 40% of the simulation time. These observations underscore the importance of MD simulations in capturing conformational states and interactions not observed in static crystal structures. Notably, in both EphA2/ephrinA1 and EphA2/A40s, residues Asp53 and Arg103 play a pivotal role in mediating intermolecular interactions. This finding suggests that these residues may serve as conserved interaction sites, potentially underlying a well-defined mechanism of molecular recognition or binding affinity.

A



B

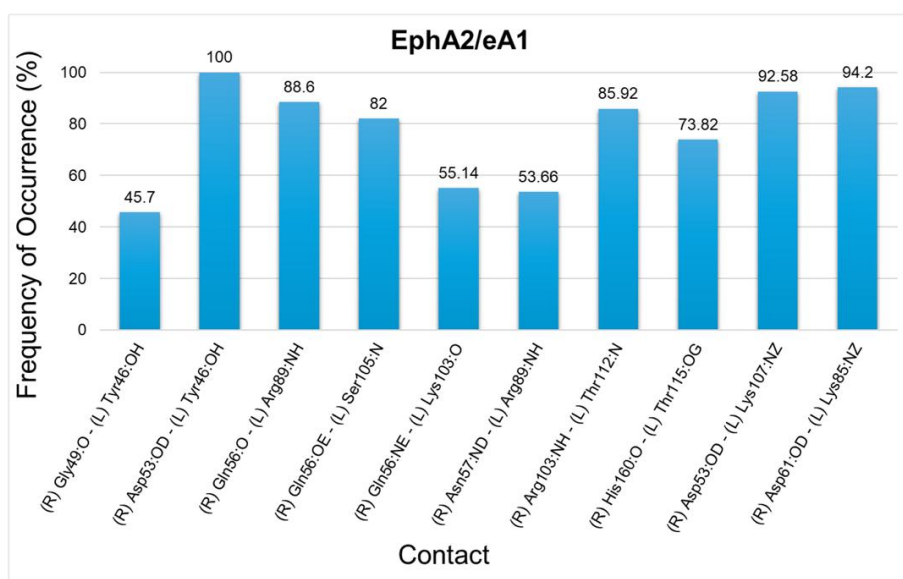


Figure 3.16. A) Histograms showing the frequency of occurrence of the most relevant EphA2/A40s (A) and contacts over the MD trajectory. B) Histograms showing the frequency of occurrence of the most relevant EphA2/eA1 contacts over the MD trajectory. For the identification of the non-covalent biomolecular interactions PyContact¹⁴ was employed. The frequency of occurrence was obtained using the

Hydrogen Bonds tool of VMD¹⁵ where the donor-acceptor (heavy atoms) distance cutoff was set at 3.5 Å and the angle cutoff at 40°.

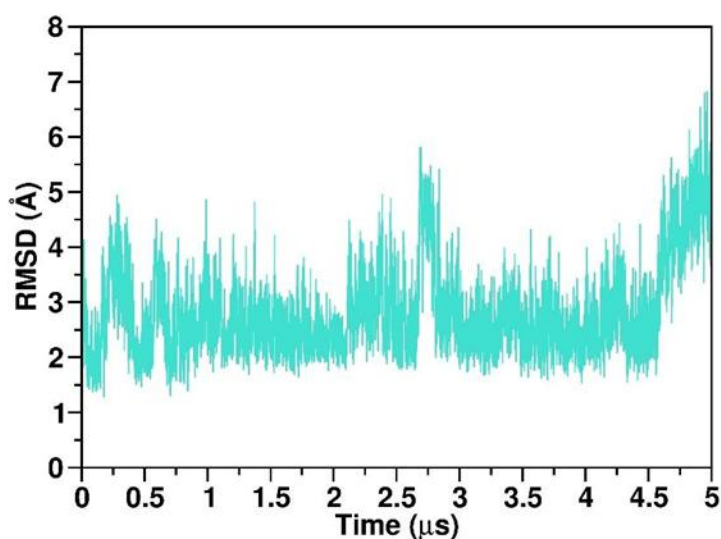


Figure 3.17. RMSD (calculated on the secondary C α atoms) of eA1 over the MD simulations on the EphA2/eA1 complex. The trajectory was aligned on the secondary structure C α atoms of EphA2.

3.2. RNA Force Field Evaluation

MD simulations represent a valuable tool for investigating the structure and dynamics of RNA at atomic detail, offering a genuine complement of experimental methods. Nevertheless, MD simulations also have considerable limitations. The reliability of MD simulations critically depends on the accuracy of the force field (FF), which defines the potential energy surface governing molecular behavior. Neglecting limits of MD methodology can lead

to overinterpretation of the simulation data and deviations from experimentally observed ensembles.

In this section, we address this challenge by performing a comprehensive evaluation of RNA FF performance. Specifically, we investigated four different FFs across a set of nine representative RNA motifs. For each FF, we first evaluate the global stability, followed by a comprehensive analysis of hydrogen bonds and dihedral angles. Finally, where experimental data are available, we compute the NOE-derived distances and analyze their agreement with NMR observables.

This approach allows us to identify the strengths and limitations of current RNA FFs, assess which FF formulations most reliably reproduce RNA behavior, and unveil important paths for ongoing and future FF development.

3.2.1. *Selection of FFs*

Four FFs among the most commonly used or most recently refined, were selected for evaluation. These include i) OL3²⁰⁹, an AMBER FF that contains updates such as parmbse0²³⁰ (reference FF for DNA) with α/γ torsion correction and a refit of the glycosidic torsion (χ -OL3) for backbone and base orientations improvement; ii) OL3 cP-ghbfix21^{231,232} (hereafter OL3-fix), which augments OL3 with the case phosphate (CP) nonbonded parameters softening the phosphate-base contacts through revised charges and matched dihedrals, and by adding the gHBfix21 correction, a data-driven modification that slightly enhances canonical base-pair hydrogen bonding; iii) Shaw's FF²³³ (also referred to as DESRES in the literature, here Shaw) that represents a full reparameterization including extensive refitting of torsional angles (χ , γ , ϵ/ζ ,

β , and sugar-pucker), replacement of standard phosphate vdW terms with CP phosphate parameters, implementation of a large NBfix table to fine-tune base-stacking and base-phosphate contacts, and adjustments of selected nucleobase charges and Lennard-Jones (LJ) radii to better balance hydrogen bonding and π -stacking; and iv) DES-AMBER²³⁴ (hereafter DES), which provides a charge-damped approach, in which all partial charges are uniformly scaled to 0.9, assuming that this would improve errors derived from the lack of electronic polarization in Lifson's like FFs.^{235,236} Moreover, the χ and ζ torsions are refitted using MP2-level quantum data, and a minimal NBfix patch is applied to selectively adjust LJ interactions between phosphate oxygens and aromatic carbons with water, contributing to preserve base stacking and solvation behavior under the softened electrostatics (**Table 3.1**).

Table 3.1. Comparison of parameter treatments across the tested RNA FFs.

FF term	OL3	OL3-fix	Shaw	DES
Backbone torsions	parmbse0 α/γ updates, χ -OL3 tuning	Same as OL3	Full χ , γ , ϵ/ζ , β , sugar pucker refit, extensive reparameterization	Refitted χ and ζ using MP2 data, rest largely AMBER-like
Glycosidic torsion χ	χ -OL3 correction	Same as OL3	Fully reparameterized	refitted (MP2-level)
Nonbonded corrections (LJ, Nbfix, CP parameters)	Standard AMBER	Standard AMBER	Refitted	minimal NBfix patch (phosphate oxygens ↔ aromatic carbons & water); modified LJ implicitly

				softened via charge scaling
Electrostatics	standard AMBER RESP charges	standard AMBER RESP charges	standard charges; some nucleobase charges modified	uniform charge scaling to 0.9 (charge-damped) to partially mimic missing polarization
Hbond corrections	none	gHBfix21: modest strengthening of canonical base-pairs & selected H-bond patterns	implicit tuning through LJ, torsions, NBfix; no explicit H-bond term	softened electrostatics reduce H-bond over-stabilization; no explicit H-bond term

3.2.2. *Selection of RNA motifs*

The state-of-the-art FFs were tested on nine RNA structures with different topological architectures including duplexes, internal loop, hairpin, sarcin-ricin, tetraloop, junction, G-quadruplex (G4), kissing-loop, and tetraloop-tetraloop receptor (TTR) motifs (**Figure 3.18, Table 3.2**). This diversity ensures that the FFs are tested across both short and long sequences and against structurally distinct RNA motifs that challenge different aspects of RNA energetics. Motifs such as i-motifs and triplexes were excluded as they

involve protonated cytosines, for which most current RNA FFs do not yet provide suitable parameters.

Table 3.2. List of the selected RNA structures for FF evaluation.

<i>PDB code</i>	<i>RNA motif</i>	<i>Experimental method</i>	<i>Number of nt</i>
<i>2VUQ</i>	Duplex	X-RAY	14
<i>4U37</i>	Duplex	X-RAY	14
<i>2V7R</i>	Duplex	X-RAY	14
<i>2H49</i>	Internal-loop	NMR	18
<i>1Q9A</i>	Sarcin-ricin (Hairpin)	X-RAY	27
<i>6Y0T</i>	Sarcin-ricin (Hairpin)	X-RAY	27
<i>2KOC</i>	Tetraloop	NMR	14
<i>8ITS</i>	3-way-junction	X-RAY	46
<i>7PS8</i>	Q4	NMR	23
<i>8UTG</i>	Q4	X-RAY	42
<i>2MI0</i>	Kissing-loop	NMR	43
<i>2D1B</i>	Kissing-loop	NMR	78
<i>6DVK</i>	TTR	X-RAY	95

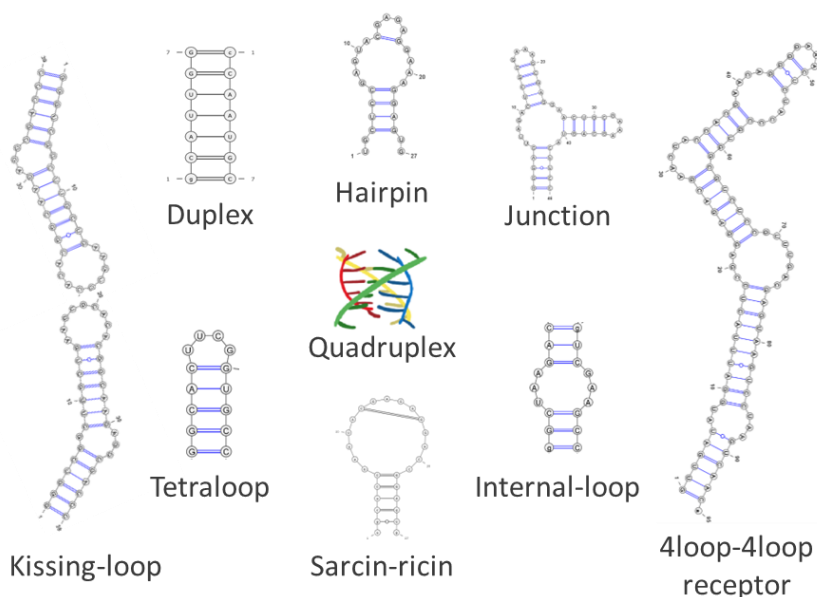


Figure 3.18. RNA motifs evaluated using OL3, OL3-fix, Shaw, and DES FFs.

The study began with the evaluation of three duplexes (PDB IDs: 2VUQ²³⁷, 4U37²³⁸, 2V7R²³⁹), the simplest RNA architecture, in order to establish a robust baseline. Establishing consistent performance across several duplexes ensures that each FF accurately reproduces foundational features such as A-form geometry, base-pair stability, and backbone torsion behavior, before its performance is examined on more complex RNA architectures.

We next evaluated an internal loop RNA (PDB ID: 2H49²⁴⁰), which represents a double-stranded stem where both strands contain unpaired nucleotides, creating a bulge on each side. These interruptions in regular base pairing, influence the RNA's overall shape, flexibility, and stability, and often serve as sites for tertiary interactions or protein binding.²⁴⁰

Next, we included a hairpin motif (PDB ID: 6Y0T), an essential secondary structure of RNA as it can determine interactions in a ribozyme, protect

mRNA from degradation, serve as a recognition motif for binding proteins or act as substrate for enzymatic reactions.²⁴¹ Among hairpin RNAs, the sarcin-ricin loop (PDB ID: 1Q9A²⁴²) is a highly conserved motif found in all large ribosomal subunits. It forms part of the GTPase-associated center and is crucial for binding and activating translational GTPases. Structurally, it is a distorted hairpin containing a GAGA tetraloop, a G-bulged cross-strand A-stack, a flexible region and a terminal A-form duplex. It is named after its sensitivity to the sarcin and ricin toxins, which cleave a specific adenine residue within this loop, thereby inactivating the ribosome and halting protein synthesis.

We then examined the UUCG tetraloop (PDB ID: 2KOC²⁴³), which possesses a well-defined native structure with canonical base pairing in the stems and characteristic noncanonical interactions involving sugar-base and base-phosphate hydrogen bonds (H-bonds) in the loop. We also evaluated a kink-turn junction (hereafter junction), a motif commonly found in duplex RNA consisting of a short (typically 3 nt) bulge-loop followed by conserved G:A base pairs (PDB ID: 8ITS²⁴⁴).

Among higher-order RNA architectures, G4s represent a particularly interesting structural class. They consist of stacked guanine quartets stabilized by Hoogsteen hydrogen bonding and monovalent cations. G4s influence processes such as replication, translation, and splicing.^{245,246} Here, two distinct G4s have been investigated, one comprising of a single RNA chain and the other composed of two identical chains (PDB IDs 7PS8 and 8UTG²⁴⁷, respectively).

Kissing-loop motifs are formed between complementary base pairings between the loops of two hairpins. Two variants were explored: one formed

by two identical long hairpins of ca. 40 nt (PDB ID: 2D1B²⁴⁸) and another formed by two shorter, distinct hairpins of ca. 20 nt (PDB ID: 2MI0²⁴⁹).

Finally, a mini TTR RNA (PDB ID: 6DVK²⁵⁰) has been included to assess how well FFs handle long-range tertiary interactions.

3.2.3. *MD Simulation Protocols*

The systems were simulated for 1 μ s using Amber 20²¹⁰ for OL3, Shaw, and OL3 OL3 cP-ghbfix21 FFs, while DES-AMBER was simulated using GROMACS 2023.²⁰⁸ All systems run with Amber 20 were solvated in an octahedral OPC water box built with a 15 Å padding parameter, neutralized using K⁺ ions, and set to a near-physiological 0.15 M concentration of KCl. The systems underwent standard optimization procedures, beginning with a six-stage energy minimization protocol. The first five stages consisted of 3,000 steps of steepest-descent minimization, starting with a 25 kcal/mol restraint on heavy atoms and reducing the restraint by 5 kcal/mol at each subsequent stage. This was followed by an additional 20,000 steps of unrestrained minimization. Thermalization of 500,000 steps (1 ns) was then performed using Langevin dynamics.¹⁶³ During the first 0.7 ns the temperature was gradually increased to 298 K, which was maintained for the remaining 0.3 ns. Thermalization was followed by 10 ns equilibration using the Monte-Carlo¹⁷¹ barostat. Finally, 1 μ s of plain MD simulations was carried out under NPT conditions (1 atm, 298 K) with a 4 fs timestep. OL3 cP-ghbfix21 was applied only in the OL3 cP simulations and PLUMED²⁵¹ coupling in Amber was activated during production to generate the sequence-specific ghbfix21 inputs.

The DES-AMBER systems, run with GROMACS 2023, were solvated in a dodecahedral box of TIP4P-D water (15 Å padding), and ionized to ca. 150 mM KCl (K^+/Cl^- carry $\pm 0.9e$ under DES charge scaling). Because the total box charge is not always divisible by $\pm 0.9e$, residual net charge was minimized using gromologist²⁵² and handled by PME's uniform neutralizing background. Gromologist workflow was used to prepare the described system and perform energy minimization of 1,000 steps using the steepest-descent algorithm. 1 ns thermalization was then carried out to reach 298 K temperature using the v-rescale thermostat, followed by 10 ns equilibration with 2 fs timestep. 1 μ s simulations with GROMACS were performed under NPT conditions (300 K, 1 atm) with the v-rescale thermostat ($\tau = 0.1$ ps, 300 K), a Berendsen isotropic barostat ($\tau = 2.0$ ps, 1 bar), and a 4 fs timestep. PME electrostatics employed a 1.0 nm real-space cutoff, while van der Waals interactions used a cutoff of 1.0 nm (dispersion correction off).

3.2.4. *MD Simulation Results and Analysis*

RMSD Analysis

- Overall, throughout the simulations, the RNA exhibited stable dynamics for duplexes, internal-loop, sarcin-ricin, and the three-way junction motifs across all FFs. For duplexes, RMSD values ranged from 1.058 ± 0.257 Å to 1.555 ± 0.397 Å. The internal-loop motif showed RMSDs between 1.175 ± 0.202 Å and 1.418 ± 0.247 Å, while the sarcin-ricin structures varied from 1.195 ± 0.198 Å to 1.691 ± 0.417 Å. The three-way junction motif exhibited higher flexibility, with RMSDs ranging from 3.235 ± 0.769 Å to $5.044 \pm$

1.778 Å, possibly due to its larger size (duplexes, internal-loop, and sarcin-ricin: 14-27 nt; junction: 46 nt).

- The tetraloop presented stability in OL3 and DES FFs, with RMSD values not exceeding 1.6 Å. In contrast, OL3-fix and Shaw displayed structural rearrangements at approximately 250 ns and 800 ns, respectively, resulting in increased average RMSD of 2.266 ± 0.635 Å for OL3-fix and 1.918 ± 0.792 Å for Shaw.
- For the G4 structures, only the OL3 FF applied to the single-chain G4 (7PS8) exhibited global stability with RMSD 3.053 ± 0.315 Å. In all other simulations, structural deviations led to increased RMSDs ranging from 3.943 ± 0.504 Å to 4.556 ± 0.461 Å. The increased stability of OL3 for the 7PS8 G4 compared to the 8UTG G4 can be attributed to differences in size and topology: 7PS8 consists of a single 23-nt RNA chain, whereas 8UTG is composed of two identical 21-nt RNA chains.
- Notably, kissing-loops presented higher stability in the 2D1B RNA when simulated with the OL3 and OL3-fix FFs, with RMSD values of 4.236 ± 0.849 Å and 4.301 ± 0.843 Å, respectively. In contrast, the remaining simulations of both 2MI0 and 2D1B showed substantially higher RMSDs, ranging from 6.206 ± 0.862 Å to 9.472 ± 1.044 Å, indicating greater structural deviations under those force fields.
- Finally, the TTR motif presented enhanced stability when simulated with the OL3 and OL3-fix FFs with RMSD of 3.741 ± 0.571 Å and 4.397 ± 0.604 Å, respectively. In contrast, the Shaw and DES FFs induced conformational changes, leading to increased average RMSDs of 11.808 ± 7.794 Å at approximately 410 ns (Shaw) and 19.768 ± 7.735 Å around 150 ns (DES). These

results suggest that OL3 and OL3-fix better preserve the structural integrity of the TTR motif compared to Shaw and DES.

Notably, OL3 FF exhibited the best performance in reproducing the crystal structures, achieving agreement in 12/13 structures with the lowest mean RMSD (2.722 ± 0.549 Å). OL3-fix closely followed, reproducing 10/13 structures with a mean RMSD of 2.982 ± 0.573 Å. Shaw and DES showed lower performance, both reproducing 8/13 structures, with Shaw exhibiting structural deviations in the tetraloop, the G4s, and the 78-nt kissing-loop 2D1B, and the TTR, and DES in the G4s, the kissing-loops, and the TTR. The corresponding mean RMSDs were 3.676 ± 0.923 Å and 4.555 ± 1.440 Å, respectively. Moreover, DES exhibited faster structural collapse and greater fluctuations compared to the other force fields (**Table 3.3**).

Table 3.3. RMSD (Å) of RNA heavy atoms excluding terminal nt using the OL3, OL3-fix, Shaw, DES FFs. The mean values of each FF are also provided along with their standard error.

Structure	OL3 (Å)	OL3-fix (Å)	Shaw (Å)	DES (Å)
2VUQ	1.061 ± 0.260	1.058 ± 0.257	1.147 ± 0.282	1.411 ± 0.352
4U37	1.196 ± 0.278	1.173 ± 0.269	1.259 ± 0.283	1.408 ± 0.303
2V7R	1.333 ± 0.558	1.172 ± 0.301	1.270 ± 0.324	1.555 ± 0.397
2H49	1.247 ± 0.226	1.175 ± 0.202	1.274 ± 0.241	1.418 ± 0.247
1Q9A	1.282 ± 0.246	1.265 ± 0.232	1.435 ± 0.316	1.691 ± 0.417
6Y0T	1.195 ± 0.198	1.210 ± 0.206	1.262 ± 0.254	1.558 ± 0.11

2KOC	1.422 ± 0.233	2.266 ± 0.635	1.918 ± 0.792	1.530 ± 0.285
8ITS	3.565 ± 0.776	5.044 ± 1.778	3.235 ± 0.769	4.663 ± 0.914
7PS8	3.053 ± 0.315	3.943 ± 0.504	4.166 ± 0.370	3.974 ± 0.674
8UTG	4.338 ± 0.506	4.249 ± 0.636	4.206 ± 0.580	4.556 ± 0.461
2MI0	7.721 ± 0.868	7.518 ± 1.171	7.954 ± 1.005	6.206 ± 0.862
2D1B	4.236 ± 0.849	4.301 ± 0.843	6.854 ± 1.048	9.472 ± 1.044
6DVK	3.741 ± 0.571	4.397 ± 0.604	11.808 ± 7.794	19.768 ± 7.735
Mean	2.722 ± 0.549	2.982 ± 0.573	3.676 ± 0.923	4.555 ± 1.440

Hydrogen Bond (H-bond) Analysis

We performed H-bond analysis and categorized interactions into the following groups: WC, sugar–phosphate (s–ph), base–phosphate (b–ph), sugar–sugar (s–s), sugar–base (s–b), and non–WC base–base interactions. This classification allowed us to track which interactions were disrupted, in what order, and at which stage of the simulation they occurred.

To distinguish interactions, present in the experimental structures from those generated during the simulations, we used a reference set: interactions from crystal structures or, in the case of NMR structures, interactions present in more than 50% of the reported conformations. When plotting H-bonds, reference interactions are indicated in bold on the y-axis. It should be noted that some interactions, particularly s–ph, b–ph, s–s, and s–b, may be real but absent from the reference set, as they might require energy minimization to be captured. For RNA structures with more

than 40 nucleotides, H-bond plots are not shown due to the large number of interactions, which would reduce clarity. For the description of molecular interactions, AMBER atom nomenclature is used (see **Figure 3.19**).

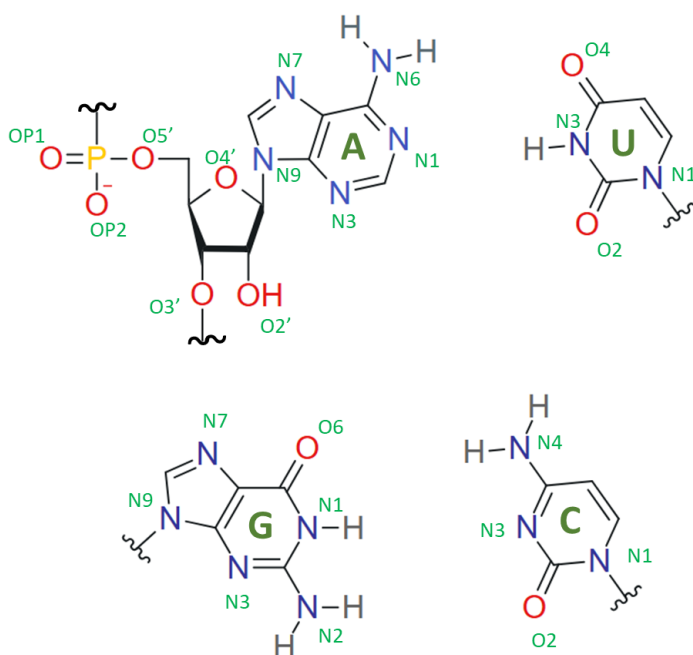


Figure 3.19. AMBER atom nomenclature used for the description of RNA molecular interactions.

- Consistent with the RMSD results, duplexes, internal-loop, sarcin-ricin, and junction motifs largely preserved their H-bonds throughout the 1 μ s simulations. Some canonical (e.g., A15–U4 in the internal-loop, A27–U43 in the junction) and non-canonical base pairs (e.g., U1–G27 in the sarcin-ricin 1Q9A motif) occasionally lost contact during the simulations; however, these

disruptions did not significantly affect the overall structural stability of the RNA (**Figure 3.20**, **Figure 3.21**, **Figure 3.22**).

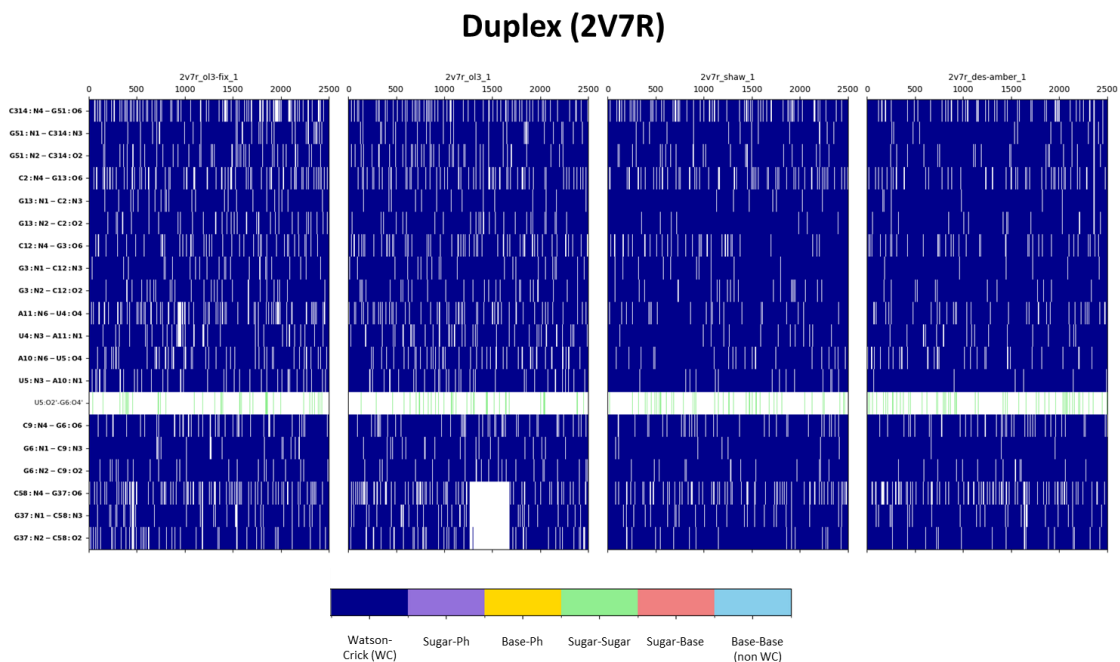


Figure 3.20. H-bond analysis of the duplex motif (2V7R) in simulations using OL3, OL3-fix, Shaw, and DES force fields. The canonical base pair G7-C8 occasionally lost contact during the simulations.

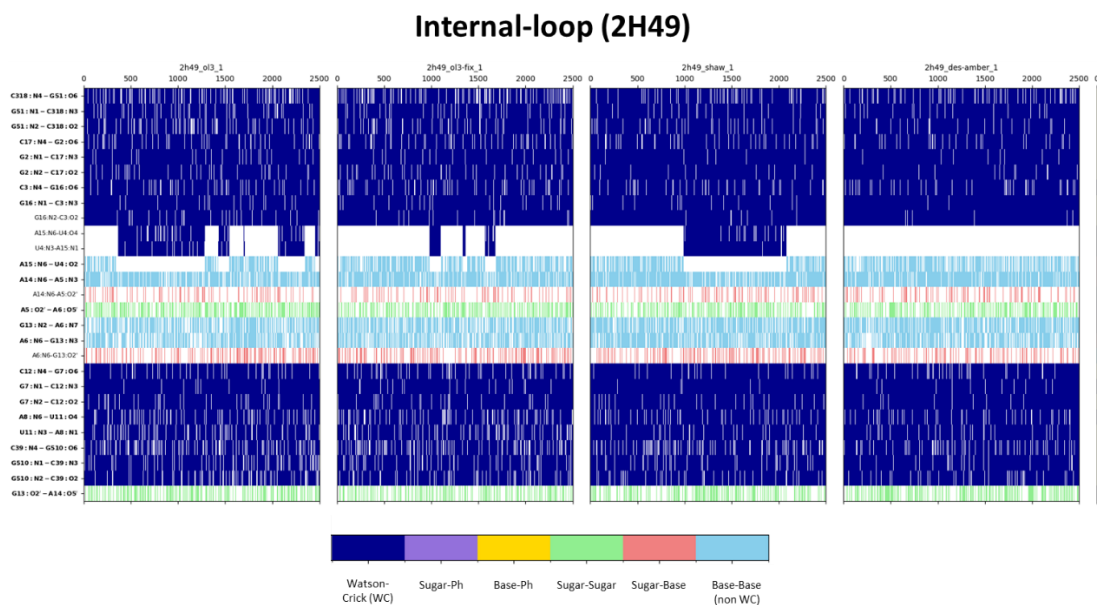


Figure 3.21. H-bond analysis of the internal-loop motif (2H49) in simulations using OL3, OL3-fix, Shaw, and DES force fields. The canonical base pair A15–U4 occasionally lost contact during the simulations.

