Design and implementation of molecular circuits for mitigating genetic diseases

Vom Promotionsausschuss der

Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Ana Karen Velázquez Sánchez

aus

Ensenada, Mexiko

2022

1. Gutachter: Prof. Dr. Dr. habil. Karl-Heinz ZIMMERMANN
Institut für Eingebettete Systeme, Technische Universität Hamburg (TUHH)

2. Gutachter: Prof. Dr. Zoya IGNATOVA
Institut für Biochemie und Molekularbiologie, Universität Hamburg (UHH)

Tag der mündlichen Prüfung: 15.12.2021

Vorsitzender des Prüfungsausschusses:
Prof. Dr. An-Ping ZENG
Institut für Bioprozess- und Biosystemtechnik, Technische Universität Hamburg (TUHH)

DOI: https://doi.org/10.15480/882.4094
ORCID: orcid.org/0000-0001-9374-4006

Abstract

A systematic view of disease, where networks of interacting entities characterize physiological states related to disease, allows for innovation of novel therapeutic approaches. Genetic circuits emulate digital circuits behavior, where cellular signals act as input signals, and their processing controls cellular activities, such as production of a protein as an output signal. Genetic disorders are given by alterations or mutations in chromosomes and/or genes or by defects in gene expression. Around 11% of all inherited genetic diseases are caused by nonsense mutations, and to date, no cure has yet been developed. Nonsense mutations are single base substitutions within the gene sequence, consequently introducing in-frame premature termination or stop codons (PTCs).

This study focuses on *de novo* gene regulation by means of artificial small RNAs (sRNAs), and the design of molecular circuits for diagnosis and mitigation of genetic diseases, particularly, nonsense mutation-associated diseases. For this, the combined use of RNA-based components (i.e. sRNAs, miRNAs, riboswitches, tRNAs) for development of molecular circuitry to tackle the devastating genetic disease, cystic fibrosis (CF) is proposed. For this, artificial sRNAs are designed and built to dynamically regulate the expression of transcription factors, and in turn expression levels of a model easy-to-detect output protein, thus allowing for tuning previously built genetic constructs. Furthermore, synthetic sRNAs are suggested to negatively regulate a nonsense suppressor tRNA.

MicroRNAs (miRNAs) are key components of the cell, in both normal and pathological states, and serve as potential indicators of disease conditions as well as in therapeutic approaches. A thorough investigation on the most commonly reported miRNAs that are dysregulated in CF is performed. Here, a prototype of a semi-autonomous synthetic RNA-based circuit for non-invasive diagnosis and mitigation of PTC-triggered diseases is proposed and its applicability particularly illustrated for CF. A proof-of-concept for its implementation in *E. coli* is provided. Two diagnostic models of CF disease are presented: one for CF-specific diagnosis and another for differential diagnosis from chronic obstructive pulmonary disease. These are formulated by logically describing the miRNA biomarker profile as variables with a Boolean expression. Furthermore, a scheme for a theranostic approach for CF is depicted, which represents an example for toehold strand displacement-based circuits for solving the diagnostic rules.

Ribosomopathies are other type of genetic disorders linked to the differential composition of ribosomes or modifications in ribosomal proteins (RPs). The detection of fluctuations of RP composition with possible functional implications remains a challenging task. In this study, we present an approach to analyze the RP composition and our results show an invariant RP composition of ribosomes across different murine tissues (brain and liver) and ages (juvenile, adult, and middle-aged), suggesting that if there is any ribosomal heterogeneity, it may be the result from other processes, rather than variation on RP composition.

Acknowledgments

First of all, I would like to express my deep gratitude and appreciation to my supervisors, Prof. Dr. Dr. Karl-Heinz Zimmermann and Prof. Dr. Zoya Ignatova for all their guidance and support, and for encouraging me to follow my research interests. I want to thank Prof. Zimmermann for accepting me as his student, for his support, and for encouraging me to learn more about molecular biology. I want to thank Prof. Ignatova for welcoming me into her lab and taking me as one of her students, for her scientific guidance and mentoring, for trusting in my abilities, and for encouraging me to do experimental work.

I would also like to thank the Mexican Council of Science and Technology (CONACYT) and the German Academic Exchange Service (DAAD) for awarding me with a scholarship to pursue my doctorate studies.

My thanks and appreciation to all former and present members of the Embedded Systems Institute (TUHH) and from Ignatova's Lab (UHH). Thanks to everyone who took out time to discuss an issue with me or to help me clarify something. Thanks to Elke Nissen for all her help during my first months in Hamburg. Thanks to my officemate Ji Youn Hong for making me feel welcome from day one. Special thanks to Susan Amirbeigiarab for taking the time to introduce me and support me through the basics of experimental work, and for our scientific collaboration. Thanks to Dr. Suki Albers for sharing the plasmids and details of the nonsense suppressor tRNAs, and for the scientific support. Thanks to all the students who supported me during my project, special thanks to Bjarne Klopprogge who helped me greatly with the experimental work during the first stage. Thanks to the amazing technicians Christine Polte and Patrick John for their countless help along these years. Thanks to Katrin Krögger for all her help with administrative matters.

Finally, I am grateful to everyone who has encouraged me, believed in me, and supported me throughout this journey, especially to my mother for always being there for me.

Contents

List of Figures

List of Tables

List of Abbreviations

ΔG	Gibbs free energy
H	Enthalpy
ori	Origin of replication
S	Entropy
A	Adenine
bp	Basepair
C	Cytosine
cDNA	Complementary DNA
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane regulator
DSD	DNA strand displacement
dsDNA	Double stranded DNA
eGFP	Enhanced green fluorescent protein
FACS	Fluorescence-activated cell sorting
G	Guanine
GFP	Green fluorescent protein
LB medium	Luria-Bertani medium
MFE	Minimum free energy structure
miRNA	microRNA
mRNA	Messenger RNA
ncRNA	Non-coding RNA
nt	Nucleotide
ORF	Open reading frame

PTC	Premature termination codon
RNA	Ribonucleic acid
RPs	Ribosomal proteins
rRNA	Ribosomal RNA
sRNA	small RNA
ssDNA	Single stranded DNA
T	Thymine
TF	Transcription factor
tRNA	Transfer RNA
U	Uracil
UTR	Untranslated region
wt	Wildtype

Chapter 1

Introduction

Genetic circuits and molecular devices emulate digital circuits behaviour, where environmental or cellular signals (i.e. specific molecules) act as input signals, and their processing can control certain cellular activities, such as production of a reporter protein as output signal. Potential applications of these mechanisms include *de novo* systems for gene regulation [1, 2], therapeutics [3, 4], cellular computation [5, 6], cell reprogramming [7, 8], and bioremediation [9], to name a few. This study focuses on *de novo* gene regulation and cellular computation for diagnosis and mitigation of genetic diseases, particularly on nonsense mutation-associated diseases.

Programming individual cells that interact with each other and perform desired functions is possible through a collection of sequences encoding RNA or protein (biological parts), known as genetic circuits [10, 11]. Moreover, toehold-mediated strand displacement, is one of the most important paradigms for programming nucleic acid (DNA or RNA) devices; mainly because it relies on the complementarity of nucleotide sequences without the need for additional components (e.g. enzymes) [12]. DNA devices can be built on sequential cascading of strand displacement reactions [12–17], and RNA devices can be programmed by designed base-pairing reactions by means of toehold switches (or riboswitches) [18–20].

Genetic disorders are given by alterations or mutations in: *i)* chromosomes, *ii)* a single gene (monogenic disorder), *iii)* in multiple genes together with environmental conditions (multifactorial inheritance disorder) or *iv)* defects in gene expression [21]. Around one-third of the genetic diseases are caused by nonsense mutations [22], and to date a cure has not yet been developed [23]. Nonsense mutations are single base substitutions within the gene sequence, consequently introducing in-frame premature termination or stop codon (PTC: TTA, TGA or TAG). Therefore, investigation and development of novel therapies targeting these anomalies is of significant importance.

Cystic fibrosis (CF) is a genetic disease originated by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Approximately 10% of CF cases are caused by PTCs within the CFTR gene [24], resulting in truncated dysfunctional CFTR proteins. This is physiologically deleterious since the CFTR protein is responsible for controlling the balance of salt and water in lungs and other tissues.

Ribosomopathies are prejudicial genetic disorders that result in clinical phenotypes that occur when genetic abnormalities cause defects in ribosome biogenesis and function. The differential composition of ribosomes, e.g. changes or modifications in ribosomal proteins (RPs), termed ribosome heterogeneity, is linked to ribosomopathies [25]. The detection of fluctuations of RP composition with possible functional implications remains a challenging task.

The definition of diseases at a molecular level, that is, unfolding associations between mutations on the DNA and clinical conditions facilitates a systematic view of disease [26], where networks of interacting molecules characterize physiological states related to diseases. Recently, miRNAs appear as a promising type of biomarkers for diagnosis and prognosis or as predictors of drug response [27]. miRNAs regulate gene expression by binding to the 3' UTR of mRNAs, and are key components of the cell in both normal and pathological states. They can be found within cells, as well as extracellularly enclosed within vesicles, which can be isolated from biological fluids (e.g. cerebrospinal fluid, blood plasma and serum). Dysregulated miRNA expression profiles have been observed in several genetic disorders (e.g. cystic fibrosis [28]) and disease conditions [29], from cancers [30] to neurodegenerative diseases (e.g. amyotrophic lateral sclerosis [31]), and even in hypertension [32].

The aim of this thesis is to address the feasibility of designing (*in silico*) and implementing (*in vivo*) molecular circuits applied towards semi-autonomous diagnosis and mitigation (therapy) of genetic diseases, particularly cystic fibrosis.

The research questions that were sought to elucidate: *i)* Are synthetic biology/molecular computing based approaches feasible to diagnose and/or mitigate genetic diseases?; *ii)* Can a semi-autonomous diagnose and therapy for CF be achieved simultaneously, by means of RNA-based sensors and novel therapeutics?; *iii)* What are the most suitable biomarkers for designing an RNA-based diagnostic sensor for CF?; *iv)* Are synthetic sRNAs efficient for dynamically regulating genetic molecular/circuits?

The research hypotheses of the study are: *i)* Molecular circuits, coordinated with well-established molecular biomarkers (e.g. miRNAs), and novel therapeutics (e.g. nonsense suppressor tRNAs or miR-mimics), can be applied for the semi-autonomous diagnosis and mitigation (therapy) of cystic fibrosis. *ii)* Rationally designed synthetic sRNAs can modulate genetic circuits dynamically. *iii)* A multilayered (transcription, translation) molecular circuit, tuned by synthetic sRNAs that is capable of sensing a CF biomarker (diagnosis) and activating nonsense suppressor tRNAs (therapy) can be designed, and a proof-of-concept can be implemented in *E. coli*.

The major contributions of this work are the following

1. Modulating synthetic small RNAs (sRNAs) will be shown to be useful in tuning previously built genetic constructs. Synthetic sRNAs will be proposed for dynamically regulating transcription factors, and in turn the expression levels of the gene (green fluorescent protein) under their control.

2. Based on point 1, synthetic sRNAs designed to negatively regulate (i.e. turn off) a previously reported [33] constitutively expressed nonsense suppressor tRNA system in *E. coli* will be presented. This will be further used for the development of point 4.

3. An investigation on the literature reporting dysregulated miRNAs in CF will be outlined. Furthermore, a classification of these biomarkers into potential biomarkers for diagnosis and/or therapy will be depicted. This effort will facilitate the identification of components for efficient

design and development of molecular circuits and devices against CF. This assessment will serve for accomplishing points 4, 5, and 6.

4. The design of the first prototype (to the best of our knowledge) of a semi-autonomous synthetic RNA-based circuit for non-invasive diagnosis and mitigation of PTC-related diseases, particularly CF will be illustrated. A proof-of-concept for its implementation *in vivo* in *E. coli* will be provided. Prokaryotic and eukaryotic diagnostic riboswitches for CF based on miR509 detection will be demonstrated.

5. Two simplified models (diagnostic rules) of CF disease will be depicted. These will be formulated by logically describing (by means of a Boolean expression) the profile of its miRNA biomarkers (variables): *i)* a CF-specific diagnosis and *ii)* a differential diagnosis of CF disease and clear distinction from chronic obstructive pulmonary disease. In theory, these can be implemented by adapting any previously reported modular molecular logic gates.

6. A theranostic approach for CF will be defined, suggesting the use *de novo* toehold strand-displacement based circuits for solving diagnosis rules (based on point 5), and releasing a therapeutic RNA molecule on detection (based on point 3).

7. Results from our collaborative work will show an invariant RP composition of ribosomes across murine tissues (cortex, hippocampus, and cerebellum from brain, and liver) from juvenile, adult, and middle-aged mice groups; suggesting that if there is any ribosomal heterogeneity, it may be the result from other processes, rather than variation on RP composition.

The dissertation is organized as follows.

Chapter 2 integrates the background knowledge from different fields required for the development of this work. Across the sections of this chapter, important concepts are reviewed: from central dogma of gene expression, all the way to thermodynamics of nucleic acids and its *in silico* prediction tools, and finally introducing the state-of-the-art technology for designing and implementing molecular circuitry.

Chapters 3, 4, 5, 6 and 7 contain new research results and contributions.

Chapter 3 proposes the rational design (*in silico*) of modulating (i.e. different repression levels) synthetic sRNAs for tuning genetic constructs, and its methodology for implementation in *E. coli*.

Chapters 4, 5, 6 revolve around efforts to tackle CF disease. Chapter 4 presents a survey on the most known dysregulated miRNAs in CF; furthermore, these are reviewed and classified as potential biomarkers for diagnosis and/or therapy. Chapter 5 summarizes the efforts to design and build a prototype synthetic RNA circuit for the purpose of non-invasive diagnosis and mitigation of CF. Chapter 6 proposes theranostic molecular systems for the mitigation of CF, and discusses the feasibility of its implementation within human cells.

Chapter 7 aims to elucidate whether ribosomes differ in protein composition across tissues and over the lifespan by means of quantitative proteomics. The content is published (PNAS, Volume 116, No. 45, pp.22567–22572, 2019, ref: [34]) within a collaboration of Prof. Dr. Ignatova's group (University of Hamburg, Germany) with Prof. Dr. Hartmut Schlüter's group (University Medical Center Hamburg Eppendorf, Germany).

Lastly, conclusions and overall discussion are presented in Chapter 8.

Chapter 2

Synthetic biology

The primary goal of synthetic biology is novel designs and engineering of living matter. This is a multidisciplinary field that brings together concepts and expertise within several areas, particularly computer science and engineering, molecular biology, and biotechnology. This work explores feasibility of designing and implementing nucleic acid circuits in living cells that exhibit multi-layered dynamic properties, which are suitable for performing Boolean logic, and could be applicable in molecular diagnosis context. This chapter summarizes earlier concepts involved in the development of this work, and provides a concise review of published relevant previous research.

2.1 The genetic information flow: central dogma and gene expression

Living organisms are composed by fundamental structural and functional units −cells−, which fall into two categories: eukaryotic (they contain a nucleus) or prokaryotic cells [35]. Molecular biology is the science that studies the mechanisms that lay behind the processes that occur in the cell, and its overall functioning. In the cell, small monomers (e.g. single amino acids, sugars, and nucleotides) assemble together in a repeated manner forming larger molecules known as polymers, such as nucleic acids and proteins.

Nucleic acids (deoxyribonucleic and ribonucleic acids, or DNA and RNA, respectively) are polynucleotides, or chains of nucleotides covalently bound together [36]. Nucleotides contain a nitrogenous base attached to a sugar (ribose in RNA and deoxyribose in DNA) and at least one phosphate group. Bases are classified as purines: adenine (A) and guanine (G); and pyrimidines: cytosine (C), uracil (U) and thymine (T). Each sugar molecule contains 5 atoms of carbon, conventionally numbered from 1′ to 5′, and from which atoms 5′ and 3′ determine the direction of the DNA or RNA molecule, usually referred to as either the 5′-end (five prime end) or the 3′-end (three prime end), respectively. The backbone of the nucleic acids is given by the bonds of each of its bases with a sugar molecule: the phosphate group (attached to the 5′-end) from one nucleotide, allows for addition of the subsequent nucleotide, binding covalently to the 3′-hydroxyl group from the latter, forming thus a phosphodiester bond (Figure 2.1).

Nucleic acids are crucial for storage, retrieval and processing of information in a summary process called gene expression [37]. DNA is the means by which organisms transfer genetic information to their offspring. In eukaryotic organisms, DNA resides in the nucleus and in prokaryotes it is dispersed in the cytoplasm forming a coalescent membraneless structure. The genetic information is stored within the sequential bases of DNA, which in an oversimplified manner, can be seen as sequences

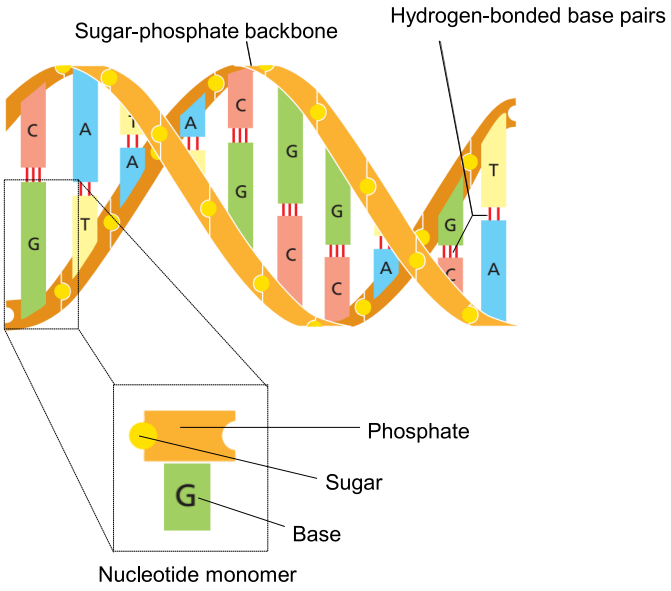


Figure 2.1: DNA structure schematics. The DNA molecule is composed of two strands twisted around each other forming a double helix-a structure. Each single strand of DNA contains a sequence of bonded smaller building blocks called nucleotides. Nucleotides from each complementary strand are bond together by hydrogen bonds, whereas nucleotides on the same strand are held together by covalent bonds (adopted and modified from [35]).

composed of elements from a 4 letter alphabet: C (cytosine), G (guanine), T (thymine) −exchanged to U (uracil) in RNA− , A (adenine); certain permutations of this letter code determine which of the different amino acids (20 natural occurring) is to be synthesized by the translational machinery. DNA is normally double-stranded, unlike RNA. The two strands are antiparallel, that is, oriented oppositely (one goes from 5′-3′ and the other 3′-5′). These strands are fully complementary, this is given by the principle of base-pairing. A base from one strand binds to a base from the other strand by a hydrogen bond in the following manner: AT forms two hydrogen bonds, while the basepair GC forms three hydrogen bonds. All the reactions that take place between nucleic acids during gene expression process are governed by base-pairing interactions between complementary sequences.

Proteins are essential for the stability and functioning of the cells, since they intervene in mostly every process, from growth and development, to cellular uptake regulation and storage of crucial molecules [35]. Proteins consist of at least one long chain of amino acid residues (polypeptide). There are 20 proteinogenic amino acids, and combinations between them make for different proteins. An example of proteins are enzymes, which are responsible for catalyzing biochemicals reactions in the cell. A catalyst is an entity that decreases the activation energy of a chemical reaction, which in turn increases the speed of the reaction.

The collection of all the information that makes for the properties of a cell from a particular organism, is contained in its DNA (or RNA in viruses), and it is referred as the genome. The regions of the genome that allow for the synthesis of structural or regulatory components (proteins or non-coding RNAs) are called genes [36]. The process of synthesizing the information code within the genes into a product is referred to as gene expression, and involves two main processes: transcription and

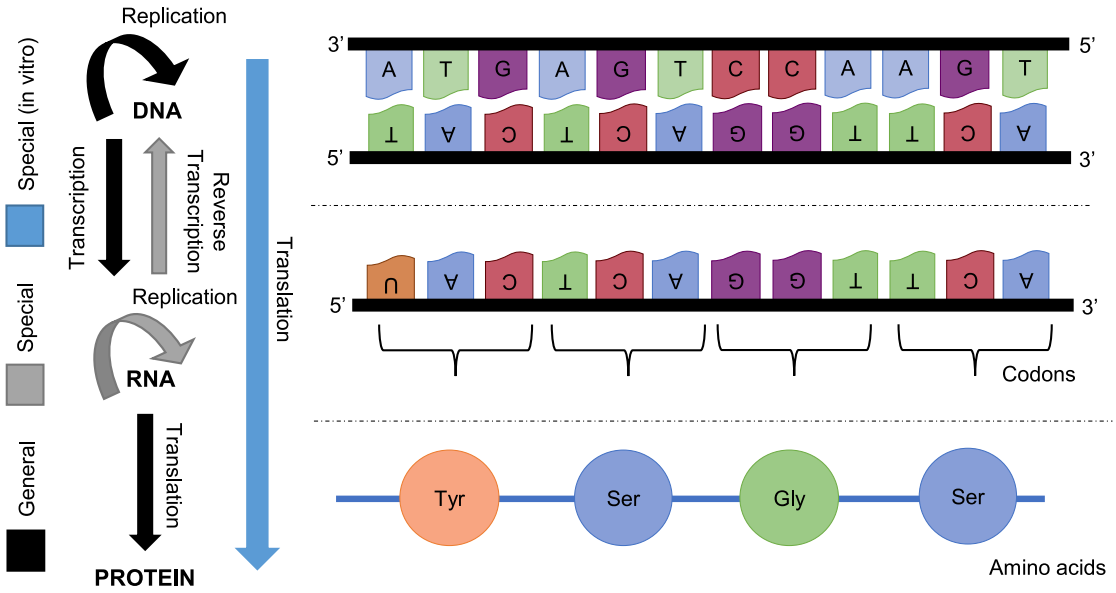


Figure 2.2: Schematic overview of the central dogma of molecular biology. Arrows indicate the flow of information between the polymers.

translation.

The explanation of the sequential flow of genetic information lies in the central dogma of molecular biology (Figure 2.2), proposed by Francis Crick in 1958 [38]. It defines the relations and information transfer between the genetic information-carrying polymers: the DNA code is processed into an intermediate RNA copy (transcription) which in turn is used as template for protein synthesis (translation). DNA to DNA relation is called DNA replication, which involves the synthesis of two identical molecules of DNA from one template. This process occurs in all organisms during cell division (e.g. during growth or damage repair), and provides each new cell with its own identical DNA copy. Transcription is the first step of protein synthesis, and refers to the transfer of the information from DNA into RNA in the form of messenger RNA (mRNA), which is synthesized by the enzyme RNA polymerase. The second step is translation, in which the mRNA is processed and further converted into a protein by ribosomes. Although the general information flow is "DNA to RNA to Proteins", there are some exceptions. For example, certain viruses (like hepatitis B) can perform reverse transcription, synthesizing DNA from RNA by means of the reverse transcriptase enzyme (which acts as a polymerase); this enzyme is particularly used in molecular biology research to perform reverse transcription polymerase chain reaction (RT-PCR). The synthesis of RNA from an RNA template, or RNA replication, is also a special information transfer usually happening in RNA viruses. The direct synthesis of proteins by a DNA template can also be performed by *in vitro* in the test tube; it can be either coupled (transcription and translation at the same time) in a 2-step manner.

The terms upstream and downstream are employed to indicate direction, for example, in which direction transcription develops [37]. In this case, upstream is the direction towards the 5′-end of the molecule in concern, whereas downstream is towards the 3′-end. In case of dsDNA, upstream is used to indicate the direction towards 5′-end of the strand that encodes the gene of interest, and downstream to the 3′-end. Since the DNA molecule is antiparallel, the 3′-end of the template strand is said to be upstream of the gene, and downstream is the 5′-end.

2.1.1 Transcription overview

Transcription (Figure 2.3) is the first of the two main processes that comprise gene expression. It involves the synthesis of RNA by the enzyme RNA polymerase using one strand of DNA (gene coding strand or open-reading frame) as a template (Figure 2.3A). The availability of a gene to be transcribed is determined by regulatory proteins. For transcription to start, the RNA polymerase has to dock in a special region at the beginning of a gene, called the promoter [39]. The first basepair that is transcribed is called the start point, and from there, the RNA polymerase follows the coding strand until it finds a stopping sequence, called terminator.

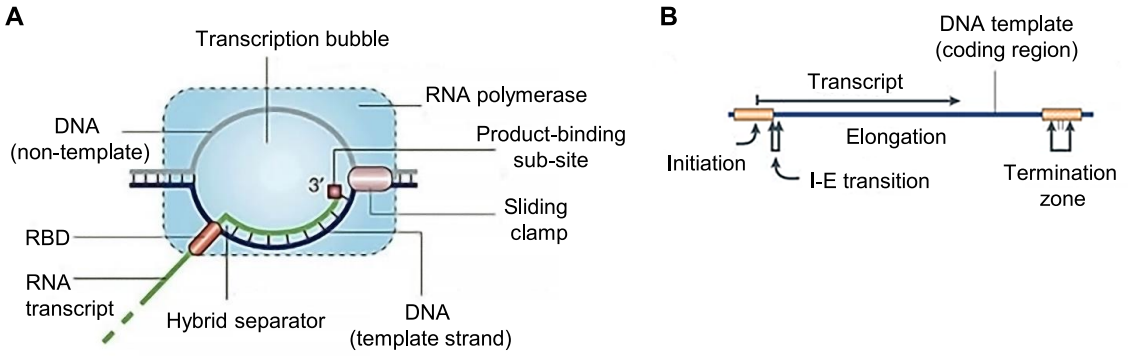


Figure 2.3: Schematic overview of transcription. (A) Transcription complex; **(B)** general scheme for transcription cycle for both prokaryotes and eukaryotes (adopted and modified from [39]).

The region from the promoter to the terminator is known as transcription unit, however, a transcription unit may include one gene or more. The instant product of the RNA polymerase is a primary transcript, which is processed further. This transcript contains the original 5′- and 3′- ends coming from the promoter and terminator, and it is further processed depending on the organism: in prokaryotes, it is rapidly converted into mRNA or into translator competent mature RNAs, such as rRNA and tRNA; in eukaryotes, the ends are modified to produce mRNA or cleaved into mature RNA products [36]. RNA molecules are disposable, that is, they are recycled into their constituent nucleotides, so the cell mRNAs are spliced and can transcribe other sequences as needed, to respond accordingly to dynamic conditions.

The whole process of transcription is composed of four stages (Figure 2.3B):

i) Template recognition: occurs when RNA polymerase docks in the promoter region of the dsDNA, forming a closed complex (transcription bubble); furthermore the DNA strands unwind into an open complex to expose the template strand to basepairing with ribonucleotides.

ii) Initiation: starts with the synthesis of approximately the 9 first nucleotide bonds in RNA, while RNA polymerase is still docked at the promoter, and the phase ends when it leaves it.

iii) Elongation: corresponds to the progression of the transcription bubble along the DNA by disrupting or unwinding its helix structure, the template strand of the briefly unwound region is paired with the nascent RNA chain at the growing point, afterwards the DNA strands pair again to recover the double helix. The transcription bubble is the unwound DNA region that is undergoing transcription.

iv) Termination: takes place when the RNA polymerase reaches the terminator sequence. After the addition of the last base, the enzyme detaches, DNA recovers its helix form and RNA is released.

2.1.2 Translation overview

Translation (Figure 2.4) is the most energy consuming process in the cell [40], and can be divided into four stages: initiation, elongation, termination, and recycling. Overall, these stages are almost identical in all life domains. However, there are some particularities (Figure 2.11) regarding the mechanisms of ribosome docking in mRNA and selection of start site (initiation) between prokaryotes and eukaryotes [41]. These will be further addressed in Sections 2.3 and 2.4, respectively.

Protein synthesis takes place inside the ribosome, a large machine consisting of protein and ribosomal RNA (rRNA). Translation of the mRNA sequence is carried on in triplets of nucleotides or codons, each one encoding a particular amino acid [35]. Since there are 64 naturally occurring codons and only 20 amino acids, some codons code for the same amino acid. Transfer RNAs (tRNAs), which also carry an amino acid attached to their 3′- ends, identify their corresponding codon by base-pairing from their anticodon. This basically summarizes the elongation stage: tRNAs transport the amino acids to the ribosome, which are then linked together into the protein chain.

Termination of the process occurs when a stop codon (UAG, UAA or UGA) is reached on the aminoacyl-tRNA site (A-site) from the ribosome. It is recognized by a class 1 release factor protein (RF), which cleaves the ester bond (hydrolysis) between the tRNA and the growing peptide chain; afterwards, G-protein RF32 promptly removes the RF from the ribosome. In bacteria there are two RF proteins: RF1 (UAA or UAG) , and RF2 (UAA or UGA). In eukaryotes (also in archaea), there is only one RF that can recognize all stop codon sequences, the eRF1 (aRF1 in archaea) [43].

