UNIVERSIDADE DA BEIRA INTERIOR
Ciências

Development and optimization of a microRNA purification technology by affinity chromatography, using the concept of QbD

Adriana Alexandra Fernandes Afonso

Dissertação para obtenção do Grau de Mestre em

Biotecnologia

(2º ciclo de estudos)

Orientadora: Profª. Doutora Fani Sousa
Coorientadora: Profª. Doutora Ângela Sousa

Covilhã, junho de 2013

"All our science, measured against reality, is primitive and childlike - and yet it is the most precious thing we have."

Albert Einstein

To my parents…

For all their love, support and encouragement.

I love you both.

Acknowledgments

First of all, I would like to express my sincere gratitude to my major supervisor Professor Doctor Fani Sousa and co-major supervisor Professor Doctor Ângela Sousa, for all the guidance, valuable advice and trust. I really appreciate all the dedication, the vast knowledge and scientific expertise, as well as all the criticisms and suggestions made during the development of this work.

I would like to express my sincere gratitude to Professor Doctor João Queiroz and to University of Beira Interior for the availability and provision of the necessary resources essential to the development of this research project.

I am extremely thankful to Patrícia Pereira, who with her special expertise provided me with invaluable advice, encouragement, and careful guidance. Your patience, support and friendship were crucial for the success of this work, I sincerely thank you.

I would also like to express a special thanks to Susana Ferreira who generously helped me on the analysis of my experimental design model. It was essential to rightly explore this field and to draw the right conclusions. I really appreciate her availability and sympathy whenever I contacted her.

Moreover, I also thank to all the people involved in Health Sciences Research Centre of the University of Beira Interior, especially to the Biotechnology and Biomolecular Sciences group for providing invaluable technical and administrative help, advice, support and feedback, particularly to Rita Martins and Marta Silva. Thanks for all the great moments, friendliness and support on the not so good times.

My heartfelt thanks to my friends, I appreciate your encouragement, support, and assistance. I will never forget those hours when we talked about our difficulties, and despite that we laugh about it. Your companionship made difficult times enjoyable.

To all my family, thanks for never give up on believing me, for understand my absence and for all of your kindness even when you can't understand what I am doing so many hours closed in the lab. For you acceptance I thank you.

For their endless encouragement, patience, support, love and because they are who made all of this achievable, to my parents, Mário e Zita Afonso, I am eternally thankful to you both.

To my brother, because I could not be around when you need me the most, thanks for understand. I love you.

Finally, I want to thank to Tiago, for all his support and availability whenever I needed. You kept me on the surface on the difficult times and you have shown me that when I want it enough, strength will come with the will. Thank you my love.

Resumo

Atualmente, mais de 30 milhões de pessoas sofrem de demência em todo o mundo e, sem tratamento visível num futuro próximo, os números tendem a aumentar rapidamente. De todas as doenças neurodegenerativas, a doença de Alzheimer (DA) é a mais predominante. Na Doença de Alzheimer há uma perda neuronal e sináptica generalizada que causa um declínio progressivo na memória e em outras funções cognitivas, conduzindo finalmente à demência. Esta doença neurodegenerativa está associada a dois tipos de depósitos proteicos, placas extracelulares ß-amilóides (Aß) e complexos neurofibrilares intraneuronais. Estudos recentes demonstraram que o cluster miR-29 pode diminuir a deposição das placas de Aß atuando diretamente sobre a enzima que cliva os péptidos ß-amilóides (BACE-1), a qual é responsável pela clivagem da proteína precursora e, consequentemente geração de espécies Aß tóxicas. Estes estudos apresentam a possibilidade da utilização do pré-miR-29 como um novo alvo terapêutico no tratamento da DA, para qual é necessário garantir um elevado grau de pureza, estabilidade e integridade do miRNA. Assim, o principal objetivo deste estudo é o desenvolvimento de uma nova estratégia de cromatografia de afinidade com uma matriz O-fosfo-L-tirosina comparando duas metodologias diferentes – o desenho experimental e o método "tentativa/erro" - para obter o pré-miR29 com elevado grau de pureza e rendimento, visando a sua aplicação em terapia génica.

Palavras-chave

Cromatografia de afinidade, desenho experimental, método "tentativa/erro", pre-miRNA29b-1, purificação.

Resumo Alargado

Desde a descoberta de RNAs com actividade catalítica no início de 1980 até á investigação de pequenos RNAs de interferência, a compreensão biológica do RNA evoluiu de um simples intermediário entre o DNA e uma proteína, para uma molécula dinâmica e versátil que regula as funções dos genes e das células em todos os organismos.

Os pequenos RNAs não-codificantes representam um grande grupo de diferentes RNAs, que foram identificados em plantas e animais. Inicialmente, foram descobertos em *Caenorhabditis elegans*, onde silenciavam genes específicos, mas atualmente sabe-se que eles atuam no silenciamento de outros RNAs, reorganizam o DNA e atuam na regulação pós-transcricional do RNA mensageiro através de microRNAs endógenos. Particularmente, as moléculas de microRNA (miRNAs) são constituídas por duplas cadeias, cada uma com 15 a 21 nucleótidos que, ao se ligarem ao RNA mensageiro alvo, silenciam a tradução ou induzem a sua degradação. Devido às suas funções e ao facto de a sua expressão aparecer desregulada em várias doenças, os miRNAs representam um importante alvo terapêutico.

Atualmente, mais de 30 milhões de pessoas sofrem de demência em todo o mundo e, sem tratamento visível num futuro próximo, os números tendem a aumentar rapidamente. Está previsto que em 2030 o número de pessoas afetadas suba para 66 milhões e que em 2050 este ronde já os 115 milhões de pessoas. De todas as doenças neurodegenerativas, a doença de Alzheimer (DA) é a mais predominante sendo a principal causa de demência em pessoas idosas. O aparecimento exponencial destas doenças deve-se principalmente ao envelhecimento da população, à falta de progresso na correta identificação da doença e à falta de técnicas preditivas de diagnóstico. Este tipo de doenças representa um grande encargo para a sociedade, quer a nível económico como social, tornando-se urgente encontrar uma solução para este problema.

O desenvolvimento de estratégias terapêuticas para a DA requerem o conhecimento das principais causas e efeitos desta doença. Os pacientes apresentam uma perda neuronal e sináptica generalizada que causa um declínio progressivo na memória e em outras funções cognitivas, conduzindo finalmente à demência. Os sintomas acima referidos são maioritariamente provocados pela toxicidade de dois tipos de depósitos proteicos, placas extracelulares ß-amiloides (Aß) e complexos neurofibrilares intraneuronais. Estes últimos resultam de uma forma fosforilada da proteína tau, enquanto que as placas ß-amiloides resultam da clivagem da proteína precursora amiloide (APP) que origina os pépidos-Aß que podem agregar.

Recentemente, estas características patológicas foram reconhecidas como estando ligadas direta ou indiretamente à regulação da APP, da enzima que cliva os péptidos ß-amiloides

(BACE-1) e de genes fundamentais na evolução da doença de Alzheimer. Particularmente, estes avanços indicaram que em pacientes da DA a expressão da família mir-29 (miR29a, miR29b, miR29b e miR29c) está diminuída e que os níveis de BACE1 estão aumentados. Hebbert e a sua equipa de investigadores (2008) demonstraram que o cluster miR29a/b-1 assume um papel importante na evolução da doença, dado que baixos níveis do cluster miR29a/b-1 implicavam altos níveis da proteína BACE1. Estes estudos apresentaram a possibilidade da utilização do pré-miR-29 como um novo alvo terapêutico no tratamento da DA, sendo necessário garantir um elevado grau de pureza, estabilidade e integridade do miRNA. Assim, torna-se indispensável desenvolver um método de purificação fiável, específico e reprodutível, para simplificar o processo utilizado e aumentar a eficiência na obtenção de um miRNA específico.

Recentemente, a cromatografia de afinidade com aminoácidos tem sido estudada como uma alternativa aos métodos clássicos (métodos baseados no uso de fenol e/ou matrizes de sílica) para isolar o RNA, uma vez que permitem interações específicas com este ácido nucleico. Esta técnica de afinidade baseia-se na multiplicidade de interações naturais que ocorrem entre os aminoácidos e os ácidos nucleicos no ambiente celular. A seletividade deste método pode ser explicada pelo reconhecimento da estrutura individual envolvida na interação específica aminoácido/ácido nucleico.

Como está descrito na literatura, o reconhecimento da molécula de miRNA é feito por um complexo enzimático envolvido no processo de silenciamento – complexo RISC. Este apresenta uma cavidade aminoacídica conservada que possui no seu interior uma tirosina, crucial para que o reconhecimento e a ligação entre o miRNA e o complexo RISC ocorra. Assim, o objetivo deste estudo passa por explorar esta interação natural, utilizando uma matriz de agarose O-fosfo-L-tirosina para purificar a molécula de pré-miRNA29b-1 tendo em conta todo o seu potencial uso na terapêutica da doença de Alzheimer.

O desenvolvimento de métodos cromatográficos para a purificação de ácidos nucleicos é muitas vezes complexo e demorado. Geralmente, a otimização de um método é feita de acordo com as informações recolhidas após cada experiência realizada, baseando-se nos erros ou sucessos (método "tentativa/erro") de cada ensaio. No entanto, o desenvolvimento de um processo cromatográfico não é apenas caracterizado pelos rendimentos e pela pureza obtidos, mas também tem que ter em conta a robustez do método e todos os constrangimentos económicos e ecológicos envolvidos. É necessário maximizar a eficiência dos ensaios realizados, para diminuir o desperdício e o custo. Os investigadores procuram realizar experiências que proporcionem a maior quantidade e pureza possível da molécula alvo com o menor número de ensaios realizados. Para tal, a utilização do desenho experimental em ambiente laboratorial surge como uma ferramenta importante para encontrar o melhor método de purificação.

Neste trabalho é então proposta a utilização do desenho experimental, seguindo o desenho Box-Behnken, no desenvolvimento de um método cromatográfico para purificar o pre-miR29b-1, utilizando a matriz O-phospho-L-tyrosine – explorando as interações naturais que ocorrem num ambiente celular. Por outro lado, também são discutidos os benefícios desta abordagem em comparação com a metodologia "tentativa/erro".

Abstract

Nowadays, over 30 million people suffer from dementia worldwide and, with no cure in sight, these numbers tend to increase very fast. Of all dementing disorders, Alzheimer's disease (AD) is the most predominant. In AD there is a widespread synaptic and neuronal loss that causes a progressive decline in memory and other cognitive functions ultimately leading to dementia. This neurodegenerative disease is associated with two types of protein deposits, extracellular amyloid-ß (Aß) plaques and intraneuronal neurofibrillary tangles (NFTs). Recent studies demonstrated that the miR-29 cluster can decrease the production of the Aß plaques by acting directly on the ß-amyloid cleavage enzyme (BACE-1), which is critical for the cleavage of amyloid precursor protein and the generation of toxic Aß species. These studies raised the possibility of using pre-miR-29 as a novel target for therapeutic intervention on AD, being essential to guarantee the purity, stability and integrity of the miRNA. Hence, the main goal of this study is the development of a new affinity chromatographic strategy with an O-phospho-L-tyrosine matrix, comparing two different methodologies – experimental design versus "one at a time" method - to obtain the pre-miR29 with high purity degree and yield, envisioning their application in gene therapy.

Keywords

Affinity chromatography, experimental design, "one at a time" method, pre-miRNA29b-1, purification.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AD	Alzheimer's Disease
Aß	ß-amyloid peptides
ADDLs	Amyloid ß-Derived Diffusible Ligands
AGO	Argonaute
APP	Amyloid precursor protein
BACE1	ß-site APP-cleaving enzyme 1 or ß-secretase
BBD	Box-Behnken design
bp	Base pair
casiRNA(s)	Cis-acting siRNA(s)
CCD	Central composite design
C. elegans	*Caenorhabditis elegans*
CI	Confidence interval
cm	Centimeter
DEPC	Diethyl pyrocarbonate
DM	Doehlert design
DNA	Deoxyribonucleic acid
DoE	Design of experiments
dsRNA	Double-stranded RNA
EDTA	Ethylenediaminetetraacetic acid
g	Gram
G	Gravitational constant
GC	Gas chromatography
GC-FID	Gas chromatography with flame ionization detector
GS-MS	Gas chromatography–mass spectrometry
h	Hour
HPLC	High-pressure liquid chromatography
kb	Kilobase
LC-MS	Liquid chromatography – mass spectometry
Lin-4	Lineage-4
Lin-14	Lineage-14
M	Molar
mA	Miliamper
min	Minute
miR-29	MicroRNA29

miRNA(s)	MicroRNA(s)
mL	Milliliter
mM	Milimolar
mRNA	Messenger RNA
MWCO	Molecular Weight Cut Off or MWCO
natsiRNA(s)	Natural antisense transcript derived siRNA(s)
NFTs	Neurofibrillary tangles
ng	Nanogram
nm	Nanometer
nt	Nucleotides
OD_{600}	Optical density at 600 nm
OVAT	One-variable-at-a-time
PC	Paper Chromatography
PCR	Polymerase chain reaction
pDNA	Plasmid DNA
piRNA(s)	Piwi-interacting RNA(s)
P-Tyr	O-Phospho-L-Tyrosine
rasiRNA(s)	Repeat-associated siRNA(s)
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNase(s)	Ribonuclease(s)
rRNA(s)	Ribosomal RNA(s)
rpm	Revolutions per minute
R. sulfidophilum	*Rhodovulum sulfidophilum*
s	Second
sc pDNA	Supercoiled pDNA isoform
scnRNA(s)	Small-scan RNA(s)
shRNA	Short hairpin RNAs
siRNA(s)	Small interfering RNA(s)
sRNA(s)	Small RNA(s)
SPE	Solid-phase extraction
TAE	Tris-acetate-EDTA
tasiRNA(s)	Trans-acting siRNA(s)
TBE	Tris-Borate-EDTA
tmRNA	Transfer-messenger RNA
TRBP	TAR RNA-binding protein 2, also known as TARBP2
Tris	Tris(hydroximethyl)methylamine
tRNA(s)	Transfer RNA(s)

Tyr	Tyrosine
U	Enzyme activity unit
UV	Ultraviolet
µL	Microlitre
µmol	Micromol
V	Volt

Chapter 1

Introduction

Small non-coding RNAs represent a large group of different RNAs, that have been identified in plants and animals [1]. At first, Fire and co-workers (1998) found these molecules on *Caenorhabditis elegans* where they silenced specific genes; but today it is known that they act in order to silence RNA, silence interference RNA-mediated chromatin, rearrange DNA and act on post-transcriptional regulation of messenger RNA (mRNA) by endogenous microRNA (miRNA) [2, 3].

Particularly, miRNAs are 15-21 base duplexes that bind partially to the complementary sequences in target mRNA, silencing the translation or inducing its degradation [4]. Due to their functions, miRNAs represent a therapeutic target, reflected on the number of papers that discuss this theme.

Currently, it is known that over 30 million people suffer from dementia worldwide [5] and that by 2030, this number can reach 66 million people and increase to 115 million by 2050. Within these numbers, Alzheimer's disease (AD) is the most representative of dementia disorders, being the leading cause of dementia in the elderly people [6]. These numbers are mainly due to the aging of the population, the continuing lack of progress in identifying effective treatment modalities and the lack of predictive diagnostic techniques [7]. Actually, it was estimated that in 2010, the worldwide cost of dementia to society was US$604 billion/year, which clearly shows the enormous impact on socioeconomic conditions. Hence, it is of most importance to found a solution to this problem as soon as possible [7, 8].

In order to develop an AD treatment is essential to understand the principal features of this disease. Alzheimer's patients demonstrate two main characteristic protein deposits: tau neurofibrillary tangles (NFTs) and ß-amyloid plaques. While the NFTs are a hyper-phosphorylated form of tau, the ß-amyloid plaques result from proteolytic cleavage of amyloid precursor protein (APP) that origin Aß-peptides, which can aggregate [9]. Currently, these pathological features have been recognized as linked to deregulated microRNAs (present on AD brains) [10]. Most of these molecules are direct or indirectly implicated in the regulation of key genes involved in AD, including APP and as ß-site APP-cleaving enzyme 1 (BACE1) [9]. Several reports indicate that the expression of miR-29 family (miR-29a, miR-29b, and miR-29c) is altered in AD patients, being significantly decreased, and displaying abnormally high levels of BACE1 protein [9].

Particularly, the work of Hebert and co-workers (2008) showed that the decreased levels of the miR-29a/b-1 cluster implicated increased levels of BACE1 [11]. This finding raised the chance to use pre-miRNA29 as a therapeutic target. This way, it is essential to find an effective strategy to obtain pre-miRNA29 and to guarantee the purity, stability and integrity of the molecule.

The purification method has to be as reliable and specific as possible, in order to simplify the process used to obtain a specific miRNA. Recently, amino acid-based affinity chromatographic methods have been studied as an alternative to the classic methods (phenol and/or silica matrix based methods) to isolate RNA, as they allow specific nucleic acid purification [12-14]. This affinity technique is based on the exploitation of natural interactions that occur between amino acids and nucleic acids in the cellular environment. As it is described in the literature, miRNA recognition is made by an enzymatic RNA-induced silencing complex (RISC) that presents a conserved amino acidic pocket. In particular, it has been suggested that a conserved tyrosine of this pocket is crucial to induce the recognition and binding between RISC and miRNA. So, the aim of this study is to explore this natural interaction, using an O-phospho-L-tyrosine agarose matrix to purify pre-miRNA29b-1 regarding its potential use as Alzheimer's therapeutics.

Taking into consideration the aforementioned, the present work can be divided into three main phases:

1. Extraction of small RNAs from *Rhodovulum sulfidophilum* DSM 1374;
2. Recovery of pre-miRNA29b-1 from the small RNAs samples;
3. Optimization of the purification step using different methods (one at a time and Box Behnken design).

Section 1.1 - Interference RNA / small RNA

Since the discovery of catalytic RNAs in the early 1980s and of RNA interference in the late 1990s, biological understanding of RNA has evolved from simply an intermediate between DNA and protein to a dynamic and versatile molecule that regulates the function of genes and cells in all living organisms [15].

Fire and co-workers (1998) have found that double strand RNA (dsRNA) introduced into *Caenorhabditis elegans* (*C. elegans)* silenced the expression of a homologous target gene more efficiently than the corresponding antisense RNA [3]. On another study performed by Zamore and co-workers (2000), they obtained the same interference response *in vitro* when long dsRNA was added to a Drosophila embryo extract, silencing the expression of a homologous gene by directing degradation of its mRNA [16, 17]. With the breakthrough discovery that genes could be silenced by interference RNA (RNAi), these small RNAs rapidly became the focus of studies on the mechanisms of their specific regulatory roles [1]. These shorter dsRNA found by Zamore and co-workers (2000) were named "small interference RNAs" (siRNAs). Similarly, they discovered small RNAs *in vivo* in Drosophila cells transfected with long dsRNA and in fly embryos and *C. elegans* injected with long dsRNA [16, 17].

With 19–31 nucleotides (nt) long, these non-coding RNAs behave as sequence-specific triggers for mRNA degradation, translation repression, heterochromatin formation and transposon control. They can be classified into different groups based on their origin or the components to which they are coupled: microRNAs (miRNAs), small-interfering RNAs (siRNAs), trans-acting siRNAs (tasiRNAs), small-scan RNAs (scnRNAs), cis-acting siRNAs (casiRNAs), natural antisense transcript derived siRNAs (natsiRNAs), repeat-associated siRNAs (rasiRNA) and Piwi-interacting RNAs (piRNAs) (Table 1) [1].

Briefly, miRNAs are 19 to 25 nt long and derived from hairpin-structured precursor, which facilitate translation repression in several organisms. SiRNAs are the products of long Dicer-processed dsRNAs that silence genes by cleaving their target mRNAs. They have been widely exploited as an exogenous tool to study gene function by acting over the mRNA in mammalian cells. They also mediate the innate immune response against RNA viral infection in different organisms [1]. Endo-siRNAs are composed by cis-acting siRNAs (casiRNAs), tasiRNas and natsiRNAs and induce post-transcriptional regulation of transcripts and transposons and transcriptional gene silencing [18]. TasiRNAs trigger mRNA degradation but function in trans, that is, their target transcripts are other than the original genes where tasiRNAs are derived [1, 18]. In response to stress (such high salt levels), plants and *Arabidosis* species produced natsiRNAs. On another hand the cis-acting siRNA comprise the bulk of endo-siRNAs, being 24 nt long that promote the heterochromatin formation by directing DNA methylation and histone modification at the loci from which they originate [18]. ScnRNAs were found in *Tetrahymena thermophila* acting in histone methylation, DNA elimination and also inducing

genome rearrangement [1]. Longer than siRNAs, rasiRNAs are silence repetitive and mobile genetic elements in yeast, plants, and flies [18].

Table 1: Classes of small RNAs (adapted from [4])

Class	Length (nt)	Function
Small interfering RNA (siRNA)	19-21	Target mRNA cleavage.
MicroRNA (miRNA)	19-25	Regulation of mRNA stability, translational repression.
Trans-acting siRNA (tasiRNA)	21-22	Post-transcriptional regulation, mRNA cleavage.
Natural antisense transcript-derived siRNA (natsiRNA)	21-24	Regulation of stress-response genes.
***Cis*-acting siRNA (casiRNA)**	~ 24	Chromatin modification.
Repeat-associated siRNA (RasiRNA)	24-27	Transposon control transcriptional silencing.
Small-scan RNA (scnRNA)	~ 28	DNA elimination.
Piwi-interacting RNA (piRNA)	26-31	Transposon regulation.

The development of this field of investigation has raised different types of RNA-based therapeutics that extend the range of ‘‘drug-able’’ targets further than the existing pharmacological drugs [15].

Subsection 1.1.1 - MicroRNAs

MicroRNAs are the most studied small non-coding RNA molecules that are involved in post-transcriptional regulation of target genes [19].

Ambros and co-workers (1993) as well as Ruvkun and co-workers (1993) studies revealed that the *Caenorhabditis elegans* lineage-4 (lin-4) gene produces a 21-nucleotides RNA that recognizes complementary sites in lin-14 messenger [20-22]. Accordingly, RNA downregulates the translation of lin-14 during the transition from the first to the second larval stage of development [20-22]. Thus, Ambros and co-workers (1993) settled that larval development of the nematode *C. elegans* needs a small RNA to inhibit the expression of a protein-coding gene

[20]. Now, this little lin-4 RNA is known as the founding member of an abundant class of small regulatory RNAs called microRNAs or miRNAs [22].

MiRNA-directed gene regulation is becoming more important as more miRNAs and their regulatory targets and functions are discovered. There are several roles assigned to miRNA as the control of cell proliferation, cell death, fat metabolism in flies, the neuronal patterning in nematodes, the modulation of hematopoietic lineage differentiation in mammals and the control of leaf and flower development in plants [22].

