UNIVERSIDADE
BEIRA INTERIOR

Molecular docking tools for the screening of affinity ligands for pre-miRNA-29b purification

Paulo Ricardo Esteves Almeida

Dissertação para obtenção do Grau de Mestre em
Biotecnologia
(2º ciclo de estudos)

Orientadora: Prof. Doutora Fani Pereira de Sousa
Co-orientador: Doutor Matheus Mendonça Pereira
Co-orientadora: Prof. Doutora Mara Guadalupe Freire

outubro de 2023

Declaração de Integridade

Eu, Paulo Ricardo Esteves Almeida, que abaixo assino, estudante com o número de inscrição 11693 de/o 2º Ciclo em Biotecnologia da Faculdade de Ciências, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 9 /10 /2023

“Nobody believes in you,
You’ve lost again and again and again
The lights are cut off,
But you still looking at your dream,
Reviewing it every day and say to yourself
It’s not over until I win.”
Les Brown

Acknowledgments

The last year has been of great changes in my life, both personally and professionally. I didn't cross this path alone, and so, these words are for all the people that supported and helped me grow.

Firstly, I would like to thank my supervisors Doctor Fani Sousa, Doctor Mara Freire, and Doctor Matheus Pereira for everything they've done for me. The scientific knowledge, advice, and lessons they imparted, along with their positive and kind words, have propelled me to this point. I will never forget their support. Thank you so much for always being there for me.

I was so lucky to be part of such an amazing lab group. I want to thank Priscilla Cravo, Bruno Rodrigues, Micaela Riscado, Ana Isabel, Pedro Ferreira, and Bruno Baptista for teaching me invaluable lessons and always cheering me up, even during challenging times. You transformed a grey work environment into a joyful space. Though we were few, we managed to win that Christmas door competition while singing with our amateur voices. Thank you immensely for the support and the unforgettable memories we made together.

A heartfelt thank you to Rita Carapito for the incredible support over the past year. I've learned so much with you, and I sincerely appreciate everything you've done for me. I couldn't have come this far without your invaluable support. Thank you so much for all the knowledge, attention, and confidence you transmitted. I am grateful for your continuous guidance and for consistently challenging me to improve as a professional.

To my closest friends, Alexandre Candeias, Carlos Martins, Diogo Pontes, Igor Dias, João Reis, and Rodrigo Rodrigues who have been with me since the beginning, you also have a presence on this work. Thank you so much for all the support.

A special thanks to Inês Ferraz, for her patient, support, and kindness. You have also played a significant role in shaping the person I am today. I don't have words to describe the impact you had on my personal and professional development. I am truly grateful to have embarked on this journey, knowing that you were there, watching and guiding me along the way.

I would like to thank all the members of CICS-UBI that directly or indirectly contributed to this work.

Last, but not least, I want to thank my parents, Paulo Almeida, and Laura Almeida, as well as my little sister, Diana Almeida, for all their support and love through each moment in my life.

To all of you, I will be always grateful. You owe me nothing. I owe you everything.

Resumo

Ao longo dos anos, o interesse pela utilização de ácidos nucleicos como biofármacos, na tentativa de identificar alternativas aos tratamentos tradicionais de diversas patologias como o cancro ou doenças neurodegenerativas tem aumentado. A capacidade de regular a expressão de genes pode levar a novas formas de tratamento para doenças cujos tratamentos atuais são ineficazes, como é o caso da doença de Alzheimer.

Para que possa ser considerada a possibilidade de utilização destas biomoléculas como biofármacos, é necessário que estas apresentem um elevado grau de pureza e a sua atividade biológica preservada. Recorrendo a processos biotecnológicos, a produção recombinante destas biomoléculas torna-se mais rápida, mais económica, e pode ser aplicada a nível industrial. No entanto, os passos subsequentes que estão relacionados com as etapas de separação e purificação podem tornar o processo dispendioso. A principal técnica de purificação atualmente utilizada na indústria biofarmacêutica é a cromatografia. No caso de se pretender estabelecer um processo mais seletivo, deverá considerar-se a cromatografia de afinidade, mas a seleção de ligandos específicos pode revelar-se o maior desafio desta etapa.

Neste âmbito, o presente trabalho tem como objetivo a identificação de ligandos para a purificação do pre-miRNA-29b, que apresenta potencial como biofármaco para o tratamento da doença de Alzheimer. Esta identificação e seleção de ligandos é realizada através de métodos *in silico*, o que torna o trabalho menos dispendioso, menos demoroso e com menor impacto ambiental.

Inicialmente, foram desenhados 13 oligonucleótidos (oligos) para interagirem com o pre-miRNA-29b por complementaridade de bases com um de três possíveis locais de ligação, nomeadamente a extremidade 5', a extremidade 3' ou a estrutura *hairpin*. O desenho destes ligandos contemplou a inclusão de uma cadeia carbonada ligada à extremidade 5' e um grupo amina de forma a ser possível a reação de imobilização do ligando ao suporte cromatográfico. Esta cadeia carbonada pode consistir em seis ou doze carbonos, o que faz com que cada oligonucleótido tenha duas versões, somando um total de 26 ligandos diferentes.

Após o desenho das sequências de oligos e do pre-miRNA-29b, estes foram submetidos num servidor para análise de *molecular docking* e posteriormente os resultados obtidos foram analisados no que diz respeito à energia de afinidade e tipos de interação. Após a

análise do *molecular docking score*, do número de interações estabelecidas entre o ligando e o alvo, e do local de interação, foram selecionados 4 oligos como os mais promissores. Estes ligandos permitirão perceber a influência da cadeia carbonada, assim como a diferença do potencial para purificação do pre-miRNA-29b entre oligos versáteis, que podem interagir com mais do que um local do alvo, e oligos seletivos, que apenas interagem com o local para o qual foram desenhados. Cerca de 80% das interações estabelecidas entre o ligando e o pre-miRNA-29b são pontes de hidrogénio, o que pode conferir uma certa estabilidade ao complexo (ligando-receptor).

Para obter mais informação sobre a estabilidade da interação entre os ligandos e o receptor, foram realizados ensaios de dinâmica molecular, e estes demonstraram que o complexo formado é bastante estável (RMSD < 1 Å). Desta forma, foi possível selecionar os melhores ligandos para a purificação do pre-miRNA-29b, assim como confirmar a sua estabilidade na interação com o alvo, o que é essencial para o posterior desenvolvimento de um método de purificação eficaz.

Palavras-chave

Oligonucleótidos; pre-miRNA-29b; Afinidade; *Docking* Molecular; Dinâmica Molecular

Abstract

Over the years, the interest in using nucleic acids as biopharmaceuticals to establish alternatives to traditional treatments for diseases such as cancer and neurodegenerative disorders has been increasing. The ability to regulate gene expression may lead to new forms of treatment for diseases with ineffective therapies, such as Alzheimer's disease.

To further study these biomolecules and make them usable as biopharmaceuticals, they need to have a high degree of purity and preserved biological activity. Through biotechnology, the recombinant production of these biomolecules is rapid, cost-effective, and applicable on an industrial scale. However, subsequent steps related to separation and purification make the process costly. The main purification technique currently used in the biopharmaceutical industry is chromatography. But if a more specific strategy is expected, affinity chromatography must be considered, where the selection of more selective ligands is the greatest challenge.

In this context, the present work aims to identify ligands for the purification of pre-miRNA-29b, which has potential as a biopharmaceutical for Alzheimer's disease treatment. This identification and selection of ligands was carried out through *in silico* methods, making the work less costly, time-saving, and environmentally friendly.

Initially, 13 oligonucleotides (oligos) were designed to interact with the pre-miRNA-29b through base complementarity with one of three possible binding sites: the 5' end, the 3' end, or the hairpin structure. Each of these oligos has a carbon chain linked to the 5' end and an amino group to allow the immobilization of the ligand onto the chromatographic support. This chain can have either six or twelve carbons, resulting in two versions of each oligonucleotide, summing up to a total of 26 different ligands.

After designing the oligo sequences and pre-miRNA-29b, they were submitted to a server for molecular docking analysis, and the results regarding the affinity energy and interaction capability were subsequently analyzed.

Following the analysis of the docking score, the number of connections between the ligand and the target, and the site of interaction, 4 oligos were selected as the most promising. These ligands will help understand the influence of the carbon chain and the potential differences for the pre-miRNA-29b purification between versatile oligos, which can interact with more than one target site, and selective oligos, which only interact with the site they were designed for. The results indicated that approximately 80% of the

interactions established between the ligand and pre-miRNA-29b are hydrogen bonds, which can provide some stability to the complex.

To obtain more information about the stability of the interaction between the ligands and the receptor, molecular dynamics assays were performed, demonstrating that the formed complex is quite stable (RMSD < 1 Å). This way, it was possible to select the best ligands for the purification of pre-miRNA-29b from the initial group, as well as to confirm their stability with the target.

Keywords

Oligonucleotides; pre-miRNA-29b; Affinity; Molecular Docking; Molecular Dynamics

Index

List of Figures

List of Tables

List of Acronyms

A	Adenine
AC	Affinity Chromatography
AD	Alzheimer's disease
AGO	Argonaute Proteins
APP	Amyloid Precursor Protein
asRNA	Antisense RNA
BACE1	β-site Amyloid Cleaving Enzyme 1
bp	Base Pair
C	Cytosine
DEAE	Diethyl Aminoethyl
DEPC	Diethyl Pirocarbonate
DNA	Deoxyribonucleic Acid
ds	Double-Stranded
E. coli	*Escherichia coli*
eRNA	Enhancer RNA
FDA	Food and Drug Administration
G	Guanine
gDNA	Genomic DNA
H-bond	Hydrogen Bond
HIC	Hydrophobic Interaction Chromatography
IEX	Ion Exchange Chromatography
MD	Molecular Dynamics
miRNA	Micro RNA
mRNA	Messenger RNA
ncRNA	Non-coding RNA
nt	Nucleotides
OD	Optical Density
pDNA	Plasmid DNA
piRNA	Piwi-Interacting RNA
pri-miRNA	Primary-miRNA
RISC	RNA Inducing Silencing Complex
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid

rRNA	Ribosomal RNA
SEC	Size Exclusion Chromatography
siRNA	Small-Interfering RNA
snoRNA	Small-Nucleolar RNA
snRNA	Small-Nuclear RNA
tRNA	Transfer RNA
U	Uracil

Scientific Communications

Flash Presentation

Paulo Almeida, Rita Carapito, Matheus Pereira, Mara G. Freire, Fani Sousa. Molecular docking of customized oligonucleotides as affinity ligands for pre-miRNA-29b purification. Affinity 2023. 5-7 June, Lisboa, Portugal.

Oral Presentation – Award for "Best MSc student oral communication"

Paulo Almeida, Rita Carapito, Matheus Pereira, Mara G. Freire, Fani Sousa. Molecular Docking as screening method of affinity ligands for pre-miRNA-29b purification. XVIII Annual CICS-UBI Symposium. 10-11 July, Covilhã, Portugal.

Chapter 1 - Introduction

1.1 Nucleic acids

In 1869, while studying the proteins that make up the leucocytes, the scientist Friedrich Miescher reported the discovery of a mysterious substance. It was made of elements commonly found in biomolecules like carbon, oxygen, hydrogen, and nitrogen, however, it was neither a protein nor a lipid nor a carbohydrate, but a new molecule that he called "nuclein" due to their presence in the cell nucleus [1]. Unfortunately, his discovery did not get much attention at the time, and only in the half of the 20th century, new studies demonstrated how important that molecule is as a carrier of genetic information [2,3]. The "nuclein", nowadays known as desoxyribonucleic acid (DNA), got its structure unraveled in 1953 by Watson and Crick, that suggested that DNA has two helical strands that are coiled in the same axis and held together due to the purine and pyrimidine bases [4]. Furthermore, they believed that the bases would pair following a certain pattern, that is adenine (purine) only pairs with thymine (pyrimidine), and guanine (purine) only pairs with cytosine (pyrimidine).

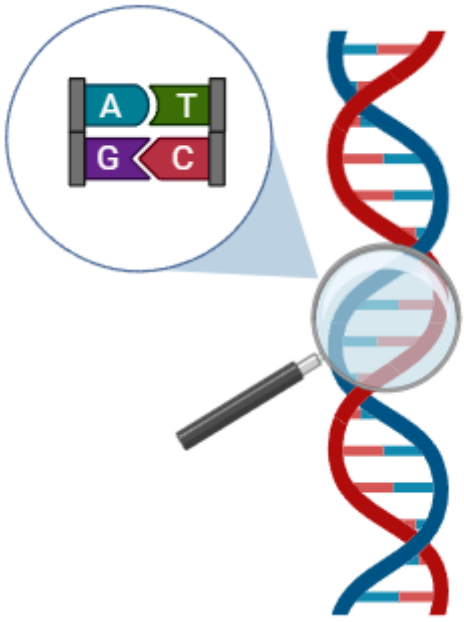


Figure 1: Structure of DNA (Adapted from [4]). The first representation of desoxyribonucleic acid proposed by Watson and Crick.

A few years later, Francis Crick proposed the central dogma of molecular biology, which refers to ribonucleic acid (RNA) as a simple intermediary between DNA and protein synthesis, in a unidirectional path, that is, starting with DNA and ending with a protein but never the opposite [5].

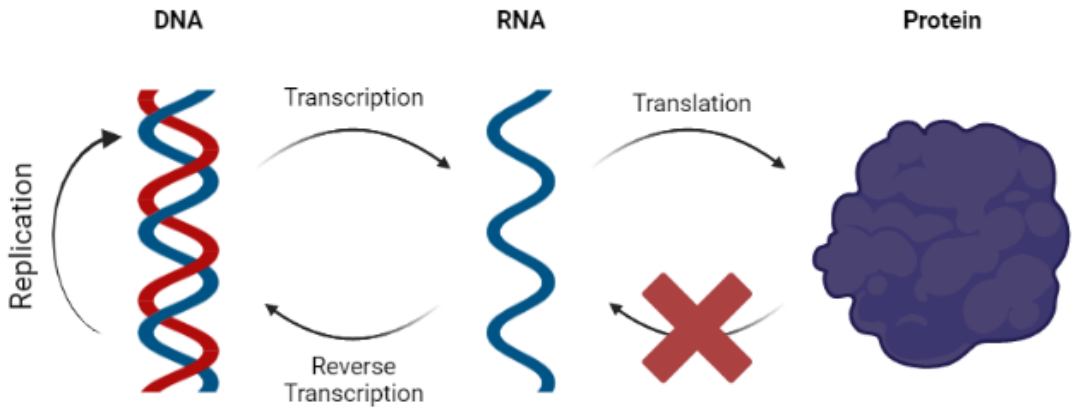


Figure 2: Central Dogma of Molecular Biology. According to the central dogma, RNA is an intermediary in the passage of DNA to proteins, while the passage of proteins to DNA is impossible.

Francis also speculated that every three bases in DNA specify one amino acid of the protein and in the 1960s, Nirenberg and his colleagues supported that idea as they cracked the genetic code [6]. Furthermore, from the sixty-four possible combinations of three nucleotides, known as codons in messenger RNA (mRNA), only 20 different amino acids are translated, meaning that a single amino acid could be encoded by different nucleotide triplets, therefore the natural code was classified as redundant. In addition, the same codon sequence obtained from distinct species translates for the same amino acid, which makes the genetic code universal [6].

For decades these molecules were recognized only as simple carriers of genetic information, however, the discovery of new molecules like non-coding RNAs (ncRNAs) with regulatory function in the 1980s would change the view of the scientific community [7]. In 1990, the Human Genome Project was initiated to determine the whole nucleotide sequence of the human genome as it could accelerate biomedical research and give a deep understanding of the nature of genetic information. After 10 years, the “complete” human genome was published, however, it covered only the euchromatic fraction, lacking 8% of the remaining genome, which was later addressed [8-10].

Nowadays, with a better understanding of genetics and the discovery of new nucleic acids with the ability to silence gene expression such as micro RNAs (miRNAs) [11], the research is turning to the use of nucleic acids as potential biopharmaceuticals against various diseases, including Alzheimer's disease [12].

1.1.1 Properties of RNA

Identically to DNA, the nucleotide is the simplest RNA unit, consisting of a phosphate group, a pentose sugar, and a nitrogenous base. In RNA, each nucleotide has one of four different bases: adenine (A), guanine (G), cytosine (C), or uracil (U) attached to the 1' carbon of the pentose, and a phosphate group also linked to the 5' carbon of the sugar moiety. A and G are classified as purines due to the double-ring structure whereas C and U are pyrimidines with only one ring. Moreover, the pentose sugar is the ribose in RNA, which in contrast to deoxyribose (present in DNA), has a hydroxyl group on the 2' carbon instead of hydrogen (Figure 3) [13-15]. This hydroxyl group makes the ribose more reactive than deoxyribose, being an easier target for a nucleophilic attack by hydroxide ions (S_N2 reaction), resulting in a less stable and short-life molecule compared to DNA [16]. In addition, for a pH > 4, RNA is a negatively charged molecule due to the phosphate groups in the backbone [17] and may present a more hydrophobic character compared to DNA, since its aromatic moieties are more exposed in the single chain [18].

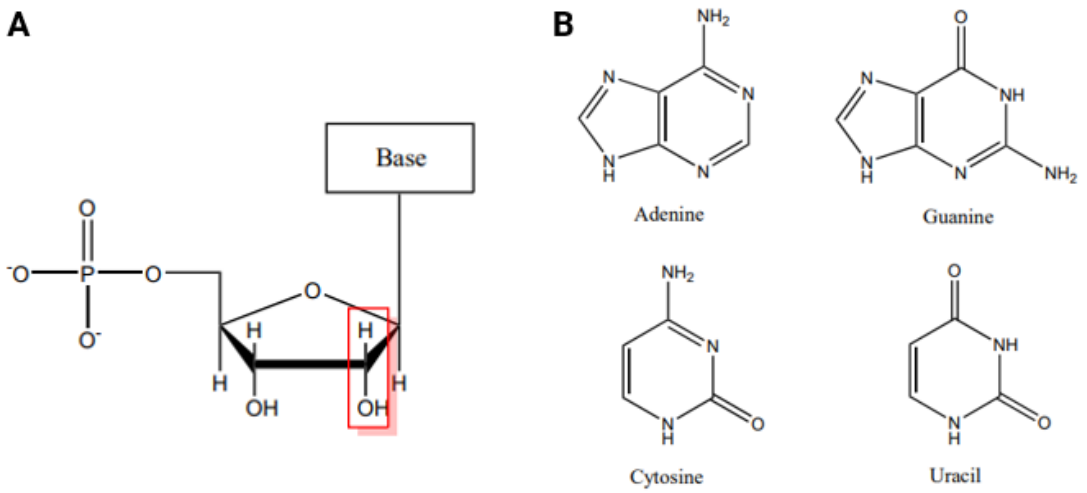


Figure 3: Structure of a ribonucleotide and its nitrogenous bases. A**:** Structure representation of a nucleotide in RNA, with the hydroxyl group (-OH) in the 2' carbon of the pentose highlighted by the red rectangle. B: structures of the ribonucleotide bases.

Nucleic acids are constituted by polynucleotide strands, where each nucleotide forms a phosphodiester bond between its 5'-phosphate group and the 3'-OH group of the consecutive nucleotide, therefore, DNA and RNA sequences are represented from the 5' to the 3' direction. Distinct from DNA, RNA is a single-stranded molecule, however, it has the potential to fold and form helical structures through the same mechanism as DNA. Hydrogen bonds (H-bonds) formed between the pairs A-U and C-G, as well as base π-stacking allows the RNA molecule to fold on itself resulting in RNA secondary structures (Figure 4). On edge, the interaction between secondary structure motifs may conduce to a tertiary structure of RNA [19,20]. As is common in proteins, the variability in the structure could lead to different behaviors, which is also true for nucleic acids. In fact, as a double-stranded molecule, DNA is conformationally restricted and its disposal to present different structures is limited, consequently, it has fewer roles in the cell than RNA [13] which will be discussed in the next section.

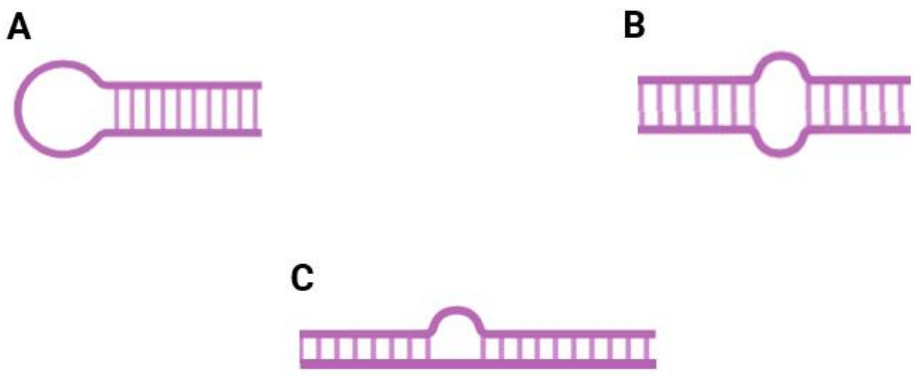


Figure 4: Secondary Structures of RNA (Adapted from [19]). A) hairpin is formed when the RNA strand folds back on itself. B) internal loops are formed wherever single or more bases are unpaired on each strand. C) bulges are formed wherever single or more bases are unpaired on only one strand.

Furthermore, the location of each nucleic acid can differ within the eukaryotic cells, while DNA can be found in the cell nucleus and mitochondria, RNA is present in the nucleus and cytoplasm [21]. Table 1 displays the main differences between DNA and RNA.

Table 1: Properties comparison between DNA and RNA.

	DNA	**RNA**
Bases	A, G, C, T	A, G, C, U
Strand	Double stranded	Single stranded
Structure	Antiparallel helix	Hairpin and loops
Sugar	Deoxyribose	Ribose
Location	Nuclear or mitochondrial	Nuclear or cytoplasmic
Lifetime	Long	Short

1.1.2 Coding and Non-coding RNA

For a long time, RNA was recognized only as a simple intermediate in protein synthesis, passing the genetic information from DNA to proteins. However, nowadays it is well known that are different types of RNAs with various roles in the cell. RNA can be divided into coding RNA and non-coding RNA (ncRNAs). Coding RNA refers to an RNA sequence involved in protein synthesis, whose role is encoding the protein's building blocks, which are the amino acids in the right order. Protein synthesis begins with the transcription of a DNA template by RNA polymerase in the nucleus, resulting in a complementary strand to the template, that is mRNA. However, this mRNA needs further processing to accomplish protein translation. At this point, various classes of ncRNAs are required, including small nuclear RNAs (snRNAs) that are involved in splicing events of mRNAs. The splicing process, described in eukaryotic cells, entails the removal of small sequences named introns from the mRNA during its maturation. Transfer RNAs (tRNAs) are also necessary, as they identify the codons in the mRNA sequence and recruit amino acids. Additionally, ribosomal RNAs (rRNAs) are essential as they are the most abundant RNA molecules in cells and form the ribosome, a crucial structure for protein synthesis. [14,22]. Therefore, ncRNAs are RNA molecules that do not encode proteins but have other functionalities in the cell and can be classified as housekeeping ncRNAs or regulatory ncRNAs. Housekeeping ncRNAs are usually small sequences, constitutively expressed and necessary for cell viability; they include rRNAs, tRNAs, snRNAs whose role was mentioned before, and small nucleolar RNAs (snoRNAs) that carry chemical modifications on these housekeeping ncRNAs [23]. On the other hand, regulatory ncRNAs are only expressed at specific cell types, at certain stages of cell differentiation, or in response to external stimuli. There are two categories of regulatory ncRNAs based on the strand length: if a sequence has less than 200 nucleotides (nt) then is a short ncRNA, otherwise is a long ncRNA (Figure 5). Short ncRNAs include microRNAs (miRNAs), small interfering RNAs (siRNAs), Piwi-interacting RNAs

(piRNAs), antisense RNAs (asRNAs), and enhancer RNAs (eRNAs), while long ncRNAs usually comprehend asRNAs and eRNAs [24,25].