Sarcin-ricin (1Q9A)

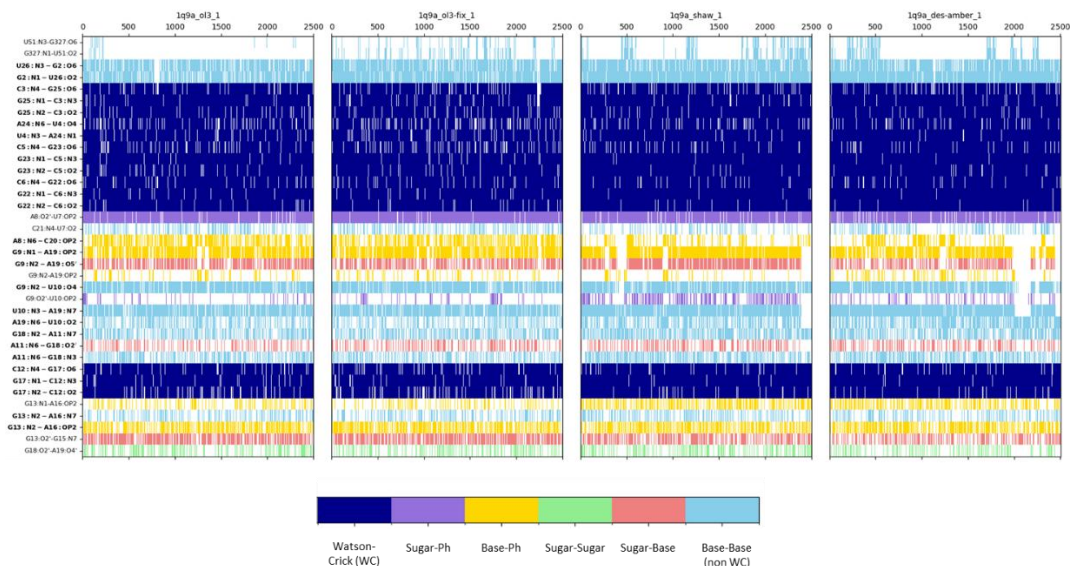


Figure 3.22. H-bond analysis of the sarcin-ricin motif (1Q9A) in simulations using OL3, OL3-fix, Shaw, and DES force fields. The non-canonical base pair U1–G27 occasionally lost contact during the simulations.

- As previously mentioned, the tetraloop simulated with the OL3-fix and Shaw FFs underwent structural deviations, resulting in increased RMSDs. H-bond analysis indicated that this change was primarily due to the loss of a s-b interaction and one non-canonical b-b interaction between U6-G9 in both FFs (**Figure 3.23**). Despite these conformational changes, the final 20–40 ns of the 1 μ s simulation with OL3-fix suggest partial reformation of these interactions, indicating that the RNA may be attempting to return to its native conformation. Extending the MD simulations would be necessary to verify whether the RNA fully recovers its initial structure.

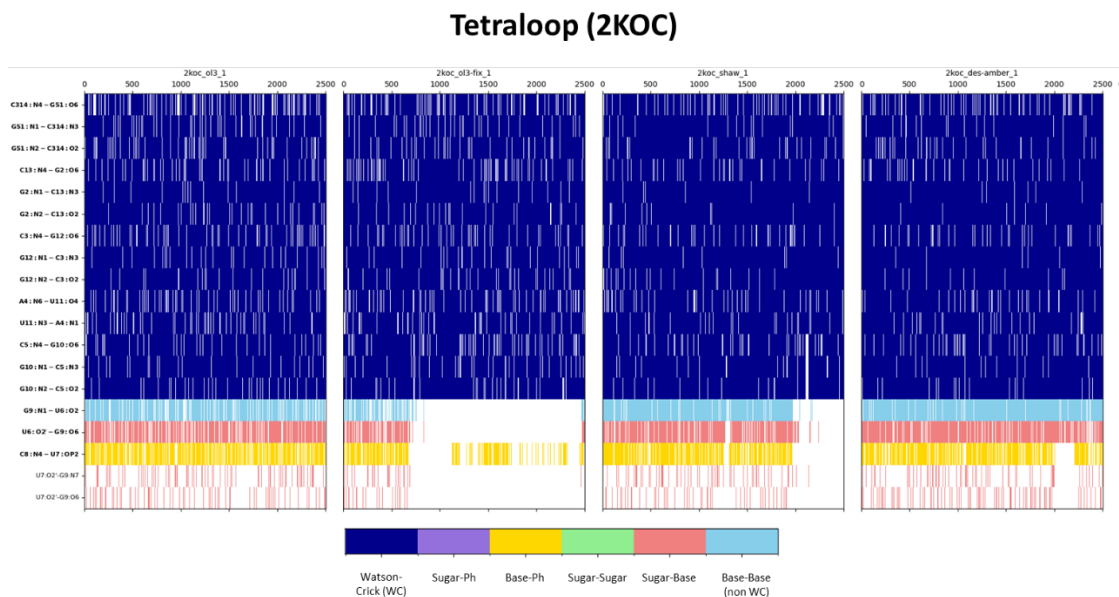


Figure 3.23. H-bond analysis of the tetraloop motif (2KOC) in simulations using OL3, OL3-fix, Shaw, and DES force fields.

- H-bond analysis of the G4 RNA motifs revealed differences in interactions across the various force fields (**Table 3.4**). In particular, the 7PS8 G4 with OL3 and OL3-fix FFs maintained a s-ph interaction between A2:O2'-G3:OP2, which was present in the experimental structure, as well as a WC and a non-WC interaction between A2:N6-U16:O2, which are absent in the crystal structure. These interactions were lost at approximately 300 ns in simulations with Shaw, and were completely absent in simulations with DES. Furthermore, DES failed to maintain several non-WC base pairs (G3:N1-G7:O6, G3:N2-G7:N7, G7:N1-G17:O6, G7:N2-G17:N7, G17:N1-G21:O6, G17:N2-G21:N7), while OL3-fix missed a WC base pair between G19-C20 (present in the experimental structure) and a base-phosphate (b-ph) interaction

between C20:N4–G23:OP2 (absent in the crystal structure). Despite the additional loss of non-WC H-bonds, the average RMSD of DES FF does not exhibit notable differences with respect to OL3-fix or Shaw FFs.

Regarding the 8UTG G4 simulations, the Shaw FF failed to maintain two s-b interactions (G4:O2'-A9:N1 and G8:O2'-G5:N3) as well as a s-ph H-bond between A7:O2'-A9:OP2. Notably, the G8-G5 s-b interaction was present in the crystal structure. In addition, both Shaw and DES exhibited a b-ph interaction between C12:N4-G31:OP1, which is absent in the NMR structure. In general, interactions present in the NMR structure were largely preserved during the 1 μ s simulations. Exceptions include the non-WC G4:N2–A9:N7 interaction, which was lost in all force fields at approximately 100 ns (OL3, Shaw) or 400 ns (OL3-fix, DES), and non-WC G25:N2–A30:N7, present in the NMR structure but lost within the first 50 ns of the OL3 simulation.

Table 3.4. Interactions of G4 structures that were lost during the 1 μ s MD simulations. Each category is colored following the color scheme of the H-bond plots.

	OL3	OL3-fix	Shaw	DES
7PS8		<u>WC</u> G19–C20 <u>b-ph</u> C20:N4–G23:OP2	<u>s-ph</u> A2:O2'-G3:OP2 <u>WC</u> A2-U16 <u>Non-WC</u> A2:N6-U16:O2	<u>s-ph</u> A2:O2'-G3:OP2 <u>WC</u> A2-U16 <u>Non-WC</u> A2:N6-U16:O2 G3:N1–G7:O6 G3:N2–G7:N7 G7:N1–G17:O6 G7:N2–G17:N7 G17:N1–G21:O6

				G17:N2–G21:N7
8UTG	<u>WC</u> G4:N2–A9:N7 Non-WC G25:N2–A30:N7	<u>WC</u> G4:N2–A9:N7	<u>WC</u> G4:N2–A9:N7 <u>s-b</u> G4:O2'–A9:N1 G8:O2'–G5:N3 <u>s-ph</u> A7:O2'–A9:OP2	<u>WC</u> G4:N2–A9:N7

Although detailed H-bond analysis highlighted differences across force fields, the interactions present in the crystal structures were generally preserved throughout the 1 μ s simulations. This preservation alone, however, does not fully explain the observed RMSD increases. The initial rise in RMSD is primarily due to the flexible corner regions of the G4, which exhibit high instability early in the simulations, while the central tetrad core remains well-preserved, maintaining its native conformation during this period (**Figure 3.24**). This indicates that these structural deviations could arise from loss of stacking interactions. Our results illustrate that H-bond analysis alone is insufficient to capture all conformational changes, emphasizing the need for a combined approach, including trajectory visualization, RMSD calculations, and H-bond analysis, to accurately assess MD performance for RNA structures.

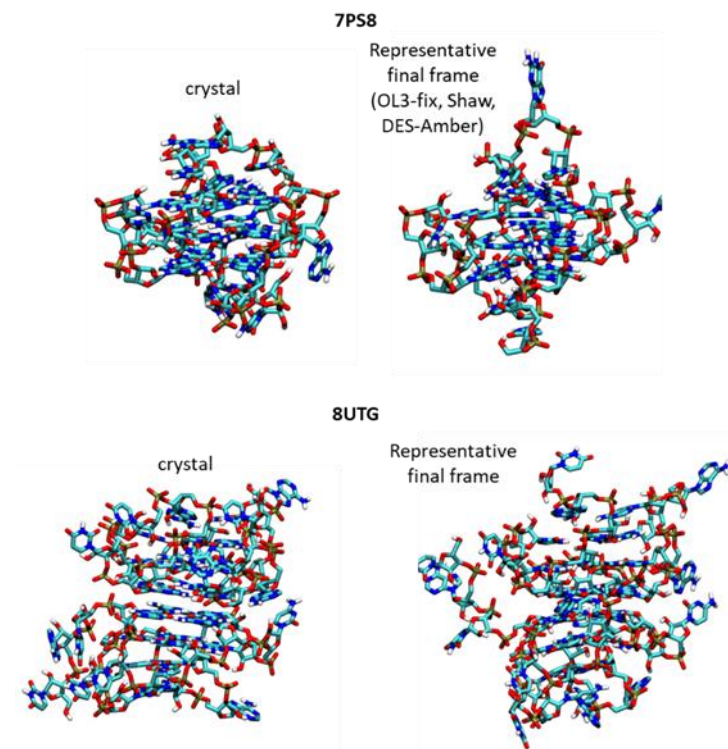


Figure 3.24. Comparison between the crystal structures of G4 7PS8 and G4 8UTG (left) and a representative conformation from the molecular dynamics trajectory (right). The trajectory snapshot highlights the pronounced flexibility of the peripheral corner regions, while the central G4 core remains structurally stable.

- Investigation of the kissing-loop motifs revealed higher stability only for the 2D1B RNA when simulated with the OL3 and OL3-fix force fields, whereas the remaining simulations exhibited increased RMSD values (**Table 3.3**). H-bond analysis further explained the differential stability of the longer 2D1B kissing-loop compared to the shorter 2MI0 system. The 2D1B stability relied mainly on non-WC and s-ph interactions, whereas 2MI0 involved a broader set of H-bond types, making it more difficult to maintain

its native interaction network.

H-bond analysis of 2D1B highlighted notable differences among the force fields. For example, DES lacked a s-ph interaction between A15:O2'–G14:OP2, as well as several non-WC interactions: A15:N6–A23:N1 (present in experimental structure), G71:N2–G45:O6 (absent in experimental structure), A70:N6–G46:O6, G46:N1–A70:N1 (present in experimental structure), C73:N4–G72:N2, and C73:N4–G72:N3 (present in the crystal structure). However, DES formed a s-b interaction between G46:N1–G71:O2' and a s-s interaction between G71:O2'–G72:O5'. Shaw also missed the s-ph interaction A15:O2'–G14:OP2 and the non-canonical b-b interaction A15:N6–A23:N1 but formed a new non-WC interaction A55:N6–A23:N1.

Interestingly, despite the apparent instability suggested by RMSD calculations, trajectory visualization revealed that the kissing-loop motifs converged to a stable conformation over the 1 μ s simulations for OL3, OL3-fix, and Shaw. This apparent discrepancy arises because RMSD was initially calculated relative to the crystal structure. During equilibration, the kissing-loop motifs underwent conformational rearrangements, reaching a new stable structure that persisted throughout the simulations. This observation is further supported by RMSD calculations using the equilibrated structure as a reference (**Table 3.5, Figure 3.25**).

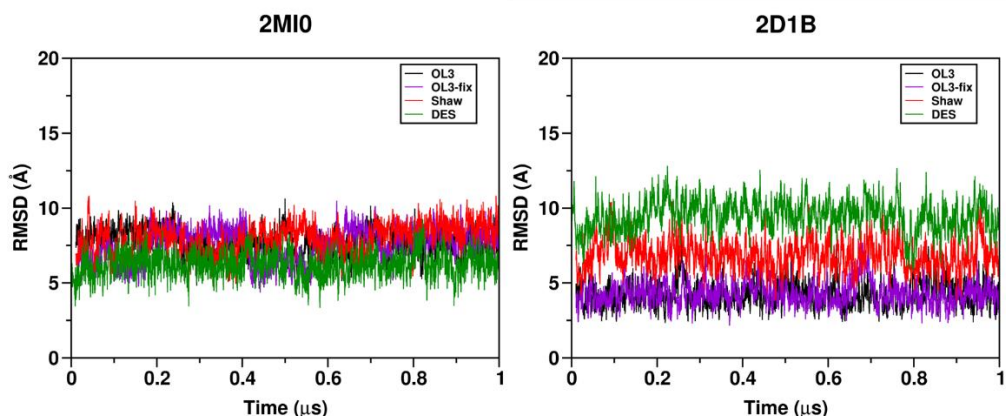


Figure 3.25. RMSD plots of kissing-loop motifs: 2MI0 (left) and 2D1B (right).

In contrast, simulations with the DES force field revealed an artifact for the 2D1B RNA: the RNA grooves gradually broadened, leading to elongation of the global structure (**Figure 3.26**). RMSD calculations relative to the equilibrated structure showed little effect for other PDB IDs, indicating that continuous conformational changes occurred during the 1 μ s simulations. This effect was more pronounced for 2D1B (78 nt) than for 2MI0 (43 nt), likely due to the longer sequence length.

Table 3.5. Average RMSD ($\text{\AA}$) of the kissing-loop motifs (2MI0, 2D1B) with reference to the crystal and the equilibrated structure.

	2D1B		2MI0	
Force Field	Equilibrated structure	Crystal structure	Equilibrated structure	Crystal structure

OL3	4.471 ± 1.027	4.236 ± 0.849	2.188 ± 0.532	7.721 ± 0.868
OL3-fix	4.025 ± 1.049	4.301 ± 0.843	4.783 ± 1.028	7.518 ± 1.171
Shaw	3.659 ± 0.690	6.854 ± 1.048	3.413 ± 0.573	7.954 ± 1.005
DES	7.180 ± 1.174	9.472 ± 1.044	5.791 ± 0.871	6.206 ± 0.862

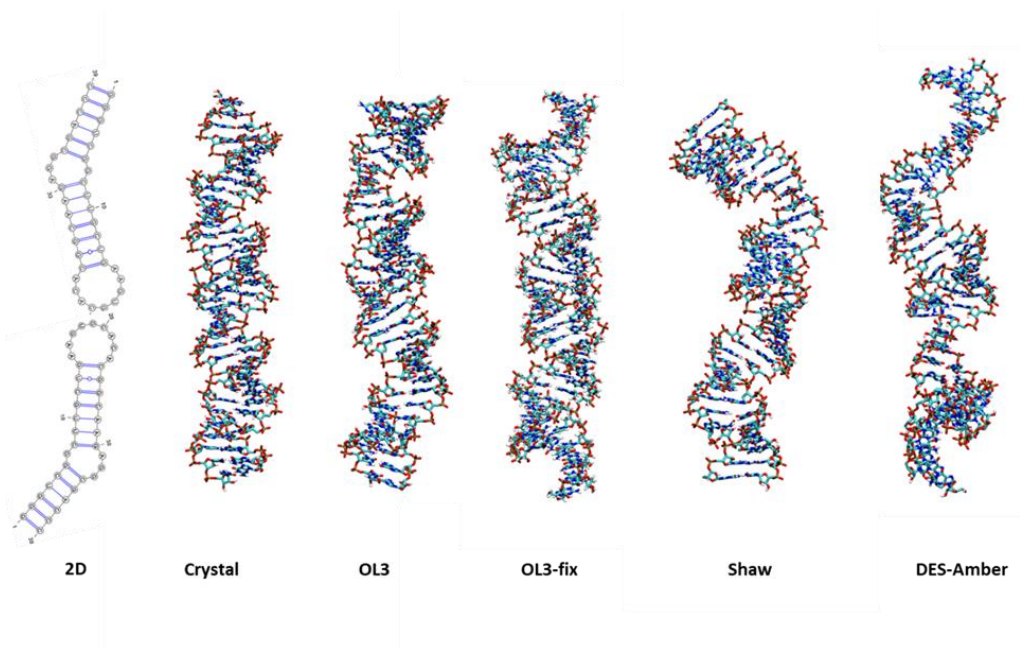


Figure 3.26. Comparison of kissing-loop (PDB ID: 2D1B) crystal structure and average structure of OL3, OL3-fix, Shaw, and DES FFs. OL3 and OL3-fix maintain a conformation similar to the crystal structure, Shaw adopts a alternative conformation with respect to the crystal structure after equilibration, while DES broadened its grooves leading to an elongated conformation.

- For the TTR motif, tertiary interactions, specifically two non-canonical contacts between A7–A47, a non-canonical G9–A49

interaction, and a s-s contact between A48–U86, were critical for maintaining the RNA’s structural integrity. Simulations with the Shaw and DES FFs lost these contacts, but via distinct mechanisms. Shaw exhibited a straightforward loss of tertiary interactions, whereas DES gradually widened the RNA grooves beginning at approximately 100 ns, elongating the overall structure. This deformation disrupted long-range contacts and ultimately led to complete opening of the TTR (**Figure 3.27**). Notably, a similar artifact was observed previously for a kissing-loop motif, suggesting a recurring issue with DES-AMBER in simulations of longer RNA constructs (>40 nt).

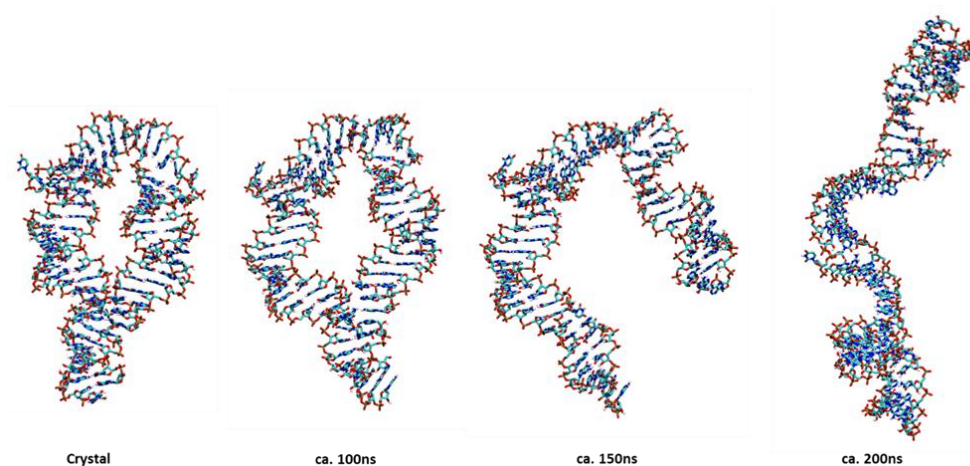


Figure 3.27. Loss of long-range tertiary interactions in the TTR using DES FF. DES gradually widened the RNA grooves beginning at approximately 100 ns, elongating the overall structure.

Dihedral Angle Analysis

All dihedral angles were monitored throughout the 1 μ s simulations. To streamline the analysis and reduce figure complexity, dihedrals that remained within a 60° range for at least 98% of the simulation were excluded from the plots. In addition, the remaining dihedrals were grouped by type to facilitate interpretation and improve overall readability.

- Representative dihedral-angle plots are shown for the duplex, internal-loop, and sarcin–ricin motifs. As expected from the RMSD and H-bond analyses, these structures maintain (along with the junction) canonical backbone conformations, and no substantial dihedral transitions were observed across force fields (Figure 3.28, Figure 3.29, Figure 3.30).

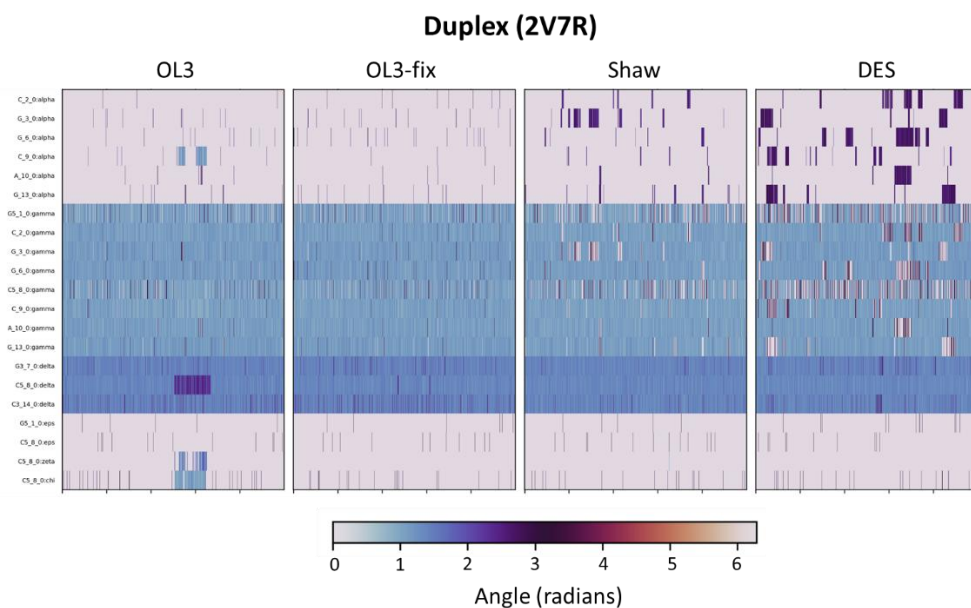


Figure 3.28. Dihedral angle time series for the duplex motif (2V7R).

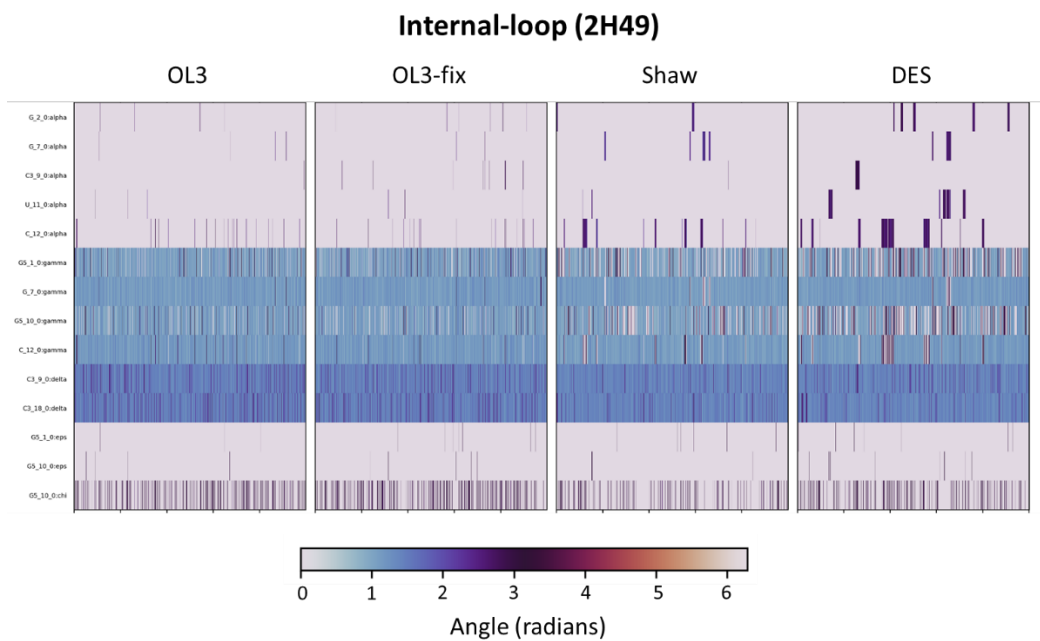


Figure 3.29. Dihedral angle time series for the internal-loop motif (2H49).

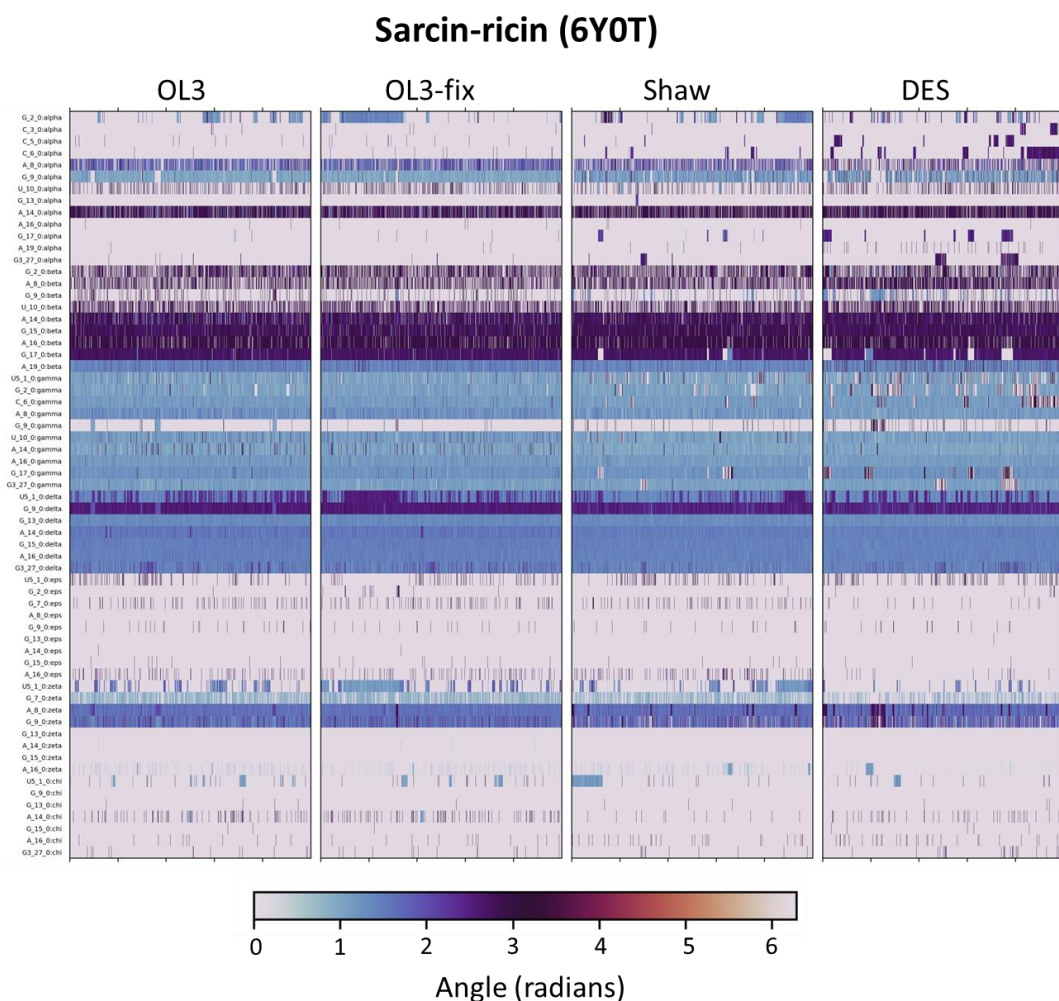


Figure 3.30. Dihedral angle time series for the duplex motif (6Y0T).