With 19–25 nt long, mature miRNAs are single stranded RNA molecules derived from a 70–100 nt long hairpin-precursor named pre-miRNA. Usually, RNA polymerase II transcribe miRNAs from chromosomal DNA regions (intragenic or intergenic) into primary transcripts of various lengths (usually 1–3 kb) called pri-miRNAs (Figure 1) [19]. The resulting pri-miRNA has his hairpin recognized by an RNase complex consisting in a RNase III-type enzyme Drosha and its cofactor DiGeorge syndrome critical region gene 8 (DGCR8) that cleaves each hairpin at ~11 nt from its base [19, 23]. This step releases an approximately 70 nt long stem-loop (hairpin structure) precursor of miRNA (pre-miRNA). The pre-miRNA is then transported from the nucleus to the cytoplasm [4, 19]. Once in the cytoplasm, the endoribonuclease protein Dicer – a different RNAse III enzyme – with its co-factors TRBP (TAR RNA-binding protein 2, also known as TARBP2) cleaves off the terminal loop, resulting in a mature double-stranded miRNA of variable length (~20–25 nt) [4, 19, 23].

After Dicer cleavage, the short RNA duplex is loaded into a multisubunit complex termed RISC (RNA-induced silencing complex) by an Argonaute (AGO) protein. There, one of the two strands is released and degraded – passenger strand (also referred to as miR*), whereas the other strand is incorporated into RISC - guide strand or miR [4]. This strand selection might be determined by the relative stability of the two ends of miRNA duplexes, thus, the strand less stable at the 5' end is loaded into RISC. The miRNA strand is bound to a conserved basic pocket inside the complex RISC that, with four invariant amino-acid residues (Y529, K533, Q545 and K570), is positively charged [1]. After the selection of the guide strand, it targets RISC to mRNAs with partially complementary sequences and cleaves, degrades or suppresses translation of target mRNAs [4, 23]. These sequences are named of 'seed' region and are 2–8 nucleotides from the 5'-end of a miRNA. They are crucial for target recognition; hence, miRNAs that share the seed region but differ outside are frequently considered to form a 'family' of miRNAs with largely overlapping sets of targets [4].

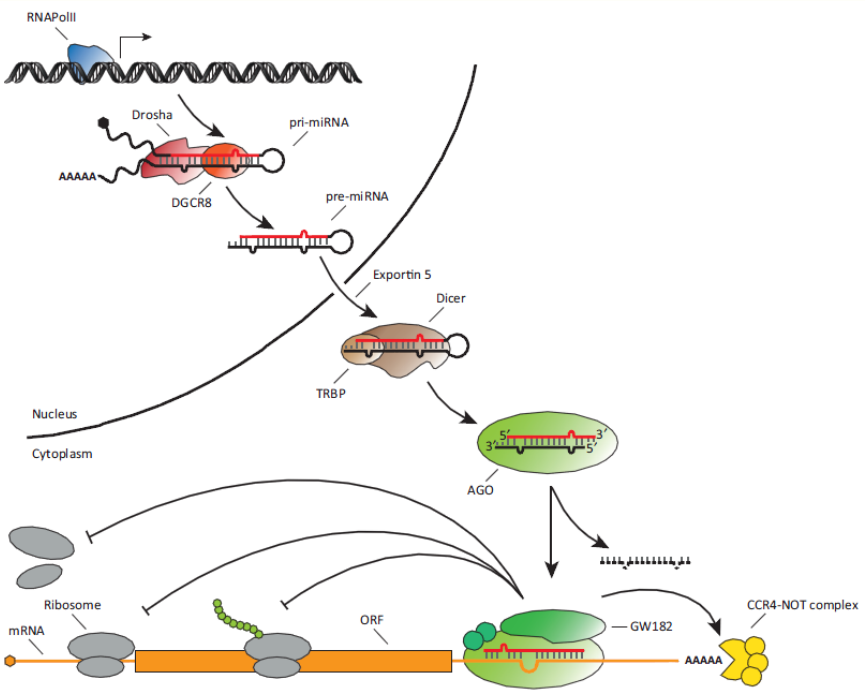


Figure 1: Translation inhibition (adapted from [4]).

Subsection 1.1.2 - MicroRNAs and their potential as therapeutic product

The capacity to inhibit the expression of a specific mRNA is a great advantage that can be used as a therapeutic application to some diseases that have origin or are involved in deregulated levels of mRNA and miRNA or when the prevention of gene expression is required.

Several studies demonstrated that abnormal miRNA expression seems to characterize many diseases, thus, miRNA expression profiles are a potential application as diagnostic and prognostic tools or also therapeutic targets. Most of the existing methods for modulating the de-regulated levels of miRNAs are adapted from existing gene therapy and antisense technology [23].

Currently, there are two types of therapeutic strategies being explored *in vivo*: miRNA mimics and anti-miRNAs. The first strategy is used for miRNAs whose expression is reduced in the disease state, and the re-introduction of the mature miRNA into the proper tissue could provide a therapeutic benefit by restoring regulation of target gene(s). That is, they are used to down-regulate the expression of specific target proteins [19, 23]. So far, it is known that siRNAs resemble the mature miRNA duplex and are functionally interchangeable, relatively to RISC action against target mRNAs. Consequently, the introduction of miRNA-mimics molecules that mimic the Dicer-processed miRNA duplex is a potential method to increase a particular expression (Figure 2b) [23].

However, when the problem comes from the over-expression of a specific type of miRNA, a loss-of-function in miRNA should be created. Injecting a complementary RNA sequence that binds to a miRNA target will achieve its inactivation, blocking its activity. For that role, it was developed the "antagomirs", which are RNA snippets with cholesterol molecules that facilitate the entrance of RNA into the cells (Figure 2a) [19].

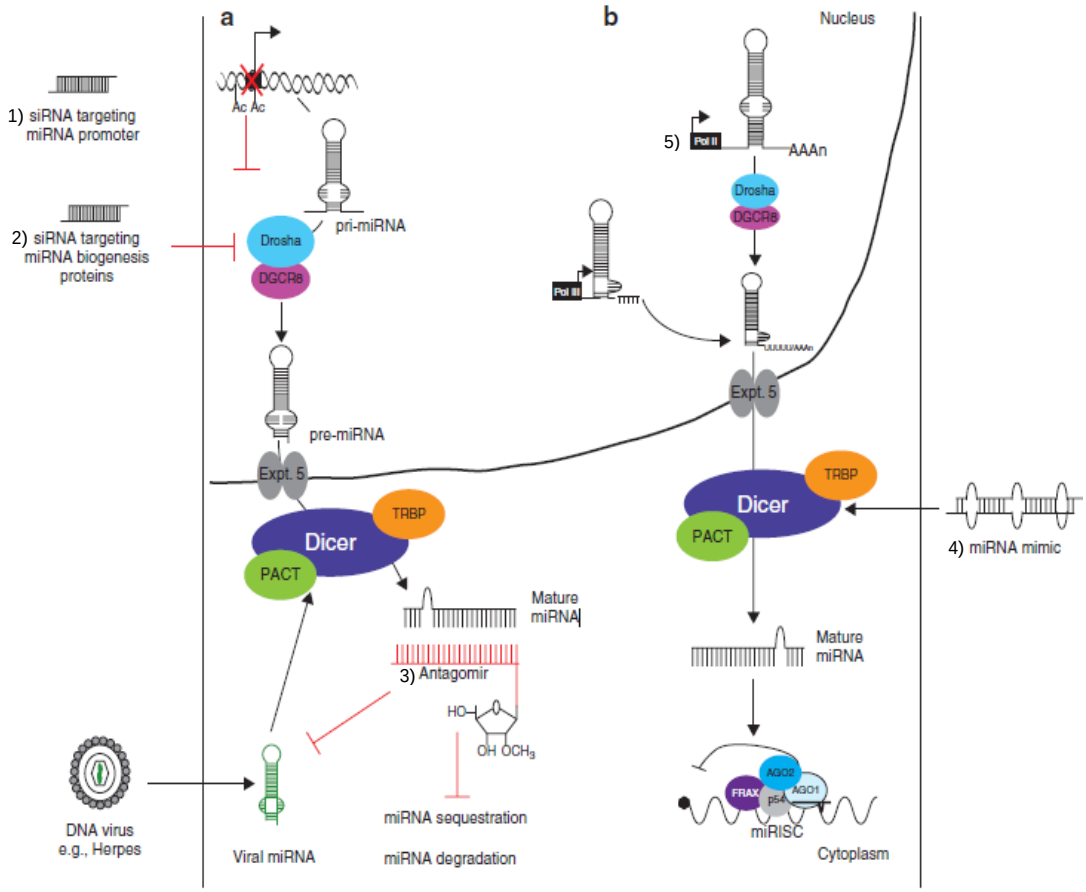


Figure 2: Therapeutic strategies to re-equilibrate miRNA levels. (a) Therapeutic strategies to reduce the miRNA expression. 1) Inhibition of miRNA transcription by introducing siRNA directed against the genomic miRNA promoter which induces chromatin changes, such as histone acetylation (Ac). 2) Introduction of siRNA directed against the miRNA biogenesis components such as Drosha, reducing processed miRNA available for target gene silencing. 3) Antagomirs (modified antisense oligonucleotides), through complementary base pair interactions, can sequester and/or degrade the mature miRNA. They can also act against incoming viral miRNA. (b) Therapeutic strategies to increase miRNA expression. 4) The introduction of a synthetic miRNA mimic into the cytoplasm that can be processed and loaded into RISC by Dicer, can restore miRNA function. 5) RNA Pol II promoter can express an entire pri-miRNA, leaving open the possibility for tissue-specific or induced ectopic miRNA expression [23].

Section 2 – Alzheimer's Disease

Neurodegenerative diseases such Alzheimer's disease (AD), Parkinson's disease, Huntington's disease among other, are becoming increasingly prevalent. Actually, it is estimated that neurodegenerative diseases affect nearly 30 million people worldwide [5]. The prevalence of neurodegenerative diseases is directly linked to the current trends: aging societies resulting from the increased life expectancies which cause more people living long enough to be affected, augmenting significant societal, emotional and economic burdens [24].

Particularly, AD is predicted to increase dramatically over the next 40 year; being a slowly progressing, age-related neurodegenerative disease, it already affects ~2% of the population in industrialized countries [25]. Alzheimer's disease can be termed as Sporadic Alzheimer's disease or Familial Alzheimer's disease. The first type accounts for more than 90% of cases and it can take up to 20 years for the disease to develop. Adults of all ages can be affected, although most cases occur in people over the age of 65 years [26]. Otherwise, familial Alzheimer's disease is a rare type caused by a genetic mutation, which appears sooner than the Sporadic type (40 – 50 years). Generally, it is diagnosed in families that have more than one member affected in more than one generation, and when it is indistinguishable from Sporadic form [26].

In general, Alzheimer's disease affects cholinergic neurons being characterized by the accumulation of plaques formed of short ß-amyloid (Aß) peptides (derived from amyloid precursor protein (APP)), which occur primarily in 40mer and 42mer forms, designated Aß40 and Aß42 [7]) in the hippocampus region of the brain and by deposits of intraneuronal neurofibrillary tangles (NFTs) [5, 27], constituted by insoluble hyperphosphorylated forms of tau, which is a microtubule-associated protein [27]. Aß peptides are produced upon endoproteolytic cleavage of APP by ß-secretase, also known as ß-site APP-cleaving enzyme 1 (BACE1), which contributes to the formation of these plaques [5, 27]. The ß-secretase cuts APP first, producing the N-terminus of Aß and then the γ-secretase cleaves the C-terminal fragment to release the mature Aß peptide. On the other hand, a α-secretase acts within the Aß domain on APP, interrupting Aß formation (Figure 3) [28].

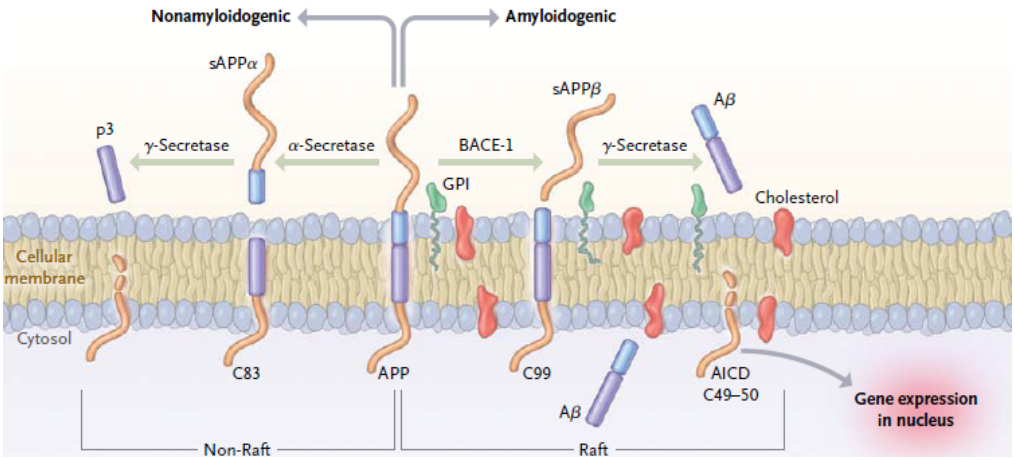


Figure 3: Schematic representation of APP processing [29].

Because γ-secretase produces several Aß peptides with heterogeneous C termini ranging from 38-43 nt long, and because ß-secretase is a site-specific protease that cleaves exactly between specific amino-acids of Aß, the last enzyme inhibition would decrease production of all forms of Aß peptides, including the pathogenic Aß42. Thus, this enzyme is recognized as a prime drug target for therapeutic inhibition of Aß production in AD [5, 27, 30].

While a lot of efforts have been concentrated on deregulated proteins and protein-encoding genes with an important role on Alzheimer's disease, more recently significant attention has been focused on miRNAs (Figure 4) [31].

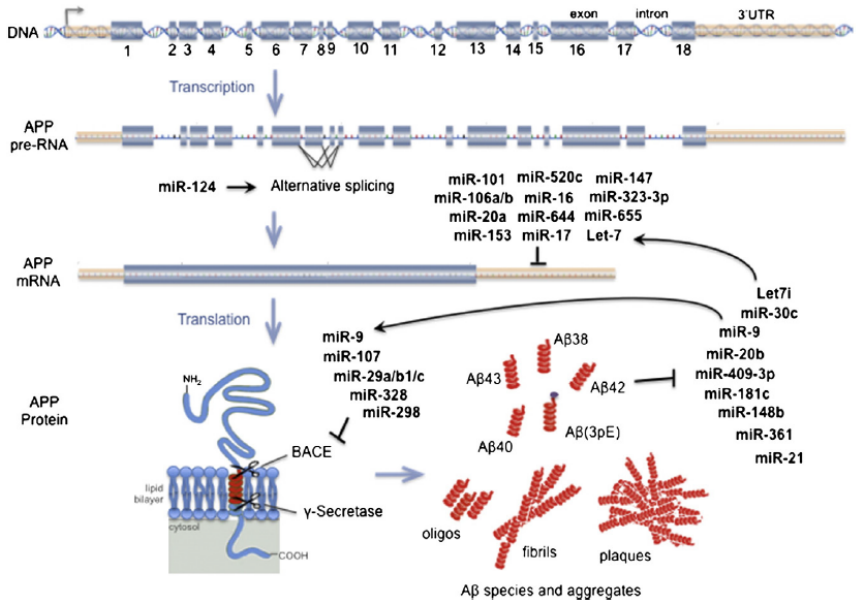


Figure 4: MiRNA network surrounding APP, that affect direct or indirectly the APP processing and Aß metabolism. Different types of Aß species are produced, which differ in their degrees of toxicity. Aß40 is the major species, while Aß42 is the major pathologic species (adapted from [5]).

Several papers mention that the deregulation of miRNA expression plays a key role in AD pathogenesis [32]. MiRNAs stand out as a novel therapeutic, as in the case of APP, they regulate APP mRNA levels and have a role in alternative splicing. They also indirectly regulate APP processing via the ß-secretase BACE1 that has been identified as a miRNA target, with the miRNA cluster containing miR-29a,-29b1,and-9 playing a crucial role [31].

In Table 2 are described some of the de-regulated miRNAs founded to have a role in the Alzheimer's disease.

Table 2: De-regulated miRNA related to Alzheimer's disease (adapted from [2, 25]).

miRNA	Upregulation (Up) or downregulation (Down)	Target	Pathological implications	Year	[Ref.]
miR-9, miR-128	Up	ND	General neuropathology of AD.	2007	[33]
miR-146a	Up	Complement factor H (CFH) and IL-1 receptor-associated kinase-1 (IRAK1).	Sustained inflammatory response.	2008	[34]
miR-107, miR-29a/29b-1, miR-9, miR-15a, miR-19b	Down	BACE1 APP (only miR-15a)	Increased production of Aß.	2008	[11, 35]
miR-106b	Down	APP	Increased production of Aß.	2009	[36]
miR-29a	Down	Neuron navigator 3 (NAV3).	A putative compensatory mechanism against neurodegenerative events.	2010	[37]
miR-15a	Down	Extracellular signal regulated kinase 1 (ERK1).	Hyperphosphorylation of tau.	2010	[38]
miR-485-5p	Down	BACE1	Increased production of Aß.	2011	[39]
miR-124	Down	PTBP1	Aberrant APP mRNA alternative splicing.	2011	[40]

Subsection 2.1 – Alzheimer's disease and miR-29

As described on Table 2, Hebert and co-workers (2008) showed that specific loss of miRNA cluster miR-29a/b-1 was correlated with increased BACE1 expression in sporadic AD (Figure 5) and that the introduction of pre-miR-29 reduces secretion of Aß peptides [5, 11]. These observations raised the opportunity to use pre-miR29 as a therapeutic product for AD, based on the hypothesis that decreasing the specific protein level itself is a protective therapeutic strategy [19].

The use of pre-miR instead of using the mature miRNA represents an advantage to the therapeutic process/purpose. Pre-miR molecule is longer than the mature form facilitating the purification step. Furthermore, in cellular environment the pre-miRNA is more easily recognized by cells, which will help the entrance and activity [41, 42].

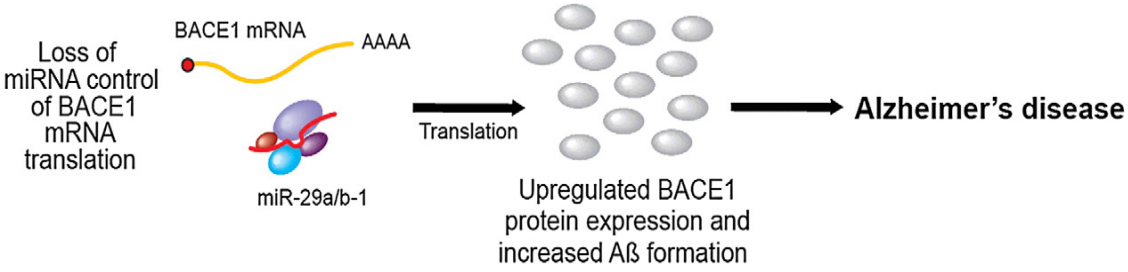


Figure 5: Summary of miR-29a/b-1 role on Alzheimer's disease (Adapted from [25]).

Recent studies have provided significant insights into the biology of the miR-29 family of miRNAs and now it is known that comprises four closely related members: miR-29a, miR-29b-1, miR-29b-2 and miR-29c [43]. All sharing identical sequences at nucleotide positions 2-8 of 5' end and the same seed region, they highly overlap in predicted mRNA targeting as previously described [43, 44]. Some characteristics of miR-29 family members are described on the Table 3.

Table 3: Characteristics of miR-29 family members (adapted from [44]).

Characteristics	
Mature sequences	
miR-29b-1 and miR-29b-2	Identical
miR-29a and miR-29c	More distantly removed from miR-29b, distinguished by only a single different nucleotide outside the seed sequence.
Bi-cistronic clusters	
miR-29a/b-1	Located on the antisense strand of chromosome 7 of the human genome.
miR-29b-2/c	Located on the antisense strand of chromosome 1 of the human genome.
Compartmentalization	
miR-29b and miR-29c	Concentrated in nucleus.
miR-29a	Mainly cytoplasmatic.

Section 3 - Purification of pre-miR29

In order to use miR-29 as a therapeutic product it is important to ensure its purity, integrity and stability. Thus, the development of an efficient biotechnological process, namely the optimization of the extraction and purification methods is crucial to guarantee the miR-29 quality.

Although RNA is a sensitive molecule and has a very short half-life, once extracted from cells or tissues, is especially unstable due to the ubiquitous presence of RNases (RNA-degrading enzymes), which are present in aqueous buffers, labware and associated with human handling. These RNAses are difficult to inactivate as they do not require cofactors and because they are heat-stable and refold following heat denaturation. Therefore, RNA extraction relies on good laboratory and RNase-free techniques [12].

There are two commonly used methods to isolate RNA from natural sources (e.g. tissue samples, whole organisms, cell cultures, bodily fluids), namely a phenol-based extraction method and silica matrix or glass fiber filter-based binding method, often referred to as solid-phase extraction (SPE) [45]. Using phenol-based isolation procedures is possible to recover

RNA species with 10 to 200 nucleotides, and it is the recommended procedure for isolating total RNA from biological samples. This method uses a phenol-based reagent, such as TRIzol®/TRI Reagent®, to obtain a good quality total RNA (pure and intact) from a small amount of tissue [45]. The second method, solid-phase extraction, relies on using high salt or salt and alcohol to decrease the affinity of RNA for water and increase its affinity for the solid support used. This method is based on the formation of complexes between nucleic acids and the silica gel matrix, in the presence of chaotropic salts [45, 46]. The silica-based methods are less laborious and time-consuming (~30 minutes), need less manipulation, involve no pellet handling and therefore, result to a reproducible high RNA yield [45, 46].

Nowadays, there is a variety of commercially available methods for the isolation of samples enriched in small RNAs (supplied by Ambion, Stratagene, Roche, Sigma–Aldrich, Invitrogen; etc. (Table 4)) from which mirVana miRNA Isolation kit (Ambion) is the most used [47].

Table 4: Examples of commercially available kits to isolate microRNA.

Company	Product	Smal RNAs	Phenol	Solid-phase extraction	[Ref.]
Ambion	**mirVana miRNA Isolation kit**	< 200 nt	Yes. Uses organic extraction (Acid-Phenol:Chloroform extraction).	Yes. Uses two sequential filtrations with a glass-fiber filter.	[48]
Qiagen	**miRNeasy Micro Kit**	< 200 nt	Yes. Phenol/guanidine-based lysis.	Yes. Silica membrane–based purification.	[49]
Sigma-Aldrich	**mirPremier microRNA Isolation Kit**	Small RNAs, not specified nt	No. Employs a novel purification chemistry, without using hazardous organic extractions.	Yes.	[50]
Stratagene	**Absolutely RNA miRNA Kit**	Small RNAs, not specified nt.	Yes. Phenol/ guanidine-based lysis for cells, although for lysates prepared from cultured cells, phenol extraction is not required.	Yes. Employs silica-based fiber matrix.	[51]
Invitrogen	**Ambion® PureLink® miRNA Isolation Kit**	< 200 nt	No. No need for hazardous reagents such as phenol.	Yes. Silica-based two-column system.	[52]
Roche	**High Pure miRNA Isolation Kit**	<100 nt	No. No need for organic solvent extraction.	Yes.	[53]

In addition to these commercial kits, further research has been performed in order to purify small RNAs. Many other methods are referred on literature, such as crush and soak method, electroelution using a dialysis bag, electroelution onto cellulose membrane, electrophoresis of RNA molecules onto diethylaminoethyl paper, centrifugal filtration through filter membranes or siliconized sterile glass wool and elution of RNAs from denaturing polyacrylamide gel electrophoresis. These are mostly time consuming, expensive and tedious methodologies [46]. To overcome these disadvantages, new techniques were explored, like the application of magnetic oligo(dT) beads for the purification of poly(A)$^+$ RNA from total RNA sample, anion exchange HPLC to purify chemically synthetized RNAs oligonucleotides or size exclusion chromatography [46, 54]. Still, these methods are laborious and the use of synthetic RNAs represents a limitation [54].