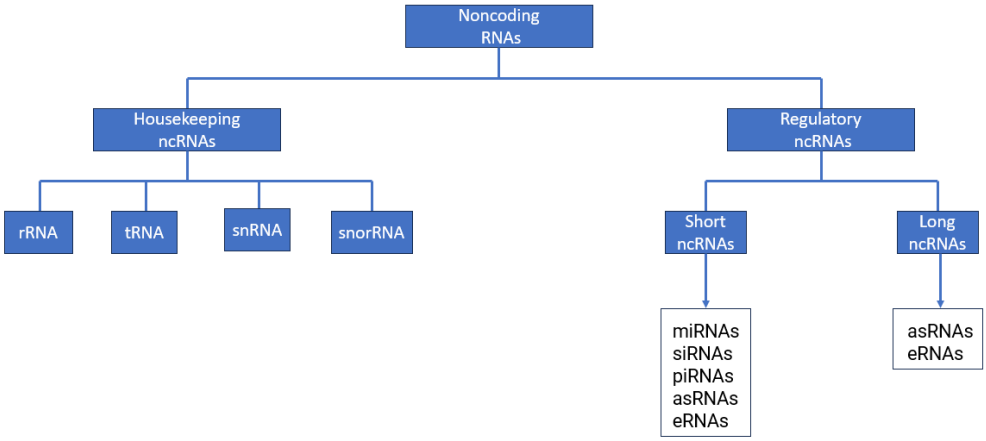


Figure 5: Types of noncoding RNAs (Adapted from [24]). Noncoding RNAs are classified as housekeeping or regulatory noncoding RNAs. Housekeeping ncRNAs include rRNA, tRNA, snRNA, and snoRNA. Depending on sequence length, regulatory can fit into short (<200 nt) or long (>200 nt) ncRNAs.

Long ncRNAs are involved in high-order chromosomal dynamics, telomere biology, and subcellular structural organization. They mediate epigenetic changes by recruiting chromatin remodeling complexes to specific genomic loci and participate in the regulation of transcription by directly binding to DNA sequences and inhibiting gene expression. Besides that, recent studies reported that long ncRNAs could encode short biological active peptides, for example, with transport function and that would allow the movement of proteins during essential stages of development in cells [26,27].

Short ncRNAs such as miRNAs, siRNAs, and piRNAs are responsible for various functions in the cell, including regulating gene expression, defense against viruses, suppressing transposons, and modification of chromatin structure [13]. MiRNAs are small single-stranded molecules with 20 to 24 nt, that have the ability to regulate the expression of nearly half of the genes in a cell at the post-transcriptional level. The biogenesis of miRNA starts with the transcription of its genes into a primary miRNA (pri-miRNA) and then proceeds to a pre-miRNA after the interaction of DGCR8 and Drosha. Pre-miRNA is then exported to the cytoplasm by a complex (XPO5/RanGTP) and processed again by the Dicer enzyme which removes the characteristic hairpin, resulting in a mature miRNA duplex. In the end, one of the strands, usually the strand with lower 5' stability, is loaded into the Argonaute family of proteins (AGO) and forms the RNA-

induced silencing complex (RISC). The miRNA then pairs with mRNAs, with a preference for the 3' untranslated region and induces their translational repression and degradation [28]. SiRNAs are 21 to 23 nt long double-stranded RNA molecules that also act at the post-transcriptional level. Identical to miRNA, siRNAs duplexes arise when the precedent is cleaved by Dicer which then follows a similar pathway until gene silencing [29]. PiRNAs are 24 to 32 nt non-coding RNAs that form complexes with PIWI proteins and are primarily derived from the transcripts of protein-coding genes, transposons, tRNA, and rRNA, possibly contributing to the targeted sequence specificity observed in mature piRNAs. Distinct from miRNAs and siRNAs, the maturation process of the piRNAs occurs in a Dicer-independent way. Mature piRNAs can either reenter the nucleus to induce transposon silencing and gene regulation or initiate secondary biogenesis known as the ping-pong cycle which produces piRNAs until the transcription of transposons is diminished [30,31]. Due to the ability to regulate protein expression by silencing the genes at a post-transcriptional level, ncRNAs, particularly miRNAs, and siRNAs are being explored as drugs for the treatment of various diseases including cancer [32] and neurological disorders [12,33].

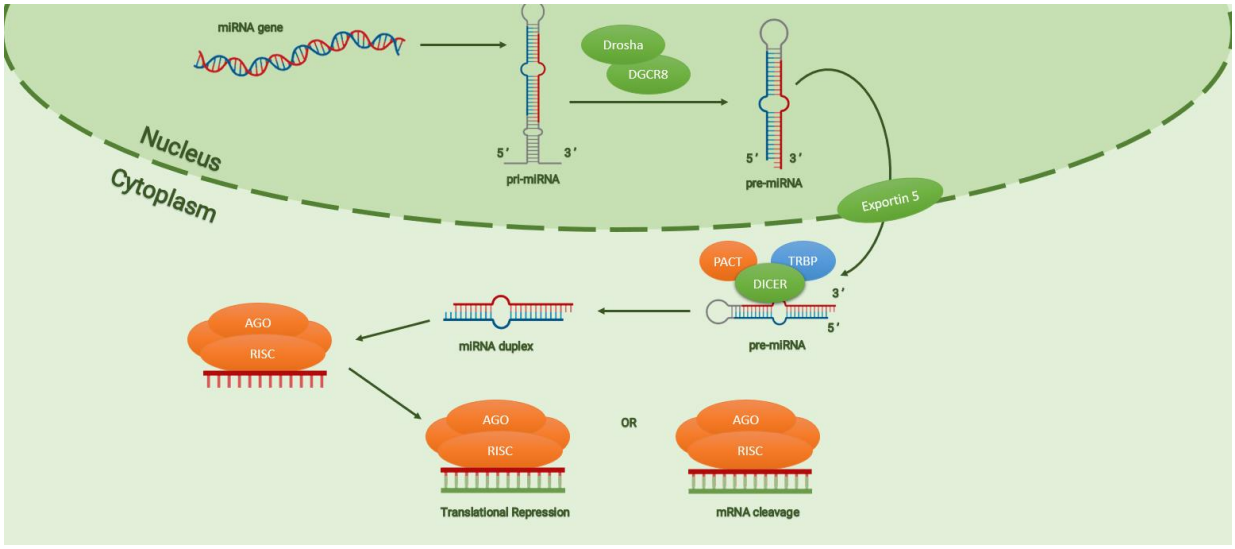


Figure 6: miRNA biogenesis pathway (Adapted from [34]).

1.1.3 Small ncRNAs vs Small Molecules as therapeutic agents

Over the past years, miRNAs and siRNAs have been receiving more attention as promising therapeutic agents for various diseases. Distinct from conventional drugs or small molecules, which usually target proteins, these biomolecules modulate gene expression to induce therapeutic effects [35]. This mode of action allows the inhibition of target molecules before their synthesis *in vivo*, thus interacting in a much closer relation to the origin of the disease [36]. This is an advantage over the traditional method

of delivering small molecules to bind at target proteins since its potential is limited by various factors regarding the protein structure, including the presence of binding pockets and the degree of polarity [37]. Moreover, it is estimated that of approximately 20000 human proteins, only 3000 can be treated by this traditional method [38]. In contrast, these biomolecules can block the protein expression itself by targeting the genes instead of proteins, thus dodging these problems. The ncRNAs are highly potent and can be designed to interact with any gene of interest including the genes that translate to proteins incapable of being inhibited by small molecules [39]. For instance, βIII-tubulin (encoded by the TUBB3 gene) is a cytoskeletal protein involved in cell mitosis that is found at high levels of expression in various types of cancer, thus being a target in chemotherapy. Despite the efforts of finding small molecules for this protein, βIII-tubulin is one of seven different isotypes that share high sequence homology, consequently, off-target toxicity is a major drawback of this approach [40]. Given the high analogy between structures and the lack of selectivity by small molecules, βIII-tubulin is characterized as an "undruggable" target, however, the genes that encode the different isotypes are distinct, and so, using ncRNAs to target and inhibit the specific gene expression could potentially circumvent the previous problem. Following this though, Teo and co-workers showed that βIII-tubulin siRNA reduced pancreatic tumor growth and metastases *in vivo* [41,42]. Table 2 shows some diseases and targets tested with this gene silencing approach.

Table 2: Examples of miRNA and siRNA molecules with therapeutical potential.

miRNA			
Molecule	**Target**	**Disease**	**Reference**
miRNA-29b	BACE1	Alzheimer's Disease	[43]
miRNA-34a	MYCN	Neuroblastoma	[44,45]
miRNA-181c	TLR4	Ischemic Stroke	[46,47]
miRNA-206	HDAC4	Amyotrophic Lateral Sclerosis	[48,49]
siRNA			
Molecule	**Target**	**Disease**	**Reference**
Patisiran	Transthyretin	Hereditary Transthyretin Amyloidosis	[50]
Inclisiran	PCSK9	Dyslipidemia	[51]
Givosiran	ALAS1	Acute Hepatic Porphyria	[52]
Lumasiran	Glycolate Oxidase	Primary Hyperoxaluria 1	[53]

1.1.4 MiRNA-29 for Alzheimer's Disease

Alzheimer's disease (AD) is considered the leading cause of dementia and one of the most lethal and burdensome diseases of this century. It is a neurodegenerative disorder characterized by a slow but progressive loss of memory, the inability to form new memories, and at later stages, the incapability of performing day-to-day tasks [12,54]. Various studies indicate that the accumulation of amyloid-β peptides (Aβ) in neurons is one of the main causes of this disease, as it leads to synaptic blockage and neuronal death [12]. As these phenomena are also found in the brains of healthy elderly individuals, it is widely accepted that the biological processes related to AD are present long before symptoms are expressed [55].

Aβ is a peptide composed of 40 or 42 amino acids, derived from sequential cleavages of amyloid precursor protein (APP) by β-site amyloid precursor protein cleaving enzyme 1 (BACE1). This protein is found to be overexpressed in patients with AD and so is considered a main target for therapeutic inhibition of Aβ production [12]. However, treatments aiming to reduce production or removing already produced proteins have not yet proven to be successful strategies. Currently available treatments on the market can only slow down the progression of the disease, without completely halting its progression [12].

On the other hand, recent studies have shown that AD patients present low levels of miRNA-29b and high expression of BACE1 protein. This supports the idea of a relationship between miRNA-29b and AD, as miRNA-29b inhibits BACE1 expression and consequently reduces the Aβ accumulation in neuronal cells [43,56]. Therefore, it is suggested that miRNA-29b can be considered as a potential biopharmaceutical for AD [43].

1.2 Production of RNA

As the scientific community keeps exploring and elucidating the potential of RNA in therapeutics, new challenges start to arise. Large amounts of RNA are required to deeply understand these new biopharmaceuticals' structure, biophysical properties, function, and mechanism [57] and further consider its use for therapy. Particularly for miRNA, there are different approaches to obtain this biomolecule, such as chemical synthesis, enzymatic reaction, or via recombinant production [58].

The basic principle of chemical synthesis involves the step-by-step addition of nucleotides to a growing RNA chain, using chemical reactions to form phosphodiester bonds between adjacent nucleotides. This process usually starts from the 3'-end

nucleotide and grows towards the 5'-end, adding each nucleotide at a time by repeating a four-step procedure (deblocking-coupling-capping-oxidation) until the full sequence is synthesized [59,60]. Chemical synthesis is considered to be the fastest method to produce miRNA and offers the possibility to include modifications such as fluorescent labels or chemical groups, however, several drawbacks are associated with this technique such as the length of the sequence that can be accurately synthesized and the expensive equipment required [60].

Distinct from chemical synthesis, enzymatic methods rely on the fundamentals of molecular biology, that is, using RNA polymerases and a DNA template to produce RNA transcripts *in vitro* [60]. The RNA polymerase is highly specific for its promotor sequence; thus, the common strategy is to place the target DNA sequence downstream from the promotor [61]. Due to the increased transcriptional activity, the bacteriophage systems (RNA polymerase and promotor) are usually selected for this process, being T7 the most used system [58,62]. In addition, a bacteriophage-derived RNA polymerase has a single polypeptide chain, which allows its simple recombinant production [63]. However, even with the high transcription rates, this method is not efficient for large-scale production of RNA, considering the amount of enzyme necessary and its costs [63]. Furthermore, low yields and multiple purification steps to remove impurities such as enzymes and DNA are associated with this method [60].

On the other hand, recombinant production is the only method for the synthesis of RNA in which all the process occurs *in vivo*. The principle of this method is to deliver the DNA sequence for the target RNA in a host that will use its cellular machinery to produce the target. The sequence to be transcript is inserted into a vector that will be used to transform the host cells so that the target starts to be produced and accumulated inside the cell [58]. This approach allows the production of large quantities of sample from cheap biomass fermentation, thus, being considered the most cost-effective method from the previous [60,64]. In addition, most of the native RNA have post-transcriptional modifications that are important for their structure and function, and these may not be reproducible under chemical or enzymatic synthesis [65]. Then, the cells are harvested, lysed and the RNA is purified through chromatography techniques [58].

1.3 Chromatography for RNA purification

Regardless of the sample amount, to fulfill the quality specifications required by pharmaceutical regulatory agencies like the Food and Drug Administration (FDA), which demands that the product for therapeutical use needs to be highly pure, many downstream steps are usually performed which increase the cost of the biotechnological process [66-68]. It is estimated that 50-80% of the global cost of a bioprocess is associated with downstream steps [69,70]. Among all the methods for biomolecules separation, chromatography is predominantly used in biopharmaceutical industries since it can isolate the desired component from a complex mixture [71].

Chromatography is a separation process in which the components of a mixture are distributed between the stationary and the mobile phases. The preference of each molecule in the mixture for a phase relies on its physicochemical properties, resulting in different retention times, thus achieving molecule separation [72]. A chromatographic process can be divided into 4 steps: 1) Loading phase, in which the sample is loaded onto the column in a buffer that promotes adsorption of the target molecule (binding buffer); 2) Washing phase, where non-binding or weakly bound species are washed off the column with either the binding buffer or another suitable mobile phase that promotes desorption of only the weakly bound species; 3) Elution phase, where the adsorbed species are eluted from the column by either completely or gradually replacing the binding buffer with the eluting buffer; 4) Regeneration phase, where the column must be re-equilibrated with the binding buffer before the next sample is applied [73].

The size, shape, charge, and hydrophobicity are some properties that vary within the biomolecules. Depending on the properties that are mostly explored, a set of chromatography techniques have been developed: Size Exclusion Chromatography (SEC), Ion-Exchange Chromatography (IEX), Hydrophobic Interaction Chromatography (HIC), and Affinity Chromatography (AC) [74].

1.3.1 Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) or gel filtration consists in a separation technique that depends on the relative size or hydrodynamic volume of a macromolecule and the average pore size of the packing [75]. Distinct from other chromatography techniques, the separation of biomolecules in SEC does not rely on interactions between the ligand and the analyte, but rather on the diffusion and distribution of the different size molecules in the pores of the stationary phase [76,77]. Due to the dimensions of the pores, only the molecules with a similar or lower size will have access to the pores and will present a longer retention time in the stationary phase, while the larger molecules

will elute more promptly from the column, thus achieving separation between different species (Figure 7). Therefore, the separation profile of this method is the elution of the molecules in descending order of their size [77].

Regarding RNA purification, Kim and co-workers showed that SEC was able to separate the produced RNA, after *in vitro* transcription, from the reagents of the enzymatic reaction, whereas the separation of different size fragments of RNA was also accomplished [78]. In another study, Chillón and co-workers also accomplished the purification of long ncRNAs for further biophysical studies through SEC [79].

Despite its separation capacity, SEC is considered a low-resolution chromatography due to its lack of discerning between similar species and so, is mostly used in the final step or polishing step of a purification process [77].

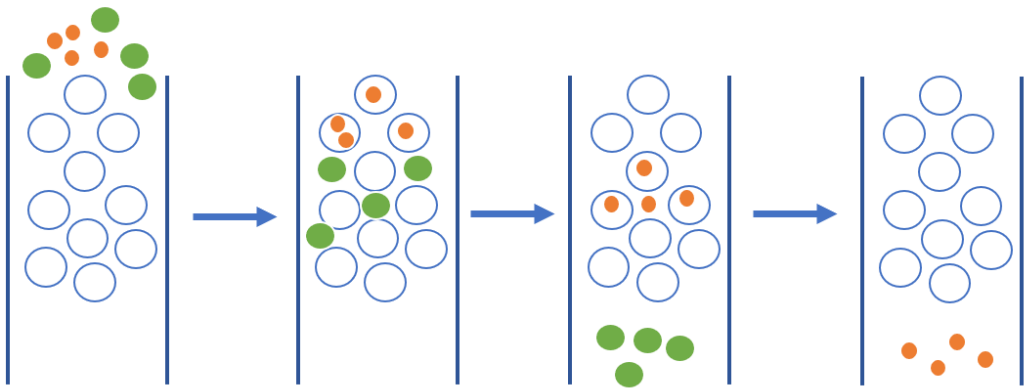


Figure 7: Schematic of a SEC separation showing the separation of smaller (orange dots) and larger (green circles) molecules (Adapted from [77]).

1.3.2 Ion-Exchange Chromatography (IEX)

Ion-exchange chromatography (IEX) is a method that allows the separation of ionizable molecules based on differences in charge properties. Distinct from SEC, in the IEX matrix, some charged ligands are immobilized, that will interact with the sample. The separation phenomena occur due to the attraction of oppositely charged molecules, as the charge of the ligands in the stationary phase is often the opposite of the analyte in the mobile phase [81]. Since a charged molecule can either present a positive or negative charge, two modes of IEX can be explored: cation-exchange chromatography and anion-exchange chromatography. The IEX mode relies on which charged molecules the immobilized ligands will interact with. For instance, if the ligands are negatively charged,

they will mostly interact with positive analytes, and so, a cation-exchange chromatography mode will be explored [82].

After the analyte adsorption, its desorption can be achieved by increasing the counterion concentration within the mobile phase, that is, if the analyte is negatively charged, the increase of anions concentration within the mobile phase will result in the elution of the analyte (Figure 8) [82].

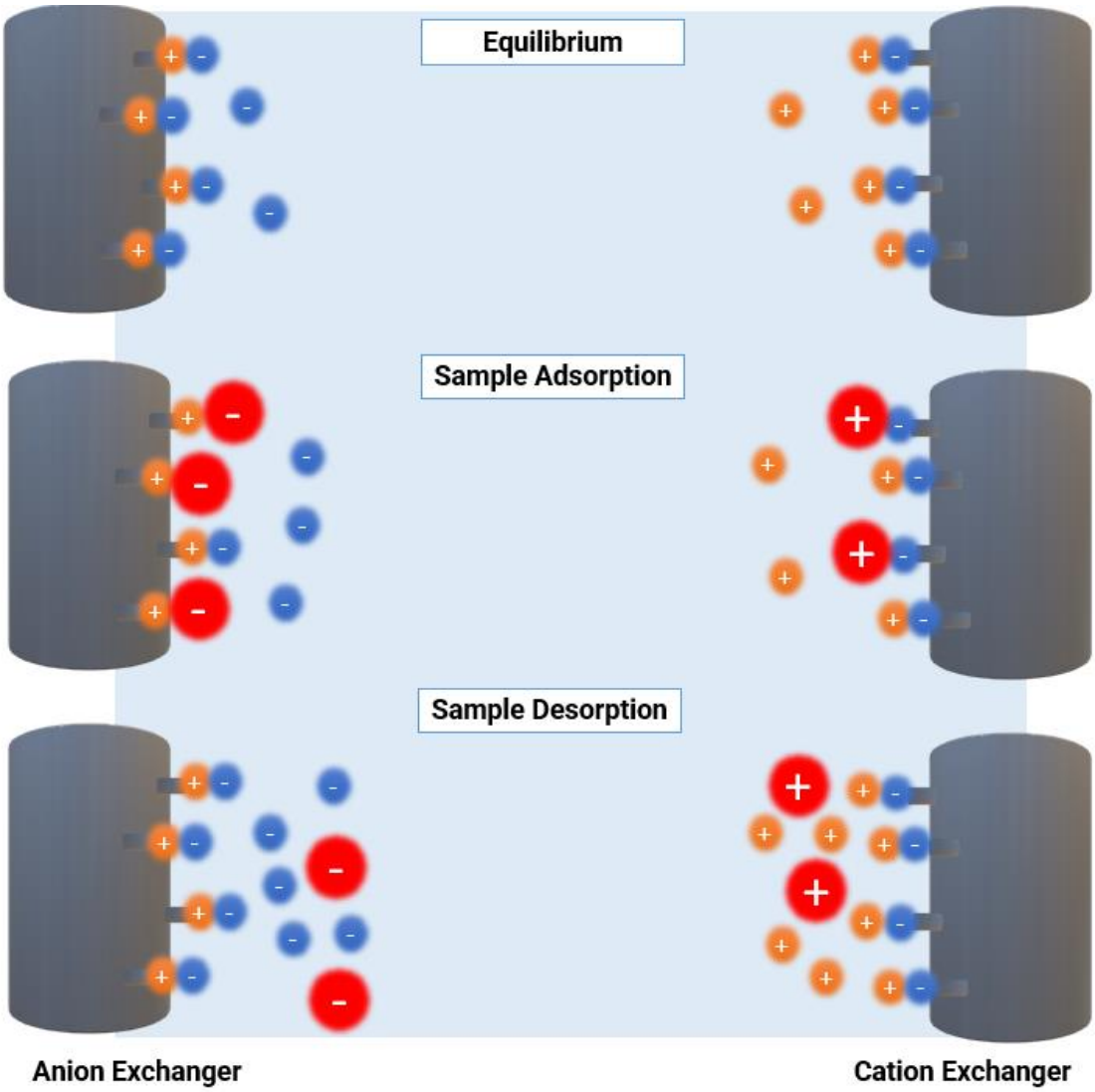


Figure 8: Representation of the chromatographic steps in both Ion-exchange chromatography modes (anion exchange and cation exchange chromatography) (Adapted from [83]).

Due to the presence of phosphate groups, nucleic acids are negatively charged molecules and so, anion-exchange chromatography is widely used for RNA purification. Easton and co-workers reported the purification of target RNA from a mixture of enzymatic reaction reagents, including T7 RNA polymerase and ribonucleotides in excess, and other RNA

fragments, by performing a weak anion-exchange chromatography using diethyl aminoethyl (DEAE) ligands immobilized in Sepharose [84]. In another study, Koubek and co-workers reported RNA purification from a similar mixture but using strong anion-exchange chromatography [85]. The difference between weak and strong IEC relies on the capacity of the ligands to keep the same charge despite the buffer pH, as strong ion exchanger can retain their selectivity and capacity over a wide pH range [82]. Another group described a method for siRNA purification using anion-exchange chromatography. The authors also believe it could be useful for manufacturing other nucleic acids such as miRNA [86]. Overall, IEX offers a reliable way for downstream steps without using denaturing conditions, with a relatively low cost. However, some drawbacks regarding IEX are that the sample loaded must be at low ionic strength, as well as small changes in pH can alter the adsorption profile of the target molecule on the ligand [82,87].

1.3.3 Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography (HIC) is a method widely used for the purification of biological compounds based on differences in their hydrophobicity [88]. It exploits the interaction between hydrophobic regions of the target and the ligands immobilized on the support. Contrary to IEX, the sample loading is performed at high to moderate salt concentrations, allowing the biomolecules to expose their hydrophobic regions during adsorption [88,89]. When two non-polar compounds are put into water, they spontaneously associate due to hydrophobic interactions. This phenomenon is explained by the increase in entropy due to the movement of water molecules that were previously around the compounds, which consequently results in a negative change in free energy and a thermodynamically favorable process [89]. After adsorption, the elution of binding species is possible by decreasing the salt in the mobile phase.

Regarding nucleic acid purification, RNA presents a more hydrophobic profile than DNA due to its aromatic moieties being more exposed, thus being possible to separate those molecules. A study reported the purification of plasmid DNA (pDNA) from a mixture of genomic DNA (gDNA) and RNA. After alkaline lysis, the gDNA that was double stranded becomes denatured, and so, it gets as hydrophobic as RNA. For this reason, gDNA and RNA have more potential to be retained in the stationary phase due to their hydrophobicity, and it is possible to purify pDNA by negative HIC [90]. Negative chromatography is an approach where the focus is to bind the impurities on the stationary phase, meanwhile, the target flows through the column, thus achieving separation [91]. A main advantage of HIC is the potential to separate biomolecules while

preserving their biological activity given that mild conditions are used, in comparison to reversed-phase chromatography [88,89]. Nevertheless, the high amount of salt associated with this method is usually seen as a disadvantage either for the cost of the process or in an environmental way [92].

In addition, no reports are available on the use of HIC for RNA purification, as RNA is usually seen as an impurity in the process. This might be due to the lack of selectivity, considering that other hydrophobic biomolecules such as proteins and endotoxins are usually present in the lysates [93].

1.3.4 Affinity Chromatography

Affinity chromatography is known to be the most selective and powerful method among all the chromatographic techniques. Distinct from the previous approaches that only rely on a single molecule property such as size, charge, or hydrophobicity, AC relies on multiple interaction types simultaneously [70,94,95]. This technique stimulates the biorecognition between the target molecule and the ligand, just as it happens with an enzyme and its substrate or with an antigen and its antibody [70,94,96]. Given this, AC allows a much more selective separation and has the potential to achieve high purity in only one step as well as reduce or even eliminate further downstream stages, which increases the process yield and reduces the costs of the process [70,94].