In all domains of life, the translation of the mRNA into proteins takes place on the ribosome ($\sim$ 5 megadaltons in eukaryotes [44] and $\sim$ 2.5 megadaltons in bacteria [45]). The ribosome (Figure 2.5A) consists of two subunits: small and large, which in turn are composed of rRNA and ribosomal proteins (RPs). The mRNA is sandwiched between the small and large subunit. There are three sites (P, peptidyl; A, acceptor; E, exit) in which the tRNA moves through.

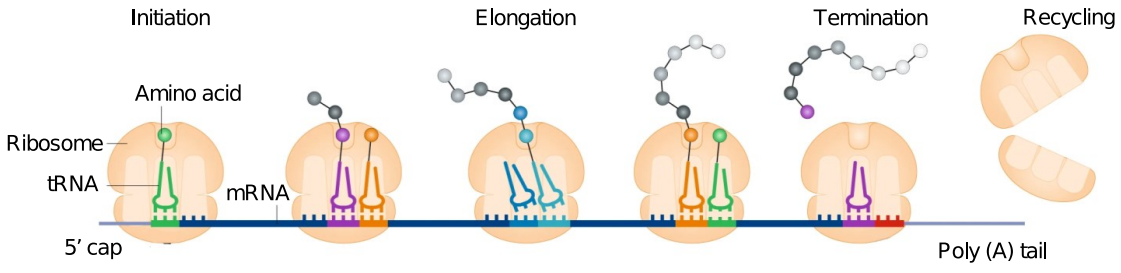


Figure 2.4: Schematic overview of translation. Initiation of translations process begins with the assembly of the ribosomal units, docking in the mRNA and a complex coordination of translation initiation factors (eIFs). The stage of elongation comprises the synthesis of the peptide chain. Termination occurs with the peptide release. Afterwards, for the cycle to be reset, the ribosome subunits are recycled by ATP-binding cassette subfamily E member 1 (ABCE1) (adopted and modified from [42]).

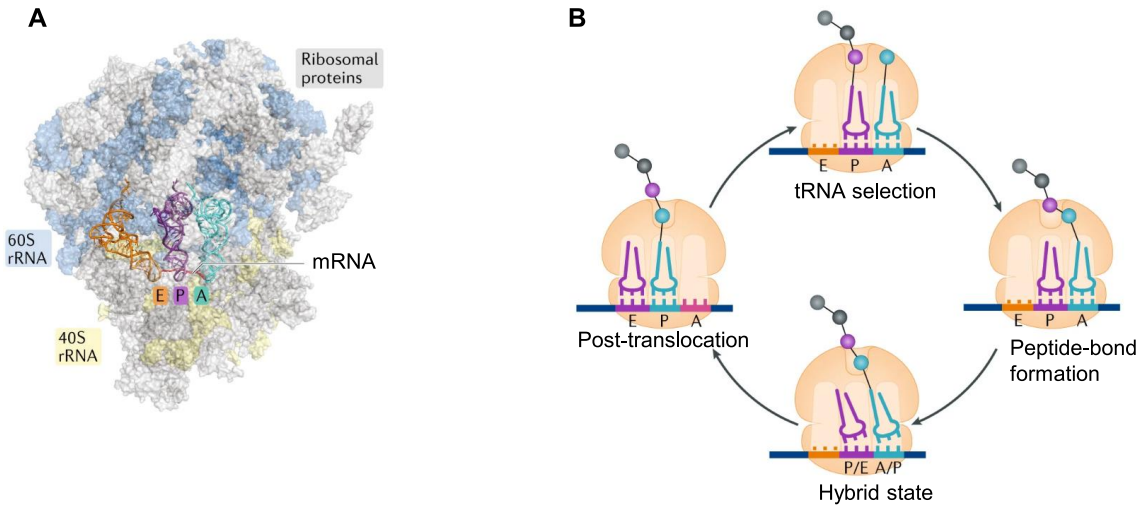


Figure 2.5: The ribosome. **(A)** Model of the ribosome (yeast), highlighting its components (ribosomal proteins and 60S and 40S rRNA), and its 4 bindings sites: the tRNA binding sites E (exit), P (peptidyl) and A (aminoacyl), and the mRNA binding site. **(B)** Schematic overview of the tRNA transit inside the ribosome during elongation phase (adopted and modified from [42]).

For the full assembly of the ribosome to happen, first the small subunit has to dock on an mRNA, then the large subunit comes together with the small, holding the mRNA stable and in proper orientation during protein synthesis. The ribosomes and their subunits are different across kingdoms [46]: in eukaryotes (ribosome 80S), the small subunit (40S) consists of the 18S rRNA and 33 RPs and its large subunit (60S) is composed by 28S rRNA, 5S rRNA, 5.8S and 49 proteins. In bacteria (ribosome 70S), the small subunit (30S) consists of the 16S rRNA and 21 RPs and its large subunit (50S) is composed by 23S rRNA, 5S rRNA, and 31 proteins. In archaea (ribosome 70S), the dimensions and composition of the ribosome are similar to bacteria, but more similar to eukaryotes regarding sequences. The small subunit (30S) and large subunit (50S) are composed by 16S, 23S and 5S, and 50-78 protein, depending on the species [47].

The tRNA and mRNA transit across the ribosome is a complicated process that combines high speed with high accuracy [48]. As seen in Figure 2.5B, first, the tRNA selection takes place on site A which holds the new amino acid to be added to the polypeptide chain. The P-site holds the tRNA with the growing polypeptide attached; when peptide-bond formation occurs, it transfers peptide to the tRNA in A-site. The tRNAs adopt a hybrid state simultaneously with peptide-bond formation, which allows for translocation and decoding of the next codon. The E-site holds the empty tRNA that will exit the ribosome.

2.1.3 Non-coding RNA (ncRNA)

There is a special subset of RNAs that does not encode proteins, which is denominated as non-coding, and plays important roles in regulation of gene expression as well as cell proliferation [49]. Recently, it has been understood that large portions of genomes from eukaryotes (e.g. mammals) correspond to ncRNAs.

There are several classes of already identified ncRNAs, and yet more to be elucidated, from the well studied rRNA and tRNA that play important roles in protein synthesis, to small nuclear

RNAs (snRNAs) with roles in intron splicing and RNA biogenesis, and small nucleolar RNAs (snoRNAs) aiding in chemical modifications of RNA, particularly rRNA [50].

Short/small interfering RNAs (siRNAs) and microRNAs (miRNAs) are other important ncRNA classes in eukaryotes, which are involved in gene regulation and aiding in preventing damage of the genome from exogenous nucleic acids [51]. Likewise, in prokaryotic (i.e. bacteria and archaea) genomes, small RNAs (sRNAs) are present and are involved in the regulation of many pathways, including stress response and metabolism [52].

2.1.3.1 Transfer RNA (tRNA)

The transfer ribonucleic acid's (tRNAs) general cloverleaf structure is highly conserved in prokaryotes and eukaryotes [53]: *i)* a 5′-terminal phosphate group, *ii)* a CCA tail at 3′- end, *iii)* the acceptor stem, *iv)* 3 stem loops (D loop, anticodon loop and T loop), and *v)* a variable loop.

These molecules are synthesized by the RNA polymerase by transcripting tRNA genes. After transcription, tRNAs are heavily modified and further processed by splicing, addition of CCA (in eukaryotes), and posttranscriptional modifications before they undergo maturation. Primary transcripts are trimmed from 5′ and 3′ by nucleases [54]. Also, introns are found within the tRNAs genes in almost all kingdoms of life, which are further spliced by a tRNA-specific endonuclease, and then the processed fragments joined by tRNA ligase. All tRNAs contain the conserved sequence CCA at their 3′-terminal, essential for the recognition of tRNA enzymes and from where the amino acid attaches (to A). This sequence is encoded in the genes in prokaryotes, and added posttranscriptionally by the CCA tRNA nucleotidyltransferase in eukaryotes. Some nucleotides of the tRNA molecules are chemically modified by specific enzymes (e.g. pseudouridine synthases) or guide RNAs. These modifications are essential for the molecule stability, tRNA quality control, decoding, subcellular localization, and immune response.

The tRNA molecules have to be charged, that is, an adenylated amino acid is transferred to the tRNA by an aminoacyl tRNA synthetase (aaRS or ARS) [54]. The adenylation of the amino acid occurs when it reacts with ATP and a pyrophosphate is consequently released. There is a dedicated aaRS for each of the 20 amino acids.

The concentrations of tRNA species in the cells are not only related to codon-bias in the ORFs of protein-encoding genes but also regulated by the copy number of tRNA genes, transcriptional activity, RNA degradation and environment conditions [53].

2.1.3.2 Small RNA (sRNA)

Small non-coding RNAs (sRNAs) are prevalent across all life domains, as they are found in bacterial and eukaryotic organisms, involved in regulation of diverse physiological processes and responses, such as environmental stresses [55].

Bacterial sRNAs, with a length of $\sim$ 40–500 nt, function as regulation mechanisms that allow bacteria to adapt and respond rapidly to harsh conditions. Many of these sRNAs function as antisense regulatory molecules. The location of sRNAs in the genome is not restricted to intergenic regions, there has been found that they can be present in the 3′- and 5′-untranslated regions (UTRs) of bacterial mRNAs [57].

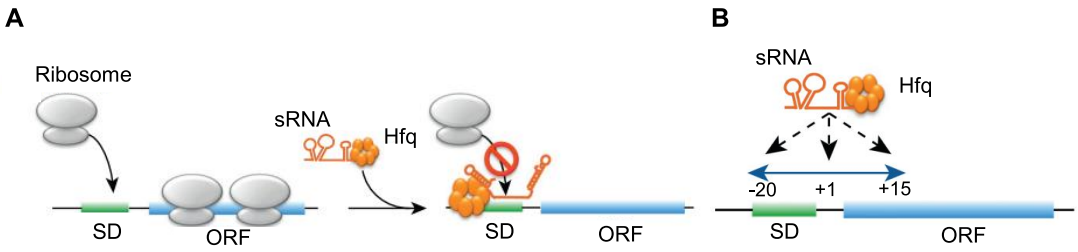


Figure 2.6: Scheme of canonical translation repression by small RNAs. (A) Hfq aids the sRNA in preventing translation initiation; the sRNA binds to the 5′- UTR of its target mRNA, blocking the Shine-Dalgarno (SD) region. **(B)** Schematics of an sRNA binding site on its target mRNA (adopted and modified from [56]).

Hfq-dependent sRNAs are the most studied class of bacterial sRNAs. In E. coli, there are ~ 80 identified Hfq-dependent sRNAs [58], for example, the well characterized RyhB, which aids iron homeostasis by down-regulating the translation of iron related proteins [59]. This class of sRNAs bind to target mRNAs to act as posttranscriptional regulators, in a process mediated by the RNA chaperone Hfq. Both sRNA and mRNA bind to different regions of the Hfq (Figure 2.6A).

Generally, repression of a target mRNA by a sRNA results in accelerated degradation of the target mRNA. It can also result in changes in translational efficiency and destabilization of target mRNAs dependent on the RNase E-including degradosome complex [60]. In *E. coli*, degradation is dependent on RNase E, a single strand-specific endoribonuclease [56].

As seen in Figure 2.6B, the base-pairing between the sRNA and its target mRNA occurs between nucleotides -20 and +15 (relative to the first codon), which also corresponds to the boundaries of the 30S subunit-binding region (blue arrow) [56].

Hfq is a bacterial Sm-like protein that acts as a general mediator across RNA interactions [61]. The roles of Hfq in sRNA regulation are diverse, one of them is to protect sRNAs form degradation by RNases. The sRNA-Hfq complex competes with the ribosome for the RBS region of the target mRNA [62]. Hfq also interacts with RNase E, hence another role is to recruit degradation machinery to the target mRNA [63]. It has also been found that Hfq could also be the direct responsible for translational repression of the target mRNA, whereas the sRNA is only a factor for its recruiting [64], representing the non-canonical path. The roles of Hfq in the cell are many others apart from aiding sRNA regulation. For example, it has been found to be important for the biogenesis of the prokaryotic ribosome, due to its necessary intervention for the correct processing and folding of the 16S rRNA [65]. Furthermore, it is implicated in polyadenylation-dependent mRNA decay, Rho-dependent transcription termination, and transposition. Curiously enough, it is also implicated in the autoregulation of its own gene expression [66].

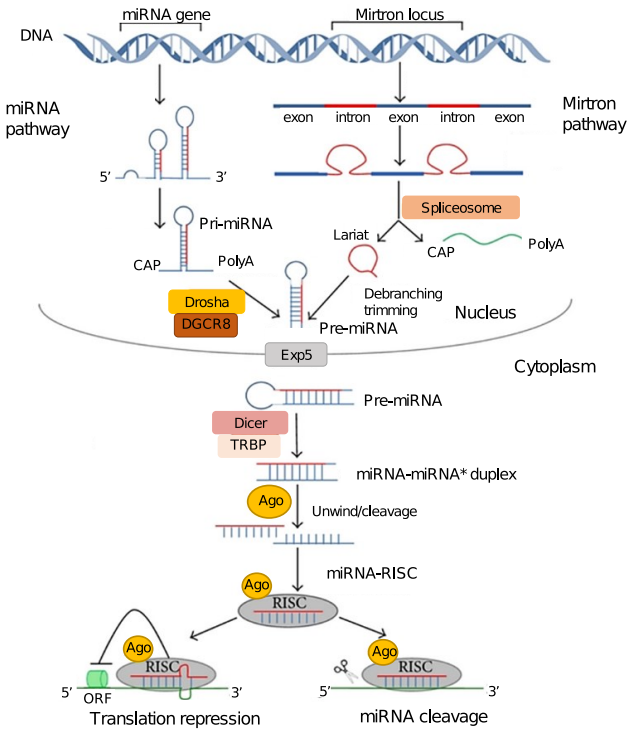


Figure 2.7: Biogenesis pathways of miRNAs. In the canonical biogenesis pathway, miRNA are transcribed from miRNA genes (pri-miRNA) in the nucleus, then processed into pre-miRNAs by Drosha/DGCR8 complex; transcripts from mirtron (miRNAs encoded within introns) follow a noncanonical pathway, they are transcribed by RNA polymerase II, and trimmed by splicing (spliceosome) and debranching (lariat enzyme) without Drosha intervention. In both pathways, the pre-miRNAs are exported to the cytoplasm by Exportin-5 were they are further processed into double-stranded-RNAs called miRNA/ miRNA* by DICER/TRBP. Argonaute separates miRNA/ miRNA* complex into a guide strand (miRNA) and the passenger strand (miRNA*). Furthermore, the miRNA bound to Ago is loaded into the RISC complex, which regulates gene expression by interacting with the 3′-UTR of the target mRNA either by translation repression or degradation (adopted and modified from [67]).

2.1.3.3 Micro RNA (miRNA)

MicroRNAs (miRNAs) are functional single-stranded molecules with a length of approximately 22 nt. Their main function is to aid gene expression process by regulating mRNAs through imperfect basepairing with their 3′-UTR, thus disrupting the translation output [68]. Besides regulating gene expression at the posttranscriptional level, they are involved in important processes like cell development, fate determination, proliferation, and apoptosis [69]. Additionally, it has been found that miRNA expression profiles in disease conditions suffer alterations compared to normal conditions, which leads to conclude that they also play a role in abnormal cellular function (further discussed in Section 2.5.1).

The miRNAs (Figure 2.7) are encoded within exonic, intronic, or intergenic regions of the genome. They are synthesized by RNA polymerase II, which produces a primary miRNA (pri-miRNA) transcript. The pri-miRNA is further cleaved by the Microprocessor complex (DROSHA and two DGCR8 molecules) near the base of the hairpin, and hence turned into precursor miRNA (pre-miRNA), a hairpin with an approximate size of 55-70 nt. Next, pre-miRNA is

recognized from $\sim$ 2 nt 3′ overhangs by Exportin-5 and transported from the nucleus to the cytoplasm where it undergoes maturation. This final processing involves assemblage with a complex of DICER and the Hsp90 chaperone, and generation of miRNA duplexes (only one strand is matured) by DICER cleavage. Subsequently, the mature miRNA is loaded into the RNA-induced silencing complex (RISC) together with an Argonaute 2 (AGO2) protein [70].

2.2 Generalities on the energy flow in biological systems: thermodynamics

The laws of thermodynamics provide general principles applicable to all biological and physical processes which determine the conditions under which a specific process may or may not occur. Two basic concepts in the laws of thermodynamics are the system and its environment. The system consists of the matter included in a defined region of space, and the environment corresponds to the remaining matter. Other important concepts are entropy and enthalpy. Entropy (S) is a thermodynamic quantity representing the unavailability of a system's thermal energy for conversion into mechanical work, often interpreted as the degree of disorder or randomness in the system. Enthalpy (H) can be defined as thermodynamic quantity equivalent to the total heat content of a system [71]. It is equal to the internal energy of the system plus the product of pressure and volume. Changes in enthalpy are associated with a particular chemical process.

The First Law of Thermodynamics states that the total energy of a system and its environment remain constant, that is, the energy content of the universe is constant, since the energy can not be created or destroyed; however, it can take different forms (e.g. heat or potential energy) [72]. In chemical systems, potential energy is related to the likelihood that atoms will react to each other. In sum, it implies that the energy that is released from the formation of chemical bonds can be released in a form of heat, can be used to break other bonds, or simply stored as another source.

The Second Law of Thermodynamics states that the total entropy of a system and its environment always increases [72]. However, it is possible for the entropy to decrease locally, as long as it increases in other parts of the universe by an equal or greater magnitude. This local decrease usually leads to a release of heat which in turn increases the entropy of the environment. The degree of entropy of a system can be expressed more conveniently as the free energy (G) of the system, in other words, the available energy to produce a reaction. The change in free energy or Gibbs free energy (ΔG), is the thermodynamic potential that is minimized when a system reaches chemical equilibrium at constant pressure and temperature. Its derivative with respect to the reaction coordinate of the system vanishes at the equilibrium point. Therefore, it determines whether a reaction is energetically favorable (or spontaneous) or not. For a process to occur spontaneously at constant pressure and temperature, the free energy must decrease, or in other words, the change in G of the system must be negative.

Regarding chemical reactions, the change in free energy ΔG is a way to measure the amount of entropy that a particular reaction will create, and hence a measurement of how likely it will occur spontaneously [35]. For example, given a reversible reaction $A \leftrightarrow B$, if $\Delta G < 0$, then the reaction is expected to occur on $A \rightarrow B$ direction. Concentrations of the molecules (denoted by [ ]) involved in the reaction, affect ΔG. Considering the same reversible reaction, if $[A] > [B]$, then the direction of the reaction tends to $A \rightarrow B$; as the ratio of [A] with respect to [B] increases, $\Delta G(A \rightarrow B)$

decreases and $\Delta G(B \rightarrow A)$ increases. For reaction $A \rightarrow B$, if there are equal concentrations of reactants and products $[A] = [B]$, then $\Delta G(A \rightarrow B) < 0$, and thus the reaction is energetically favorable; conversely $\Delta G(B \rightarrow A) > 0$. However, still some occurrences of $B \rightarrow A$ will take place. As the concentrations of the reactants decrease, products increase, and the ratio of [A] with respect to [B] increases as well, which in turn increases ΔG value, which ultimately equals 0 (reaches equilibrium), or shifts to positive, and thus the reaction turns unfavorable.
The equilibrium constant for the reaction $A \rightarrow B$ is given by

$$K = \frac{[A]}{[B]} \tag{2.1}$$

The change in free energy ΔG, is dependent on the aggregated stored energy of all the molecules involved in the reaction, as well as their concentrations; therefore it does not allow for direct comparison of the energy from different reactions. In order to perform this comparison, the change of free energy under a standard condition, also known as the standard free-energy change (ΔG°), has to be calculated for each reaction. ΔG° allows for quantitative prediction of biological reactions [73]. All the concentrations of reactants are set to 1 mole/liter (mole = $6 \cdot 10^{23}$ molecules of a substance), consequently depending only on reacting molecules.
Hence ΔG° of a reaction $A \rightarrow B$ is given as follows

$$\Delta G = \Delta G^\circ + RT \ \ln \frac{[A]}{[B]} \tag{2.2}$$

where ΔG units are kJ/mol, $[A, B]$ correspond to the concentrations of reactants in moles/liter, $\ln$ is the natural logarithm, the gas constant $R = 8.314 \ \mathrm{J}K^{-1}\mathrm{mole}^{-1}$, and T is the temperature in Kelvin. If $[A] = [B]$, then ($\Delta G = \Delta G^\circ$). Plugging in Eq. (2.1), the equation for the standard free energy change is

$$\Delta G^\circ = -RT \ \ln \ K \tag{2.3}$$

To sum up, the ΔG° of a reaction gives a good indication of its equilibrium. Generally, large negative values for ΔG° relate to large values for equilibrium constant.

2.2.1 Thermodynamics of nucleic acids

The primary structure of nucleic acids consists of nucleosides (G, C, A, T/U; for DNA and RNA, respectively) linked together by phosphodiester bonds, forming the covalent structure of the molecule. The complementary Watson–Crick hydrogen bonding of intermolecular single strands (DNA) or intramolecular basepairs (i.e. tRNA cloverleaf structure) defines the secondary structure. The tertiary structure indicates the three dimensional architecture of the nucleic acid molecule, or in other words, its arrangement in space [74].

Biologically functional nucleic acid duplexes include either two DNA strands, a DNA and RNA strand (hybrids), or two RNA strands (or an intramolecular complex, e.g. in rRNAs and tRNAs) [71]. Prediction of nucleic acid secondary structures can be done by calculating the minimum free energy (MFE) of folding (lowest free energy value), using dynamic programming algorithms [74]. Although in theory this structure is the most likely to form, alternative forms are also possible (at least transiently).

The evaluation of oligonucleotide-duplexes stability and estimation of thermodynamics can be achieved by using the nearest-neighbor parameters [75]. For the estimation of the stability in *in vivo*, the optimal temperature of the organism is considered (e.g. *E. coli*, 37°C). the thermodynamic parameters (ΔH°, ΔS°, and $\Delta G^\circ 37$) for a duplex formation with the nearest-neighbor model are comprised by the terms: (*i*) free energy change as the sum of each subsequent basepair; (*ii*) free energy change for initiation; and (*iii*) free energy change of a mixing entropy term if a duplex is a self-complementary strand.

There are different methods for predicting hybridization reactions according to the type of a complex. For instance, in the case of DNA/DNA complexes, the most used is the nearest-neighbor model from SantaLucia and Hicks [76], mainly because it is quite accurate and simple to calculate; some models used in the case of RNA/RNA complexes are the ones proposed by Xia *et al.* [77] or Zucker *et al.* [78]. For DNA/RNA that of Sugimoto *et al.* [79] or Watkins *et al.* [80].

Table 2.1: Thermodynamic nearest neighbor (NN) parameters for DNA Watson Crick basepairs in 1 M NaCl at 37°C (adopted from [76]).

Propagation sequence	ΔH°	ΔS°	ΔG°_{37}
AA/TT	-7.6	-21.3	-1.00
AT/TA	-7.2	-20.4	-0.88
TA/AT	-7.2	-21.3	-0.58
CA/GT	-8.5	-22.7	-1.45
GT/CA	-8.4	-22.4	-1.44
CT/GA	-7.8	-21.0	-1.28
GA/CT	-8.2	-22.2	-1.30
CG/GC	-10.6	-27.2	-2.17
GC/CG	-9.8	-24.4	-2.24
GG/CC	-8.0	-19.9	-1.84
Initiation	+0.2	-5.7	+1.96
Terminal penalty (AT)	+2.2	+6.9	+0.05
Symmetry correction	0.0	-1.4	+0.43

2.2.1.1 Thermodynamic calculations for *in silico* prediction of DNA interactions

When a dsDNA molecule (native form) is heated, the hydrogen binding forces between the two strands are broken and eventually the strands will separate. This process is known as denaturation [73]. Under certain conditions, a ssDNA molecule (denatured) can form hydrogen bonds in a complementarity-driven manner and adopt a double helix structure again, in a process called renaturation. When the renatured DNA is built from DNA molecules of different origin, or between one DNA molecule and a RNA molecule, the process is known as hybridization. It can be defined as a phenomenon in which ssDNA or RNA molecules anneal DNA or RNA driven by the complementarity (full or partial) of their sequences.

The nearest-neighbor model [76] indicates that the stability of a dsDNA molecule depends on the basepairs that make up the nearest-neighbor. Table 2.1 shows the 10 possible basepair interactions, the symbol "/" in the sequence indicates where the first strand ends and that both are in antiparallel orientation; symmetry correction only applies to self-complementary duplexes. Additionally, to each duplex that ends with the sequence AT a penalty of +1.0 kcal/mol is applied to ΔG°. The units for ΔH° and ΔG° are kcal/mol and the units for ΔS° is J/K.

An example calculation of the change of energy of a dsDNA molecule with the nearest-neighbor model [76], as follows:

5' - CGTTGA - 3'
3' - GCAACT - 5'

Calculation of its change of free energy at $37°C$ is given by

$$\Delta G^{\circ}_{37}(total) = \Delta G^{\circ}_{37i} + \Delta G^{\circ}_{37s} + \sum \Delta G^{\circ}_{37k} + \Delta G^{\circ}_{37t} \tag{2.4}$$

where ΔG°_{37i} corresponds to the initiation energy of the molecule, ΔG°_{37s} refers to symmetry correction energy, $\sum \Delta G^{\circ}_{37k}$ is the energy generation by the duplex molecule and, ΔG°_{37t} represents the terminal AT penalization, then

$$\begin{aligned}\Delta G^{\circ}_{37}(total) =& \Delta G^{\circ}_{37i} + \Delta G^{\circ}_{37s} + \Delta G^{\circ}_{37k}(\text{CG/GC}) + \Delta G^{\circ}_{37k}(\text{GT/CA}) + \\ & \Delta G^{\circ}_{37k}(\text{CG/GC}) + \Delta G^{\circ}_{37k}(\text{TT/AA}) + \Delta G^{\circ}_{37k}(\text{TG/AC}) + \\ & \Delta G^{\circ}_{37k}(\text{GA/CT}) + \Delta G^{\circ}_{37t}\end{aligned}$$

plugging in parameter values (Table 2.1) accordingly

$$\Delta G^{\circ}_{37}(predicted) = 1.96 + 0 - 2.17 - 1.44 + 1.0 - 1.45 - 1.30 + 0.05$$

hence, the change of free energy is given by

$$\Delta G^{\circ}_{37}(predicted) = -5.35 \text{ kcal mol}^{-1}$$

2.2.1.2 Thermodynamic calculations for *in silico* prediction of RNA interactions

For the study of antisense interactions (e.g. RNAi or miRNA-mRNA), the reliable prediction of secondary structures is essential. A general approach for those predictions is by free energy minimization [74, 78], which is based on the assumption that the addition of the values from stacking basepairs and loop entropies defines the free energy [81]. The thermodynamics of RNA-RNA interactions can be seen as the addition of: (*i*) the required energy for opening/exposing the binding site (complementary sequence), and (*ii*) the energy from hybridization. RNA folding problem and calculations for the thermodynamic model via partition-function are solved by dynamic programming algorithms. Most of them are based on the assumption that RNA secondary structures are free from pseudoknot (crossing basepairs) formations. In other words, they ignore the secondary structures between monomers and only account for intermolecular basepairs, hence favoring computational speed.

Accuracy in prediction of secondary structures can be enhanced by only taking canonical structures (they do not contain isolated bases or enclosed basepairs) into account. Therefore, co-folding model is biophysically more accurate [81]. It is based on the concatenation of the sequences from both RNA molecules, and uses particular energy parameters for the loop that contains the cut-point (first nt of second strand) between both sequences. Energy parameters are dependent on the sequence, length and type of loops, and they have been extensively measured an collected in the last three decades [82–84]. Stacking probabilities are calculated as post-processing of dynamic programming tables entries, and have been previously computed and stored in the folding algorithms (ViennaRNA).

For computational treatment, RNA secondary structures are formatted as outer-planar graphs. The vertices represent the nucleotides, and the edges the basepairs and covalent backbone of the molecule. The idea behind the algorithm is that each basepair within the evaluated secondary structure, divides the structure into an exterior and interior section, each solved separately, and further added together to the energy model. Energy loops (Figure 2.8A) can be biophysically distinguished in stacked basepairs, hairpins, bulges, true interior loops, and multi(branched) loops [85]; algorithmically they can be treated as stacked pairs (although in some cases these require separate handling), bulges and true interior loops as subtypes of interior loops.