In addition, several affinity chromatographic strategies have been developed in order to purify miRNAs (Table 5). Among which, it can be mentioned the association of biotinylated RNAs with streptavidin matrices, hybridization of poly-A sequences inside the target RNA to anchored oligo-U or oligo-dT, the immobilization of RNA aptamers with affinity for streptavidin or streptomycin, junction of RNAs with cyanogen bromide activated sepharose or the coupling of periodate-oxidized RNA with adipic acid dihydrazide-agarose beads [55]. Table 5 resumes some of the latest studies that have been done in order to separate RNAs using affinity chromatography as main method.

Table 5: Affinity chromatographic methods to achieve RNA purification.

Method	Matrix	RNA	Advantages	Disadvantages	[Ref.]
Association of biotinylated RNAs with streptavidin-bearing matrices	Streptavidin-agarose matrix	Biotinylated RNAs	High specificity; Streptavidin-agarose beads are commercially available and the elution with biotin is done under mild conditions.	Need RNA aptamers production; Use synthetic RNAs; Aptamer tag insertion into an accessible region of the RNA can fail due to misfolding.	[55-57]
Hybridization of poly-A sequences within the target RNA to anchored oligo-dT	Oligo-dT cellulose	poly(A) tail at the 3'end of pri-miRNA	Available commercially. Specific method.	Considerable amount of a given pri-miRNA does not contain 5' cap or poly(A) tail; miRNA samples need to be poly adenhylated.	[58]
High capacity affinity column with covalently linking an amino-modified RNA to a Sepharose support.	Hi-Trap-NHS Sepharose converted on a alkyl-thiol column	Amine-modified RNA	Ease of performing the cystamine conversion of HiTrap columns; Reusable column.	Low efficiency of direct coupling of RNA to matrix, hence, inefficient for large scale; Need to modify the initial matrix for larger efficiency; Synthetic RNA.	[55]

Even with all the progresses accomplished, there still is a long way to achieve a reliable, fast and robust method to get a good and substantial separation of a specific RNA. For example, the use of biotinylated synthetic miRNAs pulled-down on streptavidin beads presents high levels of background, because RNAs biotinylation errors can occur, also the biotin molecule stay permanently liked to RNAs and it always has impurities resulting from reaction [57]; some tagged RNAs do not cleave even after prolonged incubation which can significantly degrade the molecule and oligomeric RNA species formed during transcription can not be separated from the monomeric species [59, 60].

In order to overcome some existing drawbacks, over the last years, amino acid-based affinity chromatography has been exploited as a resource to achieve the isolation and purification of different nucleic acid, based on the natural occurrence of many different interactions between proteins and nucleic acids in biological organisms [61]. This selectivity can be explained by the recognition of the individual structure involved on amino acid/nucleic acid specific interaction [62].

First, amino acid affinity chromatography appears applied to the purification of plasmid DNA (pDNA) where histidine, arginine and lysine succeeded to isolate supercoiled pDNA isoform (sc pDNA). These studies proved that some specific interactions exist between those amino acid matrices and DNA [61]. With histidine affinity chromatography and using high salt concentration, the sc pDNA was efficiently isolated from other isoforms and host impurities, since this molecule stay strongly retained in column [12]. Although this matrix showed a good result, arginine and lysine affinity-chromatography represented a good alternative because the total retention is established with low salt concentration and species are eluted with a slight increase of ionic strength. Other study with this support allowed the 6S RNA purification with high selectivity, denoting that RNA can also be purified with amino acid-based chromatography, as well [61].

In the present work, the isolation of the miR-29 on an O-phospho-L-tyrosine matrix is proposed. This matrix was chosen given that inside the cell the miRNA binds to the complex RISC through a conserved tyrosine residue [63-65], essential to the right cleavage of mRNA.

As described on literature, the phosphorylated 5′-terminal nucleotide of the guide RNA is unpaired and stacks over a highly conserved basic pocket, anchoring to the aromatic ring of an invariant tyrosine (Tyr). This conserved Tyr has an essential role, as its absence attenuate mRNA cleavage activity [63, 64]. These studies raised the chance to explore this natural interaction, which is expected to occur as equal between the miRNA and the chromatographic matrix.

Subsection 3.1 - One at a time method and Design of experiments

The optimization of chromatographic methods for nucleic acid purification is often intricate and can be time consuming. Usually, the optimization is achieved by changing parameters one by one which is commonly called "one variable at a time" (OVAT) methodology or simply "one at a time" method [66]. However, purification step is not only defined by high yields and product purity. A separation step has also to take into account additional demands such process robustness, financial and ecological constraints. The need to maximize the efficiency of scientific experimentations, in order to minimize waste and cost, has caused researchers to do a set of experiments that give the most information possible with the fewest assays performed [67]. Thus, the experimental design appears as an important tool to help scientists to find the best purification method.

Although the classic theories of experimental design appear around the middle of twentieth century, it was on the past decade that they had an exponential growth as the number of papers increased [67]. Hibbert (2008) shows how this type of method has grew in the last few years and it is possible to see that "Design of Experiments" (DoE) is mostly applied on high-performance liquid chromatography (HPLC) and gas chromatography (GC) systems [67]. The DoE methodology is a multivariate statistical analysis that allows evaluating multiple variables which affect a particular process, instead one at a time [68]. With it, the multivariate data can be fitted to an empirical function, usually linear or quadratic, which can provide information about the system [67].

Most chromatographic studies performed with DoE can be classified as either method validation robustness studies or optimizing method conditions [67]. Typically, the user has to define the factors of study (normally from two to four factors) like mobile phase composition, gradient parameters, pH, temperature, injection volume or flow rate. Then, the minimum and maximum values for each factor in the experimental domain should be set. These values represent the design levels needed to establish the group of experiments to be done when the multivariate technique is employed [67, 69]. The majority of the designs used to determine response surfaces are the full and fractional factorial designs, the central composite design, Box–Behnken, Doehlert and mixture designs. Factorial designs acquire more importance when it is necessary to investigate the most important factors or which do not significantly affect the experimental responses [69]. However, central composite designs are commonly used also for optimization, even when Box–Behnken or Doehlert designs, with greater efficiency, might be better [67].

The Box-Behnken design is a three-level incomplete factorial design, which means that has three levels or more and can be used on problems with three or more factors (Figure 6). This design consists in three parts of four runs. Within each part, two factors are arranged in a full

two-level design, while the level of the third factor is set at zero. The points lie on the surface of a sphere centered at the origin of the coordinate system and tangential to the midpoint of each edge of the cube [67, 69].

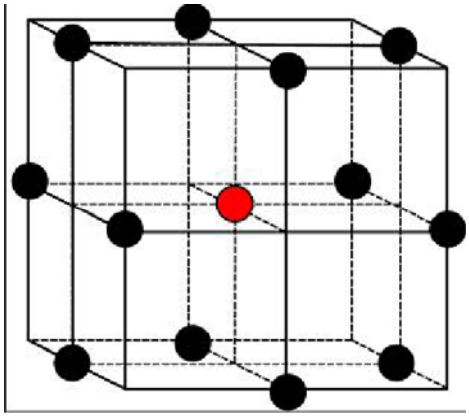


Figure 6: The Box-Behken design in three factors. In red, within of the design, are the optional numbers of center-points. This design is useful when experiments at the extreme levels are undesirable or impossible (adapted from [70]).

As this design does not contain combinations where all the factors are at their higher or lower levels, it may be advantageous for assays under extreme conditions, for which are predicted unsatisfactory results. Thus, this design is not indicated for experiments that need to evaluate the responses at the vertices of the cube, that is, at the extreme levels [71].

Ferreira and co-workers (2007) promoted the use of Box–Behnken designs over central composite design and mostly over the three-full factorial designs where the efficiency is compromised when the factor number is higher than 2 [71]. The efficiency of a model can be obtained by the number of coefficients in the estimated design, divided by the number of experiments. Table 6 shows the different efficiencies for Box-Behken designs, central composite design and Doehlert design [71].

Table 6: Comparison of efficiency of Box-Behken designs (BBD), central composite design (CCD) and Doehlert design (DM) (adapted from [71]).

Factors	Number of coefficients	Number of experiments			Efficiency		
		CCD	DM	BBD	CCD	DM	BBD
2	6	9	7	--	0.67	0.86	--
3	10	15	13	13	0.67	0.77	0.77
4	15	25	21	25	0.60	0.71	0.61
5	21	43	31	41	0.49	0.68	0.46

So far, Box–Behnken design has been employed for the determination of critical conditions in extraction steps, derivatization reactions and also separation steps in chromatographic methods like paper chromatography (PC), gas chromatography with flame ionization detector (GC–FID), gas chromatography–mass spectrometry (GC–MS), liquid chromatography – mass spectrometry (LC-MS) and HPLC. The Table 7 summarizes these applications of this design in optimization of chromatographic methods [69].

Table 7: Box-Behnken design applied to chromatographic studies (adapted from [67, 71]).

Chromatographic technique	Sample	Analyte	Optimized process
GC-ECD	Wine	2-4-6 TCA, 2-4-6 TBA	Extraction step
HPLC	--	Aliphatic aldehydes	Derivatization step
GC-MS	Sediments	Organochlorine pesticides	Extraction step
Hydrophobic interaction C-DAD	--	Organic acids	--
GC-FID	Castor oil	Fatty acid composition	Extration step
HPLC	Pharmaceutical tablets	Captopril	Separation step
GC-MS	--	Aminoglycoside antibiotics	Derivatization reaction
PC	--	Aminoacids hydroxamates	Separation step

In this present work, the Box-Behnken design was used to achieve the optimization of the purification method. A method to purify pre-miR29b-1 was also developed, using a new matrix – O-phospho-L-tyrosine – exploiting the natural interactions that occur on a natural environment. This brings the advantage of purify the pre-miR29b-1 from other small RNAs present on the sample from a previous extraction. As the Box Behnken design represents a useful technique to optimize a purification step, here it will be discussed the benefits compared to the "one at a time" methodology.

Chapter 2

Materials and Methods

Section 2.1 - Materials

For *Rhodovulum sulfidophilum* DSM 1374 culture, it was used tryptone and yeast extract from Bioakar (Beauvais, France), $CaCl_2.2H_2O$, NaCl and K_2HPO_4 from Panreac (Barcelona, Spain), $MgSO_4$, glucose, KH_2PO_4 and $FeSO_4.7H_2O$ from Sigma-Aldrich (St Louis, MO, USA), kanamycin and H_3BO_3 obtained from Thermo Fisher Scientific Inc. (Waltham,USA) and agar was from Pronalab (Mérida, Yucatán). The reagents used to cell lysis were ß-Mercaptoethanol, sodium acetate and chloroform acquired from Merck (Whitehouse Station, USA), isopropanol from Thermo Fisher Scientific Inc. (Waltham,USA) and guanidinium thiocyanate salt and sodium citrate from Sigma-Aldrich (St Louis, MO, USA). All solutions were freshly prepared using 0.05 % Diethyl pyrocarbonate (DEPC) treated water. DEPC was purchased in Fluka (Sigma-Aldrich - St Louis, MO, USA). The matrix O-phospho-L-tyrosine agarose was purchased from Sigma-Aldrich (St Louis, MO, USA) and the salts used on buffers, Tris(hydroximethyl)methylamine (Tris) and ammonium sulfate, were bought from Nzytech (Lisboa, Portugal) and VWR (Radnor, Pennsylvania, USA) respectively. All buffers where filtered through a 0.20 µm pore size membrane (Schleicher Schuell, Dassel, Germany). All the materials used in the experiments were RNase-free.

Section 2.2 – Methods

Subsection 2.2.1 - Bacterial *R. sulfidophilum* DSM 1374 growth conditions

The RNA used in this study was obtained from the marine photosynthetic bacterium *R. sulfidophilum* DSM 1374 strain, it being acquired a high production of RNAs enriched in small RNAs and particularly pre-miR29b-1. This bacterium was previously transformed by heat-shock with pBHSR1-RM vector (genetically modified with pre-miR29b-1 sequence). The host selection was made considering that it does not produce RNases and that it selectively secretes the several nucleic acids species [72]. For pre-cultivation, a stride from a selective plate (5 g/L yeast extract, 10 g/L glucose, 20 g/L NaCl, 4.1 g/L $MgCl_2$, 10 g/L polypeptone and 15 g/L agar supplemented with 1 mg/L $ZnSO_4.H_2O$, 10 mg/L $MnO_4S.4H_2O$, 10 mg/L $FeSO_4{\cdot}7H_2O$ and 30 µg/mL kanamycin) was inoculated into a 250 mL shake flask at 30 ºC and 250 rpm with 50 mL culture medium (10 g/L tryptone, 500 mg/L yeast extract, 1 g/L KH_2PO_4, 4 g/L K_2HPO_4, 30 g/L NaCl, 10 g/L glucose supplemented with 1 g/L $(NH_4)_2SO_4$, 0.2 g/L $MgSO_4$, 0.05 g/L $CaCl_2.2H_2O$ and 30 µg/mL kanamycin). All liquid medium were also supplemented with a salt solution (5.56 g/L $FeSO_4.7H_2O$, 3.96 g/L $Cl_2CuH_4O_2$, 5.62 g/L $CoSO_4.7H_2O$, 0.34 g/L $CuCl_2.2H_2O$, 0.58 g/L $ZnSO_4.7H_2O$, 0.6 g/L H_3BO_3, 0.04 g/L $NiCl_2.6H_2O$ and 0.06 g/L $Na_2MoO_4.2H_2O$ dissolved on 0.5 N HCl), for 100 mL of medium was added 100 µL of salt solution. Cell growth was suspended in the exponential phase, ~ 2.6 OD_{600} (optical density of the culture medium at a wavelength of 600 nm). All fermentations where started with the appropriate amount of pre-culture, calculated to have an initial OD_{600} of approximately 0.2 in 100 mL of culture medium (present on a 500 mL shake flasks). After 72 h, all cells were harvested, recovered by centrifugation and stored at -20 ºC.

Subsection 2.2.2 - Lysis and small RNA isolation

Small RNAs samples were obtained by using the acid guanidinium thiocyanate–phenol–chloroform method based on what Chomczynski and Sacchi (2006) described [73]. The bacterial pellets (from 100 mL of culture medium) were resuspended by vortexing in 20 mL of 0.8 % NaCl solution and centrifuged at 6000 G during 10 min at 4 ºC. The resulting pellets were resuspended by successive pipeting using 5 mL of Solution D (4 M guanidiniumthiocyanate, 25 mM sodium citrate, pH 4.0, 0.5% (w/m) N-laurosylsarcosine (sarcosyl) and 0.1 M ß-mercaptoethanol) to achieve cell lysis. After incubating on ice for 10 min, 10 mL of water-saturated phenol and 1 mL of 2 M sodium acetate (pH 4.0) are gently added and mixed to the previous mixture. This step allows the precipitation of cellular debris, genomic DNA and proteins. The small RNAs were then isolated by adding 2 mL of chloroform/isoamyl alcohol (49:1) followed by a vigorously mixing (until two immiscible

phases were obtained), 15 min on ice and a 20 min centrifugation at 10 000 G and 4 ºC. The aqueous phase was recovered and concentrated by the addition of 5 mL of isopropanol (-20 ºC), followed by a centrifugation step (10 000 G for 20 min at 4 ºC). The obtained pellet, which contains sRNA, was resuspended in 1.5 mL of Solution D by vortexing and concentrated again by adding 1.5 mL of isopropanol (-20 ºC). After the centrifugation of 10 min at 10 000 G and 4 ºC, was added 2.5 mL of 75% ethanol to the resulting pellet to wash all the sRNAs present and were incubated at room temperature for 10 min, followed by a 5 min centrifugation at 10 000 G (4 ºC). Pellets were air-dried and then dissolved in 1 mL of 0.05 % DEPC-treated water. Finally, the RNA concentration and purity was determined using a Nanodrop spectrophotometer and realizing an agarose gel electrophoresis.

Subsection 2.2.3- Affinity chromatography

Based on the natural interactions that occur between tyrosine and siRNas on a natural environment, this type of chromatography was performed with the samples obtained after small RNAs extraction from *Rhodovulum sulfidophilum* DSM 1374 cells. Affinity chromatography was performed in an AKTA Avant system with UNICORN™ 6.1 software (GE Healthcare, Sweden). A 20 mL column was packed with the commercial O-phospho-L-tyrosine agarose gel. This support is characterized as a cross-linked 4% beaded agarose matrix with an 1-atom spacer and an extent of labelling between 5 and 15 µmol/mL. The column was differently equilibrated, from 2 to 1.2 M $(NH_4)_2SO_4$ in 10 mM Tris–HCl (pH 7.5) buffer, depending on the experiments performed. It was used a flow rate of 1 mL/min and the sample was injected into the column using a 200 µL loop at the same flow rate and buffer concentration of the equilibrium step. All the chromatographic process was monitored at 260 nm. After elution of unbound species with the initial equilibrium buffer, the ionic strength of the buffer was decreased stepwise in a range of 1.4 M to 0.2 M of $(NH_4)_2SO_4$ in 10 mM Tris–HCl (pH 7.0) buffer, according to the experiments performed at the time. Finally, the column was washed, removing tightly bound RNA species by changing to simple 10 mM Tris–HCl (pH 7.0) buffer. All peaks were pooled according to the obtained chromatograms. The pooled fractions were then concentrated and desalted, using RNA precipitation after a PD-10 desalting column (GE Healthcare) or only using the Vivaspin concentrators (Vivascience). The PD-10 column isolates the salt molecules from the miRNA sample based on different sizes, using the gel filtration chromatography principle. The sRNA molecules elute first, because they are selectively excluded from entering the porous gel phase, being collected for further analysis. On the other hand, salt molecules will penetrate the pore volume and therefore they elute after the large molecules. At the end, all samples were reserved (at -20 ºC) for further analysis.

Subsection 2.2.4 - Agarose gel electrophoresis

To estimate the purity and stability of RNAs acquired from extraction an agarose gel electrophoresis can be made. This technique is performed on a 15-cm-long 1 % agarose gels (Grisp, Oporto, Portugal) stained with green safe (0.3 µL/mL) and visualized under UV light in a Vilber Lourmat system (ILC Lda, Lisbon, Portugal). Electrophoresis is carried out in Tris-acetic acid (TAE) buffer (40 mM Tris base, 20 mM acetic acid and 1 mM Ethylenediaminetetraacetic acid (EDTA), pH 8.0) in DEPC-treated water and run at 110 V and 400 mA for 30 min.

Subsection 2.2.5- RNA precipitation

In order to recover and further analysis RNA from different peak fractions, the samples were desalted and concentrated. To achieve the concentration of the samples, the nucleic acids were precipitated (after desalting with a PD-10 column). The RNAs fraction (300 µL) was precipitated by adding 60 µL of 1 mg/mL of glycerol and 1500 µL of ice-cold ethanol absolute. After an overnight incubation at - 80 ºC, the pellets were recovered by centrifugation at 16 000 G, for 15 min at 4 ºC. The obtained pellets were air-dried for 10 min at room temperature and their solubilization was performed in 40 µL of 0.05 % DEPC-treated water. As the work evolved, this step was substituted by the utilization of concentrators.

Subsection 2.2.6- cDNA synthesis

cDNA was synthesized using the "RevertAid™ First Strand cDNA Synthesis Kit" (Fermentas, , Thermo Fisher Scientific Inc.), according to manufacturer's instructions. A total of 0.5 µg of RNA of samples collected after the chromatographic purification process with tyrosine–agarose column was used to initiate cDNA synthesis, so that it was possible to perform PCR amplification of samples. To 0.5 µg of RNA were added 20 pmol of gene specific primers and the mixture was incubated at 65 ºC for 5 min on thermocycler. After cooling the samples on ice, vortex, and cooling again, the following components for the respective order were added: 4 µL 5X Reaction Buffer, 1 µL Ribolock™ Rnase Inhibitor (20 U/µL), 2 µL of 10 mM dNTP Mix and 1 µL Revert Aid™ M-MuLV Reverse Transcriptase (200 U/µL). The mixture was processed on a two steps program on a thermocylcler. The program consists of 1 hour at 42 ºC followed of 5 minutes at 70 ºC. At the end, cDNA is obtained and ready to be used on a PCR amplification.

Subsection 2.2.7- Polymerase chain reaction

To know if pre-miRNA eluted on a specific peak, the fractions collected were transformed on cDNA and then amplified by a polymerase chain reaction (PCR). First, a mix is prepared with 0.125 U of Supreme DNA polymerase (NZYtech, Lisboa, Portugal), 50 mM of magnesium chloride (NZYtech, Lisboa, Portugal) and 150 nM of each primer. Specific primers (Forward – 5'- GGA AGC TGG TTT CAT ATG GTG -3' and Reverse – 5'- CCC CCA AGA ACA CTG ATT TC -3') for cDNA whose design was achieved on RNA database were used to amplify a fragment of 145 bp. The polymerase chain reaction was already optimized for the primers involved and the program followed was: 5 minutes at 95 ºC for desnaturation of cDNA, 35 cycles with three different steps – 95 ºC for 30 seconds, 64 ºC for 30 seconds and 72 ºC for 15 seconds- for cDNA amplification with 64 ºC as annealing temperature and 5 minutes at 72 ºC for final elongation. To confirm the presence and purity of amplicons, PCR products were analyzed by 1% agarose gel electrophoresis.

Subsection 2.2.8- Polyacrylamide electrophoresis

The integrity and identification of small RNAs present on the collected fractions was checked by denaturing urea polyacrylamide gel electrophoresis. All samples were denatured with a 97.5% formamide, 0.3% bromofenol, 10 mM EDTA at pH 7.5 solution and the denatured conditions were kept in the gel due to the presence of 8 M urea. Twenty microliters from the concentrated peaks were then resolved into a 10% polyacrylamide TBE-urea gel that was carried out at 120 V for 120 min with Tris-Borate-EDTA (TBE) buffer (0.84 M Tris base, 0.89M boric acid and 0.01 M EDTA, pH 8.3). After the run, all gels were stained with ethidium bromide (0.5 mg/mL) and the sRNAs in the gel were visualized using a Vilber Lourmat system (ILC Lda). To run the electrophoresis, an Amersham Biosciences system (GE Healthcare, Sweden) with a Bio-Rad font were used.

Chapter 3

Results and Discussion

Alzheimer's disease represents, at this time and in a near future, a burden to all societies. In health, economically and in quality of life this disease affects many people, being necessary to find a useful therapeutic strategy to overcome all these problems. Recent studies have demonstrated that miRNAs are fundamental on the development of this disease. The expression patterns are modified on AD brains, and some are directly or indirectly related with the hallmarks symptoms [74]. MiRNA therapeutics are evolving, leading to miRNA regulation through different strategies like miRNA mimics molecules [75].