Different types of interactions can be involved, such as electrostatic, hydrophobic, van der Waals, dipole-dipole, π-π interactions, or hydrogen bonds, however, depending on what conditions the separation occurs, it is possible to favor particular interactions [97,98]. For instance, Pereira and co-workers reported the purification of pre-miRNA-29 by arginine-affinity chromatography using different conditions. Firstly, due to the negative charge of the RNA, the authors applied typical conditions to mainly favor ionic interactions. The chromatographic run started at low ionic strength with 280 mM NaCl, which after adsorption was increased to 360 mM NaCl to elute the target, accomplishing the recovery of 96.5% of pre-miRNA-29. Then, to mostly exploit hydrophobic interactions, the group started the process with 2M of ammonium sulfate, which resulted in total pre-miRNA-29 retention. Decreasing the salt concentration to 1.1 M promoted the target's desorption. In the end, the group accomplished the purification of the target using the same ligand (arginine) but with two different approaches, either by preferentially establishing ionic or hydrophobic interactions [98].

Regarding the elution of the target molecule, there are 2 approaches, the non-specific and specific elution [96]. In non-specific elution, solvent conditions are manipulated to

reduce the target and ligand interactions. Some manipulation on the mobile phase include the increase or decrease in the salt concentration, pH adjustment, and chemical components of buffers that may affect the interaction between the target and the ligand without compromising its biological activity [96]. The elution strategy carried out by Pereira and co-workers is an example of non-specific elution [98]. For specific elution, the mobile phase has a certain agent that will compete for either the ligand or the target, resulting in the release of the target [96]. For example, Martins and co-workers tried to elute the RNA adsorbed in arginine-affinity chromatography using arginine as a competitive agent in the mobile phase [97]. Besides arginine, other ligands have been used in affinity chromatography for RNA purification. In Table 3 are presented some examples.

Table 3. Examples of affinity ligands for RNA purification.

Ligand	Target RNA	Binding Mechanism	Reference
Oligo(dT)	Eukaryotic mRNA	Involves the base pairing between the poly(A) tail found at the 3' end of almost every eukaryotic mRNA and the thymine sequence from the ligand.	[99,100]
Histidine	rRNA and 6S RNA	Hydrophobic interactions and histidine-RNA direct H-bonds.	[101]
Boronate	RNA	Esterification of boronate with cis-diol moiety of ribonucleotide.	[102]

Despite their potential to purify RNA, some limitations are associated with these ligands. For instance, oligo (dT) chromatography allows the purification of mRNA through a base pairing mechanism with its 3'-end poly(A) tail, however, this tail can be short or even absent in some mRNAs as in bacteria [99]. In addition, other RNA types such as miRNA don't have these tails by nature, which restrains the transversal application of this strategy. On the other hand, for histidine-based AC, high amounts of salt are usually used to elute the target RNAs, which is the main drawback of this approach since it increases the cost of the process and it is not environmentally friendly [34]. Regarding boronate

affinity chromatography, the pH is a key factor for the binding mechanism, as the esterification reaction between boronate and the cis-diol group of the target molecule occurs in basic conditions and is reversible in acidic pH. Thereby, the sample needs to be manipulated in basic pH, which is a limitation not only because of the pH adjustment but also because it may increase the degradation of labile molecules [103].

So far, the affinity ligands described can be used to separate RNA from other biomolecules such as DNA and proteins, however, if different RNA sequences are present in the sample, the selectivity may not be suitable to capture only the target sequence.

Therefore, a simple approach for the purification of the target RNA can be the functionalization of the chromatographic support with oligonucleotides complementary to the target. If either the 3' or 5' end of the target sequence is unpaired, a small sequence can be designed to interact with the target in a base pairing mechanism, which results in high specificity [104]. For instance, Pereira and co-workers described a method for purifying RNA using a ssDNA capture strand that binds to the target RNA. This oligonucleotide forms a 20-base pair (bp) duplex with the target sequence, which resulted in over 95% purity [105].

1.4 Molecular Docking

Molecular docking is a computational method to predict a complex between two given molecules. Usually, one of the molecules is designated the ligand and other the receptor. These concepts were first introduced in 1978 by Janin and Wodak [106]. The ligand molecule is often a smaller molecule that is "docked" into the receptor one, which is a larger molecule such as a protein or a DNA/RNA sequence (Figure 9) [106,107].

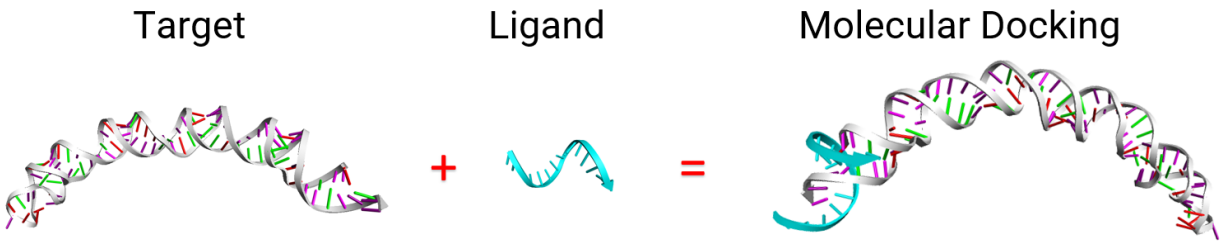


Figure 9: Main elements in molecular docking (Adapted from [107]).

Due to the ability to estimate the binding affinity and the interactions established within the complex, molecular docking has been used in various fields such as molecular biology and drug discovery [108]. Particularly, it has been used to perform virtual screening of a

ligand library in order to highlight the ones that interact the most with the target and/or best fit in the binding pocket, which is appropriate when finding drug candidates or affinity ligands, significantly reducing the experimental costs [109].

Docking can be divided into two stages. In the first stage, a binding function is applied to generate ligand-receptor poses by searching for ligand positions around the target, while in the second stage, a scoring function is used to evaluate and rank the poses obtained in the first stage. The ranks are usually shown in descending order of binding affinity scores [110].

1.4.1 Protein Vs Nucleic Acids Docking

Since the beginning of molecular docking, several functions and methods have been developed over the years, however, the majority are designed for protein-ligand interactions, whereas for nucleic acids fewer software are available [111]. This may be because only recently the scientific community has begun to focus on nucleic acids as therapeutic targets. Even if some protein-ligand programs can be used for nucleic acid docking, they lack accuracy due to structural differences between proteins and nucleic acids, which compromises the studies viability [112]. Given that nucleic acids are negatively charged due to the phosphate groups in the backbone, they will attract metal ions and water molecules, which consequently affects the interactions between the ligand and the receptor. These types of effects may be miscalculated by protein-based functions resulting in inaccurate binding scores or erroneous ligand positions. In addition, nucleic acids present special flexibility and may change their initial conformation after the ligand binding, which usually is not considered by traditional docking methods [112-114]. Therefore, appropriate functions to deal with the solvation effect and flexibility are in development to increase the accuracy of nucleic acids docking.

1.4.2 Conformational Searching Functions

Regarding the first step in molecular docking, there are up to 4 different search functions categories: shape matching, systematic search, random search, and simulation methods [115].

The shape matching method is one of the simplest algorithms and allows a rapid generation of ligand-binding modes by matching the geometric features between the ligand and the target's binding pocket. For instance, Kuntz and co-workers developed a method where both ligand and binding site were represented as spheres. Then, the algorithm finds a good match whenever the ligand set of spheres fits in the receptor set

of spheres. However, the number of poses generated is limited since only the translational and rotational degrees of freedom of the ligand are considered [115,116].
On the other hand, systematic search methods are more complex and are usually associated with high computational costs, given that their algorithm tries to exhaustively explore the pocket space for ligand positions. This method allows the generation of all possible ligand binding conformations since it explores all degrees of freedom of the ligand [115,117,118].
Another method is the random search approach, wherein ligand binding orientations and conformations are sampled by introducing random changes to the ligand molecule. These changes are performed at each step in both the conformational space, that is, the possible shapes or conformations the ligand can adopt, and the translational/rotational space of the ligand, which are the possible positions and orientations that the ligand can occupy relative to the receptor site. These changes are then evaluated with a predefined probability function and the favorable ones are accepted [117,118].
Another way to derive ligand binding conformations is by simulation methods such as molecular dynamic simulations. In this approach, the complex trajectory is simulated over time using Newtonian mechanics, and the forces are calculated on each atom from the small change in potential energy between the current position and the previous one. However, due to the high computational cost and low-efficiency nature, this method is rarely used by itself to sample ligand receptor poses [117].

1.4.3 Scoring Functions

Currently, scoring functions for nucleic acid-ligand interactions can also be distributed into four types: force field-based, empirical, knowledge-based, and machine learning (ML).
Force field scoring functions evaluate the interaction energy from each pose as the sum of diverse terms such as van der Waals interactions, electrostatic interactions, and hydrogen bonding [115,117]. For instance, the equation for the force field scoring function implemented on DOCK is described as follows:

$$E = \sum_{ij}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}}\right) \qquad (1)$$

where r_{ij} represents the distance between receptor atom i and ligand atom j. A_{ij} and B_{ij} are the van der Waals parameters, and the third term stands for the electrostatic interaction between two charged atoms. This type of function has the advantage of being based on well-understood physical principles and concepts, however, it may require more computational and time efforts than other methods [115,117].

Empirical scoring functions calculate the binding energy by adding up several energy terms such as van der Waals energy, electrostatics, hydrogen bonding, desolvation, entropy, and hydrophobicity.

The empirical scoring function can be described as follows:

$$\Delta G = \sum_i w_i * \Delta G_i \qquad (2)$$

where ΔG_i represents the different energy terms, and w_i is the corresponding coefficient, which is obtained by fitting the binding scores with experimental binding affinity data for a given training set of receptor-ligand structures. Although this function is fast and easy to use, the dependence on real experimental data is a disadvantage due to the limited number of nucleic acid-ligand complexes [115,117].

Compared to the force field and empirical scoring functions, knowledge-based functions have a good balance between accuracy and speed. These functions derive from statistical observations of interactions in experimental structures complexes. The interaction potential is given by the equation:

$$u_{ij}(r) = -k_B T ln[\frac{p_{ij(r)}}{p^*_{ij}(r)}] \qquad (3)$$

where k_B is the Boltzmann constant, $p_{ij(r)}$ is the observed density of atom pair i and j at distance r, and $p^*_{ij}(r)$ is the corresponding atom pair density in a reference state in which the interatomic interactions are zero. In addition, since these functions are derived from a large and diverse structural database, there will be fewer training set dependencies compared with empirical functions [115,117].

The last type of scoring functions is machine learning-based functions, which represent the most recent advancements in artificial intelligence over the years. Distinct from the previous scoring function types, machine learning does not rely on specific rules or functions and is able to capture nonlinearities in the training set. However, artificial intelligence is known for requiring large amounts of data and has a high dependency on the training set, which combined with the limited nucleic acid-ligand structures, makes the application of machine learning uncommon [115].

1.4.4 Nucleic Acid-Ligand Docking Programs

Despite the different types of functions in both stages of docking, current docking programs often combine more than one strategy in order to improve performance [115]. So far, different programs can be used in nucleic acid-ligand docking, however, where some of them have been specifically designed for that purpose, there are other programs originally created for protein-ligand docking that can be also applied to nucleic acids or got some optimization in a way that can treat protein and nucleic acid molecules. As

mentioned above, despite the versatility of some programs to be used for both molecules, it is expected that programs developed for nucleic acids will have more accuracy. In addition, depending on the ligand library size, some of the programs may not be appropriate to be used for virtual screening. For example, MOlecular Recognition with a Driven dynamics OptimizeR (MORDOR) uses molecular simulation techniques that take account of the nucleic acid and ligand flexibility, which increases the prediction accuracy, however, such an approach is computationally extensive and so, its usage in virtual screening might be limited to a small ligands group [119]. In Table 4 are discriminated various programs that can be used in nucleic acid-ligand docking.

Table 4. Summary of docking programs that can be used for nucleic acid-ligand docking.

Docking Program	Development aim	References
MORDOR	Developed for Nucleic acids	[120]
rDock	Developed for Nucleic acids	[121]
RLDOCK	Developed for Nucleic acids	122]
NLDOCK	Developed for Nucleic acids	[123]
HNADOCK	Developed for Nucleic acids	[124]
DOCK 6	Adapted for Nucleic acids	[125]
FITTED	Adapted for Nucleic acids	[126]
CDOCKER	Adapted for Nucleic acids	[127]
AutoDock	Can be used/Adapted for Nucleic acids	[128]
AutoDock Vina	Can be used for Nucleic acids	[129]
GOLD	Can be used for Nucleic acids	[130]
Glide	Can be used for Nucleic acids	[131]
ICM	Can be used for Nucleic acids	[132]
Surflex	Can be used for Nucleic acids	[133]

1.5 Molecular Dynamic Simulation

Molecular dynamic (MD) simulation is another *in silico* tool well established and helpful to gain insight into how biomolecules work. It predicts how every atom in a biological system will move over time based on general physic properties regarding interatomic interactions [134,135]. Through these simulations, it is possible to retrieve different information about the system, including conformational change, ligand binding, and

possible perturbations which then allows the understanding of its stability over time. The forces acting on each atom are estimated through the equation presented in Figure 10. Similar to the molecular docking, the force field function implemented in MD also uses non-bonded forces such as van der Waals and electrostatic interactions to estimate the atomic forces that lead to movement. However, it also considers bonded force properties such as bond length, the angle between two subjacent bonds, and torsional angles. Then, once the forces acting on each atom have been calculated, their tridimensional positions are updated according to Newton's laws of motion. After the update, the simulation time is increased typically by only 1 or 2 quadrillionths of a second, and the process is then repeated millions of times. Such number of calculations significantly increases the complexity and computational demands of the entire process [134,135].
Consequently, in screening studies, MD simulations are usually performed after molecular docking. This approach allows for the selective testing of only the most promising candidates from the initial set, leading to a significant reduction in the overall time required for analysis [134,135]. Common force fields used in MD are AMBER [136], CHARMM [137], GROMOS [138], and NAMD [139].

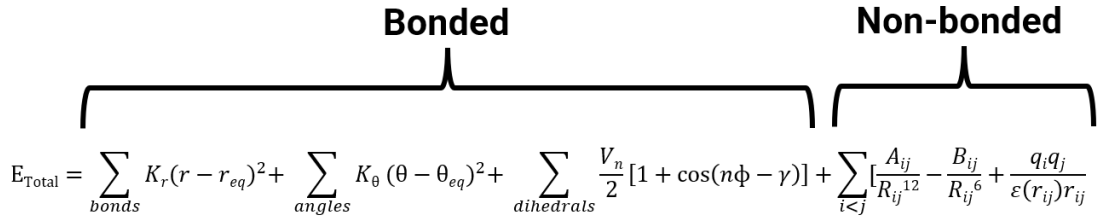


Figure 10: An example of a Force Field function used in Molecular Dynamics. (Adapted from [134])

Although NMR and X-ray Crystallography produce a static model of the biological system that provides valuable insights into structures, biological phenomena such as molecular recognition and drug binding constitute dynamic processes. For instance, when a ligand meets its target in solution, it encounters a macromolecule in constant motion instead of a frozen structure. Therefore, MD simulations play a fundamental role in providing complementary information [134,135].

These *in silico* tools can be valuable for the initial screening of ligands. They provide crucial information about how the candidates interact with the target, giving insight into the stability of each complex. This data is beneficial for the selection of affinity ligands that can be further experimentally analyzed for the purification of target molecules, such as pre-miRNA-29b.

Chapter 2 – Global aims

Over the last years, nucleic acids have been receiving considerable attention due to their roles in cell pathways but also for their ability to regulate gene expression. These new tools have the potential to provide novel treatments for various diseases such as cancer and Alzheimer's disease, which is on the basis of the recent interest from the biopharmaceutical industry. With biotechnology, the production of these biopharmaceuticals is relatively easy and inexpensive, however, the subsequent steps of separation significantly increase the costs of the process. The main purification technique used in the biopharmaceutical industry is chromatography, which is a well-known and established method, but with many gaps that need to be improved. Despite being the most powerful method, affinity chromatography remains considerably expensive and requires different ligands for different targets, whose identification can be very challenging. With the world requiring more environmental care, *in silico* methods such as molecular docking represent a novel approach to conduct scientific research without wasting valuables resources.

Therefore, the aim of this work is to develop new affinity ligands for the specific purification of pre-miRNA-29b, that may present beneficial effects in Alzheimer's disease. The ligands are constituted by oligonucleotides sequences designed to be complementary to the target and will be screened using molecular docking. After this initial screening, the Molecular Dynamic studies will be performed with the most promising ligands. Then, the selected oligonucleotides will be immobilized onto a Sepharose support and evaluated regarding their potential to capture and purify the RNA target. Different conditions of binding and elution will also be tested in a way to understand which interactions are mostly present during the adsorption.

In general, the aim of this work is to design and evaluate new affinity ligands for the purification of potential biopharmaceuticals, with a central focus of developing highly specific purification strategies.

Chapter 3 – Materials and Methods

3.1 Materials

For the activation of Sepharose, it was used Sepharose CL-6B with a particle size of 45-165 μm from Cytiva (Sweden), sodium hydroxide (NaOH) and acetone from LabChem (Portugal), sodium borohydride ($NaBH_4$) from Acros Organics (Waltham, USA), 1,4-butanediol diglycidyl ether ($C_{10}H_{18}O_4$) from Sigma-Aldrich (St. Louis, Missouri, USA) and sodium bicarbonate ($NaHCO_3$) from Fisher Scientific Inc. (Waltham, USA). The 4 selected oligonucleotides were purchased from Eurogentec (Belgium), all with an amino modifier in the 5'-end.

For the culture of bacterial cells, the reagents used were tryptone and yeast extract from Bioakar (Beauvais, France), glycerol from Himedia, dipotassium hydrogen phosphate (K_2HPO_4) from Panreac (Barcelona, Spain), potassium dihydrogen phosphate (KH_2PO_4) from Sigma-Aldrich (St. Louis, Missouri, USA), "Luria-Broth Agar" from Pronalab (Mérida, Yucatán, Mexico) and kanamycin from Thermo Fisher Scientific Inc. (Waltham, USA).

For the extraction of nucleic acids, the reagents used were guanidine thiocyanate, N-Lauroylsarcosine sodium salt, sodium citrate and isoamyl alcohol all from Sigma-Aldrich (St. Louis, Missouri, USA), isopropanol from Thermo Fisher Scientific Inc. (Waltham, USA), and β-mercaptoethanol from Merck (Whitehouse Station, USA).

For the oligonucleotide immobilization protocol it was used 1,4-butanediol diglycidyl ether from Sigma-Aldrich (St. Louis, Missouri, USA). In the chromatographic assays it was used sodium chloride (NaCl) and ammonium sulphate ($(NH_4)_2SO_4$) both commercialized by Panreac (Barcelona, Spain), tris(hydroxymethyl)aminomethane (Tris) from Merck (Darmstadt, Germany), monosodium phosphate (NaH_2PO_4) from CARLO ERBA (France), disodium phosphate (Na_2HPO_4) from Merck (Darmstadt, Germany), and ethylenediaminetetraacetic acid (EDTA) from LABKEM (Spain). All solutions were prepared with Milli-Q water treated with 0.01% of diethyl pyrocarbonate (DEPC) from Sigma-Aldrich (St. Louis, Missouri, USA).

3.2 Methods

3.2.1 Molecular Docking

Firstly, the pre-miRNA-29b sequence was inputted in the ViennaRNA Web services [140] to obtain information about the minimum free energy and the code of the structure.

Then, the code was inputted in the RNAComposer web server [141,142] to obtain the pdb file of the pre-miRNA-29b. Meanwhile, all the 26 oligonucleotides to be tested were constructed in the BIOVIA Discovery Studio [143], this includes the attachment of the carbon chain and the amine group. Then, all the 26 oligonucleotides were opened in the ChemDraw3D program (PerkinElmer Informatics), and their initial conformation was altered for a minimum energy conformation using the MM2 protocol. Finally, the pre-miRNA-29b was inputted as receptor and each oligonucleotide was inputted as ligand in the HNADOCK web server [144], that performed molecular docking and generated the top 10 models and respective docking score for each oligonucleotide.

3.2.2 Molecular Dynamics

To perform molecular dynamic simulations, selected models belonging to the best ligands screened by molecular docking were inputted in the MDWeb server [145]. The dynamic method selected was NAMD and the total simulation time was configured to be 500 ps and a temperature of 300 K. After the simulation, some operations within the platform can be performed to analyze the results. For this work, the B-factor per residue, Root Mean Square Deviation (RMSD) per residue, trajectory RMSD and MD trajectory were analyzed.

3.2.3 Ligand Immobilization in Sepharose

For the functionalization of Sepharose, three steps are required. First, 10 mL of Sepharose CL-6B were filtered and washed with approximately 500 mL of Milli-Q water. After decanting the excess water, the next step involved the crosslinking of the 1,4-butanediol diglycidyl ether with Sepharose and activation of its epoxy group. The washed Sepharose was incubated with 10 mL of solution 1 (1.32 mM $NaBH_4$ and 0.6 M NaOH) under agitation at 300 rpm, during 15 min at room temperature. Then, 16.6 mL of 1,4-butanediol diglycidyl ether were added and the mixture was kept under agitation at 300 rpm, for 6 h at room temperature. After this, the mixture was filtered and washed with approximately 500 mL of Milli-Q water. In the second step, which involves the immobilization of the ligands onto the Sepharose, 10 mL of solution 2 (containing 1.24 μM of the specific oligonucleotide and 2 M $NaHCO_3$) was added to the previous solution and agitated during 16 h at 300 rpm and 55 °C. In the third step, which includes the washing of functionalized Sepharose to remove the oligonucleotides that were not immobilized, the mixture was filtered and washed multiple times using different ratios

of Milli-Q water and acetone, with the proportion of acetone increasing at each wash. In the end, the product was washed and stored at 4 °C in Milli-Q water, or 20% ethanol for longer periods. The entire process was repeated for each type of ligand, as well as a control without any ligand. This process is summarized in Figure 11.

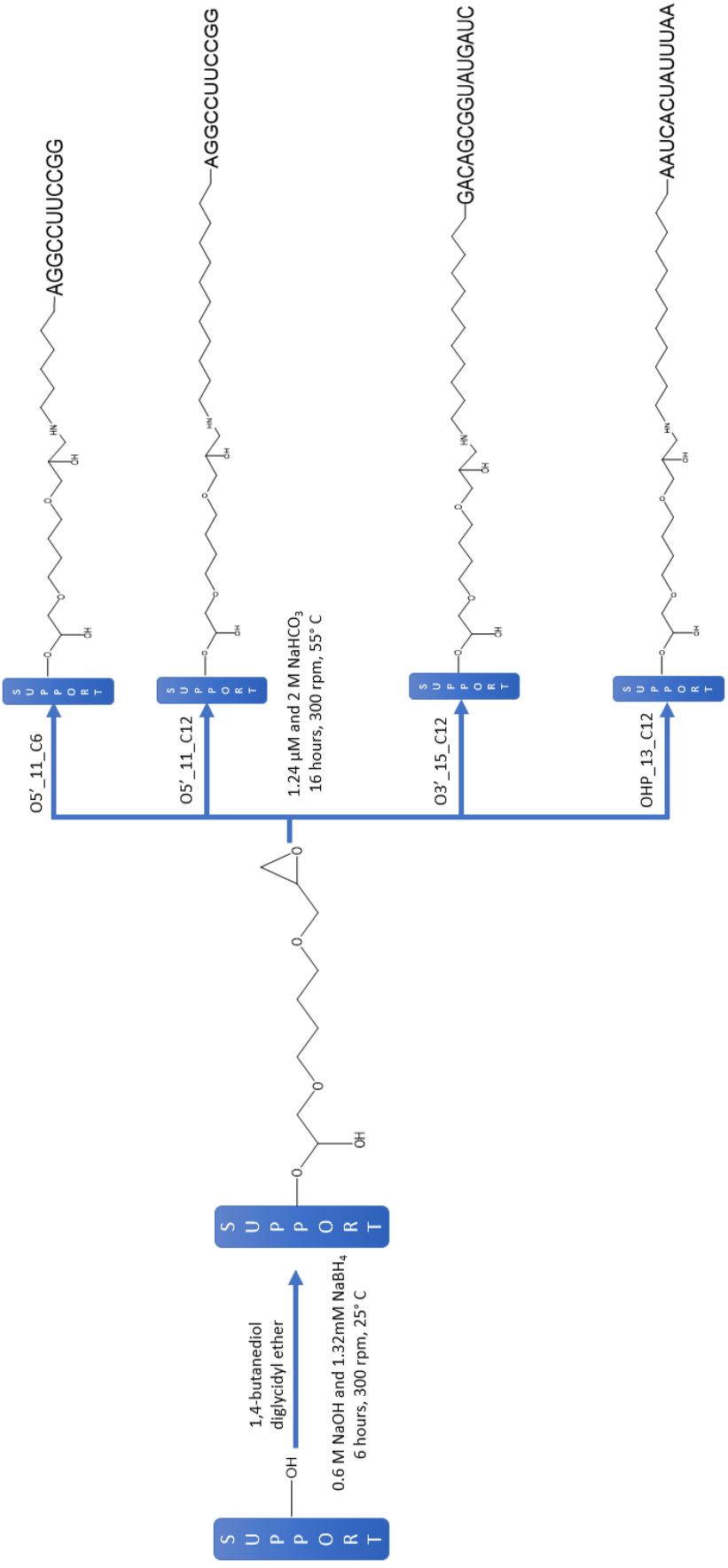


Figure 11: Scheme of the reactions conducted for the functionalization of the sepharose supports with oligonucleotides.