- In contrast, the tetraloop exhibited clear dihedral rearrangements in the OL3-fix and Shaw simulations, occurring at exactly the same time as the RMSD increase. In the OL3-fix simulation, the G10 β dihedral shifted by 1.5 rad, while the γ dihedrals of G9 and G10 transitioned from ca. 3 rad to ca. 1.5 rad. The C8 δ angle increased from ca. 2 rad to ca. 3 rad but returned to its original value after

approximately 50 ns, without affecting RMSD. Additionally, the glycosidic χ angle of G9 changed from ca. 2 rad to ca. 1 rad, and later to ca. 0 rad at approximately 450 ns.

In the Shaw simulation, similar transitions occurred: the G10 β dihedral shifted from its initial ca. 1.5 rad value (exact final value not specified), the G9 and G10 γ dihedrals decreased from ca. 3 rad to ca. 1 rad, and both the C8 δ and ζ angles changed by ca. 1 rad. Shaw also exhibited a shift in the U7 δ dihedral from ca. 2.5 rad to ca. 1.5 rad.

Notably, OL3-fix and Shaw share a consistent pattern of γ and ζ transitions, and these common dihedral changes correlate with the structural collapse of the tetraloop (**Figure 3.31**).

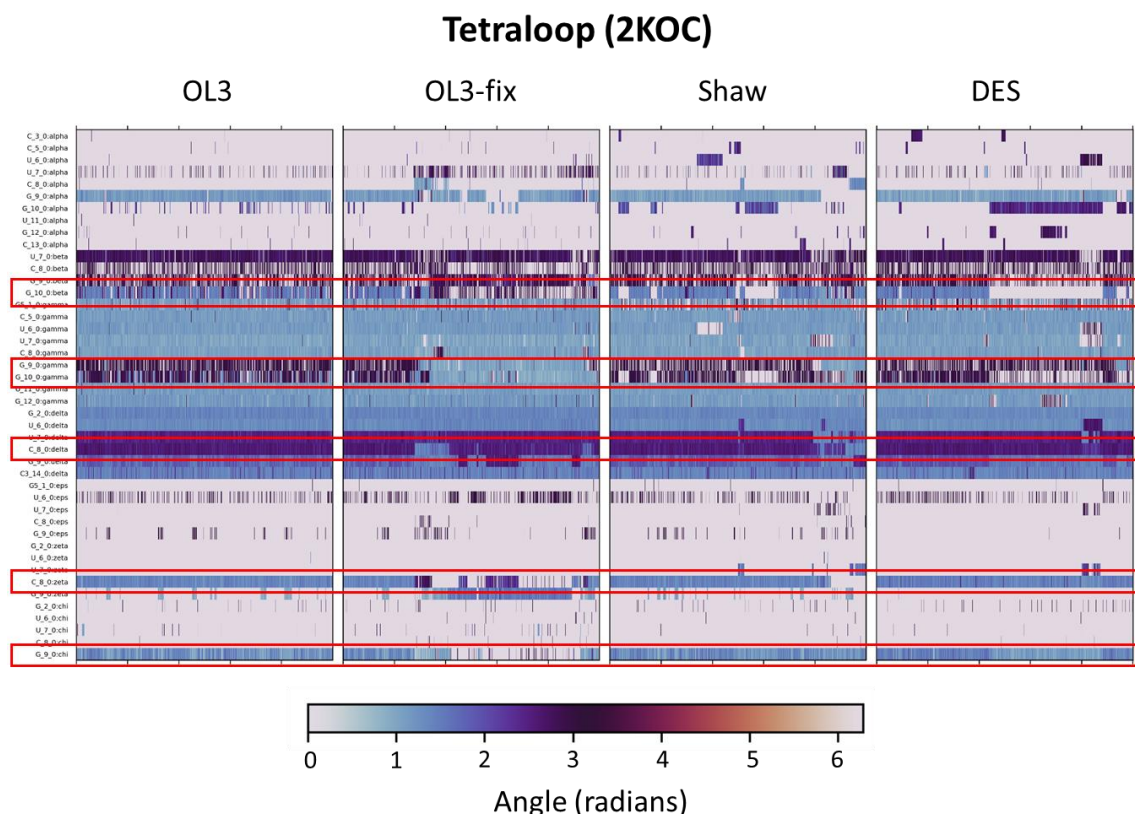


Figure 3.31. Dihedral angle time series for the tetraloop motif (2KOC).

- For the G4 motifs, dihedral-angle plots were substantially longer due to the large number of monitored dihedrals (100–200 per system), making detailed per-dihedral interpretation impractical. Nonetheless, a qualitative assessment revealed consistent trends across force fields.

In both 7PS8 and 8UTG G4 structures, the ϵ dihedrals remained remarkably stable throughout the simulations, indicating preservation of the backbone geometry within the tetrad core. In contrast, the α , β , δ , and χ dihedrals exhibited extensive fluctuations, reflecting the intrinsic flexibility of the peripheral

loops and corner regions. Although the current dihedral analysis is insufficient to pinpoint which dihedral changes drive the G4 deformations, particularly the dynamic corner regions, the overall pattern supports the notion that backbone flexibility in peripheral regions, rather than disruptions in the tetrad stack, contributes to the RMSD increases. To explore in detail which dihedral transitions drive these dynamic structural changes, a more rigorous and automated analysis would be required to quantitatively link dihedral time series with the observed conformational rearrangements.

- Regarding the kissing-loop motifs, all force fields except DES maintained structural stability after equilibration. In the 2MI0 kissing-loop, dihedrals remained largely stable with OL3, OL3-fix, and Shaw, exhibiting only minor deviations: in OL3-fix, two α dihedrals, one γ , two δ , and one ζ showed changes of ca. 1.0–1.5 rad. In contrast, the DES FF, which displayed the highest RMSD, exhibited a pronounced shift in the δ dihedral of G15, potentially contributing to the observed structural rearrangement.

For the 2D1B motif, OL3 and OL3-fix also preserved dihedral stability. Shaw maintained generally stable dihedrals as well, but adopted a distinct conformation compared to OL3 and OL3-fix (**Figure 3.26**). By contrast, DES exhibited continuous fluctuations in RMSD throughout the 1 μ s simulations. Notably, despite these fluctuations, χ , ϵ , and γ dihedrals remained relatively stable, suggesting that backbone flexibility primarily involves α , β , δ , and ζ angles.

- Finally, for the TTR motif, δ and ζ dihedral angles exhibited fluctuations across all force fields, despite the fact that OL3 and OL3-fix FFs preserved stability, suggesting that these variations did not directly impact the loss of tertiary interactions. In contrast, γ and χ dihedrals appeared largely stable throughout the simulations. These observations are qualitative, given the large size of the dataset, and as previously noted, a more robust and automated analysis would be required to quantitatively assess the contribution of specific dihedral transitions to the observed structural changes.

NOE Analysis

For the NMR RNA structures of our study and for those that experimental NOE values were available (internal loop 2H49, tetraloop 2KOC, G4 7PS8, kissing-loop 2MI0), we computed the NOE-derived distances and analyzed their agreement with NMR observables. The analysis was performed separately for intra-nucleotide and inter-nucleotide NOEs to assess force field performance at different levels, and a total agreement metric was calculated to summarize overall accuracy.

- The internal-loop motif exhibited high agreement with experimental NOE data, with intra-nucleotide NOEs ranging from 95.7% to 98.9% and inter-nucleotide NOEs showing slightly lower agreement between 82.4% and 86.8%. Notably, Shaw showed the best performance for intra-nucleotide NOEs (98.9%), OL3 performed best for inter-nucleotide NOEs (86.8%), and DES

exhibited the lowest agreement for inter-nucleotide NOEs (79.4%). Despite these variations in intra- and inter-nucleotide performance, OL3, OL3-fix, and Shaw all achieved the same overall agreement of 91.9% (Figure 3.32).

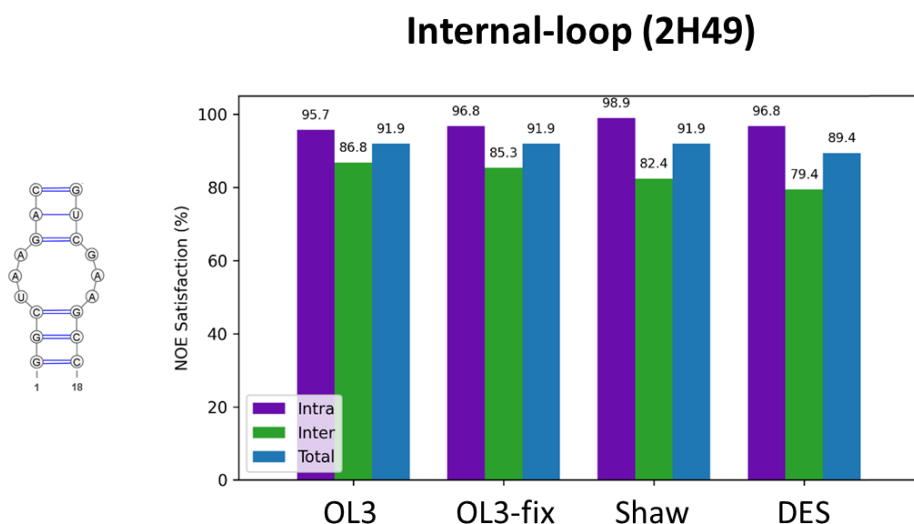


Figure 3.32. Agreement between MD-derived NOE distances and experimental NOEs.

- The tetraloop exhibited broadly comparable NOE satisfaction across all force fields. Intra-residue agreement ranged from 85.1% to 87.9%, while inter-residue agreement varied between 80.5% and 87.3%. The overall NOE performance for the tetraloop was similarly consistent, falling within 83.3–86.9%. Among the force fields, OL3-fix showed the lowest total agreement, consistent with the early structural rearrangement observed in its MD trajectory. In contrast, the conformational change seen with the Shaw FF did not markedly affect its NOE agreement, as the structural deviation occurred late in the simulation (after ca. 800 ns), leaving most of

the trajectory in good agreement with the experimental NOEs. Notably, the DES FF achieved the highest overall NOE satisfaction for this motif (**Figure 3.33**).

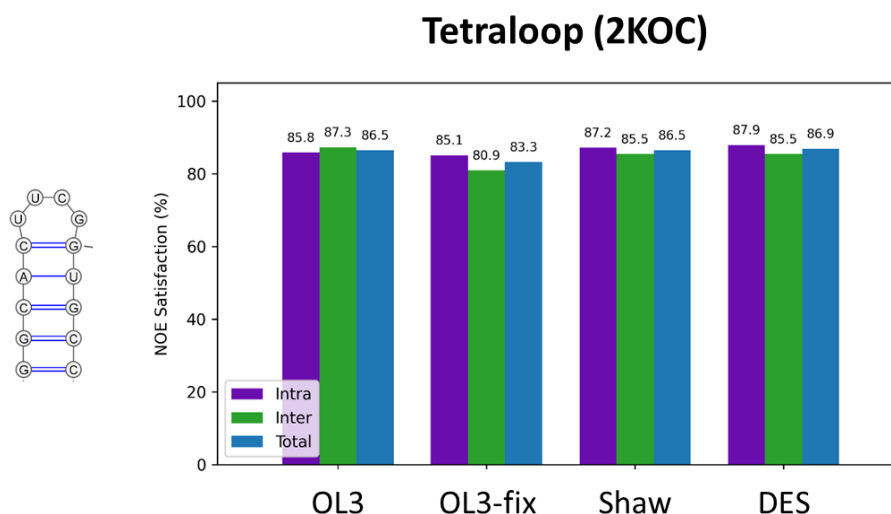


Figure 3.33. Agreement between MD-derived NOE distances and experimental NOEs.

- The G4 motif displayed the lowest NOE agreement among the systems analyzed. Intra-residue agreement ranged from 71.5% to 75.1%, whereas inter-residue agreement was substantially lower, between 52.5% and 62.6%, with OL3 showing the best performance in both the inter-residue and total categories (total agreement: 68.6%). Interestingly, the Shaw FF achieved the highest intra-residue satisfaction but simultaneously the lowest inter-residue agreement, suggesting that maintaining local (intra-residue) contacts may come at the expense of longer-range interactions in this force field. Overall, the reduced NOE agreement is consistent with previous observations: while the

central G4 tetrad core remains stable, the peripheral loops and corner regions are highly dynamic, likely disrupting many inter-residue NOEs. The stable core alone accounts for approximately 60% of the satisfied NOEs, whereas the flexible outer regions reduce the overall agreement with experimental data (**Figure 3.34**).

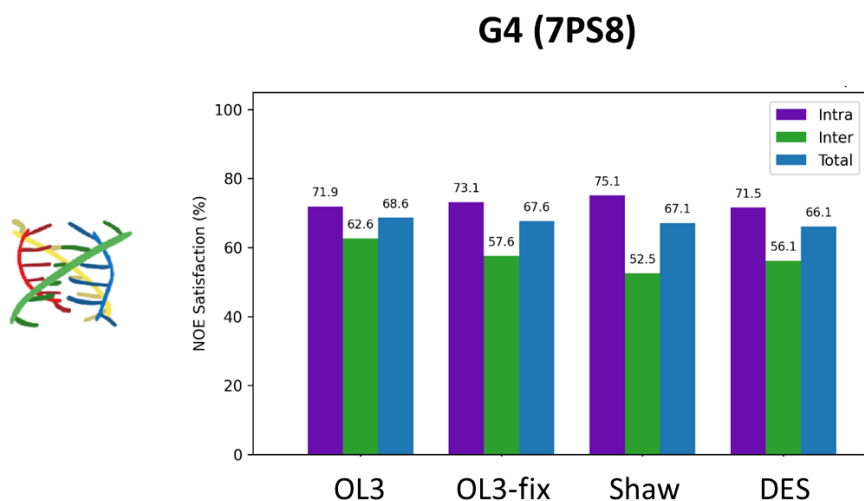


Figure 3.34. Agreement between MD-derived NOE distances and experimental NOEs.

- The kissing-loop motif 2MI0 exhibited excellent agreement with the experimental NOEs, with intra-residue, inter-residue, and total satisfaction values all exceeding 95% across the force fields. This high level of agreement indicates that, despite the structural rearrangements observed after equilibration, the resulting conformations remain well within the range supported by the experimental data. In other words, the post-equilibration structures adopted by the simulations are not artifacts but fall comfortably

within the ensemble of experimentally allowed NOE-compatible conformations (**Figure 3.35**).

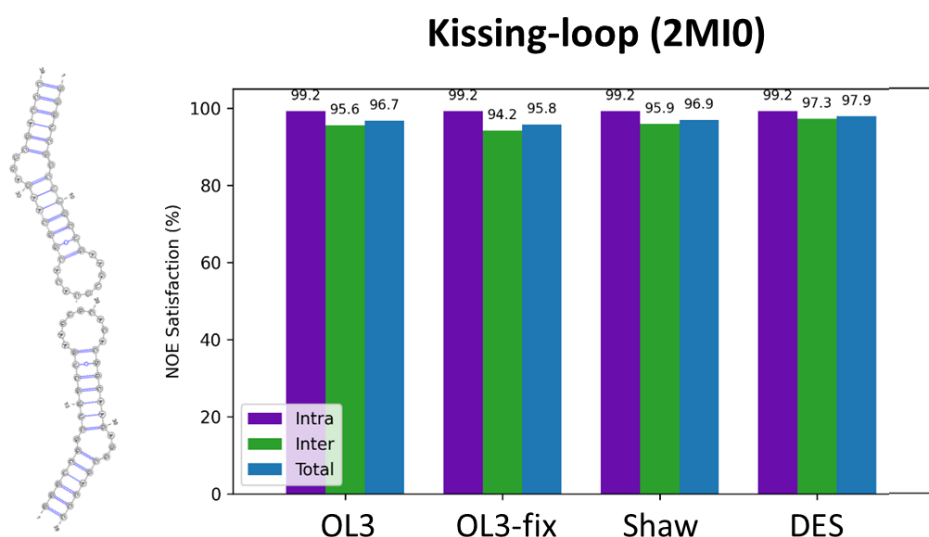


Figure 3.35. Agreement between MD-derived NOE distances and experimental NOEs.

Across all RNA motifs evaluated, OL3 generally showed the most consistent agreement with the experimental NOEs, particularly in the inter-residue and total categories for the internal loop, G4, and tetraloop structures. Shaw frequently achieved the highest intra-residue agreement, but this often coincided with lower inter-residue performance, suggesting that maintaining local geometry sometimes occurred at the expense of longer-range contacts. OL3-fix performed comparably to OL3 in most cases but showed reduced agreement for the tetraloop due to its structural rearrangement. DES exhibited the lowest overall NOE satisfaction, particularly for inter-residue distances, consistent with its tendency to destabilize or over-flexibilize larger RNA

motifs. Taken together, NOE agreement trends complement the RMSD, H-bond, and dihedral analyses, reinforcing OL3 as the most reliable performer, while highlighting DES as the least consistent, especially for RNA constructs with >40 nt.

3.3. Targeting tau pre-mRNA with Small Molecules

In this chapter, we focus on targeting tau (tubulin associated unit) pre-mRNA using small molecules. The microtubule-associated protein tau (MAPT) undergoes alternative splicing of exon 10 to generate the 3R- and 4R-tau isoforms. Proper neuronal function requires a balanced 3R/4R ratio, as tau plays a critical role in stabilizing microtubules in neurons. However, inclusion of exon 10, which depends on the hairpin stability of the following intron 10, leads to overproduction of 4R-tau. This aggregation has been linked to a group of neurodegenerative diseases, known as tauopathies.

Here, we aim to target the hairpin of tau pre-mRNA using small molecules to stabilize it and restore the equilibration between 3R and 4R isoforms. To achieve this, we first screened a ligand library by performing an AI-based similarity search on a known binder.⁷¹ The top-ranked candidates were then examined by NMR spectroscopy (1D¹H NMR and WaterLOGSY) to identify interaction with the target. The most promising compounds were biologically evaluated (by experimental collaborators) and the best candidate was examined in cellular assays. Then, this compound was further characterized *in silico* to elucidate its interaction pattern with the target RNA.

Our results show that this pipeline successfully identified a ligand that binds the target RNA with a richer interaction network than the reference molecule. This approach demonstrates that combining AI-based screening, NMR validation, and computational characterization provides an effective strategy for discovering RNA-targeting small molecules.

3.3.1. AI-based Similarity Screening of Ligand Library

The ligand library used for the structure-based drug design (SBDD) targeting tau pre-mRNA consisted of an in-house guanylylhydrazone (GH) collection provided by our collaborator Prof. Seneci (University of Milan, Italy). The GH library contains 96 structurally diverse molecules including monoaryl, bisaryl, polyaryl, bicyclic, and tricyclic scaffolds. The GH library was selected for exploration as it was originally inspired by GM118,⁷¹ a carbazole analogue known to bind the A-bulge of tau pre-mRNA hairpin, stabilizing it primarily *via* stacking interactions.

To prioritize compounds for further experimental testing, we performed an AI-based similarity analysis aimed at identifying GH molecules structurally related to GM118. The selected hits were subsequently subjected to NMR spectroscopy for experimental validation.

Methods

To assess molecular similarity we leveraged MolProphet,¹⁸⁴ a DL-based online platform, developed for accelerating lead optimization of small

molecules. The GH library was provided as an input in SMILES format and GM118 was selected as the reference structure.

The software allows filtering compounds based on physicochemical properties such as molecular weight (MW), liposolubility (LogP), topological polar surface area (TPSA), number of rings (NR), rotatable keys (Nrot), and number of hydrogen bond donors (HBD) and acceptors (HBA). In this initial screening, property thresholds were adjusted to ensure inclusion of all molecules in the library.

Similarity analysis was performed in two stages: first using 2D similarity, followed by 3D similarity assessment, enabling comprehensive evaluation of structural resemblance to GM118.

Results

The DL-based similarity analysis on the GH library (96 compounds) yielded ranked lists for both 2D and 3D similarity to the reference ligand GM118. The 2D similarity values ranged between 0.562-0.935, whereas the 3D similarity values between 0.659-0.949. Notably, fifteen of the compounds were consistently appearing among the top 20 hits in both 2D and 3D similarity analysis. Thus, these 15 compounds were prioritized for experimental validation using $1D^1H$ NMR and WaterLOGSY experiments. A summary of the 15 NMR candidates along with their 2D, 3D similarity scores, and their ranking is reported in **Figure 3.36**.

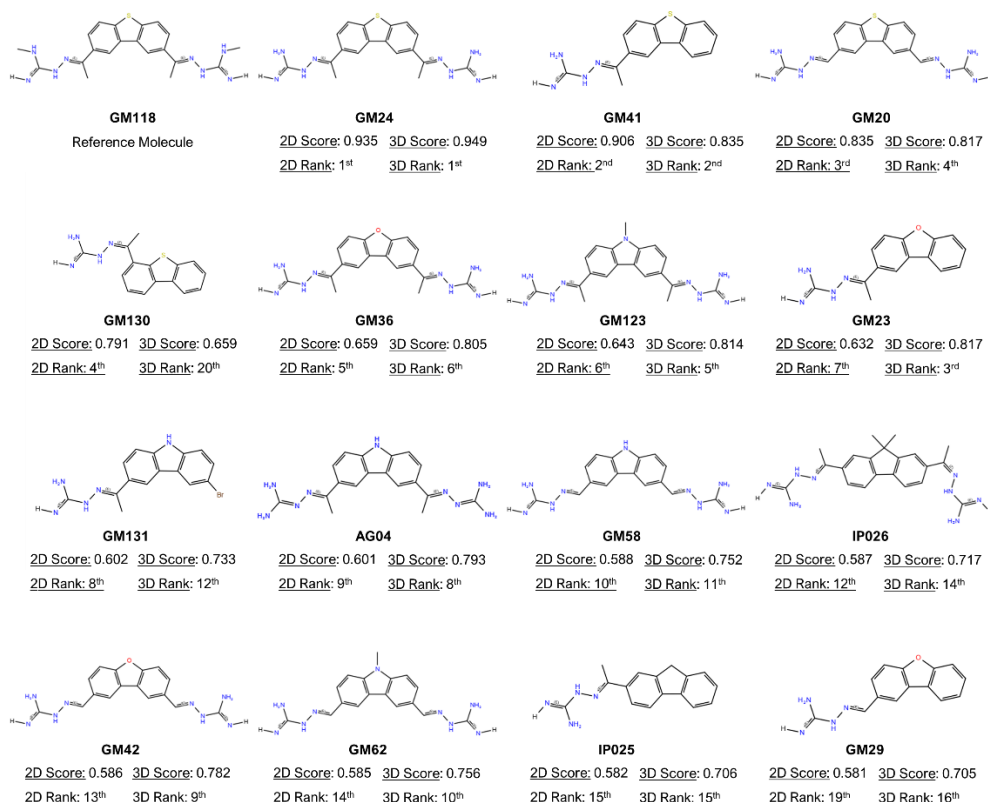


Figure 3.36. Top 15 ligands selected for NMR screening based on 2D and 3D similarity to GM118.

All 15 selected molecules from the GH library share a tricyclic scaffold but exhibit important structural variations. These differences include: i) bridgehead atom identity, with S (as in GM118) present in 4/15 compounds, O in 4/15, NH in 3/15, *N*-Me in 2/15, and CH₂ in 1/15; ii) number of GH groups, with 9/15 compounds bearing two symmetric GH units (as in GM118) and 6/15 containing a single asymmetric GH group; and iii) GH substitution pattern, where 10/15 molecules incorporate a methyl at the α -position (as in

GM118), while 5/15 lack this modification (**Figure 3.36**). Consequently, these features indicate that the selected set captures substantial chemical diversity within the GH library, making it a well-balanced panel for subsequent NMR-based screening.

3.3.2. 1D¹H NMR On Selected 15-ligand Library

1D ¹H NMR was used to directly monitor tau pre-mRNA hairpin. Because imino protons (from G and U) resonate only when involved in base-paired regions, changes in RNA signals confirm interaction with a ligand.

Methods

For RNA the 1D ¹H NMR experiments, constructs were purchased from Sigma-Aldrich. Oligoribonucleotides that form the duplex, including r(5'-CCGGCAGUGUG-3') and r(5'-CACACGUCGG-3') to mimic the stem of tau exon 10-intron hairpin, were deprotected and desalted with PD-10 columns according to the manufacturer's recommended protocol. Equimolar amounts of each strand were combined with NMR buffer (10 mM KH₂PO₄/K₂HPO₄ and 0.05 mM EDTA, pH 6.0) to give final concentrations of 25 μM. 10% D₂O was included in each sample to provide a lock signal. To form duplexes, RNA samples were reannealed by heating to 80 °C for 3 min and were slowly cooled to room temperature.

For all NMR experiments Shigemi NMR tubes (Shigemi, Inc.) were used. The 1D ¹H NMR spectra were acquired with a Bruker Avance NMR spectrometer operating at 700 MHz ¹H Larmor frequency equipped with a z-gradient 5 mm

triple-resonance probe head. The spectra were acquired at 282 K in 90% H₂O/10% D₂O. 1D ¹H NMR spectra were collected for RNA alone, RNA in complex with each of the 15 selected ligands, and RNA in complex with the reference ligand GM118. Each compound was titrated to a final ligand/RNA molar ratio of 2:1 (50 μM ligand:25 μM RNA).

Results

Our 1D ¹H NMR studies on the tau RNA hairpin alone, in complex with each of the 15 selected ligands, and in complex with the reference binder GM118 were collected to identify the chemical shift and line width changes to the imino proton resonances were monitored as an indication of interaction. The spectrum of the RNA alone and the RNA/GM118 spectrum were compared with previously published data, confirming consistent imino proton patterns.⁷¹ Such validation ensures that our protocol reliably reproduce expected RNA folding and ligand-induced perturbations, thereby providing a solid basis for the NMR screening of the 15-ligand library.

RNA alone & RNA/GM118 Reference Spectra

The 1D ¹H spectra acquired for the RNA alone displayed the expected imino resonances for G and U nt, fully consistent with the literature. Upon addition of GM118, spectral changes were provoked. Specifically, nt in close proximity to the binding site such as G4, G7, U8, and G17 showed strong decreases in the signal intensity, while U18 experienced a chemical shift from ca. 12.1 ppm to ca. 11.9 ppm. Furthermore, more distal nt including G3, G9, G20, G21 also displayed reduced intensities with G9 and G20 presenting a line broadening,

whereas U10 revealed an increase signal intensity relative to tau RNA alone (**Figure 3.37**).

These perturbations, including effects on residues distant from the canonical binding site, suggest that GM118 binding propagates structural changes in the broader RNA stem, in agreement with previously reported behavior.

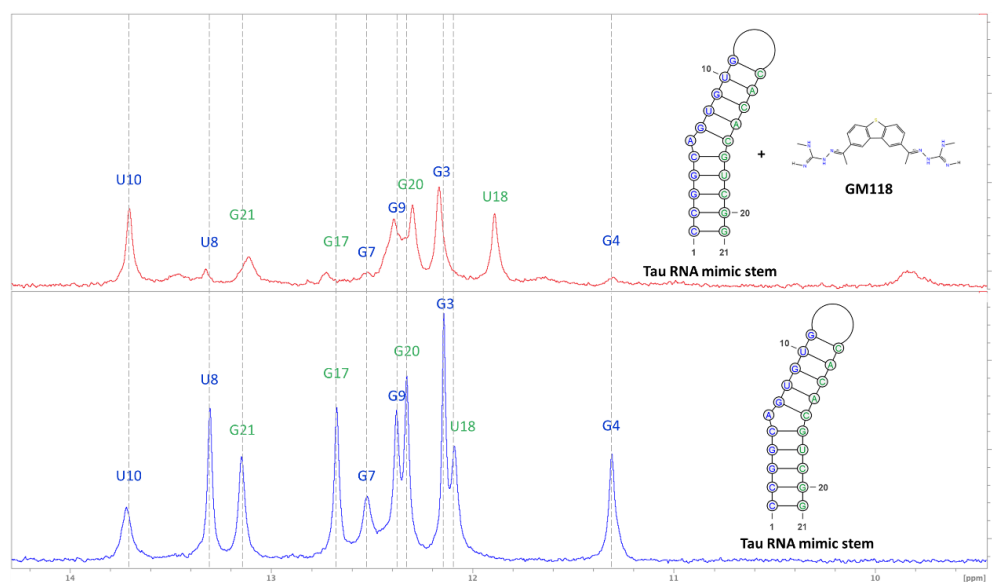


Figure 3.37. 1D ¹H spectra acquired of tau RNA stem mimic duplex alone (bottom) and in complex with the positive control GM118. Our spectra are consistent with those reported in the literature.⁷¹

RNA/ligand library Spectra

We then examined the 1D ¹H spectra of tau RNA in complex with each of the 15 selected ligands. Notably, all 15 ligands displayed chemical shifts, line broadening, or complete disappearance of G and U signals, indicating interaction (**Figure 3.38**).