Dynamic programming algorithms for solving energy-directed RNA folding decompose the secondary structure into further sub-structures defined on subsequences (Figure 2.8A,B). RNAcofold [78], which is included in ViennaRNA Package [86] (used for calculations in this work), is an implementation of the model described in [87] for the calculation of MFE. As a first step, the evaluated secondary structure is decomposed into smaller substructures or energy loops (Figure 2.8A). This implementation only distinguishes 4 types of loops: hairpin, interior, multibranch and exterior. This is due to the bulges, stacked pairs and true interior loops can be treated as a subtype of an interior loop, for the purpose of the algorithm [85]. Furthermore, the free energy (stability) is individually calculated (Figure 2.8B) for each loop. Finally, the free energy values from the decomposed loops are added together, resulting in the total free energy of the evaluated secondary structure. The decomposition scheme for a secondary structure is outlined by the RNA folding grammar (Figure 2.8C) or secondary structure folding recurrences. For the evaluation of the free energies, a set of nearest neighbor parameters is used, particularly the one compiled by Turner (2004) [88]. These parameters are divided by the type of the structure (i.e. Watson-Crick helices, GU pairs, dangling ends, terminal mismatches, hairpin loops, bulge loops, internal loops, coaxial stacking, multibranch loops and exterior loops) [89]; these experimental parameters were overall measured for $T = 37°C$ and 1 M sodium chloride solutions [87].

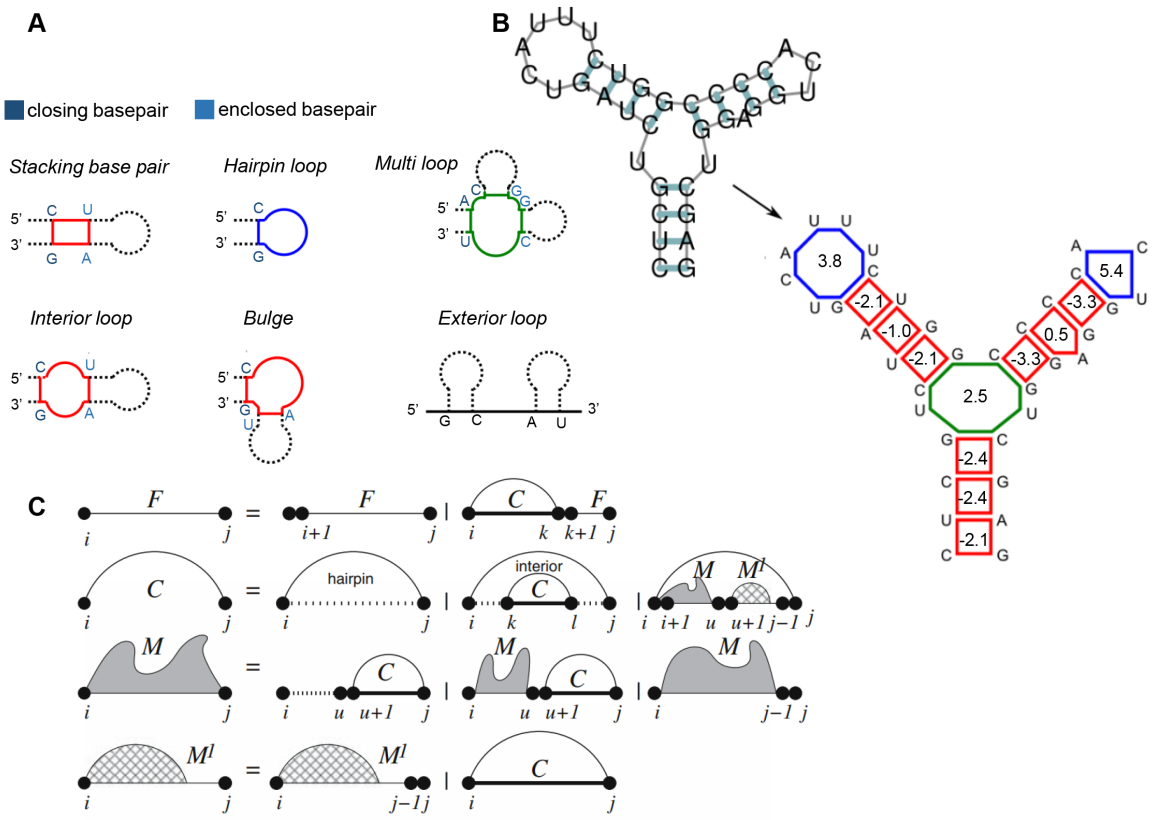


Figure 2.8: Energy loop decomposition of an RNA secondary structure. (A) Loop types are determined by the number of basepairs delimiting the formation of the loop. The exterior loop is the exception, since is a special case that lacks a closing pair. For example, hairpin loops are 1-loops (single closing basepair); interior loops are 2-loops (closing and enclosed basepair). When more than two basepairs are involved, they are called multibranch loops. **(B)** The secondary structure of the RNA molecule is subdivided into loops, each of them is scored in terms of free energy and lastly the sum of all individual scores, makes for the free energy of the evaluated structure (adopted and modified from [90]). **(C)** Graphical representation of secondary structure folding recursions; thin arcs represent substructures closed by a single basepair, whereas thick arcs denote substructures enclosed by a stacked pair. Decomposition rules correspond to disjoint sets of cases (Row 1-4). Row 1: a structure on $[i, j]$ starts either with an unpaired base or with a paired base. On Row 2: a structure enclosed in a base pair is either a hairpin loop, or delimited by an interior loop, or branches in a multiloop (i and j do not form a basepair); a multiloop is composed of two parts, one with one or more components (M) and another with exactly one component (M^1). Row 3 and 4: show recursions for the two types of multiloop components. The structures denoted by C are enclosed by a basepair that become canonical when enclosed by an additional exterior basepair (adopted from [91]).

The loop-based energy model description is as follows [85]:
Given an RNA sequence x with a length of n, the nucleotide at position k is x_k. Then, the subsequence from indexes (basepairs) k to l $(x_k, ..., x_l)$ is denoted by $x[k, l]$.

Hairpin loops (Figure 2.8A) are determined by their closing basepair (k, l), with ℓ representing the length of the loop (no. of unpaired nt), then its energy parametrization is:

$$\mathcal{H}(k, l) = \mathcal{H}(x[k, l]) = \mathcal{H}(x_k, x_{k+1}, \ell, x_{l-1}, x_l) \tag{2.5}$$

Interior loops (Figure 2.8A) are determined by their enclosing basepairs and its energy parameters are given as:

$$\mathcal{I}(k,l;p,q) = \mathcal{I}(x[k,p],x[q,l]) = \mathcal{I}(x_k,x_{k+1};\ell_1;x_{p-1},x_p;x_q,x_{q+1};\ell_2;x_{l-1},x_l) \tag{2.6}$$

where $\ell_1 = p-k+1$ and $\ell_2 = l-q+1$ denote the length of the unpaired strands between k and p, and q and l, respectively.

A stacked pair (Figure 2.8A) is given if $\ell_1 = \ell_2 = 0$, and a bulge only if one of either ℓ_1 or ℓ_2 is missing. The energy model is symmetrical: $\mathcal{I}(k,l;p,q) = \mathcal{I}(q,p;l,k)$.

The energy of multiloops (Figure 2.8A) is independent of the sequence of the closing pair, it follows an additive model of the form $\mathcal{M} = a + b \times B + c \times \ell$, where ℓ is the length of the multiloop (no. of unpaired nt), $B \geq 2$ the number of branches (the one containing the closing pair of the loop is not counted).

The decomposition rules (Figure 2.8C) indicate that for the evaluated secondary structure, each one of the of the possible sub-structures (energy loops) corresponds to exactly one of the sub-cases. The following arrays for $i < j$, correspond to the types of substructures in Figure 2.8C [85]:

F_{ij} - free energy of the optimal substructure on the subsequence $x[i,j]$.

C_{ij} - free energy of the optimal substructure on the subsequence $x[i,j]$ when i and j are a basepair.

M_{ij} - free energy of the optimal substructure on the subsequence $x[i,j]$ subject to the constraint that $x[i,j]$ is part of a multiloop and has at least one sub-sequence enclosed by a basepair.

M^1_{ij} - free energy of the optimal substructure on the subsequence $x[i,j]$ subject to the constraint that $x[i,j]$ is part of a multiloop and has exactly one sub-sequence that has (i,h) as a closing pair for some h satisfying $i < h\ j$.

Arrays F, C and M are the ones that have to be stored for the computation of the minimum free energy structure; M^1 full array is required for other more elaborate backtracking subprocesses in Vienna RNA Package (e.g. RNAsubopt, that produces all RNA secondary structures in an energy interval). The recursions of the energy minimization algorithm [87] as implemented in Vienna RNA package (graphical representation in Figure 2.8C) are as follows [85]:

$$\begin{aligned}
F_{ij} &= \min\Big\{F_{i+1,j},\quad \min_{i<k\leq j} C_{ik} + F_{k+1,j}\Big\} \\
C_{ij} &= \min\Big\{\mathcal{H}(i,j),\quad \min_{i<k<l<j} C_{kl} + \mathcal{I}(i,j;k,l),\quad \min_{i<u<j} M_{i+1,u} + M^1_{u+1,j-1} + a\Big\} \\
M_{ij} &= \min\Big\{\min_{i<u<j}(u-i+1)c + C_{u+1,j} + b,\quad \min_{i<u<j} M_{i,u} + C_{u+1,j} + b,\quad M_{i,j-1} + c\Big\} \\
M^1_{ij} &= \min\Big\{M^1_{i,j-1} + c,\quad C_{ij} + b\Big\}
\end{aligned} \tag{2.7}$$

Decomposition is necessary for partition-function algorithms, for obtaining the partition-functions, operators are replaced in the recursions, minimum operations with sums, and additions with multiplications. The inverse thermal energy is denoted by $\beta = 1/RT$, and the initialization for the algorithm is $F_{ii} = 0, C_{ii} = M_{ii} = M^1_{ii} = +\infty$ [91]:

$$
\begin{aligned}
Z_{ij} &= Z_{i+1,j} + \sum_{i<k\leq j} Z^B_{ik} Z_{k+1,j} \\
Z^B_{ij} &= e^{-\beta \mathcal{H}(i,j)} + \sum_{i<k<l<j} Z^B_{kl} e^{-\beta \mathcal{I}(i,j;k,l)} + \sum_{i<u<j} Z^M_{i+1,u} Z^{M1}_{u+1,j} e^{-\beta a} \\
Z^M_{ij} &= \sum_{i<u<j} e^{-\beta(u-i+1)c} Z^M_{u+1,j} + \sum_{i<u<j} Z^M_{i,u} Z^B_{u+1,j} e^{-\beta b} + Z^M_{i,j-1} e^{-\beta c} \\
Z^{M1}_{ij} &= Z^{M1}_{i,j-1} e^{-\beta c} + Z^B_{i,j} e^{-\beta b} \\
Z_{ii} &= 1, \quad Z^B_{ii} = Z^M_{ii} = Z^{M1}_{ii} = 0
\end{aligned}
\tag{2.8}
$$

A canonical secondary (C) structure does not contain basepairs that are not stacked on another pair, and its partition-function is given by Z^*_{ij}.

2.2.1.3 Tools for *in silico* prediction of nucleic acid secondary structures and interactions

A brief introduction to the software, tools and libraries that have been used in the development of this work:

Python and R are interpreted high-level programming languages suitable for scripting. Python is considered as a general-purpose programming language whereas R is mostly used for statistical and data analysis purposes.

ViennaRNA package is a C code library and a set of programs related to evaluation of RNA secondary structures (e.g. optimal secondary structure prediction, RNA-RNA interaction analysis, energy evaluation, folding path identification, consensus structure prediction, and structural alignment) [86]. It is available as webservices or can be downloaded and used with customized scripts. The scripting language interfaces for Python were used in this work, for calling the program RNAcofold using in-house written scripts.

NUPACK is software suite for the analysis and design of nucleic acid structures, devices, and systems [92]. It offers a web-based application where analysis and design algorithms can be used, as well as the possibility to download the algorithm implementations, which can be directly used from console or further used within customized scripts. For this work, the tools for the calculation of probability of formation of a MFE structure, and the ensemble defect were of particular interest.

Biopython is a freely available set of libraries [93] for Python language offering tools for computational molecular biology. The tools for handling nucleic acid sequences were used.

RNApredator is a software for prediction sRNA-mRNA interactions in bacteria [94]. It uses a dynamic programming approach to compute the possible targets. Furthermore, post-processing of selected targets aid to identify differences on the target mRNA before and after the binding of the sRNAs. Also, a Gene Ontology(GO) and pathways (KEGG) term enrichment analysis can be performed on a set of selected targets. The functionality for identification of possible targets for an

sRNA was used.

2.2.1.4 General notation for *in silico* representation of nucleic acid sequences, structures and interactions

The FASTA format (.fas, .fasta) is the most used [95] for representing both nucleic acids and amino acid sequences. It is a text-based format, which begins with a single line containing the symbol '>', immediately (without space) followed by the identifier of the sequence. Next lines correspond to string of the sequence (A,T,G,C), which ends when another '>' is found; this indicates the input of another sequence (or the end of the file), in this manner multiple sequences can be defined in a single file:

```
>name_sequence1
CCTGCGGAAGATCGGCACTAGAATAGCCAGAACCGTTTCTCTGAGGCTTCCGGCCTTCCC
TCCCACTAATAATTCTGAGG
>name_sequence2
CCATCGGTAGCGCATCCTTAGTCCAATTAAGTCCCTATCCAGGCGCTCCGCCGAAGGTCT
ATATCCATTTGTCAGCAGACACGC
>name_sequence3
CCACCCTCGTGGTATGGCTAGGCATTCAGGAACCGGAGAACGCTTCAGACCAGCCCGGAC
TGGGAACCTGCGGGCAGTAGGTGGAAT
```

The dot-bracket-notation was early introduced for the usage of ViennaRNA package algorithms [87], and even nowadays is the most common way of representing the secondary structure of nucleic acids for algorithmic purposes. As seen in Figure 2.9, matching brackets represent paired bases and dots are used for the unpaired bases. Different strands can be indicated with an ampersand or with a plus sign in case of NUPACK [92].

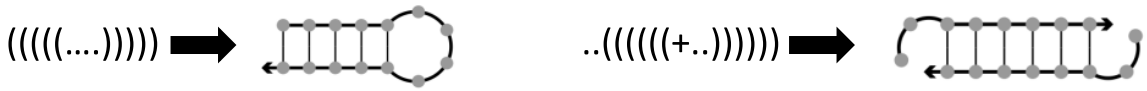


Figure 2.9: Examples of notation for secondary structure representation and corresponding plots. Notation of a single-stranded hairpin (left) and notation for a semi paired double strand (right). Structure plots generated with NUPACK.

2.3 Gene expression in prokaryotes

In prokaryotes, transcription and translation can occur synchronously in the cytoplasm. The regulation of gene expression involves complex networks, in the case of transcriptional regulation, promoters and transcription factors (TFs) are fundamental elements. Regulatory elements that are encoded within the same gene they act upon, are called cis-regulatory; trans-regulatory elements belong to a different gene than the one they regulate.

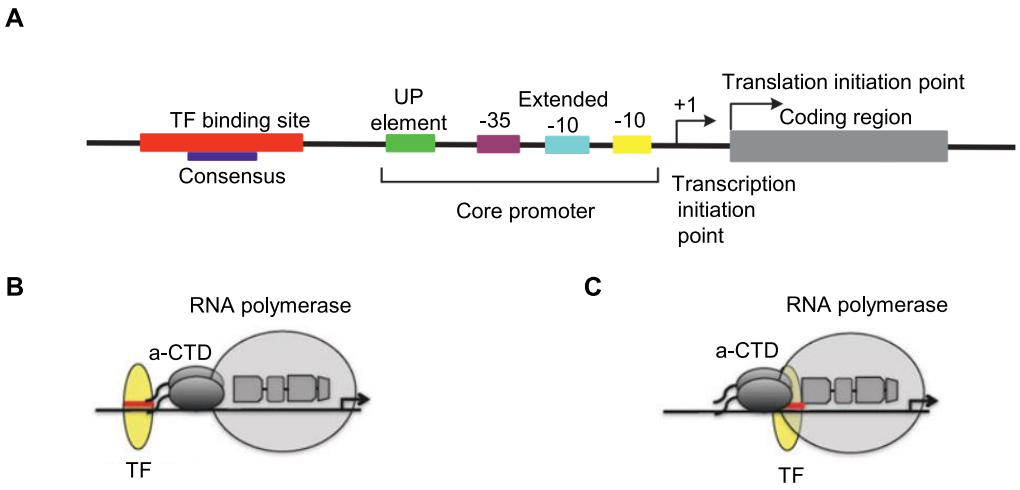


Figure 2.10: Scheme depicting the structure of a prokaryotic promoter and the transcription factors. Transcription begins with the RNA polymerase (RNAP) binding to a promoter region, this process is regulated by transcription factors. **(A)** Structure of promoter, the basepairs are numbered, the transcription start point is indicated by +1; the basepairs upstream of +1 are positive, while those downstream are negative. The core of the promoter correspond to the –10, –35 and extended –10 and UP elements. RNA polymerase σ subunit domain 2 recognizes –10 from the core promoter, and domains 4 and 3 the –35 and extended –10 elements, respectively; the C-terminal domains of the RNA polymerase α subunits recognize the UP elements (adopted and modified from [96]). Transcription factors assist the RNA polymerase in regulating transcription and can be classified into **(B)** activators or **(C)** repressors, regarding their binding site relative to the transcriptional start site (adopted and modified from [97]).

2.3.1 Transcription

In prokaryotes, the transcription process can be divided into three phases: initiation, elongation and termination. Initiation phase corresponds to the binding of the RNA polymerase holoenzyme to a promoter region (40-60 bp in length) that contains two conserved elements: the -10 sequence (Pribnow box) and the -35 sequence with the consensus TATAAT and TTGACA, respectively (Figure 2.10). Promoter sequences are not included in the RNA synthesis, it begins 35 base pairs downstream of the first consensus region and ten base pairs downstream of the second [96]. In nature, almost none of the promoters exactly matches the consensus sequence, but the degree of similarity to it correlates with the promoter strength (how actively the RNA polymerase transcribes from it); the transcription frequency of strong promoters is very high, conversely weak promoters are transcribed less.

Transcription factors (Figure 2.10B-C) are proteins that by binding to the cis-regulatory sequence (or TF-binding site) within a promoter, can either aid (activators) or hinder (repressors) transcription initiation depending on its binding location with respect to the transcription start site. The TF binding site contains a consensus sequence (-10 region), for example in E. coli a common sequence is TATAAT [98], it is also frequent that the TF binding sites overlay the core sequence from the promoter. TFs can bind to cis-acting DNA sequences, other TFs, or to both. Activators (Figure 2.10 B) perform positive control by stimulating RNA polymerase and thus gene expression, whereas repressors (Figure 2.10C), perform negative control by binding and blocking the RNA polymerase

from starting transcription. TFs can either be activators or repressors, or dual regulators that hold both functions. There is a plethora of transcription factors, for example, in the Escherichia coli K-12 genome there are $\sim$ 314 TFs, where 35% are activators, 43% repressors, and 22% exhibit dual function [99].

The core prokaryotic RNA polymerase ($\alpha_2\beta\beta'$, two alpha subunits, one beta subunit, and one betaprime subunit) can synthesize RNA but it is not capable of recognizing the promoter [100]. The sigma subunit (σ) plays an important role in this, when binding to the core enzyme, allows for the now holoenzyme ($\alpha_2\beta\beta'\omega\sigma$) to recognize the promoter region. The σ factor of RNA polymerase interacts directly with the Pribnow sequence.

In the elongation phase, σ factor is decoupled from the transcriptional complex, staying only the core enzyme ($\alpha_2\beta\beta'\omega$), in charge of the elongation of the RNA transcript [100]. RNA is produced from the template DNA strand (antisense), thus having the same sequence as the coding strand (sense), with U exchanges on T.

RNA synthesis stops with the encounter of a termination sequence, the most common is a GC-rich palindromic region followed by an AT-rich region [101]. The RNA from the palindromic region forms a hairpin structure followed by a U stretch at the 3′ terminus of the nascent RNA. The other type of termination signal sequences (no hairpin structure) require the aid of a rho (ρ) protein for recognition of the termination site.

Around 20% of the dry weight of the prokaryotic cell corresponds to RNA [102]. RNA transcripts from protein genes (mRNA) barely undergo modifications in prokaryotes. Many mRNAs are translated simultaneously along transcription before their synthesis is complete. Nonetheless, precursor rRNA and tRNA require posttranscriptional processing.

2.3.2 Translation

The prokaryotic mRNA is polycistronic, that is, many proteins are coded under a single promoter. mRNA translation ($\sim$ 10-20 amino acid residues per second) takes place in the cytoplasm. Half life of prokaryotic mRNA is short (seconds to minutes) [103]. Translation can occur coupled with transcription [104], as soon as the newly transcribed RNA emerges from the RNA polymerase. Guanosine triphosphate (GTP), a purine nucleotide, is the source of energy during translation, especially on translocation of the ribosome and elongation start. The prokaryotic translation cycle (Figure 2.11A) can be divided into four phases, each one aided by protein factors: initiation, initiation factors (IFs); elongation of the polypeptide chain, elongation factors (EFs); termination, release factors (RFs); and recycling, recycling factors (RRFs). The details on the cycle presented here particularly correspond to *E. coli*, workhorse of molecular biology (also used in this work), and a representative example of prokaryote organisms.

In the initiation phase (Figure 2.11A steps 1-3), (1) the 30S (small subunit) binds to mRNA on the Shine-Dalgarno sequence (AGGAGG), located 8-10 nt upstream of the start codon (AUG). (2) IF1 binds to 30S, preventing at this point any entry of tRNAs to the A site. IF2-GTP aids on the binding of 30S with fMet-tRNA, which is a special tRNA that carries a formylated methionine (fMet); it recognizes and binds the start codon, for this reason, fMet is found at the N terminus (start) of every polypeptide chain. IF3 binds to 30S, preventing early binding of 50S, and further aiding in mRNA binding. The 30S, mRNA, initiation factors (IF1, IF2, and IF3), and fMet-tRNA primed in the P site,

form together the initiation complex. (3) The 50S (large) subunit binds to the initiation complex, assembling the 70S (complete ribosome) and initiating translation [105].

The elongation phase (Figure 2.11A step 4) is the most conserved in both bacteria and eukaryotes, and is aided by homologous elongation factors: EF-Ts make for active EF-Tu; EF-Tu brings the aminoacyl-tRNA (aa-tRNA) to the A site in the ribosome, EF-G aids in translocation of the ribosome, and EF-P aids in the synthesis of difficult-to-synthesize peptide bonds [106]. During this step, polysomes are assembled, that is, multiple ribosomes simultaneously translating one mRNA, forming large helical complexes: the inner frame is occupied by the small ribosomal subunit and mRNA, whereas the outer frame is composed by the large ribosomal subunit (the nascent peptide emerges here). The nascent peptide folds cotranslationally, in the restricted space of the polypeptide exit tunnel of the ribosome (Figure 2.5A).

Termination (Figure 2.11A step 5) occurs when encountering a termination codon (UAA, UAG, or UGA) by release factors RF1 (UAG/UAA) and RF2 (UGA/UAA). RF1 and RF2 also aid in hydrolysis of the ester bond of the peptidyl-tRNA (by their conserved GGQ motif) [106], which is catalyzed by the peptidyl transferase center on the ribosome. RF3 aids in R1/R2 dissociation from the 70S.

In the recycling (steps 6 and 7) phase, with the aid of RRF (that binds to the A site on the ribosome) and EF-G [107], (6) the 30S and 50S subunits dissociate from each other and EF-G, and (7) from the mRNA and tRNA. All of these elements are recruited into a new translation initiation complex virtually instantly.

2.4 Gene expression in eukaryotes

Gene expression in prokaryotes and eukaryotes share some basic similarities. Nonetheless, eukaryotes posses larger and significantly more complex genomes than prokaryotes, and mainly differ regarding its organization. Eukaryotic genes are monocystronic with several introns and exons. The former are spliced out in mRNA maturation process. Prokaryotic genes do not undergo splicing and often are organized in a polycistronic manner (i.e. multiple genes grouped in operons with common controlling elements). Furthermore, eukaryotic gene expression regulation is combinatorial, since it requires the coordinated interactions of multiple molecules (i.e. proteins, transcription factors, ncRNAs) [108].

2.4.1 Transcription

RNA synthesis in eukaryotes takes place in the nucleus and essentially follows the same phases as in prokaryotes (see Section 2.3.1). However, the organization of protein coding genes in eukaryotes is different. Coding sequences within the genes are called exons, whereas non-coding sequences are called introns (ranging in size from 80 bp to $\sim$ 10 000 bp) [36].

Eukaryotes have three RNA polymerases [109]: RNA Pol I (nucleolar) transcribes rRNAs (28S, 18S and 5.8S); RNA Pol II (nucleoplasmic) transcribes mRNAs, snoRNAs and most snRNAs; and RNA Pol III (nucleoplasmic) transcribes tRNAs, 5S rRNA, U6 snRNA and the 7S RNA of the signal recognition particle (SRP) involved in protein translocation across the membrane of the endoplasmic reticulum. The eukaryotic RNA polymerases are composed by at least 12 subunits, sharing some homology with the *E. coli* RNA polymerase; around 4-7 subunits in the enzyme are unique, holding

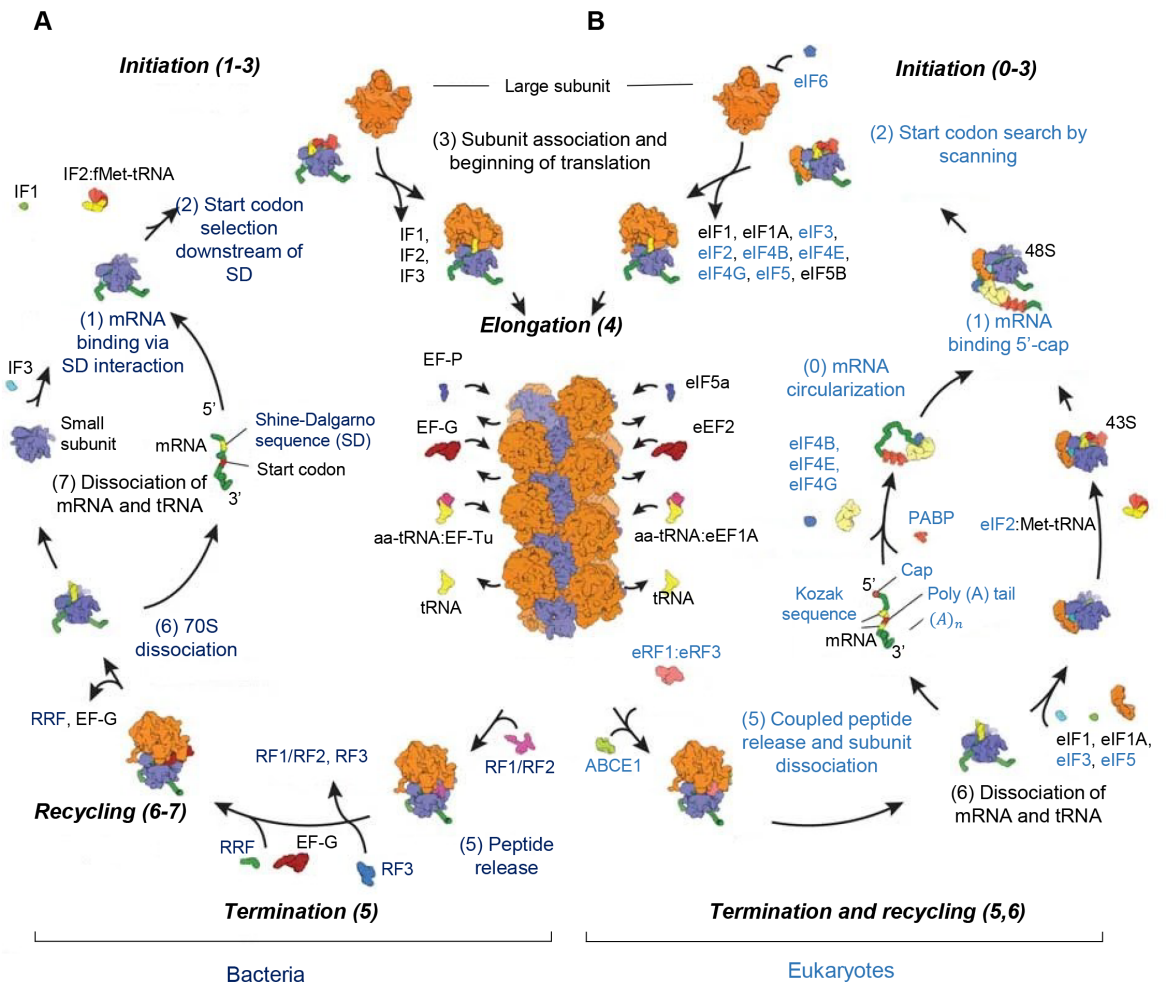


Figure 2.11: Schematic of translation cycle in prokaryotes and eukaryotes. Differences on the 4 phases of the translation cycle between **(A)** bacteria and **(B)** eukaryotes are depicted. In dark blue the steps and nonhomologous protein factors unique to bacteria, in light blue for eukaryotes, in black homologous factors and common steps occurring in both (adopted and modified from [46]).

no similarities to bacterial or other eukaryotic RNA polymerases.