3.2.4 Support Characterization

All the synthetized supports were analyzed by elemental analysis in order to check if the ligand was immobilized onto the Sepharose matrix. Elemental analysis was conducted using the TruSpec 630-200-200 equipment, with a sample size of approximately 2 mg. Therefore, the carbon and hydrogen percentages were determined through infrared radiation absorption after combustion of the sample, while the nitrogen content was quantified by thermal conductivity.

3.2.5 Nucleic Acids Production in *Escherichia coli* DH5α

The production of pre-miRNA-29b was performed using a strain of *Escherichia coli* DH5α (*E. coli*) that was previously transformed with the plasmid pBHSR1-RM. In the first step, *E. coli* was inoculated in a solid growth medium "Luria-Broth Agar" (LB-Agar) with 50 μg/mL of kanamycin. The bacteria was allowed to growth overnight at 37 °C. The pre-fermentation and fermentation were conducted in "Terrific Broth" medium (TB) which has 12 g/L of tryptone, 24 g/L of yeast extract, 5.5×10^{-5} M glycerol, 0.017 M KH_2PO_4 and 0.072 M K_2HPO_4. Then, *E. coli* was transferred from solid growth medium to pre-fermentation and incubated at 37 °C on an orbital shaker operating at 250 rpm. During incubation, the optical density (OD) was measured at 600 nm using the spectrophotometer Pharmacia Biotech Ultraspec 3000 UV/Visible equipment at regular intervals until it reached a value of 2.6. Then, to ensure that the fermentation could start at an OD of 0.2, the required volume of pre-fermentation was calculated using the following equation:

$$V_{to\ tranfer\ from\ pre-fermentation} = \frac{(V_{to\ transfer} + V_{fermentation}) * OD_{fermentation}}{OD_{pre-fermentation}} \quad (4)$$

After the transference of the pre-fermentation volume to the fermentation vessel, this was incubated at 37 °C on an orbital shaker set to 250 rpm for 8 h. At the end, the medium was centrifuged at 3900 g and 4 °C for 10 min and the resultant cell pellets were stored at -20 °C. In Figure 12 it is summarized the general procedure.

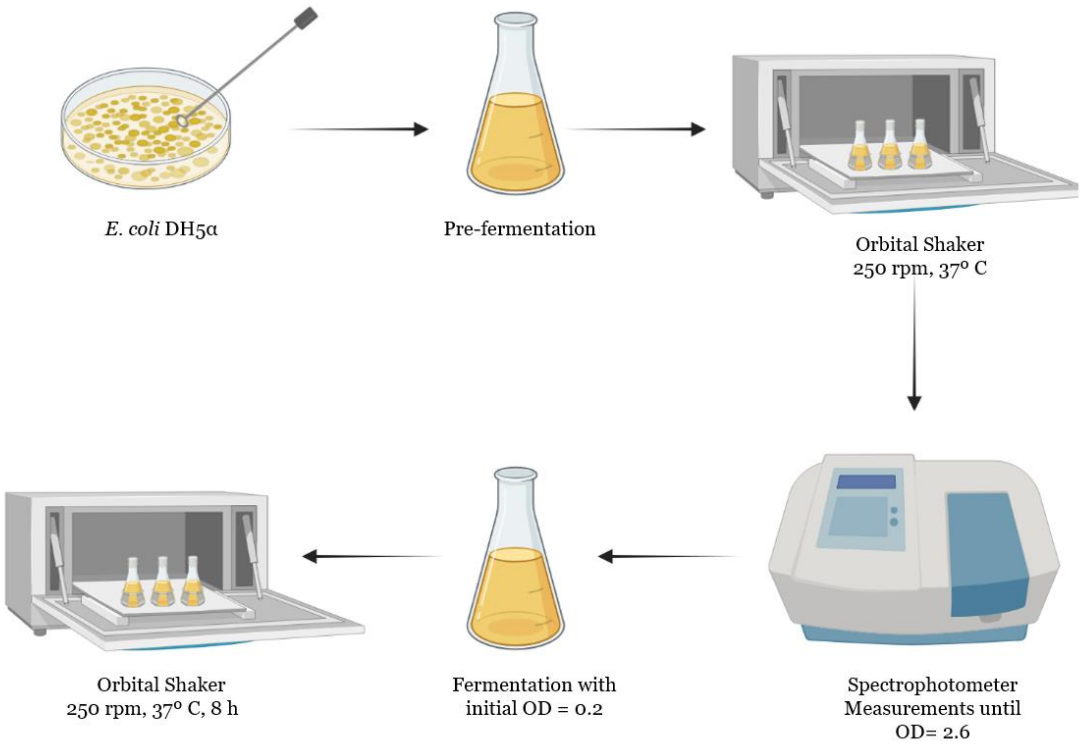


Figure 12: General process for *E. coli* culture to produce pre-miRNA-29b.

3.2.6 RNA Extraction

RNA extraction was conducted through the method of acid guanidium thiocyanate-phenol-chloroform. Initially, the *E. coli* pellets that had been produced and stored at -20 °C were thawed and resuspended in 0.8% NaCl. Subsequently, centrifugation was carried out at 6000 g for 10 min at 4 °C. The resulting supernatant was discarded, and the pellets were then resuspended in 5 mL of solution D (4 M guanidinium thiocyanate, 0.025 M sodium citrate and pH 7, 0.5% sodium N-lauroylsarcosinate and 0.1 M β-mercaptoethanol) and incubated on ice for 10 min. After incubation, 0.5 mL of 2 M sodium acetate, pH 4 and phenol were added to the mixture, which was carefully homogenized. Subsequently, 1 mL of a solution containing chloroform and isoamyl alcohol in a ratio of 49:1 was added to the mixture. The mixture was vigorously shaken and then incubated on ice for 15 min. After incubation, the mixture was centrifuged at 10000 g for 20 min at 4 °C. Two aqueous phases were formed after centrifugation: the upper phase, which is rich in RNA, and the bottom phase, which contains DNA. To prevent DNA contamination, the upper phase was carefully collected and transferred to a new tube. Then, 5 mL of isopropanol is added to precipitate the RNA, and the mixture is centrifuged at 10000 g for 20 min at 4 °C. Once again, the resulting supernatant was discarded, and 1.5 mL of solution D and 1.5 mL of isopropanol were added to the RNA

pellets. Then, the mixture was carefully homogenized and centrifuged at 10000 g for 10 min at 4 °C. The supernatant was discarded, the pellets were resuspended in 2.5 mL of 75% ethanol in DEPC-treated water and then incubated at room temperature for 10-15 min. After incubation, the mixture was centrifuged at 10000 g for 5 min at 4 °C. Supernatant was discarded, and the pellets were air-dried for 5-10 min at room temperature. Then, the pellets were resuspended in 1 mL of DEPC-treated water and incubated for 10-15 min at room temperature. The concentration of RNA was measured in the Nano Photometer (IMPLEN, United Kingdom). The integrity and the purity of RNA was verified by agarose gel electrophoresis, and samples were stored at -80 °C.

3.2.7 Agarose Gel Electrophoresis

Visualization of RNA integrity and purity was conducted by horizontal electrophoresis in 1 % agarose gel. The gel was prepared with 0.012 µL/mL of Green Safe, which can intercalate with nucleic acids and allow the visualization under ultraviolet light (UV). Electrophoresis run was performed at 120 V for 30 min in TAE buffer (40 mM Tris base, 20 mM acetic acid, and 1 mM EDTA and pH 8). After the run, the gel was revealed under UV light using the Uvitec Cambridge Fire-Reader equipped with a UV chamber (UVITEC Cambridge, Cambridge, United Kingdom).

3.2.8 Chromatographic Assays

To gain insight into the RNA binding and elution profile to each functionalized support, the synthetized materials were packed into empty columns for subsequent screening assays. In this way, each support was dispersed in Milli-Q water and transferred into the empty column until it reached a height of 3 cm. Throughout the process, a constant flow of running water was maintained. At the end of each assay, the column was washed multiple times with Milli-Q water.

Chromatographic assays where hydrophobic interactions were mainly favored:

Firstly, the column is washed multiple times with DEPC-treated water. Then, 15 mL of binding buffer (2.5 M $(NH_4)_2SO_4$, 10 mM Tris-HCl and pH 8) is added to the column as the equilibration step. After the equilibrium is established, 50 µg of RNA sample were loaded onto the column. Then, 10 mL of the previous buffer were applied to the column to promote binding of the sample and washing out the non-bound species. 1 mL-fractions were collected for subsequent analysis. Elution step was performed by applying 10 mL of elution buffer (10 mM Tris-HCl and pH 8) to the column and the fractions were also collected for analysis. After that, 10 mL of DEPC-treated water was added to the column and collected for analysis. In the end, the column was washed multiple times with DEPC-

treated water. The fractions absorbance was measured in the Nano Photometer (IMPLEN, United Kingdom).

Chapter 4 - Results and Discussion

4.1 Molecular Docking Studies

Aiming at identifying specific ligands for the purification of pre-miRNA-29b, a series of oligonucleotides were designed taking into consideration the pre-miRNA-29b sequence for base pairing interaction mechanism. The original sequence of pre-miRNA-29b is illustrated in figure 13. However, for recombinant production purposes, the plasmid was constructed in a way that the target gene sequence is found between two ribozyme sequences. Therefore, after plasmid transcription in bacteria cells, the ribozymes will induce autocleavage and release the target pre-miRNA-29b in the cell. This process adds a small number of bases to both 5' and 3' ends, which must be considered for the design of the complementary oligonucleotides' ligands. The additional bases are displayed in figure 14.

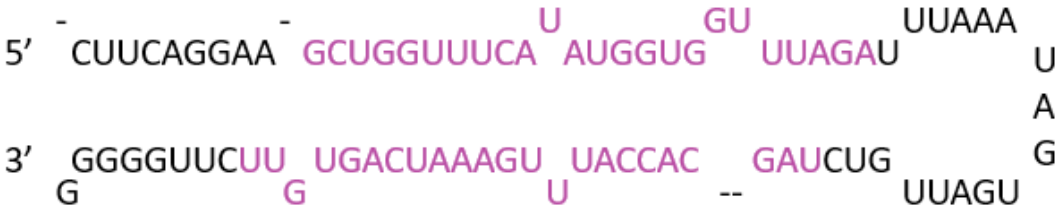


Figure 13: Pre-miRNA-29b sequence. In purple are marked the mature sequences of miRNA-29b.

Figure 14: The additional sequences present on both 5' and 3' ends of the pre-miRNA-29b.

Therefore, the ligands were designed to be complementary either to the hairpin structure of pre-miRNA-29b or to the additional sequences from the ribozymes. In addition, different lengths of oligonucleotides were used to evaluate the effect of the size of the oligo on the interaction established with the pre-miRNA-29b. The nomenclature of each oligonucleotide derives from its complementary site, the number of bases and the length of the carbon chain. For instance, the oligonucleotide O3'_15_C12 is complementary to the 3' region of the target, it has 15 bases, and the carbon chain is 12 carbon long. In table 5 are displayed some examples of the oligos under study.

Table 5: Nomenclature, sequence, binding target, number of nucleotides, and carbon chain length of each selected oligonucleotide.

Name	Sequence 5'-3'	Target	Nº Bases	Carbon Chain Length
O3'_15_C12	GACAGCGGUAUGAUC	3' end	15	12
O5'_11_C6	AGGCCUUCCGG	5' end	11	6
O5'_11_C12	AGGCCUUCCGG	5' end	11	12
OHP_13_C12	AAUCACUAUUUAA	Hairpin	13	12

Then, the full sequence of the target (pre-miRNA-29b flanked by the ribozyme sequences) was inputted in the ViennaRNA to obtain information regarding the structure and then in the RNAComposer to obtain the receptor file required for docking studies. The structure obtained from ViennaRNA is displayed in the figure 15.

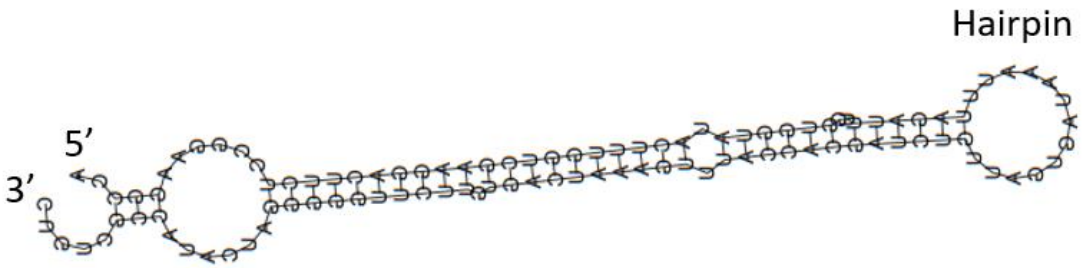


Figure 15: Target Sequence of pre-miRNA-29b flanked by the ribozyme sequences obtained from ViennaRNA.

Then, all designed ligands got their conformation modified for the minimum energy conformation using the MM2 protocol. Figure 16 illustrates the difference between the initial and minimum energy conformation of the ligand O5'_11_C12 as an example.

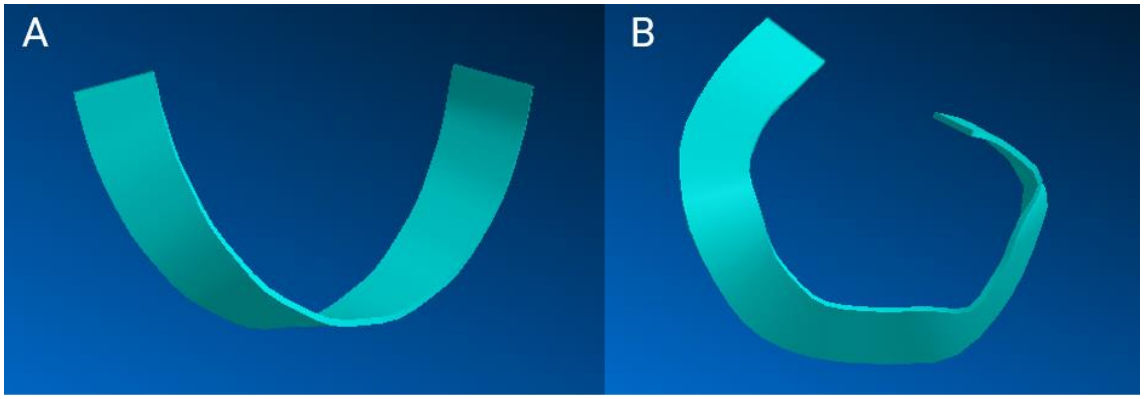


Figure 16: Conformations of O5'_11_C12 before and after energy minimization protocol MM2 was applied. A- Initial conformation; B- Minimum energy conformation

To evaluate the affinity of the novel ligands towards the nucleic acid target, the pre-miRNA-29b, virtual screening was performed using molecular docking. This process was conducted in the HNADOCK server, which outputted the top 10 complex models between the target and each ligand, sorted by affinity scores. However, even with the complementary region of the ligands for the target, some models outputted by the docking program show the ligand interacting in different parts of the receptor, rather than only in the matching region. Therefore, every model of each ligand was analyzed in a way to understand which affinity score is referred to a pose that has the ligand interacting in the complementary region. In addition, since the ligands were designed to interact with the pre-miRNA-29b in a base pair mechanism, only the affinity scores of poses involving the complementary regions were used to compare the ligands. The resulting docking scores are displayed in Table 6.

Table 6: Docking scores and the number of interactions of each ligand with the target, pre-miRNA-29b.

Ligand	**C6 Version**		**C12 Version**	
	Score	**Interactions**	**Score**	**Interactions**
O5'_10	-292.59	12	-318.56	10
O5'_11	-327.03	19	-322.27	12
O5'_12	-314.22	18	-258.12	12
O3'_10	-289.12	9	-289.87	10
O3'_11	-337.69	16	-329.06	19
O3'_12	-309.78	9	-306.02	10
O3'_13	-286.54	18	-295.47	9
O3'_14	-276.04	11	-319.49	13
O3'_15	-317.32	15	-313.30	15
OHP_10	-299.04	9	-276.53	21
OHP_11	-284.36	23	-249.25	18
OHP_12	-269.48	17	-276.66	9
OHP_13	-280.04	9	-279.93	14

Commonly in docking studies, low values of docking scores are associated with high affinity values between the receptor and the ligand. It should be mentioned that these docking scores do not represent the real binding affinities, but a relative ranking between models, as the scoring function used by the HNADOCK docking program, DITScoreRR, was not adjusted with experimental binding data [124]. In general, the ligands that were designed to interact in the 5'/3' region exhibit better scores than ligands projected to associate with the hairpin structure.

Furthermore, when the group of models was analyzed for each ligand, it was noticed that some ligands present a model in the complementary region, but also in the region intended for other designed ligands. For instance, the ligand O3'_15 that was originally

designed to interact in the 3' region of the target, presents models for that specific region but also in the hairpin structure (Figure 17).

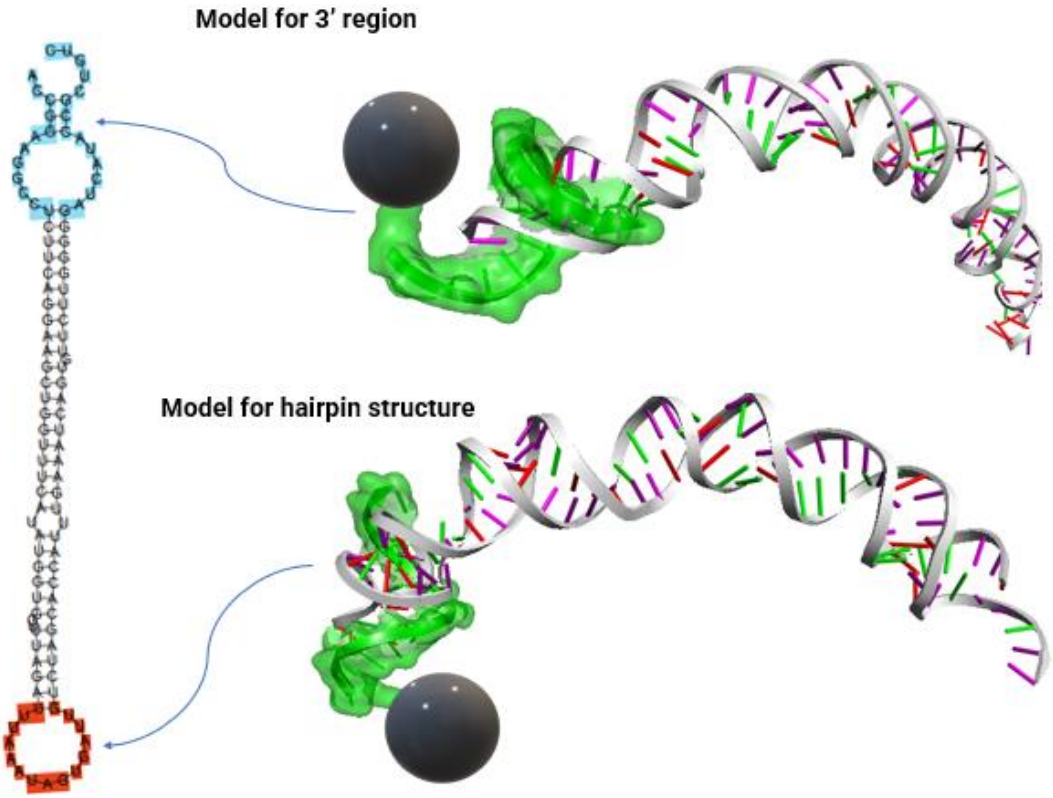


Figure 17: Different models outputted by the HNADOCK program, where in one model the ligand O3'_15_12C is interacting in the complementary region (3') and the other interacting in the hairpin structure.

This means that some ligands can be more versatile than others, which could be advantageous since the ligand can interact with multiple regions of the target being more likely to capture it. In the Table 7 it is summarized the information regarding the areas where each ligand can interact, if only for the region they were designed, or for two different regions.

Table 7: Presence of at least one model in a binding site.

Ligand	C6 Version	C12 Version
O5'_10	✓ ✓	✓
O5'_11	✓	✓
O5'_12	✓	✓ ✓
O3'_10	✓	✓ ✓
O3'_11	✓	✓ ✓
O3'_12	✓	✓ ✓
O3'_13	✓ ✓	✓ ✓
O3'_14	✓ ✓	✓ ✓
O3'_15	✓ ✓	✓ ✓
OHP_10	✓	✓
OHP_11	✓	✓ ✓
OHP_12	✓	✓
OHP_13	✓ ✓	✓ ✓

✓ At least one model for 5'/3' region ✓ At least one model for the hairpin structure

It was found that all the designed ligands presented more interactions for the 5'/3' region than the hairpin structure, even those that were designed to base pair with the sequence present in the hairpin. This might be due to the 3D structure of the target, which seems to have more accessibility to interact with ligands on the 5'/3' region than the hairpin structure, once the bases are more exposed. In addition, the same ligand with a longer carbon chain is more likely to interact in another region besides the complementary part, as the versions with C12 present more models for two different regions, comparatively to the C6 versions. This is probably due to the higher flexibility achieved for the ligand with the longer carbon chain, resulting in an increased ability to adjust and interact with the pre-miRNA-29b. Through the analysis of Table 7, we conclude that the ligands O3'_13, O3'_14, O3'_15, and OHP_13 are the most versatile, since both versions of shorter and longer carbon chains present models in both binding sites.

Therefore, in order to understand the versatility of these ligands, a careful look was taken over their sequences considering that they could have common complementary sequences in the 3' region and the hairpin. Indeed, it was found that there are small subsequences in common between the two groups of oligonucleotides (Figure 18).

O3'_13 Sequence:

5' GACAGCGGUAUGA 3'

O3'_14 Sequence:

5' GACAGCGGUAUGAU 3'

O3'_15 Sequence:

5' GACAGCGGUAUGAUC 3'

OHP_13 Sequence:

5' AAUCACUAUUUAA 3'

Figure 18: Sequence of nucleotides of the four more versatile ligands. The common sequences are highlighted in yellow.

However, these small subsequences are also present in other ligands such as O3'_11, O3'_12, OHP_10, OHP_11, and OHP_12 (Figure 19), and these did not show the same versatility as the prior group.

O3'_11 Sequence:

5' GACAGCGGUAU 3'

OHP_11 Sequence:

5' AAUCACUAUUU 3'

OHP_10 Sequence:

5' AAUCACUAUU 3'

O3'_12 Sequence:

5' GACAGCGGUAUG 3'

OHP_12 Sequence:

5' AAUCACUAUUUA 3'

Figure 19: Sequence of nucleotides of ligands that also have the same subsequences in common. The common sequences are highlighted in yellow.

After concluding that the versatility was not achieved particularly by the sequence, the number of nucleotides was investigated, considering that the groups of ligands intended to interact in the 3' end or the hairpin structure were the longer ones. For this purpose, new ligands were designed to have one or two more nucleotides than the previous ones and tested with molecular docking following the same methodology. The added nucleotides are the complementary bases in the pre-miRNA-29b (Figure 20).

O3'_16 Sequence:

5' GACAGCGGUAUGAUCC 3'

OHP_14Sequence:

5' AAUCACUAUUUAAA 3'

O3'_17 Sequence:

5' GACAGCGGUAUGAUCCC 3'

OHP_15 Sequence:

5' CAAUCACUAUUUAAA 3'

Figure 20: Sequence of nucleotides of the new ligands to be tested, in which additional bases were inserted, complementary to the pre-miRNA29b sequence. The added nucleotides are highlighted in yellow.

After the output by HNADOCK, the models for each ligand were analyzed following the same methodology. The docking process for the ligand O3'_17 did not give valid results since the ligand was broken in half in all models, thus this ligand was not considered for the analysis. Table 8 displays which binding sites each ligand has at least one model.