Some representative examples include:

- RNA/GM36, which exhibited similar changes to the positive control GM118, suggesting a comparable binding mode.
- RNA/IP025, which induced perturbations in the binding site region (disappearance of G4, U8, and G17, shift of U18).
- RNA/GM58, which also presented changes in the binding site region (U8 and G17 disappearance, G4 shift, and G9, U18 perturbations)

Across the library, the most commonly perturbed signals were G4, U8, G17, and U18, consistent with ligands interacting in the vicinity of the known GM118 binding site, supporting the hypothesis that most ligands engage a similar pocket.

More distal nucleotides, such as U10, maintained its peak for all ligands but exhibited a reduce of intensity in 4/15 cases; G21 remained visible in 10/15 cases, disappeared in 4/15, and shifted only in the complex with ligand IP025; G20 was present in 13/15 spectra, whereas in GM29 and GM113 it disappeared entirely and in other complexes displayed reduced intensity, broadening, or chemical-shift changes. These effects suggest propagation of structural changes through the duplex.

Although the 1D ^1H spectra indicated perturbations across the full length of our selected ligands, we must consider the limitations of this method, which can produce false positives. For example, aggregation or partial degradation can broaden imino resonances and obscure true ligand-induced changes. Additionally, strong line broadening may arise from nonspecific interactions or sample instability rather than genuine binding. In our study, ligands GM29, GM23, AG04 lost multiple resonances, such as G4, U8, U10, G17, G20, and G21, probably due to RNA precipitation. For these reasons, we complemented

1D ^1H NMR with Water-LOGSY, a ligand-based NMR technique to elucidate the interaction in more detail and distinguish real binders from false positives.

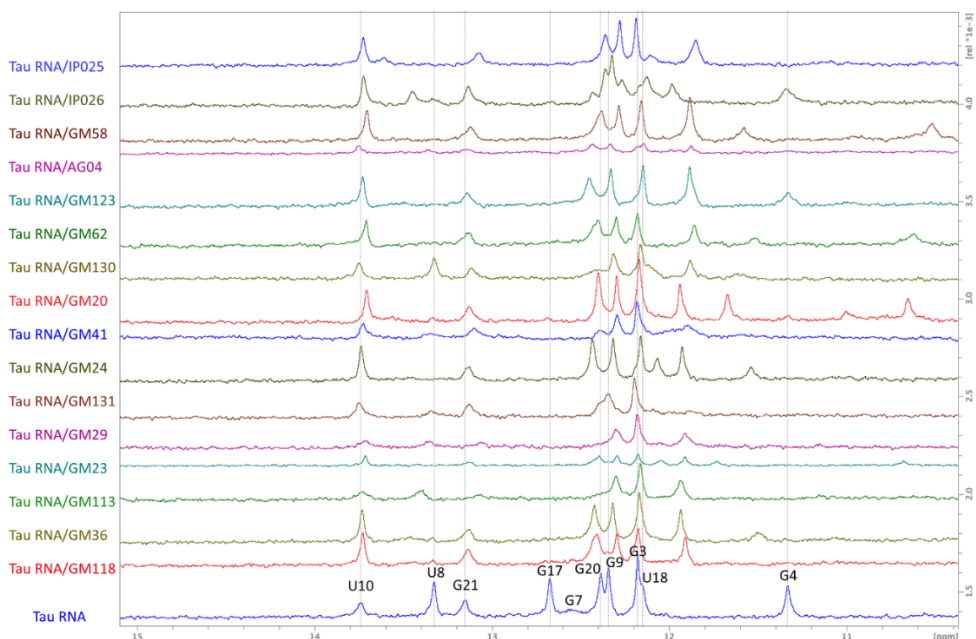


Figure 3.38. 1D ^1H NMR screening of the 15-ligand GH library against tau pre-mRNA.

3.3.3. *WaterLOGSY On Selected 15-ligand Library*

WaterLOGSY is another NMR spectroscopy method used to investigate the interactions between small molecules and biomolecules, such as RNA, in aqueous solution. In this approach, magnetization transfer is mediated by water molecules that interact with the binding site. Ligand resonances that receive magnetization through binding to the biomolecule appear with opposite signal phase compared to resonances of non-binding ligands. This technique is particularly suitable for our study as it complements the 1D ^1H

NMR experiment, giving us the opportunity to explore tau RNA-ligand interaction by focusing on both the macromolecule RNA (1D ^1H) and the ligand (WaterLOGSY). With WaterLOGSY, perturbations in ligand signals will confirm binding. In this way, a more reliable assessment of ligand–RNA interactions will be accomplished.

Methods

The WaterLOGSY NMR spectra were acquired with a Bruker Avance NMR spectrometer operating at 600 MHz ^1H Larmor frequency equipped with a z-gradient 5 mm triple-resonance probe head. WaterLOGSY (water-ligand observed via gradient spectroscopy) experiments were carried out on samples of the WT tau stem mimic duplex, prepared as described in the previous subsection. For each of the 15 ligands, two spectra were acquired: i) the ligand alone, and ii) the ligand in the presence of the RNA, ensuring direct comparison of bound and unbound states. Samples for WaterLOGSY experiments contained 25 μM RNA and 500 μM ligand, corresponding to a 1:20 RNA:ligand ratio. The buffer used was the same as described in the previous subsection, whereas the spectra were acquired at 298 K.

Results

Tau RNA/GM118 Reference Spectra

We first acquired the WaterLOGSY spectrum of the reference ligand GM118 to validate our experimental setup (**Figure 3.39**). As expected, GM118 displayed negative WaterLOGSY signals, fully consistent with previously reported data.⁷¹ This agreement confirms both the correct folding of the tau

RNA stem mimic and the reliability of our WaterLOGSY experimental protocol.

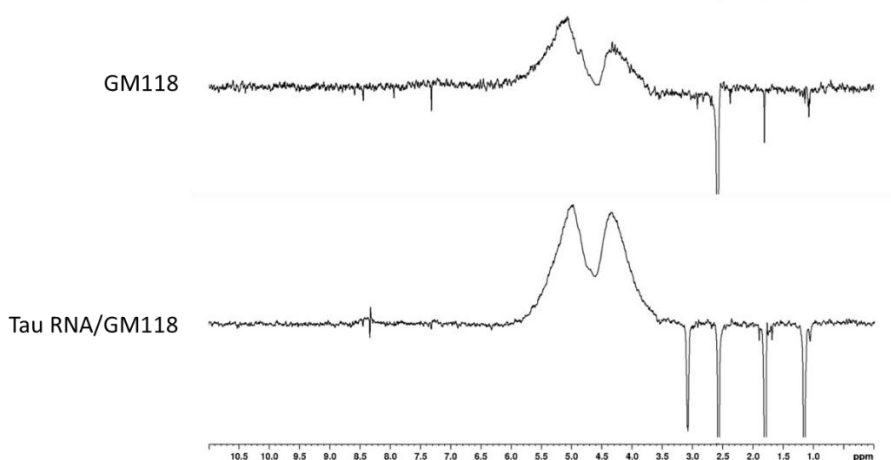


Figure 3.39. WaterLOGSY NMR spectra of the positive control GM118 alone (top) and tau RNA/GM118 complex (bottom). Ligand resonances appear between 7 and 9 ppm, water signals between 4 and 6 ppm, and buffer signals between 1 and 3.5 ppm.

Tau RNA/Ligand Library Spectra

The WaterLOGSY spectra of tau RNA in complex with each of the 15 selected ligands were then recorded to evaluate binding. Based on the presence of positive signals, the compounds were classified as positive binders, or non-binders.

Positive binders (6 ligands)

GM36, GM131, GM113, IP025, GM29, and GM41 displayed clear, unambiguous positive WaterLOGSY signals, demonstrating robust binding to the tau RNA hairpin. Chemically, this group is enriched in: Mono-GH asymmetric scaffolds (5/6), with GM36 as the only 2-GH symmetric ligand;

α -methyl substitution (5/6), with GM29 lacking the α -Me group; and O at the bridgehead (3/6), with the remaining examples containing S, NH, or CH₂ (IP025). These structural elements suggest that both the GH motif orientation as well as the nature of the bridgehead atom influence RNA recognition. Representative spectra of GM113, GM36, and IP025 are illustrated in **Figure 3.40**.

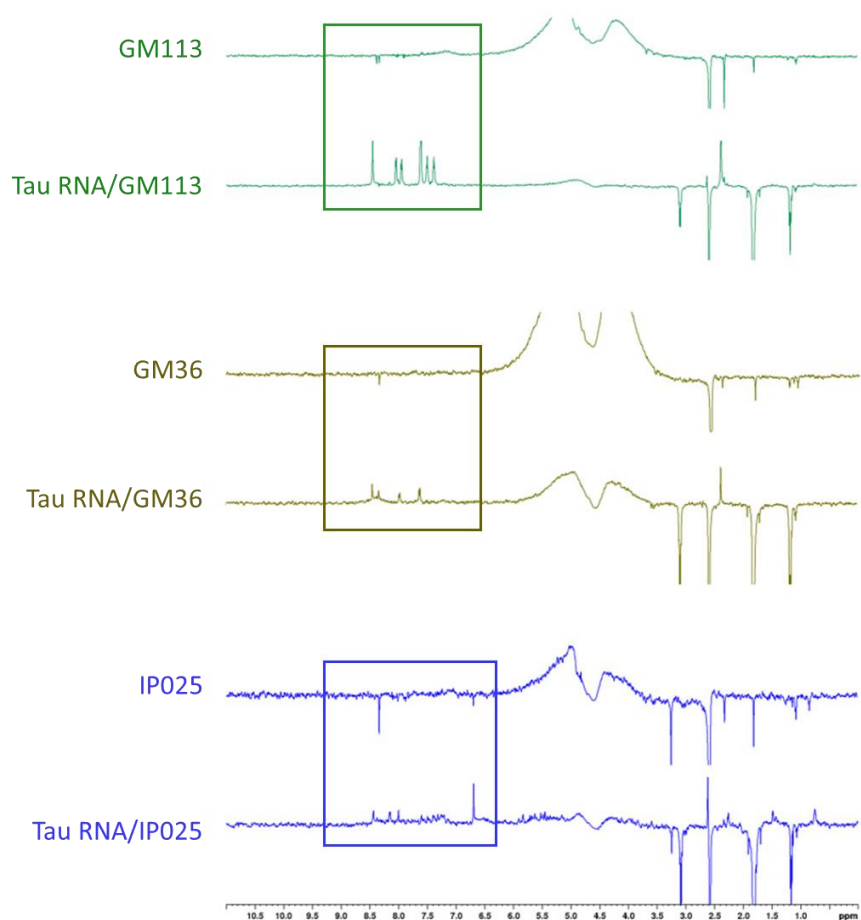


Figure 3.40. WaterLOGSY NMR spectra of the positive binders IP025 (blue), GM36 (olive), and GM113 (green), recorded both alone and in the presence of tau RNA.

Ligand resonances appear between 6.3 and 8.5 ppm, water signals between 4 and 6 ppm, and buffer signals between 1 and 3.5 ppm.

Non-binders (9 ligands)

The remaining compounds exhibited signals in the same phase as the buffer or flat WaterLOGSY signals, suggesting that these compounds did not interact with RNA or faced challenges due to solubility issues. A representative spectrum for GM24, where the signals appear in the same phase as the buffer is illustrated in **Figure 3.41**.

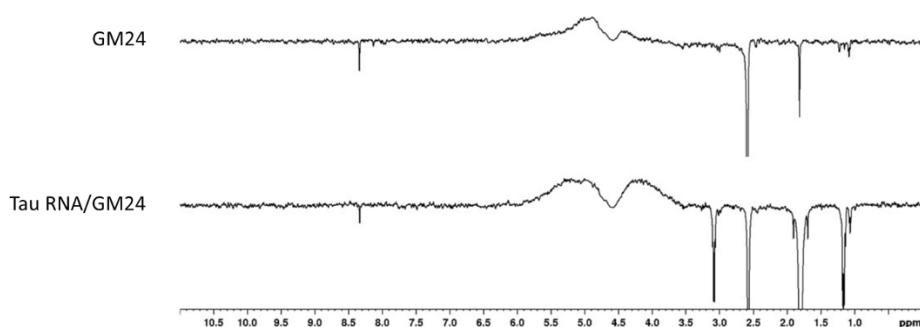


Figure 3.41. WaterLOGSY NMR spectra of GM24 recorded alone and in the presence of tau RNA. Ligand resonances remain in the same phase, indicating lack of interaction.

The WaterLOGSY results reinforce our earlier observations that 1D¹H NMR alone can overestimate binding, mainly because alter of imino signals can arise from aggregation or instability. Thus, WaterLOGSY NMR is an effective strategy to differentiate binders from false positives.

Based on the WaterLOGSY evaluation, the six most reliable binders were selected for further evaluation through biological assays to assess their activity.

3.3.4. Biological Screening on 6 Ligands Derived from NMR Experiments

Following the NMR screening, six ligands (GM36, GM131, GM113, IP025, GM29, and GM41), which suggested interaction in both 1D ^1H and WaterLOGSY spectroscopy, along with the positive control GM118, were selected for biological evaluation. These assays were performed by our collaborator Dr. Vania Broccoli and Dr. Simone Alberti (IRCCS Ospedale San Raffaele, Italy) to assess their ability to modulate tau exon 10 alternative splicing.

For activity evaluation, the ligands were screened in HeLa cells transfected with a WT plasmid, and the 3R/4R tau isoform ratio fold-change was quantified using two different calculation methods: i) Fiji-based image analysis, and ii) an in-house quantification method.

Biological Activity Results

Across both analytical pipelines, IP025 demonstrated the most pronounced effect, showing an increase in the 3R/4RR ratio to 1.1 and 1.7 in the Fiji and in the in-house analysis, respectively. This effect was consistent between three replicates, supporting the biological effect.

In contrast, GM36 showed inconsistent behavior. In the in-house pipeline, the two replicates displayed large disagreements, one close to the baseline and one approaching a value of 1.5, artificially elevating the mean value with respect to the vehicle DMSO. As for the Fiji calculations, GM36 exhibited consistency within the two replicates, both close to the DMSO control, suggesting minimal or no biological activity. This discrepancy indicates that the apparent positive effect in the in-house calculation is likely to represent a technical artifact rather than true activity.

The remaining ligands including GM29, GM41, GM113, GM131, and the reference molecule GM118 exhibited consistently values close to the baseline using both pipelines (ca. 1.0 in Fiji analysis and 1.0-1.2 with in-house method). These results indicate low or no effect on the 3R/4R tau isoform ratio (Figure 3.42).

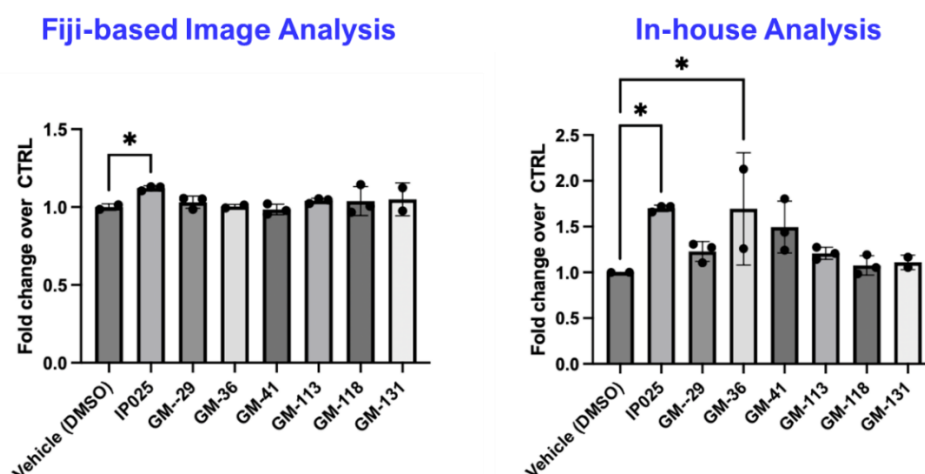


Figure 3.42. 3R/4R % tau isoform ratio fold-change quantification over DMSO control in HeLa transfected with WT plasmid using Fiji-based image analysis and in-house analysis.

Overall, the reference molecule GM118, as well as GM29, GM41, GM113, and GM131 clustered around DMSO values with both computational pipelines, demonstrating no significant activity. GM36 appeared inactive in the more reliable Fiji analysis and produced inconsistent values in the in-house method, suggesting that its apparent activity is not biologically meaningful. Only IP025 produced consistent, reproducible increases in the 3R/4R isoform ratio across both pipelines and was therefore selected for further computational investigation to elucidate its interaction with the tau RNA hairpin.

In addition, following the experiments performed on the WT tau pre-mRNA, which allowed for initial identification of binders of the A-bulge pocket of tau pre-mRNA, further experimental studies will be carried out on the mutated specific variant +14 DDPAC. Specifically, 1D ^1H and WaterLOGSY NMR studies are ongoing on the stem of DDPAC tau in presence of the identified compound IP025, along with biological studies in HeLa cells that feature the mutated form of the gene. These additional experiments will allow us to more accurately assess the activity of IP025 in a disease-relevant context.

3.3.5. *In-silico Analysis of Tau-pre-mRNA/IP025*

After the biological assays, in which IP025 was identified as the only effective ligand capable of increasing 3R/4R tau isoform ratio, its interaction with the tau RNA hairpin was investigated using computational methods. Specifically, IP025 was docked into the A-bulged cavity of tau pre-mRNA, followed by an analysis of its predicted intermolecular interactions.

Methods

To identify an appropriate 3D structure of tau pre-mRNA for docking, we searched the PDB and identified five available structures: 6VA1, 6VA2, 6VA3, 6VA4, and 2KGP. These include one NMR *apo* structure (6VA1) of tau pre-mRNA exon 10 stem, and four NMR *holo* structures complexed with small molecules GM118 (6VA2), MQC (6VA3), MIP (6VA4), or MIX (2KGP).

Comparative studies have demonstrated that *holo* receptor structures yield better enrichment in docking (ca. 70%) compared to the *apo* (ca. 20%) or modelled forms (ca. 10%).²⁵³ Therefore, the *apo* structure 6VA1 was excluded. To identify the most suitable structure among the remaining structures, we compared the 2D and 3D similarity of IP025 to each co-crystallized ligand using MolProphet (**Table 3.6**). The reference ligand GM118 (referred also as MH5 in PDB), bound in 6VA2, exhibited the highest similarity to IP025 in both metrics (3D Score 0.653, 2D score 0.560), thus it was selected for the docking calculations.

Table 3.6. 2D and 3D similarity scores with reference to IP025 compound using MolProphet.

Ligand	3D Similarity Score	2D Similarity Score
GM118 (MH5)	0.653	0.560
MQC	0.625	0.329
MIP	0.631	0.393

MIX	0.519	0.333
------------	-------	-------

The 6VA2 entry corresponded to an NMR structure from *Homo sapiens* and included a 20-conformer ensemble. The lowest energy conformer (model 1) was selected for our study.

Protein & Ligand Preparation

The Protein Preparation Wizard tool in Maestro 14.3 (Schrödinger Suite 2024.3)²⁵⁴ was used to prepare the RNA structure. Because the tool is designed for proteins and not nucleic acids, only missing hydrogens were added, and structure minimization was performed ignoring other options designed for proteins such as protonation state assignment and optimization.

IP025 was built in 2D sketch of the same software and was prepared using LigPrep with standard parameters. The lowest-energy conformer was used for subsequent docking.

Docking Setup

Glide (Schrödinger Suite 2024.3) was employed for docking calculations. Although Glide was originally designed for protein-small molecule docking, starting with Schrödinger 2024.2 release, an alternative Glide scoring function is available that is tuned for RNA receptors, with reweighted Coulomb, vdW, and internal energy terms.

To generate the grid box for docking, the “Receptor Grid Generation” was employed within Maestro. The co-crystallized ligand (GM118) was excluded from the grid box and used to define the binding site. Receptor grids were generated in Glide¹²⁰ using OPLS3e atom typing. Binding-site cavities were identified on a 26×26×26 initial grid with a 1.0 Å spacing (125 cavity points detected). RNA vdW radii scaling was set by 1.0, polarity cutoff was set to 0.25 by default, and the final docking grid dimensions were produced at (nx, ny, nz) = (72, 72, 72) grid points.

Glide docking was finally performed using the default parameters (Standard Precision scoring function) and the RNA receptor scoring.

Docking Results

Docking of IP025 on the tau RNA A-bulged pocket yielded nine poses. Although all poses occupy the same planar binding region as the reference GM118 (**Figure 3.43A**), they can be organized into three orientation clusters based on the relative directionality of the tricyclic core and methyl substituent. In the most populated cluster I, 5/9 poses (poses 1, 2, 3, 4, and 8) converged to a similar orientation, in which the methyl group directed toward the interior of the RNA binding site. In cluster II, 3/9 poses adopted a mirror orientation along the x-axis with respect to the first cluster, in which the methyl group is facing the solvent-exposed region. Finally, cluster III includes a single pose (pose 9) that is approximately parallel to the y axis, with the methyl group facing towards the binding pocket. Its orientation deviates from both cluster 1 and cluster 2, suggesting a less common but geometrically permissible alternative binding mode.

Interaction Analysis

A representative structure from each cluster was selected for detailed interaction analysis. Non-covalent interactions were evaluated using the following distance cutoffs: 3.5 Å for hydrogen bonds, 4.0 Å for salt bridges, and 5.0 Å for π - π stacking and cation- π interactions.

Despite the different orientation of the methyl substituent that distinguishes Cluster I and Cluster II, their interaction patterns are largely similar. These include a salt bridge between the positively charged guanidinium nitrogen of the of IP025 and the phosphate oxygen of G17, a cation- π interaction between the same guanidinium nitrogen and the aromatic ring of C16, as well as three π - π stacking interactions between the six-members rings of the tricyclic core of IP025 and C5, G7, and C16 (**Figure 3.43C, D**).

Cluster III exhibits the same number and types of interactions as cluster I and II, however, with a distinct spatial arrangement that engages different nucleotides. Specifically, the salt bridge occurs with the phosphate group of C5, the cation- π interaction with C5, and the π - π stacking interactions with G7, C17, and G17 (**Figure 3.43E**). Thus, while the chemical interactions are conserved, the binding register shifts relative to the A-bulged pocket.

In contrast, the co-crystallized reference ligand GM118 displays a different interaction profile. GM118 forms five π - π stacking interactions between the full tricyclic system and nucleotides G7, C16, and G17. However, both guanidinium groups remain solvent-exposed and do not establish salt bridges or cation- π interactions with the RNA, unlike IP025 (**Figure 3.43B**).

Overall, both ligands secure the binding to tau RNA using stacking interactions, explaining the NMR experiments. However, IP025 forms unique salt bridge and cation- π interactions in all binding poses across all clusters,

interactions absent from the reference GM118. These additional stabilizing contacts can explain the distinct biological behavior of IP025, particularly its ability to modulate the 3R/4R tau ratio. Such subtle differences are characteristic of small-molecule RNA ligands and can influence selectivity, specificity, and biological activity.

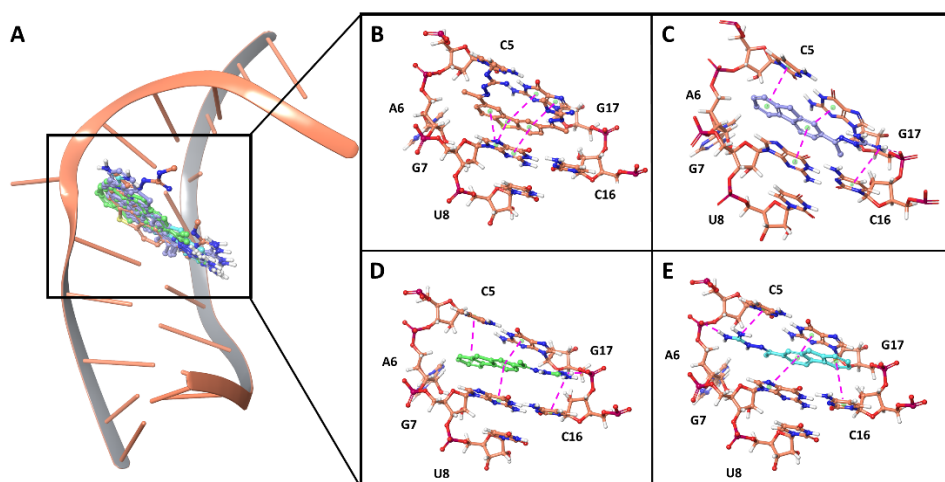


Figure 3.43. Docking results. A) Superposition of all nine IP025 docking poses together with the co-crystallized reference ligand GM118. Cluster I is shown in lilac, Cluster II in green, Cluster III in cyan, and GM118 in peach. B) Interactions between the tau pre-mRNA and GM118. C-E) Interactions between the tau pre-mRNA and the representative pose of IP025 from Cluster I (C) Cluster II (D) and Cluster III (E).

4. DISCUSSION

The fundamental objective of pharmaceutical research is to develop safe and effective medicines for treating diseases and disorders. This endeavor hinges on understanding how drugs interact with complex biological macromolecules, including proteins, polysaccharides, lipids, and nucleic acids. Historically, drug discovery has focused predominantly on proteins as therapeutic targets, using structure-based (SBDD) and ligand-based (LBDD) design strategies, which have led to the majority of clinically approved small-molecule drugs. In fact, approximately 70–90 % of approved therapeutics target proteins, whereas only ~1–2 % directly target nucleic acids.¹⁷ This discrepancy indicates a major scientific gap and also emphasizes the potential of RNA as a therapeutic target. The crucial role of RNA in molecular biology along with its ability to adopt a wide range of well-defined three-dimensional motifs, provide highly specific recognition sites that can support innovative drug design strategies. In this thesis, we approached RNA-based therapeutics from three complementary perspectives.

First, we explored RNA aptamer A40s as a ligand to the EphA2, a protein overexpressed in GSCs. These cells promote therapy resistance and drive tumor relapse in GBM, the most aggressive tumor in adults. Despite extensive efforts to target EphA2 using small molecules, peptide-based approaches, and antibodies, these strategies are hindered by limitations such as lack of specificity, toxicity and delivery inefficiencies.

In our research pipeline we utilized the 2-fluorinated-30-nt A40s RNA aptamer, leveraging its potential advantages, including high specificity, low toxicity and low susceptibility to degradation by nucleases (achieved through

the fluoro modification, with minimal disruption to local and overall helix geometry^{62,63}), the small size, and the ease of synthesis and modification. A40s selectively binds to GSCs by EphA2 recognition and effectively inhibits GSCs' growth, stemness, and migration.^{64,65} These properties addressed the shortcomings of previous strategies and highlighted the potential of A40s as a targeted and effective therapeutic approach for EphA2-positive cancers. Indeed, after the FDA approval of pegaptanib (Macugen), an RNA aptamer for the treatment of neovascular (wet) age-related macular degeneration (AMD),²⁵⁵ and avacincaptad pegol (Izervay), a complement inhibitor for the geographic atrophy (GA) treatment,²⁵⁶ strategies focusing on RNA aptamers hold great promise for targeting a wide range of human diseases.^{256,257}

To develop a structural model of A40s as well as of the aptamer bound to its target EphA2, we integrated advanced computational and experimental methods. First, we predicted the secondary and tertiary structure of A40s, which resulted in a hairpin consisting of an 8-base-paired stable stem and a more flexible 14-nucleotide loop. Notably, MD simulations supported this structure, revealing additional base pairs and stacking interactions, which were absent in the initial software predictions. These findings were corroborated by ILP and RNase T1 cleavage experiments, providing a reliable A40s model for further investigations into its interaction with EphA2.

This underlines the importance of integrating advanced computational and experimental methods in understanding RNA molecules and biological systems. Simple computational tools built upon different underlying algorithms and prediction strategies offer powerful insights, however combined with advanced computational methods reveal crucial details that simple software or experiments might fail to predict. On the other hand, experimental methods validate predictions and unveil structural information

that computational methods might overlook. By capitalizing on both computational and experimental strengths, we advance our ability to study, understand and design complex biomolecular structures.

We then performed molecular docking studies between A40s and EphA2, which proposed a potential binding mode where A40s aligns horizontally atop EphA2, with nucleotides 19–22 in the aptamer loop engaging the LBD surface pocket. This configuration resembles the experimental binding pose of the endogenous ligand ephrinA1, suggesting a shared interaction mechanism. AlphaFold3 identified a similar binding region, while it was unable to accurately predict the aptamer's secondary structure, highlighting the limitations of current AI-based tools and the need for further refinement to improve predictive accuracy. On the other hand, MD simulations on the EphA2/A40s complex suggest that the aptamer could undergo structural post-binding adjustments to stabilize into a conformation consistent with its predicted structure. This finding hints at the dynamic nature of RNA-protein interactions and underscores the potential importance of MD simulations in revealing conformational shifts that static models might miss.