In order to use miRNA-based therapies it is required a pure, stable and complete molecule. So, several methodologies have been tested to achieve its purification, namely anion exchange chromatography and many types of affinity chromatography, using a diversity of matrices, mostly involving target RNA modification.

In this work an O-phospho-L-tyrosine was used to achieve the pre-miR29b-1 purification. Exploiting the natural and specific interactions between miRNA and the matrix brings several advantages like: it is not necessary to modify the recombinantly produced pre-miR in order to bind it to the matrix; it is not used a synthetic pre-miR, because this method allows the purification after its extraction from cells; and it is adaptable to an industrial process. To reach this objective the optimization of the chromatographic method is crucial. Through a large number of experiments, using the method "one at a time", some conditions were established. Furthermore, the Box Behnken design approach was evaluated towards the conventional methodology, as a useful laboratory technique to achieve faster pre-miR purification. The design of experiments approach brings the advantage of testing several conditions that interfere with the purification process, accessing their influence on purification performance, and optimizing it through a mathematical model.

Section 3.1 - Bacterial *R. sulfidophilum* DSM 1374 lysis

The *R. sulphidophilum* bacterium has particular features that make it a promising host for RNA production. As this microorganism secrets RNAs (predominantly sRNAs) with no RNases production at the same time, it has been studied to produce RNAs of defined sequence [72]. Suzuki and coworkers have already demonstrated an extracellular RNA aptamer production [76, 77]. Although, the bacterium differently secrets nucleic acids, the quantity that is produced is not totally clear. In addition, it is not known what adverse conditions these molecules face outside the cell membrane, even if this host does not secrete RNases. Thus, this work has focused on small RNAs produced intracellularly, as there are in a natural environment, reducing the risk of degradation or destabilization of pre-miR molecule.

On chromatographic processes it is important to characterize the initial sample. The small RNAs samples were recovered from *R. sulphidophilum* using the acid guanidinium thiocyanate–phenol–chloroform method. After the extraction, the characterization of the RNAs samples, regarding the integrity and types of RNAs present in the sample, was achieved by doing an agarose gel electrophoresis.

All the samples prepared were enriched on small RNAs of different types, as it is visible in Figure 7. Agarose gel electrophoresis shows the intense band corresponding to sRNAs, whilst the polyacrylamide gel presents numerous bands corresponding to the different sRNAs species. After confirming the extraction was successful achieved, which means that the obtained sample were enriched on small RNAs with no DNA present, these samples were used on the chromatographic process to purify the pre-miR.

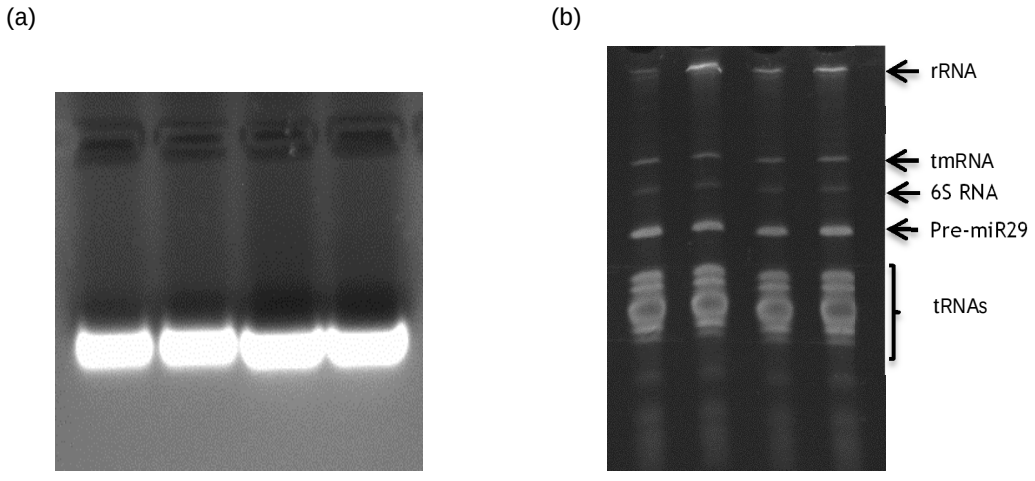


Figure 7: Characterization of four different sRNAs samples extracted form *R. sulphidophilum*: a) Agarose gel electrophoresis analysis, b) Polyacrylamide gel electrophoresis.

Section 3.2 – Pre-miR29b-1 purification

Subsection 3.2.1 - Main interactions occurring between RNA and O-phospho-L-tyrosine agarose matrix

Amino acid-affinity chromatography uses biological-based selectivity, which occurs between amino acid ligands and nucleic acids to isolate some particular biomolecules. This type of selectivity is explained by the biological recognition or by an individual chemical structure, which favors the interaction [78]. The use of an O-phospho-L-tyrosine (P-Tyr) brings the chance to explore the natural interaction that occurs between complex RISC, particularly with a conserved tyrosine residue inside it, and miRNA molecules [63, 64]. The interactions between the ligand and the target molecule can be mostly based on electrostatic or hydrophobic interactions, van der Waals forces and/or hydrogen bonding. To promote the biomolecule elution, the interaction can be reversed by using a competitive ligand, or by changing the pH, ionic strength or polarity of the mobile phase [79].

Initially, it was important to determine the main interactions occurring between RNAs and the matrix. For that, 200 µL of sample where injected onto the column and different binding and elution conditions were tested. Therefore, electrostatic interactions were explored by using a stepwise gradient from Buffer A – 10 mM Tris-HCl at pH 7.5 to Buffer B – 1 M NaCl (in 10 mM Tris-HCl at pH 7.5). The chromatographic profile obtained under these conditions is presented in Figure 8.

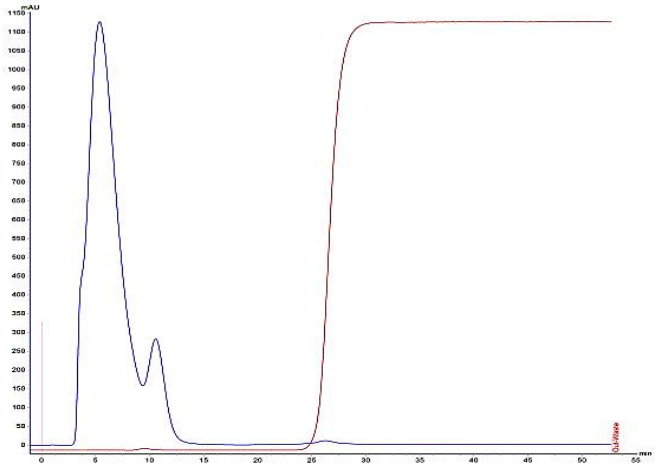


Figure 8: Chromatographic profile of RNAs using a stepwise gradient of 10 mM Tris-HCl at pH 7.5 to 1 M NaCl in 10 mM Tris-HCl at pH 7.5.

After sample injection onto the column, most RNAs immediately eluted in the flowthrough with Buffer A, allowing to conclude that there was no retention of RNAs and it was not possible to achieve selectivity for the target miRNA under ionic conditions.

Considering the previous result, it was evaluated the binding behavior of RNA to the matrix when hydrophobic interactions were predominantly favored. In this case, it was used as Buffer A a 2 M $(NH_4)_2SO_4$ (in 10 mM Tris-HCl, pH 7.5) solution and as Buffer B a 10 mM Tris-HCl at pH 7.5 solution. The experiment was performed using the Buffer A to establish the equilibrium condition, and after RNA sample injection the majority of RNA species were retained, only eluting from the column when the ionic strength of the buffer decreased (Figure 9).

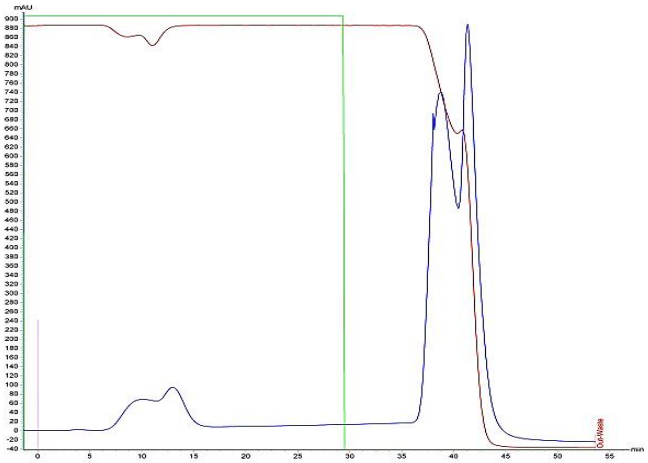
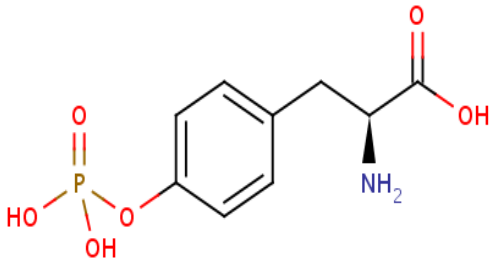


Figure 9: Chromatographic profile of RNAs using a stepwise gradient of 2 M $(NH_4)_2SO_4$ in Tris-HCl, pH 7.5 to 10 mM Tris-HCl, pH 7.5.

The chromatographic profile obtained under these conditions shows the presence of some interactions occurring between the ligand and the RNA sample. Furthermore, it is visible a different retention pattern, evident from the non resolved peaks eluting with Tris-HCl buffer. This result indicated that not all biomolecules bound with the same strength to the matrix, allowing to explore this difference to purify pre-miR.

The interactions established between RNA and the amino acid ligand can be explained by the presence of hydrophobic groups on P-Tyr matrix, as the aromatic group as well as the phosphate group (Figure 10) [62]. Using a high ammonium sulfate concentration it was possible to favor the hydrophobic interactions with RNA molecules. The salt can capture the water surrounding the biomolecule, exposing their hydrophobic groups, favoring the interaction with the ligand.

(a) (b)



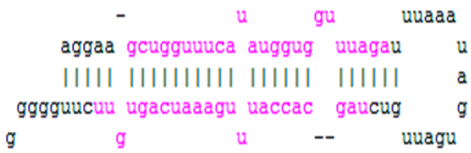
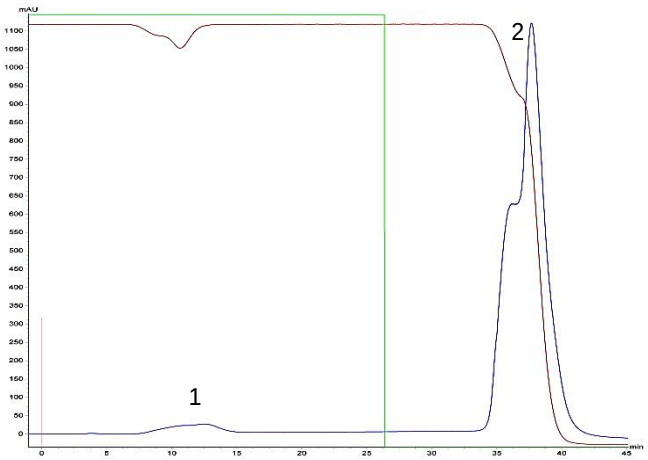


Figure 10: P-Tyr ligand (a) immobilized on the agarose matrix and pre-miR29b-1(b).

In addition to the hydrophobic interactions previously mentioned, the tyrosine ligand can also induce other elementary interactions with RNA. Because of the existence of an amine group and a carboxylic group, this ligand presents polar interactions that can also interfere on the binding process. Thus, it is supposed that the conjugation of all these interactions can induce some affinity and selectivity towards the target RNA.

In order to characterize the RNAs eluting in each peak, the chromatographic experiment was used to recover the retained and not retained biomolecules. Thus, both peaks were collected, desalted and concentrated to produce cDNA and perform a PCR (Figure 11). In this case, the desalting process was achieved by using a PD-10 desalting column (GE Healthcare) and the concentration was accomplished by precipitation with ice-cold absolute ethanol. After the preparation of the RNA samples, the PCR technique was used to verify if the target pre-miR29 was recovered and in which peak it was eluted.

(a) (b)



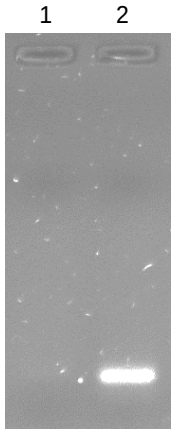


Figure 11: (a) Chromatographic profile of RNAs elution with a stepwise gradient of 2 M $(NH_4)_2SO_4$ in Tris-HCl, pH 7.5 to 10 mM Tris-HCl, pH 7.5. (b) Agarose gel electrophoresis analysis of cDNA synthesized using each chromatographic fraction (1 and 2) as template.

It was verified that the pre-miR bound to the column and eluted by decreasing the ionic strength, changing the buffer A to buffer B, as it is visible from the pre-miR band resultant from the PCR amplification, in lane 2 of the agarose gel electrophoresis (Figure 11b).

To explore this result little further, trying to isolate the pre-miR from other sRNA species, it was tested a three step gradient: 2 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5; 1.6 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5 and finally 10 mM Tris-HCl, pH 7.5.

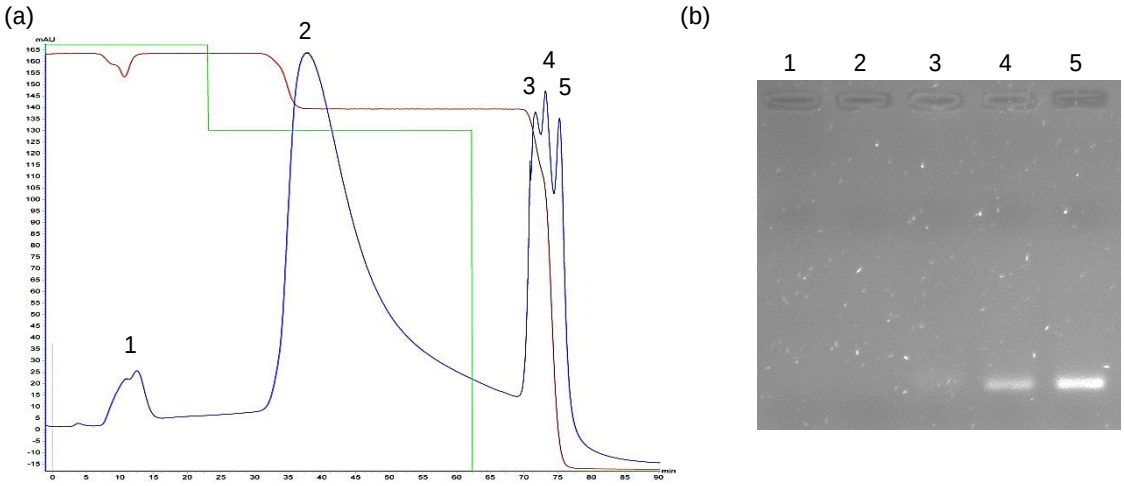


Figure 12: (a) Chromatographic profile of RNAs using a stepwise gradient: 2 M $(NH_2)_4SO_2$ in 10mM Tris-HCl, pH 7.5; 1.6 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5 and finally 10 mM Tris-HCl, pH 7.5b) Agarose gel electrophoresis of cDNA synthesized using each chromatographic fraction (1-5) as template**.**

After analysis of Figure 12, it was possible to conclude that the molecule of interest eluted on the last non-resolved peak, distributed between the fractions 4 and 5. This demonstrated that pre-miR elution occurs when the ammonium sulfate concentration is lower than 1.6 M in 10 mM Tris-HCl, pH 7.5. This result was important because it gives some crucial information to optimize the gradient in order to accomplish the pre-miR purification, as discussed below.

Subsection 3.2.2 - O-phospho-l-tyrosine affinity chromatography for pre-miR29b-1 purification using "one at a time" approach.

To optimize a chromatographic separation process, conventionally, researchers use their results to plan the next experiment. Mistakes or successes lead the investigation path, till the optimum result is obtained. This method is commonly called of "one-variable-at-a-time" (OVAT) methodology [80] or simply "one at a time". Here, the objective was to optimize the pre-miR29b-1 purification process based on the results mentioned on subsection 3.2.1 and using new elution strategies.

Thus, in order to isolate the pre-miR molecule, the initial conditions were maintained and an additional elution step was included, trying to separate the species co-eluting in the last peak. Consequently, the gradient steps were: 2 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5; 1.6 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5; 500 mM $(NH_2)_4SO_2$ in Tris-HCl, pH 7.5 and finally 10 mM Tris-HCl, pH 7.5 (Figure 13). In this experiment, the first peak was not collected because it does not contain the pre-miR, as previously verified.

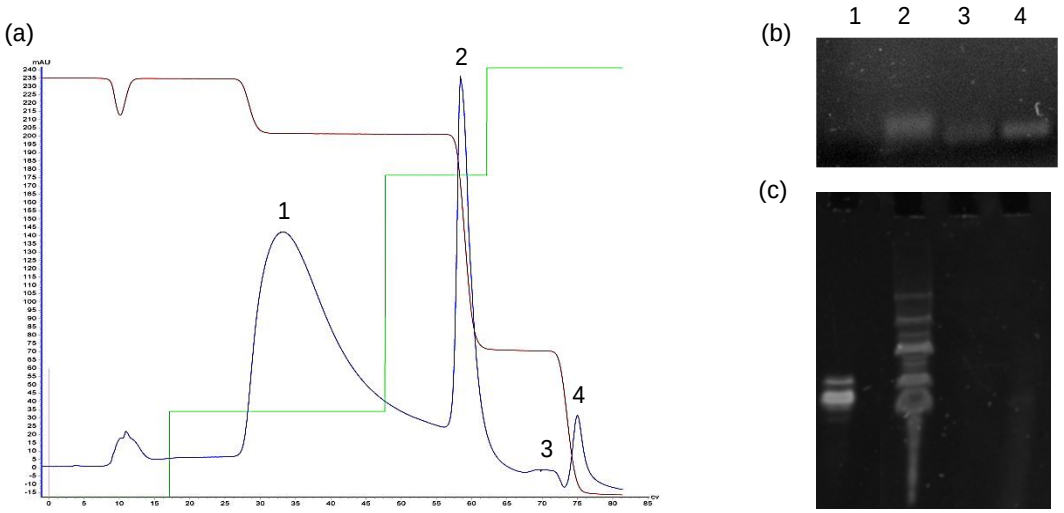


Figure 13: (a) Chromatographic profile of RNAs eluting with a stepwise gradient: 2 M $(NH_2)_4SO_2$ in 10mM Tris-HCl, pH 7.5; 1.6 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5; 500 mM $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5 and finally 10 mM Tris-HCl, pH 7.5. (b) PCR products obtained using the fractions 1-4 as template and (c) polyacrylamide gel electrophoresis of corresponding chromatographic peaks.

Although samples were desalted in a PD-10 column, this method was not totally efficient, causing some distortion, interference and failure to see all the bands on polyacrylamide gel electrophoresis profile. This effect was caused by the remaining salt on samples, and also by the loss of RNAs present on small peaks [81]. In addition, this desalting method can be considered an inefficient technique for RNA, because all the samples were treated at room

temperature which can cause denaturation and degradation of the RNAs and also influence the quantification on the further step. At this point, it was not possible to access the full RNAs content in the samples by electrophoresis (Figure 13c) and only cDNA allowed to verify that the pre-miR was eluted in the fractions 2, 3 and 4, corresponding to the last two gradient steps (500 mM $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5 and 10 mM Tris-HCl, pH 7.5).

Despite these limitations and taking into count the previous results, the next strategy was developed in order to elute more contaminants in the second gradient step and recovering pre-miR with higher purity. As pre-miR eluted in the final peaks, the second gradient step was changed to 1.4 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5. Figure 14 shows the results obtained with these conditions.

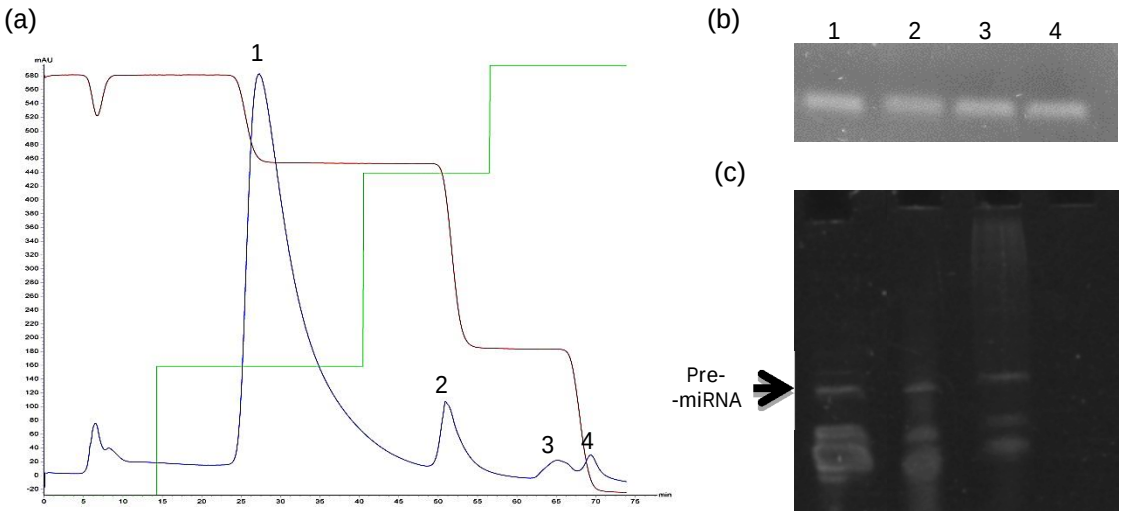
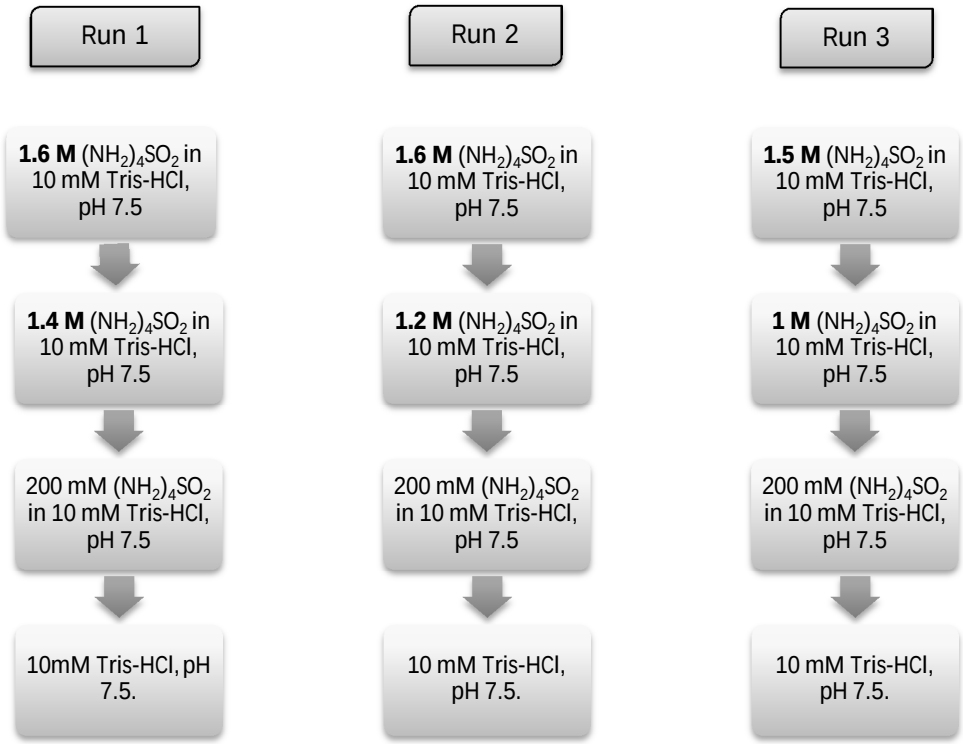


Figure 14: (a) Chromatographic profile of RNAs eluting with a stepwise gradient: 2 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5; 1.4 M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5; 500 mM $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5 and finally 10 mM Tris-HCl, pH 7.5. (b) PCR products obtained using the fractions 1-4 as template. (c) Polyacrylamide gel electrophoresis analysis of chromatographic peaks.