Table 8: Presence of at least one model in a binding site for the newer ligands.

Ligand	C6 Version	C12 Version
O3'_16	✓ ✓	✓ ✓
O3'_17	--	--
OHP_14	✓ ✓	✓
OHP_15	✓ ✓	✓

✓ At least one model for 5'/3' region ✓ At least one model for the hairpin structure

With the addition of nucleotides, the versatility was lost given that version C12 of the ligands OHP_14 and OHP_15 did not display any interaction with the hairpin structure. These results followed a different pattern from before, indicating that for this specific target, there is no advantage in increasing the ligands nucleotide length.

Therefore, it has been concluded that the versatility of some ligands does not rely on a particular factor such as sequence or number of bases, but rather on the effect of multiple variables where those can be involved.

Then, the docking scores, the number of interactions, and the type of interactions were investigated for these 8 ligands to understand which one could be more suitable to capture the pre-miRNA-29b target.

Table 9: Docking scores and number of interactions between the target and each ligand of the versatile group of ligands.

Ligand	C6 Version		C12 Version	
	Score	Interactions	Score	Interactions
O3'_13	-286.54	18	-295.47	9
O3'_14	-276.04	11	-319.49	13
O3'_15	-317.32	15	-313.30	15
OHP_13	-280.04	9	-279.93	14

After opening the output files provided by the HNADOCK server in BIOVIA Discovery Studio, it was possible to count and identify which interactions occur between the target and each ligand. For all ligands, the main interactions established in the complex are hydrogen bonds, representing about 80-90% of the total interactions. This can be beneficial, considering that, although the affinity between the ligand and the target is important, the interaction should be relatively easy to destabilize with mild conditions. In addition, it was expected that the H-bonds were the major interaction, since the ligands are designed to establish interactions with the target by a base pairing mechanism. Other studies have also demonstrated the significant occurrence of these interactions between nucleic acid targets and a complementary ligand. For instance, Schluep and co-workers purified plasmid DNA that formed a triplex with immobilized oligonucleotides. The ligands interacted with the DNA in a complementary manner, resulting in the formation of the triplex, with H-bonds as the major driving force [146]. In another research study, Nadal and co-workers purified mRNA using the same triplex affinity chromatography method. The establishment of H-bonds between strands was once again essential for the triplex formation, which then allowed for selective purification [147]. The ligand O3'_15 was selected from this group due to its lower docking score and the number of interactions established with the target.

Regarding the length of the carbon chain, the version with twelve carbons was selected. Despite the docking scores between the two versions being similar, the longer version could have more advantages, given that the number of interactions in the hairpin model is higher than the C6 version (data not displayed) as well as a longer spacer arm may increase the accessibility of the ligand to the target and reduce non-specific binding in the posterior experimental studies [148].

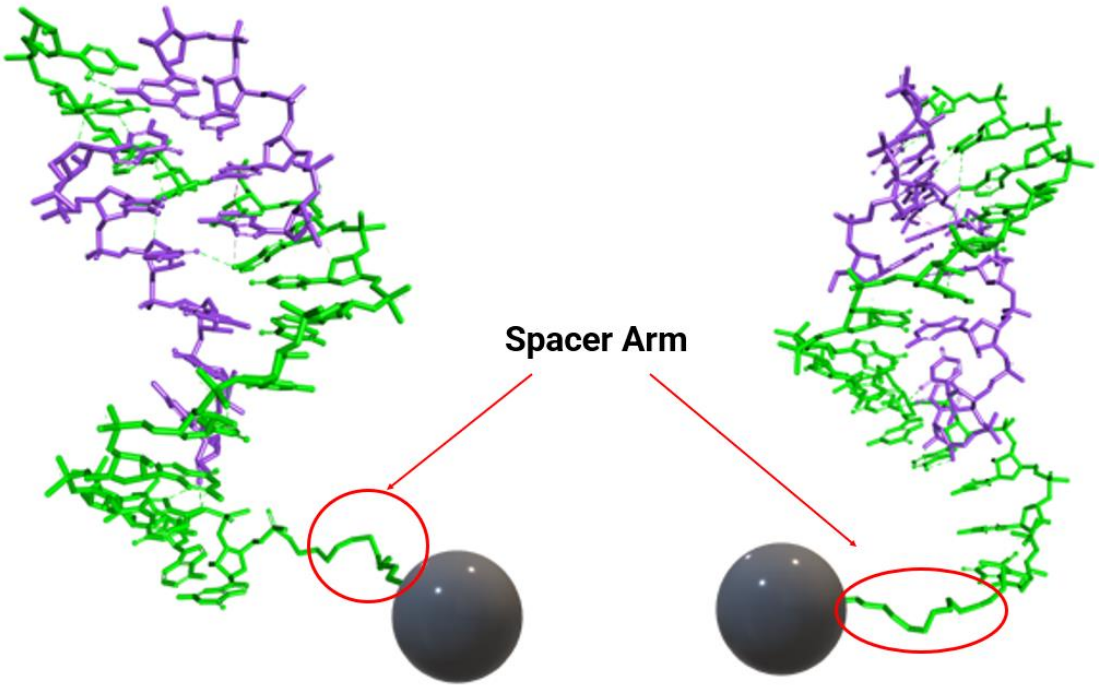


Figure 21: Interactions between the target (purple) and ligand O3'_15_C12 (green), visualized on the Discovery program. Both images represent the same complex but with a 180º horizontal rotation.

In Figure 21 are represented the interactions between the target and the ligand O3'_15_C12, and it is possible to verify that none of them occurred between the spacer arm and the target. Therefore, even if the carbon chain increases the global hydrophobic character of the ligand, the spacer arm does not directly affect the docking score obtained, considering that the C6 version has a similar score.

At this point, one ligand has been selected, O3'_15_C12, due to its good docking score and versatility. Nevertheless, this versatility could also have a negative effect, if it results in the loss of selectivity, increasing the ability to interact with impurities such as other RNAs during the purification process. Thus, another ligand that showed a good affinity value with only a model for the intended complementary region was also selected for further studies.

Considering the good docking score, the number of interactions, and the fact that it only interacts in one binding site, the ligand O5'_11 was also selected for experimental studies, to evaluate its selectivity. Furthermore, in a way to understand the real effect of the carbon chain length, the two versions of this ligand were selected for extended experimental studies. The interactions between each of these ligands and the target are displayed in Figure 22 and Figure 23.

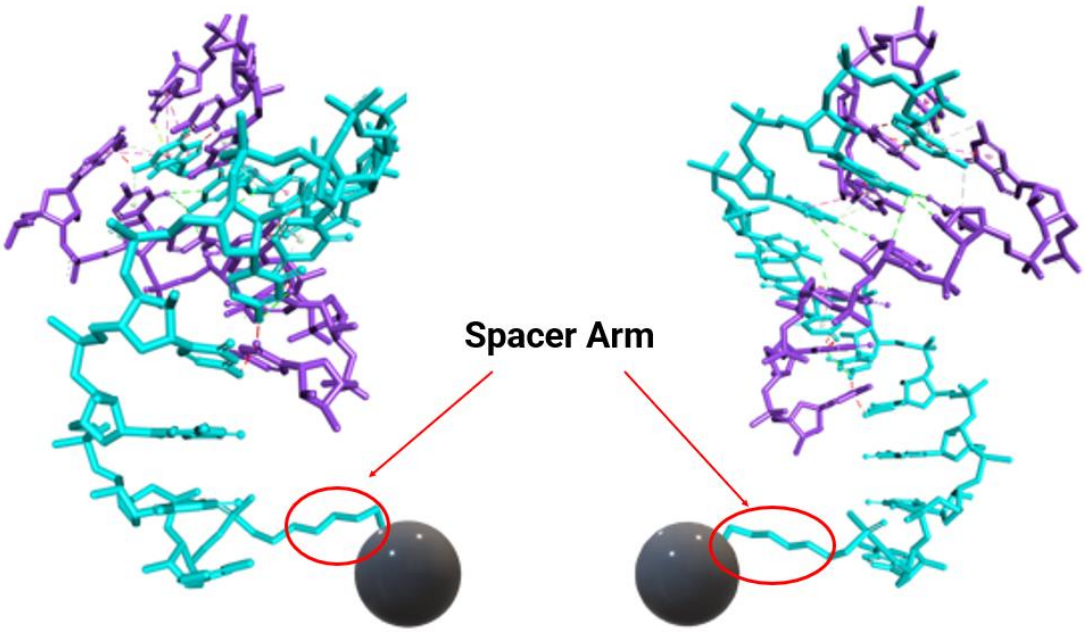


Figure 22: Interactions between the target (purple) and ligand O5'_11_C6 (light blue), visualized in the Discovery program. Both images represent the same complex but with a 180º horizontal rotation.

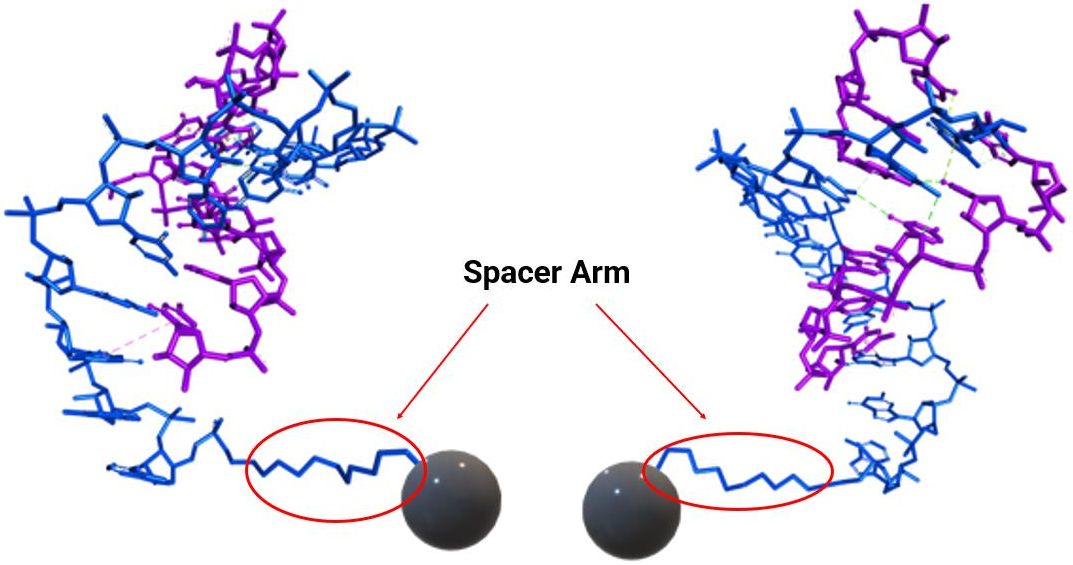


Figure 23: Interactions between the target (purple) and ligand O5'_11_C12 (dark blue), visualized in the Discovery program. Both images represent the same complex but with a 180º horizontal rotation.

Once more, no interactions were found between the spacer arm of the ligands and the target, which could mean that the presence of the carbon chain induces minimal effect on the affinity. Finally, since a ligand from the 5' group and a ligand from the 3' group were selected, another candidate from the OHP group was also chosen in order to compare the capture efficiency between ligands that were designed for different binding sites. In the OHP group, only the ligand OHP_13 was versatile, so it was chosen over the rest. The longer version of carbons was preferred over the shorter one for the same previous reasons. The interactions between this ligand and the target are displayed in Figure 24.

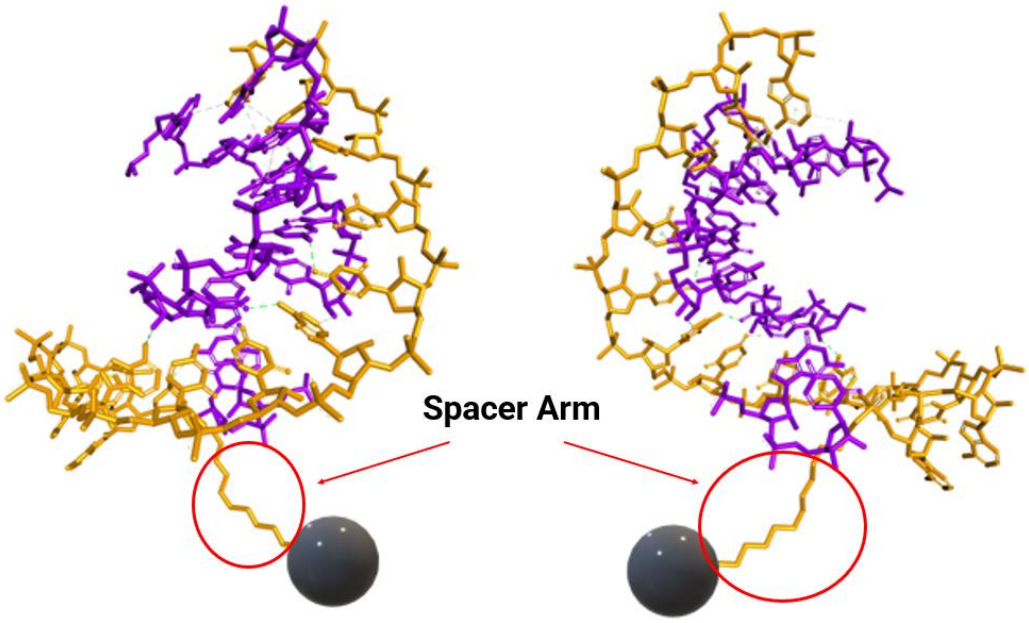


Figure 24: Interactions between the target (purple) and ligand OHP_13_C12 (yellow), visualized in the Discovery program. Both images represent the same complex but with a 180º horizontal rotation.

After performing this virtual screening using molecular docking, and based on the results of interactions and versatility, 4 ligands were selected from the 26 initial candidates and used in further studies. The ligands chosen are O3'_15_C12, O5'_11_C6, O5'_11_C12, and OHP_13_C12.

4.2 Molecular Dynamic Studies

In order to get a better understanding of each complex stability and to validate the docking results, molecular dynamic simulations were performed. The best models for each selected ligand were inputted in the molecular dynamic simulations' server. The program used was NAMD, with a total time of simulation of 500 ps and at a temperature of 300 K. After the simulation was performed, the RMSD per residue, Trajectory RMSD, B-factor per residue, and MD trajectory were analyzed. The RMSD per residue values of all complexes are displayed in the Figure 25. Each individual graphic can be found in the attachments.

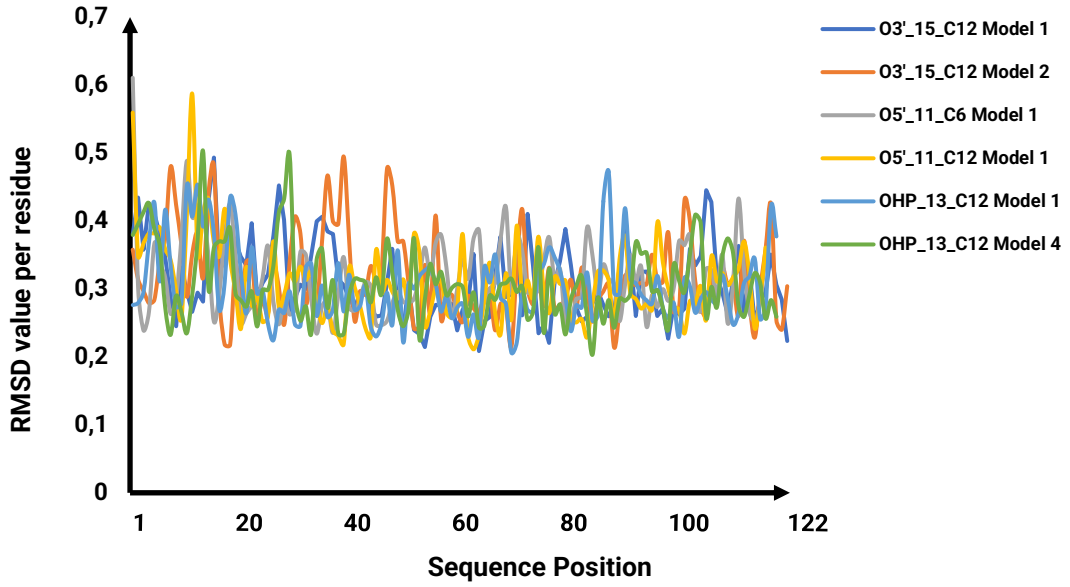


Figure 25: RMSD per residue for all complexes.

In Figure 25, the largest RMSD values for each nucleotide of the ligand and the receptor are observed. Typical RMSD values below 2 Å suggest a high level of structural similarity between two conformations and high stability [149,150]. For instance, Ghosh and co-workers identified 3 potential drug candidates for the main protease of SARS CoV-2 using *in silico* tools. The RMSD obtained for each complex ranged from 1.43 Å to 1.61 Å, indicating their stability [151]. In another study for the same protein, Kumar and co-workers also identified 3 promising inhibitors using molecular docking and dynamic simulations. The authors claimed that the complexes formed were stable, with RMSD values ranging from 1.5 Å to 2.45 Å [152]. Thus, it can be observed that all complexes

tested in this work exhibit very low RMSD values and show minimal fluctuations. This suggests that all complexes are stable, and it is unlikely that both the ligand and receptor undergo conformational changes.

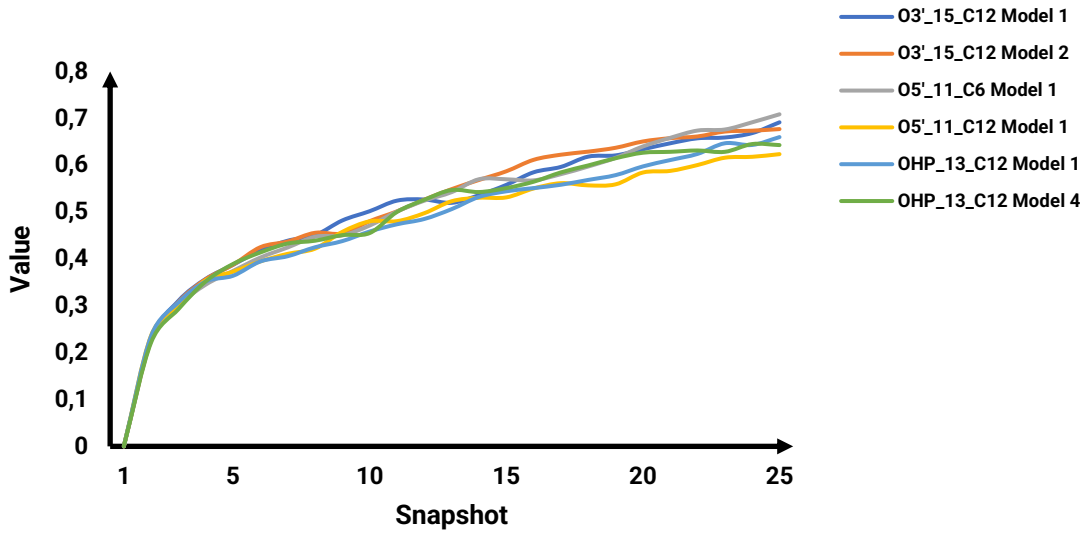


Figure 26: RMSD Trajectory.

Contrary to Figure 25, Figure 26 illustrates the dynamic changes in RMSD of the complex through the simulation. Each snapshot corresponds to a specific moment during the simulation. The RMSD values of all the complexes remained under 1 Å from the beginning until the end of the simulation, indicating their good stability [149,150]. Therefore, it is unlikely for these complexes to change their conformation without the presence of a destabilizing factor.

Based on the RMSD analysis for individual nucleotides as well as the overall structure, we can draw the conclusion that each one of these ligands forms a stable complex with the target. The stability of these complexes holds significant importance for the study, particularly in the context of future chromatography assays. It is important that complex is successfully formed during the assay, but the interaction should not be so strong that only highly aggressive strategies would allow the targets elution, since it could compromise its activity and stability. So, these results of molecular dynamics also indicate that these ligands can be good candidates for the purification of the pre-miRNA-29b.

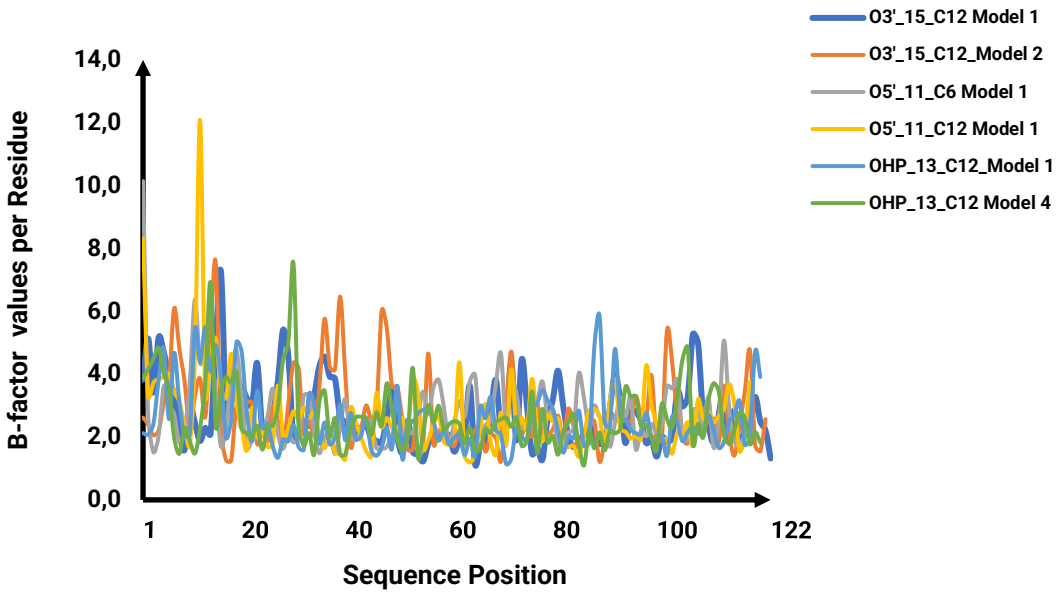


Figure 27: B-factor of all the complexes.

In Figure 27 are represented the B-factor values for each nucleotide of the ligand and the receptor, in which the B-factor reflects the thermal motion of a structure. The simulation was performed at 300 K which is the room temperature. B-factor have been also used as a parameter to understand the stability of the complex over time. For instance, Prajapati and co-workers found sterenin M as potential inhibitor of SARS CoV-2 main protease through molecular docking and molecular dynamic simulation studies. The group have shown that the B-factor values for the complex ranged from 15 to 75 $Å^2$, but for most of the complex residues the values were around 30 $Å^2$, indicating that the ligand formed a stable complex with the target [153]. The values obtained in this work (Figure 29) are low B-factor values, which indicates that the atom's position of the complex is well defined and has limited thermal motion [154,155]. This supports the previous RMSD results, that suggest that all complexes are stable.

The program also provided twenty-five snapshots of each simulation. Considering that all simulations have similar behaviors in terms of stability, only the O5'_11_C6 complexed with the target is illustrated in Figure 28. Four snapshots of the simulation were selected for demonstration purposes: the initial moment (A), two intermediary moments (10^{th} and 20^{th} snapshots, (B) and (C) respectively) and the final moment (D). It is difficult to observe any minor differences between the structures due to its high complex stability. The snapshots of all the simulations conducted were analyzed (data

not shown) and all suggest that the structures do not undergo significant conformational changes, which is in accordance with the previous information.

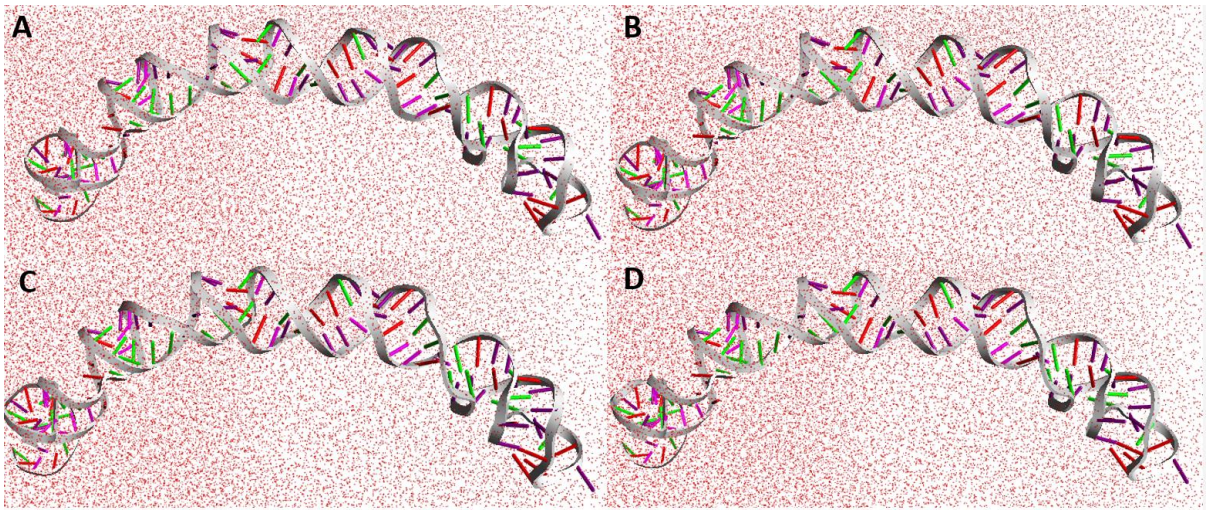


Figure 28: Different moments of the simulation with the O5'_11_C6 complexed with the pre-miRNA-29b. A- Initial Moment; B- 10th Snapshot; C- 20th Snapshot; D- Final moment.