Comparative analysis of simulations of EphA2/A40s and EphA2/ephrinA1 complexes enhances the hypothesis that the aptamer mimics ephrinA1's binding mode, thus restoring EphA2 canonical signaling. A40s and ephrinA1 could share key EphA2 interacting residues (Asp53 and Arg103), however, the broader divergence in their binding profiles suggests opportunities for optimization. Residues uniquely engaged by ephrinA1 (Gly49, Gln56, Asn57, Asp61, and His160) could guide the design of A40s variants with enhanced mimicry of ephrinA1's binding mode. For instance, nucleotides 19–22, which are critical for binding within the EphA2 surface pocket, could be further modified to strengthen interactions with these specific amino acids. Such

refinements might improve the aptamer's capacity to activate EphA2's canonical signaling, potentially unlocking therapeutic applications. Structural modifications informed by our model could also improve the aptamer's 2D structure stabilization, resistance to nuclease degradation, or enhance its pharmacokinetics without compromising its binding efficiency. Additionally, these insights can guide the rational optimization of A40s, including strategies such as site-specific conjugation to enable multivalent binding. In recent years, multimeric aptamers targeting membrane receptors have been successfully developed, including a dimeric DNA aptamer that functions as a VEGFR2 agonist by promoting receptor dimerization,²⁵⁸ and two multivalent aptamers that act as CD28 agonists.²⁴ Likewise, conjugated aptamers could be engineered to promote EphA2 clustering, potentially influencing receptor internalization, activation, or downstream signaling. This concept is supported by previous studies where dimeric ligands on EphA2 led to higher-order EphA2 oligomers, increasing receptor tyrosine phosphorylation and downstream kinase-dependent signaling through two dimerization interfaces, whereas monomeric ligands dimerized via a single one.²⁵⁹ Therefore, our results not only shed light on the EphA2/A40s interface but also offer a foundation for the design of aptamers with improved biological activity.

Altogether, this study offers valuable hints for rational design strategies of modified aptamer sequences before experimental validation. This computational approach enables the development of optimized A40s variants as it offers crucial insights into its interaction with EphA2. By rationally modifying this aptamer, we pave the way for more effective and specific therapeutic aptamers targeting EphA2 and related proteins, while minimizing the need for extensive trial-and-error experiments, saving time and resources.

In the second project, we studied the performance of four FFs namely OL3, OL3-fix, Shaw, and DES. Accurate FFs are essential for studying RNA structure and dynamics, as computational approaches can provide insights that are difficult to access experimentally due to limitations in time, cost, and complexity. However, RNA simulations are still limited by two major factors: accessible timescales and FF accuracy. Evaluating state-of-the-art RNA FFs is therefore critical to assess their ability to reproduce and sample realistic RNA structures. The aim of this project was to compare these FFs across a range of RNA motifs, identify their strengths and limitations, and provide guidance for future FF development.

We selected 13 RNA structures spanning a broad range of motifs, including duplexes, internal loops, hairpins, sarcin-ricins, tetraloops, junctions, G4s, kissing loops, and TTRs, covering both short and long sequences. For each structure, 1 μ s MD simulations were performed using the OL3, OL3-fix, Shaw, and DES FFs. Trajectories were analyzed in terms of visual inspection, global stability (RMSD), H-bond interactions, and dihedral-angle behavior. For systems with available NMR data, NOE-derived distances were also computed and compared with experimental observables.

RMSD analysis indicated that OL3 showed the best agreement with the experimental crystal structures, maintaining native-like conformations in 12 of the 13 systems. Structural rearrangements were observed for the two 21-nt G4 8UTG chains. OL3-fix followed closely, reproducing 10 of 13 structures, with deviations observed in the tetraloop and both G4 systems. Shaw and DES each reproduced 8 of 13 structures: Shaw showed deviations in the tetraloop, G4s, the 78-nt kissing-loop (2D1B), and the TTR, whereas DES failed to maintain native structures for the G4s, kissing loops, and the TTR over the 1

μ s simulations. Notably, DES lost the native-like geometry and exhibited larger fluctuations compared to the other force fields.

Our results indicate that all force fields perform well for short RNA motifs (<40 nt), including duplexes and internal loops, as well as for the larger junction motif. This conclusion is further supported by the comparison between experimental NOE restraints for the internal-loop system and MD-derived observables, which showed agreement exceeding ca. 90%.

H-bond and dihedral angle analysis revealed that tetraloop collapse in OL3-fix and Shaw simulations was initiated by the loss of s-b interactions, together with transitions in the γ and ζ dihedrals. The structural deviations of G4s from their experimental structures in simulations across all FFs was primarily due to dynamic corners of the structures, whereas the main core remained stable. The H-bond analysis indicated loss of all contact types (W-C, non-WC, b-ph, s-ph, and s-b) and generally preserved interactions present in the crystal structures, while dihedral analysis also exhibited fluctuations in several dihedrals, including α , β , δ , and χ . These findings suggest that the observed deviations are primarily driven by altered stacking interactions. Consistently, NOE analysis showed only ca. 68% agreement with experimental data, indicating that the flexible corner conformations sampled in the simulations are unlikely to be fully realistic.

The high RMSD values observed for kissing-loops systems in simulations with OL3, OL3-fix, and Shaw arise by structural rearrangements occurring during the equilibration phase, leading to conformations with slightly narrowed grooves that remain stable for the rest of the simulations. Despite these rearrangements, NOE analysis showed high agreement (>95%), indicating that the resulting conformations remain compatible with

experimental restraints. Among these force fields, Shaw induced a stronger groove bending, increasing the average RMSD by approximately 2 Å relative to OL3 and OL3-fix.

In contrast, the DES FF promoted groove widening, leading to elongation of the global structure. In these simulations, several non-WC and s-ph interactions present in the crystal structure were completely missing or lost during the simulations, while dihedral analysis highlighted that the deviations could arise from backbone flexibility involving α , β , δ , and ζ angles. H-bond analysis further explained the differential stability of the longer 2D1B kissing-loop compared to the shorter 2MI0 system. The 2D1B stability relied mainly on non-WC and s-ph interactions, whereas 2MI0 involved a broader set of H-bond types, making it more difficult to maintain its native interaction network.

H-bond analysis revealed that deviations in the TTR simulations with Shaw and DES FFs were provoked by the loss of tertiary interactions, particularly non-WC and s-s contacts, driven, however, by distinct mechanisms. Shaw exhibited a straightforward loss of tertiary interactions, whereas DES gradually widened the RNA grooves, elongating the overall structure. Dihedrals γ and χ remained stable, while δ and ζ fluctuated without affecting the RMSD, suggesting that changes in the α , β , and ϵ dihedrals might have a more significant impact on destabilizing the TTR fold.

Overall, all OL3, OL3-fix, Shaw, and DES FFs reliably reproduce simple RNA architectures including duplexes, internal-loops, and junctions. In contrast, OL3-fix and Shaw show limitations for short but structurally complex tetraloops, where loss of s-b interaction as well as γ and ζ dihedral deviations lead to structural instability. All FFs also struggle with accurately reproducing G4 architectures, likely due to an imperfect balance of stacking

interactions, resulting in highly dynamic corner regions. For long kissing-loop systems (ca. 80 nt), OL3, OL3-fix, and Shaw consistently induce groove narrowing, with Shaw producing the strongest bending, while DES promotes groove widening and global elongation. A similar behavior is observed for the TTR motif, suggesting that DES introduces a systematic bias toward elongated conformations in longer RNA sequences (>40 nt). This tendency may explain the larger fluctuations, earlier structural collapse, and lower overall NOE agreement observed for DES across multiple motifs.

As mentioned in Chapter 1, the accurate treatment of ions in RNA systems is critical due to the highly negatively charged nature of RNA, yet limitations in ion parametrization and compatibility with different FFs remains a main challenge. Although our simulations included K^+ , Cl^- , and Mg^{2+} ions (the latter for G4s, where their presence is essential), proper RNA folding often requires ions to occupy specific positions. Our 1 μs simulations may be too short for ions to fully equilibrate and reach these positions, limiting their stabilizing effect. For instance, the TTR crystal structure contains Mg^{2+} ions at key interaction sites, which were not fully sampled in our simulations with Shaw and DES. This suggests that the observed structural collapse could, at least in part, result from the absence of ions in their proper positions.

This raises an important consideration regarding OL3. While OL3 demonstrates the best performance in terms of global stability, hydrogen bonds, dihedrals, and agreement with NOE data, its parametrization restricts backbone flexibility. As a result, OL3 is particularly effective at reproducing experimental structures; however, one of the primary purposes of MD simulations is to explore RNA dynamics and the range of conformational ensembles that the molecule can adopt to understand its functional behavior

This study offers crucial insights regarding strengths, limitations, and potential concerns about current RNA FFs. However, given the inherent RNA complexity, further analysis is warranted. Longer μ s-scale simulations with multiple replicates would help determine the statistical robustness of our observations. Additional analyses, such as root-mean-square fluctuation (RMSF), could further quantify residue-level flexibility and provide deeper insights into local conformational dynamics that complement our current results. Including ions at specific binding sites observed in experimental PDB structures, followed by a comparative analysis with simulations lacking such ions, could help differentiate FF-related effects from ion-mediated stabilization and clarify whether some conformational deviations arise from the absence of bound ions. Development of automated scripts for H-bond and dihedral analysis, integrated with RMSD metrics, could facilitate the identification of features that critically impact stability, structural rearrangements, or failures. Additional investigations of ions effects in RNA systems would offer crucial insights into RNA structure and dynamics. Finally, advanced computational approaches, such as metadynamics and energy landscape mapping, could more thoroughly explore conformational ensembles and assess their compatibility with experimental data.

In summary, this study provides key-insights into the performance of state-of-the-art OL3, OL3-fix, Shaw, and DES FFs, highlighting their strengths, limitations and general concerns. These findings are valuable not only for evaluating current FF behavior but also for guiding the development of novel, more accurate RNA FFs. In particular, the data will inform the ongoing efforts of Prof. Modesto Orozco's group, which is developing a novel RNA force field using ML techniques. This new approach integrates experimental NOE data from RNA tetramers, nucleoside-specific information, and features

identified in the systems analyzed here, paving the way toward improved accuracy in RNA simulations.

In the third project, we focused on targeting tau pre-mRNA using small molecules. Tau pre-mRNA produces 3R isoform (with exon 10 exclusion) and 4R isoforms (with exon 10 inclusion) in 1:1 ratio. Destabilization of exon 10-intron hairpin leads to overproduction and aggregation of the 4R isoform, which has been linked to tauopathies. The aim of this project was to identify ligands capable of binding stabilizing the tau pre-mRNA hairpin, with the goal of restoring the physiological 3R:4R isoform ratio by decreasing the 4R tau production.

To achieve this, we started by screening a GH library of approximately 100 compounds using DL-based 2D and 3D similarity metrics, using a known tricyclic binder as the reference ligand. Notably, 15 compounds appeared within the top 20 hits in both 2D and 3D similarity analysis. While the GH library included a structural variety of molecules including monoaryl, bisaryl, polyaryl, bicyclic, and tricyclic scaffolds, all 15 selected compounds shared a tricyclic core differing only in the bridgehead atom, the number of GH moieties, and the presence or absence of an α -methyl group. The inclusion of all 15 molecules in the top 20-ranking solutions using both similarity metrics indicates that 2D and 3D metrics capture related chemical features, and that compounds sharing common core motifs with the reference scaffold tend to be prioritized. However, the fact that these compounds appeared with different ranking highlights that 2D and 3D descriptors emphasize different aspects of molecular similarity. Specifically, 2D metrics depend on topological and substructural relationships, while 3D metrics are more sensitive to the spatial arrangement with the reference ligand. As a result, the divergent rankings

suggest that even though the compounds share key structural elements, their 3D shape can influence their potential binding to the RNA in a way that cannot be captured by 2D similarity alone.

$1D^1H$ NMR analysis of the selected 15 compounds exhibited positive binding for all of them. This could suggest that the tricyclic scaffold, which was present in all NMR tested compounds is essential for interaction with the A-pocket of tau RNA hairpin. Some compounds shared similar patterns of spectral perturbations such as GM36 with GM118 and GM62 with GM123, which indicates similar binding mode. Others, however, displayed alternative perturbations including signal broadening, disappearance, increased intensity, or chemical shifts. While these observations may reflect variations in binding behavior, the fact that all 15 compounds produced positive signatures also raises concerns about the specificity of the method. $1D^1H$ NMR can generate false positives due to aggregation, compound instability, or nonspecific interactions, all of which can alter spectral profiles independently of true RNA binding. Therefore, while the results are encouraging, careful interpretation is required.

The ligand-based WaterLOGSY approach complemented the macromolecule-based $1D^1H$ approach as it focuses on changes in the signals of the ligands rather than the RNA, providing an orthogonal readout of binding. Using this method, six compounds were unambiguously characterized as binders (GM36, GM131, GM113, IP025, GM29, GM41), while the remaining compounds suggested no interaction. These results confirmed the concern raised by the $1D^1H$ experiments, while they helped refine the hit list by identifying which NMR perturbations reflected true binding and which artefacts. Overall, combining $1D^1H$ and WaterLOGSY provided a more reliable framework for identifying genuine RNA binders than either method alone. This integrated

approach enabled us to filter out false positives and prioritize structurally promising ligands.

The six most promising compounds were then biologically evaluated by measuring their ability to modulate 3R:4R tau isoform ratio relative to the DMSO control, using two independent analysis methods. Among all compounds tested, IP025 consistently demonstrated the strongest effect across three replicates. Specifically, IP025 increased the 3R:4R ratio to 1.1 in the Fiji analysis and to 1.7 in the in-house analysis. In contrast, the reference ligand GM118, a previously reported small-molecule binder of this hairpin, exhibited minimal activity, yielding ratio fold-ranges of 1.05 (Fiji) and 1.1 (in-house), both very close to the baseline. The remaining compounds also exhibited values close to the control condition, suggesting limited or negligible activity. These results suggest that our screening protocol successfully identified a ligand with substantially higher biological activity than the GM118 reported in literature.

These findings arise from experiments performed on the WT tau pre-mRNA, which allowed for initial identification of binders of the A-bulge pocket of tau pre-mRNA. In addition to these, further experimental studies will be carried out on the mutated specific variant +14 DDPAC. Specifically, 1D ^1H and WaterLOGSY NMR studies are ongoing on the stem of DDPAC tau in presence of the identified compound IP025, along with biological studies in HeLa cells that feature the mutated form of the gene. These additional experiments will allow us to more accurately assess the activity of IP025 in a disease-relevant context.

Molecular docking of IP025 on tau RNA hairpin revealed three different binding poses. The first and most populated group (5/9 poses) positioned the methyl group towards the interior of the tau RNA binding site. The second

group (3/9 poses) displayed a flip orientation, with the methyl substituent facing the solvent. The final pose (Cluster III) adopted a configuration mirrored along the y axis, with the methyl group still directed toward the binding pocket. Despite these orientation differences, all poses occupied the same planar binding region as the reference ligand GM118, indicating a conserved binding mode driven by the geometry of the tricyclic scaffold. Interestingly, the interaction profiles of all IP025 poses were highly similar. Each pose formed a salt bridge between the positively charged guanidinium nitrogen and a nearby phosphate oxygen, a cation- π interaction with an aromatic nucleotide base, and three π - π stacking interactions mediated by the two outer rings of the tricyclic core. Cluster III reproduced the same set of interactions but engaged a different nucleotide triad, consistent with its flipped orientation. Comparison to the co-crystallized reference ligand GM118 revealed key mechanistic differences. GM118 lacks both the salt-bridge and the cation- π interaction that are consistently observed across all IP025 poses, forming an additional π - π stacking interaction involving its central aromatic ring, instead. In addition, the two symmetric GH groups of GM118 are solvent exposed, where they contribute an energetic penalty rather than stabilizing the complex. These distinctions offer a plausible explanation for the markedly lower 3R:4R fold-change observed for GM118 in the biological assays, and for the strong effect of IP025. Beyond explaining experimental trends, the docking-derived interaction map offers a rational path for optimization. For instance, based on the geometry of Cluster I, the aromatic ring opposite the GH group could be functionalized with a small substituent capable of forming productive interactions with the sugar moiety of C5 or the phosphate group of A6. Such targeted modifications could enhance RNA stabilization, improve binding affinity, and potentially increase both selectivity and specificity,

which are the main challenges in small-molecule RNA-targeted drug discovery.

Overall, our workflow, encompassing DL-based screening, complementary 1D1H and WaterLOGSY NMR experiments, biological assessment, and computational analysis, successfully identified a novel small molecule capable of binding the tau RNA hairpin with a more favorable interaction profile than the reported GM118 and exhibiting higher biological activity. This integrated strategy provides a blueprint for the design and optimization of selective ligands targeting the tau exon 10-intron hairpin, with the potential to stabilize its structure and ultimately modulate the splicing event underlying tauopathies.

5. CONCLUSIONS

RNA plays a central role in molecular biology as it is involved in a huge number of functions and mechanisms. Its ability to fold into diverse and well-defined 3D motifs provides highly specific recognition sites that can support innovative drug design strategies. In this thesis, we examined RNA-based therapeutics from three complementary perspectives that together illustrate the potential of RNA in drug discovery.

In the first project, we investigated RNA as a ligand for the EphA2 protein receptor with the aim to target GSCs. Through extensive computational analysis including 2D/3D motif prediction, molecular docking, and MD simulations, guided also by ILP experiments, RNase T1 cleavage assays, and western blot data, we unveiled the structure of A40s aptamer and investigated its interaction with EphA2. The generated model offers a blueprint to accelerate the development of optimized A40s variants, advancing next-generation EphA2-targeted anticancer therapies.

In the second project, we focused on the current computational infrastructure available for RNA modelling. In particular, we performed MD simulations and we examined how state-of-the-art MD FFs, including OL3, OL3 cP-ghbfix21, Shaw, and DES-AMBER, reproduce key structural features of RNA motifs. This assessment revealed the strengths and limitations in the ability of current force fields to capture RNA dynamics and conformational preferences, providing useful guidelines for future computational studies and for the development of improved RNA force fields.

In the final project, we investigated small molecules as RNA-targeting agents, focusing on the stabilization of disease-relevant tau RNA exon 10-intron

hairpin. For this effort we initially employed a multi-tiered workflow combining AI-based screening, and NMR spectroscopy (1D¹H and WaterLOGSY). Subsequently, biological assays performed by collaborators enabled us to functionally evaluate the screened ligands, and the most promising one was further characterized through molecular docking and interaction analysis. This integrative strategy enabled us to identify structural patterns associated with RNA binding and stabilization, offering insights into the rational design of small-molecule modulators of tau RNA.

Across all three projects, computational approaches were integrated with experimental data, either generated through collaboration (ILP, RNase T1 cleavage, western blot analysis, cellular assays) or obtained from literature (i.e. NOEs). This synergistic strategy enhanced the robustness of our structural models and provided mechanistic insights that would not have been accessible through experiments alone. Overall, this thesis demonstrates the value of combining computational and experimental approaches for advancing RNA-focused drug discovery and offers methodological frameworks that may accelerate the development of RNA-based therapeutic strategies.

ACKNOWLEDGEMENTS

I acknowledge the Piano Nazionale di Ripresa e Resilienza (PNRR) for funding my three-year PhD scholarship, and the CINECA ISCRA B project “*Modeling A40s Aptamer for Advancing Glioblastoma Therapy*” (IsB30_AptaMod) for providing computational resources. I thank Prof. Gerolama Condorelli (Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II) for performing the Western blot experiments; Prof. Mario Mörl (Institute for Biochemistry, Leipzig University) for the probing experiments; and Prof. Seneci (University of Milan) for providing the GH library. I also thank Dr. Vania Broccoli and Dr. Simone Alberti (IRCCS Ospedale San Raffaele, Italy) for the biological experiments on tau pre-mRNA.

I am grateful to my supervisors, Prof. Luciana Marinelli, Prof. Francesco Saverio Di Leva, and Prof. Modesto Orozco, for giving me the opportunity to work on these projects and for their continuous guidance and support throughout my PhD. I also thank my referees, Dr. Zoe Cournia (Biomedical Research Foundation, Academy of Athens) and Prof. Vittorio Limongelli (USI Lugano, Switzerland), for their valuable feedback and constructive suggestions that contributed to improving this thesis. I would also like to thank all the lab members for creating a pleasant and collaborative environment.

Technical Details

Figures and graphical representations were generated using BioRender (<https://app.biorender.com/>), VARNA,²⁰³ ChimeraX,²⁶⁰ VMD,²²⁸ and Maestro (Schrödinger Suite).²⁵⁴

ABBREVIATIONS

RNA	Ribonucleic acid
mRNA	Messenger RNA
tRNA	Transfer RNA
rRNA	Ribosome
snoRNA	Small nucleolar RNA
riboswitches	RNA switches
ncRNA	Noncoding RNA
A	Adenine
U	Uracile
G	Guanine
C	Cytosine
2'OH	Ribose 2'-hydroxyl
2D	Secondary structure
3D	Tertiary structure
WC	Watson-Crick
SBDD	Structure-based drug design
LBDD	Ligand-based drug design
ASOs	Antisense oligonucleotides
siRNAs	Small interfering RNA
FDA	US Food and Drug Administration
RNAi	RNA interference
SELEX enrichment	Systematic evolution of ligands by exponential
PS	Phosphorothioate
PMO	Phosphorodiamidate morpholino oligomer

2'-F	2'-fluoro
2'-OMe	2'-O-Methyl
2'-MOE	2'-O-methoxyethyl
5'-mC	5'-methylcytosine
MD	Molecular dynamics
FF	Force field
GBM	Glioblastoma multiforme
GSCs	Glioblastoma stem cells
EGFR	Epidermal growth factor receptor
VEGF	Vascular endothelial growth factor
PDGFR	Platelet-derived growth factor receptor
EphA2	Erythropoietin-producing hepatocellular carcinoma A2
RTK	Receptor tyrosine kinase
LBD	Ligand binding domain
CRD	Cysteine-rich domain
nFN3	n-Fibronectin type III domain
cFN3	c- Fibronectin type III domain
JM	Juxtamembrane
TanCAR	Tandem CAR-T cell
BBB	Blood brain barrier
2'-F-RNA	2'-fluoro-modified ribonucleotides
Tau	Tubulin associated unit
MAPT	Microtubule-associated protein tau
NTD	N-terminal projection domain
PRR	Proline-rich region
MTBD	Microtubule-binding domain
snRNP	Small nuclear ribonucleoprotein

FTDP-17 chromosome 17	Frontotemporal dementia and parkinsonism linked to
GH	Guanylylhydrazone
AI	Artificial intelligence
ML	Machine-learning
DL	Deep-learning
MSAs	Multiple sequence alignments
ΔG_b	Binding free energies
ΔG_{solv}	Solvation free energy
ΔG_{conf}	Contribution of conformational changes
ΔG_{int}	Free energy, due to ligand-protein interactions
$\Delta G_{t/r}$ energy	Penalty for the loss of translational and rotational free
ΔG_{vib} modes	Free energy of the system due to changes in vibrational
QM	Quantum mechanics
CG	Coarse-grained
BD	Brownian dynamics
QM/MM	Quantum mechanics/Molecular mechanics
MM/CG	Molecular mechanics/Coarse-grained
vdW	Van der Waals
PBCs	Periodic boundary conditions
PDB	Protein Data Bank
NMR	Nuclear magnetic resonance spectroscopy
NOE	Nuclear Overhauser effect
CSPs	Chemical shift perturbations
RMSD	Root mean square deviation
PCA	Principal component analysis

PME	Particle-mesh Ewald
RMSF	Root mean square fluctuations
ILP	In-line probing
OL3-fix	OL3 cP-ghbfix21
DES-AMBER	DES
LJ	Lennard-Jones
G4	G-quadruplex
TTR	Tetraloop-tetraloop receptor
H-bonds	Hydrogen bonds
Nt	Nucleotide
s-ph	Sugar-phosphate
b-ph	Base-phosphate
s-b	Sugar-base
non-WC	Non canonical WC
MW	Molecular weight
LogP	Liposolubility
TPSA	Topological polar surface area
NR	Number of rings
Nrot	Rotatable keys
HBD	Hydrogen bond donors
HBA	Hydrogen bond acceptors
DMSO	Dimethylsulfoxyde

List of Figures

Figure 1.1. Scheme of RNA multiple functions.....	2
Figure 1.2. Schematic representation of RNA. The phosphate, ribose sugar, and base are marked by green, purple, and cyan squares, respectively. A nucleoside is referred to a base linked to a sugar, while a nucleotide is a nucleoside with a phosphate group attached.....	4
Figure 1.3. A) canonical base pairs (A-U, G-C), B, C) Interacting edges of nucleobases.	5
Figure 1.4. RNA 2D key substructures.	7
Figure 1.5. RNA torsion angles α , β , γ , δ , ϵ , ζ , χ	8
Figure 1.6. Advanced 3D RNA architectures, including triple helices, kissing-loops, G-quadruplexes, and i-motif-like structures.....	9
Figure 1.7. A) Structural organization of EphA2. B) Canonical and non-canonical pathways of EphA2. During the canonical pathway eA1 binds to EphA2 enabling signaling through the receptor EphA2 (forward signaling) and the ligand eA1 (reverse signaling). Upon activation by eA1, EphA2 oligomerizes and provokes autophosphorylation of tyrosine residues in the intracellular domain. In contrast, in the non-canonical pathway, EphA2 heterodimerizes with other RTKs, leading to phosphorylation of Ser897....	17
Figure 1.8. Tubulin associated unit (tau) protein binds to microtubule, leading to the proper shape of neurons.	22
Figure 1.9. The alternative splicing of MAPT pre-mRNA, which leads to 4R-tau and 4R tau isoforms.	23
Figure 1.10. Hairpin structure formed by the exon 10–intron junction, which regulates access of the U1 snRNP splicing factor and thereby controls alternative splicing. Mutations at positions +3, +10, +13, +14, and +16 (numbered relative to the first nucleotide of the splice-donor site, +1)	

destabilize the hairpin, promoting U1 snRNP binding and increased production of 4R-tau. Targeting the A-bulge binding pocket of the hairpin with small molecules represents a promising strategy to stabilize the structure and restore physiological splicing. 25

Figure 2.1. Methods for RNA 2D structure prediction.	31
Figure 2.2. Fragment-based approach for RNA 3D prediction.....	32
Figure 2.3. Coarse-grained approach for RNA 3D prediction. Adapted from Boniecki et al. ⁹²	33
Figure 2.4. Four categories of the scoring functions for receptor–ligand docking. Adapted from Jin et al. ¹³⁴	41
Figure 2.5. Schematic representation of multiscale simulation methods in computational chemistry, illustrating the relationship between system size, spatial resolution, and accessible time scales. Adapted from Dans et al. ¹⁴⁴ .	46
Figure 2.6. Symbolic representation of the bonded interactions: bond stretching r (upper left), bond angle bending θ (bottom left), dihedral ϕ (upper right) and improper dihedral ψ (bottom right) and the small out-of-plane angle α . Adapted from Adcock and McCammon. ¹⁴⁶	48
Figure 2. 7. Definition of bond stretching between two atoms.....	48
Figure 2. 8. Definition of angle bending between three atoms.	49
Figure 2. 9. Definition of dihedral torsion between four atoms.....	50
Figure 2.10. The improper dihedral.	51
Figure 2.11. The Lennard-Jones potential, where σ is the collision parameter and ε the well depth.	52
Figure 2.12. Schematic representation of electrostatic interactions between three charged atoms q_i , q_k , q_j	53
Figure 2.13. A periodic system illustrating the periodic boundary conditions used in MD simulations.	55

Figure 2.14. Representation of one-dimensional potential energy surface. Minimization methods move downhill to the nearest minimum.	56
Figure 2.15. NMR spectrum in frequency-domain obtained after the Fourier Transform of the detected signal FID in time-domain.....	76
Figure 2.16. Water-LOGSY principle. The target is shown with the buried cavities and the active binding site. The ligand is shown in the bound and free states. Excitation of bulk water (circles) is shown with a solid arrow and some of the different magnetization transfer pathways are shown with dashed lines.	82
Figure 3.1. Hairpin and hairpin-bulged (bulged) models derived from clustering of 2D RNA structures. The 2D RNA representations shown were generated using VARNA. ²⁰⁰	87
Figure 3.2. Clustering of 2D structure predictions of A40s using RNAstructure 6.5, RNAfold 2.6.3, CentroidFold 0.0.16, NUPACK 4.0, Vfold2D 2.5, Ipknott++ 2.2.1, MC-Fold 8, KineFold 3, pKiss 2.3.0, sFold 2.2, and AlphaFold 3. All available parameter settings provided by each tool were considered for 2D structure prediction. For AlphaFold 3, which allowed for both isolated RNA and RNA-protein complex predictions, we included 2D structures of A40s predicted alone as well as those predicted in complex with EphA2. In cases where multiple structures with varying free energies were obtained, we selected the top three solutions, resulting in a total of 46 predicted 2D structures.	87
Figure 3.3. Ensemble of 3D conformations of hairpin structure (left, violet cartoon) and bulged structure (right, dark green cartoon) obtained using RNAcomposer, 3dRNA, and Vfold3D.	90
Figure 3.4. eRMSD-based analysis of 3D A40s structure predictions, alternatively including (A, B, D) or excluding (C, E) the aptamer's loop	