There are different conserved promoter elements for the genes transcribed by each polymerase [109]: for RNA polymerase I genes have two GC-rich promoter sequences in the -45 to +20 region; RNA polymerase III has upstream promoters (within the genes); most promoter sites for RNA polymerase II have a TATA box (TATAAA) $\sim$ 25 bp upstream +1.

Initiation phase starts with the identification of the promoter (5′ of +1) by basal (or general) transcription factors (named with prefix TFII). More specifically, the transcription factor IID (TFIID) binds to the TATA box in the promoter by means of its TATA box binding protein (TBP) to aid in the recruit of the appropriate polymerase [110]. When TFIID is bound, another TF (TFIIA) binds to stabilize the TFIID-TATA box interaction. Next, TFIIB intervenes as a bridge protein, which binds to TFIID, and also can bind to RNA polymerase II. TFIIE and TFIIH further bind to complete the transcription pre-initiation complex (PIC), which contains at least 40 polypeptides.

RNA polymerase transcribes only the antisense strand of the DNA template in 5′ to 3′ direction. TFIIH (helicase) unwinds the DNA helix (with the use of ATP) to allow for transcription, it also

phosphorylates the C-terminal domain (CTD) of RNA polymerase II, inducing a conformational change and dissociation of the TFs [111]. The RNA polymerase II starts elongation phase, moving along the template and synthesizing RNA.

Termination phase occurs at various distances downstream of the gene, depending on each of the different RNA polymerases, followed by the release of the primary transcript (pre-mRNA). There is no specific signal for RNA Polymerase II, it can transcribe from a few to a thousand bp after the end of the gene, but the transcript is cleaved at an internal site before the end of transcription [112]. In case of protein genes, the cleavage site locates upstream an AAUAAA sequence and downstream a GU-rich sequence, which is separated by about 40-60 nt in the synthesized RNA. In humans, the CPSF protein binds the AAUAAA sequence, and CstF the GU-rich sequence, forming a complex; CPSF cleaves the nascent pre-mRNA at 10-30 nt downstream the AAUAAA site [113]. Poly(A) polymerase is also part of the complex (together with CPSF and CstF), which adds a 3′ poly-A tail on the pre-mRNA ($\sim$ 200 residues are incorporated).

The pre-mRNA has to be further processed to turn into mature mRNA: intron sequences are removed (RNA splicing) and exons joined together (7 methyl guanosine), a 5′ cap is added (in all mRNAs), polyadenylation (in most of the mRNAs) and 3′ cleavage [111]; translation is enhanced when both the 5′ cap and poly (A) tail are added to a mRNA [114]. Elongation is connected to RNA processing, since the processing enzymes associate with the RNA transcript through the CTD of RNA polymerase II, and the transfer to the growing pre-mRNA.

Enhancers, and silencers also help to regulate the frequency of RNA synthesis. Enhancers are positive transcriptional control elements (100–200 bp) located either upstream or downstream of the target gene [115]. Repressors inhibit transcription of specific genes by binding either to control elements near to the target gene or to silencers (long distance sequences).

2.4.2 Translation

In eukaryotes, ribosomes are located in the cytoplasm, and it is where translation ($\sim$ 3-5 amino acid residues per second) takes place [116]. The half life of the eukaryotic mRNA is longer than that of prokaryotes (few hours to few days). Eukaryotic translation cycle (Figure 2.11B) is also divided into four phases: initiation, elongation of the polypeptide chain, termination and recycling of ribosomes. The protein factors which aid in the processes are initiation factors (eIFs), elongation factors (eEFs), release factors (eRFs) and recycling factors [117]. Eukaryotic mRNA is monocistronic, that is, there is only one proteinogenic gene under each promoter.

There are $\sim$ 12 eIFs involved in the initiation phase (Figure 2.11B steps 0-3) [118]. It begins with the (0) mRNA circularization, as the result of the interaction of eIF4G simultaneously binding to the poly(A) binding protein (Pab1p) and eIF4E [114]. The complex composed by eIF4G and eIF4E recruits 40S through an interaction with eIF3. The eIF3 is associated with 40S, and prevents the 60S from premature binding; it also interacts with eIF4F complex (eIF4A, eIF4E and eIF4G). The complex composed by eIF4G and eIF4E recruits 40S through an interaction with eIF3. The eIF3 is associated with 40S, and prevents the 60S from premature binding; it also interacts with eIF4F complex (eIF4A, eIF4E and eIF4G). The eIF2 selects the not formylated initiator tRNA (Met-tRNA$_i$) into the ternary complex (eIF2:GTP:Met-tRNA$_i$), this binds to the 40S subunit, forming the pre-initiation complex (43S, or 40S and mRNA) [117]. Factors eIF4A and eIF4B and eIF4F aid

in the binding of the 43S to the 5′ cap of the mRNA. (1) The 5′ cap of the mRNA is recognized by the cap binding protein eIF4E. The 43S moves toward the mRNA 3′ end (2) scanning for the start codon (AUG), usually contained within the Kozak consensus (ACCAUGG). The 43S bound to the mRNA on the start codon by the anticodon of the initiator tRNA (on the P site of the ribosome) forms the initiation complex (48S). The ATP-dependent RNA helicase eIF4A helps unwind secondary structures within the mRNA transcript, eIF1A enhances eIF1 in the dissociation of aberrant complexes from the mRNA, and further mediates the assembly of 48S complex on the start codon. Factors are displaced from the 48S and (3) 60S (large) subunit binds to the complex. For the assembly of 80S (ribosome), eIF5B intervention is required, since it has a ribosome-dependent GTPase activity; the eIF5 induces the hydrolysis of eIF2:GTP. In some mRNAs, translation initiation is cap-independent, and occurs instead by internal ribosomal entry [119].

During elongation phase (Figure 2.11B step 4), eEF1A in ternary complex with GTP brings the aminoacyl-tRNA (aatRNA) to the A site on the ribosome. EEIF5A aids the synthesis of peptide bonds and eEF2 is involved on the mRNA–tRNA complex translocation [120].

Termination and recycling (Figure 2.11B steps 5 and 6) begins with the recognition of a termination codon (UAA, UAG and UGA) by eRF1. With the aid of eRF3 (a special GTPase), and IF5A (binds to site E on the ribosome), eRF1-mediated hydrolysis releases the (5) coupled peptide. After termination, recycling of the subunits occurs when ABCE1 dissociates the 80S by ATP binding hydrolysis, aided by eRF1 [42].

2.5 Molecular approach of diseases

Advances in molecular biology offer the possibilities to monitor DNA, RNA, metabolite and protein variation among organisms in a wide scale. This has enabled to define diseases at molecular level, and to find associations between mutations on DNA and clinical conditions like cystic fibrosis and Duchenne muscular dystrophy. This has facilitated a systematical view of disease [26], where networks of interacting molecules characterize physiological states related to diseases.

Molecular diagnostics together with biomarkers is used for screening and diagnosis of several genetic disorders, for example Down's syndrome, muscular dystrophy or psoriasis [121].

miRNAs are a promising type of biomarkers, since not only are expressed among cells and tissues, but also found extracellularly. It has been shown that dysregulated miRNA expression profiles correlate with specific disease states [28], especially in cancers.

2.5.1 Biomarkers of diseases: miRNAs

As discussed in Section 2.1.3.3, the main function of miRNAs is to negatively regulate gene expression by binding to 3′ UTR of mRNAs. Since they are key components of the cell in normal and pathological states, they hold potential for diagnostic and prognostic biomarkers, and as predictors of drug response [29].

The study of miRNAs as biomarkers of disease has been heavily explored in cancer research. It has been shown that miRNAs dysregulation is related to tumor growth , invasion, angiogenesis, and immune invasion [30]. Moreover, the first anticancer therapeutic against several cancers, a liposomal miRNA mimic (MRX34) entered clinical phase 1 trial studies on 2013; however the first-in-human

clinical trial failed [122]; particularly, an optimal delivery strategy and toxicity remain as strong obstacles to overcome.

Furthermore, miRNA dysregulation has been observed in several disease conditions apart from cancers. For example, genetic diseases like cystic fibrosis [28], neurodegenerative such as amyotrophic lateral sclerosis [31], and even in hypertension [32], to name a few.

MicroRNAs are not only found inside cells, but also extracellularly enclosed in vesicles, more precisely in exosomes, which prevents them from enzymatic degradation [123]. Extracellular vesicles (EVs) are remarkable stable nucleus-absent and lipid bilayer particles, originated in the endosome, released into the extracellular environment, and absorbed by adjacent cells [124]; they are thought to be released from all cell types and to intervene either positively or negatively in the function of its recipient. Exosomes are lipid vesicles (30–200 nm) that contain a non-uniform composition of mRNA, miRNA, tRNA, RNA, sRNA, DNA, lipids and proteins [125]. These vesicles can be isolated from biological fluids, for example cerebrospinal fluid (CSF), blood plasma and serum, breast milk, lymph, bile and saliva. Biological fluctuations of circulating miRNAs within EVs, can occur as a disease progressed or as a response to treatment [30].

The fact that miRNAs can be detected in biological fluids, opens the possibility for developing less-invasive techniques for monitoring and diagnosis of diseases. To date there are nearly 40,000 entries in the miRBase (Release 22.1, accessed on April 2021) [126] from annotated and published sequences of miRNAs from different organisms, from which almost 2,000 were found in human.

2.6 Digital circuits: binary variables, Boolean algebra and logic gates

Boolean algebra is the foundation of digital circuit design [127]. This branch of algebra is named after George Boole who in 1847 formulated its bases in his book *'The Mathematical Analysis of Logic'* [128]. Boolean logic is performed by Boolean operators (Figure 2.12A): conjunction (AND), disjunction (OR) and negation (NOT) applied to binary variables. Binary variables are categorical variables that have only two possible outcomes, either TRUE (also referred as 1 or ON) or FALSE (also referred as 0 or OFF). These operators can be further combined into more elaborated Boolean functions or expressions.

Digital circuits are the basis of all digital computers, this electric circuits are based on two discrete voltage levels to determine their state, one ~ 0 volts (OFF), and other at a higher level (ON) depending on the supply voltage that is being used (e.g. ~ 5 volts in TTL or transistor-transistor logic, widely used in computers) [129] in each technology. The building blocks of digital circuits are logic gates. A logic gate (Figure 2.12B) is a device (model or physical electronic) that implements a Boolean function, and has one or more inputs and only one output [130]. Their output is given by the values of their inputs according to the function they implement. The truth table of a digital logic circuit lists all possible combinations of its input values and the only possible value of the output of each particular case.

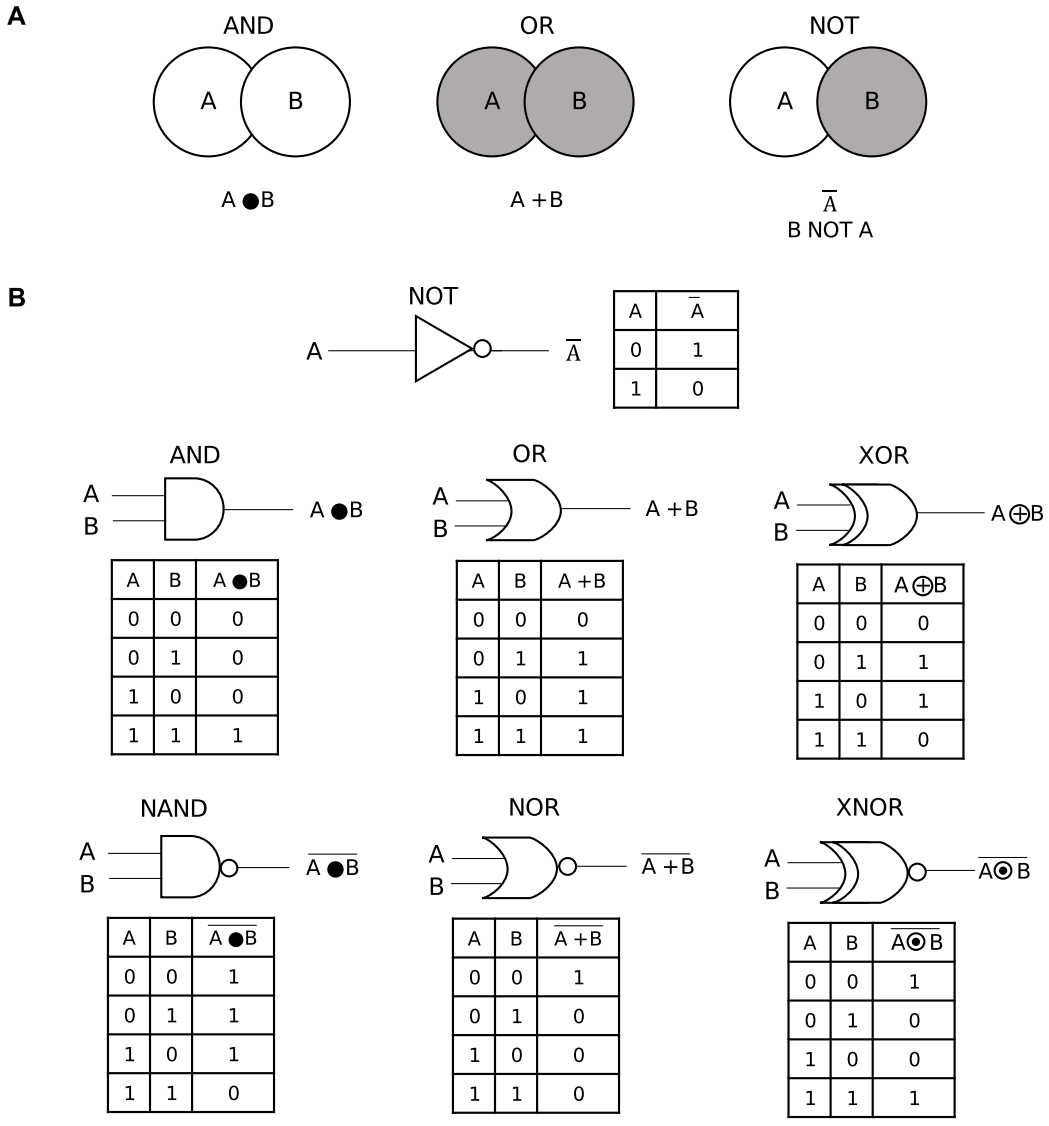


Figure 2.12: Digital circuits overview. **(A)** Graphical representation of Boolean operators. **(B)** Basic logic gates with their corresponding diagrams, algebraic or switching functions and truth tables.

2.7 State of the art: molecular computing and synthetic biology

The original idea of harnessing molecules to build computing devices dates from 1959, when Richard Feynman discussed this possibility in his famous lecture *'There's Plenty of Room at the Bottom'*, which was later published in 1961 [131]. Nonetheless, Tom Head was one of pioneers to provide the basis for understanding the mathematical theory behind operations with biomolecules. Particularly, important to mention his work on formal language theory for operations carried out by means of restriction enzymes [132]. However, it was not until 1994, that the first proof-of-principle of molecular computing was reported, when Adleman used nucleic acids, namely DNA, as a computation element to solve an instance of 7 vertices of the Hamiltonian path problem (HPP) [133].

The early classical models of biomolecular computation (e.g. filtering models, sticker systems and splicing systems) are considered non-autonomous [134], that is, it is necessary for a person to monitor the process and perform the corresponding laboratory operations. Initially, molecular computation was used to solve NP-Complete problems (until now impossible to solve with conventional computing). However, operations were slow and usually susceptible to errors and the volume of DNA that is required to codify a problem grows exponentially according to its complexity [135]. Due to this, autonomous computing emerged as the second generation of the molecular computing (e.g. self-assembly, finite state automaton, DNA hairpin models). This is based on self-assembly properties of molecules (sometimes modulated by enzymes) which gives it autonomy and the possibility of being partially programmable. Molecular computing research can be divided into two main areas: the first one studies new computational paradigms based on the characteristics of molecules; the second one, includes the implementation of molecular devices that emulate elements of traditional computing, such as carbon-based transistors or molecular logic gates [134].

DNA computing is a branch of biomolecular computing concerned with the use of DNA as a carrier of information to make arithmetic and logic operations. Since every DNA sequence has a complement, for example sequence S= 5′- GTCG -3′, and its complement, S^*= 3′- CAGC -5′, both sequences (S and S') could hybridize to form double stranded DNA. This property of complementarity makes DNA a perfect data structure for performing computation and can be exploited in many ways. DNA computing methods can be employed to address two kind of problems: finite state problems, such as arithmetic and Boolean operations and combinatorial problems, for instance, the Hamilton path problem (HPP), satisfiability problem (SAT), among others. DNA computing techniques can be divided in three classes: (*i*) intramolecular, (*ii*) intermolecular and (*iii*) supramolecular [136]. The first one focuses on performing operations with single stranded DNA molecules. The second, on the hybridization between different DNA molecules to perform computation. The third one, makes use of self-assembly of DNA molecules with different sequences. Conventional computing has a basic suite of operations to perform calculations, like addition, bit-shifting, logical operators (AND, OR, NOT, etc); similarly, DNA has cutting, copying, pasting, repairing, among others. The DNA properties and techniques that are used for performing computation are: Watson-Crick Based Pairing, hybridization and denaturation, ligation, polymerization, Polymerase Chain Reaction (PCR), gel electrophoresis (for result visualization), DNA extraction, and parallel operations.

Synthetic biology is a field that combines expertise in biology and engineering to design and implement *de novo* functions *in vivo* [137]. For the design, building and characterization of the systems, key concepts have been borrowed and adapted from the engineering disciplines, for example, abstraction, standardization, modularity, and datasheet specifications. One of the aims of the field is to provide alternative solutions to the common standards in other important fields such as pharmaceutical and petroleum industries by engineering living organisms into technologies [138]. The publication of Francois Jacob and Jacques Monod (1961) paved the road for the field, which unveiled the existence of cellular regulation and molecular networks, particularly by studying the *lac* operon in *E. coli* [139]. Advances in the filed can be divided chronologically in three periods [140]: foundational (2000-2003) early studies concentrated on building gene regulator circuits that exhibited similar behavior as electric circuits; intermediate (2004-2007), characterized expansion of the field

but lack in engineering advances, and recent (2008 to date), with advances towards practical applications in biotechnology and medicine.

Molecular logic systems (such as logic gates or circuits) apply to many areas of considerable importance such as life sciences or medicine [141]. Due to its small size, molecular logic systems enable detection or recognition of microparticles, specially in the intracellular environment, which could eventually lead to make intelligent diagnostics possible. In addition, many approaches have been employed to construct molecular based logic gates and circuits, among them we can mention computational genes [142], DNAzymes [143–145], restriction enzymes [146], deoxyoligonucleotides [147], DNA self-assembly [148], DNA strand displacement [12], and RNA-based ribozyme-aptamers [149] and toehold switches [19].

2.7.1 Genetic circuits

Regulation of gene expression is carried out by genetic regulatory systems, which are complex networks of interactions between many interconnected components (including DNA, RNA, proteins, and small molecules), engaged in positive and negative feedback loops [150].

A genetic circuit is an interconnected set of biological parts (DNA sequences encoding regulatory or protein genes) that is introduced to a biological chassis (host organism). There are two main ways for performing the genetic manipulation in cells and introducing the synthetic constructs, by modifying the genome (e.g. with CRISPR) or by adding extrachromosomal plasmids [151]. Using the transcriptional and translational machinery from the host, these parts together enable the cells to calculate a particular response or emulate a logical function [137]. These constructs have applications in bioenergy (converting waste-to-fuel), environment (biosensors or degradation of pollutants) or medicine (diagnosis or production of drugs).

Living organisms can be programmed to perform electronic inspired behavior such as oscillators, bistable switches, waveform generators, logic functions, or high/low/band-pass filters by means of genetic circuits [11]. A very well known example is the repressilator [10] network, consisting of a feedback loop with three genes; each genes from the circuit expresses a regulatory protein that represses the following gene in the construct (Figure 2.13).

Although genetic circuits are the core of synthetic biology, there is a well known problem in this type of constructs, that is the expression of the reporter gene even when the inducer is absent, also known as leakiness. To date there are still ongoing efforts to tackle the leakiness on genetic circuits. A recent example is the adaptation of a nonsense suppressor tRNA based system [1] to existing biosensors (e.g. lactate and arabinose).

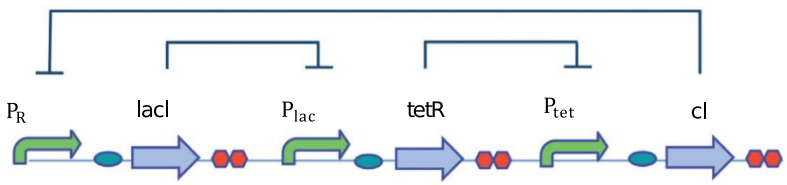


Figure 2.13: The repressilator: a milestone in genetic circuits. This biological circuit acts as a ring oscillator [10], it consists of a network of genes and regulatory elements, which act as 3 interconnected NOT gates with feedback (adopted and modified from [137]).

2.7.2 Toehold-mediated strand displacement

At its core, a strand displacement reaction can be described as the process by which two partially or completely complementary strands hybridize to each other, displacing one or more pre-hybridized strands (Figure 2.14A). In strand displacement designs, the DNA sequences are abstracted domains, this means that a segment of the sequence is represented by a certain number or letter (i.e. 1), and its complementary domain with the same letter accompanied by an asterisk (i.e. 1^*). The kinetics of strand displacement is controlled by certain mechanisms called toeholds, which are short regions of DNA sequences responsible for initiating the reactions. Strand displacement reactions are programmable due to the meticulous design of the used sequences, for example, to obtain a fast and reliable reaction the toehold domains should be short (about 6 nt), while the displacing domains must be longer (about 20 nt) [13].

DNA strand displacement (DSD) is one of the most important paradigms in the field of DNA nanotechnology [14], particularly because it allows programming computation only in terms of nucleotide sequences. In other words, this paradigm allows the implementation of computational devices without the need for additional components, for example enzymes [12].

Toehold-mediated strand displacement requires additional external single stranded DNA signals to operate. To avoid this, the strand displacement reactions may be performed in cascades. Cascading implies that the output from one reaction serves as input for another reaction, thereby eliminating the need to add single strands that trigger the reactions in each step, so it is possible to construct complex autonomous systems. An example of this is given in [16] were cascade strand displacement reactions and building blocks (gates) are proposed for the construction of a four-bit square-root molecular circuit that comprises 130 DNA strands. Each block consists of a single gate (seesaw) that is configured in such a way that an input catalytically converts the fuel into an output if the concentration of the input exceeds the threshold level [17]. Input, output and fuel signals are DNA single strands, whereas the gates correspond to partially hybridized double stranded DNA structures. Although it is suggested that this architecture can be scalable to circuits with thousands of gates, this has not been shown. Some drawbacks of this approach is that for constructing more complicated circuits that require more interconnected gates, many more different DNA sequences have to be generated, then effectiveness of the thresholds would decrease and crosstalk between strands becomes more likely. This would in turn slow down the desired reactions and the overall the circuit performance [15–17].

Another obstacle to tackle is implementing reversibility or recycling in DNA logic gates, since most of the existing approaches are not reusable, meaning once the gate releases the output, it becomes waste. This characteristic is sought not only to optimize the gate performance, but also would bring a step closer to interesting *in vivo* applications in biomedicine such as molecular diagnosis.

2.7.2.1 Toehold-switches

RNA interactions can be programmed by designed base-pairing reactions by means of toehold switches [153]. A toehold switch (Figure 2.14B) is a construct in which an mRNA molecule that encodes a protein of interest, holds a second structure formation (hairpin) preventing the ribosome from reaching the ribosomal binding site (RBS), hence translation is initially turned off [18]. This device detects a target RNA that contains the complementary sequence of the toehold in the left half of the hairpin. When the target RNA binds, the hairpin unwinds and exposes the RBS allowing the

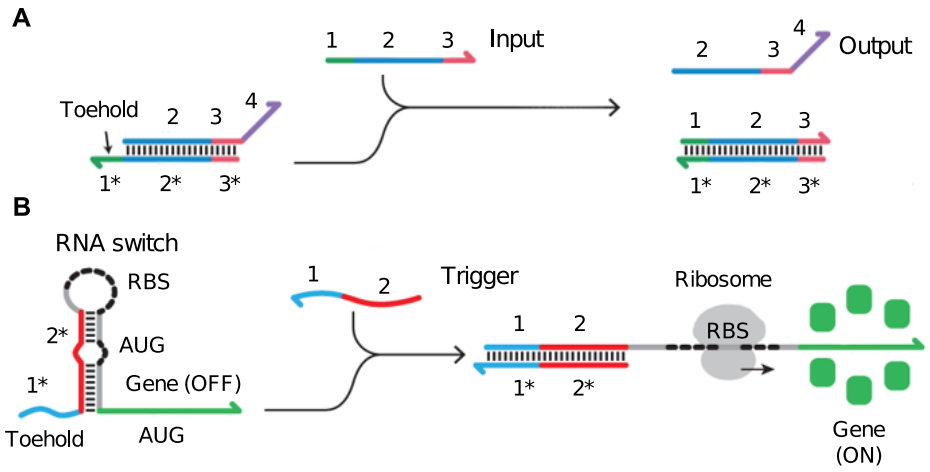


Figure 2.14: Toehold-mediated strand displacement reactions. **(A)** An example of a toehold-mediated reaction. A double stranded complex contains an exposed toehold domain (1^*), which is complementary to domain a from the input strand; the input strand binds to the toehold, and by a branch migration process (domains 2 and 3), it displaces the pre-hybridized output strand. The result of the reaction is a double stranded complex and the released output, that can be further employed in other reactions. **(B)** The toehold switch is a designed translational regulator. The switch makes use of a 5′ toehold domain designed to be complementary to a trigger sequence (or input). Once the trigger binds to the toehold on the switch, the hairpin structure unwinds, exposing the RBS and AUG codon, giving space for the ribosome to dock and translate the output gene (adopted and modified from [152]).

ribosome to proceed translation [154].

This approach can be used to build Boolean logic gates, however it presents some limitations [19]. The gates available (AND, OR, AND-NOT) are sufficient to compute any Boolean logic expression, but they must be transformed to into disjunctive normal form (DNF). Therefore, to avoid this transformation step, it would be desirable to have designs for all the basic gates (AND, OR, NOT, XOR, NOR, NAND, XNOR), additionally it would allow for circuit simplification. Having a only a complete set (e.g. AND,OR, NOT; NAND, NOR, AND, NOT; OR, NOT) will have the same issue of a conversion step. A set of gates is a complete set if we can implement any logical function using only the type of gates that belong to it. Another observation regarding the approach is that the gates require further optimization for scaling the complexity of the circuits, particularly, their OR gate has low efficiency compared to the other two gates, this will prevent overall high efficiency in more complex circuits [155]. Last, to enable the design of layered circuits, each gate should create an output that serves as the input for the next gate (cascading); the gates would then go back to their usual state, ready to process the next input, this is the basis for performing complex calculations. This behavior might be achievable by expressing transcription factors as output proteins that can control the expression of the gates in the next layer.

More importantly, since toehold switches can be used as sensors for different types of external stimuli or inputs (e.g. miRNA or RNA molecules derived from pathogens like virus) they can be applied for development of sensing technologies. Examples of this are proof-of-concept studies for paper-based synthetic circuits for diagnostics of Zika [156] or Ebola virus [20], and more recently, efforts regarding the development of biosensors for detection of SARS-CoV-2 [157].

Chapter 3

Engineering synthetic sRNAs as dynamical modulators of gene expression

In this chapter, a multilayered (i.e. both at transcriptional and posttranscriptional levels) tunable (i.e. dynamically regulated by a network of transcription factors and synthetic sRNAs) platform for the regulation of genetic circuits is presented. It is contemplated to be used along with existing assemblies or constructs, and as a proof of concept, it is adapted to negatively regulate (i.e. turn off) an existing constitutively expressed nonsense suppressor tRNA system in *E. coli*.