4.3 Synthesis and Characterization of functionalized supports

Once selected the most promising ligands for the interaction with pre-miRNA-29b, different conditions were tested for the immobilization of these ligands onto the chromatographic supports, to study their potential for pre-miRNA-29b purification. After the immobilization of ligands onto Sepharose support, a small portion were used for elemental analysis. This analysis was performed at the University of Aveiro, but in parallel, it was evaluated the capacity of the different synthesized supports to capture the pre-miRNA-29b.

For these assays, different conditions were tested in a way to favor different interactions between the ligands and the target. Besides the supports functionalized with ligands, two other supports with just Sepharose or activated Sepharose were also studied as controls of the functionalization.

Figure 29 shows the results regarding the binding/elution experiments, when using the experimental conditions to mainly favor hydrophobic interactions between the new ligands and RNA. To favor these interactions, the binding step was accomplished using 2.5 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl pH 8.0, which increased the ionic strength and exposed the non-polar regions of RNA for possible interactions. As for the elution step, the ionic strength was decreased using 10 mM Tris-HCl pH 8.0 buffer. The results of

RNA capture and elution were analyzed by absorbance measurement at 260 nm, and the elution percentage was calculated in each step of the chromatographic assay.

It can be observed that every functionalized support retained all the RNA in these conditions (Figure 29, A, B, and C) but the same behavior was found for the supports without ligands (Figure 29, D and E). In addition, all the supports presented a similar binding/elution profile, suggesting that there were no differences between the support's constitution.

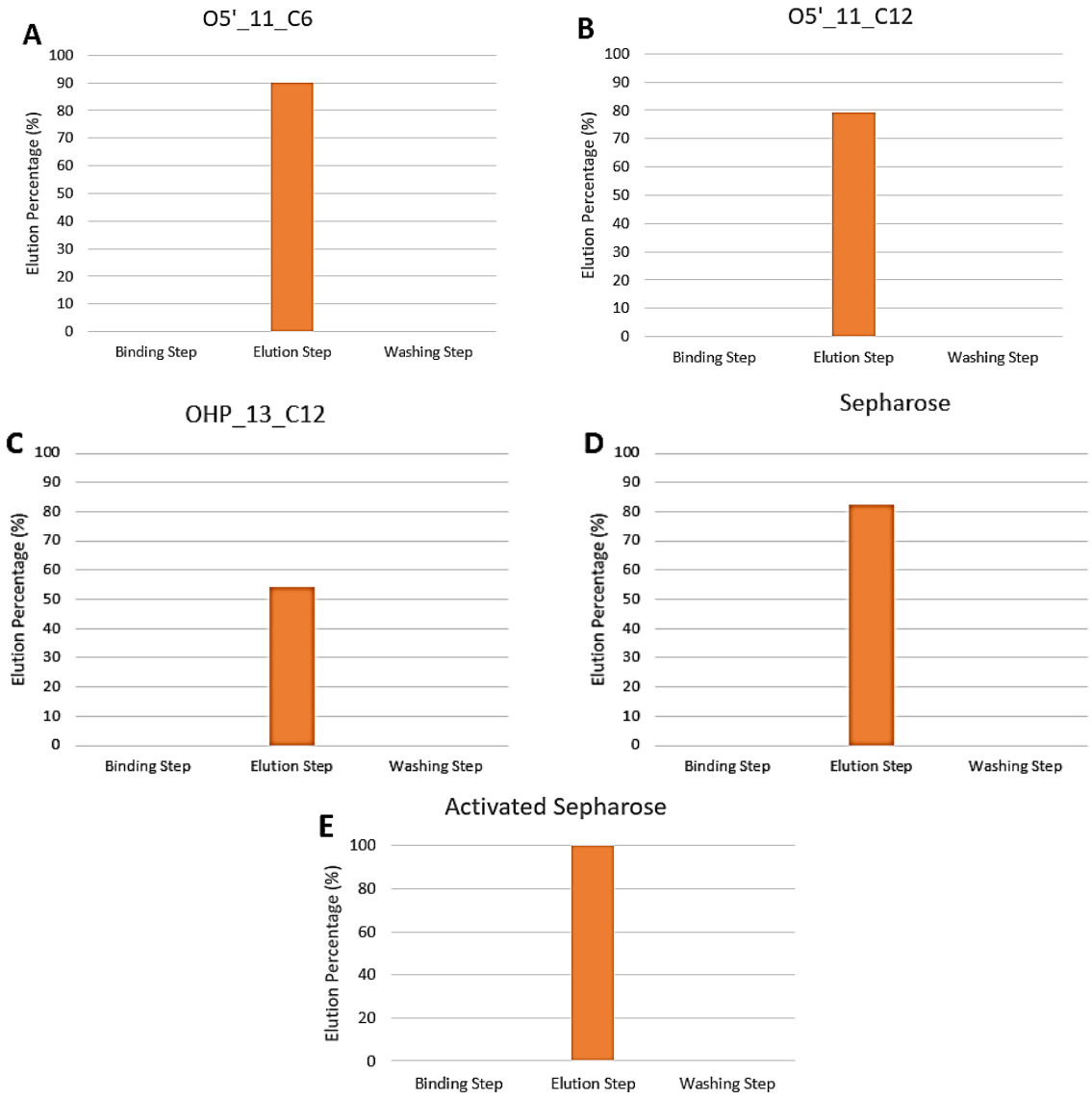


Figure 29: Binding / Elution behavior of RNA at each step of the chromatographic assay in conditions that favor hydrophobic interactions. A – O5'_11_C6; B – O5'_11_C12; C – OHP_13_C12; D – Sepharose; E – Activated Sepharose

In fact, these results were corroborated by the elemental analysis (Table 10), where it was possible to confirm that the supports were not effectively modified with the ligands. The N content, that would confirm the immobilization with the oligonucleotides, is very residual, indicating the absence of ligands.

Table 10: Elemental analysis of the first synthetized supports.

Oligonucleotide	Mass (mg)	% C	% H	% N
O5'_11_C6	1.364	41.519	6.560	0.113
O5'_11_C12	1.268	43.567	6.960	0.128
O3'_15_C12	1.339	41.200	6.397	0.094
OHP_13_C12	1.213	41.261	6.709	0.114

The Sepharose backbone has carbon and hydrogen in its constitution, so only the content of nitrogen can confirm if there is any ligand immobilized in the support. Given that a support functionalized with ligands would typically have an amine group attached to the carbon tail, along with various nitrogen atoms associated with the nucleotides, the low nitrogen percentage indicates that the immobilization step was not efficient.

Then, we proceeded to optimize the immobilization protocol. Firstly, to achieve better control temperature during the process, we replaced the previous heat provider with a paraffin bath. Then, we conducted two additional support synthesis: one maintaining the oligonucleotides mass but reducing the volume of Sepharose by half compared to the previous synthesis, to test if more ligand per Sepharose would result in better immobilization. For the other support, it was maintained the ratio used in the initial synthesis to verify if the paraffin bath was beneficial for the reaction.

Elemental analysis of the second synthetized supports was not possible due to external reasons, however, the new supports were tested in the same way as previously. Chromatography assays were conducted using the conditions that favor hydrophobic interactions to compare the RNA retention behavior of the new functionalized supports with the supports without ligands (non-modified).

The RNA elution profiles of the new supports are illustrated in Figure 30. The support with O5'_11_C6 ligand contains the same amount of oligonucleotide mass but uses only half the volume of Sepharose (5 mL of Sepharose and 1.24 μM of ligand). On the other hand, the support with O3'_15_C12 oligo maintains the same proportions as in the first synthesis (5 mL of Sepharose and 0.62 μM of ligand). It is shown that these differences do not affect the RNA retention behavior of both supports (Figure 30 – A and B) when

compared to the results with the first synthesis. Additionally, the profiles are similar to the supports without ligands (Figure 29 – E and D). From this, we concluded that the optimization of temperature control, and the manipulation of the ratio between the oligonucleotides and Sepharose was not sufficient for an efficient immobilization of ligands onto the support.

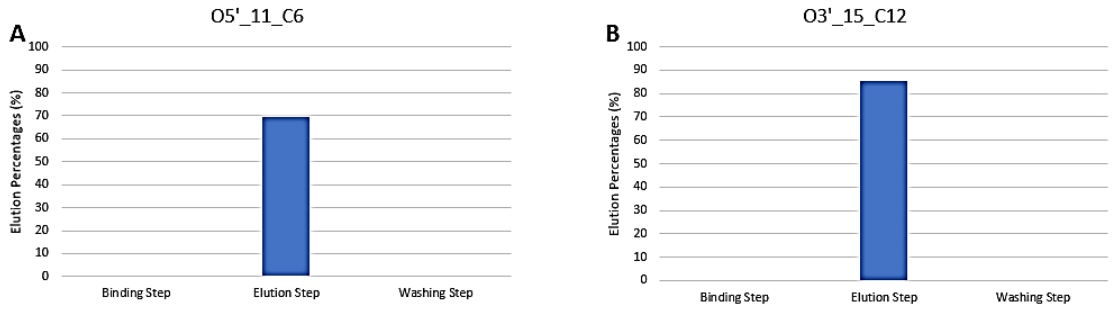


Figure 30: Binding/Elution behavior of RNA at each step of the chromatographic assay in conditions that favor hydrophobic interactions. A – O5'_11_C6; B – O3'_15_C12;

Then, we proceeded with a third synthesis, following the same protocol and incorporating a new optimization related to the reduction of sepharose volume from 5 mL to 2.5 mL, which resulted in an increase of oligonucleotide per quantity of support.

All conditions estabilished on each synsthesis are displayed in table 12.

Table 12: Summarized Conditions estabilished on each synthesis round.

Synthesis	[oligonucleotide]		Sepharose Volume	Functionalization Step
1º	1.24 µM		10 mL	The 16 h reaction was performed at 55º C using a magnetic hot plate stirrer.
2º	1.24 µM	0.62 µM	5 mL	The 16 h reaction was performed at 55º C in a paraffin bath.
3º	1.24 µM		2.5 mL	The 16 h reaction was performed at 55º C in a paraffin bath.

The oligonucleotides used for the third synthesis were the same that were used on the second, O5'_11_C6 and O3'_15_C12. After the support synthesis, a small portion was also used for elemental analysis. Once again, the supports were tested with hydrophobic

conditions while waiting for elemental analysis results. The RNA retention behavior of one of the newly supports is displayed in Figure 31. So far, the elution profile is similar to the previous supports that were in fact, non-functionalized. This suggests that the third synthesis was not aso effective in the immobilizing of the ligands onto the support as well. However, we decided to modify the salt concentration used, since we could be working at extreme conditions. Given that the assay was perfomed following the same steps but with a 1.5 M $(NH_4)_2SO_4$ for the binding step. The results for the support with activated Sepharose and O5'_11_C6 are displayed in Figure 32.

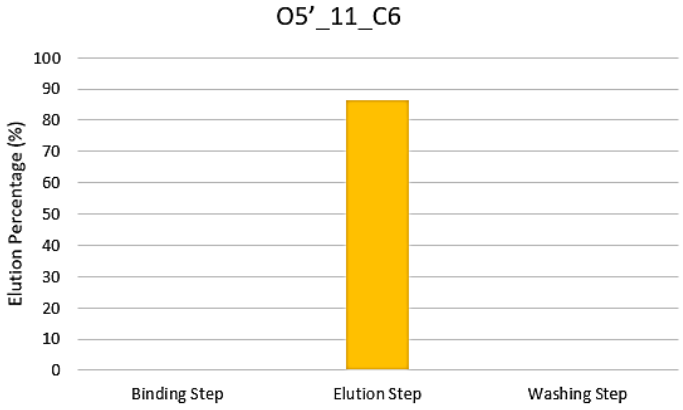


Figure 31: Binding/Elution behavior of RNA at each step of the chromatographic assay in conditions that favor hydrophobic interactions. Third synthesis of the support with the O5'_11_C6 oligo.

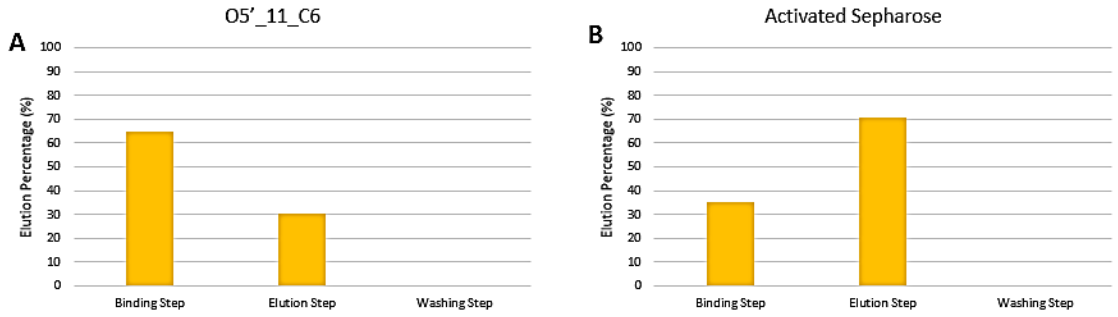


Figure 32: Binding/Elution behavior of RNA at each step of the chromatographic assay with the modification of the binding step from 2.5 M to 1.5 M $(NH_4)_2SO_4$. A- O5'_11_C6; B- Activated Sepharose.

With the reduction of ammonium sulfate concentration from 2.5 M to 1.5 M, we observed different behaviors in RNA retention between the two supports. Considering that the ionic strength has decreased, the RNA sample was not totally bound to the support, resulting in some RNA elution during the binding step. Once again, the results were

similar between the supports subjected to the immobilization protocol and the non-modified ones, which suggests that the ligand immobilization was ineffective. These results were posteriorly confirmed by the elemental analysis, displayed in Table 13. We can observe that, despite the optimizations and increased proportion of ligand per Sepharose, the immobilization reaction failed. The nitrogen percentages remain very low, similar to the first synthesis.

Table 13. Elemental analysis of the third synthetized functionalized supports.

Oligonucleotide	Mass (mg)	% C	% H	% N
O5'_11_C6	1.554	42.776	6.363	0.096
O3'_15_C12	1.587	45.931	6.710	0.096

Besides hydrophobic conditions, all the assays illustrated were also repeated with conditions that favor ionic interactions, and conditions usually used for mRNA purification with Oligo(dT) columns. The data is not shown, but the results were all equivalent to the Sepharose support since the functionalization was not effective. Various protocol parameters can be adjusted to attempt a more efficient reaction, such as increasing the amount of oligonucleotide used, raising the temperature, using a catalyst, or substituting the epoxy group with a carbonyl group, which is also reactive with primary amines. While the functionalization protocol requires optimization, the molecular docking and dynamic simulation studies have already reduced the amount of experimental work needed for identifying the most promising ligands.

Chapter 5 – Conclusion and Future Perspectives

New approaches using nucleic acids are emerging for the treatment of various diseases that currently lack effective therapies, including Alzheimer's disease. Although the recombinant production of these biomolecules is cost-effective and suitable for industrial scaling, further improvements are still required in later stages of the biotechnological process. In order to study and bring these biopharmaceuticals to the clinic, it is required that its integrity and biological activity are guaranteed, as well as its purity must be in accordance with regulatory agencies requisites. The method of choice for biomolecules purification within biopharmaceutical industries is liquid chromatography, particularly, affinity chromatography, which is the most powerful approach as it relies in multiple, specific, and reversible interactions. However, a major drawback is the need for a specific ligand for each target, and the process of finding the right ligand can be laborious and expensive. This is where *in silico* methods can support the ligands selection process. Instead of testing each ligand on the laboratory, what would require a lot of resources and time, *in silico* methods such as molecular docking can provide similar information in a faster and cheaper way.

With this work, it was possible to identify four ligands from the original twenty-six group and confirm its stability with the pre-miRNA-29b through molecular dynamic simulations. Firstly, the twenty-six designed ligands and the pre-miRNA-29b were prepared and inputted in a molecular docking program. Then, through the analysis of its docking score, binding site, number of interactions, and specificity, the ligands O5'_11_C6, O5'_11_C12, O3'_15_C12, and OHP_13_C12 were found to be the most promising ones. This group of ligands was selected to be further studied through chromatographic assays to explore the influence of the carbon chain and the purification potential of versatile and specific ligands. Molecular Dynamic Simulations were also performed to gain a better understanding of each ligand stability when interacting with the target. The results of RMSD and B-factor, along with the simulation images, suggest that all ligands form stable complexes with pre-miRNA-29b. To perform chromatographic studies and evaluate the purification capacity of these ligands, the immobilization onto the supports is necessary. Therefore, a known synthesis protocol was conducted in order to modify the Sepharose matrix with the oligonucleotides. Despite all the optimizations attempted, none of the synthesis was efficient, as the percentages of nitrogen obtained from elemental analysis were too low, meaning that this is a step that still need optimization to be succeeded.

Therefore, this work demonstrates how *in silico* methods can save time and resources while retrieving important information for research. Particularly, it has been demonstrated that molecular docking and molecular dynamic simulations can be

implemented in new fields besides their common use in drug discovery. The number of candidates has been reduced by 85%, advancing with only four ligands for the experimental studies, which represents a more cost-effective method to select affinity ligands.

As for future perspectives, the immobilization protocol needs improvement or replacement with a suitable one. Following successful functionalization of the supports, chromatographic studies should be conducted with samples of different complexities to understand the capture potential of these ligands for pre-miRNA-29b. Additionally, it is essential to gain insight into the influence of arm length and the advantages of using a versatile ligand over a more specific one. Given that the ligands designed for the 5' and 3' regions derive their complementarity due to the ribozyme product sequences, they should undergo testing for the purification of various RNA molecules produced by the same expression system. Furthermore, these tools must be tested with other pairs of oligonucleotides/RNA to understand the general applicability of this method.

References

1. Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Developmental biology*, *278*(2), 274-288. doi:10.1016/j.ydbio.2004.11.028
2. McCarty, M., & Avery, O. T. (1946). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: II. Effect of desoxyribonuclease on the biological activity of the transforming substance. *The Journal of experimental medicine*, *83*(2), 89.
3. Hershey, A. D., & Chase, M. (2017). *Independent functions of viral protein and nucleic acid in growth of bacteriophage* (pp. 121-139). Springer Berlin Heidelberg. doi: 10.1085/jgp.36.1.39.
4. Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737-738.
5. Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561-563. doi:10.1038/227561a0
6. Nirenberg, M., Caskey, T., Marshall, R., Brimacombe, R., Kellogg, D., Doctor, B., ... & Anderson, F. (1966, January). The RNA code and protein synthesis. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 31, pp. 11-24). Cold Spring Harbor Laboratory Press. doi:10.1101/SQB.1966.031.01.008
7. Delihas, N. (2015). Discovery and characterization of the first non-coding RNA that regulates gene expression, micF RNA: A historical perspective. *World journal of biological chemistry*, *6*(4), 272. doi:10.4331/wjbc.v6.i4.272.
8. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... & Kalush, F. (2001). The sequence of the human genome. *science*, *291*(5507), 1304-1351. doi:10.1126/science.1058040
9. International Human Genome Sequencing Consortium. Correction: Initial sequencing and analysis of the human genome. *Nature* **412**, 565–566 (2001). doi:10.1038/35087627
10. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., ... & Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44-53. doi:10.1126/science.abj6987
11. Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *cell*, *75*(5), 843-854. doi: 10.1016/0092-8674(93)90529-y.
12. Angelucci, F., Cechova, K., Valis, M., Kuca, K., Zhang, B., & Hort, J. (2019). MicroRNAs in Alzheimer's disease: diagnostic markers or therapeutic agents?. *Frontiers in pharmacology*, *10*, 665. doi:10.3389/fphar.2019.00665
13. Pierce, B. A. (2012). Genetics: A Conceptual Approach WF Freeman and Company. *New York, USA. 324pp*.

14. Nelson, D. L., Lehninger, A. L., & Cox, M. M. (2008). *Lehninger principles of biochemistry*. Macmillan.
15. Suzuki, D. T., Griffiths, A. J., Miller, J. H., & Lewontin, R. C. (2007). *An introduction to genetic analysis* (No. Ed. 9). WH Freeman and Company.
16. Soukup, G. A., & Breaker, R. R. (1999). Relationship between internucleotide linkage geometry and the stability of RNA. *Rna*, *5*(10), 1308-1325. doi:10.1017/S1355838299990891
17. Thaplyal, P., & Bevilacqua, P. C. (2014). Experimental approaches for measuring pKa's in RNA and DNA. In *Methods in enzymology* (Vol. 549, pp. 189-219). Academic Press. doi:10.1016/B978-0-12-801122-5.00009-X
18. Neves, M. C., Pereira, P., Pedro, A. Q., Martins, J. C., Trindade, T., Queiroz, J. A., ... & Sousa, F. (2020). Improved ionic-liquid-functionalized macroporous supports able to purify nucleic acids in one step. *Materials Today Bio*, *8*, 100086. doi:10.1016/j.mtbio.2020.100086
19. Wyatt, J. R., Puglisi, J. D., & Tinoco Jr, I. (1989). RNA folding: pseudoknots, loops and bulges. *BioEssays*, *11*(4), 100-106.doi:10.1002/bies.950110406
20. Higgs, P. G. (2000). RNA secondary structure: physical and computational aspects. *Quarterly reviews of biophysics*, *33*(3), 199-253. doi:10.1017/S0033583500003620
21. Cheng, L., Zhang, D. Y., & Eble, J. N. (Eds.). (2008). *Molecular genetic pathology*. Totowa: Humana. doi:10.1007/978-1-59745-405-6
22. Hombach, S., & Kretz, M. (2016). Non-coding RNAs: classification, biology and functioning. *Non-coding RNAs in colorectal cancer*, 3-17. doi:10.1007/978-3-319-42059-2_1
23. Morey, C., & Avner, P. (2004). Employment opportunities for non-coding RNAs. *FEBS letters*, *567*(1), 27-34. doi:10.1016/j.febslet.2004.03.117
24. L Losko, M., Kotlinowski, J., & Jura, J. (2016). Long noncoding RNAs in metabolic syndrome related disorders. *Mediators of inflammation*, *2016*. doi:10.1155/2016/5365209
25. Szymański, M., Erdmann, V. A., & Barciszewski, J. (2003). Noncoding regulatory RNAs database. *Nucleic acids research*, *31*(1), 429-431. doi:10.1093/nar/gkg124
26. Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature reviews genetics*, *10*(3), 155-159.doi:10.1038/nrg2521
27. Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., & Shen, F. (2019). Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *International journal of molecular sciences*, *20*(22), 5573. doi:10.3390/ijms20225573