The polyacrylamide gel and the PCR results showed that the pre-miR started to elute on the second gradient step. It was concluded that this salt concentration was too low and some pre-miR was lost at this step. This indicates that further optimization is required to obtain pure pre-miR.

As the initial step did not contain pre-miR, it was hypothesized that it could be eliminated from the chromatographic strategy, and the process could become simpler. However, it was verified that changes on the first step interfere directly on the binding of RNAs to the matrix. Meanwhile, it was intended to understand how the first step affected the pre-miR elution, and for that different gradients where defined, as represented below.



The resulting chromatograms and polyacrylamide gel electrophoresis for each run are presented in follow (Figures 15-17).

Run 1:

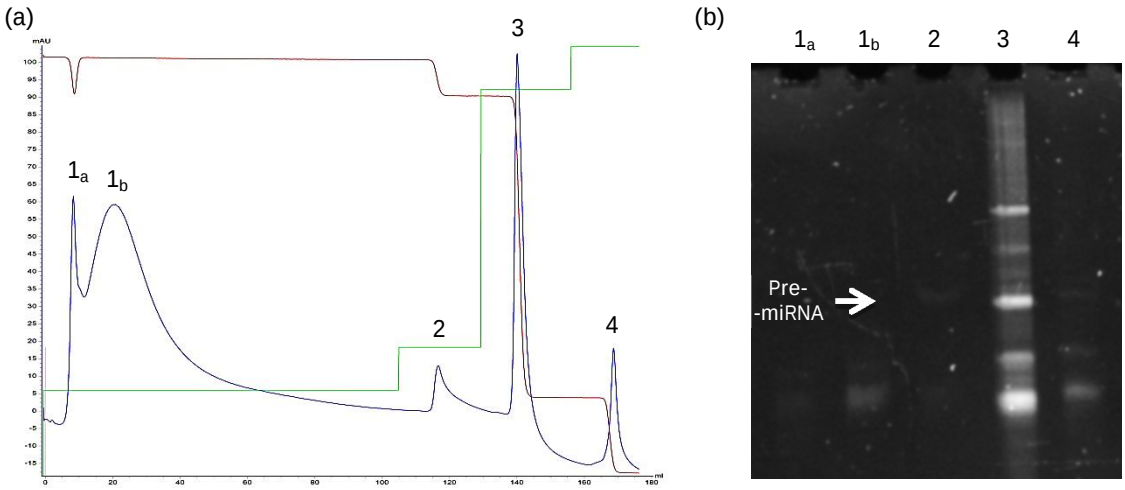


Figure 15: (a) Chromatographic profile of RNAs elution resulting from Run 1. (b) Polyacrylamide gel electrophoresis analysis of 1a, 1b and 2 to 4 chromatographic peaks.

Run 2:

(a) (b)

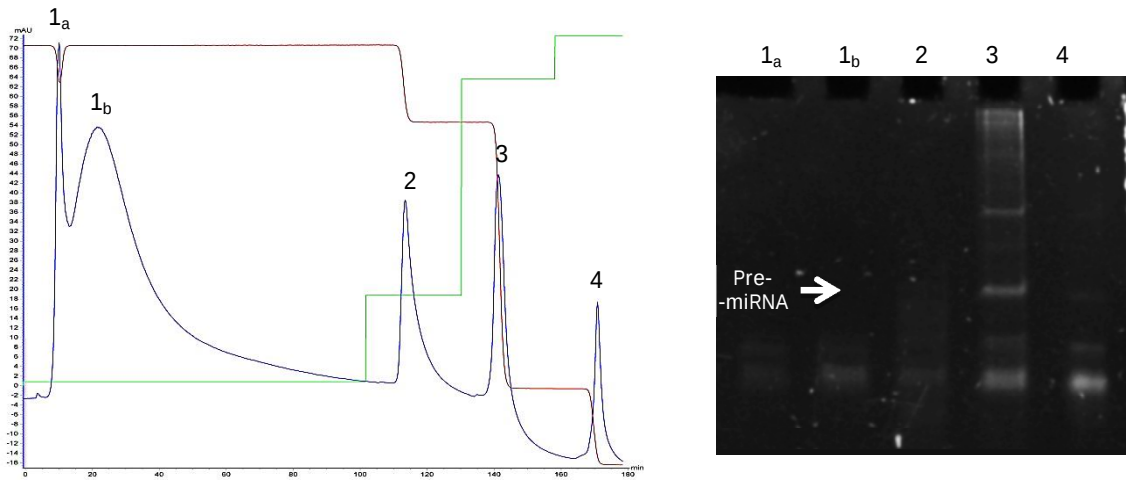


Figure 16: (a) Chromatographic profile of RNAs elution resulting from Run 2. (b) Polyacrylamide gel electrophoresis analysis of 1a, 1b and 2 to 4 chromatographic peaks.

Run 3:

(a) (b)

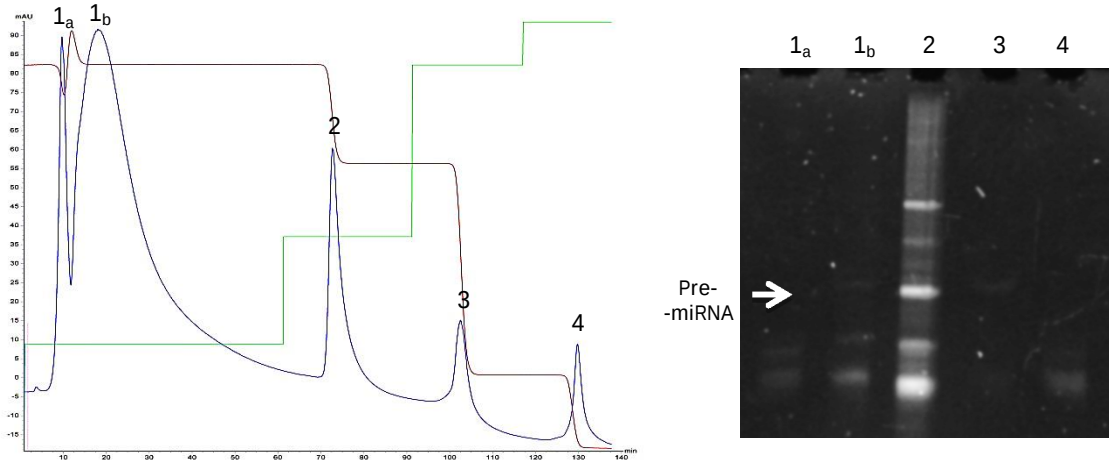


Figure 17: a) Chromatographic profile of RNAs elution resulting from Run 3. (b) Polyacrylamide gel electrophoresis analysis of 1a, 1b and 2 to 4 chromatographic peaks.

As the PD-10 column showed to be inefficient to concentrate the samples, these collected samples were desalted and concentrated with concentrators (6 mL volume) of 10 000 MWCO. First, these concentrators suffer a pseudo-fouling of their membrane in order to maintain all RNAs retained. The fouling was performed with 2M $(NH_2)_4SO_2$ in 10 mM Tris-HCl, pH 7.5.

The first peak was hardly to be observed on electrophoresis (referred to fractions 1_a and 1_b), probably because the smallest RNAs could have crossed the membrane and so, they appeared with less intensity. Nevertheless, the decrease of the ionic strength in the first gradient step proved to be a good strategy, as miRNA binds onto the column similarly to when a 2 M $(NH_2)_4SO_2$ step was performed. In addition, the reduction of the number of steps on a chromatographic procedure is essential, because more steps involve more time, more instability to the RNAs, more quantity of buffers and so a more expensive and less useful method.

Comparing the results from runs 1 and 2 it was possible to conclude that both, 1.6 M and 1.5 M, are good initial conditions to provide the binding of miRNA to the matrix. Furthermore, the result of run 2 also shows that when using 1.5 M of salt the pre-miR is bound to the column and more contaminants are eliminated, not interfering with the target elution. These experiments also revealed that pre-miR only eluted at 0.2 M $(NH_2)_4SO_2$. On the other hand, experimental conditions used in Run 3 caused the elution of most pre-miR on the second step, using 1 M of ammonium sulfate. Overall, the results indicated that pre-miR is mostly retained at 1.4 M of ammonium sulfate and 1 M of this salt is sufficient to induce its total elution.

In this way, the next strategy was defined in order to establish an initial condition with the lowest salt concentration possible to retain the pre-miR, but able to eliminate more contaminants. The second step was defined trying to find a condition that selectively bind and elute pre-miR. Therefore, a new set of experiments were developed, as represented bellow.

The resulting chromatograms and polyacrylamide gel electrophoresis for each run are presented in Figures 18, 19 and 20.

Run 4:

(a) (b)

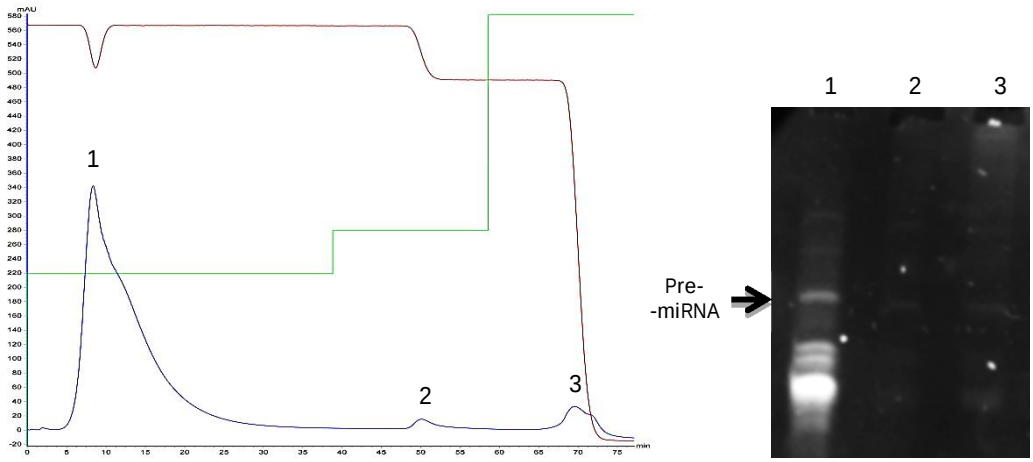


Figure 18: Chromatographic profile of RNAs elution resulting from Run 4. (b) Polyacrylamide gel electrophoresis analysis of 1-3 chromatographic peaks.

Run 5:

(a) (b)

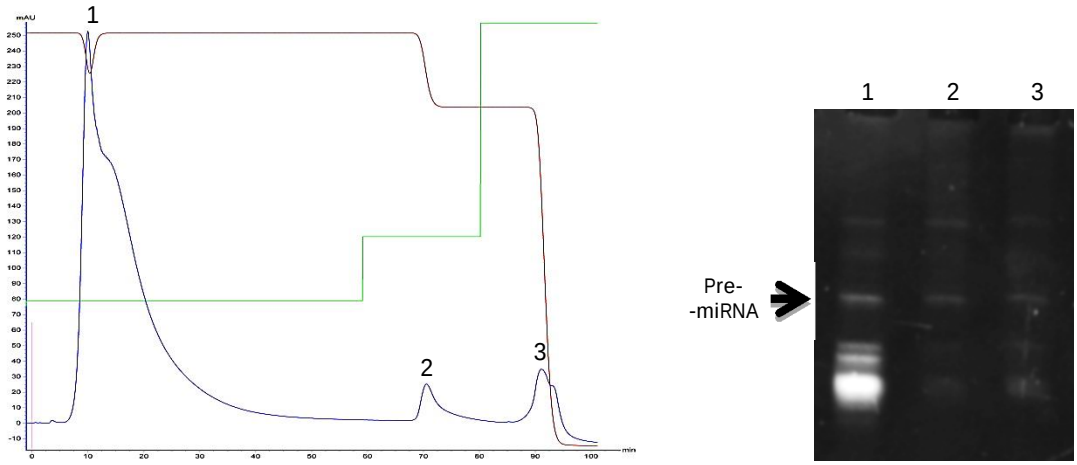


Figure 19: (a) Chromatographic profile of RNAs elution resulting from Run 5. (b) Polyacrylamide gel electrophoresis analysis of 1-3 chromatographic peaks.

Run 6:

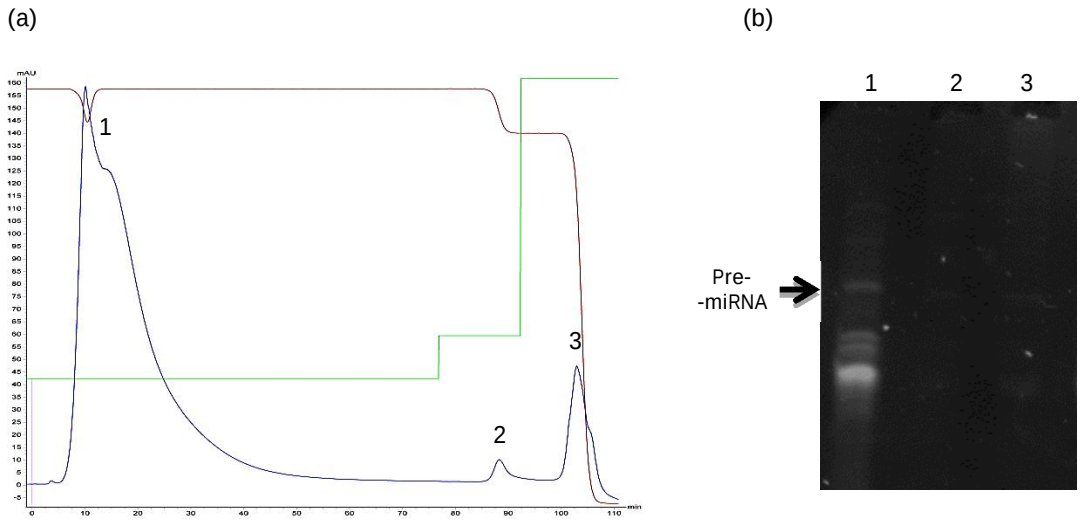


Figure 20: (a) Chromatographic profile of RNAs elution resulting from Run 6. (b) Polyacrylamide gel electrophoresis analysis of 1-3 chromatographic peaks.

With these new results it was verified that lowering the salt concentration of the initial step did not provide strength enough to bind the pre-miR to the column and it eluted in the flowthrough with other contaminants. Analysing these results it is concluded that the first condition had to be set to 1.6 M or 1.5 M of ammonium sulfate. In order to compare these two conditions, two experiments were performed with an identical start concentration, but testing a different condition on the second step (1.25 and 1.3 M of salt).

The resulting chromatograms and polyacrylamide gel electrophoresis for each run are presented in Figures 21, 22 and 23.

Run 7:

(a) (b)

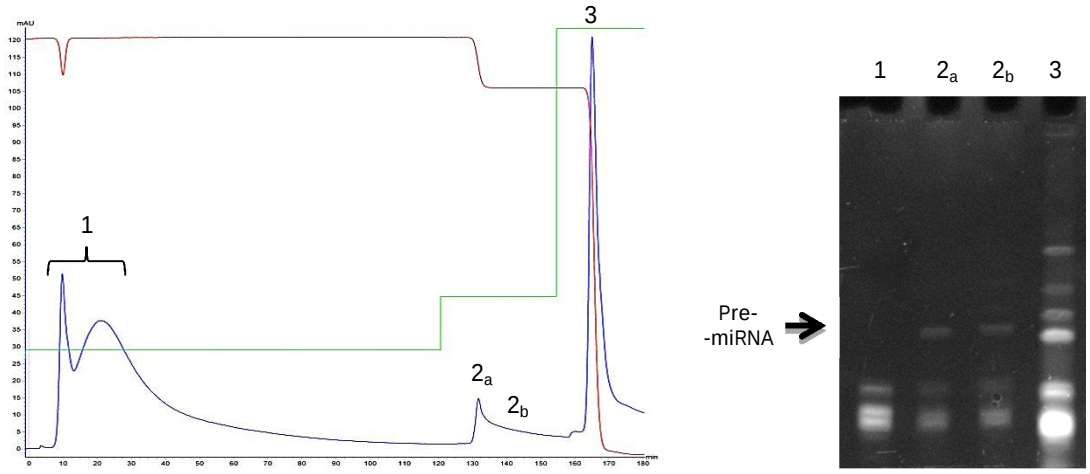


Figure 21:(a) Chromatographic profile of RNAs elution resulting from Run 7. (b) Polyacrylamide gel electrophoresis analysis of 1, 2 (a,b) and 3 chromatographic peaks.

Run 8:

(a) (b)

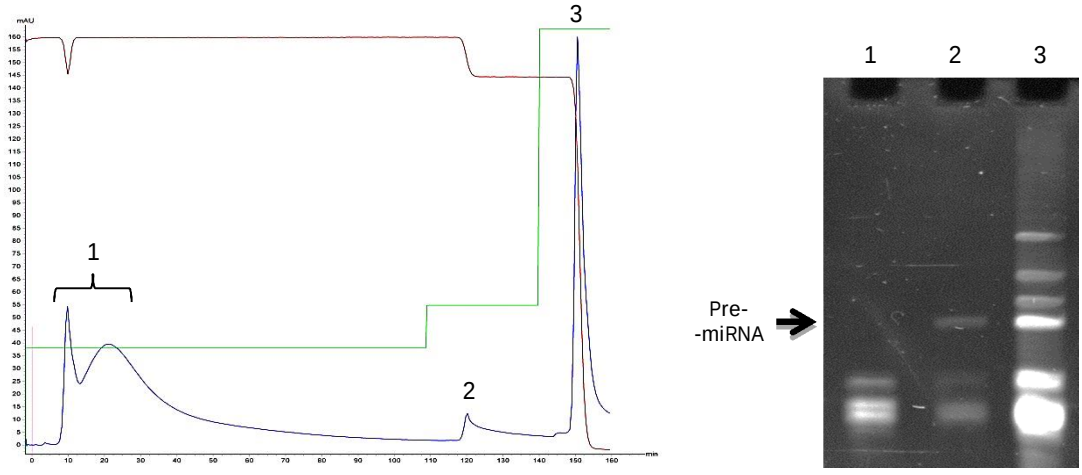


Figure 22: (a) Chromatographic profile of RNAs elution resulting from Run 8. (b) Polyacrylamide gel electrophoresis analysis of 1-3 chromatographic peaks.

Run 9:

(a) (b)

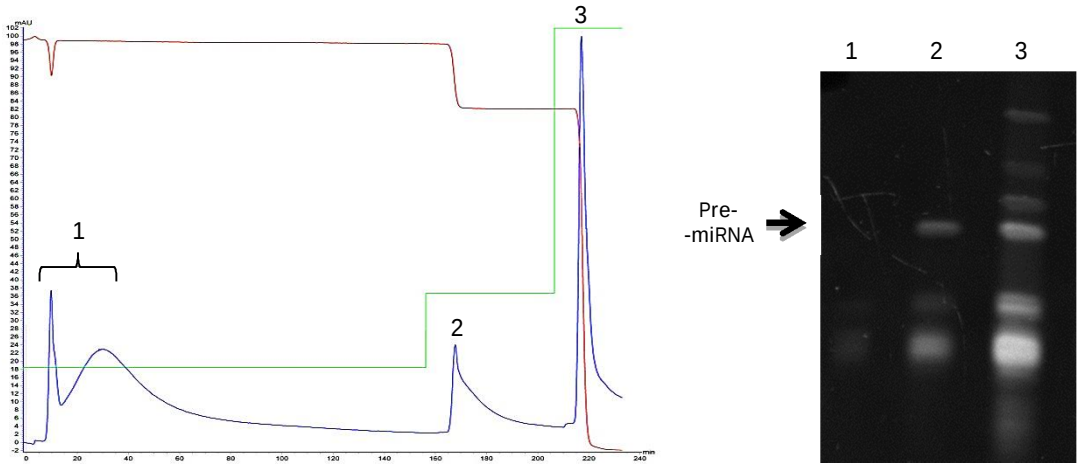


Figure 23: (a) Chromatographic profile of RNAs elution resulting from Run 9. (b) Polyacrylamide gel electrophoresis analysis of 1-3 chromatographic peaks.

In the run 7, the intermediate peak was collected on top (2_a) and also on its tail (2_b) to see if there was different species of RNAs eluting. The polyacrylamide electrophoresis showed that the RNAs eluting on the top and on the tail of the second peak are similar.

Analysing all the chromatograms, 1.6 M of $(NH_2)_4SO_2$ was chosen to be the initial condition on this purification process. This decision was made because this condition provided a good ionic strength to efficiently bind pre-miR to the matrix and it appeared to be essential to the selectivity. Regarding the second step, the concentration of 1.25 M was not sufficient to totally elute the pre-miR and a large quantity was retained only eluting on the final peak. For a better result, that is, to have lower contamination present on pre-miR peak it was tried to use a cleaning step of 1.5 M $(NH_2)_4SO_2$ as a second step and then a 1.1 M step was chosen instead of the 1.25M, since this condition was not able to efficiently elute pre-miR.

Figure 24 shows the result for the final run characterized by the following gradient: 1.6 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl, pH 7.5; 1.5 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl; 1.1 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl and 10 mM Tris-HCl, pH 7.5.