3.1 Synthetic genetic regulation and artificial sRNAs

The intrinsic machinery from the cell performs targeted knockdown of genes by RNA interference (RNAi) [158] and can be harnessed [159–161] to achieve programmable downregulation. Across all organisms, small noncoding RNAs (sRNAs) are present as *trans*-acting regulators of translation and stability of messenger RNAs (mRNAs) [162].

Engineering synthetic molecules that mimic naturally occurring molecules in the cell that already perform regulatory tasks [151, 163–166], expands the biotechnology and synthetic biology toolkit. Programmable and flexible downregulation of gene expression might aid in elucidating fundamental biological processes [167, 168] and re-engineering cell pathways or various applications, from metabolic engineering [169–172] to synthetic circuits [173, 174].

Synthetic biology (SynBio) aims to introduce *de novo* functions to living organisms by rationally designing and interconnecting biological parts (e.g. promoters, transcription factors, regulatory proteins and RNAs) from different sources (e.g. bacteria, bacteriophages, yeast) [137, 140, 150, 175]. Most synthetic constructs are powered primarily by transcription networks, where regulatory or reporter proteins are under control of promoters, which in turn respond to one or more transcription factors (TFs) [176]. Frequently these systems present variations on rates of transcription and translation *in vivo*, depending on multiple intrinsic and extrinsic factors [176], from the growth rate of the organism and availability of the cells resources, to experimental environment or measurements performed [174, 177, 178]. Altogether, fine tuning genetic regulatory components for achieving the desired functions or outputs is often complicated and overall a persistent challenge in the field [2, 178, 179]. Thus, the development of highly versatile and programmable genetic expression regulators is convenient.

In particular, ligand (i.e. transcription factor) inducible genetic constructs or systems must exhibit proper and well defined ON and OFF functions, and the difference (i.e. dynamic range) between these states is more often than not difficult to control [180]. The switch of an ON state to an OFF state, is referred as the NOT function or gate, and it is often achieved by negatively inducible promoters to repress a gene transcription on the presence of a specific ligand (e.g. P_{tet}) promoter is constitutively ON, and on its binding with TetR, transcription is hindered, or OFF); eventually leading to lack of production of the final protein output. Traditional gene editing knockout strategies (i.e. CRISPR-CS9, or introducing mutations) also offer the possibility to selectively stop the production of certain molecules, by suppressing or turning OFF genes; however in an irreversible way, since the genome has to be modified.

Bacterial sRNAs regulate gene expression positively and negatively, by pairing with their target mRNAs [181], modulating their rates of translation initiation and degradation [182]. In *Escherichia coli*, a large group of negative regulator sRNAs contain a consensus secondary structure, or scaffold to recruit the RNA chaperone Hfq [168]. Hfq promotes the annealing between sRNAs and target mRNAs [183], aids on the intracellular stability of the sRNAs [184], and guides RNase E for the destabilization and degradation of the mRNAs [60]. sRNAs negatively regulate their targets by binding to the 5′ region of the mRNAs [184], and covering either the Shine-Dalgarno (SD) or start codon (AUG) sequence, thus preventing the 30S ribosome entry, and in turn, translation initiation [185].

Interactions between sRNA and mRNAs occurring in posttranscriptional regulation are governed by the Watson–Crick base pairing principle [186, 187]. Therefore, they are predictable and can be programmable in a rational fashion by means of calculations of the hybridization energy of the molecules [188, 189]. Additionally, naturally occurring bacterial sRNAs present modular characteristics: an antisense sequence targeting its cognate mRNA followed by the Hfq recruiting sequence, which allows for synthetic sRNAs to be engineered following those principles [164, 167, 169, 190**?** –192]. Artificial sRNAs can be combined with other approaches, such as CRISPRi [173], or riboswitches [2] to coordinate and control synergistically transcriptional and translational regulation. Previous strategies for obtaining different expression levels from synthetic sRNAs included fluctuating their abundance by controlling their production by exchanging promoters with varying strengths [169, 192], rather than focusing on the sRNAs sequence.

Here, a rational design of synthetic sRNAs is proposed to dynamically modulate gene expression exhibiting different levels of repression, based on the binding affinity value of its target binding antisense region with their target mRNA. Since the transcription factors are the main effectors on both engineered and natural gene networks, synthetic posttranscriptional regulation is applied to dynamically modulate the expression of the gene they activate. To test the modulation activity of synthetic sRNAs, three different TFs that activate a common promoter are selected. First, genetic circuits are designed and characterized with exogenous parts in *E. coli*: a fluorescent protein reporter under the control of an positively inducible promoter (P_F from P2 phage) that can be activated by three different exogenous transcription factors (Delta, Pag, Ogr), each with different transcriptional activity, and classified them according their of reporter expression yield (weak, moderate, strong). Modulating synthetic sRNAs (repression levels: H > M > L) are designed to target transcription factor mRNAs to show that different ranges of repression to fine-tune gene expression can be achieved

independently of the construct. Furthermore, synthetic sRNA modulation is applied to turn off (highly repress) a constitutively expressed nonsense suppressor tRNA that was previously reported in the literature [33], only by the addition of a tailored sRNA, without altering the original design of the system. Overall, results suggest successful modulation of the dynamics of gene expression by the incorporation of synthetic sRNAs to existing genetic constructs, without modifications of genetic parts (e.g. exchanging promoters).

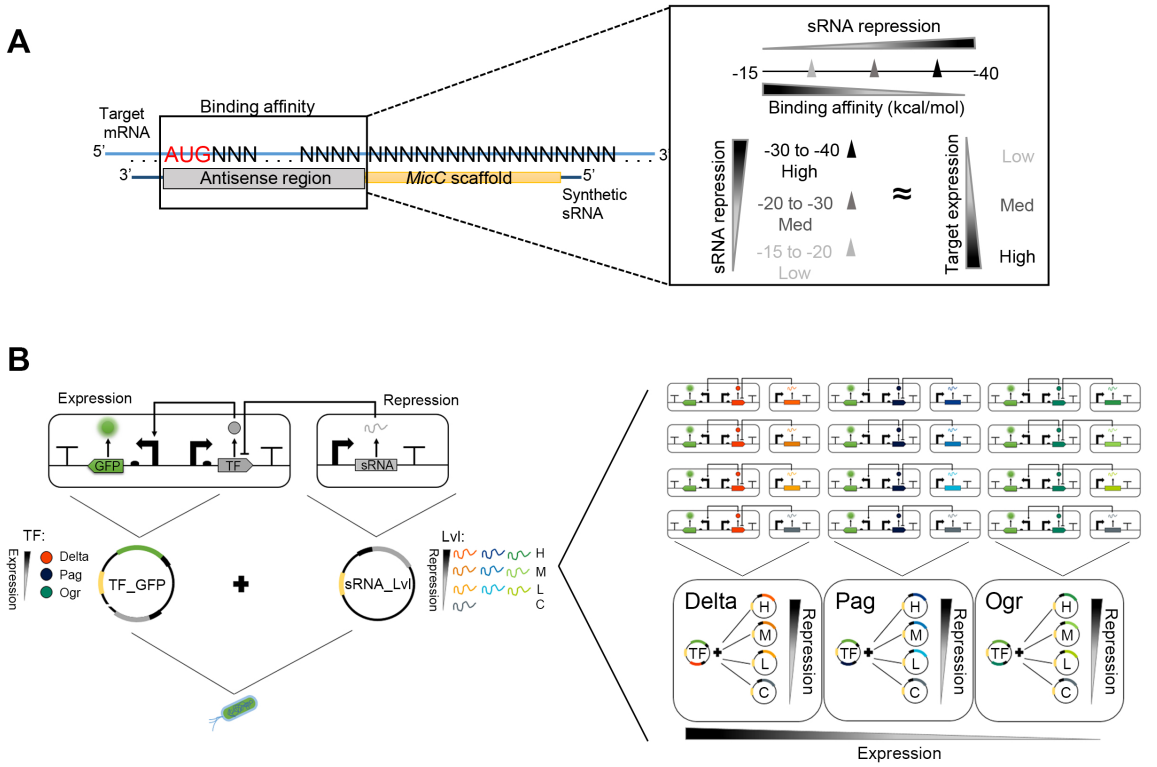


Figure 3.1: Overview of engineering modulating synthetic sRNAs. (A) Conceptual design of the synthetic sRNAs. **(B)** Conceptual design of the tunable expression-repression system. Transcriptional regulation strength: TF, high expression; TF, medium expression; TF, low expression. Translational regulation strength: H, high repression; M, medium repression; L, low repression.

3.2 Results

3.2.1 Designing synthetic sRNAs with modulating repression levels

Synthetic sRNAs mimicking bacterial regulators can be rationally designed following a modular architecture, by fusing a Hfq-recruiting motif from naturally occurring sRNAs (e.g. Spot42 [190], MicF [167], MicC [169, 193], DsrA [190, 194]) with the antisense sequence targeting the gene of interest [167, 184, 185, 190, 193**?**]. The *MicC* scaffold sequence is present on sRNAs that regulate the expression of OmpC in *E.coli* [195]. It has been previously shown [193] that artificial sRNAs can be built successfully by exchanging the *ompC* binding sequence from *MicC* with the complementary sequence to the translation initiation region (TIR) of the target mRNA (i.e. from the start codon (AUG)). Thus, customized sRNAs were designed with the antisense sequence to the desired mRNA to repress, immediately followed by the sequence of the *MicC* Hfq-scaffold (Figure 3.1A). The antisense

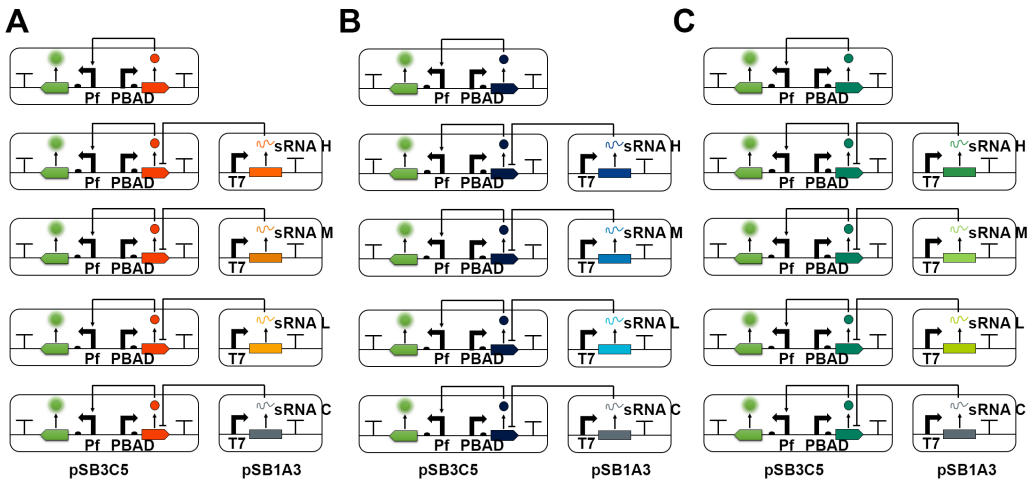


Figure 3.2: Expression-repression tandem system. **(A)** Delta, **(B)** Pag, and **(C)** Ogr constructs. Translational regulation strength by sRNAs: H, high repression; M, medium repression; L, low repression; C, control.

sequence, with length for optimal repression between 20-30 nt [**?**] ($\sim$ 24 nt [193]) corresponds to a partially complementary sequence to the TIR of the target gene, depending on desired the repression efficiency.

The binding energy value between the antisense (target-binding) sequence of an artificial sRNA and its target mRNA, correlates with the repression activity [169], enabling to predict the sRNA efficiency. Values ranging from -40 to -30 kcal/mol have shown empirically a high translation repression (up to 90%), with -15 kcal/mol as the minimum value able to repress translation ($\sim$ 10%) [193]. On the basis of this previous knowledge, thresholds were established within these known values to obtain different tuning levels: -40 to -30 kcal/mol for high, -30 to -40 kcal/mol for medium, and -20 to -15 kcal/mol for low repression (Figure 3.1A). Nucleotides from the antisense sequence were either deleted, added or randomly exchanged until the calculated binding energy value felt within the threshold for the expected repression efficiency.

3.2.2 Designing and characterizing synthetic circuits with different transcriptional activity and protein production yield

Instead of directly repressing a reporter gene, functional molecule genes were targeted, particularly TFs, since they are the base of cascading reactions in molecular circuits. *E. coli* strain BL21(DE3) was selected as the host, and the circuits were completely synthetic, as the parts are exogenous to the host. To determine if modulation levels of repression activity of the sRNAs holds independently of the genetic circuit or the target mRNA, different circuits (Figure 3.1B) were constructed with an invariable reporter construct and a variable TF gene under control of P_{BAD} promoter: δ TF (Delta) from P4-related retronphage ϕR73 [196], Pag TF from PSP3 bacteriophage [197], Ogr TF from P2 phage [196]. The reporter construct consists of the green fluorescent protein (GFP) gene under control of the exogenous promoter P_F from phage P2 [198]. P_F is one from the four promoters controlling transcription of late genes from P2 coliphage (which are responsible for head and tail formation), and it is intrinsically activated by the small phage-encoded protein Ogr [197].

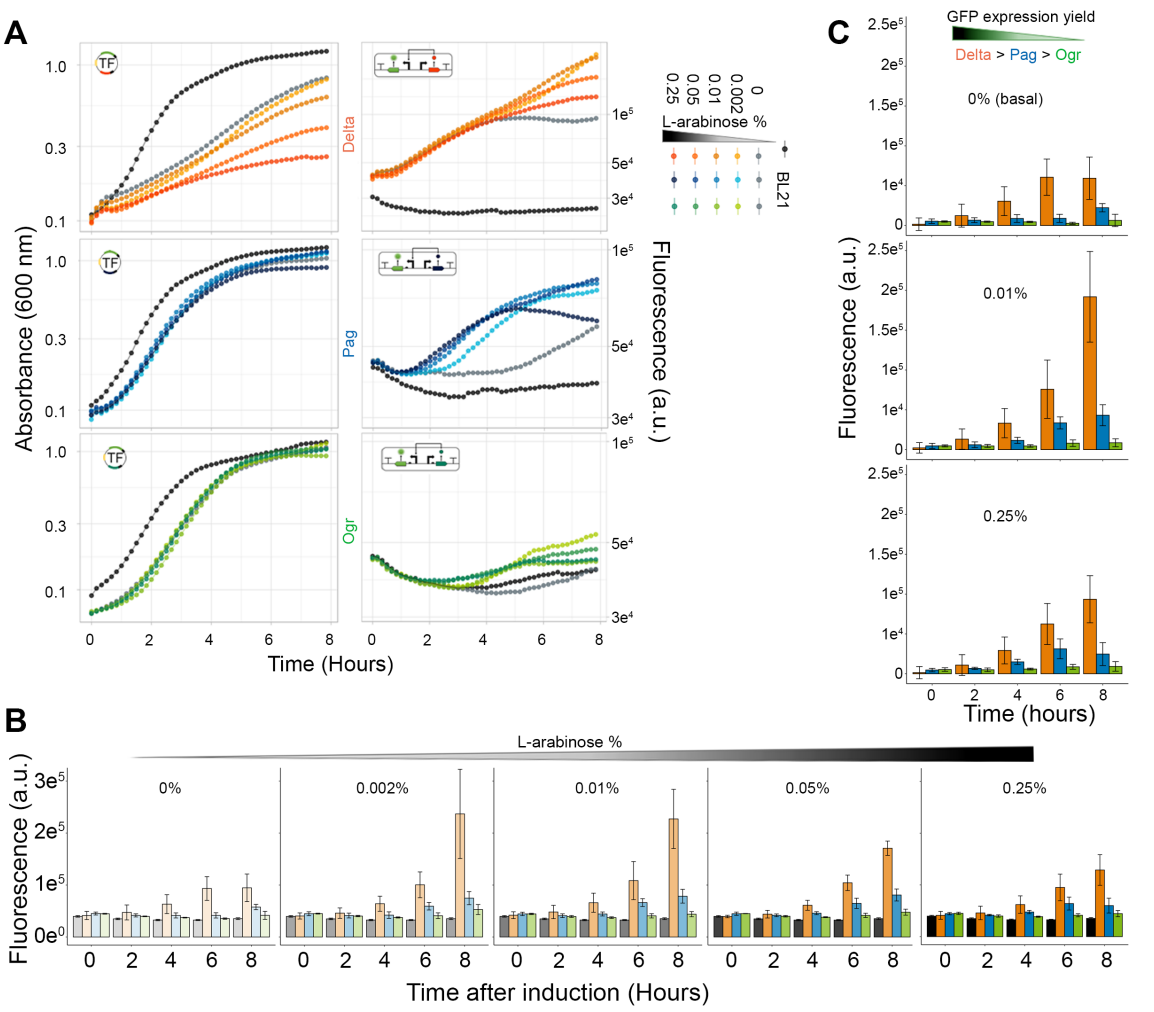


Figure 3.3: Characterization of genetic circuits. (A) Real-time monitoring and analysis of burden of the circuit by cell growth (right) and GFP yield (left) for each construct analysis of the dose dependence on the inducer (L-arabinose) of the TF gene expression on the cascade expression. **(B)** Raw measurements from the GFP yield dependent on dose of L-arabinose. **(C)** Normalized measurements from basal levels (0%), 0.01% and 0.25% dosages over time after induction.

All selected TFs are positive regulators of transcription, and activate the P_F promoter, differ however in transcription activity and consequently reporter protein (GFP) yield (strong, moderate, weak). Modulating sRNAs were designed to target each TF gene with different repression levels (high, medium, low, control), and co-expressed with the corresponding reporter circuit, ultimately resulting in 4 different circuits for each TF (Figure 3.1B). By adding L-arabinose, the P_{BAD} promoter is activated for transcription of the TF gene; the TF protein in turn, activates P_F to allow for transcription of GFP gene, ultimately resulting a a fluorescent signal that can be detected (Figure 3.2).

The estimation of quantitative measurements of a SynBio part is referred as characterization [176]. Each TF-construct was then characterized: the burden they impose to the host by evaluation of cell growth, fluorescence signal from the GFP expression cascade, and finally classified them according to the strength of the GFP yield (Figure 3.3). To determine the L-arabinose (inducer of P_{BAD} promoter) dose dependence on the transcription of the TF genes, and moreover the yield of GFP, a real-time evaluation of the fluorescence signal over a time span of 8 hours after induction was performed

(Figure 3.3A). The raw data from experiments, comparing control (BL21(DE3)) to the constructs across dose of L-arabinose (Figure 3.3B), showed that overall Delta reached the highest yield of GFP (Figure 3.3B) (however also the most detrimental to the cells), followed by Pag, and the lowest yield interestingly was from the intrinsic inducer Ogr. Furthermore, normalized data to control (BL21(DE3)) showed that the most proficient concentrations of L-arabinose inducer were 0.01% and 0.25% (Figure 3.3C), therefore, these concentrations were used for the further experiments for the sRNAs evaluation. It is also important to note, that constructs exhibited a basal fluorescence even without inducer, which is addressed in Section 3.2.4.

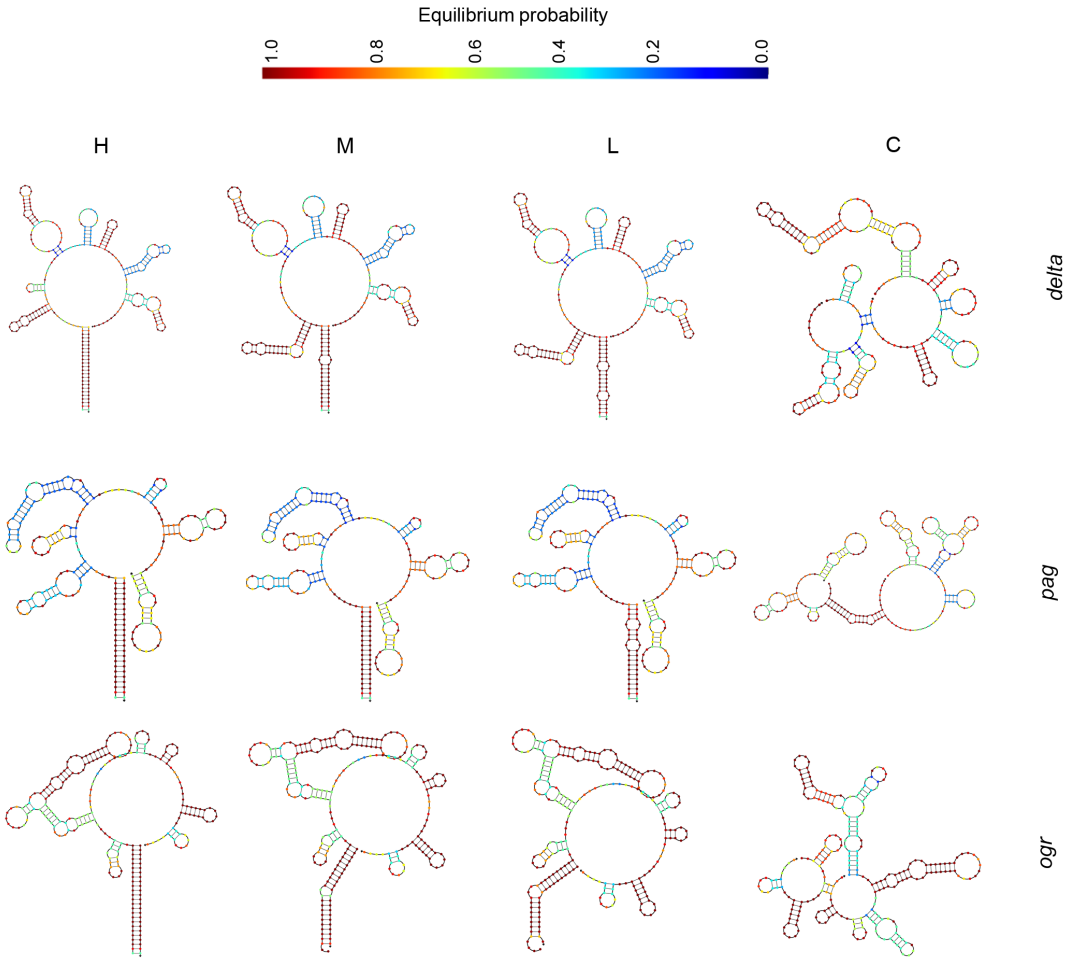


Figure 3.4: Minimum Free Energy structures for the TF mRNA-sRNA duplexes and basepair equilibrium probability. Predictions of *delta*, *pag*, *ogr* mRNAs paired with their cognate sRNAs (translational regulation strength: H, high repression; M, medium repression; L, low repression; C, control).

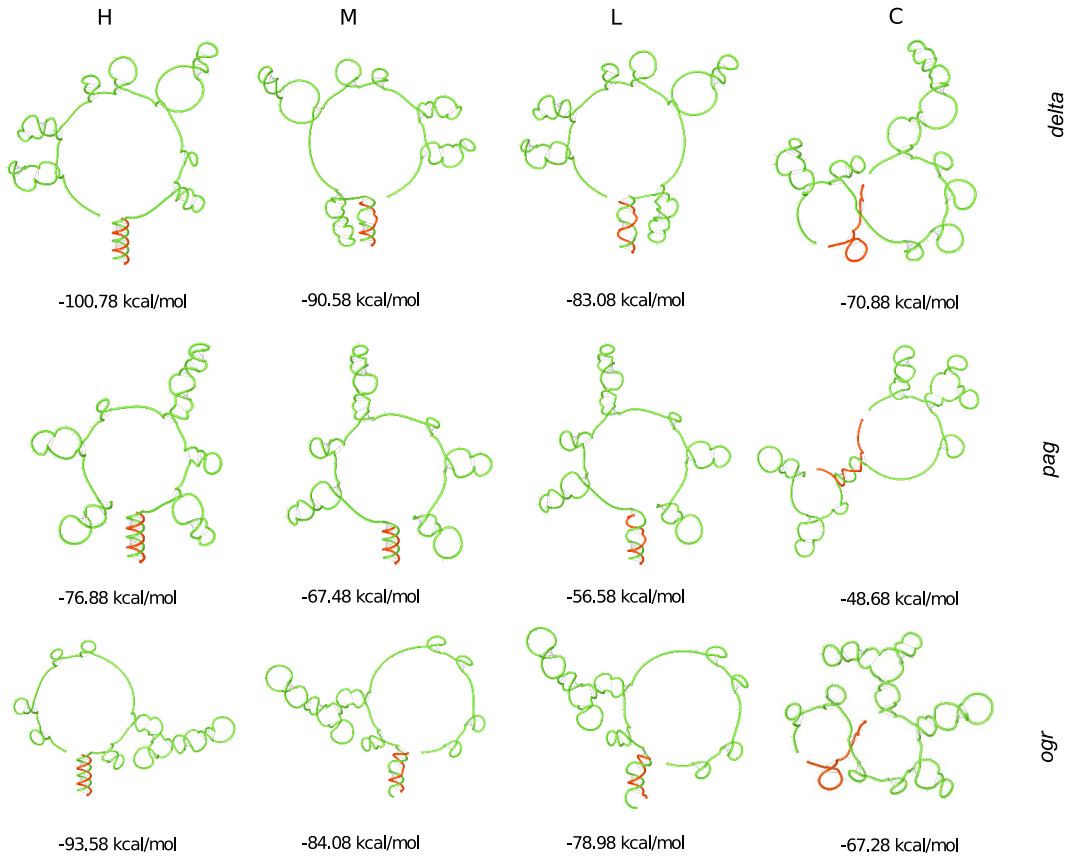


Figure 3.5: Ideal helical geometry of the TF mRNA-sRNA duplexes. As binding affinity and regulation strength of the sRNAs decrease, the stability of duplexes is lower. Translational regulation strength by sRNAs: H, high repression; M, medium repression; L, low repression; C, control.

3.2.3 Designing synthetic sRNAs for modulating genetic circuits

Synthetic sRNAs were designed to modularly inhibit TF activity (Table 3.1) on the P_F promoter, and consequently the GFP yield by targeting the TF genes. A dysfunctional sRNA (with a randomized target sequence) was used as a negative control. Furthermore, the duplex formation of the TF mRNAs with their correspondent set of sRNAs (H, M, L, C) was simulated. Predicted MFEs for all TF-sRNA duplexes showed an expected stable bonding between the TIR of miRNAs and the sRNAs (Figure 3.4). Additionally, the predicted helical geometry (Figure 3.5) illustrated graphically the binding affinity of the modulating sRNAs with the mRNAs, and together with the thermodynamic analysis indicated consistently that as the modulating level decreases, the sRNAs were less affine to the mRNA, and the predicted duplexes were also less stable (Figures 3.4, 3.5).

The sRNAs were expressed under control of T7 promoter with a different plasmid (pSB1A3 backbone) from the TF-GFP constructs. Both plasmids were expressed in tandem in *E. coli* (Figure 3.2).

Table 3.1: Synthetic sRNAs sequence details. The 79 nt sequence of the MicC scaffold (5′- TTTCTGTTGGGCCATTGCATTGCCACTGATTTTCCAACATATAAAAAGACAAGCCCGA ACAGTCGTCCGGGCTTTTTTT -3′) has to be concatenated to the target binding sequence for completing each sequence of the sRNAs, for simplicity only the target sequence is denoted.

Name	Target binding sequence (DNA, 5′- 3′)	Length (nt)	Binding energy (kcal/mol)	Details	Total length (nt)
Delta_H	ACGACAGAAAGG GCAGCGCATCAT	24	-36.70	High repression sRNA against Delta TF (strongest TF)	103
Delta_M	ACAGAAAGAGCA GCGCATCAT	21	-25.28	Medium repression sRNA against Delta TF (strongest TF)	100
Delta_L	CAGAAAGAGCAG CGAATCAT	20	-18.03	Low repression sRNA against Delta TF (strongest TF)	99
Pag_H	TTGGCATAACGG GCAGTGCATCAT	24	-32.81	High repression sRNA against Pag TF (medium TF)	103
Pag_M	CATAACGGGCAG TGCATCAT	20	-28.78	Medium repression sRNA against Pag TF (medium TF)	99
Pag_L	CATATCGGACAG TGCATCAT	20	-18.23	Low repression sRNA against Pag TF (medium TF)	99
Ogr_H	GCTGGCATAAAG GACAATGAAACAT	25	-33.49	High repression sRNA against Pag TF (medium TF)	105
Ogr_M	CTAGCATAAAGG ACAATGAAACAT	22	-23.84	Medium repression sRNA against Ogr TF (weakest TF)	104
Ogr_L	CTAGCATAAAGG ACAATAAAACAT	20	-18.94	Low repression sRNA against Ogr TF (weakest TF)	104
Control	CAGAACTGAATC TGCCGGTA	20	NA	Dysfunctional sRNA. Randomized sequence Pag_M.	99

3.2.4 Assembling and debugging of the constructs

TF genetic circuits plasmids went through 3 iterations of the 'design-build-test' cycle, these are referred as generation 1, 2, 3 (G1, G2, G3) plasmids (Figure 3.6A). Plasmids from G1, correspond to the circuits cloned into pGGA backbone. After tandem expression of G1 and sRNA plasmids, FACS experiments showed noise that was difficult to track (Figure 3.7A).