(nucleotides 9-22). A) Projection onto PC1 and PC2 of G-vectors computed from the PDBs provided by the 3D-conformation predictor tools for bulged (left) and hairpin (right) conformations. B) and C) Cross-eRMSD matrix between the PDBs generated by the 3D-conformation predictor tools for bulged (left) and hairpin (right) conformations. D) and E) Cross-RMSD matrix between the PDBs generated by the 3D-conformation predictor tools for bulged (left) and hairpin (right) conformations, based on the GROMOS algorithm..... 91

Figure 3.5. RMSD (top panels), eRMSD (central panels), and RMSF (bottom panels) plots of the hairpin (A, C, E) and bulged (B, D, F) models. The values are computed on the A40s's heavy atoms with respect to the initial frame. For the RMSD calculations, the terminal nucleotides were excluded..... 96

Figure 3.6. Time evolution of the eRMSD¹² and corresponding density histograms computed over three independent RNA-only MD simulations of the A40s aptamer, initiated from either the hairpin (upper panels) or bulged (lower panels) conformations. The initial offset in eRMSD values reflects the use of the respective cluster centroids (from hairpin or bulged ensembles) as reference structures for eRMSD calculation. These centroids undergo slight rearrangements during the initial equilibration phase, which is not included in the plotted trajectories..... 97

Figure 3.7. Cross-eRMSD evolution (upper row) and density histograms (lower row) computed over the A40s aptamer in three independent replicas of RNA-only simulations initialized from hairpin (upper panels) and bulged (lower panels) conformations. The reference conformations for eRMSD calculations are the centroids of the hairpin structure (when the simulation was initialized from the bulged conformation) and the bulged structure (when initialized from the hairpin conformation), both obtained through clustering. 98

Figure 3.8. A) Changes of the A40s's secondary structures during MD simulations for both the hairpin and the bulged conformation. B) Structure probing results on the full length of the aptamer (A40L), where signals indicate unpaired nucleotides. The unpaired nucleotides of the shorter aptamer A40s are highlighted with green lines. C) Base stacking interactions in the hairpin model..... 102

Figure 3.9. Semi-quantitative analysis of RNase T1 cleavage at G19. (Left) Autoradiograph showing cleavage at G residues, with G19 displaying markedly weaker cleavage compared to G15–G16. (Right) Densitometric profile of the gel, where peaks correspond to cleavage events. The weak signal at G19 supports transient unpairing at this position..... 103

Figure 3.10. Western blot analysis of phosphorylation of EphA2 (left) and associated kinases RSK2 and AKT (right). 104

Figure 3.11. A) Prediction of EphA2/A40s complex using AlphaFold 3. B) 2D structure of A40s on EphA2 predicted by AlphaFold 3. The structure was not supported by the experimental data and was therefore excluded from further studies..... 106

Figure 3.12. A) Superposition of the EphA2/A40s complex predicted by HADDOCK 2.4 and the EphA2/eA1 X-ray complex (PDB code: 3CZU) showing that A40s and the endogenous ligand occupy the same binding surface. Graphical representation of the most important interactions between EphA2 and A40s (B), and between EphA2 and eA1 (C) are shown as insets. The main interacting residues are illustrated as sticks, with specific interactions highlighted by dashed lines. EphA2 is colored in beige, A40s in violet, and eA1 in cyan. 106

Figure 3.13. RMSD of A40s (calculated on heavy atoms, excluding terminal nucleotides) relative to EphA2 on the EphA2/aptamer complex. The trajectory was aligned on the EphA2 secondary structure C α atoms..... 109

Figure 3.14. RMSD of A40s (calculated on heavy atoms, excluding terminal nucleotides) over the MD simulations on the EphA2/aptamer complex. The trajectory was aligned on the A40s heavy atoms, excluding the terminal nucleotides.	109
Figure 3.15. DRPScore ¹³ computed for the EphA2/A40s complex over MD timescale. Representative frames for DRPScore calculation were extracted each 500 ns.....	110
Figure 3.16. A) Histograms showing the frequency of occurrence of the most relevant EphA2/A40s (A) and contacts over the MD trajectory. B) Histograms showing the frequency of occurrence of the most relevant EphA2/eA1 contacts over the MD trajectory. For the identification of the non-covalent biomolecular interactions PyContact ¹⁴ was employed. The frequency of occurrence was obtained using the Hydrogen Bonds tool of VMD ¹⁵ where the donor-acceptor (heavy atoms) distance cutoff was set at 3.5 Å and the angle cutoff at 40°.	113
Figure 3.17. RMSD (calculated on the secondary Cα atoms) of eA1 over the MD simulations on the EphA2/eA1 complex. The trajectory was aligned on the secondary structure Cα atoms of EphA2.	114
Figure 3.18. RNA motifs evaluated using OL3, OL3-fix, Shaw, and DES FFs.	119
Figure 3.19. AMBER atom nomenclature used for the description of RNA molecular interactions.	126
Figure 3.20. H-bond analysis of the duplex motif (2V7R) in simulations using OL3, OL3-fix, Shaw, and DES force fields. The canonical base pair G7-C8 occasionally lost contact during the simulations.	127
Figure 3.21. H-bond analysis of the internal-loop motif (2H49) in simulations using OL3, OL3-fix, Shaw, and DES force fields. The canonical base pair A15–U4 occasionally lost contact during the simulations.....	128

Figure 3.22. H-bond analysis of the sarcin-ricin motif (1Q9A) in simulations using OL3, OL3-fix, Shaw, and DES force fields. The non-canonical base pair U1–G27 occasionally lost contact during the simulations.....	129
Figure 3.23. H-bond analysis of the tetraloop motif (2KOC) in simulations using OL3, OL3-fix, Shaw, and DES force fields.	130
Figure 3.24. Comparison between the crystal structures of G4 7PS8 and G4 8UTG (left) and a representative conformation from the molecular dynamics trajectory (right). The trajectory snapshot highlights the pronounced flexibility of the peripheral corner regions, while the central G4 core remains structurally stable.	133
Figure 3.25. RMSD plots of kissing-loop motifs: 2MI0 (left) and 2D1B (right).	135
Figure 3.26. Comparison of kissing-loop (PDB ID: 2D1B) crystal structure and average structure of OL3, OL3-fix, Shaw, and DES FFs. OL3 and OL3-fix maintain a conformation similar to the crystal structure, Shaw adopts a alternative conformation with respect to the crystal structure after equilibration, while DES broadened its grooves leading to an elongated conformation.	136
Figure 3.27. Loss of long-range tertiary interactions in the TTR using DES FF. DES gradually widened the RNA grooves beginning at approximately 100 ns, elongating the overall structure.	137
Figure 3.28. Dihedral angle time series for the duplex motif (2V7R).	138
Figure 3.29. Dihedral angle time series for the internal-loop motif (2H49).	139
Figure 3.30. Dihedral angle time series for the duplex motif (6Y0T).	140
Figure 3.31. Dihedral angle time series for the tetraloop motif (2KOC)...	142
Figure 3.32. Agreement between MD-derived NOE distances and experimental NOEs.	145

Figure 3.33. Agreement between MD-derived NOE distances and experimental NOEs.....	146
Figure 3.34. Agreement between MD-derived NOE distances and experimental NOEs.....	147
Figure 3.35. Agreement between MD-derived NOE distances and experimental NOEs.....	148
Figure 3.36. Top 15 ligands selected for NMR screening based on 2D and 3D similarity to GM118.....	152
Figure 3.37. 1D ¹ H spectra acquired of tau RNA stem mimic duplex alone (bottom) and in complex with the positive control GM118. Our spectra are consistent with those reported in the literature. ⁷¹	155
Figure 3.38. 1D ¹ H NMR screening of the 15-ligand GH library against tau pre-mRNA.....	157
Figure 3.39. WaterLOGSY NMR spectra of the positive control GM118 alone (top) and tau RNA/GM118 complex (bottom). Ligand resonances appear between 7 and 9 ppm, water signals between 4 and 6 ppm, and buffer signals between 1 and 3.5 ppm.	159
Figure 3.40. WaterLOGSY NMR spectra of the positive binders IP025 (blue), GM36 (olive), and GM113 (green), recorded both alone and in the presence of tau RNA. Ligand resonances appear between 6.3 and 8.5 ppm, water signals between 4 and 6 ppm, and buffer signals between 1 and 3.5 ppm.	160
Figure 3.41. WaterLOGSY NMR spectra of GM24 recorded alone and in the presence of tau RNA. Ligand resonances remain in the same phase, indicating lack of interaction.	161
Figure 3.42. 3R/4R % tau isoform ration fold-change quantification over DMSO control in HeLa transfected with WT plasmid using Fiji-based image analysis and in-house analysis.	163

Figure 3.43. Docking results. A) Superposition of all nine IP025 docking poses together with the co-crystallized reference ligand GM118. Cluster I is shown in lilac, Cluster II in green, Cluster III in cyan, and GM118 in peach. B) Interactions between the tau pre-mRNA and GM118. C-E) Interactions between the tau pre-mRNA and the representative pose of IP025 from Cluster I (C) Cluster II (D) and Cluster III (E)..... 169

List of Tables

Table 1.1. The 12 basic families of non-canonical interactions according to Leontis-Westhof classification. ¹²	6
Table 1.2. Overview of RNA-based therapeutics.	12
Table 2. 1. Comparison Table: 1D ¹ H NMR vs Water-LOGSY	83
Table 3.1. Comparison of parameter treatments across the tested RNA FFs.	116
Table 3.2. List of the selected RNA structures for FF evaluation.....	118
Table 3.3. RMSD (Å) of RNA heavy atoms excluding terminal nt using the OL3, OL3-fix, Shaw, DES FFs. The mean values of each FF are also provided along with their standard error.	124
Table 3.4. Interactions of G4 structures that were lost during the 1 μs MD simulations. Each category is colored following the color scheme of the H-bond plots.....	131
Table 3.5. Average RMSD (Å) of the kissing-loop motifs (2MI0, 2D1B) with reference to the crystal and the equilibrated structure.	135
Table 3.7. 2D and 3D similarity scores with reference to IP025 compound using MolProphet.....	165

REFERENCES

- (1) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, 227 (5258), 561–563. <https://doi.org/10.1038/227561a0>.
- (2) Sharp, S. J.; Schaack, J.; Cooley, L.; Burke, D. J.; Soil, D. Structure and Transcription of Eukaryotic tRNA Gene. *Critical Reviews in Biochemistry* **1985**, 19 (2), 107–144. <https://doi.org/10.3109/10409238509082541>.
- (3) Pereira, M. J. B.; Harris, D. A.; Rueda, D.; Walter, N. G. Reaction Pathway of the Trans-Acting Hepatitis Delta Virus Ribozyme: A Conformational Change Accompanies Catalysis. *Biochemistry* **2002**, 41 (3), 730–740. <https://doi.org/10.1021/bi011963t>.
- (4) Cech, T. R. The Ribosome Is a Ribozyme. *Science* **2000**, 289 (5481), 878–879. <https://doi.org/10.1126/science.289.5481.878>.
- (5) Sharp, P. A. The Centrality of RNA. *Cell* **2009**, 136 (4), 577–580. <https://doi.org/10.1016/j.cell.2009.02.007>.
- (6) Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M. F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B. W.; Cassady, J. P.; Cabili, M. N.; Jaenisch, R.; Mikkelsen, T. S.; Jacks, T.; Hacohen, N.; Bernstein, B. E.; Kellis, M.; Regev, A.; Rinn, J. L.; Lander, E. S. Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals. *Nature* **2009**, 458 (7235), 223–227. <https://doi.org/10.1038/nature07672>.
- (7) Belinky, F.; Bahir, I.; Stelzer, G.; Zimmerman, S.; Rosen, N.; Nativ, N.; Dalah, I.; Iny Stein, T.; Rappaport, N.; Mituyama, T.; Safran, M.; Lancet, D. Non-Redundant Compendium of Human ncRNA Genes in GeneCards. *Bioinformatics* **2013**, 29 (2), 255–261. <https://doi.org/10.1093/bioinformatics/bts676>.
- (8) Piovesan, A.; Antonaros, F.; Vitale, L.; Strippoli, P.; Pelleri, M. C.; Caracausi, M. Human Protein-Coding Genes and Gene Feature Statistics in 2019. *BMC Res Notes* **2019**, 12 (1), 315. <https://doi.org/10.1186/s13104-019-4343-8>.
- (9) Willyard, C. New Human Gene Tally Reignites Debate. *Nature* **2018**, 558 (7710), 354–355. <https://doi.org/10.1038/d41586-018-05462-w>.

- (10) Elkon, R.; Agami, R. Characterization of Noncoding Regulatory DNA in the Human Genome. *Nat Biotechnol* **2017**, *35* (8), 732–746. <https://doi.org/10.1038/nbt.3863>.
- (11) Bagger, F. O.; Borgwardt, L.; Jespersen, A. S.; Hansen, A. R.; Bertelsen, B.; Kodama, M.; Nielsen, F. C. Whole Genome Sequencing in Clinical Practice. *BMC Med Genomics* **2024**, *17* (1), 39. <https://doi.org/10.1186/s12920-024-01795-w>.
- (12) Leontis, N. B. The Non-Watson-Crick Base Pairs and Their Associated Isostericity Matrices. *Nucleic Acids Research* **2002**, *30* (16), 3497–3531. <https://doi.org/10.1093/nar/gkf481>.
- (13) Leontis, N. B.; Westhof, E. Geometric Nomenclature and Classification of RNA Base Pairs. *RNA* **2001**, *7* (4), 499–512. <https://doi.org/10.1017/S1355838201002515>.
- (14) Leontis, N. B.; Westhof, E. Analysis of RNA Motifs. *Current Opinion in Structural Biology* **2003**, *13* (3), 300–308. [https://doi.org/10.1016/S0959-440X\(03\)00076-9](https://doi.org/10.1016/S0959-440X(03)00076-9).
- (15) Sato, K.; Kato, Y.; Hamada, M.; Akutsu, T.; Asai, K. IPknot: Fast and Accurate Prediction of RNA Secondary Structures with Pseudoknots Using Integer Programming. *Bioinformatics* **2011**, *27* (13), i85–i93. <https://doi.org/10.1093/bioinformatics/btr215>.
- (16) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; HersHKovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M.; RNA Ontology Consortium. RNA Backbone: Consensus All-Angle Conformers and Modular String Nomenclature (an RNA Ontology Consortium Contribution). *RNA* **2008**, *14* (3), 465–481. <https://doi.org/10.1261/rna.657708>.
- (17) Balakumar, P.; Mulukuri, N. V. L. S.; Jagadeesh, G. A Map of Molecular Drug Targets and Therapeutics for the US FDA-Approved Drugs: The Impact of Expedited Regulatory Pathways and First-in-Class Drug Approvals on Drug Innovation. *Pharmacology & Therapeutics* **2026**, *277*, 108945. <https://doi.org/10.1016/j.pharmthera.2025.108945>.
- (18) Mundlia, P.; Singh, S. P.; Kaushal, A.; Chatterjee, M.; Ranawat, P.; Pandey, A.; Singh, G.; Barnwal, R. P. RNA-Targeting Therapeutics: Drugging the Undruggable. *Future Medicinal Chemistry* **2025**, *17* (11), 1213–1215. <https://doi.org/10.1080/17568919.2025.2520152>.

- (19) Rinaldi, C.; Wood, M. J. A. Antisense Oligonucleotides: The next Frontier for Treatment of Neurological Disorders. *Nat Rev Neurol* **2018**, *14* (1), 9–21. <https://doi.org/10.1038/nrneurol.2017.148>.
- (20) Sang, A.; Zhuo, S.; Bochanis, A.; Manautou, J. E.; Bahal, R.; Zhong, X.-B.; Rasmussen, T. P. Mechanisms of Action of the US Food and Drug Administration-Approved Antisense Oligonucleotide Drugs. *BioDrugs* **2024**, *38* (4), 511–526. <https://doi.org/10.1007/s40259-024-00665-2>.
- (21) Hu, B.; Zhong, L.; Weng, Y.; Peng, L.; Huang, Y.; Zhao, Y.; Liang, X.-J. Therapeutic siRNA: State of the Art. *Sig Transduct Target Ther* **2020**, *5* (1), 101. <https://doi.org/10.1038/s41392-020-0207-x>.
- (22) Padda, I. S.; Mahtani, A. U.; Patel, P.; Parmar, M. Small Interfering RNA (siRNA) Therapy. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
- (23) Germer, K.; Leonard, M.; Zhang, X. RNA Aptamers and Their Therapeutic and Diagnostic Applications. *Int J Biochem Mol Biol* **2013**, *4* (1), 27–40.
- (24) Pastor, F.; Soldevilla, M. M.; Villanueva, H.; Kolonias, D.; Inoges, S.; De Cerio, A. L.; Kandzia, R.; Klimyuk, V.; Gleba, Y.; Gilboa, E.; Bendandi, M. CD28 Aptamers as Powerful Immune Response Modulators. *Molecular Therapy - Nucleic Acids* **2013**, *2*, e98. <https://doi.org/10.1038/mtna.2013.26>.
- (25) Catuogno, S.; Esposito, C. L.; De Franciscis, V. Developing Aptamers by Cell-Based SELEX. In *Nucleic Acid Aptamers*; Mayer, G., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2016; Vol. 1380, pp 33–46. https://doi.org/10.1007/978-1-4939-3197-2_3.
- (26) Fitzwater, T.; Polisky, B. [17] A SELEX Primer. In *Methods in Enzymology*; Elsevier, 1996; Vol. 267, pp 275–301. [https://doi.org/10.1016/S0076-6879\(96\)67019-0](https://doi.org/10.1016/S0076-6879(96)67019-0).
- (27) Mullard, A. FDA Approves Second RNA Aptamer. *Nat Rev Drug Discov* **2023**, *22* (10), 774–774. <https://doi.org/10.1038/d41573-023-00148-z>.
- (28) Bost, J. P.; Barriga, H.; Holme, M. N.; Gallud, A.; Maugeri, M.; Gupta, D.; Lehto, T.; Valadi, H.; Esbjörner, E. K.; Stevens, M. M.; El-Andaloussi, S. Delivery of Oligonucleotide Therapeutics: Chemical Modifications, Lipid Nanoparticles, and Extracellular Vesicles. *ACS Nano* **2021**, *15* (9), 13993–14021. <https://doi.org/10.1021/acsnano.1c05099>.

- (29) Anand, P.; Zhang, Y.; Patil, S.; Kaur, K. Metabolic Stability and Targeted Delivery of Oligonucleotides: Advancing RNA Therapeutics Beyond The Liver. *J. Med. Chem.* **2025**, acs.jmedchem.4c02528. <https://doi.org/10.1021/acs.jmedchem.4c02528>.
- (30) Paunovska, K.; Loughrey, D.; Dahlman, J. E. Drug Delivery Systems for RNA Therapeutics. *Nat Rev Genet* **2022**, 23 (5), 265–280. <https://doi.org/10.1038/s41576-021-00439-4>.
- (31) Connelly, C. M.; Moon, M. H.; Schneekloth, J. S. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **2016**, 23 (9), 1077–1090. <https://doi.org/10.1016/j.chembiol.2016.05.021>.
- (32) Disney, M. D.; Winkelsas, A. M.; Velagapudi, S. P.; Southern, M.; Fallahi, M.; Childs-Disney, J. L. Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem. Biol.* **2016**, 11 (6), 1720–1728. <https://doi.org/10.1021/acschembio.6b00001>.
- (33) Panei, F. P.; Torchet, R.; Ménager, H.; Gkeka, P.; Bonomi, M. HARIBOSS: A Curated Database of RNA-Small Molecules Structures to Aid Rational Drug Design. *Bioinformatics* **2022**, 38 (17), 4185–4193. <https://doi.org/10.1093/bioinformatics/btac483>.
- (34) Taylor, O. G.; Brzozowski, J. S.; Skelding, K. A. Glioblastoma Multiforme: An Overview of Emerging Therapeutic Targets. *Front. Oncol.* **2019**, 9, 963. <https://doi.org/10.3389/fonc.2019.00963>.
- (35) Bao, S.; Wu, Q.; McLendon, R. E.; Hao, Y.; Shi, Q.; Hjelmeland, A. B.; Dewhirst, M. W.; Bigner, D. D.; Rich, J. N. Glioma Stem Cells Promote Radioresistance by Preferential Activation of the DNA Damage Response. *Nature* **2006**, 444 (7120), 756–760. <https://doi.org/10.1038/nature05236>.
- (36) Bovenberg, M. S. S.; Degeling, M. H.; Tannous, B. A. Advances in Stem Cell Therapy against Gliomas. *Trends in Molecular Medicine* **2013**, 19 (5), 281–291. <https://doi.org/10.1016/j.molmed.2013.03.001>.
- (37) Liu, G.; Yuan, X.; Zeng, Z.; Tunici, P.; Ng, H.; Abdulkadir, I. R.; Lu, L.; Irvin, D.; Black, K. L.; Yu, J. S. Analysis of Gene Expression and Chemoresistance of CD133+ Cancer Stem Cells in Glioblastoma. *Mol Cancer* **2006**, 5 (1), 67. <https://doi.org/10.1186/1476-4598-5-67>.
- (38) Brennan, C. W.; Verhaak, R. G. W.; McKenna, A.; Campos, B.; Nounshmehr, H.; Salama, S. R.; Zheng, S.; Chakravarty, D.; Sanborn, J. Z.; Berman, S. H.; Beroukhi, R.; Bernard, B.; Wu, C.-J.; Genovese, G.; Shmulevich, I.; Barnholtz-Sloan, J.; Zou, L.; Vegesna, R.; Shukla, S. A.; Ciriello, G.; Yung, W. K.; Zhang, W.; Sougnez, C.; Mikkelsen, T.; Aldape,

K.; Bigner, D. D.; Van Meir, E. G.; Prados, M.; Sloan, A.; Black, K. L.;
 Eschbacher, J.; Finocchiaro, G.; Friedman, W.; Andrews, D. W.; Guha, A.;
 Iacocca, M.; O'Neill, B. P.; Foltz, G.; Myers, J.; Weisenberger, D. J.; Penny,
 R.; Kucherlapati, R.; Perou, C. M.; Hayes, D. N.; Gibbs, R.; Marra, M.;
 Mills, G. B.; Lander, E.; Spellman, P.; Wilson, R.; Sander, C.; Weinstein, J.;
 Meyerson, M.; Gabriel, S.; Laird, P. W.; Haussler, D.; Getz, G.; Chin, L.;
 Benz, C.; Barnholtz-Sloan, J.; Barrett, W.; Ostrom, Q.; Wolinsky, Y.; Black,
 K. L.; Bose, B.; Boulous, P. T.; Boulous, M.; Brown, J.; Czerinski, C.; Eppley,
 M.; Iacocca, M.; Kempista, T.; Kitko, T.; Koyfman, Y.; Rabeno, B.; Rastogi,
 P.; Sugarman, M.; Swanson, P.; Yalamanchii, K.; Otey, I. P.; Liu, Y. S.;
 Xiao, Y.; Auman, J. T.; Chen, P.-C.; Hadjipanayis, A.; Lee, E.; Lee, S.; Park,
 P. J.; Seidman, J.; Yang, L.; Kucherlapati, R.; Kalkanis, S.; Mikkelsen, T.;
 Poisson, L. M.; Raghunathan, A.; Scarpacci, L.; Bernard, B.; Bressler, R.;
 Eakin, A.; Iype, L.; Kreisberg, R. B.; Leinonen, K.; Reynolds, S.; Rovira, H.;
 Thorsson, V.; Shmulevich, I.; Annala, M. J.; Penny, R.; Paulauskis, J.;
 Curley, E.; Hatfield, M.; Mallery, D.; Morris, S.; Shelton, T.; Shelton, C.;
 Sherman, M.; Yena, P.; Cuppini, L.; DiMeco, F.; Eoli, M.; Finocchiaro, G.;
 Maderna, E.; Pollo, B.; Saini, M.; Balu, S.; Hoadley, K. A.; Li, L.; Miller, C.
 R.; Shi, Y.; Topal, M. D.; Wu, J.; Dunn, G.; Giannini, C.; O'Neill, B. P.;
 Aksoy, B. A.; Antipin, Y.; Borsu, L.; Berman, S. H.; Brennan, C. W.;
 Cerami, E.; Chakravarty, D.; Ciriello, G.; Gao, J.; Gross, B.; Jacobsen, A.;
 Ladanyi, M.; Lash, A.; Liang, Y.; Reva, B.; Sander, C.; Schultz, N.; Shen,
 R.; Socci, N. D.; Viale, A.; Ferguson, M. L.; Chen, Q.-R.; Demchok, J. A.;
 Dillon, L. A. L.; Shaw, K. R. M.; Sheth, M.; Tarnuzzer, R.; Wang, Z.; Yang,
 L.; Davidsen, T.; Guyer, M. S.; Ozenberger, B. A.; Sofia, H. J.; Bergsten, J.;
 Eckman, J.; Harr, J.; Myers, J.; Smith, C.; Tucker, K.; Winemiller, C.; Zach,
 L. A.; Ljubimova, J. Y.; Eley, G.; Ayala, B.; Jensen, M. A.; Kahn, A.; Pihl,
 T. D.; Pot, D. A.; Wan, Y.; Eschbacher, J.; Foltz, G.; Hansen, N.; Hothi, P.;
 Lin, B.; Shah, N.; Yoon, J.; Lau, C.; Berens, M.; Ardlie, K.; Beroukhim, R.;
 Carter, S. L.; Cherniack, A. D.; Noble, M.; Cho, J.; Cibulskis, K.; DiCara,
 D.; Frazer, S.; Gabriel, S. B.; Gehlenborg, N.; Gentry, J.; Heiman, D.; Kim,
 J.; Jing, R.; Lander, E. S.; Lawrence, M.; Lin, P.; Mallard, W.; Meyerson,
 M.; Onofrio, R. C.; Saksena, G.; Schumacher, S.; Sougnez, C.; Stojanov, P.;
 Tabak, B.; Voet, D.; Zhang, H.; Zou, L.; Getz, G.; Dees, N. N.; Ding, L.;
 Fulton, L. L.; Fulton, R. S.; Kanchi, K.-L.; Mardis, E. R.; Wilson, R. K.;
 Baylin, S. B.; Andrews, D. W.; Harshyne, L.; Cohen, M. L.; Devine, K.;
 Sloan, A. E.; VandenBerg, S. R.; Berger, M. S.; Prados, M.; Carlin, D.;
 Craft, B.; Ellrott, K.; Goldman, M.; Goldstein, T.; Grifford, M.; Haussler, D.;
 Ma, S.; Ng, S.; Salama, S. R.; Sanborn, J. Z.; Stuart, J.; Swatloski, T.;
 Waltman, P.; Zhu, J.; Foss, R.; Frentzen, B.; Friedman, W.; McTiernan, R.;
 Yachnis, A.; Hayes, D. N.; Perou, C. M.; Zheng, S.; Vegesna, R.; Mao, Y.;
 Akbani, R.; Aldape, K.; Bogler, O.; Fuller, G. N.; Liu, W.; Liu, Y.; Lu, Y.;

Mills, G.; Protopopov, A.; Ren, X.; Sun, Y.; Wu, C.-J.; Yung, W. K. A.; Zhang, W.; Zhang, J.; Chen, K.; Weinstein, J. N.; Chin, L.; Verhaak, R. G. W.; Noushmehr, H.; Weisenberger, D. J.; Bootwalla, M. S.; Lai, P. H.; Triche, T. J.; Van Den Berg, D. J.; Laird, P. W.; Gutmann, D. H.; Lehman, N. L.; VanMeir, E. G.; Brat, D.; Olson, J. J.; Mastrogiannakis, G. M.; Devi, N. S.; Zhang, Z.; Bigner, D.; Lipp, E.; McLendon, R. The Somatic Genomic Landscape of Glioblastoma. *Cell* **2013**, *155* (2), 462–477. <https://doi.org/10.1016/j.cell.2013.09.034>.

(39) Furnari, F. B.; Cloughesy, T. F.; Cavenee, W. K.; Mischel, P. S. Heterogeneity of Epidermal Growth Factor Receptor Signalling Networks in Glioblastoma. *Nat Rev Cancer* **2015**, *15* (5), 302–310. <https://doi.org/10.1038/nrc3918>.

(40) Plate, K. H.; Breier, G.; Millauer, B.; Ullrich, A.; Risau, W. Up-Regulation of Vascular Endothelial Growth Factor and Its Cognate Receptors in a Rat Glioma Model of Tumor Angiogenesis. *Cancer Res* **1993**, *53* (23), 5822–5827.