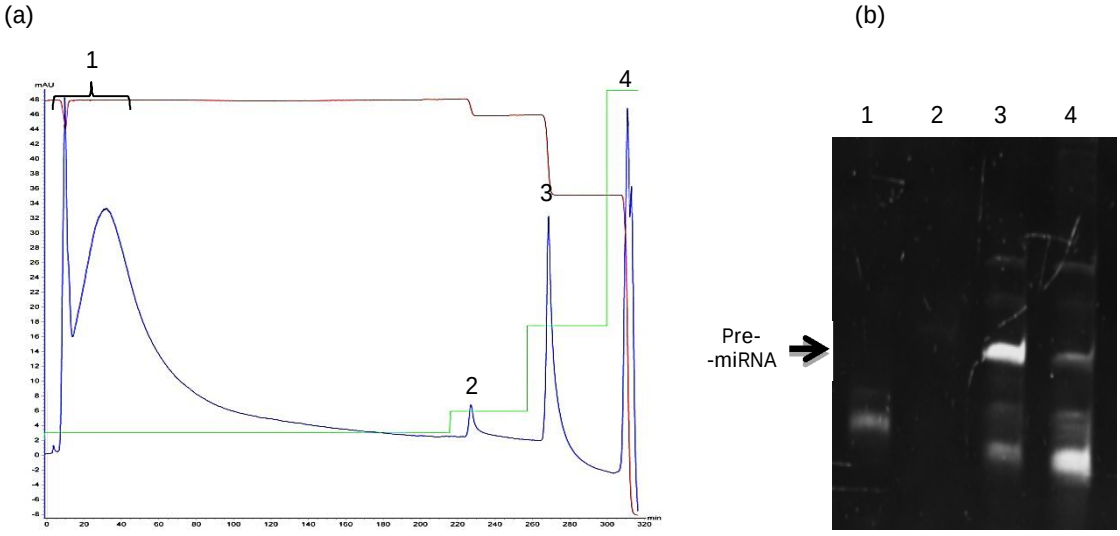


Figure 24: (a) Chromatographic profile of RNAs elution resulting of 1.6 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl, pH 7.5; 1.5 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl; 1.1 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl and 10 mM Tris-HCl, pH 7.5 (b) Polyacrylamide gel electrophoresis analysis of chromatogram peaks.

Despite the presence of a residual quantity of pre-miR in the last peak, this was the experiment conducting to the best result.

Using Phoretix 1D software was possible to achieve the percentage of recovery and purity of pre-miR29b-1. The relative recovery (% relative Recovery) was obtained by measuring the intensity of the pre-miR band in the peak 3 relatively to other pre-miR bands obtained in other peaks. On the other hand, the relative purity (% relative Purity) was acquired measuring the intensity of the pre-miR band relatively to all the other bands present within the pre-miR peak. In this case, after 59 experiments done, this experiment resulted in a relative Recovery of 71% and a purity of 52% of pre-miR29b-1. This was the best result achieved by one at a time method.

Subsection 3.2.3 - O-phospho-L-tyrosine affinity chromatography for pre-miR29b-1 purification using a design of experiments approach.

In a second approach to achieve the pre-miR29 purification, the Box-Behnken Design (BBD) was used to identify optimum conditions for the separation process. It was necessary to define which factors should be studied, concerning their influence on the purification strategy. Thus, the selected factors to be studied were the equilibrium and elution conditions and column temperature. This selection was based on the three most critical factors that affect the chromatographic process, and were evaluated at three levels. To choose the range of each level it was used some of the knowledge obtained from the first experiments of "one at a time method". Each variable was coded by the letters A, B and C respectively, and each one had 3 levels which were -1, 0 and 1, as shown in Table 8.

Table 8: Factors and levels introduced on BBD design.

Code	Factor	Levels		
		Low	Medium	High
		-1	0	1
A	**Equilibrium (M)** ($(NH_4)_2SO_4$ concentration in 10 mM Tris-HCl, pH 7.5)	2	1.7	1.4
B	**Elution (M)** ($(NH_4)_2SO_4$ concentration in 10 mM Tris-HCl, pH 7.5)	1.3	1.15	1
C	**Temperature (º C)**	4	12	20

The concentrations were established according to % of Buffer B (10 mM Tris-HCl, pH 7.5) needed to achieve the concentration defined in the table levels. So, as higher the quantity of buffer B added, the lower was the ammonium sulfate concentration and that is why the low level corresponds to a high concentration of $(NH_4)_2SO_4$.

According to BBD and using UNICORN™ 6.1 software, the following designed matrices of the trial experiments were generated, as it shown in Tables 9 and 10.

Tables 9 and 10: Box-Behken design for three factors at left and the chosen three factors on three levels BBD used for the optimization of pre-miR purification process at right. In red is represented the central point that were replicated two times (experiments 13, 14 and 15). Equilibrium and elution represent the $(NH_4)_2SO_4$ concentration (M) in 10mM Tris-HCl, pH 7.5. Temperature was defined in ºC.

Exp.	Factors		
	A	B	C
1	-1	0	-1
2	1	0	-1
3	-1	0	1
4	1	-1	1
5	-1	-1	0
6	1	-1	0
7	-1	1	0
8	1	1	0
9	0	-1	-1
10	0	-1	1
11	0	1	-1
12	0	1	1
13	0	0	0

Exp.	Factors		
	Equilibrium	Elution	Temp.
1	2	1.15	4
2	1.4	1.15	4
3	2	1.15	20
4	1.4	1.15	20
5	2	1.3	12
6	1.4	1.3	12
7	2	1	12
8	1.4	1	12
9	1.7	1.3	4
10	1.7	1.3	20
11	1.7	1	4
12	1.7	1	20
13	1.7	1.15	12

Tables presented above show that the optimization was based on 15 combined experiments, with two central point replicates. All runs consisted on three step gradients, first, the one given by equilibrium, second, the one given by elution and finally a 10 mM Tris-HCl, pH 7.5 step. To exclude any bias all experimental designs were randomized.

After the conclusion of all experiments, two types of responses were analyzed, the pre-miR29b-1 relative purity within the electrophoresis profile corresponding to the intermediate peak and the relative recovery of pre-miR29b-1 relatively to other collected peaks. The result of a central point experiment is presented in Figure 25.

(a) (b)

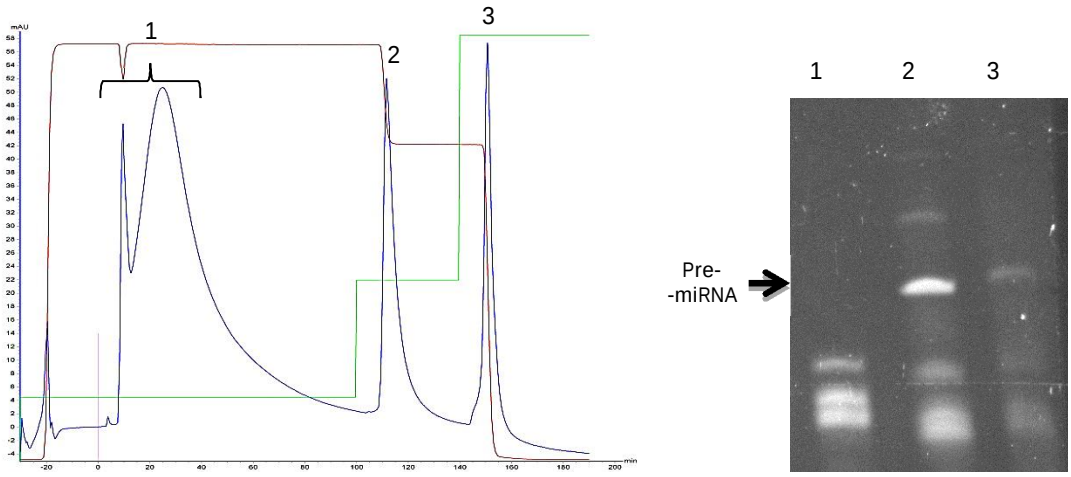


Figure 25: (a) Chromatographic profile of RNAs elution of experiment 13. The applied gradient was 1.7 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl, pH 7.5; 1.15 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl and finally 10 mM Tris-HCl, pH 7.5. The experiment was carried at 12 ºC. (b) Polyacrylamide gel electrophoresis analysis of chromatographic peaks.

Using the Phoretix 1D software in all resulting electrophoresis, the intensity of the bands was analyzed and the purity and recovery of pre-miR measured, as it is depicted in Table 11.

Table 11: Responses obtained for each run defined by BBD design.

Run	Equilibrium (M)	Elution (M)	Temperature (ºC)	% relative Recovery	% relative Purity
1	2	1.15	4	66.23	12.82
2	1.4	1.15	4	6.93	14.20
3	2	1.15	20	96.12	7.26
4	1.4	1.15	20	4.9	15.85
5	2	1.3	12	69.84	13.73
6	1.4	1.3	12	0.20	0
7	2	1	12	96.12	7.26
8	1.4	1	12	11.39	21.33
9	1.7	1.3	4	36.17	8.21
10	1.7	1.3	20	55.42	16.38
11	1.7	1	4	76.86	16.31
12	1.7	1	20	98.93	21.23
13	1.7	1.15	12	78.27	21.25
14	1.7	1.15	12	85.36	31.21
15	1.7	1.15	12	77.7	24.16

The results were introduced as responses on UNICORN™ 6.1 software allowing to access to the statistical data. Below, both results and the resulting model will be discussed.

3.2.3.1 - Goodness of fit

The goodness of fit is given by the statistical coefficients resumed on Table 12. These different statistical calculations, by which the adequacy of a model can be judged, measure how good the model is, because they consider the contribution of all values together [70].

Table 12: Statistical coefficients of the model.

	R^2	R^2 Adj	Q^2	RSD	Model validity	Reproducibility
% relative Recovery	0.9803	0.9449	0.7130	8.386	0.5271	0.98857
% relative Purity	0.9301	0.8044	0.7755	3.359	0.9953	0.5453

The graphical representation of the previous table is presented next in Figure 26.

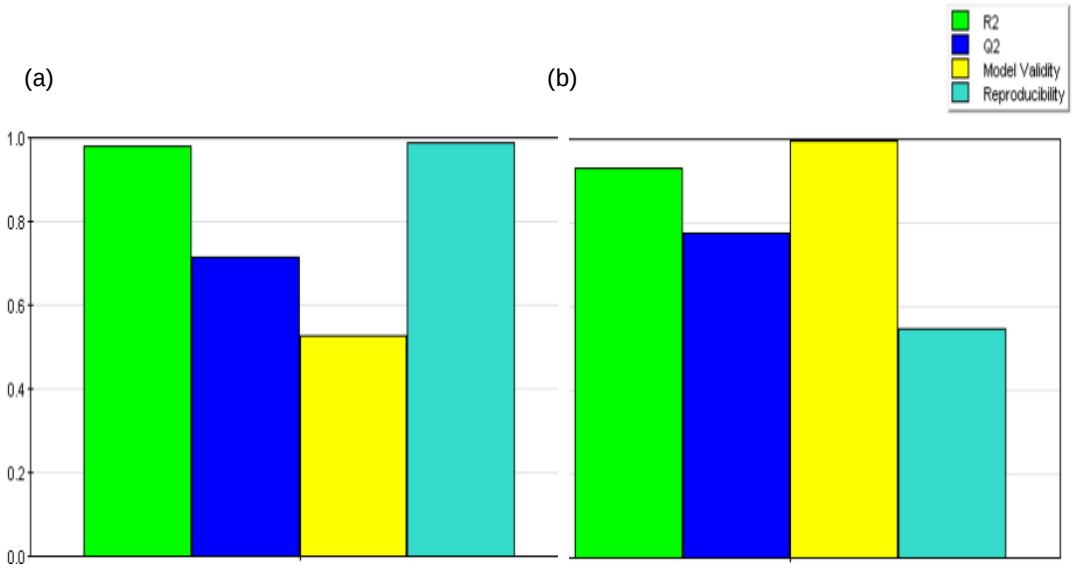


Figure 26: (a) Graphical representation of statistical coefficients obtained for both responses: % relative Recovery response (a) and % relative Purity response (b).

According to the presented analysis the model fitted the data. The explained variation, R^2, has to be between 0 and 1, with larger values being more desirable. Here, the relative Recovery value (0.9803) was better than the relative Purity value (0.9308), but both represented good coefficients, which indicated a high significance of the model [70, 82, 83].

On the other hand the "adjusted" R^2 (R^2 Adj) is a variation of the ordinary R^2 and is adjusted for the "size" of the model, that is the number of the factors. If the adjusted R^2 decreases, non significant terms are added to the model. The present results show that the model had a good adjusted R^2 on both responses [70, 82, 84, 85].

Furthermore, the predicted variation values (Q^2) were slightly better for the % relative Purity. So, the model fitted better the first response data and for the second response, the power of prediction was higher. However, the number was near of one revealing the model adequacy to the introduced data [70].

Regarding reproducibility, the percentage of relative Recovery originated a good reproducible model, while the percentage of Purity revealed a lower experimental control where data could be more variant. These statistical values guaranteed a good model, because the model validity, which reflects if the right type of model was chosen from the beginning in the problem formulation, was above 0.25 for both responses [70].

3.2.3.2 - Residual plots

The following normal probability plots (Figure 27) show the standardized residuals and normal percent probability plot. Residual plots are very effective for detecting outliers as they appear far above or below the bulk of the plotted residuals [86].

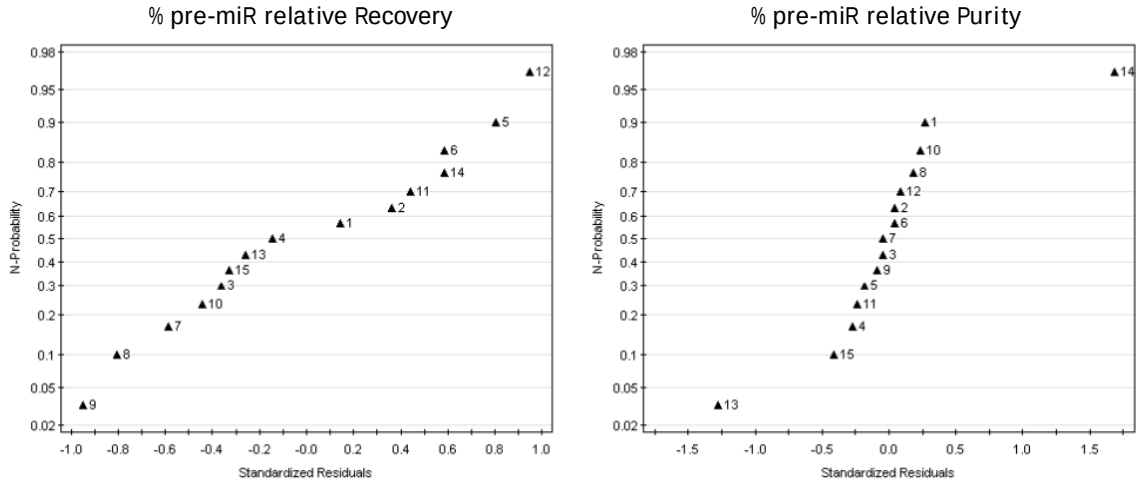


Figure 27: Normal probability plots for residuals of the two responses. The experiment numbers are labelled on the graphic.

Here, these graphical representations suggested that the main reason of a lack of reproducibility was the existence of two outlying experiments. Nevertheless, these values were maintained on the model because they fitted well in the modeling of the other response variable. All of the other standardized residuals were random and normally distributed,

having their points lying on a straight line within a standardized residuals range of -4 to +4 standard deviations [70]. Thus, according to UNICORN™ 6.1 software, the errors are normally distributed, supporting the fact that the model fits the data adequately.

3.2.3.3 - Observed versus Predicted for each response

These plots (Figure 28) show the observed versus predicted values of the different responses allowing to see the difference between the observed value of a response measurement and the value that is fitted under the theorized model [83]. With a good model all the points should fall on the 45 degree line [70].

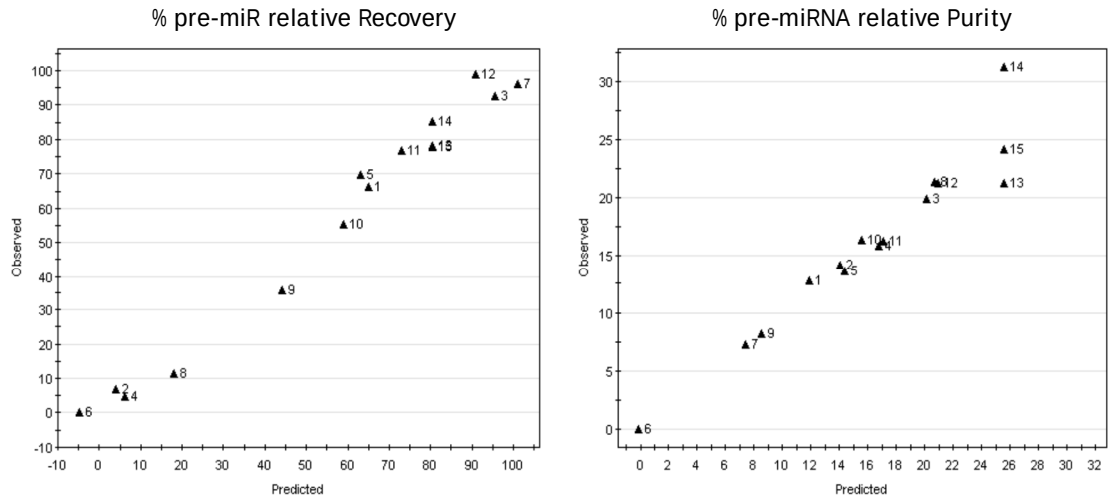


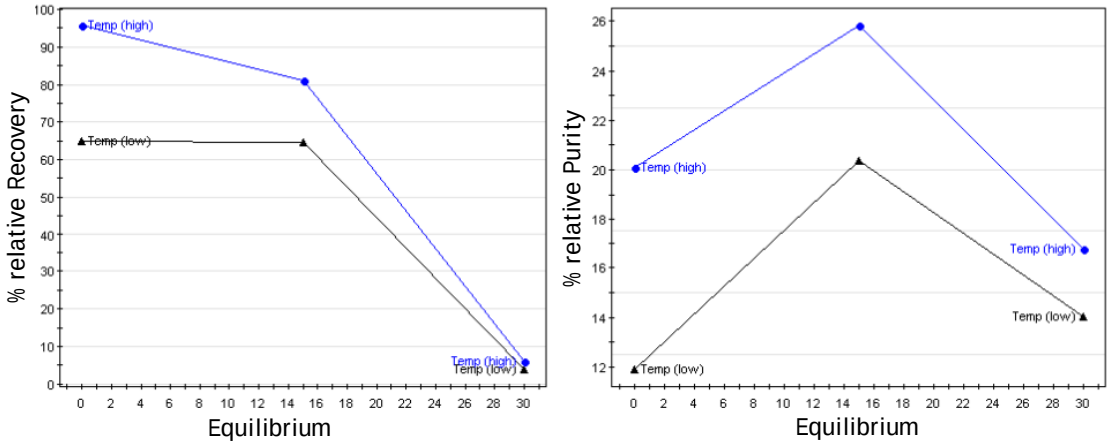
Figure 28: Observed versus predicted values for relative Recovery and Purity.

According to the data resulting from UNICORN™ 6.1 software (Figure 28), the observed and predicted percentage of miRNA plotted values suggested a satisfactory correlation between experimental to predicted value, because the data points are located close to the diagonal line. The outliners were confirmed on this representation as they were the unique points distant from the expected result.

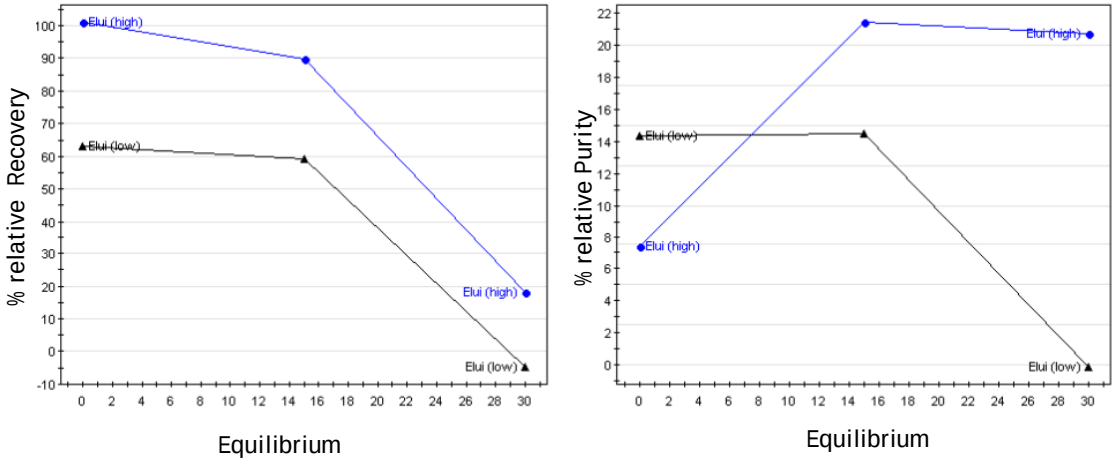
3.2.3.4 - Interaction plots

An interaction plot shows the impact of changing the settings of one factor in another factor, as it is presented in Figure 29. Because an interaction can magnify or diminish main effects, the evaluation of interactions is extremely important.

(a)



(b)



(c)

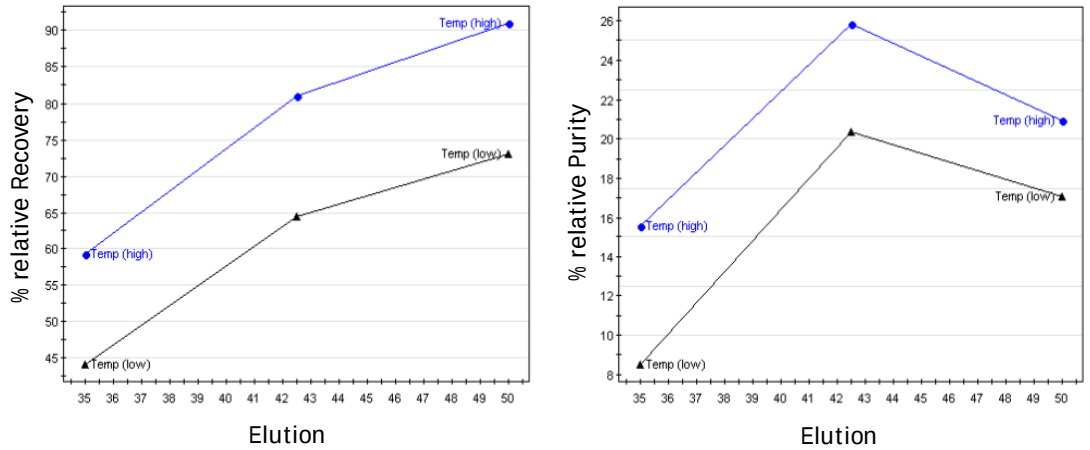


Figure 29: Interactions plots of the factors towards the added responses: (a) Interaction plot of equilibrium*temperature; (b) interaction plot of equilibrium*elution; (c) interaction plot of temperature*elution.

The Table 13 briefly resumes the interaction expressed on the interaction plots.

Table 13: Interaction between factors relative to both responses.

	Interaction	
	% relative Recovery	**% relative Purity**
Equilibrium*Temperature	Yes	Yes
Equilibrium*Elution	No	Yes Strong
Temperature*Elution	Yes Weak	No

The interaction between factors seems to be more intense relatively to % relative Purity response, where the equilibrium factor interacts more significantly with other two variables. Particularly, when equilibrium*temperature interaction is analyzed (Figure 29 (a)) it is possible to conclude that these two conditions interact in both responses, because the draw lines were not parallel. However, the lines do not cross each other, which means that the interaction were not too strong. Regarding the graphics profiles, it is observable that the higher temperature level result on better responses values. On the equilibrium*elution representation (Figure 29 (b)) there is only interaction in the % relative Purity response. Because the lines cross each other, these two factors are intimately related to the resulting response. To the % relative Recovery response there is no interaction as the graphic lines are parallel. Relative to temperature*elution (Figure 29 (c)) there is a weak interaction, on % relative Recovery response, as the lines are almost parallel but not totally. On the other hand there is no interaction when was considered the % purity response.