To investigate whether the noise was generated from the sRNA or the genetic circuit plasmid, a synthetic sRNA (high repression) was designed for targeting OmpR, an intrinsic TF from *E. coli*. The binding affinity between the sRNA (5′- ACGACAGAAAGGGCAGCGCATCAT -3′) with the *ompR* mRNA resulted in -36.70 kcal/mol. The functionality of the sRNA was tested by performing FACS after the tandem expression of the sRNA plasmid with wtGFP under control of the positively regulated *ompR* promoter. *ompR* promoter activates transcription upon binding of phosphorylated OmpR [199]. Clear signal peaks were detected without noise, however we did not observe a suppression by this sRNA, since the fluorescence signals from wtGFP and the sRNA-construct tandem were almost identical (Figure 3.7B).

Analysis of the TF genetic constructs was then performed. The TF mRNAs were screened for hairpins through the TIR that could prevent the successful binding of the synthetic sRNAs; such possibility was discarded (Figure 3.6B). According to the predictions, a secondary structure formation for *delta* mRNA is likely possible, however the equilibrium probability of it was fairly low to confirm persistence.

Furthermore, the TF expression was verified by RT-PCR (Figure 3.8) with two different melting temperatures for the primer pairs. Products with the expected size (100-200 bp) from all primer pairs for the three transcription factors were observed at both low and high melting temperatures. However, unexpected products for the uninduced sample of Ogr TF were also detected (Figure 3.8A).

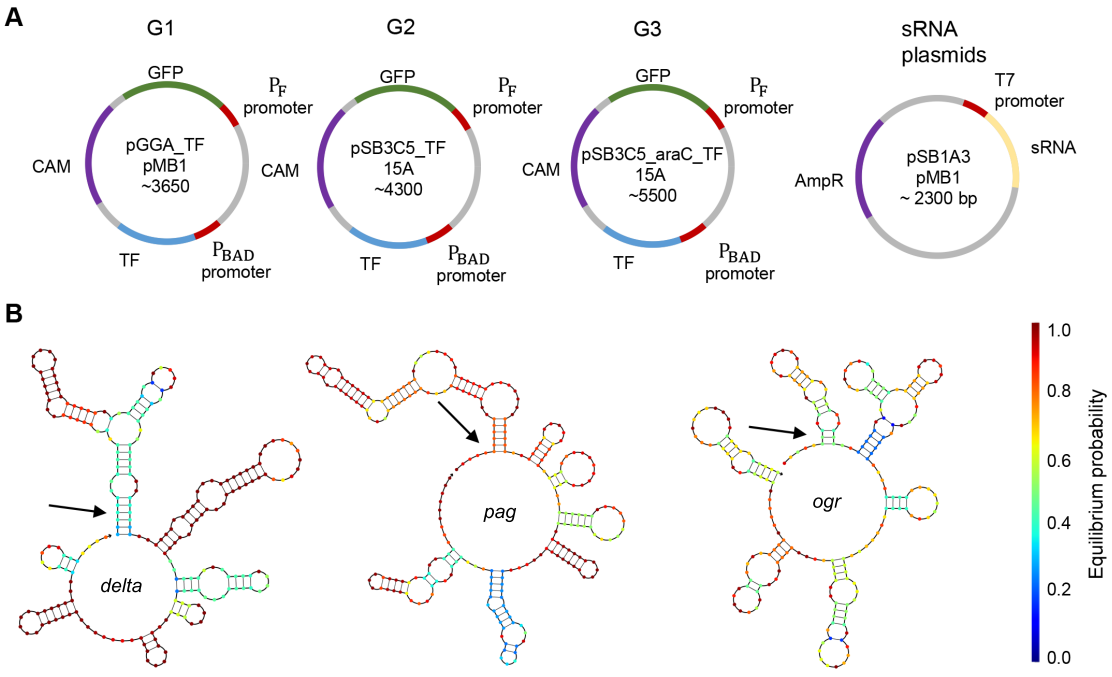


Figure 3.6: TF genetic circuit plasmids: overview. (A) Generations of TF genetic circuit plasmids. **(B)** Secondary structure formations (hairpins) discarded in the TIR (black arrow) of the TF mRNAs.

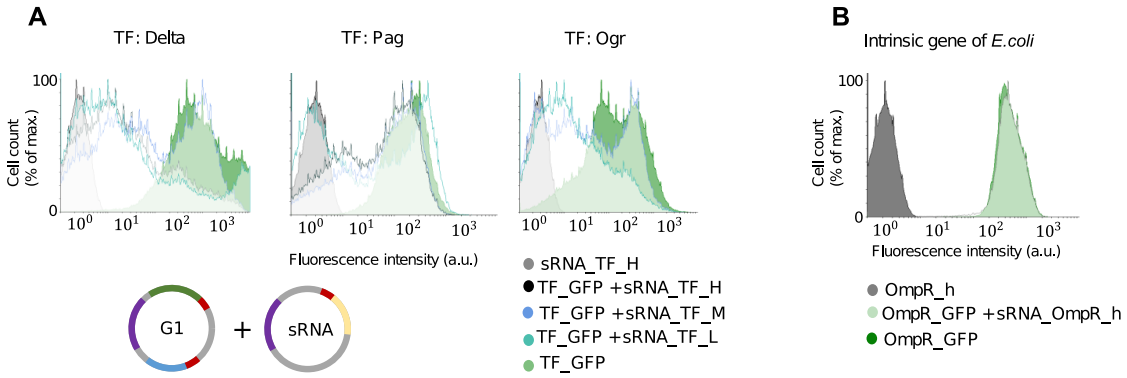


Figure 3.7: FACS analysis of tandem expression G1 TF genetic circuit plasmids with sRNA plasmids. **(A)** Noisy results from FACS analysis of tandem expression of sRNA plasmids and genetic circuits G1. **(B)** FACS analysis of tandem expression of a synthetic sRNA (highest repression threshold) targeting an intrinsic gene (*ompR*) from *E. coli*. The noise disappears, however the sRNA displays no well-defined function.

We used a second version of primers (Figure 3.8B), leading to the same result. Since bands were seen from both Ogr induced and uninduced samples with both versions of primers, their sequences were verified for homology against the *E. coli* BL21(DE3) genome by BLAST [200]. Some similarities between the primer sequences with parts of the genome were observed (Figure A.2). For the forward primer, 388 matches were found, and the most similar hit has a length of 13 nt (Figure A.2A). For the reverse primer, 1163 matches were found, the most similar alignment had 12 identical nucleotides (Figure A.2B). Thus, this was a possible explanation for the primers producing amplicons from uninduced samples. Nonetheless, both versions of the Ogr primers were also tested on *delta* cDNA samples (Figure 3.8B), producing no amplicons. Further experiments have to be performed to identify this problem, nevertheless the experimental results were as expected for all the other TF cDNA samples, it can be assumed that Ogr TF was being produced as expected as well.

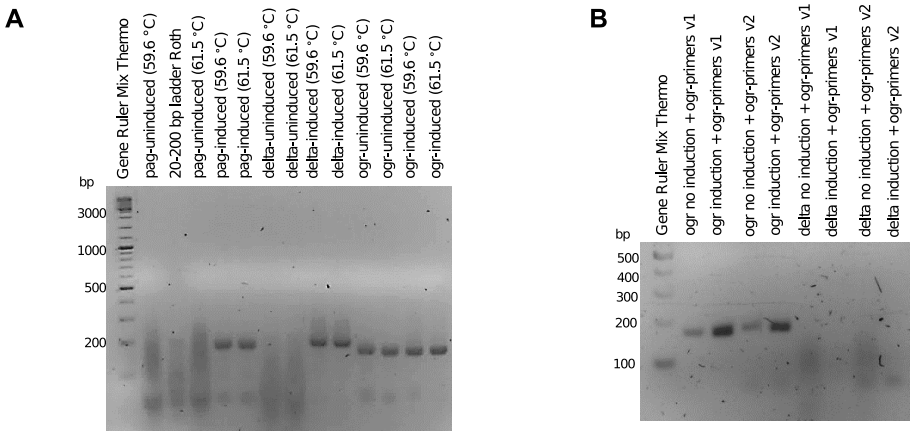


Figure 3.8: Debugging strategies for tandem expression optimization: TF transcript production verified by RT-PCR. Amplicon sizes were determined by DNA electrophoresis in a 3% Agarose gel. **(A)** Fragments of expected size for all of the TFs cDNA samples. **(B)** Second version of primers tested with *ogr* and *pag* cDNA samples.

Next step was to verify that indeed the sRNAs were being produced by the host. The sRNA production was successfully verified (Figure 3.9) by a tailored northern dot blot assay (Figure A.3B) after performing a customized RNA isolation (Figure A.3A). The optimal concentration (more than 7000 ng/µl) of total RNA from the samples was determined using fluorescent probes (Figure 3.9A) by testing different concentrations with 5S rRNA probes, since rRNA is one of the most abundant molecules in cells. The northern blot assay was performed for Ogr with sRNA_H (Figure 3.9B) under the assumption that if the verification of the production of the sRNAs was successful for the construct with the lowest yield, thus it should apply for the TFs with higher expression. The assay was performed with an uninduced sample and two induced samples with different concentrations (low and high). Methylene blue stained membrane was used as a control, the level of signal was in accordance with the amount of loaded concentration of the sample. The assay was performed with both 5S rRNA probes (as control) and sRNA probes. Two different buffers were used, Church's buffer gave the best results, as the fluorescence coming from both probes can be clearly detected. In the case of the 5S probes, the highest signal was from the sample with the highest concentration, and signals were in accordance to the total RNA concentration loaded. In the case of the sRNA probes, the only visible fluorescent signal is from the induced sample. For the assay performed with the Denhardt buffer, signal from the probes was not detected, as they seemed to be smeared away from the dots.

After discarding the possibility of failure of sRNA production, an explanation for the noisy results (Figure 3.7) is given by incompatibility between plasmid groups. Bacterial plasmids that share replication control or origin of replication (*ori*) are incompatible, since they compete for stable inheritance, therefore they cannot be used in tandem expression, leading to an unstable and

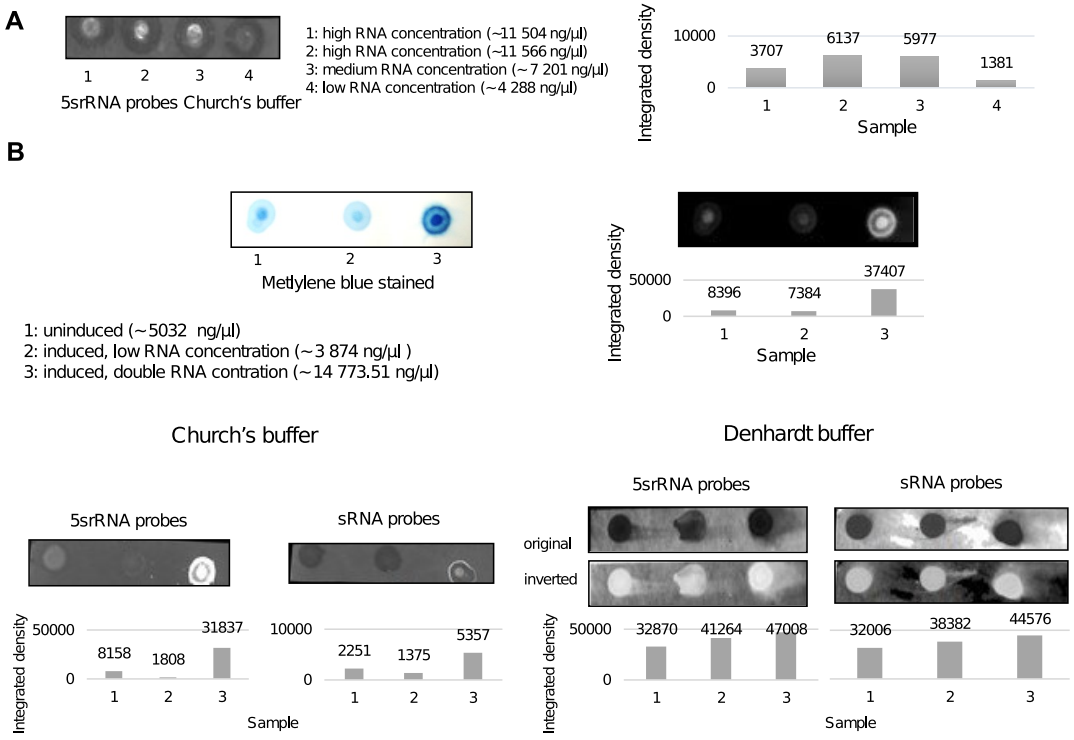


Figure 3.9: Debugging strategies for tandem expression optimization: Northern dot blot. (A) Testing different concentrations of total RNA with 5S rRNA probes. **(B)** Assay for identification of Ogr sRNA_H.

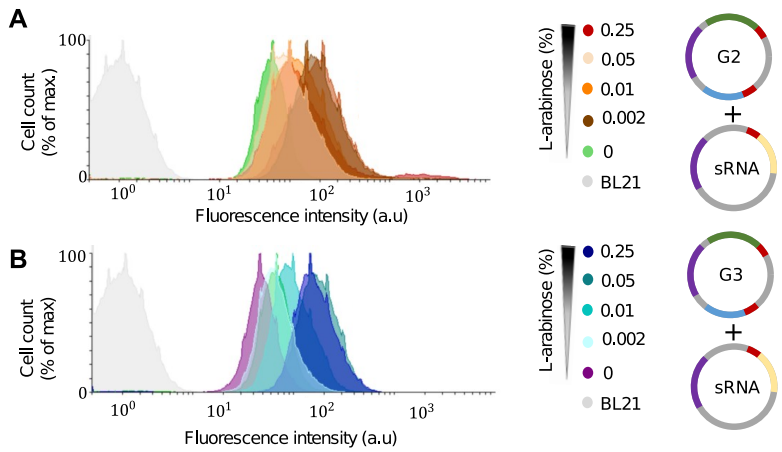


Figure 3.10: FACS: L-arabinose dosage dependence on TF-GFP cascading. (A) G2 plasmids and **(B)** G3 Ogr TF plasmids.

unpredictable expression [201, 202]. Both G1 plasmids and sRNA plasmids are high copy plasmids with a pUC19-derived pMB1 *ori*. Although, there is evidence that sometimes this matter is trivial for some constructs [203], for this case it seemed crucial. Thus, the backbone of the constructs was exchanged to a low-med copy 15A *ori* plasmid (pSB3C5). G2 plasmids were created by amplification of the TF genetic construct cassettes and cloning into pSB3C5 backbone, whose *ori* is compatible with the sRNA plasmids. After performing experiments, the TF constructs were still leaky (i.e. expression without induction), presenting high basal expression (Figure 3.10A).

Although there is an evidence that P_{BAD} promoter is strong enough (i.e. leaky transcription occurs at very low levels) [204], the addition of AraC gene exerts tight regulation on the promoter, since it represses the promoter [205]. This led to the construction of G3 plasmids. *AraC* was amplified by PCR and cloned into the G2 plasmids. Both generations of plasmids show that fluorescence intensity

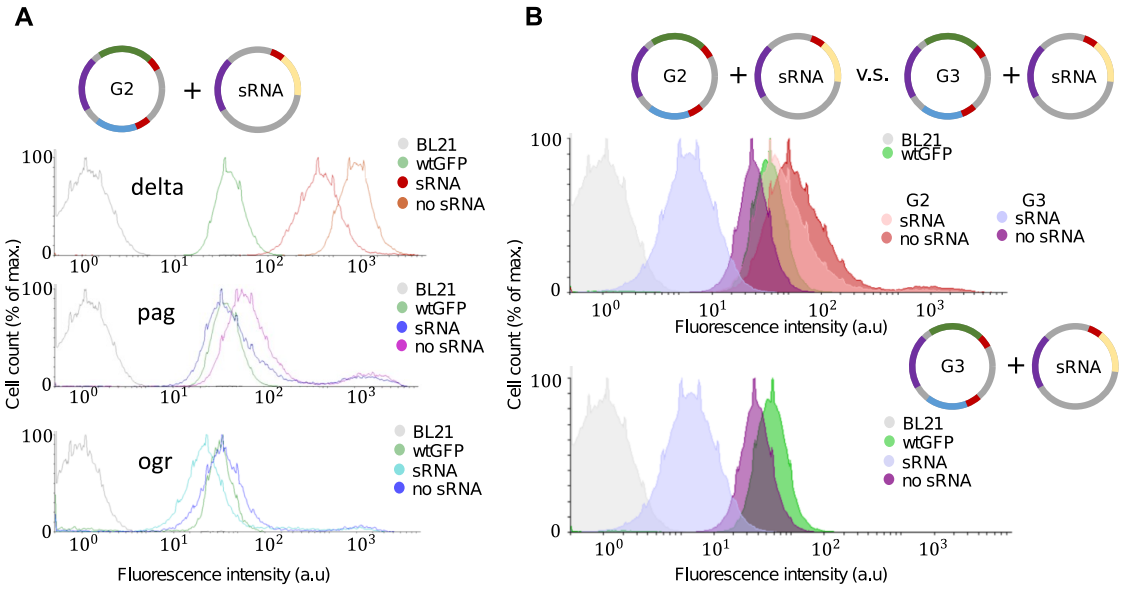


Figure 3.11: FACS verification of sRNAs (only H level) repression activity. (A) G2 TF plasmids. **(B)** G3 TF plasmids.

responds accordingly to L-arabinose dosages (Figure 3.10AB), results shown for ogr TF plasmids. Basal activity was still observed (Figure 3.10B). Also even after normalization to the host BL21 cells (Figure 3.3C), which lead to thinking that the P_F promoter controlling GFP expression, might be the problem. It has been previously shown that P_F promoter presents activity even without its activator [197], which is likely the reason for the leaky expression. In order to observe the real effect of the sRNAs, the basal activity signal was subtracted from the signals of the induced constructs (Figure 3.3C).

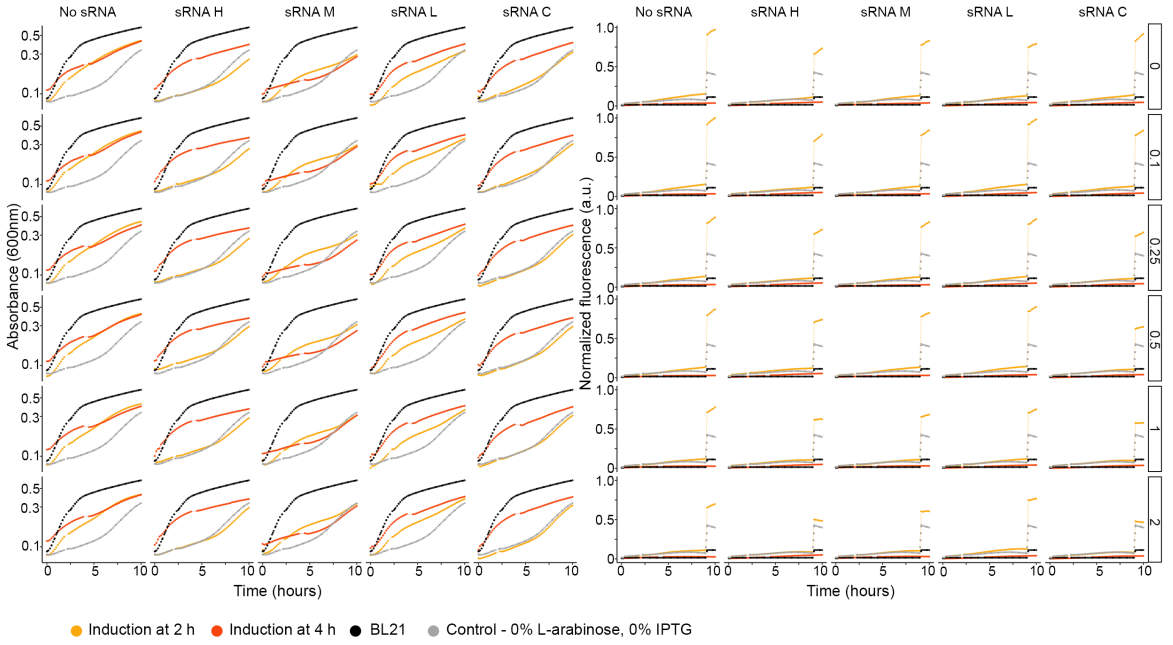


Figure 3.12: Real-time characterization of tandem expression of Delta sRNAs (IPTG at maximum) and TF-GFP construct at medium dose of L-arabinose (0.01%). Burden characterization and normalized mean fluorescence from three biological replicates.

3.2.5 Characterization of modulating sRNAs targeting TFs for gene circuits regulation

First, the repression effect of the engineered sRNAs through different sets of experiments was evaluated to elucidate the following questions: *i)* Are the sRNAs functional (i.e. was repression achieved)?, *ii)* Is the repression effect dose-dependent?, *iii)* Does the predicted repression mirrors empirical observations?, *iv)* Is the repression effect dependent on the induction timing?.

To address the first question, a FACS experiment was performed for determining the level of sRNAs (Figure 3.11), by working only with basal expression of the TF circuits (i.e. without L-arabinose induction), thus only the sRNA expression was triggered to verify if they were able to suppress at least the background fluorescence. Blank BL21(DE3) cells and BL21(DE3) with wtGFP were used as control, and samples were co-transformed with each of 3 TF circuits and its corresponding sRNA (H). Induction of sRNA expression was performed at maximum concentration of 2 mM IPTG. Notably, all sRNA showed successful repression under these conditions. For both G2 and G3, sRNAs consistently had shown repression activity.

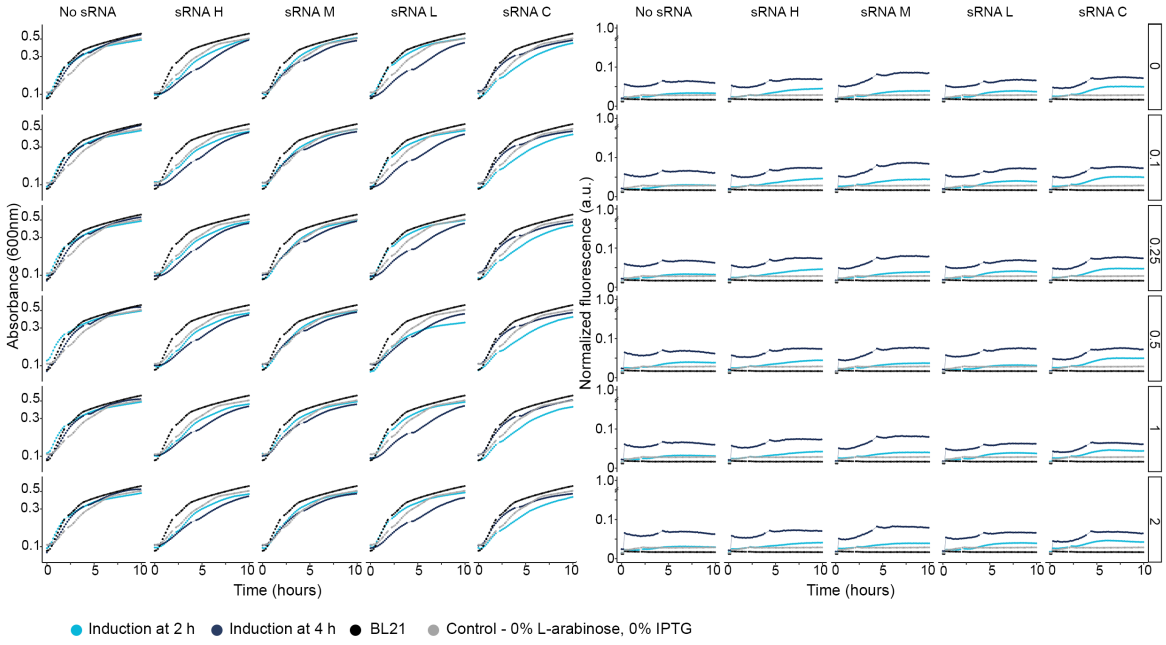


Figure 3.13: Real-time characterization of tandem expression of Pag sRNAs (IPTG at maximum) and TF-GFP construct at medium dose of L-arabinose (0.01%). Burden characterization and normalized mean fluorescence from three biological replicates.

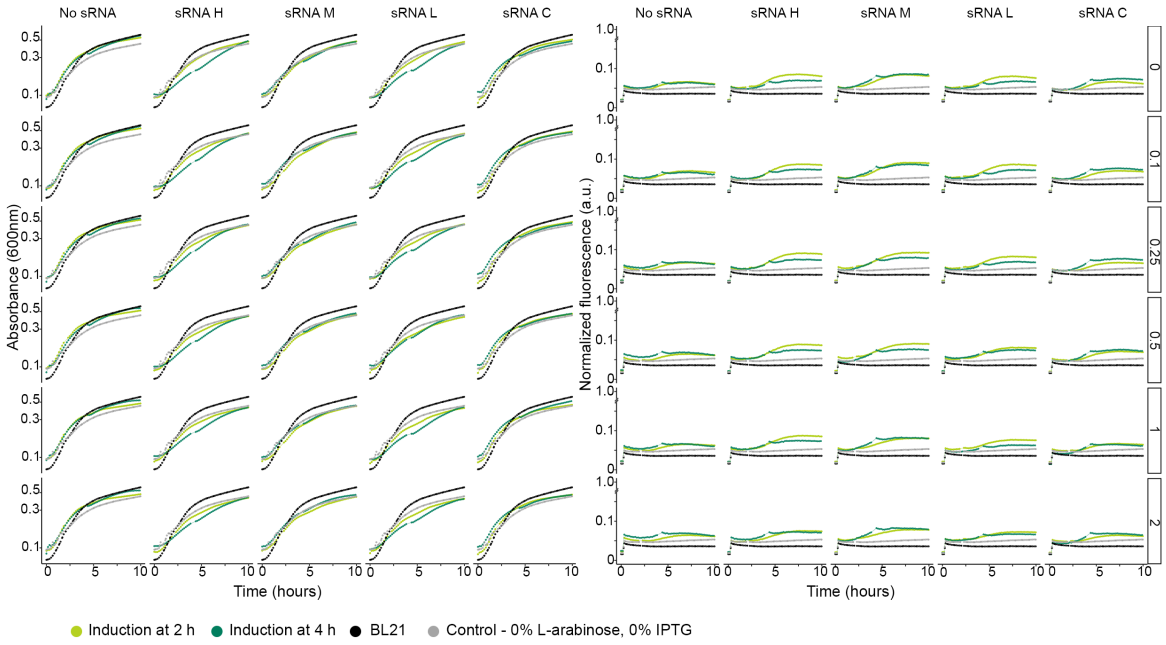


Figure 3.14: Real-time characterization of tandem expression of Ogr sRNAs (IPTG at maximum) and TF-GFP construct at medium dose of L-arabinose (0.01%). Burden characterization and normalized mean fluorescence from three biological replicates.

To answer the remaining questions, dose-dependence on both sRNA (IPTG) induction and TF-GFP cascade (L-arabinose) time-course assays including burden on the host evaluation for each tandem combinations of TF-reporter circuit and sRNA construct were performed. Two concentrations

of L-arabinose (0.01% and 0.25%) were tested while varying the IPTG concentration (0, 0.1, 0.25, 0.25, 1, 2 mM) at two different time points of IPTG induction (at two and four hours) for production of sRNA were evaluated (Figures 3.12-3.17).

Across all induction conditions tested, all constructs presented reduced growth compared to the control. This suggests that the synthetic constructs impose burden to the host (Figures 3.18-3.29), and Delta constructs were the most detrimental to the cells. Additionally, results from (Figure 3.16) suggest that there might have been something wrong with the cells and sRNA expression, thus the experiments needs to be repeated.

Regarding repression of the fluorescence signal, Delta TF was the most difficult to control (Figures 3.12 and 3.15) independently of the level and time of induction (L-arabinose); this seems appropriate since it is the construct with the highest GFP yield (Figure 3.3C). Overall the repression effect was observed better (Figures 3.24-3.26, 3.27-3.29) while inducing the TF-GFP cascade at a medium level (0.01% L-arabinose). An early induction of the sRNAs (two hours) seemed to give better results as well (Figures 3.18-3.20, 3.24-3.26).