(41) Ma, C.; Li, Y.; Zhang, X.; Zhao, G.; Xu, H. Levels of Vascular Endothelial Growth Factor and Matrix Metalloproteinase-9 Proteins in Patients with Glioma. *J Int Med Res* **2014**, *42* (1), 198–204. <https://doi.org/10.1177/0300060513481924>.

(42) Wykosky, J.; Gibo, D. M.; Stanton, C.; Debinski, W. EphA2 as a Novel Molecular Marker and Target in Glioblastoma Multiforme. *Molecular Cancer Research* **2005**, *3* (10), 541–551. <https://doi.org/10.1158/1541-7786.MCR-05-0056>.

(43) Park, J.; Son, A.; Zhou, R. Roles of EphA2 in Development and Disease. *Genes* **2013**, *4* (3), 334–357. <https://doi.org/10.3390/genes4030334>.

(44) Wilson, K.; Shiuan, E.; Brantley-Sieders, D. M. Oncogenic Functions and Therapeutic Targeting of EphA2 in Cancer. *Oncogene* **2021**, *40* (14), 2483–2495. <https://doi.org/10.1038/s41388-021-01714-8>.

(45) Qiu, C.; Sun, N.; Zeng, S.; Chen, L.; Gong, F.; Tian, J.; Xiong, Y.; Peng, L.; He, H.; Ming, Y. Unveiling the Therapeutic Promise of EphA2 in Glioblastoma: A Comprehensive Review. *Discov Onc* **2024**, *15* (1), 501. <https://doi.org/10.1007/s12672-024-01380-8>.

(46) Xiao, T.; Xiao, Y.; Wang, W.; Tang, Y. Y.; Xiao, Z.; Su, M. Targeting EphA2 in Cancer. *J Hematol Oncol* **2020**, *13* (1), 114. <https://doi.org/10.1186/s13045-020-00944-9>.

- (47) Gai, Q.-J.; Fu, Z.; He, J.; Mao, M.; Yao, X.-X.; Qin, Y.; Lan, X.; Zhang, L.; Miao, J.-Y.; Wang, Y.-X.; Zhu, J.; Yang, F.-C.; Lu, H.-M.; Yan, Z.-X.; Chen, F.-L.; Shi, Y.; Ping, Y.-F.; Cui, Y.-H.; Zhang, X.; Liu, X.; Yao, X.-H.; Lv, S.-Q.; Bian, X.-W.; Wang, Y. EphA2 Mediates PDGFA Activity and Functions Together with PDGFRA as Prognostic Marker and Therapeutic Target in Glioblastoma. *Sig Transduct Target Ther* **2022**, *7* (1), 33. <https://doi.org/10.1038/s41392-021-00855-2>.
- (48) Sharma, P.; Debinski, W. Receptor-Targeted Glial Brain Tumor Therapies. *IJMS* **2018**, *19* (11), 3326. <https://doi.org/10.3390/ijms19113326>.
- (49) Zhao, P.; Jiang, D.; Huang, Y.; Chen, C. EphA2: A Promising Therapeutic Target in Breast Cancer. *Journal of Genetics and Genomics* **2021**, *48* (4), 261–267. <https://doi.org/10.1016/j.jgg.2021.02.011>.
- (50) Chen, J.; Song, W.; Amato, K. Eph Receptor Tyrosine Kinases in Cancer Stem Cells. *Cytokine & Growth Factor Reviews* **2015**, *26* (1), 1–6. <https://doi.org/10.1016/j.cytogfr.2014.05.001>.
- (51) Binda, E.; Visioli, A.; Giani, F.; Lamorte, G.; Copetti, M.; Pitter, K. L.; Huse, J. T.; Cajola, L.; Zanetti, N.; DiMeco, F.; De Filippis, L.; Mangiola, A.; Maira, G.; Anile, C.; De Bonis, P.; Reynolds, B. A.; Pasquale, E. B.; Vescovi, A. L. The EphA2 Receptor Drives Self-Renewal and Tumorigenicity in Stem-like Tumor-Propagating Cells from Human Glioblastomas. *Cancer Cell* **2012**, *22* (6), 765–780. <https://doi.org/10.1016/j.ccr.2012.11.005>.
- (52) Biao-xue, R.; Xi-guang, C.; Shuan-ying, Y.; Wei, L.; Zong-juan, M. EphA2-Dependent Molecular Targeting Therapy for Malignant Tumors. *CCDT* **2011**, *11* (9), 1082–1097. <https://doi.org/10.2174/156800911798073050>.
- (53) Miao, H.; Gale, N. W.; Guo, H.; Qian, J.; Petty, A.; Kaspar, J.; Murphy, A. J.; Valenzuela, D. M.; Yancopoulos, G.; Hambardzumyan, D.; Lathia, J. D.; Rich, J. N.; Lee, J.; Wang, B. EphA2 Promotes Infiltrative Invasion of Glioma Stem Cells in Vivo through Cross-Talk with Akt and Regulates Stem Cell Properties. *Oncogene* **2015**, *34* (5), 558–567. <https://doi.org/10.1038/onc.2013.590>.
- (54) Zhou, Z.; Yuan, X.; Li, Z.; Tu, H.; Li, D.; Qing, J.; Wang, H.; Zhang, L. RNA Interference Targeting EphA2 Inhibits Proliferation, Induces Apoptosis, and Cooperates with Cytotoxic Drugs in Human Glioma Cells. *Surgical Neurology* **2008**, *70* (6), 562–568. <https://doi.org/10.1016/j.surneu.2008.04.031>.

- (55) Chang, Q.; Jorgensen, C.; Pawson, T.; Hedley, D. W. Effects of Dasatinib on EphA2 Receptor Tyrosine Kinase Activity and Downstream Signalling in Pancreatic Cancer. *Br J Cancer* **2008**, *99* (7), 1074–1082. <https://doi.org/10.1038/sj.bjc.6604676>.
- (56) Schiff, D.; Sarkaria, J. Dasatinib in Recurrent Glioblastoma: Failure as a Teacher. *Neuro-Oncology* **2015**, *17* (7), 910–911. <https://doi.org/10.1093/neuonc/nov086>.
- (57) Muhammad, N.; Wang, R.; Li, W.; Zhang, Z.; Chang, Y.; Hu, Y.; Zhao, J.; Zheng, X.; Mao, Q.; Xia, H. A Novel TanCAR Targeting IL13R α 2 and EphA2 for Enhanced Glioblastoma Therapy. *Molecular Therapy - Oncolytics* **2022**, *24*, 729–741. <https://doi.org/10.1016/j.omto.2022.02.012>.
- (58) Ali Zaidi, S. S.; Fatima, F.; Ali Zaidi, S. A.; Zhou, D.; Deng, W.; Liu, S. Engineering siRNA Therapeutics: Challenges and Strategies. *J Nanobiotechnol* **2023**, *21* (1), 381. <https://doi.org/10.1186/s12951-023-02147-z>.
- (59) Rocha Pinheiro, S. L.; Lemos, F. F. B.; Marques, H. S.; Silva Luz, M.; De Oliveira Silva, L. G.; Faria Souza Mendes Dos Santos, C.; Da Costa Evangelista, K.; Calmon, M. S.; Sande Loureiro, M.; Freire De Melo, F. Immunotherapy in Glioblastoma Treatment: Current State and Future Prospects. *World J Clin Oncol* **2023**, *14* (4), 138–159. <https://doi.org/10.5306/wjco.v14.i4.138>.
- (60) Lassman, A. B.; Pugh, S. L.; Gilbert, M. R.; Aldape, K. D.; Geinoz, S.; Beumer, J. H.; Christner, S. M.; Komaki, R.; DeAngelis, L. M.; Gaur, R.; Youssef, E.; Wagner, H.; Won, M.; Mehta, M. P. Phase 2 Trial of Dasatinib in Target-Selected Patients with Recurrent Glioblastoma (RTOG 0627). *Neuro-Oncology* **2015**, *17* (7), 992–998. <https://doi.org/10.1093/neuonc/nov011>.
- (61) Zhou, J.; Rossi, J. Aptamers as Targeted Therapeutics: Current Potential and Challenges. *Nat Rev Drug Discov* **2017**, *16* (3), 181–202. <https://doi.org/10.1038/nrd.2016.199>.
- (62) Patra, A.; Paolillo, M.; Charisse, K.; Manoharan, M.; Rozners, E.; Egli, M. 2'-Fluoro RNA Shows Increased Watson-Crick H-Bonding Strength and Stacking Relative to RNA: Evidence from NMR and Thermodynamic Data. *Angew Chem Int Ed Engl* **2012**, *51* (47), 11863–11866. <https://doi.org/10.1002/anie.201204946>.
- (63) Pallan, P. S.; Greene, E. M.; Jicman, P. A.; Pandey, R. K.; Manoharan, M.; Rozners, E.; Egli, M. Unexpected Origins of the Enhanced

Pairing Affinity of 2'-Fluoro-Modified RNA. *Nucleic Acids Res* **2011**, *39* (8), 3482–3495. <https://doi.org/10.1093/nar/gkq1270>.

(64) Affinito, A.; Quintavalle, C.; Esposito, C. L.; Roscigno, G.; Vilardo, C.; Nuzzo, S.; Ricci-Vitiani, L.; De Luca, G.; Pallini, R.; Kichkailo, A. S.; Lapin, I. N.; de Francis, V.; Condorelli, G. The Discovery of RNA Aptamers That Selectively Bind Glioblastoma Stem Cells. *Molecular Therapy - Nucleic Acids* **2019**, *18*, 99–109. <https://doi.org/10.1016/j.omtn.2019.08.015>.

(65) Affinito, A.; Quintavalle, C.; Esposito, C. L.; Roscigno, G.; Giordano, C.; Nuzzo, S.; Ricci-Vitiani, L.; Scognamiglio, I.; Minic, Z.; Pallini, R.; Berezovski, M. V.; De Francis, V.; Condorelli, G. Targeting Ephrin Receptor Tyrosine Kinase A2 with a Selective Aptamer for Glioblastoma Stem Cells. *Molecular Therapy - Nucleic Acids* **2020**, *20*, 176–185. <https://doi.org/10.1016/j.omtn.2020.02.005>.

(66) Islam, B.; Stadlbauer, P.; Gil-Ley, A.; Pérez-Hernández, G.; Haider, S.; Neidle, S.; Bussi, G.; Banas, P.; Otyepka, M.; Sponer, J. Exploring the Dynamics of Propeller Loops in Human Telomeric DNA Quadruplexes Using Atomistic Simulations. *J Chem Theory Comput* **2017**, *13* (6), 2458–2480. <https://doi.org/10.1021/acs.jctc.7b00226>.

(67) Di Lorenzo, D. Tau Protein and Tauopathies: Exploring Tau Protein-Protein and Microtubule Interactions, Cross-Interactions and Therapeutic Strategies. *ChemMedChem* **2024**, *19* (21), e202400180. <https://doi.org/10.1002/cmdc.202400180>.

(68) Avila, J.; Lucas, J. J.; Pérez, M.; Hernández, F. Role of Tau Protein in Both Physiological and Pathological Conditions. *Physiological Reviews* **2004**, *84* (2), 361–384. <https://doi.org/10.1152/physrev.00024.2003>.

(69) Jouanne, M.; Rault, S.; Voisin-Chiret, A.-S. Tau Protein Aggregation in Alzheimer's Disease: An Attractive Target for the Development of Novel Therapeutic Agents. *European Journal of Medicinal Chemistry* **2017**, *139*, 153–167. <https://doi.org/10.1016/j.ejmech.2017.07.070>.

(70) Zheng, S.; Chen, Y.; Donahue, C. P.; Wolfe, M. S.; Varani, G. Structural Basis for Stabilization of the Tau Pre-mRNA Splicing Regulatory Element by Novantrone (Mitoxantrone). *Chem Biol* **2009**, *16* (5), 557–566. <https://doi.org/10.1016/j.chembiol.2009.03.009>.

(71) Chen, J. L.; Zhang, P.; Abe, M.; Aikawa, H.; Zhang, L.; Frank, A. J.; Zembryski, T.; Hubbs, C.; Park, H.; Withka, J.; Steppan, C.; Rogers, L.; Cabral, S.; Pettersson, M.; Wager, T. T.; Fountain, M. A.; Rumbaugh, G.;

- Childs-Disney, J. L.; Disney, M. D. Design, Optimization, and Study of Small Molecules That Target Tau Pre-mRNA and Affect Splicing. *J. Am. Chem. Soc.* **2020**, *142* (19), 8706–8727. <https://doi.org/10.1021/jacs.0c00768>.
- (72) Mathews, D. H.; Turner, D. H. Prediction of RNA Secondary Structure by Free Energy Minimization. *Current Opinion in Structural Biology* **2006**, *16* (3), 270–278. <https://doi.org/10.1016/j.sbi.2006.05.010>.
- (73) Reuter, J. S.; Mathews, D. H. RNAstructure: Software for RNA Secondary Structure Prediction and Analysis. *BMC Bioinformatics* **2010**, *11* (1), 129. <https://doi.org/10.1186/1471-2105-11-129>.
- (74) Lorenz, R.; Bernhart, S. H.; Höner Zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms Mol Biol* **2011**, *6* (1), 26. <https://doi.org/10.1186/1748-7188-6-26>.
- (75) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and Design of Nucleic Acid Systems. *J. Comput. Chem.* **2011**, *32* (1), 170–173. <https://doi.org/10.1002/jcc.21596>.
- (76) Ding, Y.; Chan, C. Y.; Lawrence, C. E. Sfold Web Server for Statistical Folding and Rational Design of Nucleic Acids. *Nucleic Acids Research* **2004**, *32* (Web Server), W135–W141. <https://doi.org/10.1093/nar/gkh449>.
- (77) Do, C. B.; Woods, D. A.; Batzoglou, S. CONTRAfold: RNA Secondary Structure Prediction without Physics-Based Models. *Bioinformatics* **2006**, *22* (14), e90–e98. <https://doi.org/10.1093/bioinformatics/btl246>.
- (78) Hamada, M.; Kiryu, H.; Sato, K.; Mituyama, T.; Asai, K. Prediction of RNA Secondary Structure Using Generalized Centroid Estimators. *Bioinformatics* **2009**, *25* (4), 465–473. <https://doi.org/10.1093/bioinformatics/btn601>.
- (79) Sato, K.; Hamada, M.; Asai, K.; Mituyama, T. CENTROIDFOLD: A Web Server for RNA Secondary Structure Prediction. *Nucleic Acids Research* **2009**, *37* (Web Server), W277–W280. <https://doi.org/10.1093/nar/gkp367>.
- (80) Xayaphoummine, A.; Bucher, T.; Isambert, H. Kinefold Web Server for RNA/DNA Folding Path and Structure Prediction Including Pseudoknots

and Knots. *Nucleic Acids Research* **2005**, *33* (suppl_2), W605–W610. <https://doi.org/10.1093/nar/gki447>.

(81) Isambert, H.; Siggia, E. D. Modeling RNA Folding Paths with Pseudoknots: Application to Hepatitis Delta Virus Ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (12), 6515–6520. <https://doi.org/10.1073/pnas.110533697>.

(82) Xayaphoummine, A.; Bucher, T.; Thalmann, F.; Isambert, H. Prediction and Statistics of Pseudoknots in RNA Structures Using Exactly Clustered Stochastic Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (26), 15310–15315. <https://doi.org/10.1073/pnas.2536430100>.

(83) Mahen, E. M.; Watson, P. Y.; Cottrell, J. W.; Fedor, M. J. mRNA Secondary Structures Fold Sequentially But Exchange Rapidly In Vivo. *PLoS Biol* **2010**, *8* (2), e1000307. <https://doi.org/10.1371/journal.pbio.1000307>.

(84) Li, J.; Zhang, S.; Zhang, D.; Chen, S.-J. Vfold-Pipeline: A Web Server for RNA 3D Structure Prediction from Sequences. *Bioinformatics* **2022**, *38* (16), 4042–4043. <https://doi.org/10.1093/bioinformatics/btac426>.

(85) Cheng, Y.; Zhang, S.; Xu, X.; Chen, S.-J. Vfold2D-MC: A Physics-Based Hybrid Model for Predicting RNA Secondary Structure Folding. *J. Phys. Chem. B* **2021**, *125* (36), 10108–10118. <https://doi.org/10.1021/acs.jpcc.1c04731>.

(86) Akiyama, M.; Sato, K.; Sakakibara, Y. A Max-Margin Training of RNA Secondary Structure Prediction Integrated with the Thermodynamic Model. *J. Bioinform. Comput. Biol.* **2018**, *16* (06), 1840025. <https://doi.org/10.1142/S0219720018400255>.

(87) Sato, K.; Akiyama, M.; Sakakibara, Y. RNA Secondary Structure Prediction Using Deep Learning with Thermodynamic Integration. *Nat Commun* **2021**, *12* (1), 941. <https://doi.org/10.1038/s41467-021-21194-4>.

(88) Poolsap, U.; Kato, Y.; Akutsu, T. Prediction of RNA Secondary Structure with Pseudoknots Using Integer Programming. *BMC Bioinformatics* **2009**, *10 Suppl 1* (Suppl 1), S38. <https://doi.org/10.1186/1471-2105-10-S1-S38>.

(89) Popenda, M.; Blazewicz, M.; Szachniuk, M.; Adamiak, R. W. RNA FRABASE Version 1.0: An Engine with a Database to Search for the Three-Dimensional Fragments within RNA Structures. *Nucleic Acids Res* **2008**, *36* (Database issue), D386–391. <https://doi.org/10.1093/nar/gkm786>.

- (90) Biesiada, M.; Pachulska-Wieczorek, K.; Adamiak, R. W.; Purzycka, K. J. RNAComposer and RNA 3D Structure Prediction for Nanotechnology. *Methods* **2016**, *103*, 120–127. <https://doi.org/10.1016/j.ymeth.2016.03.010>.
- (91) Wang, J.; Wang, J.; Huang, Y.; Xiao, Y. 3dRNA v2.0: An Updated Web Server for RNA 3D Structure Prediction. *IJMS* **2019**, *20* (17), 4116. <https://doi.org/10.3390/ijms20174116>.
- (92) Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K. M.; Bujnicki, J. M. SimRNA: A Coarse-Grained Method for RNA Folding Simulations and 3D Structure Prediction. *Nucleic Acids Res* **2016**, *44* (7), e63–e63. <https://doi.org/10.1093/nar/gkv1479>.
- (93) Pearce, R.; Omenn, G. S.; Zhang, Y. *De Novo* RNA Tertiary Structure Prediction at Atomic Resolution Using Geometric Potentials from Deep Learning. May 15, 2022. <https://doi.org/10.1101/2022.05.15.491755>.
- (94) Wang, W.; Feng, C.; Han, R.; Wang, Z.; Ye, L.; Du, Z.; Wei, H.; Zhang, F.; Peng, Z.; Yang, J. trRosettaRNA: Automated Prediction of RNA 3D Structure with Transformer Network. *Nat Commun* **2023**, *14* (1), 7266. <https://doi.org/10.1038/s41467-023-42528-4>.
- (95) Shen, T.; Hu, Z.; Sun, S.; Liu, D.; Wong, F.; Wang, J.; Chen, J.; Wang, Y.; Hong, L.; Xiao, J.; Zheng, L.; Krishnamoorthi, T.; King, I.; Wang, S.; Yin, P.; Collins, J. J.; Li, Y. Accurate RNA 3D Structure Prediction Using a Language Model-Based Deep Learning Approach. **2022**. <https://doi.org/10.48550/ARXIV.2207.01586>.
- (96) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Židek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**. <https://doi.org/10.1038/s41586-024-07487-w>.
- (97) Levinthal, C.; Wodak, S. J.; Kahn, P.; Dadivanian, A. K. Hemoglobin Interaction in Sick Cell Fibers. I: Theoretical Approaches to the Molecular Contacts. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72* (4), 1330–1334. <https://doi.org/10.1073/pnas.72.4.1330>.

- (98) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *Journal of Molecular Biology* **1982**, *161* (2), 269–288. [https://doi.org/10.1016/0022-2836\(82\)90153-X](https://doi.org/10.1016/0022-2836(82)90153-X).
- (99) Vakser, I. A. Protein-Protein Docking: From Interaction to Interactome. *Biophysical Journal* **2014**, *107* (8), 1785–1793. <https://doi.org/10.1016/j.bpj.2014.08.033>.
- (100) Zheng, J.; Hong, X.; Xie, J.; Tong, X.; Liu, S. P3DOCK: A Protein–RNA Docking Webserver Based on Template-Based and Template-Free Docking. *Bioinformatics* **2020**, *36* (1), 96–103. <https://doi.org/10.1093/bioinformatics/btz478>.
- (101) Akram, H.; Ur Rahman, M.; Mazhar, S.; Qamer, F.; Yousaf, A. Overcoming Translational Barriers in RNA–Protein Docking: Enhancing Computational Accuracy for Targeted Drug Discovery. *Future Medicinal Chemistry* **2025**, *17* (13), 1623–1640. <https://doi.org/10.1080/17568919.2025.2533061>.
- (102) Lin, Y.-F.; Cheng, C.-W.; Shih, C.-S.; Hwang, J.-K.; Yu, C.-S.; Lu, C.-H. MIB: Metal Ion-Binding Site Prediction and Docking Server. *J. Chem. Inf. Model.* **2016**, *56* (12), 2287–2291. <https://doi.org/10.1021/acs.jcim.6b00407>.
- (103) Elokely, K. M.; Doerksen, R. J. Docking Challenge: Protein Sampling and Molecular Docking Performance. *J. Chem. Inf. Model.* **2013**, *53* (8), 1934–1945. <https://doi.org/10.1021/ci400040d>.
- (104) Pansar, T.; Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **2018**, *23* (8), 1899. <https://doi.org/10.3390/molecules23081899>.
- (105) Bentham Science Publisher, B. S. P. Docking and Scoring - Theoretically Easy, Practically Impossible? *CMC* **2006**, *13* (25), 2995–3003. <https://doi.org/10.2174/092986706778521797>.
- (106) McGovern, S. L.; Shoichet, B. K. Information Decay in Molecular Docking Screens against Holo, Apo, and Modeled Conformations of Enzymes. *J. Med. Chem.* **2003**, *46* (14), 2895–2907. <https://doi.org/10.1021/jm0300330>.
- (107) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.* **2004**, *47* (1), 45–55. <https://doi.org/10.1021/jm030209y>.

- (108) Brooijmans, N.; Kuntz, I. D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, 32 (1), 335–373. <https://doi.org/10.1146/annurev.biophys.32.110601.142532>.
- (109) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J Comput Aided Mol Des* **2001**, 15 (5), 411–428. <https://doi.org/10.1023/A:1011115820450>.
- (110) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: Fast, Fully Automated Docking of Flexible Ligands to Protein Binding Sites. *Chemistry & Biology* **1996**, 3 (6), 449–462. [https://doi.org/10.1016/S1074-5521\(96\)90093-9](https://doi.org/10.1016/S1074-5521(96)90093-9).
- (111) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *Journal of Molecular Biology* **1996**, 261 (3), 470–489. <https://doi.org/10.1006/jmbi.1996.0477>.
- (112) Liu, M.; Wang, S. MCDOCK: A Monte Carlo Simulation Approach to the Molecular Docking Problem. *J Comput Aided Mol Des* **1999**, 13 (5), 435–451. <https://doi.org/10.1023/A:1008005918983>.
- (113) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking 1 Edited by F. E. Cohen. *Journal of Molecular Biology* **1997**, 267 (3), 727–748. <https://doi.org/10.1006/jmbi.1996.0897>.
- (114) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, 19 (14), 1639–1662. [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639::AID-JCC10>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B).
- (115) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **1995**, 91 (1–3), 1–41. [https://doi.org/10.1016/0010-4655\(95\)00041-D](https://doi.org/10.1016/0010-4655(95)00041-D).
- (116) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber

Biomolecular Simulation Programs. *J Comput Chem* **2005**, 26 (16), 1668–1688. <https://doi.org/10.1002/jcc.20290>.

(117) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM : A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **1983**, 4 (2), 187–217. <https://doi.org/10.1002/jcc.540040211>.

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

(119) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, 47 (7), 1739–1749. <https://doi.org/10.1021/jm0306430>.

(120) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, 47 (7), 1750–1759. <https://doi.org/10.1021/jm030644s>.

(121) Kollman, Peter. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, 93 (7), 2395–2417. <https://doi.org/10.1021/cr00023a004>.

(122) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **2006**, 49 (21), 6177–6196. <https://doi.org/10.1021/jm051256o>.

(123) Yoshida, N. Role of Solvation in Drug Design as Revealed by the Statistical Mechanics Integral Equation Theory of Liquids. *J. Chem. Inf. Model.* **2017**, 57 (11), 2646–2656. <https://doi.org/10.1021/acs.jcim.7b00389>.

(124) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, 35 (6), 358–365. <https://doi.org/10.1021/ar010014p>.

(125) Westhead, D. R.; Clark, D. E.; Murray, C. W. A Comparison of Heuristic Search Algorithms for Molecular Docking. *J Comput Aided Mol Des* **1997**, 11 (3), 209–228. <https://doi.org/10.1023/A:1007934310264>.

- (126) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Deciphering Common Failures in Molecular Docking of Ligand-Protein Complexes. *J Comput Aided Mol Des* **2000**, *14* (8), 731–751. <https://doi.org/10.1023/A:1008158231558>.
- (127) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J Comput Aided Mol Des* **2002**, *16* (1), 11–26. <https://doi.org/10.1023/A:1016357811882>.
- (128) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein–Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42* (5), 791–804. <https://doi.org/10.1021/jm980536j>.
- (129) Lazaridis, T. Ligand and Receptor Conformational Energies. In *Protein-Ligand Interactions*; Gohlke, H., Ed.; Wiley, 2012; pp 191–205. <https://doi.org/10.1002/9783527645947.ch10>.
- (130) Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E. Empirical Free Energy Calculations of Ligand-Protein Crystallographic Complexes. I. Knowledge-Based Ligand-Protein Interaction Potentials Applied to the Prediction of Human Immunodeficiency Virus 1 Protease Binding Affinity. *Protein Engineering Design and Selection* **1995**, *8* (7), 677–691. <https://doi.org/10.1093/protein/8.7.677>.
- (131) Gohlke, H.; Hendlich, M.; Klebe, G. Predicting Binding Modes, Binding Affinities and ‘hot Spots’ for Protein-Ligand Complexes Using a Knowledge-Based Scoring Function. *Perspectives in Drug Discovery and Design* **2000**, *20* (1), 115–144. <https://doi.org/10.1023/A:1008781006867>.
- (132) Liu, H.; Hu, B.; Chen, P.; Wang, X.; Wang, H.; Wang, S.; Wang, J.; Lin, B.; Cheng, M. Docking Score ML: Target-Specific Machine Learning Models Improving Docking-Based Virtual Screening in 155 Targets. *J. Chem. Inf. Model.* **2024**, *64* (14), 5413–5426. <https://doi.org/10.1021/acs.jcim.4c00072>.
- (133) Ain, Q. U.; Aleksandrova, A.; Roessler, F. D.; Ballester, P. J. Machine-learning Scoring Functions to Improve Structure-based Binding Affinity Prediction and Virtual Screening. *WIREs Comput Mol Sci* **2015**, *5* (6), 405–424. <https://doi.org/10.1002/wcms.1225>.
- (134) Li, J.; Fu, A.; Zhang, L. An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. *Interdiscip Sci Comput*

Life Sci **2019**, *11* (2), 320–328. <https://doi.org/10.1007/s12539-019-00327-w>.

(135) Xu, M.; Lill, M. A. Induced Fit Docking, and the Use of QM/MM Methods in Docking. *Drug Discovery Today: Technologies* **2013**, *10* (3), e411–e418. <https://doi.org/10.1016/j.ddtec.2013.02.003>.

(136) Lie, M. A.; Thomsen, R.; Pedersen, C. N. S.; Schiøtt, B.; Christensen, M. H. Molecular Docking with Ligand Attached Water Molecules. *J. Chem. Inf. Model.* **2011**, *51* (4), 909–917. <https://doi.org/10.1021/ci100510m>.

(137) Roberts, B. C.; Mancera, R. L. Ligand–Protein Docking with Water Molecules. *J. Chem. Inf. Model.* **2008**, *48* (2), 397–408. <https://doi.org/10.1021/ci700285e>.

(138) Pavlovicz, R. E.; Park, H.; DiMaio, F. Efficient Consideration of Coordinated Water Molecules Improves Computational Protein-Protein and Protein-Ligand Docking. April 25, 2019. <https://doi.org/10.1101/618603>.

(139) Pinzi, L.; Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *IJMS* **2019**, *20* (18), 4331. <https://doi.org/10.3390/ijms20184331>.

(140) Kaur, T.; Madgulkar, A.; Bhalekar, M.; Asgaonkar, K. Molecular Docking in Formulation and Development. *CDDT* **2019**, *16* (1), 30–39. <https://doi.org/10.2174/1570163815666180219112421>.