28. O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in endocrinology*, *9*, 402. doi:10.3389/fendo.2018.00402
29. Alshaer, W., Zureigat, H., Al Karaki, A., Al-Kadash, A., Gharaibeh, L., Ma'mon, M. H., ... & Awidi, A. (2021). siRNA: Mechanism of action, challenges, and therapeutic approaches. *European journal of pharmacology*, *905*, 174178. doi:10.1016/j.ejphar.2021.174178
30. Rayford, K. J., Cooley, A., Rumph, J. T., Arun, A., Rachakonda, G., Villalta, F., ... & Nde, P. N. (2021). piRNAs as modulators of disease pathogenesis. *International Journal of Molecular Sciences*, *22*(5), 2373. doi:10.3390/ijms22052373
31. Sun, Y. H., Lee, B., & Li, X. Z. (2022). The birth of piRNAs: how mammalian piRNAs are produced, originated, and evolved. *Mammalian Genome*, *33*(2), 293-311. doi:10.1007/s00335-021-09927-8
32. Zhou, S., Chen, W., Cole, J., & Zhu, G. (2020). Delivery of nucleic acid therapeutics for cancer immunotherapy. *Medicine in drug discovery*, *6*, 100023. doi:10.1016/j.medidd.2020.100023
33. Amakiri, N., Kubosumi, A., Tran, J., & Reddy, P. H. (2019). Amyloid beta and microRNAs in Alzheimer's disease. *Frontiers in Neuroscience*, *13*, 430. doi:10.3389/fnins.2019.00430
34. Pereira, P., Queiroz, J. A., Figueiras, A., & Sousa, F. (2016). Affinity approaches in RNAi-based therapeutics purification. *Journal of Chromatography B*, *1021*, 45-56. doi:10.1016/j.jchromb.2016.01.022
35. Kulkarni, J. A., Witzigmann, D., Thomson, S. B., Chen, S., Leavitt, B. R., Cullis, P. R., & van der Meel, R. (2021). The current landscape of nucleic acid therapeutics. *Nature nanotechnology*, *16*(6), 630-643. doi:10.1038/s41565-021-00898-0
36. Yamada, Y. (2021). Nucleic acid drugs—current status, issues, and expectations for exosomes. *Cancers*, *13*(19), 5002. doi:10.3390/cancers13195002
37. Dammes, N., & Peer, D. (2020). Paving the road for RNA therapeutics. *Trends in Pharmacological Sciences*, *41*(10), 755-775. doi:10.1016/j.tips.2020.08.004
38. Rodgers, G., Austin, C., Anderson, J., Pawlyk, A., Colvis, C., Margolis, R., & Baker, J. (2018). Glimmers in illuminating the druggable genome. *Nature reviews Drug discovery*, *17*(5), 301-302. doi:10.1038/nrd.2017.252
39. Lam, J. K., Chow, M. Y., Zhang, Y., & Leung, S. W. (2015). siRNA versus miRNA as therapeutics for gene silencing. *Molecular Therapy-Nucleic Acids*, *4*. doi:10.1038/mtna.2015.23
40. Parker, A. L., Kavallaris, M., & McCarroll, J. A. (2014). Microtubules and their role in cellular stress in cancer. *Frontiers in oncology*, *4*, 153. doi:10.3389/fonc.2014.00153
41. Teo, J., McCarroll, J. A., Boyer, C., Youkhana, J., Sagnella, S. M., Duong, H. T., ... & Phillips, P. A. (2016). A rationally optimized nanoparticle system for the delivery of

RNA interference therapeutics into pancreatic tumors in vivo. *Biomacromolecules*, *17*(7), 2337-2351. doi:10.1021/acs.biomac.6b00185

42. McCarroll, J. A., Sharbeen, G., Liu, J., Youkhana, J., Goldstein, D., McCarthy, N., ... & Phillips, P. A. (2015). ßIII-tubulin: a novel mediator of chemoresistance and metastases in pancreatic cancer. *Oncotarget*, *6*(4), 2235. doi:10.18632/oncotarget.2946
43. Pereira, P. A., Tomás, J. F., Queiroz, J. A., Figueiras, A. R., & Sousa, F. (2016). Recombinant pre-miR-29b for Alzheimer s disease therapeutics. *Scientific Reports*, *6*(1), 19946. doi:10.1038/srep19946
44. Wei, J. S., Song, Y. K., Durinck, S., Chen, Q. R., Cheuk, A. T. C., Tsang, P., ... & Khan, J. (2008). The MYCN oncogene is a direct target of miR-34a. *Oncogene*, *27*(39), 5204-5213. doi:10.1038/onc.2008.154
45. Tivnan, A., Orr, W. S., Gubala, V., Nooney, R., Williams, D. E., McDonagh, C., ... & Stallings, R. L. (2012). Inhibition of neuroblastoma tumor growth by targeted delivery of microRNA-34a using anti-disialoganglioside GD2 coated nanoparticles. *PloS one*, *7*(5), e38129. doi:10.1371/journal.pone.0038129
46. Khoshnam, S. E., Winlow, W., Farbood, Y., Moghaddam, H. F., & Farzaneh, M. (2017). Emerging roles of microRNAs in ischemic stroke: as possible therapeutic agents. *Journal of stroke*, *19*(2), 166. doi:10.5853/jos.2016.01368
47. Zhang, L., Li, Y. J., Wu, X. Y., Hong, Z., & Wei, W. S. (2015). Micro RNA-181c negatively regulates the inflammatory response in oxygen-glucose-deprived microglia by targeting Toll-like receptor 4. *Journal of neurochemistry*, *132*(6), 713-723. doi:10.1111/jnc.13021
48. Wang, L., & Zhang, L. (2020). MicroRNAs in amyotrophic lateral sclerosis: From pathogenetic involvement to diagnostic biomarker and therapeutic agent development. *Neurological Sciences*, *41*(12), 3569-3577. doi:10.1007/s10072-020-04773-z
49. Williams, A. H., Valdez, G., Moresi, V., Qi, X., McAnally, J., Elliott, J. L., ... & Olson, E. N. (2009). MicroRNA-206 delays ALS progression and promotes regeneration of neuromuscular synapses in mice. *Science*, *326*(5959), 1549-1554. doi:10.1126/science.1181046
50. Adams, D., Gonzalez-Duarte, A., O'Riordan, W. D., Yang, C. C., Ueda, M., Kristen, A. V., ... & Suhr, O. B. (2018). Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *New England Journal of Medicine*, *379*(1), 11-21. doi:10.1056/nejmoa1716153
51. Merćep, I., Friščić, N., Strikić, D., & Reiner, Ž. (2022). Advantages and disadvantages of inclisiran: a small interfering ribonucleic acid molecule targeting PCSK9—a narrative review. *Cardiovascular Therapeutics*, *2022*. doi:10.1155/2022/8129513

52. Balwani, M., Sardh, E., Ventura, P., Peiró, P. A., Rees, D. C., Stölzel, U., ... & Gouya, L. (2020). Phase 3 trial of RNAi therapeutic givosiran for acute intermittent porphyria. *New England journal of medicine*, *382*(24), 2289-2301. doi:10.1056/nejmoa1913147
53. Garrelfs, S. F., Frishberg, Y., Hulton, S. A., Koren, M. J., O'Riordan, W. D., Cochat, P., ... & Lieske, J. C. (2021). Lumasiran, an RNAi therapeutic for primary hyperoxaluria type 1. *New England journal of medicine*, *384*(13), 1216-1226. doi:10.1056/nejmoa2021712
54. Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., ... & van der Flier, W. M. (2021). Alzheimer's disease. *The Lancet*, *397*(10284), 1577-1590. doi:10.1016/S0140-6736(20)32205-4
55. Jagust, W. (2018). Imaging the evolution and pathophysiology of Alzheimer disease. *Nature Reviews Neuroscience*, *19*(11), 687-700. doi:10.1038/s41583-018-0067-3
56. Zong, Y., Wang, H., Dong, W., Quan, X., Zhu, H., Xu, Y., ... & Qin, C. (2011). miR-29c regulates BACE1 protein expression. *Brain research*, *1395*, 108-115. doi:10.1016/j.brainres.2011.04.035
57. Nelissen, F. H., Leunissen, E. H., van de Laar, L., Tessari, M., Heus, H. A., & Wijmenga, S. S. (2012). Fast production of homogeneous recombinant RNA—towards large-scale production of RNA. *Nucleic acids research*, *40*(13), e102-e102. doi:10.1093/nar/gks292
58. Pereira, P., Pedro, A. Q., Queiroz, J. A., Figueiras, A. R., & Sousa, F. (2017). New insights for therapeutic recombinant human miRNAs heterologous production: Rhodovolum sulfidophilum vs Escherichia coli. *Bioengineered*, *8*(5), 670-677. doi:10.1080/21655979.2017.1284710
59. Caruthers, M. H. (2011). A brief review of DNA and RNA chemical synthesis. *Biochemical Society Transactions*, *39*(2), 575-580. doi:10.1042/BST0390575
60. Baronti, L., Karlsson, H., Marušič, M., & Petzold, K. (2018). A guide to large-scale RNA sample preparation. *Analytical and bioanalytical chemistry*, *410*, 3239-3252. doi:10.1007/s00216-018-0943-8
61. Beckert, B., & Masquida, B. (2011). Synthesis of RNA by in vitro transcription. *RNA: Methods and protocols*, 29-41. doi: 10.1007/978-1-59745-248-9_3
62. Milligan, J. F., & Uhlenbeck, O. C. (1989). Synthesis of small RNAs using T7 RNA polymerase. In *Methods in enzymology* (Vol. 180, pp. 51-62). Academic Press. doi:10.1016/0076-6879(89)80091-6
63. Chandler, M., Panigaj, M., Rolband, L. A., & Afonin, K. A. (2020). Challenges in optimizing RNA nanostructures for large-scale production and controlled therapeutic properties. *Nanomedicine*, *15*(13), 1331-1340. doi:10.2217/nnm-2020-0034

64. Ponchon, L., & Dardel, F. (2011). Large scale expression and purification of recombinant RNA in Escherichia coli. *Methods*, *54*(2), 267-273. doi:10.1016/j.ymeth.2011.02.007
65. Martins, R., Queiroz, J. A., & Sousa, F. (2014). Ribonucleic acid purification. *Journal of Chromatography A*, *1355*, 1-14. doi:10.1016/j.chroma.2014.05.075
66. Williams, J. A. (2013). Vector design for improved DNA vaccine efficacy, safety and production. *Vaccines*, *1*(3), 225-249. doi:10.3390/vaccines1030225
67. Stadler, J., Lemmens, R., & Nyhammar, T. (2004). Plasmid DNA purification. *The Journal of Gene Medicine: A cross-disciplinary journal for research on the science of gene transfer and its clinical applications*, *6*(S1), S54-S66. doi:10.1002/jgm.512
68. Carnes, A. E., & Williams, J. A. (2007). Plasmid DNA manufacturing technology. *Recent patents on biotechnology*, *1*(2), 151-166. doi:10.2174/187220807780809436
69. Guiochon, G., & Beaver, L. A. (2011). Separation science is the key to successful biopharmaceuticals. *Journal of Chromatography a*, *1218*(49), 8836-8858. doi:10.1016/j.chroma.2011.09.008
70. Roque, A. C. A., & Lowe, C. R. (2008). Affinity chromatography: history, perspectives, limitations and prospects. *Affinity chromatography: methods and protocols*, 1-23. doi: 10.1007/978-1-59745-582-4_1
71. Sofer, G. (1995). Preparative chromatographic separations in pharmaceutical, diagnostic, and biotechnology industries: current and future trends. *Journal of Chromatography A*, *707*(1), 23-28. doi:10.1016/0021-9673(95)00155-G
72. Scott, R. P. (2003). Principles and practice of chromatography. *Chrom-ed book series*, *1*.
73. Morton J. (1992) *Chromatography (5th Edition). Fundamentals and Applications of Chromatography and Related Differential Migration Methods. Part A: Fundamentals and Techniques (Journal of Chromatography Library, Vol. 51A)*. Vol 306.; 1992. doi:10.1016/0014-5793(92)81026-i
74. Coskun, O. (2016). Separation techniques: chromatography. *Northern clinics of Istanbul*, *3*(2), 156. doi:10.14744/nci.2016.32757
75. Barth, H. G., Jackson, C., & Boyes, B. E. (1994). Size exclusion chromatography. *Analytical chemistry*, *66*(12), 595-620. doi:10.1021/ac00084a022
76. Hong, P., Koza, S., & Bouvier, E. S. (2012). A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of liquid chromatography & related technologies*, *35*(20), 2923-2950. doi:10.1080/10826076.2012.743724
77. Giridhar, G., Manepalli, R. K. N. R., & Apparao, G. (2017). Size-exclusion chromatography. In *Thermal and Rheological Measurement Techniques for Nanomaterials Characterization* (pp. 51-65). Elsevier. doi:10.1016/B978-0-323-46139-9.00003-7

78. Kim, I., Mckenna, S. A., Puglisi, E. V., & Puglisi, J. D. (2007). Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *Rna*, *13*(2), 289-294. doi:10.1261/rna.342607
79. Chillón, I., Marcia, M., Legiewicz, M., Liu, F., Somarowthu, S., & Pyle, A. M. (2015). Native purification and analysis of long RNAs. In *Methods in enzymology* (Vol. 558, pp. 3-37). Academic Press. doi:10.1016/bs.mie.2015.01.008
80. Mori, S., & Barth, H. G. (1999). *Size exclusion chromatography*. Springer Science & Business Media. (Chapter 2)
81. Jungbauer, A., & Hahn, R. (2009). Ion-exchange chromatography. *Methods in enzymology*, *463*, 349-371. doi:10.1016/S0076-6879(09)63022-6
82. Cummins, P. M., Rochfort, K. D., & O'Connor, B. F. (2017). Ion-exchange chromatography: basic principles and application. *Protein chromatography: methods and protocols*, 209-223. doi:10.1007/978-1-4939-6412-3_11
83. Walls, D & Walker, JM. (2017). Protein Chromatography. *Protein Chromatogr*.1485:423. doi:10.1007/978-1-4939-6412-3
84. Easton, L. E., Shibata, Y., & Lukavsky, P. J. (2010). Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *Rna*, *16*(3), 647-653. doi:10.1261/rna.1862210
85. Koubek, J., Lin, K. F., Chen, Y. R., Cheng, R. P., & Huang, J. J. T. (2013). Strong anion-exchange fast performance liquid chromatography as a versatile tool for preparation and purification of RNA produced by in vitro transcription. *RNA*, *19*(10), 1449-1459. doi:10.1261/rna.038117.113
86. Noll, B., Seiffert, S., Hertel, F., Debelak, H., Hadwiger, P., Vornlocher, H. P., & Roehl, I. (2011). Purification of small interfering RNA using nondenaturing anion-exchange chromatography. *Nucleic acid therapeutics*, *21*(6), 383-393. doi:10.1089/nat.2011.0317
87. Acikara, Ö. B. (2013). Ion-exchange chromatography and its applications. *Column chromatography*, *10*, 55744. doi:10.5772/55744
88. Tomaz, C. T. (2017). Hydrophobic interaction chromatography. In *Liquid Chromatography* (pp. 171-190). Elsevier. doi:10.1016/B978-0-12-805393-5.00007-5
89. Queiroz, J. A., Tomaz, C. T., & Cabral, J. M. S. (2001). Hydrophobic interaction chromatography of proteins. *Journal of biotechnology*, *87*(2), 143-159. doi:10.1016/S0168-1656(01)00237-1
90. Diogo, M. M., Queiroz, J. A., & Prazeres, D. M. F. (2001). Studies on the retention of plasmid DNA and Escherichia coli nucleic acids by hydrophobic interaction chromatography. *Bioseparation*, *10*, 211-220. doi:10.1023/A:1016361721316
91. Lee, M. F. X., Chan, E. S., & Tey, B. T. (2014). Negative chromatography: Progress, applications and future perspectives. *Process Biochemistry*, *49*(6), 1005-1011. doi:10.1016/j.procbio.2014.02.018

92. Sousa, F., Prazeres, D. M., & Queiroz, J. A. (2008). Affinity chromatography approaches to overcome the challenges of purifying plasmid DNA. *Trends in biotechnology*, *26*(9), 518-525. doi:10.1016/j.tibtech.2008.05.005
93. Bonturi, N., Radke, V. S. C. O., Bueno, S. M. A., Freitas, S., Azzoni, A. R., & Miranda, E. A. (2013). Sodium citrate and potassium phosphate as alternative adsorption buffers in hydrophobic and aromatic thiophilic chromatographic purification of plasmid DNA from neutralized lysate. *Journal of Chromatography B*, *919*, 67-74. doi:10.1016/j.jchromb.2013.01.010
94. Lowe, C. R., Lowe, A. R., & Gupta, G. (2001). New developments in affinity chromatography with potential application in the production of biopharmaceuticals. *Journal of Biochemical and Biophysical Methods*, *49*(1-3), 561-574. doi:10.1016/S0165-022X(01)00220-2
95. Clonis, Y. D. (2006). Affinity chromatography matures as bioinformatic and combinatorial tools develop. *Journal of Chromatography A*, *1101*(1-2), 1-24. doi:10.1016/j.chroma.2005.09.073
96. Urh, M., Simpson, D., & Zhao, K. (2009). Affinity chromatography: general methods. *Methods in enzymology*, *463*, 417-438. doi:10.1016/S0076-6879(09)63026-3
97. Martins, R., Maia, C. J., Queiroz, J. A., & Sousa, F. (2012). A new strategy for RNA isolation from eukaryotic cells using arginine affinity chromatography. *Journal of separation science*, *35*(22), 3217-3226. doi:10.1002/jssc.201200338
98. Pereira, P., Sousa, Â., Queiroz, J., Correia, I., Figueiras, A., & Sousa, F. (2014). Purification of pre-miR-29 by arginine-affinity chromatography. *Journal of Chromatography B*, *951*, 16-23. doi:10.1016/j.jchromb.2014.01.020
99. Passmore, L. A., & Coller, J. (2022). Roles of mRNA poly (A) tails in regulation of eukaryotic gene expression. *Nature Reviews Molecular Cell Biology*, *23*(2), 93-106. doi:10.1038/s41580-021-00417-y
100. Minkner, R., Boonyakida, J., Park, E. Y., & Wätzig, H. (2022). Oligonucleotide separation techniques for purification and analysis: What can we learn for today's tasks?. *Electrophoresis*, *43*(23-24), 2402-2427. doi:10.1002/elps.202200079
101. Martins, R., Queiroz, J. A., & Sousa, F. (2012). Histidine affinity chromatography-based methodology for the simultaneous isolation of Escherichia coli small and ribosomal RNA. *Biomedical Chromatography*, *26*(7), 781-788. doi:10.1002/bmc.1729
102. Srivastava, A., Shakya, A. K., & Kumar, A. (2012). Boronate affinity chromatography of cells and biomacromolecules using cryogel matrices. *Enzyme and microbial technology*, *51*(6-7), 373-381. doi:10.1016/j.enzmictec.2012.08.006
103. Liu, Z., & He, H. (2017). Synthesis and applications of boronate affinity materials: from class selectivity to biomimetic specificity. *Accounts of chemical research*, *50*(9), 2185-2193. doi:10.1021/acs.accounts.7b00179

104. Panchapakesan, S. S. S., Jeng, S. C., & Unrau, P. J. (2015). RNA complex purification using high-affinity fluorescent RNA aptamer tags. *Annals of the New York Academy of Sciences*, *1341*(1), 149-155. doi:10.1111/nyas.12663

105. Pereira, M. J., Behera, V., & Walter, N. G. (2010). Nondenaturing purification of co-transcriptionally folded RNA avoids common folding heterogeneity. *PLoS One*, *5*(9), e12953. doi:10.1371/journal.pone.0012953

106. Wodak, S. J., & Janin, J. (1978). Computer analysis of protein-protein interaction. *Journal of molecular biology*, *124*(2), 323-342. doi:10.1016/0022-2836(78)90302-9

107. Hernández-Santoyo, A., Tenorio-Barajas, A. Y., Altuzar, V., Vivanco-Cid, H., & Mendoza-Barrera, C. (2013). Protein-protein and protein-ligand docking. *Protein engineering-technology and application*, 63-81. doi:10.5772/56376

108. Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Molecular modeling of proteins*, 365-382. doi:10.1007/978-1-59745-177-2_19

109. Pozzan, A. (2006). Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. *Current pharmaceutical design*, *12*(17), 2099-2110. doi:10.2174/138161206777585247

110. Luo, J., Wei, W., Waldispühl, J., & Moitessier, N. (2019). Challenges and current status of computational methods for docking small molecules to nucleic acids. *European journal of medicinal chemistry*, *168*, 414-425. doi:10.1016/j.ejmech.2019.02.046

111. Krüger, A., Zimbres, F. M., Kronenberger, T., & Wrenger, C. (2018). Molecular modeling applied to nucleic acid-based molecule development. *Biomolecules*, *8*(3), 83. doi:10.3390/biom8030083

112. Tessaro, F., & Scapozza, L. (2020). How 'protein-docking'translates into the new emerging field of docking small molecules to nucleic acids?. *Molecules*, *25*(12), 2749. doi:10.3390/molecules25122749

113. Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, A. C. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British journal of pharmacology*, *153*(S1), S7-S26. doi:10.1038/sj.bjp.0707515

114. Miroshnychenko, K. V., & Shestopalova, A. V. (2016). Molecular Docking of Biologically Active Substances to Double Helical Nucleic Acids: Problems and Solutions. In *Applied Case Studies and Solutions in Molecular Docking-Based Drug Design* (pp. 127-157). IGI Global. doi:10.4018/978-1-5225-0362-0.ch005

115. Feng, Y., Yan, Y., He, J., Tao, H., Wu, Q., & Huang, S. Y. (2022). Docking and scoring for nucleic acid–ligand interactions: Principles and current status. *Drug Discovery Today*, *27*(3), 838-847. doi:10.1016/j.drudis.2021.10.013

116. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, *161*(2), 269-288. doi:10.1016/0022-2836(82)90153-X

117. Yadava, U. (2018). Search algorithms and scoring methods in protein-ligand docking. *Endocrinol Int J*, *6*(6), 359-367. doi:10.15406/emij.2018.06.00212

118. Huang, S. Y., & Zou, X. (2010). Advances and challenges in protein-ligand docking. *International journal of molecular sciences*, *11*(8), 3016-3034. doi:10.3390/ijms11083016

119. Zhou, Y., Jiang, Y., & Chen, S. J. (2022). RNA–ligand molecular docking: Advances and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *12*(3), e1571. doi:10.1002/wcms.1571

120. Guilbert, C., & James, T. L. (2008). Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *Journal of chemical information and modeling*, *48*(6), 1257-1268. doi:10.1021/ci8000327

121. Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., ... & Morley, S. D. (2014). rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS computational biology*, *10*(4), e1003571. doi:10.1371/journal.pcbi.1003571

122. Sun, L. Z., Jiang, Y., Zhou, Y., & Chen, S. J. (2020). RLDOCK: a new method for predicting RNA–ligand interactions. *Journal of chemical theory and computation*, *16*(11), 7173-7183. doi:10.1021/acs.jctc.0c00798

123. Feng, Y., Zhang, K., Wu, Q., & Huang, S. Y. (2021). NLDock: a Fast Nucleic Acid–Ligand Docking Algorithm for Modeling RNA/DNA–Ligand Complexes. *Journal of Chemical Information and Modeling*, *61*(9), 4771-4782. doi:10.1021/acs.jcim.1c00341

124. He, J., Wang, J., Tao, H., Xiao, Y., & Huang, S. Y. (2019). HNADOCK: a nucleic acid docking server for modeling RNA/DNA–RNA/DNA 3D complex structures. *Nucleic Acids Research*, *47*(W1), W35-W42. doi:10.1093/nar/gkz412

125. Lang, P. T., Brozell, S. R., Mukherjee, S., Pettersen, E. F., Meng, E. C., Thomas, V., ... & Kuntz, I. D. (2009). DOCK 6: Combining techniques to model RNA–small molecule complexes. *Rna*, *15*(6), 1219-1230. doi:10.1261/rna.1563609

126. Corbeil, C. R., Englebienne, P., & Moitessier, N. (2007). Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *Journal of chemical information and modeling*, *47*(2), 435-449. doi:10.1021/ci6002637

127. Wu, G., Robertson, D. H., Brooks III, C. L., & Vieth, M. (2003). Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm. *Journal of computational chemistry*, *24*(13), 1549-1562. doi:10.1002/jcc.10306

128. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an

empirical binding free energy function. *Journal of computational chemistry*, *19*(14), 1639-1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

129. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2), 455-461. doi:10.1002/jcc.21334
130. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, *267*(3), 727-748. doi:10.1006/jmbi.1996.0897
131. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*, *47*(7), 1739-1749. doi:10.1021/jm0306430
132. Abagyan, R., Totrov, M., & Kuznetsov, D. (1994). ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of computational chemistry*, *15*(5), 488-506. doi:10.1002/jcc.540150503
133. Jain, A. N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry*, *46*(4), 499-511. doi:10.1021/jm020406h
134. Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, *9*(1), 1-9. doi: 10.1186/1741-7007-9-71
135. Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*(6), 1129-1143. doi:10.1016/j.neuron.2018.08.011
136. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of computational chemistry*, *25*(9), 1157-1174. doi:10.1002/jcc.20035
137. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), 187-217. doi:10.1002/jcc.540040211
138. Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., ... & van Gunsteren, W. F. (2005). The GROMOS software for biomolecular simulation: GROMOS05. *Journal of computational chemistry*, *26*(16), 1719-1751. doi:10.1002/jcc.20303
139. Nelson, M. T., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L. V., Skeel, R. D., & Schulten, K. (1996). NAMD: a parallel, object-oriented molecular dynamics program. *The International Journal of Supercomputer Applications and High Performance Computing*, *10*(4), 251-268. doi:10.1177/109434209601000401