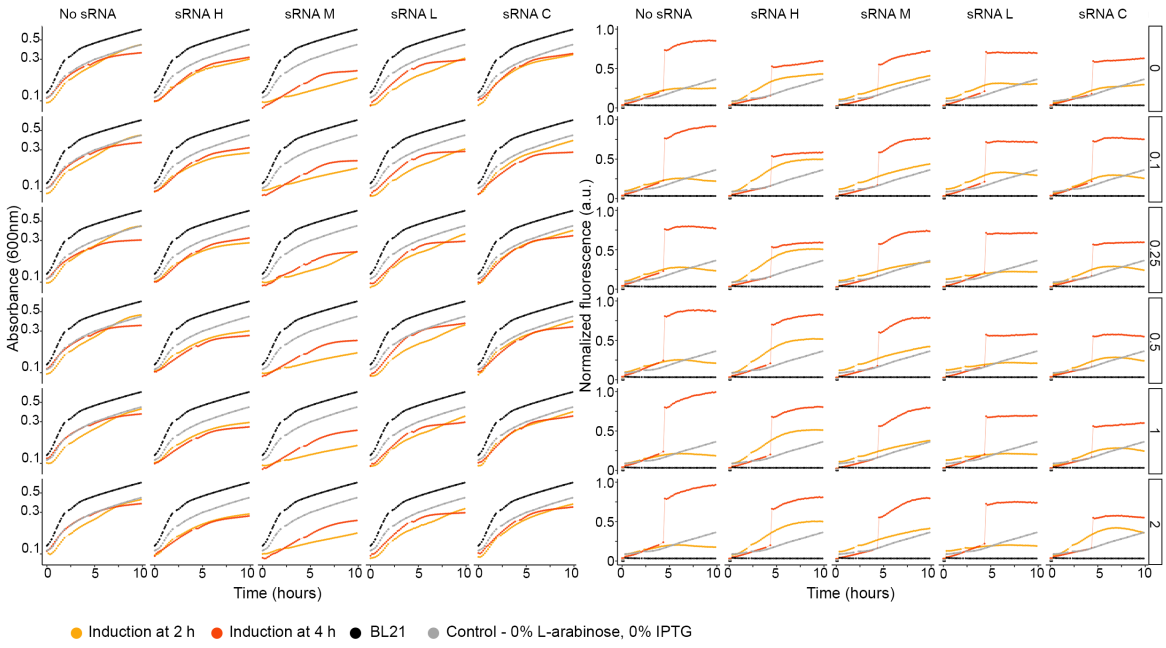


Figure 3.15: Real-time characterization of tandem expression of Delta sRNAs (IPTG at maximum) and TF-GFP construct at high dose of L-arabinose (0.25%). Burden characterization and normalized mean fluorescence from three biological replicates.

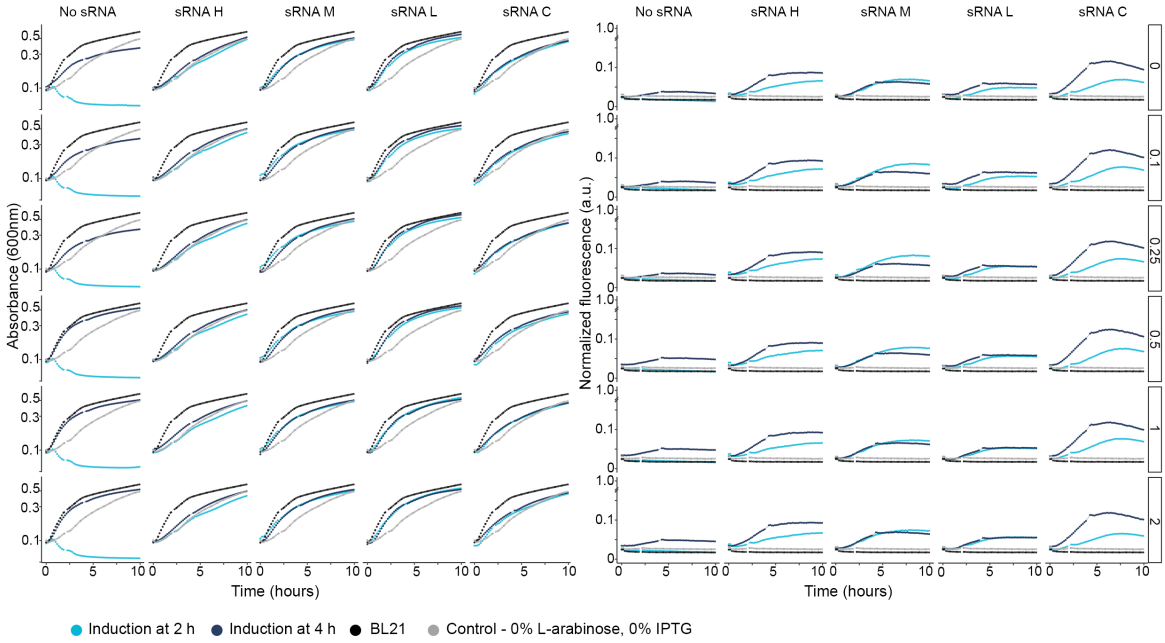


Figure 3.16: Real-time characterization of tandem expression of Pag sRNAs (IPTG at maximum) and TF-GFP construct at high dose of L-arabinose (0.25%). Burden characterization and normalized mean fluorescence from three biological replicates.

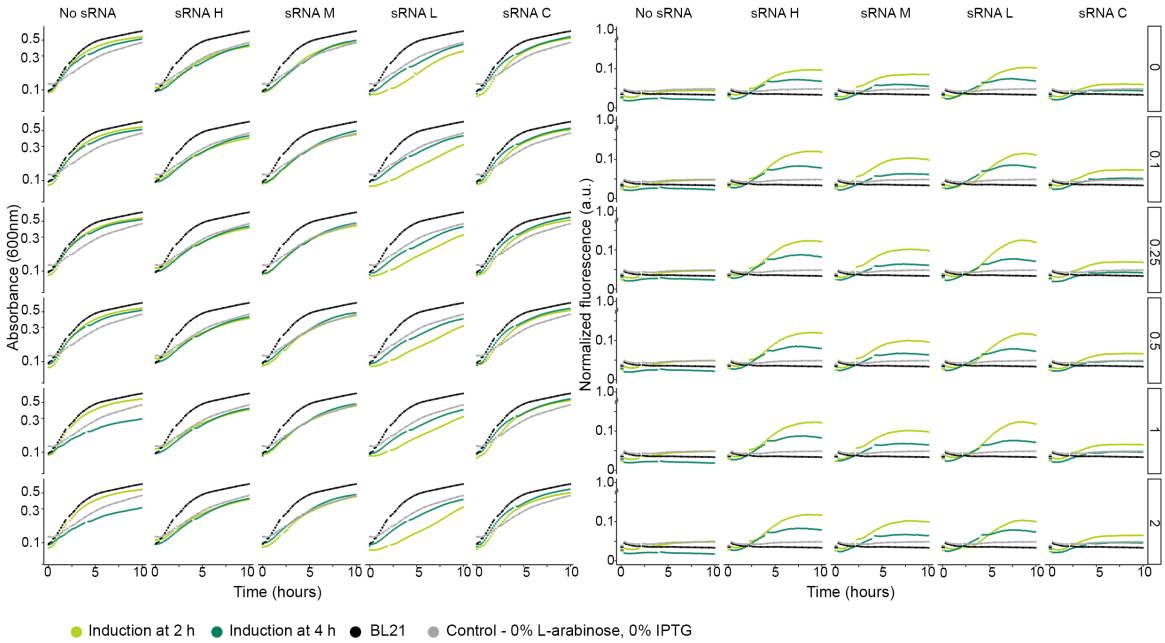


Figure 3.17: Real-time characterization of tandem expression of Ogr sRNAs (IPTG at maximum) and TF-GFP construct at medium dose of L-arabinose (0.25%). Burden characterization and normalized mean fluorescence from three biological replicates.

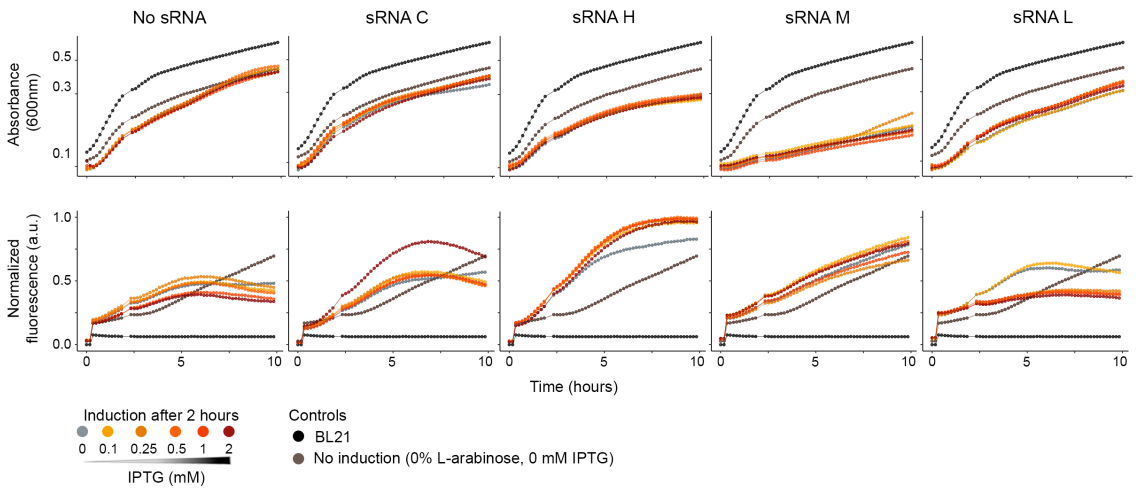


Figure 3.18: Characterization of Delta sRNAs (H, M, L, C), early induction, with high TF-GFP level. Dose dependence of IPTG (induction at two hours) on Delta sRNAs function with TF-GFP cascade at a high level (0.25% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

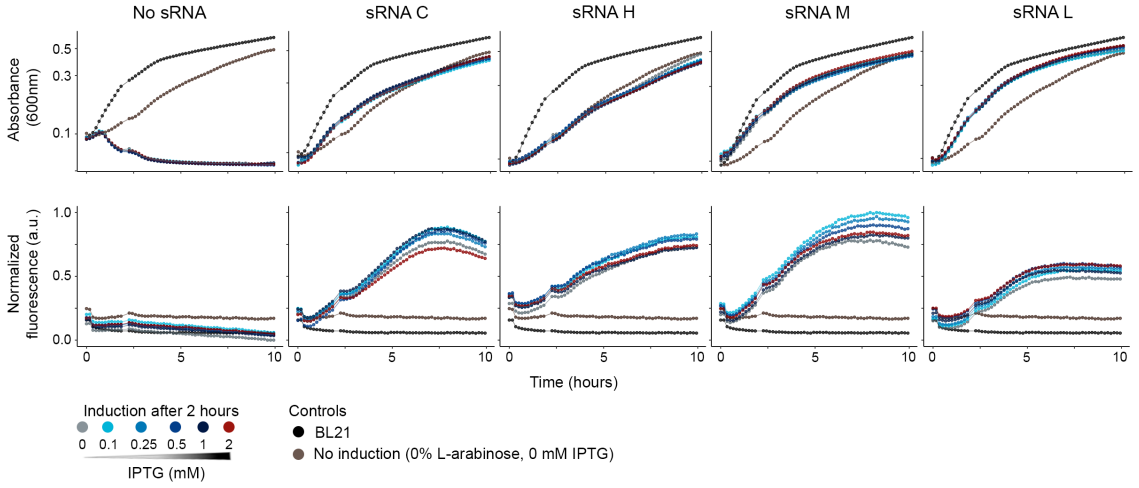


Figure 3.19: Characterization of Pag sRNAs (H, M, L, C), early induction, with high TF-GFP level. Dose dependence of IPTG (induction at two hours) on Pag function with TF-GFP cascade at a high level (0.25% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

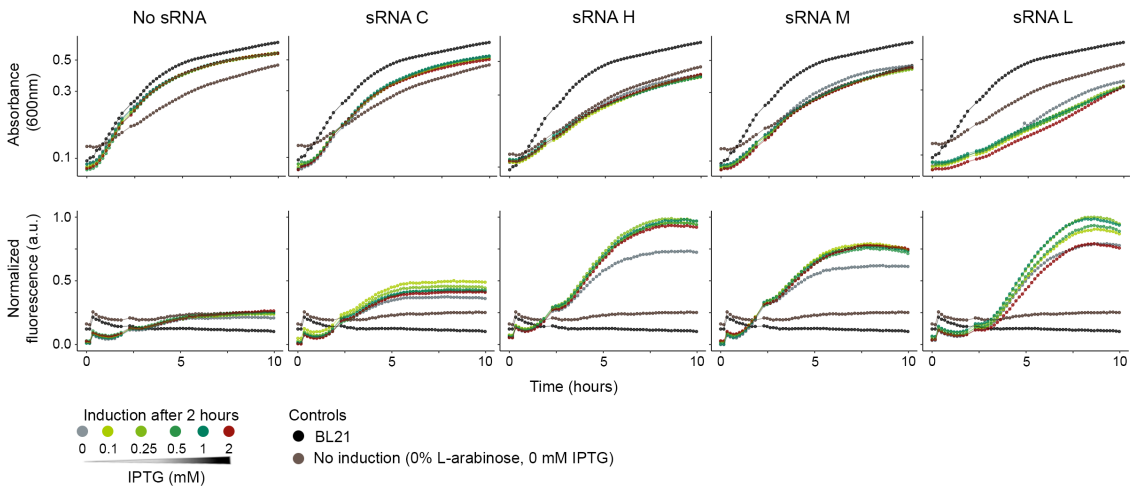


Figure 3.20: Characterization of Ogr sRNAs (H, M, L, C), early induction, with high TF-GFP level. Dose dependence of IPTG (induction at two hours) on Ogr function with TF-GFP cascade at a high level (0.25% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

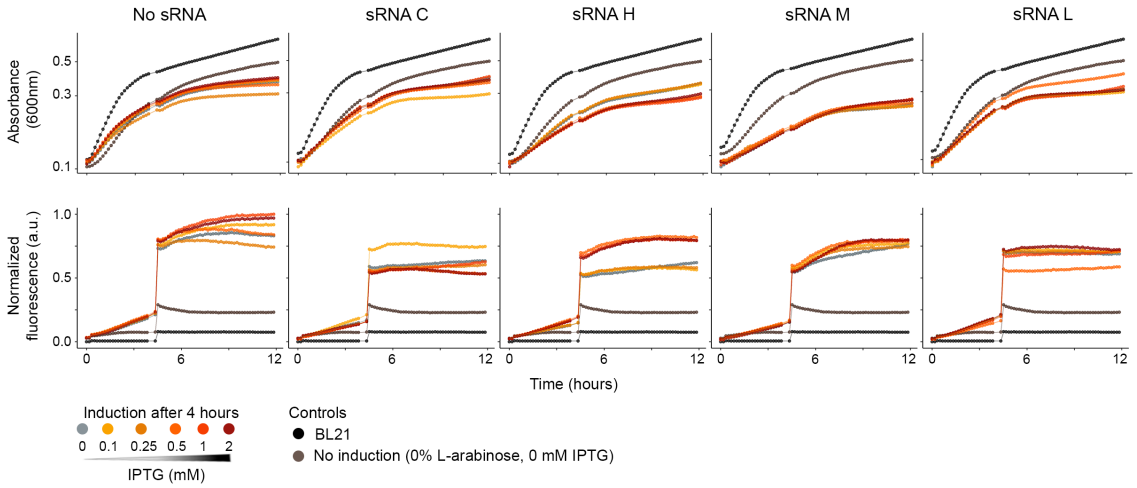


Figure 3.21: Characterization of Delta sRNAs (H, M, L, C), late induction, with high TF-GFP level. Dose dependence of IPTG (induction at four hours) on Delta sRNAs function with TF-GFP cascade at a high level (0.25% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

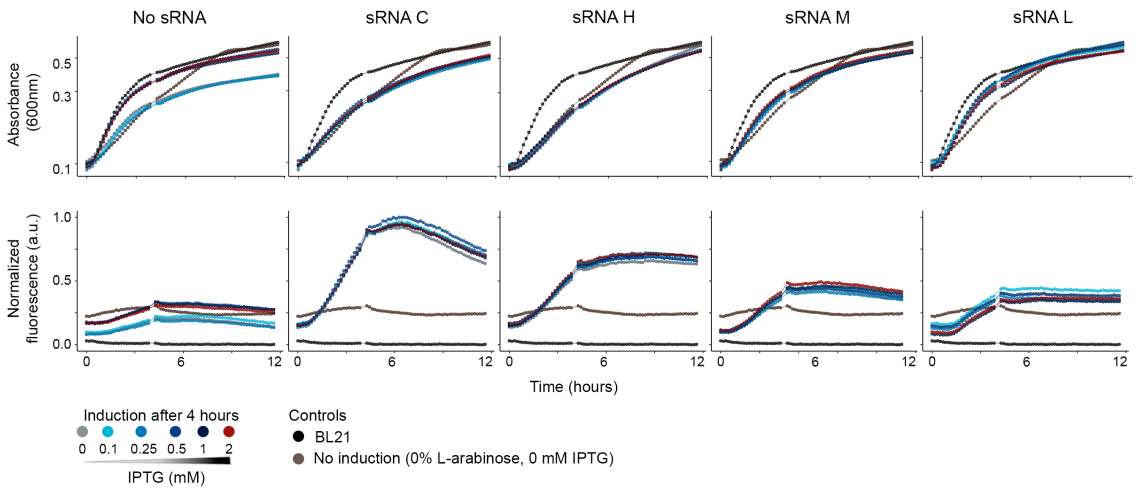


Figure 3.22: Characterization of Pag sRNAs (H, M, L, C), late induction, with high TF-GFP level. Dose dependence of IPTG (induction at four hours) on Pag sRNAs function with TF-GFP cascade at a high level (0.25% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

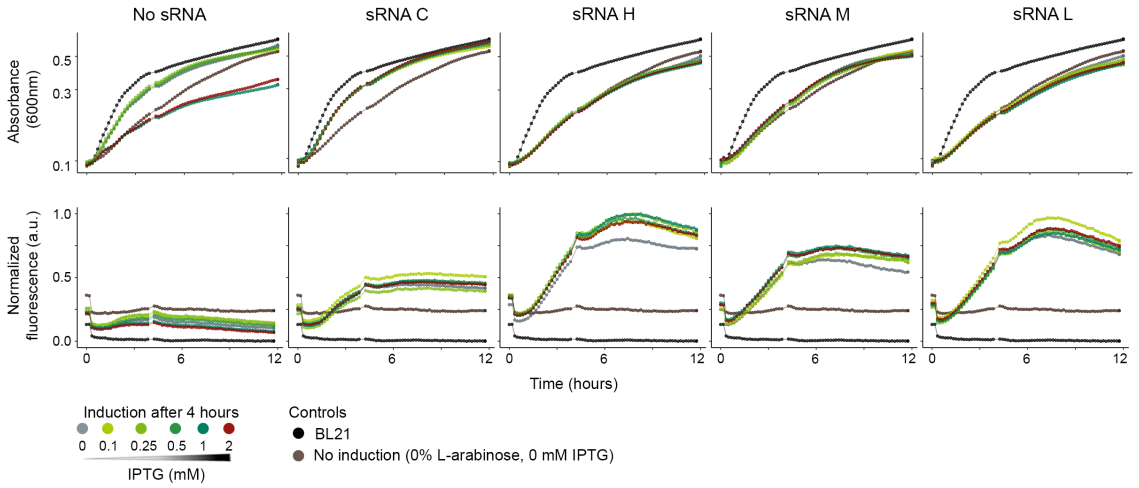


Figure 3.23: Characterization of Ogr sRNAs (H, M, L, C), late induction, with high TF-GFP level. Dose dependence of IPTG (induction at four hours) on Ogr sRNAs function with TF-GFP cascade at a high level (0.25% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

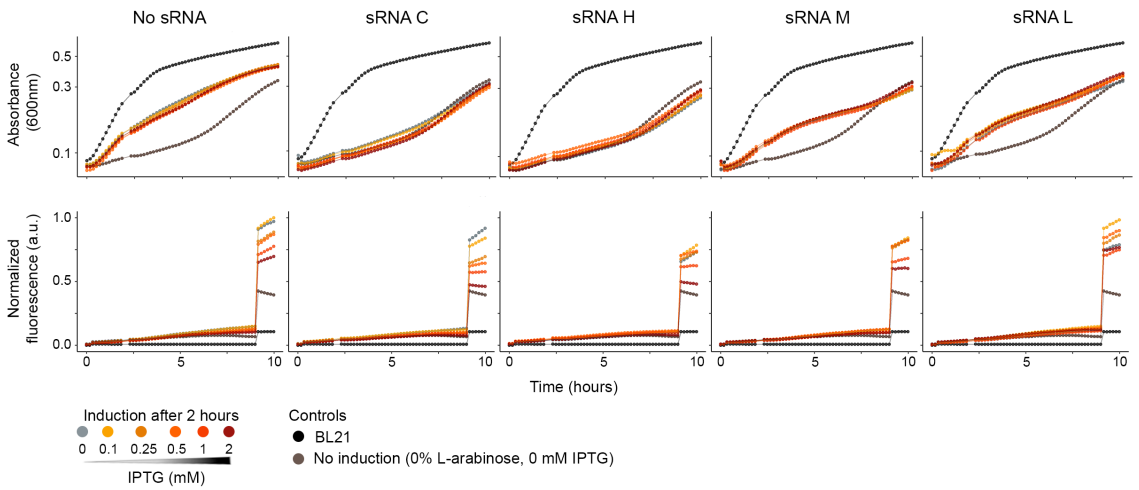


Figure 3.24: Characterization of Delta sRNAs (H, M, L, C), early induction, with medium TF-GFP level. Dose dependence of IPTG (induction at two hours) on Delta sRNAs function with TF-GFP cascade at a medium level (0.01% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

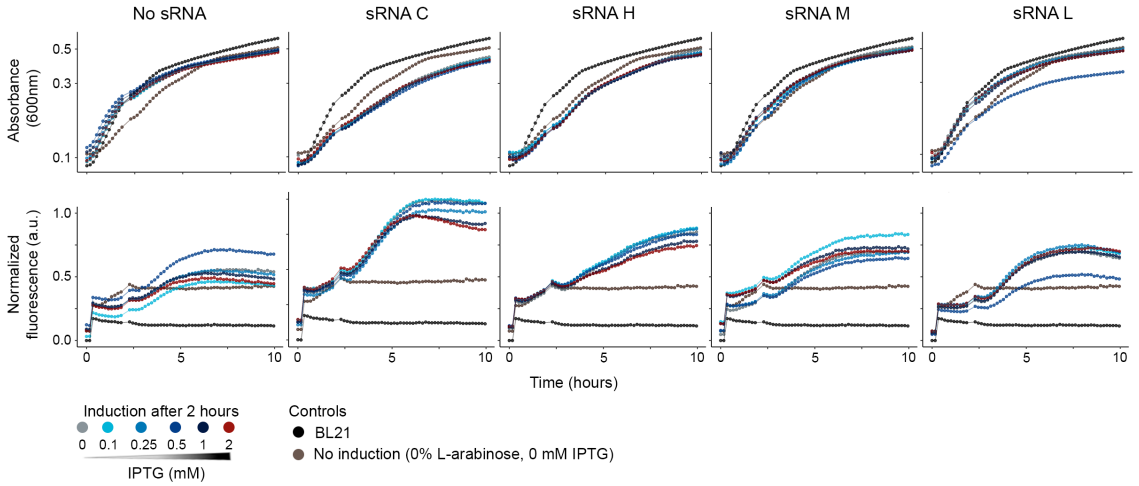


Figure 3.25: Characterization of Pag sRNAs (H, M, L, C), early induction, with medium TF-GFP level. Dose dependence of IPTG (induction at two hours) on Pag sRNAs function with TF-GFP cascade at a medium level (0.01% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

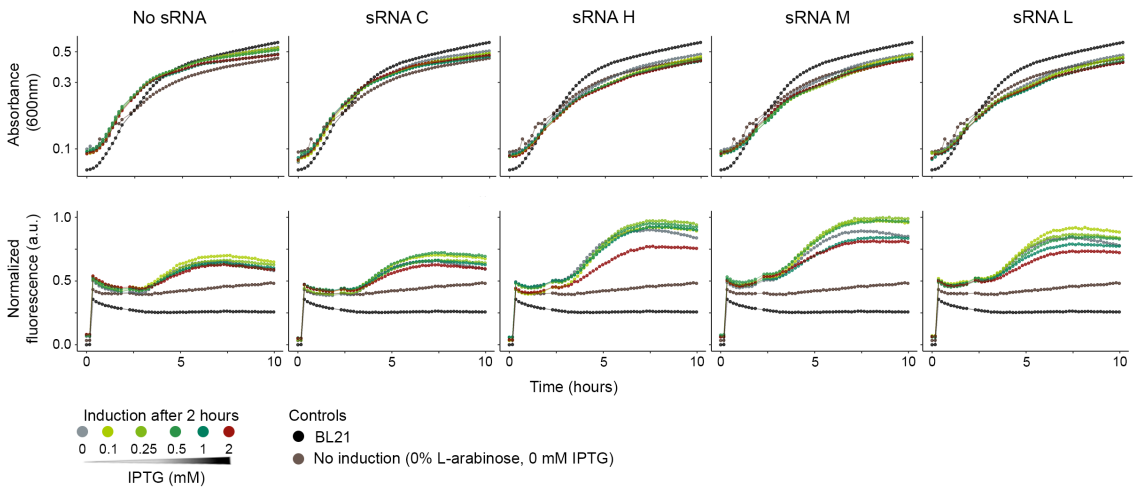


Figure 3.26: Characterization of Ogr sRNAs (H, M, L, C), early induction, with medium TF-GFP level. Dose dependence of IPTG (induction at two hours) on Ogr sRNAs function with TF-GFP cascade at a medium level (0.01% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

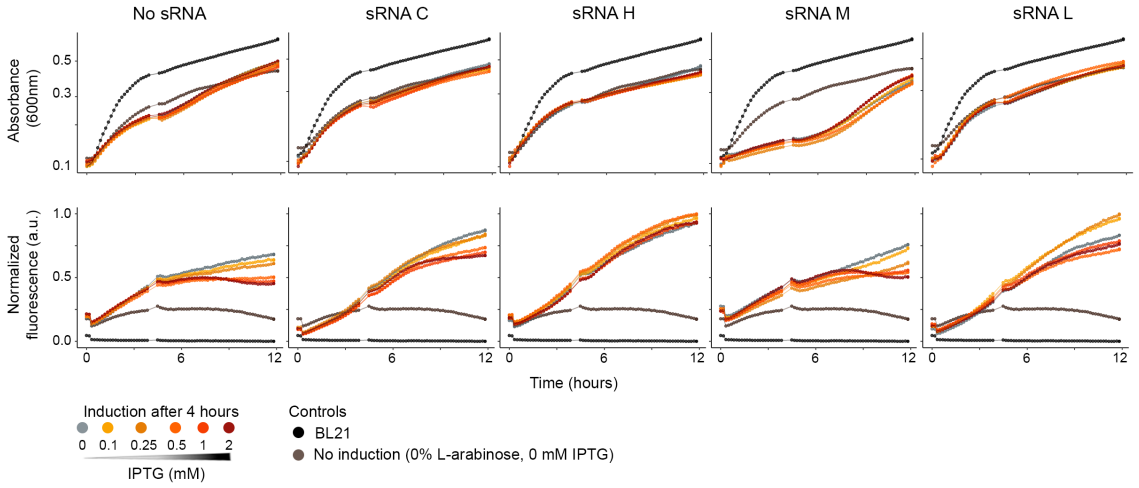


Figure 3.27: Characterization of Delta sRNAs (H, M, L, C), late induction, with medium TF-GFP level. Dose dependence of IPTG (induction at four hours) on Delta sRNAs function with TF-GFP cascade at a medium level (0.01% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

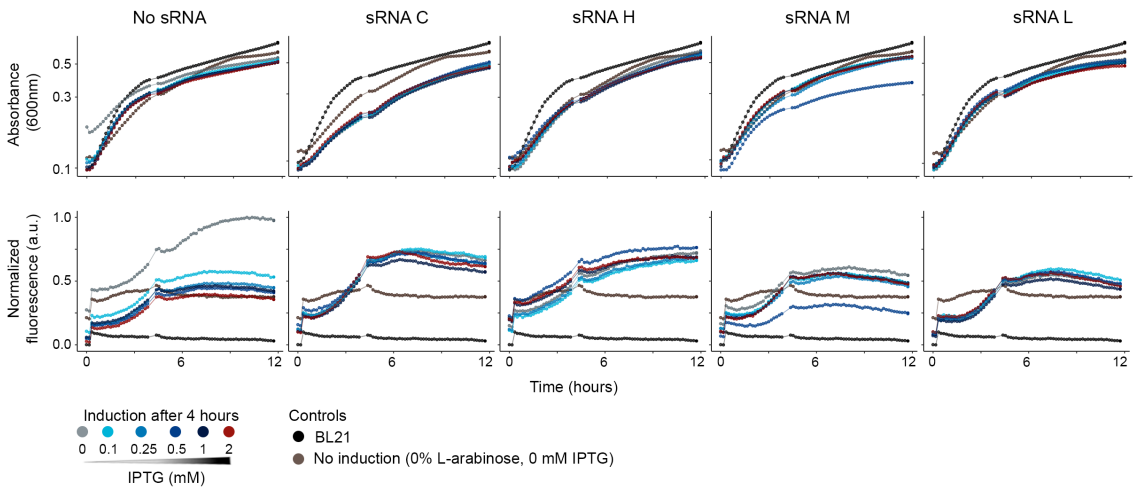


Figure 3.28: Characterization of Pag sRNAs (H, M, L, C), late induction, with medium TF-GFP level. Dose dependence of IPTG (induction at four hours) on Pag sRNAs function with TF-GFP cascade at a medium level (0.01% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

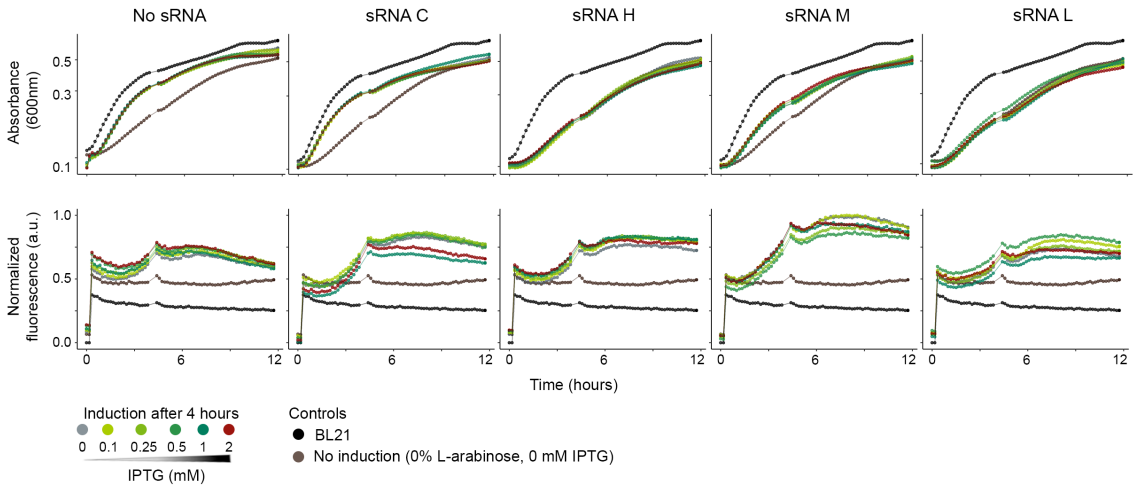


Figure 3.29: Characterization of Ogr sRNAs (H, M, L, C), late induction, with medium TF-GFP level. Dose dependence of IPTG (induction at four hours) on Ogr sRNAs function with TF-GFP cascade at a medium level (0.01% L-arabinose): H, high repression; M, medium repression; L, low repression; C, control.