(141) Gupta V.P. *Principles and Applications of Quantum Chemistry*, 1st edition.; 2015.

(142) Yang, Y.; Li, B. A Simulation Algorithm for Brownian Dynamics on Complex Curved Surfaces. *The Journal of Chemical Physics* **2019**, *151* (16), 164901. <https://doi.org/10.1063/1.5126201>.

(143) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; De Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations. *WIREs Comput Mol Sci* **2014**, *4* (3), 225–248. <https://doi.org/10.1002/wcms.1169>.

(144) Dans, P. D.; Walther, J.; Gómez, H.; Orozco, M. Multiscale Simulation of DNA. *Current Opinion in Structural Biology* **2016**, *37*, 29–45. <https://doi.org/10.1016/j.sbi.2015.11.011>.

(145) Leach R. Andrew. *Molecular Modeling Principles and Applications*, 2nd edition.; 1999.

- (146) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106* (5), 1589–1615. <https://doi.org/10.1021/cr040426m>.
- (147) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids. *J. Am. Chem. Soc.* **1995**, *117* (48), 11946–11975. <https://doi.org/10.1021/ja00153a017>.
- (148) Lennard-Jones Potential. *Chemistry Glossary*.
- (149) Ewald, P. P. Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale. *Annalen der Physik* **1921**, *369* (3), 253–287. <https://doi.org/10.1002/andp.19213690304>.
- (150) Rapaport D. C. *The Art of Molecular Dynamics Simulations*, 2nd edition.; Cabridge University Press, 2004.
- (151) *Model Box Periodic Boundary Conditions - P.B.C.* I.S.A.A.C.S. <https://isaacs.sourceforge.io/phys/pbc.html>.
- (152) Frenkel Daan; Smit Berend. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press, 2002.
- (153) Cournia, Z. Structure and Dynamics of Biomembranes Containing Cholesterol and Other Biologically-Important Sterols : A Computational Perspective. **2006**. <https://doi.org/10.11588/HEIDOK.00007044>.
- (154) Cauchy, A.-L. *Oeuvres Complètes*, 1st ed.; Cambridge University Press, 2009. <https://doi.org/10.1017/CBO9780511702396>.
- (155) Blauch David. The Maxwell Distribution Laws. In *Physical Chemistry for Life Sciences*; LibreTexts Chemistry: University of Arkansas Little Rock.
- (156) Phadungsukanan, W.; Kraft, M.; Townsend, J. A.; Murray-Rust, P. The Semantics of Chemical Markup Language (CML) for Computational Chemistry : CompChem. *J Cheminform* **2012**, *4* (1), 15. <https://doi.org/10.1186/1758-2946-4-15>.
- (157) Allen M. P. Molecular Dynamics Simulations. In *Computer Simulations of Liquids*; Oxford University Press, 2002.
- (158) Grotendorst, J.; Marx Dominik. *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*; Publication Series of the John von Neumann Institute for Computing (NIC) NIC Series Volume, 2002; Vol. 10.

- (159) Moore, C. C. Ergodic Theorem, Ergodic Theory, and Statistical Mechanics. *Proc Natl Acad Sci U S A* **2015**, *112* (7), 1907–1911. <https://doi.org/10.1073/pnas.1421798112>.
- (160) Hakala, R. W. A New Derivation of the Boltzmann Distribution Law. *J. Chem. Educ.* **1961**, *38* (1), 33. <https://doi.org/10.1021/ed038p33>.
- (161) McQuarrie D. A. *Statistical Mechanics*; University Science Books, 2000.
- (162) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *LiveCoMS* **2019**, *1* (1). <https://doi.org/10.33011/livecoms.1.1.5957>.
- (163) Davidchack, R. L.; Handel, R.; Tretyakov, M. V. Langevin Thermostat for Rigid Body Dynamics. *The Journal of Chemical Physics* **2009**, *130* (23), 234101. <https://doi.org/10.1063/1.3149788>.
- (164) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *The Journal of Chemical Physics* **1980**, *72* (4), 2384–2393. <https://doi.org/10.1063/1.439486>.
- (165) Klages, R. *Microscopic Chaos, Fractals and Transport in Nonequilibrium Statistical Mechanics*; Advanced Series in Nonlinear Dynamics; WORLD SCIENTIFIC, 2007; Vol. 24. <https://doi.org/10.1142/5945>.
- (166) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126* (1), 014101. <https://doi.org/10.1063/1.2408420>.
- (167) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690. <https://doi.org/10.1063/1.448118>.
- (168) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.
- (169) Parrinello, M.; Rahman, A. Strain Fluctuations and Elastic Constants. *The Journal of Chemical Physics* **1982**, *76* (5), 2662–2666. <https://doi.org/10.1063/1.443248>.

- (170) Najafabadi, R.; Yip, S. Observation of Finite-Temperature Bain Transformation (f.c.c. $\rightarrow$ r b.c.c.) in Monte Carlo Simulation of Iron. *Scripta Metallurgica* **1983**, *17* (10), 1199–1204. [https://doi.org/10.1016/0036-9748\(83\)90283-1](https://doi.org/10.1016/0036-9748(83)90283-1).
- (171) Yashonath, S.; Rao, C. N. R. A Monte Carlo Study of Crystal Structure Transformations. *Molecular Physics* **1985**, *54* (1), 245–251. <https://doi.org/10.1080/00268978500100201>.
- (172) Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R. K. Artificial Intelligence in Drug Discovery and Development. *Drug Discovery Today* **2021**, *26* (1), 80–93. <https://doi.org/10.1016/j.drudis.2020.10.010>.
- (173) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York: New York, NY, 2009. <https://doi.org/10.1007/978-0-387-84858-7>.
- (174) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer Texts in Statistics; Springer New York: New York, NY, 2013; Vol. 103. <https://doi.org/10.1007/978-1-4614-7138-7>.
- (175) Molnar Christoph. *Interpretable Machine Learning*; 2020.
- (176) Choi, R. Y.; Coyner, A. S.; Kalpathy-Cramer, J.; Chiang, M. F.; Campbell, J. P. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol* **2020**, *9* (2), 14. <https://doi.org/10.1167/tvst.9.2.14>.
- (177) Vemula, D.; Jayasurya, P.; Sushmitha, V.; Kumar, Y. N.; Bhandari, V. CADD, AI and ML in Drug Discovery: A Comprehensive Review. *European Journal of Pharmaceutical Sciences* **2023**, *181*, 106324. <https://doi.org/10.1016/j.ejps.2022.106324>.
- (178) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444. <https://doi.org/10.1038/nature14539>.
- (179) Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep Learning for Computational Biology. *Mol Syst Biol* **2016**, *12* (7), 878. <https://doi.org/10.15252/msb.20156651>.
- (180) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA* **2020**, *323* (9), 844. <https://doi.org/10.1001/jama.2020.1166>.

- (181) Özçelik, R.; Van Tilborg, D.; Jiménez-Luna, J.; Grisoni, F. Structure-Based Drug Discovery with Deep Learning**. *ChemBioChem* **2023**, *24* (13), e202200776. <https://doi.org/10.1002/cbic.202200776>.
- (182) Yang, Z.; Zeng, X.; Zhao, Y.; Chen, R. AlphaFold2 and Its Applications in the Fields of Biology and Medicine. *Sig Transduct Target Ther* **2023**, *8* (1), 115. <https://doi.org/10.1038/s41392-023-01381-z>.
- (183) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Židek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
- (184) Yang, K.; Xie, Z.; Li, Z.; Qian, X.; Sun, N.; He, T.; Xu, Z.; Jiang, J.; Mei, Q.; Wang, J.; Qu, S.; Xu, X.; Chen, C.; Ju, B. MolProphet: A One-Stop, General Purpose, and AI-Based Platform for the Early Stages of Drug Discovery. *J. Chem. Inf. Model.* **2024**, *64* (8), 2941–2947. <https://doi.org/10.1021/acs.jcim.3c01979>.
- (185) Wang, J.; Yi, X.; Guo, R.; Jin, H.; Xu, P.; Li, S.; Wang, X.; Guo, X.; Li, C.; Xu, X.; Yu, K.; Yuan, Y.; Zou, Y.; Long, J.; Cai, Y.; Li, Z.; Zhang, Z.; Mo, Y.; Gu, J.; Jiang, R.; Wei, Y.; Xie, C. Milvus: A Purpose-Built Vector Data Management System. In *Proceedings of the 2021 International Conference on Management of Data*; ACM: Virtual Event China, 2021; pp 2614–2627. <https://doi.org/10.1145/3448016.3457550>.
- (186) Crampon, K.; Giorkallos, A.; Deldossi, M.; Baud, S.; Steffanel, L. A. Machine-Learning Methods for Ligand–Protein Molecular Docking. *Drug Discovery Today* **2022**, *27* (1), 151–164. <https://doi.org/10.1016/j.drudis.2021.09.007>.
- (187) Tran, T. T. V.; Surya Wibowo, A.; Tayara, H.; Chong, K. T. Artificial Intelligence in Drug Toxicity Prediction: Recent Advances, Challenges, and Future Perspectives. *J. Chem. Inf. Model.* **2023**, *63* (9), 2628–2643. <https://doi.org/10.1021/acs.jcim.3c00200>.

- (188) Jarallah, S. J.; Almughem, F. A.; Alhumaid, N. K.; Fayez, N. A.; Alradwan, I.; Alsulami, K. A.; Tawfik, E. A.; Alshehri, A. A. Artificial Intelligence Revolution in Drug Discovery: A Paradigm Shift in Pharmaceutical Innovation. *International Journal of Pharmaceutics* **2025**, *680*, 125789. <https://doi.org/10.1016/j.ijpharm.2025.125789>.
- (189) Lepre, C. A.; Moore, J. M.; Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chem. Rev.* **2004**, *104* (8), 3641–3676. <https://doi.org/10.1021/cr030409h>.
- (190) Sugiki, T.; Furuita, K.; Fujiwara, T.; Kojima, C. Current NMR Techniques for Structure-Based Drug Discovery. *Molecules* **2018**, *23* (1), 148. <https://doi.org/10.3390/molecules23010148>.
- (191) Vanwetswinkel, S.; Heetebrij, R. J.; Van Duynhoven, J.; Hollander, J. G.; Filippov, D. V.; Hajduk, P. J.; Siegal, G. TINS, Target Immobilized NMR Screening: An Efficient and Sensitive Method for Ligand Discovery. *Chemistry & Biology* **2005**, *12* (2), 207–216. <https://doi.org/10.1016/j.chembiol.2004.12.004>.
- (192) Krishnan, V. Ligand Screening by Saturation-Transfer Difference (STD) NMR Spectroscopy. *CAC* **2005**, *1* (3), 307–320. <https://doi.org/10.2174/157341105774573956>.
- (193) Viegas, A.; Manso, J.; Corvo, M. C.; Marques, M. M. B.; Cabrita, E. J. Binding of Ibuprofen, Ketorolac, and Diclofenac to COX-1 and COX-2 Studied by Saturation Transfer Difference NMR. *J. Med. Chem.* **2011**, *54* (24), 8555–8562. <https://doi.org/10.1021/jm201090k>.
- (194) Wagstaff, J. L.; Taylor, S. L.; Howard, M. J. Recent Developments and Applications of Saturation Transfer Difference Nuclear Magnetic Resonance (STD NMR) Spectroscopy. *Mol. BioSyst.* **2013**, *9* (4), 571–577. <https://doi.org/10.1039/C2MB25395J>.
- (195) Pinto, I. G.; Guilbert, C.; Ulyanov, N. B.; Stearns, J.; James, T. L. Discovery of Ligands for a Novel Target, the Human Telomerase RNA, Based on Flexible-Target Virtual Screening and NMR. *J Med Chem* **2008**, *51* (22), 7205–7215. <https://doi.org/10.1021/jm800825n>.
- (196) Koos, M. R. M.; Schulz, K. H. G.; Gil, R. R. Reference-Free NOE NMR Analysis. *Chem Sci* **2020**, *11* (36), 9930–9936. <https://doi.org/10.1039/d0sc02970j>.
- (197) Bell, R. A.; Saunders, J. K. Correlation of the Intramolecular Nuclear Overhauser Effect with Internuclear Distance. *Can. J. Chem.* **1970**, *48* (7), 1114–1122. <https://doi.org/10.1139/v70-184>.

- (198) Meyer, B.; Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew Chem Int Ed* **2003**, 42 (8), 864–890. <https://doi.org/10.1002/anie.200390233>.
- (199) Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a Method for Primary NMR Screening: Practical Aspects and Range of Applicability. *J Biomol NMR* **2001**, 21 (4), 349–359. <https://doi.org/10.1023/A:1013302231549>.
- (200) Gairí, M.; Feliz, M.; García, J. Saturation Transfer Difference Spectroscopy. In *Encyclopedia of Analytical Chemistry*; Meyers, R. A., Ed.; Wiley, 2016; pp 1–17. <https://doi.org/10.1002/9780470027318.a9498>.
- (201) Parisien, M.; Major, F. The MC-Fold and MC-Sym Pipeline Infers RNA Structure from Sequence Data. *Nature* **2008**, 452 (7183), 51–55. <https://doi.org/10.1038/nature06684>.
- (202) Janssen, S.; Giegerich, R. The RNA Shapes Studio. *Bioinformatics* **2015**, 31 (3), 423–425. <https://doi.org/10.1093/bioinformatics/btu649>.
- (203) Darty, K.; Denise, A.; Ponty, Y. VARNA: Interactive Drawing and Editing of the RNA Secondary Structure. *Bioinformatics* **2009**, 25 (15), 1974–1975. <https://doi.org/10.1093/bioinformatics/btp250>.
- (204) Chen, Z.; Hu, L.; Zhang, B.-T.; Lu, A.; Wang, Y.; Yu, Y.; Zhang, G. Artificial Intelligence in Aptamer–Target Binding Prediction. *IJMS* **2021**, 22 (7), 3605. <https://doi.org/10.3390/ijms22073605>.
- (205) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; Van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, 103 (19), 3596–3607. <https://doi.org/10.1021/jp984217f>.
- (206) Bottaro, S.; Bussi, G.; Pinamonti, G.; Reißer, S.; Boomsma, W.; Lindorff-Larsen, K. Barnaba: Software for Analysis of Nucleic Acid Structures and Trajectories. *RNA* **2019**, 25 (2), 219–231. <https://doi.org/10.1261/rna.067678.118>.
- (207) Bottaro, S.; Di Palma, F.; Bussi, G. The Role of Nucleobase Interactions in RNA Structure and Dynamics. *Nucleic Acids Research* **2014**, 42 (21), 13306–13314. <https://doi.org/10.1093/nar/gku972>.
- (208) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations

- through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (209) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, 7 (9), 2886–2902. <https://doi.org/10.1021/ct200162x>.
- (210) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzes, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O’Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Rishch, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E.; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M. AmberTools. *J. Chem. Inf. Model.* **2023**, 63 (20), 6183–6191. <https://doi.org/10.1021/acs.jcim.3c01153>.
- (211) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, 79 (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (212) Joung, I. S.; Cheatham, T. E. Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, 113 (40), 13279–13290. <https://doi.org/10.1021/jp902584c>.
- (213) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, 103 (19), 8577–8593. <https://doi.org/10.1063/1.470117>.
- (214) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Molecular Simulation* **1988**, 1 (3), 173–185. <https://doi.org/10.1080/08927028808080941>.
- (215) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, 18 (12), 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).

- (216) Mlýnský, V.; Bussi, G. Understanding In-Line Probing Experiments by Modeling Cleavage of Nonreactive RNA Nucleotides. *RNA* **2017**, *23* (5), 712–720. <https://doi.org/10.1261/rna.060442.116>.
- (217) Van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; Van Dijk, M.; De Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **2016**, *428* (4), 720–725. <https://doi.org/10.1016/j.jmb.2015.09.014>.
- (218) Pierce, B. G.; Hourai, Y.; Weng, Z. Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. *PLoS ONE* **2011**, *6* (9), e24657. <https://doi.org/10.1371/journal.pone.0024657>.
- (219) Hu, W.-P.; Kumar, J. V.; Huang, C.-J.; Chen, W.-Y. Computational Selection of RNA Aptamer against Angiopoietin-2 and Experimental Evaluation. *Biomed Res Int* **2015**, *2015*, 658712. <https://doi.org/10.1155/2015/658712>.
- (220) Lee, G.; Jang, G. H.; Kang, H. Y.; Song, G. Predicting Aptamer Sequences That Interact with Target Proteins Using an Aptamer-Protein Interaction Classifier and a Monte Carlo Tree Search Approach. *PLoS One* **2021**, *16* (6), e0253760. <https://doi.org/10.1371/journal.pone.0253760>.
- (221) Ahirwar, R.; Nahar, S.; Aggarwal, S.; Ramachandran, S.; Maiti, S.; Nahar, P. In Silico Selection of an Aptamer to Estrogen Receptor Alpha Using Computational Docking Employing Estrogen Response Elements as Aptamer-Alike Molecules. *Sci Rep* **2016**, *6*, 21285. <https://doi.org/10.1038/srep21285>.
- (222) Joseph, D. F.; Nakamoto, J. A.; Garcia Ruiz, O. A.; Peñaranda, K.; Sanchez-Castro, A. E.; Castillo, P. S.; Milón, P. DNA Aptamers for the Recognition of HMGB1 from Plasmodium Falciparum. *PLoS One* **2019**, *14* (4), e0211756. <https://doi.org/10.1371/journal.pone.0211756>.
- (223) Heiat, M.; Najafi, A.; Ranjbar, R.; Latifi, A. M.; Rasaee, M. J. Computational Approach to Analyze Isolated ssDNA Aptamers against Angiotensin II. *Journal of Biotechnology* **2016**, *230*, 34–39. <https://doi.org/10.1016/j.jbiotec.2016.05.021>.
- (224) Pavan, M.; Bassani, D.; Sturlese, M.; Moro, S. Investigating RNA–Protein Recognition Mechanisms through Supervised Molecular Dynamics (SuMD) Simulations. *NAR Genomics and Bioinformatics* **2022**, *4* (4), lqac088. <https://doi.org/10.1093/nargab/lqac088>.

- (225) Kumar, J. V.; Chen, W.-Y.; Tsai, J. J. P.; Hu, W.-P. Molecular Simulation Methods for Selecting Thrombin-Binding Aptamers. In *Information Technology Convergence*; Park, J. J., Barolli, L., Xhafa, F., Jeong, H. Y., Eds.; Lecture Notes in Electrical Engineering; Springer Netherlands: Dordrecht, 2013; Vol. 253, pp 743–749. https://doi.org/10.1007/978-94-007-6996-0_78.
- (226) Himanen, J. P.; Yermekbayeva, L.; Janes, P. W.; Walker, J. R.; Xu, K.; Atapattu, L.; Rajashankar, K. R.; Mensinga, A.; Lackmann, M.; Nikolov, D. B.; Dhe-Paganon, S. Architecture of Eph Receptor Clusters. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (24), 10860–10865. <https://doi.org/10.1073/pnas.1004148107>.
- (227) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (228) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (229) Zeng, C.; Jian, Y.; Vosoughi, S.; Zeng, C.; Zhao, Y. Evaluating Native-like Structures of RNA-Protein Complexes through the Deep Learning Method. *Nat Commun* **2023**, *14* (1), 1060. <https://doi.org/10.1038/s41467-023-36720-9>.
- (230) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys J* **2007**, *92* (11), 3817–3829. <https://doi.org/10.1529/biophysj.106.097782>.
- (231) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405–4412. <https://doi.org/10.1021/ct300613v>.
- (232) Fröhling, T.; Mlýnský, V.; Janeček, M.; Kührová, P.; Krepl, M.; Banáš, P.; Šponer, J.; Bussi, G. Automatic Learning of Hydrogen-Bond Fixes in the AMBER RNA Force Field. *J. Chem. Theory Comput.* **2022**, *18* (7), 4490–4502. <https://doi.org/10.1021/acs.jctc.2c00200>.
- (233) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA Force Field with Accuracy Comparable to State-of-the-Art Protein Force Fields. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (7). <https://doi.org/10.1073/pnas.1713027115>.

- (234) Tucker, M. R.; Piana, S.; Tan, D.; LeVine, M. V.; Shaw, D. E. Development of Force Field Parameters for the Simulation of Single- and Double-Stranded DNA Molecules and DNA–Protein Complexes. *J. Phys. Chem. B* **2022**, *126* (24), 4442–4457. <https://doi.org/10.1021/acs.jpcc.1c10971>.
- (235) Hagler, A. T.; Huler, E.; Lifson, S. Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals. *J. Am. Chem. Soc.* **1974**, *96* (17), 5319–5327. <https://doi.org/10.1021/ja00824a004>.
- (236) Lifson, S.; Warshel, A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and *n* - Alkane Molecules. *The Journal of Chemical Physics* **1968**, *49* (11), 5116–5129. <https://doi.org/10.1063/1.1670007>.
- (237) Eichert, A.; Perbandt, M.; Schreiber, A.; Fürste, J. P.; Betzel, C.; Erdmann, V. A.; Förster, C. Crystal Structure of the Human tRNAGly Microhelix Isoacceptor G9990 at 1.18 Å Resolution. *Biochemical and Biophysical Research Communications* **2009**, *380* (3), 503–507. <https://doi.org/10.1016/j.bbrc.2009.01.127>.
- (238) Sheng, J.; Larsen, A.; Heuberger, B. D.; Blain, J. C.; Szostak, J. W. Crystal Structure Studies of RNA Duplexes Containing s^2 U:A and s^2 U:U Base Pairs. *J. Am. Chem. Soc.* **2014**, *136* (39), 13916–13924. <https://doi.org/10.1021/ja508015a>.
- (239) Förster, C.; Mankowska, M.; Fürste, J. P.; Perbandt, M.; Betzel, Ch.; Erdmann, V. A. Crystal Structure of a Human tRNAGly Microhelix at 1.2 Å Resolution. *Biochemical and Biophysical Research Communications* **2008**, *368* (4), 996–1001. <https://doi.org/10.1016/j.bbrc.2008.01.175>.
- (240) Shankar, N.; Kennedy, S. D.; Chen, G.; Krugh, T. R.; Turner, D. H. The NMR Structure of an Internal Loop from 23S Ribosomal RNA Differs from Its Structure in Crystals of 50S Ribosomal Subunits. *Biochemistry* **2006**, *45* (39), 11776–11789. <https://doi.org/10.1021/bi0605787>.
- (241) Svoboda, P.; Cara, A. Di. Hairpin RNA: A Secondary Structure of Primary Importance. *Cell. Mol. Life Sci.* **2006**, *63* (7–8), 901–908. <https://doi.org/10.1007/s00018-005-5558-5>.
- (242) Correll, C. C. The Common and the Distinctive Features of the Bulged-G Motif Based on a 1.04 Å Resolution RNA Structure. *Nucleic Acids Research* **2003**, *31* (23), 6806–6818. <https://doi.org/10.1093/nar/gkg908>.

- (243) Nozinovic, S.; Fürtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. High-Resolution NMR Structure of an RNA Model System: The 14-Mer cUUCGg Tetraloop Hairpin RNA. *Nucleic Acids Research* **2010**, *38* (2), 683–694. <https://doi.org/10.1093/nar/gkp956>.
- (244) Li, M.; Deng, J.; Peng, X.; Wang, J.; Wilson, T. J.; Huang, L.; Lilley, D. M. J. Structure and Ion-Dependent Folding of k-Junctions. *RNA* **2023**, *29* (9), 1411–1422. <https://doi.org/10.1261/rna.079678.123>.
- (245) Elimelech-Zohar, K.; Orenstein, Y. An Overview on Nucleic-Acid G-Quadruplex Prediction: From Rule-Based Methods to Deep Neural Networks. *Briefings in Bioinformatics* **2023**, *24* (4), bbad252. <https://doi.org/10.1093/bib/bbad252>.
- (246) Portella, G.; Orozco, M.; Vendruscolo, M. Determination of a Structural Ensemble Representing the Dynamics of a G-Quadruplex DNA. *Biochemistry* **2020**, *59* (4), 379–388. <https://doi.org/10.1021/acs.biochem.9b00493>.
- (247) Terrell, J. R.; Le, T. T.; Paul, A.; Brinton, M. A.; Wilson, W. D.; Poon, G. M. K.; Germann, M. W.; Siemer, J. L. Structure of an RNA G-Quadruplex from the West Nile Virus Genome. *Nat Commun* **2024**, *15* (1), 5428. <https://doi.org/10.1038/s41467-024-49761-5>.
- (248) Baba, S.; Takahashi, K.; Noguchi, S.; Takaku, H.; Koyanagi, Y.; Yamamoto, N.; Kawai, G. Solution RNA Structures of the HIV-1 Dimerization Initiation Site in the Kissing-Loop and Extended-Duplex Dimers. *The Journal of Biochemistry* **2005**, *138* (5), 583–592. <https://doi.org/10.1093/jb/mvi158>.
- (249) Bouchard, P.; Legault, P. Structural Insights Into Substrate Recognition by the *Neurospora* Varkud Satellite Ribozyme: Importance of U-Turns at the Kissing-Loop Junction. *Biochemistry* **2014**, *53* (1), 258–269. <https://doi.org/10.1021/bi401491g>.
- (250) Yesselman, J. D.; Eiler, D.; Carlson, E. D.; Gotrik, M. R.; d’Aquino, A. E.; Ooms, A. N.; Kladwang, W.; Carlson, P. D.; Shi, X.; Costantino, D. A.; Herschlag, D.; Lucks, J. B.; Jewett, M. C.; Kieft, J. S.; Das, R. Computational Design of Three-Dimensional RNA Structure and Function. *Nat. Nanotechnol.* **2019**, *14* (9), 866–873. <https://doi.org/10.1038/s41565-019-0517-8>.
- (251) The PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat Methods* **2019**, *16* (8), 670–673. <https://doi.org/10.1038/s41592-019-0506-8>.

- (252) Wieczor, M.; Czub, J. Gromologist: A Gromacs-Oriented Utility Library for Structure and Topology Manipulation. May 24, 2022. <https://doi.org/10.26434/chemrxiv-2022-dhswc>.
- (253) Clark, J. J.; Benson, M. L.; Smith, R. D.; Carlson, H. A. Inherent versus Induced Protein Flexibility: Comparisons within and between Apo and Holo Structures. *PLoS Comput Biol* **2019**, *15* (1), e1006705. <https://doi.org/10.1371/journal.pcbi.1006705>.
- (254) Prime, Schrödinger, LLC, New York, NY, 2021.
- (255) Ng, E. W. M.; Shima, D. T.; Calias, P.; Cunningham, E. T.; Guyer, D. R.; Adamis, A. P. Pegaptanib, a Targeted Anti-VEGF Aptamer for Ocular Vascular Disease. *Nat Rev Drug Discov* **2006**, *5* (2), 123–132. <https://doi.org/10.1038/nrd1955>.
- (256) Jaffe, G. J.; Westby, K.; Csaky, K. G.; Monés, J.; Pearlman, J. A.; Patel, S. S.; Joondeph, B. C.; Randolph, J.; Masonson, H.; Rezaei, K. A. C5 Inhibitor Avacincaptad Pegol for Geographic Atrophy Due to Age-Related Macular Degeneration. *Ophthalmology* **2021**, *128* (4), 576–586. <https://doi.org/10.1016/j.ophtha.2020.08.027>.
- (257) Li, W.; Wu, C.-C.; Wang, S.; Zhou, L.; Qiao, L.; Ba, W.; Liu, F.; Zhan, L.; Chen, H.; Yu, J.-S.; Fang, J. Identification of the Target Protein of the Metastatic Colorectal Cancer-Specific Aptamer W3 as a Biomarker by Aptamer-Based Target Cells Sorting and Functional Characterization. *Biosensors and Bioelectronics* **2022**, *213*, 114451. <https://doi.org/10.1016/j.bios.2022.114451>.
- (258) Wilson, K.; Shiuan, E.; Brantley-Sieders, D. M. Oncogenic Functions and Therapeutic Targeting of EphA2 in Cancer. *Oncogene* **2021**, *40* (14), 2483–2495. <https://doi.org/10.1038/s41388-021-01714-8>.
- (259) Ramaswamy, V.; Monsalve, A.; Sautina, L.; Segal, M. S.; Dobson, J.; Allen, J. B. DNA Aptamer Assembly as a Vascular Endothelial Growth Factor Receptor Agonist. *Nucleic Acid Therapeutics* **2015**, *25* (5), 227–234. <https://doi.org/10.1089/nat.2014.0519>.
- (260) Meng, E. C.; Goddard, T. D.; Pettersen, E. F.; Couch, G. S.; Pearson, Z. J.; Morris, J. H.; Ferrin, T. E. UCSF CHIMERAX: Tools for Structure Building and Analysis. *Protein Science* **2023**, *32* (11), e4792. <https://doi.org/10.1002/pro.4792>.