140. ViennaRNA Web Services. Available online: http://rna.tbi.univie.ac.at/ (Accessed on 13 October 2022).
141. RNAComposer. Available online: https://rnacomposer.cs.put.poznan.pl/ (Accessed on 13 October 2022).
142. Antczak, M., Popenda, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., ... & Szachniuk, M. (2016). New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochimica Polonica*, *63*(4), 737-744.
143. Dassault Systèmes BIOVIA . *Discovery Studio Modeling Environment, Release 2021*. Dassault Systèmes; San Diego, CA, USA: 2021.
144. HNADOCK Server. Available Online: http://huanglab.phys.hust.edu.cn/hnadock/ (Accessed on 10 April 2023).
145. MDWEB. Available Online: https://mmb.irbbarcelona.org/MDWeb/ (Accessed on 3 July 2023).
146. Schluep, T., & Cooney, C. L. (1998). Purification of plasmids by triplex affinity interaction. Nucleic acids research, 26(19), 4524-4528. doi:10.1093/nar/26.19.4524
147. Nadal, A., Coll, A., Aviñó, A., Esteve, T., Eritja, R., & Pla, M. (2006). Efficient sequence-specific purification of Listeria innocua mRNA species by triplex affinity capture with parallel tail-clamps. ChemBioChem, 7(7), 1039-1047. doi:10.1002/cbic.200500519
148. Bai, J. S., Bai, S., Shi, Q. H., & Sun, Y. (2014). Purification of supercoiled plasmid DNA from clarified bacterial lysate by arginine-affinity chromatography: Effects of spacer arms and ligand density. *Journal of separation science*, *37*(12), 1386-1395. doi:10.1002/jssc.201400092
149. Carugo, O., & Pongor, S. (2001). A normalized root-mean-spuare distance for comparing protein three-dimensional structures. *Protein science*, *10*(7), 1470-1473. doi:10.1110/ps.690101.
150. Liebeschuetz, J. W., Cole, J. C., & Korb, O. (2012). Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *Journal of computer-aided molecular design*, *26*, 737-748. doi:10.1007/s10822-012-9551-4
151. Ghosh, R., Chakraborty, A., Biswas, A., & Chowdhuri, S. (2021). Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors–an in silico docking and molecular dynamics simulation study. *Journal of Biomolecular Structure and Dynamics*, *39*(12), 4362-4374. doi:10.1080/07391102.2020.1779818
152. Kumar, Y., Singh, H., & Patel, C. N. (2020). In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *Journal of infection and public health*, *13*(9), 1210-1223. doi:10.1016/j.jiph.2020.06.016

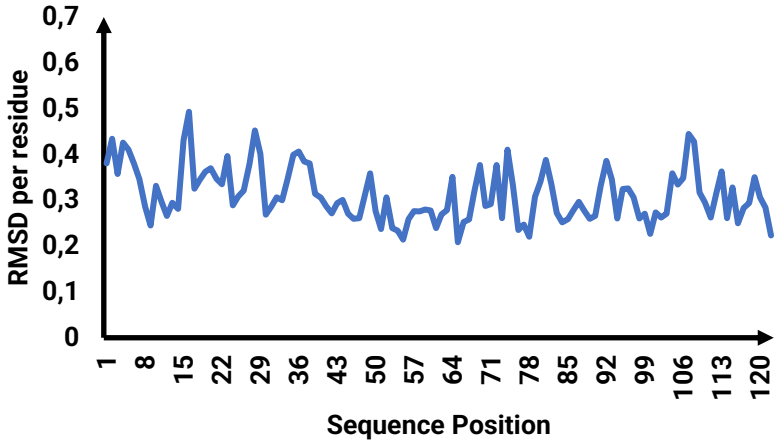
153. Prajapati, J., Patel, R., Goswami, D., Saraf, M., & Rawal, R. M. (2021). Sterenin M as a potential inhibitor of SARS-CoV-2 main protease identified from MeFSAT database using molecular docking, molecular dynamics simulation and binding free energy calculation. *Computers in Biology and Medicine*, *135*, 104568. doi:10.1016/j.compbiomed.2021.104568

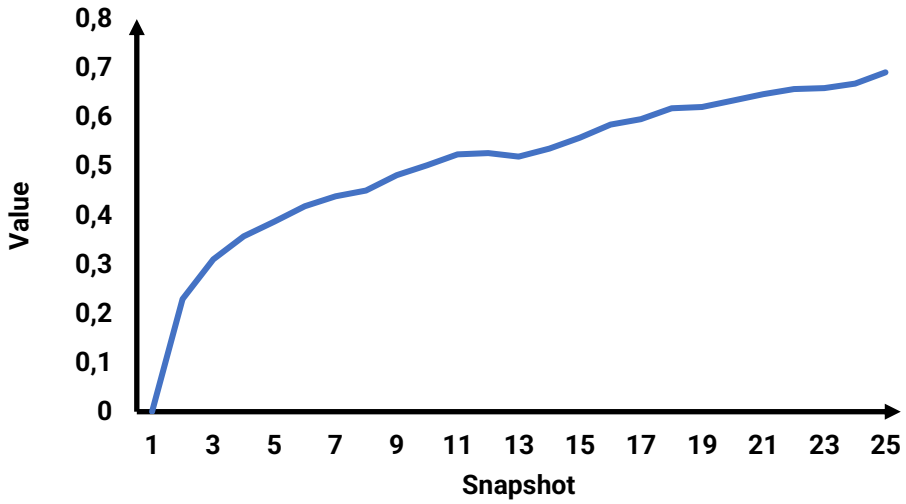
154. Carugo, O. (2018). How large B-factors can be in protein crystal structures. *BMC bioinformatics*, *19*, 1-9. doi:10.1186/s12859-018-2083-8

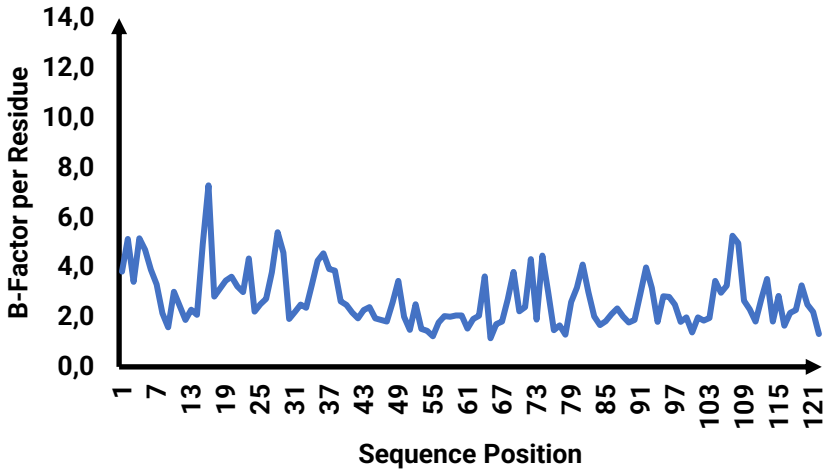
155. Carugo, O. (2018). Atomic displacement parameters in structural biology. *Amino Acids*, *50*(7), 775-786. doi:10.1007/s00726-018-2574-y

Annex

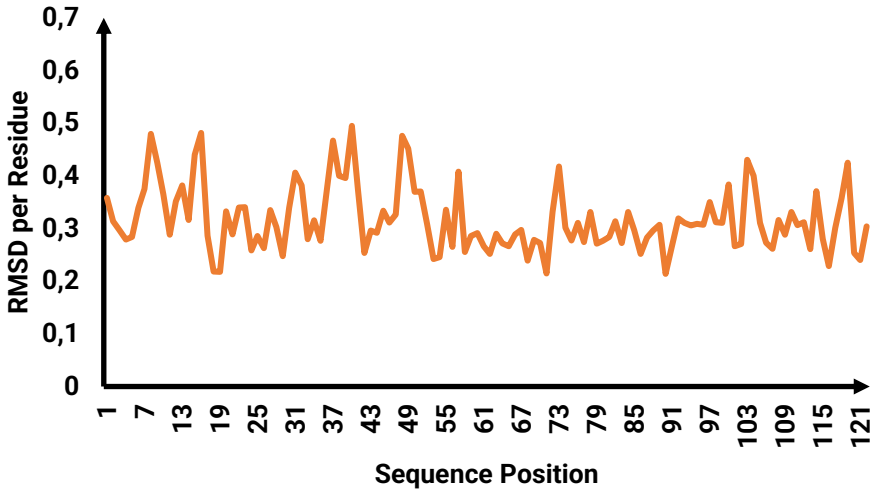
RMSD per residue, RMSD trajectory and B-factor for O3'_15_C12 Model 1.

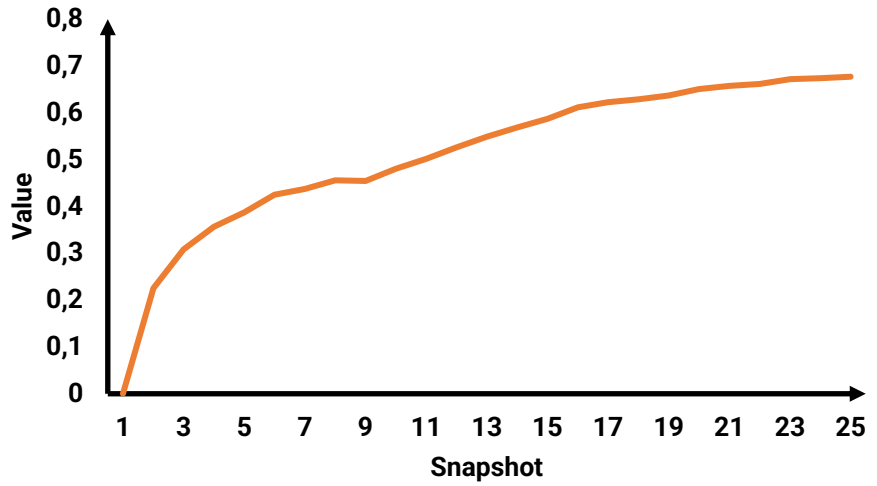


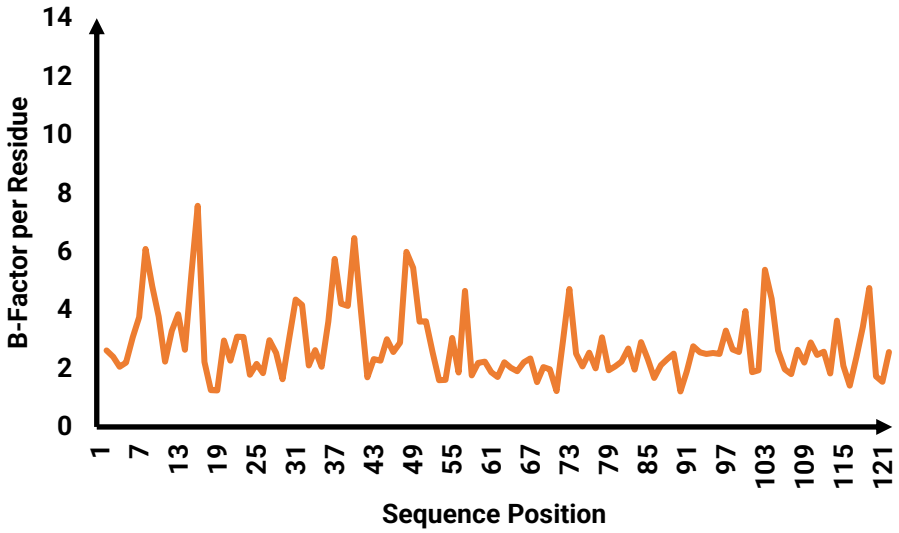




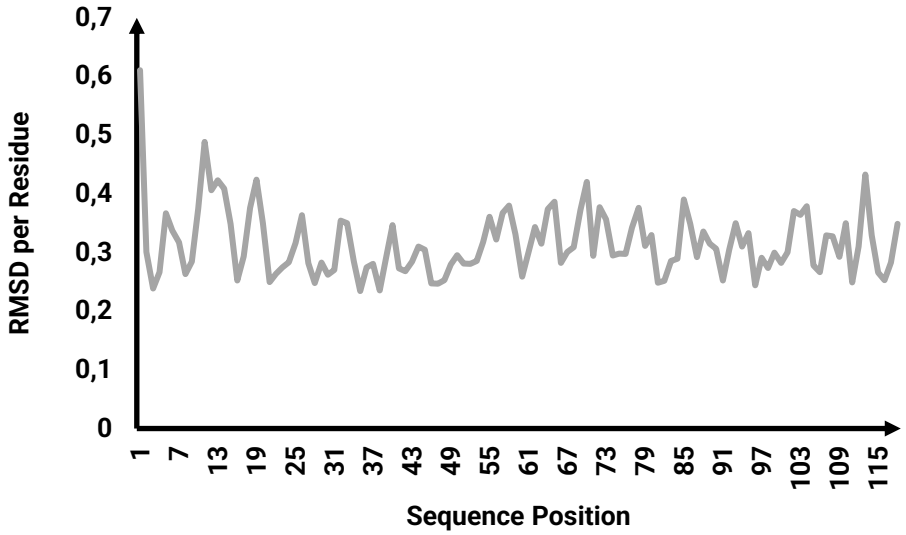
RMSD per residue, RMSD trajectory and B-factor for O3'_15_C12 Model 2.

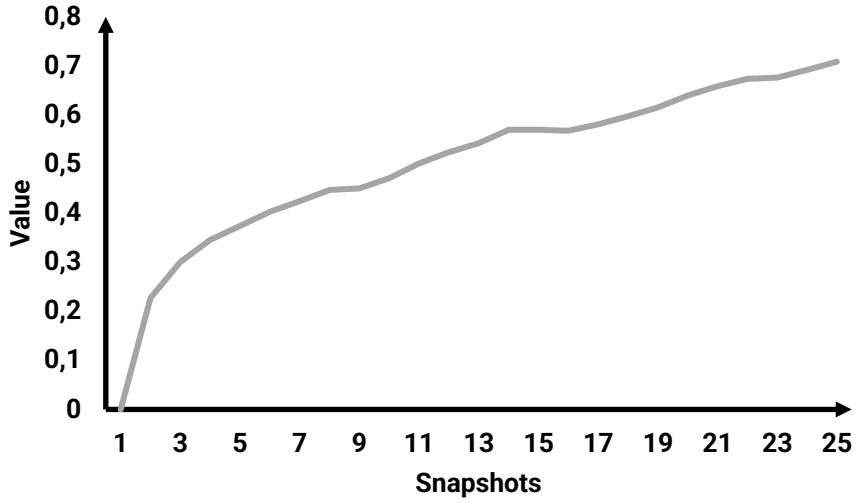


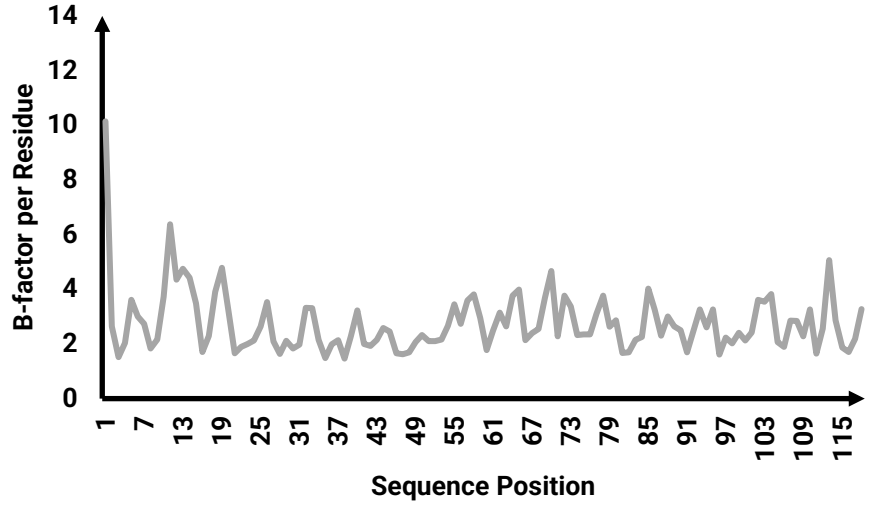




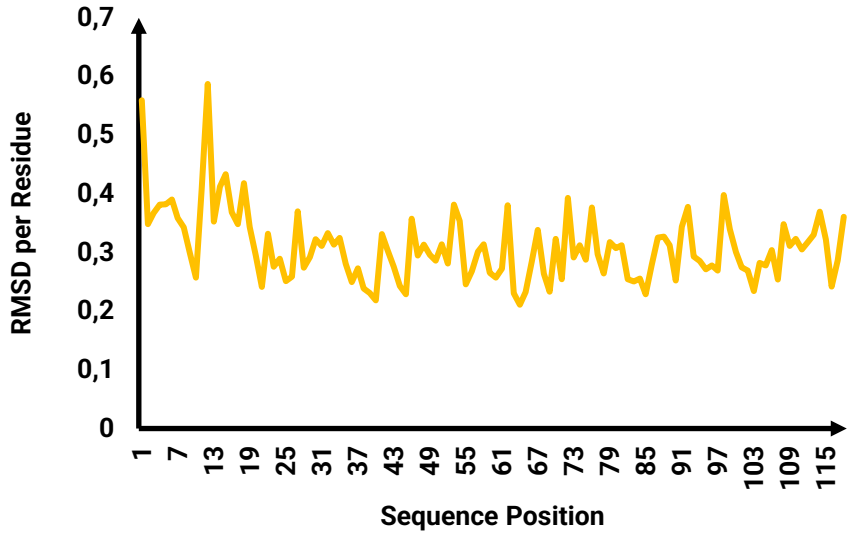
RMSD per residue, RMSD trajectory and B-factor for O5'_11_C6 Model 1.

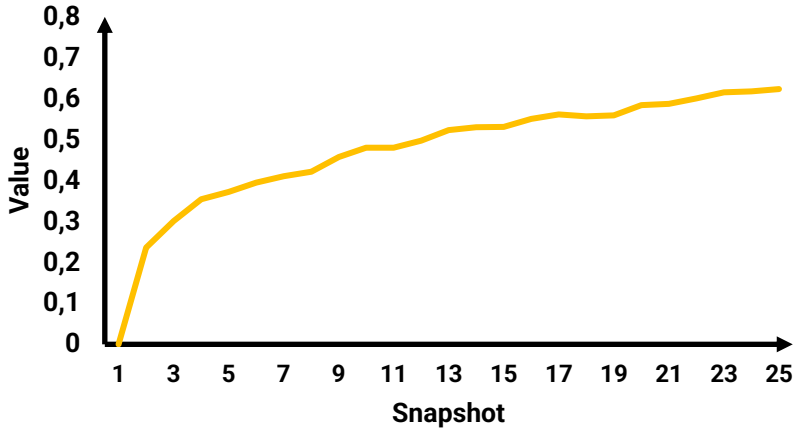


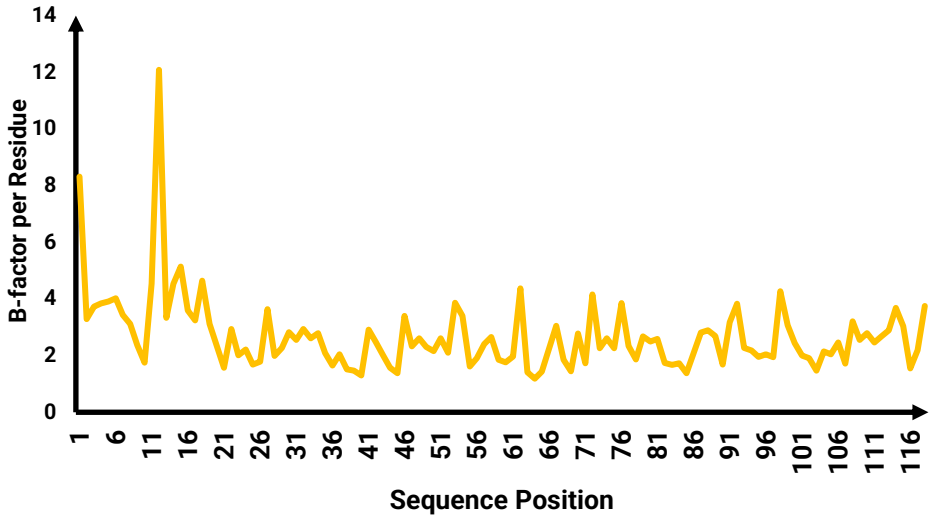




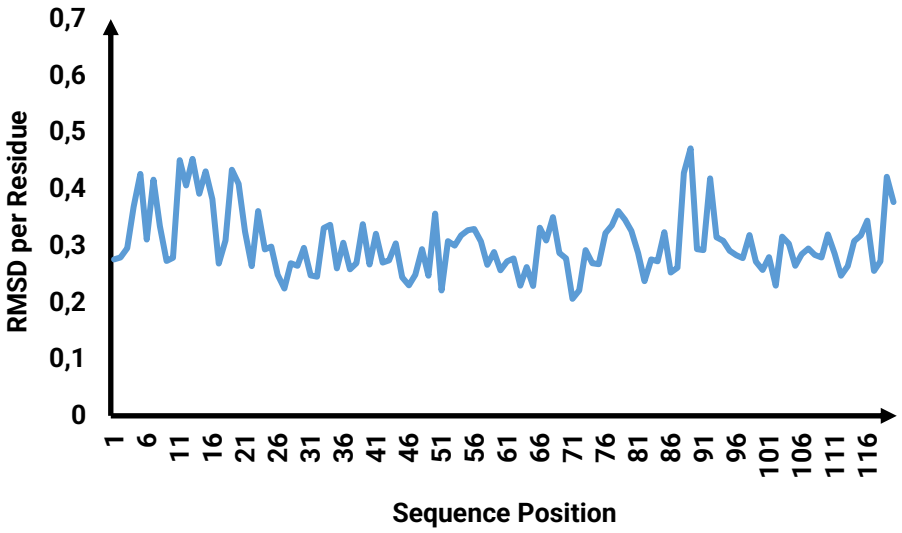
RMSD per residue, RMSD trajectory and B-factor for O5'_11_C**12** Model 1.

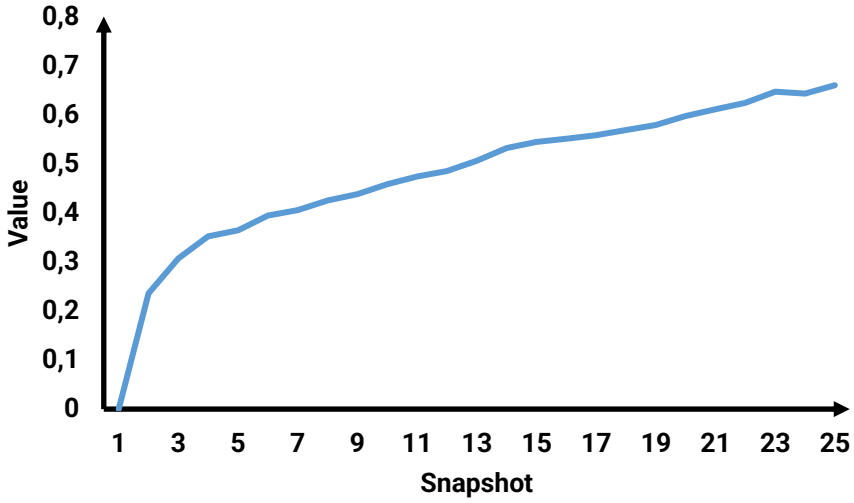


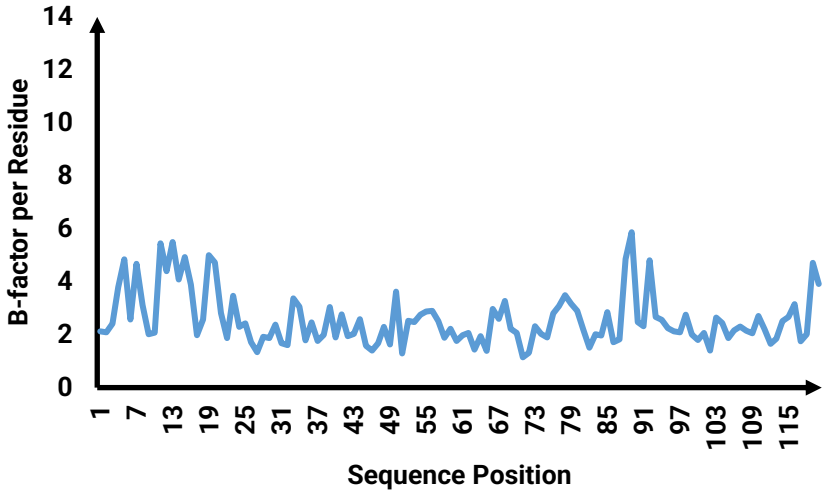




RMSD per residue, RMSD trajectory and B-factor for OHP_13_C12 Model 1.







RMSD per residue, RMSD trajectory and B-factor for OHP_13_C12 Model 4.