3.2.3.4 – Main effects

Through the multiple regression equation it is possible to access the main effects (summarized in Table 14) of the factors. Using the Design Expert 7.0 software the final coded equations (1 and 2) for the two responses (R1 - % relative Recovery and R2- % relative Purity) were obtained:

$$\mathbf{R1} = +80.44 - 37.68A + 15.21B + 8.22C - 3.78AB - 7.12AC + 0.71BC - 30.11A^2 - 5.95B^2 - 7.65C^2 \quad (1)$$

$$\mathbf{R2} = +25.54 - 0.29A + 3.48B + 2.73C + 6.95AB - 1.37AC - 0.81BC - 7.40A^2 - 7.56B^2 - 2.44C^2 \quad (2)$$

Table 14: Summary of the main effects.

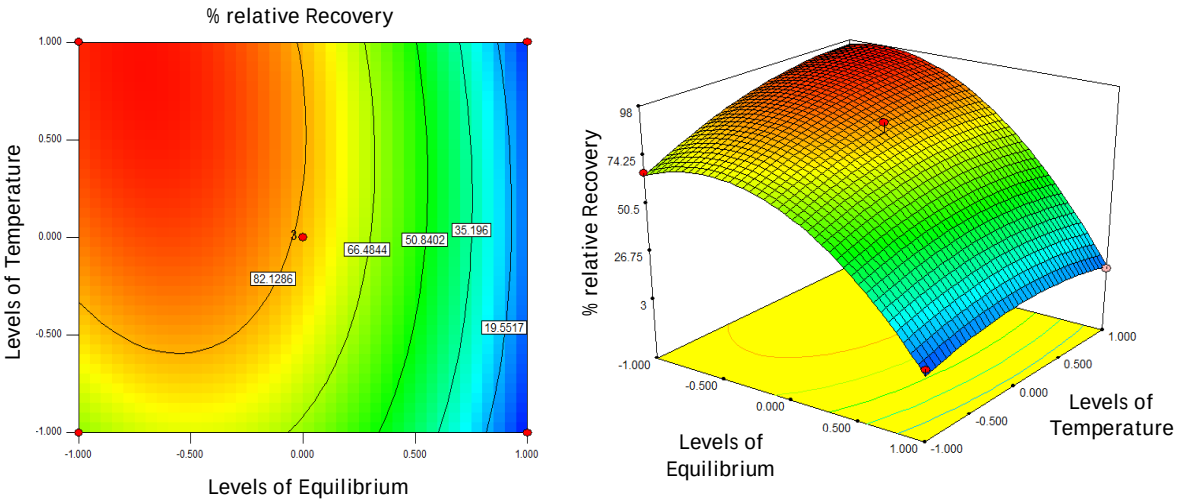
Factor	Effect R1	Effect R2
Equilibrium (A)	Negative	Negative
Eluition (B)	Positive	Positive
Temperature (C)	Positive	Positive
Equi*Eluition (AB)	Negative	Positive
Equi*T (AC)	Negative	Negative
T*Elution (BC)	Positive	Negative
Equilibrium^2 (A^2)	Negative	Negative
Elution^2 (B^2)	Negative	Negative
Temperature^2 (C^2)	Negative	Negative

Analyzing the effects, individual variables show to be related to a positive effect, that is, when the variable is on the high level the response is higher. Here, only the equilibrium has a negative effect related to R1 and R2. This is explained by the fact that, for example, when equilibrium is on the lowest level (2 M of ammonium sulfate) it causes more retention of the miRNA in the column. The elution will always contribute with more miRNA when in the first condition this target molecule does not elute.

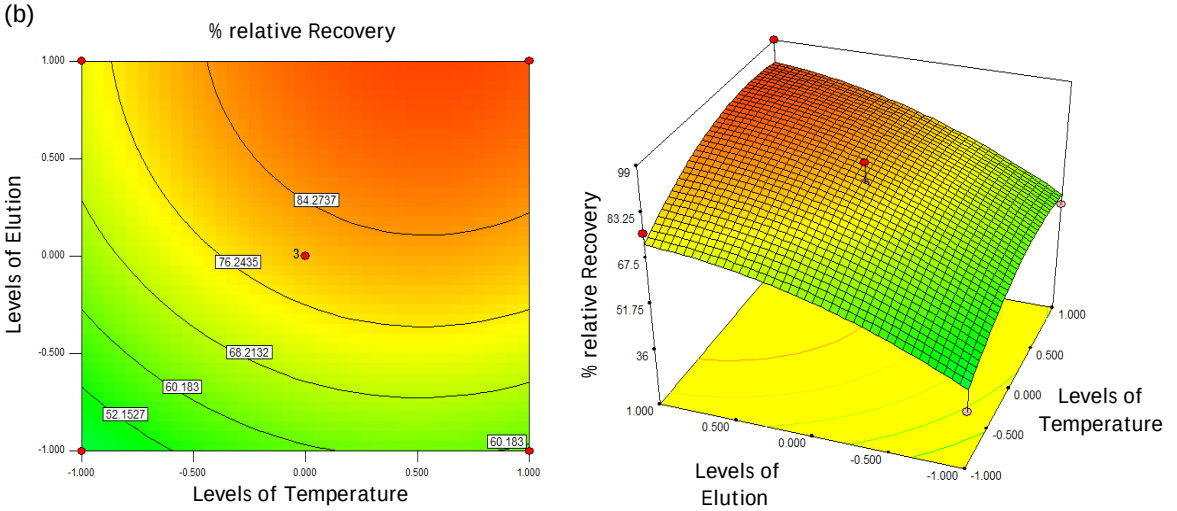
3.2.3.5 - Surface and contour plots

The Surface plots of response functions are useful in understanding the main interaction effects of the factors, but also to find their optimum levels. These graphics are the representation of the resulting model regression equation [84, 87]. The shapes of contour plots, circular or elliptical, determine if the mutual interactions between variables are significant or not [87]. When interaction is occurring the contour plot presents an elliptical nature, and if it is not a circular shape is represented. The Figure 30 shows the contour plot and the 3D response surface plot of the effect of the variables on the % relative Recovery response, while the Figure 31 shows the contour plot and the 3D response surface plot of the effect of the variables on the % relative Purity response. All graphics were obtained using the Design Expert 7.0 software.

(a)



(b)



(c)

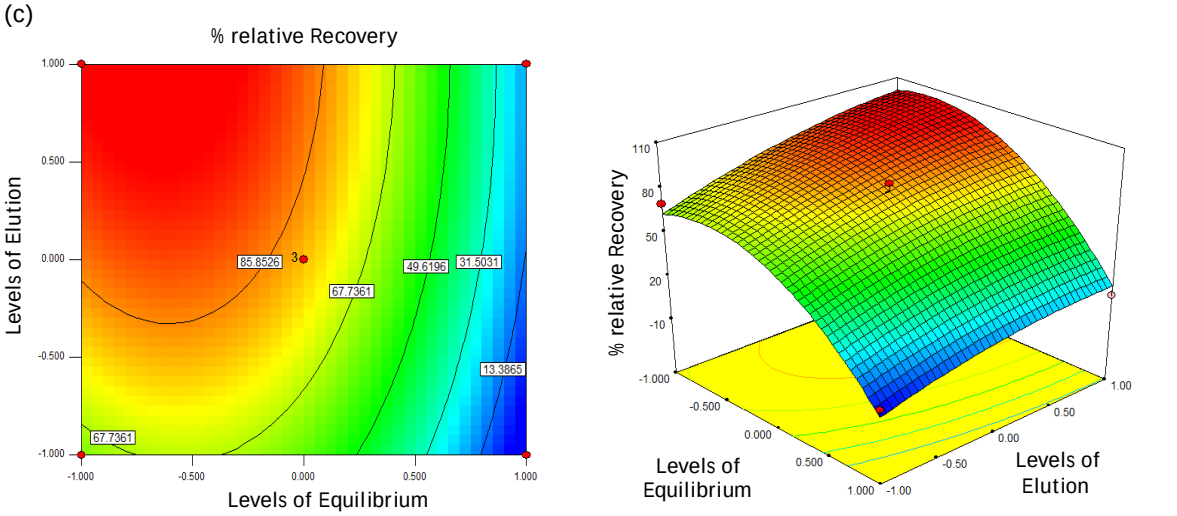
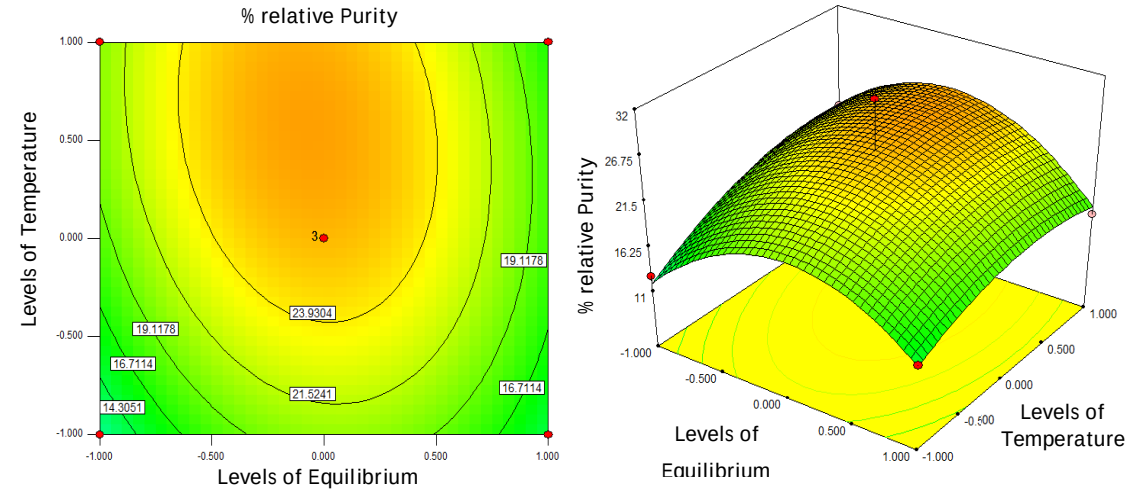
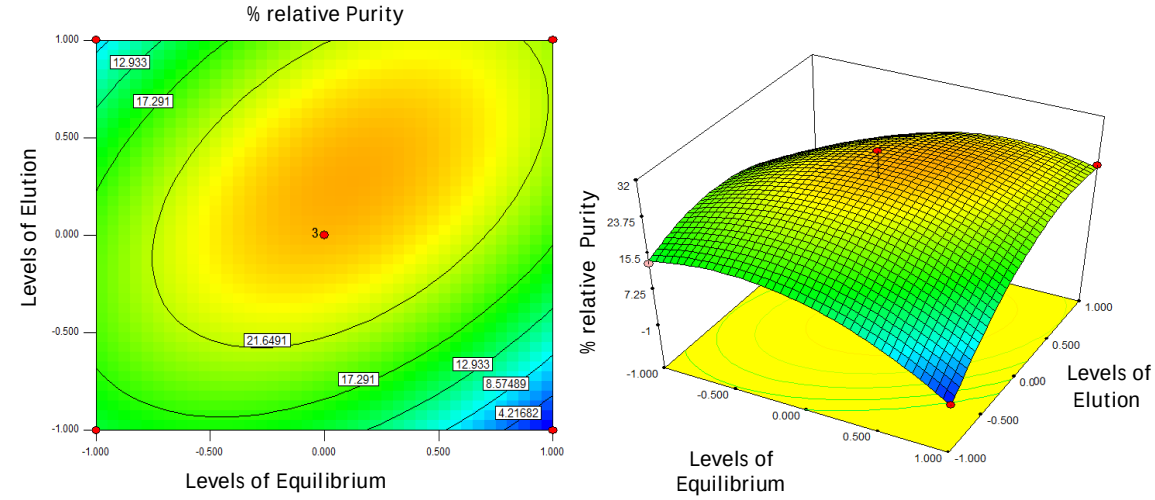


Figure 30: Three-dimensional response surface plot and contour plot of interactions of variables equilibrium, elution and temperature and its effect on % relative Recovery. One parameter for each graph is at a hold value, corresponding to the central point of the absent variable.

(a)



(b)



(c)

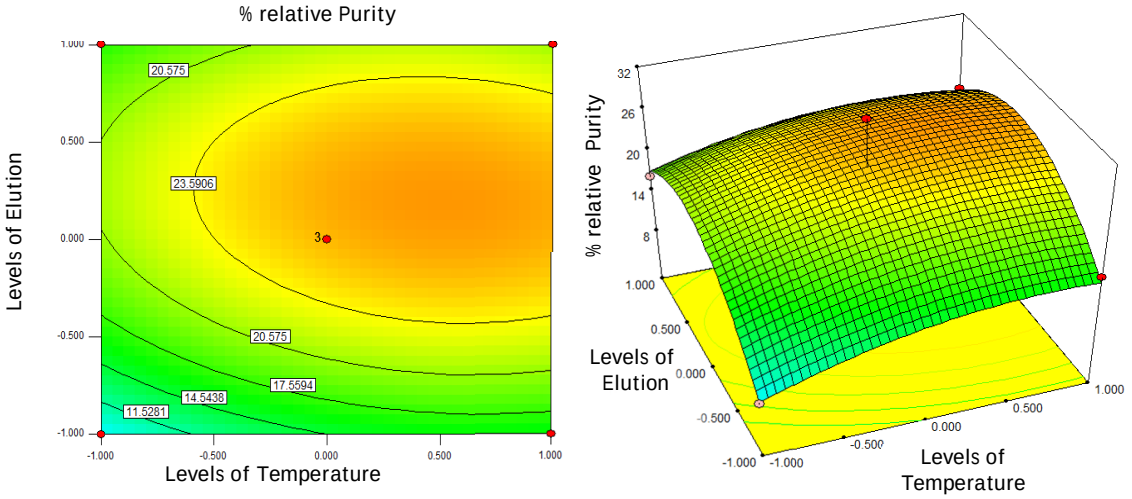


Figure 31: Three-dimensional response surface plot and contour plot of interactions of variables equilibrium, elution and temperature and its effect on % relative Purity. One parameter for each graph is at a hold value, corresponding to the central point of the absent variable.

The results providing by Figure 30 are in agreement with the interactions plots previous presented. It is clear the ellipticity of the contour plot of equilibrium versus temperature and of elution versus equilibrium, where interaction between the two variables occurs. The graphical representations of the interaction of different variables presented on Figure 31 also agree with the previous analysis made with the interaction plots. Here the ellipticity is present on the temperature versus equilibrium and equilibrium versus elution graphics, demonstrating the interaction occurring between the variables.

3.2.3.6 - ANOVA table

Analysis of variance, ANOVA, is an important diagnostic tool in regression analysis. It is based on partitioning the total variation of a selected response into one part due to the regression model and another part due to the residuals. Because were made replicates of the central point, ANOVA can also decompose the residual variation into one part related to the model error and another part linked to the replicate error [70, 86]. As it permits the estimation of different types of variability in the response data and their comparison with each other by means F-tests it reveals the significance of the model (given by p-Value) and the good fit of the data (given by the lack of fit), as it shown in Tables 15 and 16 [70].

Table 15: ANOVA table for the % relative Recovery response.

% relative Recovery	Degrees of freedom	Sum of squares	Mean square	F-value	P-value	Standard Deviation
Total	15	66817.1	4454.48			
Constant	1	48962.1	48962.1			
Total corrected	14	17855	1275.36			
Regression	9	17503.4	1944.83	27.6573	**0.001**	44.1002
Residual	5	351.593	70.3187			8.38562
Lack of fit (model error)	3	315.17	105.057	5.76873	**0.151**	10.2497
Pure error (Replicate error)	2	36.4229	18.2114			4.26749

Adeq Precision = 15.449

Table 16: ANOVA table for the % of relative Purity response.

% relative Purity	Degrees of freedom	Sum of squares	Mean square	F-value	P-value	Standard Deviation
Total	15	4772.38	318.159			
Constant	1	3964.84	3964.84			
Total corrected	14	807.541	57.6815			7.59483
Regression	9	751.129	83.4587	7.39722	**0.020**	9.13558
Residual	5	56.4122	11.2824			3.35893
Lack of fit (model error)	3	3.95481	1.31827	0.05026	**0.981**	1.14816
Pure error (Replicate error)	2	52.4574	26.2287			5.1214

Adeq Precision = 9.364

Through analysis of variances (ANOVA) quadratic regression model demonstrated that the model is highly significant. The significance probability (P-value) was considered to be significant when was minor than 0.05, when comparing modellable and unmodellable variance. Because in both responses P-value is lower than 0.02, the data is statistically significant.

The Lack of Fit expresses if the model is adequate to describe the observed data or if a more complicated model should be used. This parameter is achieved by comparing the variability of the current model residuals to the variability between observations and replicate settings of the factors [83]. That said, as the lack of fit of P-value was found to be non significant, inferior to 0.05, for both responses, it suggests that the model equation was adequate to predict the percentage of relative miRNA recovery and purity under any sets of variables combination [88].

According to Design Expert 7.0 analysis, in % relative Recovery the model F-value of 27.66 implies that the model is significant. There is only a 0.10% chance that a "Model F-Value" this large could occur due to noise. In the same way, in % relative Purity the model F-value of 7.40 implies the model is significant. There is only a 2.01% chance that a "Model F-Value" this large could occur due to noise.

The Adeq Precision value measures the signal to noise ratio, and a value greater than 4 is desirable. The % relative Recovery and % relative Purity have 15.449 and 9.364 Adeq Precision, respectively, indicating an adequate signal. So, this model can be used to navigate the design space.

3.2.3.7 - Model validation

For the model validation three assays were executed, replicates of the optimum point. The optimum separation conditions were identified introducing the data of DoE on Design Expert 7.0 software. The model was managed to predict the best purification, even if the relative Recovery was not so good. For that, the importance of these responses was established as 2 for % relative Recovery and 5 for % relative Purity.

The best conditions leading to the best predicted responses, where the pre-miR peak was more pure and more concentrated, are described on Table 17.

Table 17: Optimum conditions settled by DoE for pre-miR purification.

Factor	**Value**
Equilibrium ([$(NH_4)_2SO_4$] in 10 mM Tris-HCl, pH 7.5)	1.7 M
Elution ([$(NH_4)_2SO_4$] in 10 mM Tris-HCl, pH 7.5)	1.11 M
Temperature	16.11 ºC
Responses	**Value**
% relative Recovery	84.1992
% relative Purity	26.6146
Desirability	0.853

To better understand where the optimum conditions stand on design space, Figure 32 presents two contour graphics with the two predicted responses, obtained with the established optimum conditions.

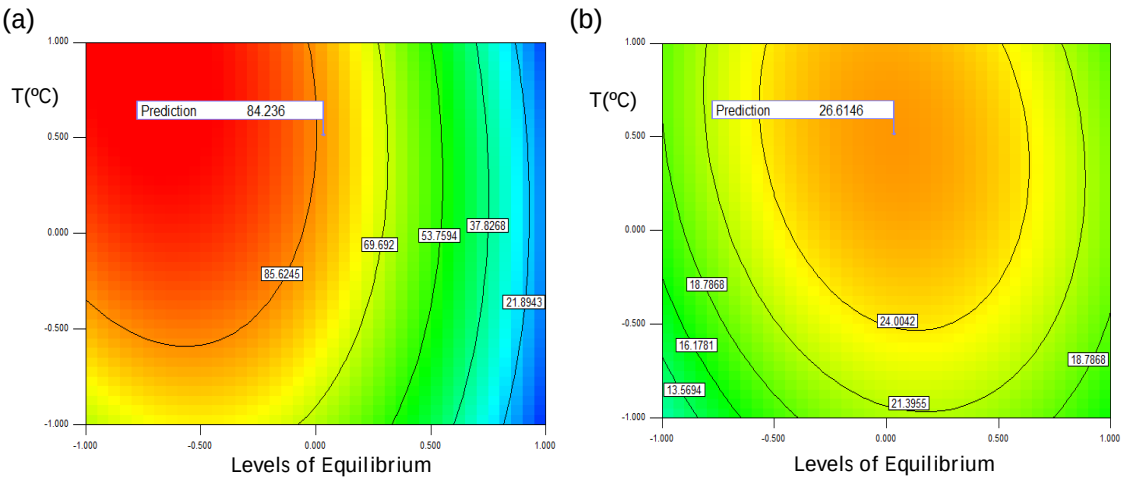


Figure 32: Contour plots for the predicted responses of the defined optimal point (a) Predicted percentage of relative Recovery response. (b) Predicted percentage of relative Purity response.

According to this data, the optimum gradient (1.7M $(NH_4)_2SO_4$ in 10 mM Tris-HCl, pH 7.5; 1.11 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl, pH 7.5; 10 mM Tris-HCl, pH 7.5) was repeated three times (Figure 33) in order to evaluate the reproducibility and suitability of the method.

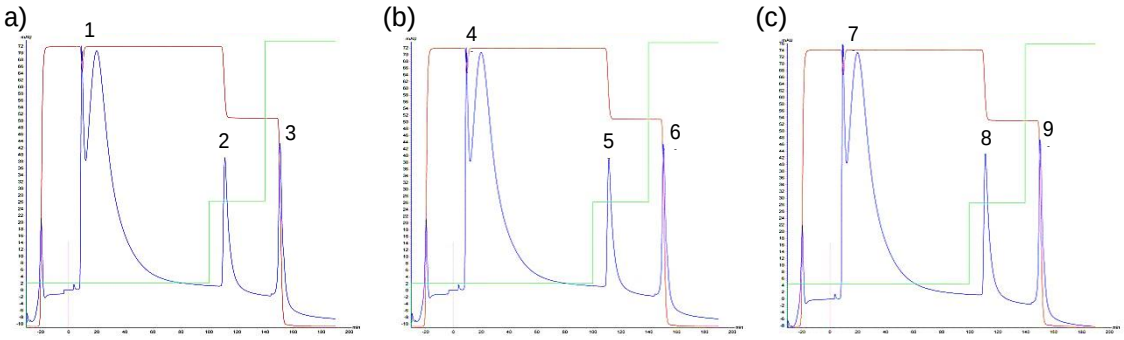


Figure 33: Chromatographic profiles of the three replicates of optimum conditions. Experiments 16 (a), 17 (b) and 18 (c).

All chromatograms revealed the same profile, suggesting that the same interactions occur in all experiments. For further analysis, the peaks fractions were examined with polyacrylamide gel electrophoresis (Figure 34) and the two different responses were determined. The % relative Recovery and the % relative Purity were established for the peaks 2, 5 and 8 (Table 18), because it was expected that the pre-miR29 elution occurred with the 1.11 M $(NH_4)_2SO_4$ in 10 mM Tris-HCl condition.

Table 18: Responses obtained from the optimization/validation of the model.

% relative Recovery	% relative Purity
83.3828	26.188
80.1292	23.843
86.4059	23.538
Mean	
83.3060	**24.523**

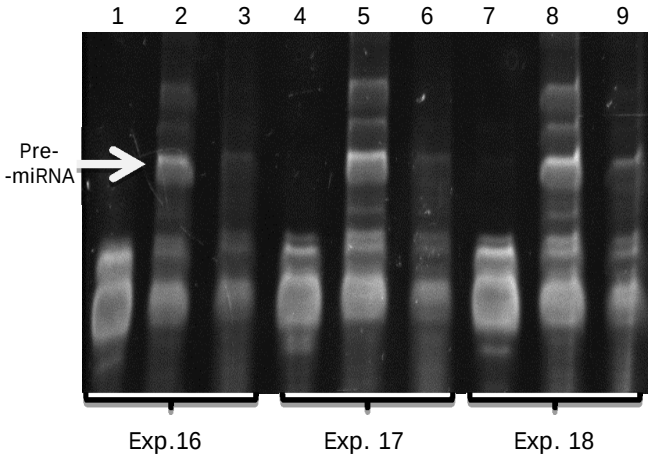


Figure 34: Polyacrylamide gel electrophoresis of the different fractions collected on 16, 17 and 18 experiments.

These results are in accordance with the chromatograms. The variation between the responses was acceptable and, despite the predicted values were not accomplished, their stayed very close and within the confidence interval (CI). With Design Expert 7.0 data, the 95% confidence interval for the two responses was given, as depicted in Table 19.

Table 19: Confidence intervals for the two responses.

Response	95% CI low	95% CI high
% Relative Recovery	68.00	92.89
% Relative Purity	20.55	30.53

The 95% CIs represent the range in which the responses should lie 95% of the time and the previous results are within the expected [86]. Thus, these data confirm that the model fitted well the data, being useful to predict the effect on responses when any factor is changed, on design space.

Chapter 4

Conclusions

Interference RNA therapeutics had an enormous development on the past few years. Nowadays, these little molecules are considered as a potential therapeutic product for many diseases as they act at several biomolecular levels, silencing genes of many ways. Because it is described that Alzheimer's disease is related to a specific cluster of miRNAs that have an important role on the disease evolution, it was proposed to take advantage of this relation and to use miRNA to the disease treatment. Therefore, in order to use pre-miR29b-1 to increase the levels on AD brain, first it is necessary to produce it, as a pure and stable molecule.

In this study the pre-miR was obtained from a recombinant host and two different methodologies were used to achieve its purification. Both "one at a time" method and design of experiments methodology were used to develop a suitable chromatographic technique enabling the recovery of pure pre-miR. Therefore, the strategy aims to exploit the natural interactions that occur between sRNAs and the O-phospho-L-tyrosine matrix, on an amino-acid affinity chromatography process.

In the one at a time method, 59 experiments were performed to achieve the best purification result provided by the matrix. This strategy required a large number of experiments, was time consuming, did not allowed to explore more than one parameter at the same time, and the interaction between the most important factors was not evaluated. Thus, this methodology revealed to be more expensive. However, in the end it was attained a relative pre-miR recovery and purity of about 71% and 52%, respectively.

As a second methodology, the Box Behnken experimental design was used in order to compare it with the method referenced above. The levels and factors were chosen based on the previous knowledge achieved with the first experiments done on the "one at a time" procedure. In this work it was intended to investigate the design space so that the pre-miR would elute on the second chromatogram peak. The experiments were done randomly, avoiding any bias.

Using informatics software, two responses were collected - % of relative pre-miR Recovery and % of relative pre-miR Purity – and introduced on Box-Behnken design to define a model that would predict the best purification conditions. The design has tested three factors (equilibrium and elution conditions and temperature) at three levels, with two replicates of

the central point. In this case, 15 experiments were developed and it was generated a good fitted model. The statistical data expressed that the model were significant and allowed to predict the optimum response. With this methodology it was obtained a result of 83% of relative pre-miR Recovery and 25% of relative Purity. Although this result was close to the predicted (84% and 27% respectively), the relative Purity obtained was significantly lower than the reached with the "one at a time method".

In this work, it was verified that the optimum point compromised the purification degree, as the quantity of miRNA recovered was increased. Hence, and considering the therapeutic application of the target biomolecule, one at a time method could be useful to manage the purification in detriment of some recovery.

The comparison of both approaches for the optimization of chromatographic processes revealed synergies that could lead to new concepts of optimization. As future remarks the design of experiments could be used as a faster approach to initially obtain the optimum conditions, revealing the importance of factors and the complexity of the problem. Additionally, it could provide information for first predictions concerning optimal factor settings and establishing new paths for the final purification optimization. On the other hand, the "one at a time" method should be used when it is known specifically what changes will lead to a greater result. As chromatographic processes are influenced by many variables, it is hard to optimize and guarantee the reproducibility of all responses. So, the DoE approach allows to know how the most important and controllable variables affects the process, and in which magnitude. The "one at a time" methodology can always be a complement to the final purification optimization, when it is clear which factors need to be changed. Therefore, based on this work, it is possible to conclude that all chromatographic procedures will benefit using the DoE approach even if more than one is needed or if the "one at a time" methodology is used as a screening or final optimization, perhaps out of the design space.

Chapter 5

Future Trends

In order to continue this work, more optimization of the chromatographic process has to be done. With the resulting data from Box-Behnken design and from the "one at a time" methodology, it would be interesting to conjugate all the information and set new experiments to be done. For example, the DoE showed that an higher temperature results on better responses, hence, the optimum result of "one at a time" could be tested at that temperature to see if a better purification will be achieved. Another interesting path would be testing other types of designs and different factors or levels, this will display new information that could lead to a better optimization process.

As pre-miR is intended to be used as a therapeutic molecule, after its purification it will be necessary to confirm the purity, stability and integrity. Thus, to achieve pre-miR purity and integrity the polyacrylamide electrophoresis can be used and the integrity could be confirmed by circular dichroism. In addition, real-time PCR could be performed to guarantee the presence of the target molecule on the collected samples.

After the validation of the purification process it is intended to assess the effect of the purified molecule in the silencing of BACE1 mRNA, and compare the results with those obtained with the pre-miR-29 produced by chemical synthesis, in order to confirm the pre-miR action and its therapeutic potential.

Chapter 6

Bibliography

[1] Chu, C. Y., and Rana, T. M. (2007) Small RNAs: regulators and guardians of the genome, *Journal of cellular physiology 213*, 412-419.

[2] Filipowicz, W., Jaskiewicz, L., Kolb, F. A., and Pillai, R. S. (2005) Post-transcriptional gene silencing by siRNAs and miRNAs, *Current opinion in structural biology 15*, 331-341.

[3] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, *Nature 391*, 806-811.

[4] Ruegger, S., and Grosshans, H. (2012) MicroRNA turnover: when, how, and why, *Trends in biochemical sciences 37*, 436-446.

[5] Schonrock, N., Matamales, M., Ittner, L. M., and Gotz, J. (2012) MicroRNA networks surrounding APP and amyloid-beta metabolism-implications for Alzheimer's disease, *Experimental neurology 235*, 447-454.

[6] Bettens, K., Sleegers, K., and Van Broeckhoven, C. (2013) Genetic insights in Alzheimer's disease, *Lancet neurology 12*, 92-104.

[7] Mullane, K., and Williams, M. (2013) Alzheimer's therapeutics: continued clinical failures question the validity of the amyloid hypothesis-but what lies beyond?, *Biochemical pharmacology 85*, 289-305.

[8] Wimo, A., Jonsson, L., Bond, J., Prince, M., and Winblad, B. (2013) The worldwide economic impact of dementia 2010, *Alzheimer's & dementia : the journal of the Alzheimer's Association 9*, 1-11 e13.

[9] Tan, L., Yu, J. T., and Hu, N. (2013) Non-coding RNAs in Alzheimer's disease, *Molecular neurobiology 47*, 382-393.

[10] Meek, A. R., Simms, G. A., and Weaver, D. F. (2012) Searching for an endogenous anti-Alzheimer molecule: identifying small molecules in the brain that slow Alzheimer disease progression by inhibition of beta-amyloid aggregation, *Journal of psychiatry & neuroscience : JPN 38*, 120166.

[11] Hebert, S. S., Horre, K., Nicolai, L., Papadopoulou, A. S., Mandemakers, W., Silahtaroglu, A. N., Kauppinen, S., Delacourte, A., and De Strooper, B. (2008) Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with

increased BACE1/beta-secretase expression, *Proceedings of the National Academy of Sciences of the United States of America 105*, 6415-6420.

[12] Martins, R., Queiroz, J. A., and Sousa, F. (2012) Histidine affinity chromatography-based methodology for the simultaneous isolation of Escherichia coli small and ribosomal RNA, *Biomedical chromatography : BMC 26*, 781-788.

[13] Martins, R., Maia, C. J., Queiroz, J. A., and Sousa, F. (2012) A new strategy for RNA isolation from eukaryotic cells using arginine affinity chromatography, *Journal of separation science 35*, 3217-3226.

[14] Martins, R., Queiroz, J. A., and Sousa, F. (2010) A new affinity approach to isolate Escherichia coli 6S RNA with histidine-chromatography, *Journal of molecular recognition : JMR 23*, 519-524.

[15] Burnett, J. C., and Rossi, J. J. (2012) RNA-based therapeutics: current progress and future prospects, *Chemistry & biology 19*, 60-71.

[16] Dykxhoorn, D. M., and Lieberman, J. (2006) Running interference: prospects and obstacles to using small interfering RNAs as small molecule drugs, *Annual review of biomedical engineering 8*, 377-402.

[17] Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals, *Cell 101*, 25-33.

[18] Ghildiyal, M., and Zamore, P. D. (2009) Small silencing RNAs: an expanding universe, *Nature reviews. Genetics 10*, 94-108.

[19] Junn, E., and Mouradian, M. M. (2012) MicroRNAs in neurodegenerative diseases and their therapeutic potential, *Pharmacology & therapeutics 133*, 142-150.

[20] Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14, *Cell 75*, 843-854.

[21] Wightman, B., Ha, I., and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans, *Cell 75*, 855-862.

[22] Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell 116*, 281-297.

[23] Soifer, H. S., Rossi, J. J., and Saetrom, P. (2007) MicroRNAs in disease and potential therapeutic applications, *Molecular therapy : the journal of the American Society of Gene Therapy 15*, 2070-2079.

[24] Dunkel, P., Chai, C. L., Sperlagh, B., Huleatt, P. B., and Matyus, P. (2012) Clinical utility of neuroprotective agents in neurodegenerative diseases: current status of drug development for Alzheimer's, Parkinson's and Huntington's diseases, and amyotrophic lateral sclerosis, *Expert opinion on investigational drugs 21*, 1267-1308.

[25] Provost, P. (2010) MicroRNAs as a molecular basis for mental retardation, Alzheimer's and prion diseases, *Brain research 1338*, 58-66.

[26] Piaceri, I., Nacmias, B., and Sorbi, S. (2013) Genetics of familial and sporadic Alzheimer's disease, *Front Biosci (Elite Ed) 5*, 167-177.

[27] Provost, P. (2010) Interpretation and applicability of microRNA data to the context of Alzheimer's and age-related diseases, *Aging 2*, 166-169.

[28] Ghiso, J., and Frangione, B. (2002) Amyloidosis and Alzheimer's disease, *Advanced drug delivery reviews 54*, 1539-1551.

[29] Querfurth, H. W., and LaFerla, F. M. (2010) Alzheimer's disease, *The New England journal of medicine 362*, 329-344.

[30] Vassar, R., Kovacs, D. M., Yan, R., and Wong, P. C. (2009) The beta-secretase enzyme BACE in health and Alzheimer's disease: regulation, cell biology, function, and therapeutic potential, *The Journal of neuroscience : the official journal of the Society for Neuroscience 29*, 12787-12794.

[31] Gotz, J., Matamales, M., Gotz, N. N., Ittner, L. M., and Eckert, A. (2012) Alzheimer's disease models and functional genomics-How many needles are there in the haystack?, *Frontiers in physiology 3*, 320.

[32] Satoh, J. (2012) Molecular network of microRNA targets in Alzheimer's disease brains, *Experimental neurology 235*, 436-446.

[33] Lukiw, W. J. (2007) Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus, *Neuroreport 18*, 297-300.

[34] Lukiw, W. J., Zhao, Y., and Cui, J. G. (2008) An NF-kappaB-sensitive micro RNA-146a-mediated inflammatory circuit in Alzheimer disease and in stressed human brain cells, *The Journal of biological chemistry 283*, 31315-31322.

[35] Wang, W. X., Rajeev, B. W., Stromberg, A. J., Ren, N., Tang, G., Huang, Q., Rigoutsos, I., and Nelson, P. T. (2008) The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of beta-site amyloid precursor protein-cleaving enzyme 1, *The Journal of neuroscience : the official journal of the Society for Neuroscience 28*, 1213-1223.

[36] Hebert, S. S., Horre, K., Nicolai, L., Bergmans, B., Papadopoulou, A. S., Delacourte, A., and De Strooper, B. (2009) MicroRNA regulation of Alzheimer's Amyloid precursor protein expression, *Neurobiology of disease 33*, 422-428.

[37] Shioya, M., Obayashi, S., Tabunoki, H., Arima, K., Saito, Y., Ishida, T., and Satoh, J. (2010) Aberrant microRNA expression in the brains of neurodegenerative diseases: miR-29a decreased in Alzheimer disease brains targets neurone navigator 3, *Neuropathology and applied neurobiology 36*, 320-330.

[38] Hebert, S. S., Papadopoulou, A. S., Smith, P., Galas, M. C., Planel, E., Silahtaroglu, A. N., Sergeant, N., Buee, L., and De Strooper, B. (2010) Genetic ablation of Dicer in adult forebrain neurons results in abnormal tau hyperphosphorylation and neurodegeneration, *Human molecular genetics 19*, 3959-3969.

[39] Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M. P., Nalls, M. A., Cookson, M. R., St-Laurent, G., 3rd, and Wahlestedt, C. (2010) Evidence for natural

antisense transcript-mediated inhibition of microRNA function, *Genome biology 11*, R56.

[40] Smith, P., Al Hashimi, A., Girard, J., Delay, C., and Hebert, S. S. (2011) In vivo regulation of amyloid precursor protein neuronal splicing by microRNAs, *Journal of neurochemistry 116*, 240-247.

[41] Zeng, Y. (2006) Principles of micro-RNA production and maturation, *Oncogene 25*, 6156-6162.

[42] Zeng, Y., Yi, R., and Cullen, B. R. (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha, *The EMBO journal 24*, 138-148.

[43] Kriegel, A. J., Liu, Y., Fang, Y., Ding, X., and Liang, M. (2012) The miR-29 family: genomics, cell biology, and relevance to renal and cardiovascular injury, *Physiological genomics 44*, 237-244.

[44] Liston, A., Papadopoulou, A. S., Danso-Abeam, D., and Dooley, J. (2012) MicroRNA-29 in the adaptive immune system: setting the threshold, *Cellular and molecular life sciences : CMLS 69*, 3533-3541.

[45] Bernardo, B. C., Charchar, F. J., Lin, R. C., and McMullen, J. R. (2012) A microRNA guide for clinicians and basic scientists: background and experimental techniques, *Heart, lung & circulation 21*, 131-142.

[46] Tan, S. C., and Yiap, B. C. (2009) DNA, RNA, and protein extraction: the past and the present, *Journal of biomedicine & biotechnology 2009*, 574398.

[47] Mraz, M., Malinova, K., Mayer, J., and Pospisilova, S. (2009) MicroRNA isolation and stability in stored RNA samples, *Biochemical and biophysical research communications 390*, 1-4.

[48] http://tools.invitrogen.com/content/sfs/manuals/fm_1560.pdf (8/5/2013).

[49] http://www.qiagen.com/resources/Download.aspx?id={9DFE7EBC-45A2-4B1E-9EA3-D7C5AC264E00}&lang=en&ver=1 (8/5/2013).

[50] http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/mirna/microrna.html (8/5/2013).

[51] http://www.chem-agilent.com/pdf/strata/400814.pdf (8/5/2013)

[52] http://tools.invitrogen.com/content/sfs/manuals/purelink_microRNA_man.pdf (8/5/2013).

[53] https://cssportal.roche.com/LFR_PublicDocs/ras/05080576001_en_05.pdf (8/5/2013)

[54] Easton, L. E., Shibata, Y., and Lukavsky, P. J. (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography, *RNA 16*, 647-653.

[55] Allerson, C. R., Martinez, A., Yikilmaz, E., and Rouault, T. A. (2003) A high-capacity RNA affinity column for the purification of human IRP1 and IRP2 overexpressed in Pichia pastoris, *RNA 9*, 364-374.

[56] Srisawat, C., and Engelke, D. R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins, *RNA 7*, 632-641.

[57] Nonne, N., Ameyar-Zazoua, M., Souidi, M., and Harel-Bellan, A. (2010) Tandem affinity purification of miRNA target mRNAs (TAP-Tar), *Nucleic acids research 38*, e20.

[58] Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., and Kim, V. N. (2004) MicroRNA genes are transcribed by RNA polymerase II, *The EMBO journal 23*, 4051-4060.

[59] Batey, R. T., and Kieft, J. S. (2007) Improved native affinity purification of RNA, *RNA 13*, 1384-1389.

[60] Kieft, J. S., and Batey, R. T. (2004) A general method for rapid and nondenaturing purification of RNAs, *RNA 10*, 988-995.

[61] Sousa, A., Sousa, F., and Queiroz, J. A. (2012) Advances in chromatographic supports for pharmaceutical-grade plasmid DNA purification, *Journal of separation science 35*, 3046-3058.

[62] Sousa, F., Cruz, C., and Queiroz, J. A. (2010) Amino acids-nucleotides biomolecular recognition: from biological occurrence to affinity chromatography, *Journal of molecular recognition : JMR 23*, 505-518.

[63] Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., and Patel, D. J. (2005) Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein, *Nature 434*, 666-670.

[64] Filipowicz, W. (2005) RNAi: the nuts and bolts of the RISC machine, *Cell 122*, 17-20.

[65] Frank, F., Sonenberg, N., and Nagar, B. (2010) Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2, *Nature 465*, 818-822.

[66] Rafamantanana, M. H., Debrus, B., Raoelison, G. E., Rozet, E., Lebrun, P., Uverg-Ratsimamanga, S., Hubert, P., and Quetin-Leclercq, J. (2012) Application of design of experiments and design space methodology for the HPLC-UV separation optimization of aporphine alkaloids from leaves of Spirospermum penduliflorum Thouars, *Journal of pharmaceutical and biomedical analysis 62*, 23-32.

[67] Hibbert, D. B. (2012) Experimental design in chromatography: a tutorial review, *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences 910*, 2-13.

[68] Bertol, G., Franco, L., and Oliveira, B. H. (2012) HPLC analysis of oxindole alkaloids in Uncaria tomentosa: sample preparation and analysis optimisation by factorial design, *Phytochemical analysis : PCA 23*, 143-151.

[69] Ferreira, S. L., Bruns, R. E., da Silva, E. G., Dos Santos, W. N., Quintella, C. M., David, J. M., de Andrade, J. B., Breitkreitz, M. C., Jardim, I. C., and Neto, B. B. (2007) Statistical designs and response surface techniques for the optimization of chromatographic systems, *Journal of chromatography. A 1158*, 2-14.

[70] L. Eriksson, E. J., N. Kettaneh-Wold, C.Wikstrom, S. Wold (2008) Design of Experiments: Principles and Applications, 3 ed., *MKS Umetrics AB.*

[71] Ferreira, S. L., Bruns, R. E., Ferreira, H. S., Matos, G. D., David, J. M., Brandao, G. C., da Silva, E. G., Portugal, L. A., dos Reis, P. S., Souza, A. S., and dos Santos, W. N. (2007) Box-Behnken design: an alternative for the optimization of analytical methods, *Analytica chimica acta 597*, 179-186.

[72] Suzuki, H., Umekage, S., Tanaka, T., and Kikuchi, Y. (2009) Extracellular tRNAs of the marine photosynthetic bacterium Rhodovulum sulfidophilum are not aminoacylated, *Bioscience, biotechnology, and biochemistry 73*, 425-427.

[73] Chomczynski, P., and Sacchi, N. (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on, *Nature protocols 1*, 581-585.

[74] Schonrock, N., and Gotz, J. (2012) Decoding the non-coding RNAs in Alzheimer's disease, *Cellular and molecular life sciences : CMLS 69*, 3543-3559.

[75] Sibley, C. R., Seow, Y., and Wood, M. J. (2010) Novel RNA-based strategies for therapeutic gene silencing, *Molecular therapy : the journal of the American Society of Gene Therapy 18*, 466-476.

[76] Suzuki, H., Ando, T., Umekage, S., Tanaka, T., and Kikuchi, Y. (2010) Extracellular production of an RNA aptamer by ribonuclease-free marine bacteria harboring engineered plasmids: a proposal for industrial RNA drug production, *Applied and environmental microbiology 76*, 786-793.

[77] Suzuki, H., Umekage, S., Tanaka, T., and Kikuchi, Y. (2011) Artificial RNA aptamer production by the marine bacterium Rhodovulum sulfidophilum: improvement of the aptamer yield using a mutated transcriptional promoter, *Journal of bioscience and bioengineering 112*, 458-461.

[78] Sousa, F., Prazeres, D. M., and Queiroz, J. A. (2008) Affinity chromatography approaches to overcome the challenges of purifying plasmid DNA, *Trends in biotechnology 26*, 518-525.

[79] Sousa, A., Sousa, F., and Queiroz, J. A. (2010) Differential interactions of plasmid DNA, RNA and genomic DNA with amino acid-based affinity matrices, *Journal of separation science 33*, 2610-2618.

[80] Dejaegher, B., and Heyden, Y. V. (2011) Experimental designs and their recent advances in set-up, data interpretation, and analytical applications, *Journal of pharmaceutical and biomedical analysis 56*, 141-158.

[81] Rio, D. C., Ares, M., Jr., Hannon, G. J., and Nilsen, T. W. (2010) Polyacrylamide gel electrophoresis of RNA, *Cold Spring Harbor protocols 2010*, pdb prot5444.

[82] Montegomery, D. C. (2001) Design and Analysis of Experiments, 5 ed., J*ohn Wiley & Sons, Inc.*

[83] Mohajeri, L., Aziz, H. A., Isa, M. H., and Zahed, M. A. (2010) A statistical experiment design approach for optimizing biodegradation of weathered crude oil in coastal sediments, *Bioresource technology 101*, 893-900.

[84] Bingol, D., Tekin, N., and Alkan, M. (2010) Brilliant Yellow dye adsorption onto sepiolite using a full factorial design, *Applied Clay Science 50*, 315-321.

[85] Peternel, I., Kusic, H., Marin, V., and Koprivanac, N. (2013) UV-assisted persulfate oxidation: the influence of cation type in the persulfate salt on the degradation kinetics of an azo dye pollutant, *Reac Kinet Mech Cat 108*, 17-39.

[86] Robert L. Mason, R. F. G., James L. Hess (2003) Statistical Design and Analysis of Experiments, 2 ed., *John Wiley & Sons, Inc.*

[87] Patil, S. A., Surwase, S. N., Jadhav, S. B., and Jadhav, J. P. (2013) Optimization of medium using response surface methodology for l-DOPA production by Pseudomonas sp. SSA, *Biochemical Engineering Journal 74*, 36-45.

[88] Garai, D., and Kumar, V. (2013) A Box–Behnken design approach for the production of xylanase by Aspergillus candidus under solid state fermentation and its application in saccharification of agro residues and Parthenium hysterophorus L, *Industrial Crops and Products 44*, 352-363.