3.2.6 Tuning gene expression with modulating sRNAs

The reaction network of the sRNA repression was modeled (Figure 3.30A) and simulated with SimBiology [206] (Figure 3.30B). Simulation results suggested a substantial however not exaggerated decrease in GFP concentration. This was in accordance to the characterization experiments, since it has to be considered that only some mRNAs are susceptible to the sRNA targeting and further degradation whereas the remaining are continued to be translated into proteins. Additionally, GFP is remarkably stable and resistant to degradation [207].

FACS analysis (Figure 3.31A) suggested that with the TF-GFP constructs at basal fluorescence the sRNAs worked as expected. The repression was observed also at lower concentrations of IPTG, and a stepped behavior shifting towards less fluorescence as the dose of IPTG induction increased confirmed that the function of the sRNAs is dose-dependent on IPTG.

Furthermore, sRNAs also repressed at lower concentrations of IPTG even when a higher production of GFP was induced (L-arabinose 0.25%). A gradual decrease of the fluorescence signal was observed as the IPTG concentration increased (Figure 3.31B). No shift from the signal was observed from the dysfunctional sRNA (C, orange) used as control, suggesting it is indeed dysfunctional, allowing to safely discard the possibility that the fluorescence decrease was due to something different than the sRNAs intervention. Together these results suggests that synthetic sRNAs are able to successfully repress the TF genes, and the effect is dose dependent on the transcriptional inducer (IPTG).

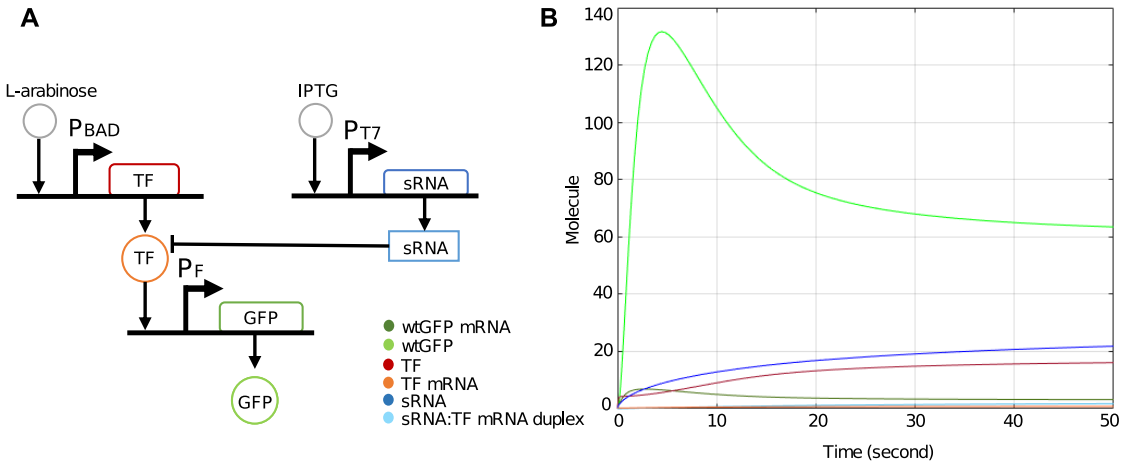


Figure 3.30: Simulation of the reaction network of the TF-GFP constructs modulated by sRNAs. (A) Reaction network. **(B)** Simulation results.

3.2.7 Controlling a characterized nonsense suppressor tRNA system with modulating sRNAs

Next step was to elucidate whether synthetic sRNAs can be used to switch (i.e. ON/OFF) existent functional molecules apart from genetic circuits. In order to do this, an artificial sRNA (Figure 3.32A) was designed to target a therapeutic tRNA [33]. The calculated binding affinity between the sRNA and the tRNA was -57.74 kcal/mol.

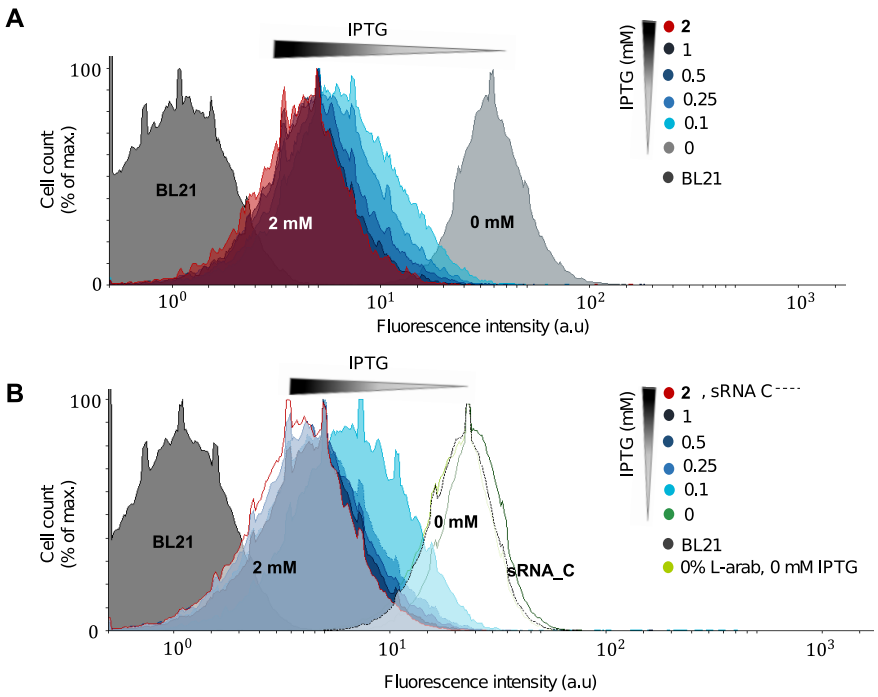


Figure 3.31: FACS analysis of dose dependence of IPTG on sRNAs (H) function for Pag TF construct. (A) TF-GFP cascade at basal level (0% L-arabinose). **(B)** TF-GFP cascade induced at medium level (0.05% L-arabinose).

A genetic circuit (Figure 3.32B) was designed for testing the function of the sRNA against the tRNA. It is conceived for it to be expressed *in vivo* in tandem with the therapeutic tRNA [33]. The sRNAs expression to be under control of a constitutive promoter (P_{tet}), with the option to be negatively regulated (i.e. turned OFF) by the addition of IPTG (since it activates the expression of the repressor of P_{tet}). The reporter construct (GFP gene with a mutation to be readthrough by the tRNA) of the therapeutic tRNAs [33] was included as a fragment of the circuit.

Primers (Table A.3) for cloning with Golden Gate assembly method were designed and used to amplify the fragments by PCR (Figure 3.33A). The original plasmid backbone contains a red fluorescence protein, the construct was cloned within this region, therefore white colonies indicated possible successful assemblies (Figure 3.33B). After performing a colony PCR (Figure 3.33C), we detected two plasmids had the expected size, and they were sent for sequencing. Alignments of the sequencing results of one of the plasmids (Figure 3.33C) showed that the sequence was highly similar, however the region corresponding to the sRNA had abundant mismatches. The assembly was found unsuccessful after performing a plasmid fingerprinting (Figure 3.33E) with two restriction enzymes. Since the obtained restricted fragments (Figure 3.33E, right) were not from the expected size (Figure 3.33E, left). Thus, the assembly of the construct is still in developmental phase.

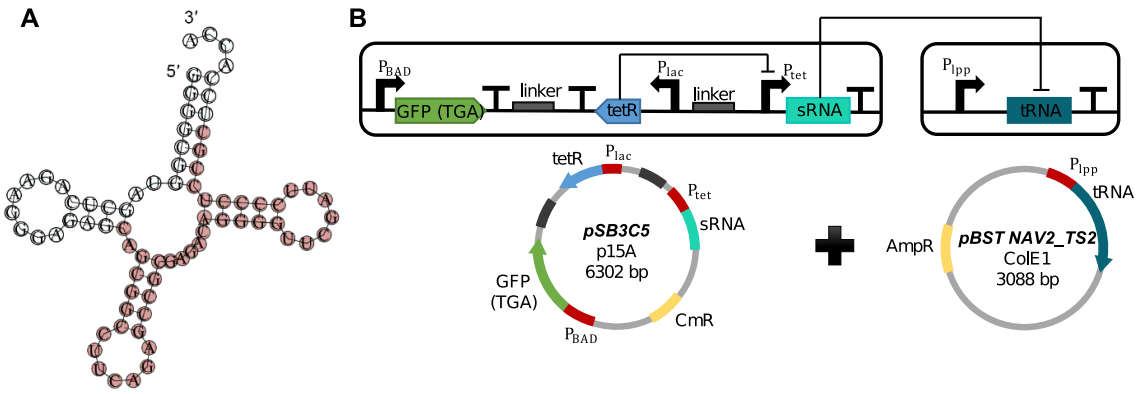


Figure 3.32: Modulating a nonsense suppressor tRNA with synthetic sRNAs. (A) Design of the sRNA (highlighted) targeting a nonsense suppressor tRNA. Sequence of the tRNA from [33]. **(B)** Design of the sRNA_t1A3T2 construct to be expressed in tandem with the tRNA plasmid from [33].

3.3 Conclusion

Artificial sRNAs can be rationally and modularly designed, and can be applied in metabolic engineering, genetic circuits and RNA-based devices, to name a few. A persistent challenge in synthetic biology is the predictable and controllable tuning of the genetic circuits to regulate gene expression as expected, including modulation of existing circuits or constructs without the need of redesign or rebuild. Here, a rational design of *de novo* sRNAs to dynamically modulate gene expression in different ranges (from high, medium to low repression) in *Escherichia coli* was proposed. Synthetic sRNAs that modulate within different ranges (low, med, high) gene expression were designed and implemented. To show that the modulation activity of the sRNAs is independent of the construct, multiple genetic circuits were designed, in which the variable element is a transcription factor inducing the transcription of a reporter protein. sRNAs were designed to target these multiple exogenous transcription factors featuring varying activity strengths (weak, moderate, strong) on a common promoter. Moreover, the approach was applied to downregulate (or turn off) an RNA-based therapeutic molecule (nonsense suppressor tRNA) [33], without altering its original design. The characterization and analysis of the sRNA library provides a rationale for achieving independent targeted modulation of gene expression, only by incorporating synthetic sRNAs, and without altering existing genetic circuit or constructs.

In summary, synthetic sRNAs were designed to modulate gene expression at different levels. Genetic circuits were designed and constructed to test the functionality of the sRNAs in *E. coli*. The design of *de novo* synthetic sRNAs was based on initial *in silico* design and verification of thermodynamic stability and viability of expression in the host cells (discarding possible off targets on intrinsic genes), paired with experimental fine-tuning of the system elements (i.e. tandem expression optimization, see Section 3.2.4). Furthermore, traditional transcriptional (promoters) together with posttranscriptional (synthetic sRNAs) regulators were applied to genetic circuits to dynamically modulate the cascade of gene expression and alter the yield of the reporter protein. This study relies on the intrinsic Hfq levels from the host *E. coli*. However, it is possible that the synthetic sRNAs functionality could be optimized by plasmid-borne overexpression of Hfq.

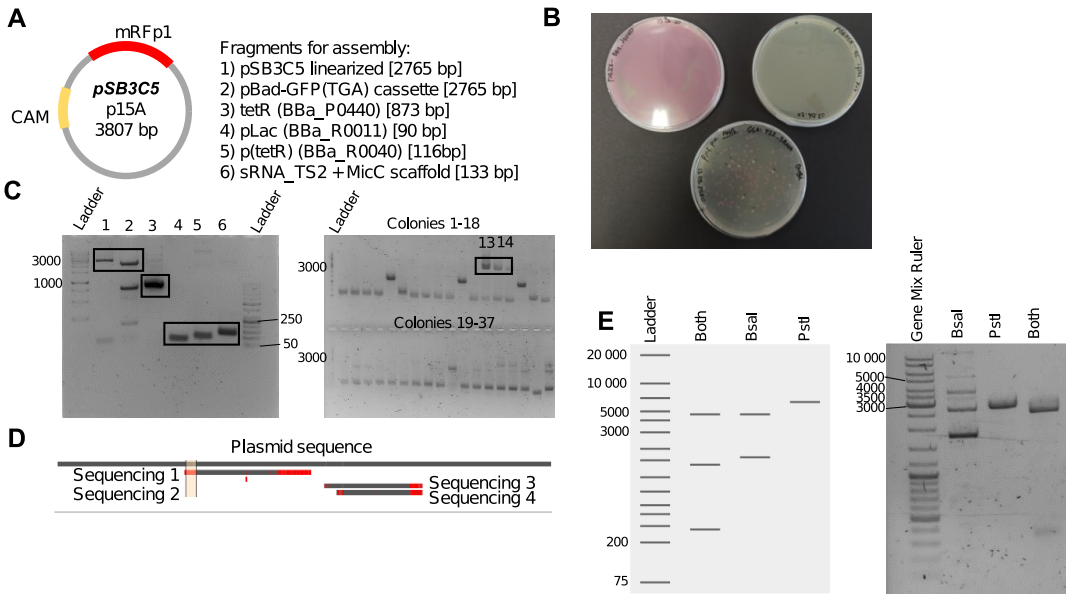


Figure 3.33: Strategy for assembling the sRNA_t1A3T2 construct. (A) Backbone vector contains a red fluorescent protein on the cloning region. **(B)** White screening of positive colonies. **(C)** Colony PCR showed two possible colonies containing the assembled plasmid. **(D)** The sRNA sequence was not identified by sequencing. **(E)** Plasmid fingerprinting with restriction enzymes, expected results (left) compared to experimental results (right).

Overall, synthetic sRNAs are capable of regulating both tailored genetic circuits and existent ones (when added as extrinsic components). The possibility of targeting functional molecules was explored, particularly TFs and an RNA-based therapeutic molecule (nonsense suppressor tRNA), with artificial sRNAs to modulate their effect on the corresponding target molecules, which reflected in the expression of the reporter protein of the constructs. TFs are often the main effectors of genetic circuits and are instrumental in multilayered RNA-based circuits [18]. Thus, to dynamically regulate the expression of these molecules, more advanced molecular devices can be further designed for specialized purposes (See Chapter 5).

3.4 Materials and methods

3.4.1 Synthetic sRNAs

Sequences were designed using in-house written scripts in Python. Binding energy calculations and 2D structure schemes were produced using RNAcofold *RNAcofold* from the ViennaRNA package [86] Python bindings. The sRNA sequence was screened for off-targets against *Escherichia coli* BL21(DE3) (Tax-id:866768) and str. K-12 substr. MG1655 (NC_000913) genomes, using RNAPredator[94]; the first ten (ranked by base pairing affinity) interactions were selected for post-processing, the sRNA was selected there were no potential interactions found. Afterwards, sRNAs sequences (DNA) were further screened with the web tool NEBcutter V2.0 [208] to discard undesired restriction enzyme sites that could interfere with the cloning procedure.

3.4.2 Thermodynamic and secondary structure analysis

Minimum Free Energy (MFE) of the TF mRNA-sRNA duplexes secondary structures, helical geometry, as well as their probability of formation were calculated by means of NUPACK [92]. Parameters for the simulations as follows: RNA energy parameter from Serra and Turner (1995); temperature was set to 37°C in correspondence to the optimum for *E. coli in vivo* experiments; values for Na^{+} = 1.0 M, and Mg^{++}= 0.0 M (set to default).

3.4.3 Simulation of the reaction pathway of gene regulation by sRNAs

Reaction pathway of the general system was modeled and simulated with SimBiology module from MATLAB [206]. Parameters and reaction rates were taken from the base model for gene expression and further modified to fit all the reactions of the system. The repression by sRNAs on the TF gene was modeled as a receptor-ligand degradation reaction. Details on the simulation can be found in Figures A.4 and A.5.

3.4.4 Synthesis of components and assembly

For each TF, a genetic circuit plasmid plus three sRNA plasmids, one for each level of repression of the sRNA (sRNA_H, sRNA_M, sRNA_L) were assembled. TF genetic circuit and sRNA plasmids were constructed using PCR and Golden Gate assembly (GAA) using BsmBI Type IIS Enzyme (New England Biolabs) in strain DH5-α. Afterwards, assemblies were confirmed by sequencing (Eurofins) (see Figure A.1). Cloning of all sRNA plasmids were assembled with *Bsa1_micC_Suf_f* primer and a reverse primer with the specific sRNA sequence. New customized plasmids for sRNA expression can be constructed by using *Bsa1_micC_Suf_f* primer and a new reverse primer containing Eco31I recognition site at 5′ end, immediately followed by the derived sRNA sequence (use reverse complement).

3.4.5 Real-time characterization of genetic circuits and synthetic sRNAs

Cell growth was measured with absorbance at OD_{600nm}. GFP expression was measured in bulk (480 nm excitation/530 nm emission). Both measurements were performed in real-time every 10 minutes over a period of 8, 12 or 14 hours (depending on induction time), with black 96-well plates with clear bottom (Corning), using a Spark (Tecan) Microplate reader. All plasmids were expressed in BL21(DE3) cells at 37°C in LB medium (Thermo Fisher Scientific) supplemented with ampicillin (100 μg/mL) and/or chloramphenicol (34 μg/mL) as selection marker(s). Expression of TF circuits was induced with 0.02, 0.05 or 0.25% (v/v) L-arabinose, and sRNAs induced after 2 or 4 hrs with 0.1, 0.25, 0.25, 1, 2 mM IPTG. To determine basal fluorescence from the TF genetic circuits, tandem transformation was performed, however only TF circuits expressed (induced with only L-arabinose) and used as a control. Furthermore, BL21(DE3) was grown to determine background fluorescence from the host, as well as normal growth in absence of the synthetic constructs, and used as control.

Min-Max normalization was applied to the data from all sRNA characterization experiments for comparison and plotted using in-house R written scripts.

3.4.6 Flow cytometry

Flow cytometry was performed on FACS Calibur with CellQuest™ Pro Software Version 6.0 (Becton Dickinson). The measurements of GFP fluorescence were taken for $\sim$ 100,000 events with the following settings: FSC = E01, SSC = 400, FL1 = 736, in log scale and threshold FSC = 52. Histograms were plotted with Flowing Software 2.5.1.

3.4.7 RNA isolation

A tailored protocol (Figure A.3A) was used and optimized to obtain a concentration of total RNA in the kilo order, that is, approximately 1300-3300 ng/µl from 1:10 diluted samples, thus expecting undiluted concentrations of 13,000-30,000 ng/µl.

Single colonies were picked and set to grow in LB media (volume 50 mL, 1:1000 antibiotics) over night with incubation at 37°C and agitation. The main culture was set to a total volume of 1.5 L (LB media). First started with 1 L + 1:1000 antibiotic and inoculation with preculture, this was further set to grow at 37°C with agitation. At OD_{600nm} = 0.3, the remaining volume of LB media was incorporated (with its correspondent 1:1000 antibiotics). When OD_{600nm} = 0.6 (recommended for protein expression), the culture was induced with 0.4 mM IPTG for sRNA expression (if required for the sample) and incubated for 2 hours with agitation.

Cell pelleting was performed in autoclaved (700 mL capacity) plastic cylinders. Centrifuging 500 mL volume at a time for 2 minutes at 5000 G, then discarding the supernatant. For the last centrifugation, after discarding supernatant, the sediment was dissolved in fresh LB-media, and further transferred to 50 mL falcons. Falcons were then centrifuged again and the supernatant was discarded, this step in order to separate the sample in a more manageable amount for RNA trizol extraction.

Each pellet was resuspended in 3 mL trizol and immediately transferred to a 15 mL falcon, then incubated at room temperature for 5 minutes. 700 µl chloroform was added, the falcon was vortexed for 15 seconds and then incubated for 2 minutes at room temperature. 15 minute centrifugation at 4 °C 12000 G was performed for phase separation. In a new 15 mL falcon, the upper phase (corresponding to the same pellet) was transferred together (volume $\sim$ 6 mL). 1 volume of 2-isopropanol was added and stored at -20°C overnight (or until required for the Northern blot experiments). The content was transferred to RNase-free eppis (1.5 mL at a time) and centrifuged at 4 °C, 12000 G for 30 minutes. Afterwards the supernatant was discarded and more volume was added until finished. The RNA pellet was washed in 500 µl 80% ice cold EtOH and afterwards resuspended in H_2O_{DEPC}.

3.4.8 Northern dot blot

To detect the production of sRNAs, Cy3 labeled oligonucleotides (Table A.2) were probed against the sRNAs, and 5S rRNA (as control) by a tailored northern dot blot assay (Figure A.3B). Cultures were set up to OD_{600nm} = 0.05, and grown for 90 minutes. One sample was induced with 0.4 mM IPTG (OD_{600nm} =0.375 at induction) and incubated for 120 minutes at 37 ° C with 200 RPM, the other sample (uninduced) was used as control.

Another RNA sample (Sample C) was obtained with the tailored RNA protocol (Figure A.3A), since a high amount of RNA was required. Induction was performed at OD_{600} = 0.511, following the recommended OD_{600nm} = 0.5-0.6 for protein expression. For each of the total RNA samples, a mixture of 10 µl of sample and 10 µl of denaturing buffer was cooked for 2 minutes at 95 °C. Samples were directly loaded to a nylon membrane (see Table A.4) 2 µl at a time, and air-dried for 5 minutes after each load. Crosslinking the membrane at max. capacity (254 nm 9999 mJ cm 2) was performed twice. The membrane was cut for using with different buffers and probes, then blocked with 10 µl in buffer (Church's and Denhardt) and incubated for 1.5 hours at 28 °C with agitation. Probes were cooked for 5 minutes at 90 °C. For hybridization, 8 µl of probe were added to its correspondent membrane in the blocking solution, then incubated overnight at 28 °C. The washing steps as follows: For Church′s buffer (signal) wash 1: 2x SSC, 0.2% SDS (10 mL), at room temperature for 10 minutes with agitation; wash 2: 1x SSC, 0.1% SDS (10 mL), at room temperature for 7 minutes with agitation. For Denhardt buffer we need 4 different washes (no clear signal using this buffer) wash 1: 6x SSC, 0.1% SDS (10 mL), at 35 °C for 10 minutes with agitation; wash 2: 6x SSC, 0.1% SDS (10 mL), at 35 °C for 10 minutes with agitation; wash 3: 6x SSC, 0.1% SDS (10 mL), at 30 °C for 5 minutes with agitation; wash 4: 6x SSC, 0.1% SDS (10 mL), at 30 °C for 5 minutes with agitation.

Blots were quantified with ImageJ [209] by measuring the integrated density of the blots and subtracting the background value.

3.4.9 Reverse transcription polymerase chain reaction (RT-PCR)

RT-PCR was used to detect the TF transcripts and therefore verify the production of TFs. Cultures were set to OD_{600}= 0.05, then grown for 90 minutes. Selected samples were induced with 0.1% L-arabinose and incubated for 120 minutes at 37°C with agitation. After culturing, induction and RNA isolation, the RNA template was converted into complementary DNA (cDNA) by reverse transcriptase (RT). The cDNA was used as a template for PCR amplification. RT-qPCR primers (Table A.2) have been used to perform a gradient PCR (Dream Taq polymerase) at two different temperatures (59.6°C and 61.5°C) for evaluating the transcripts. Conditions as follows: 35 cycles; initial degradation, 5 minutes, 95°C; denaturation 96-98°C, 35 seconds; primer annealing (59.6°C and 61.5°C), 45 seconds; extension, 30 seconds (60 seconds x kb), 74°C (optimum for Taq) 30 seconds; final extension, 10 minutes, 72°C.

PCR products were verified by a 3% agarose gel electrophoresis and RedSafe staining.

3.4.10 GGA assembly for the construct of the sRNA targeting t1A3T2 tRNA

Cloning was performed in strain DH5-α with the Golden Gate assembly (GAA) protocol using BsmBI Type IIS Enzyme (New England Biolabs). Primers (Table A.3) were used to amplify the fragments of the desired construct by PCR. Afterwards, assembly was confirmed by sequencing (Eurofins) and plasmid fingerprinting with *BsaI* and *PstI* restriction enzymes.

Chapter 4

miRNA-based diagnostics and therapeutics in cystic fibrosis

MicroRNAs dysregulation has been repeatedly observed during the progression of various diseases (e.g. cancer) [27]. Hence, miRNAs can serve as potential indicators of disease conditions as well as in therapeutic approaches. There are two types of miRNA therapeutic approaches: miRNA mimics and antagonists (antimiRs) [210]; the first one seeking to increase levels of downregulated miRNAs or miRNA replacement, and the latter corresponding to miRNA inhibitors, used mainly when miRNAs appear upregulated during the course of the disease (similar to siRNAs).

In this Chapter, dysregulated miRNAs in cystic fibrosis (CF) are reviewed and classified as potential biomarkers for the diagnosis or therapy of said genetic disease, with the aim to facilitate the identification of components for further design and development of molecular circuits/devices against CF.

4.1 Dysregulation of miRNA in cystic fibrosis

Cystic fibrosis transmembrane conductance regulator (CFTR) is positively regulated by several transcription factors (e.g. C/EBP proteins and FOXA factors) [211] and posttranscriptionally regulated by miRNAs [212]. Mutations in CFTR are implicated in the devastating genetic disease cystic fibrosis (CF).

CF is a multisystem disease, where lungs, digestive system, sweat glands and reproductive tract of the patient are compromised [213]. Disease-causing variants (Figure 4.1) have a different effect on the expression and function of CFTR, which in turn determines the degree of alteration of the epithelial ion transport in an individual with CF [214]. Therapeutics for CF that act at the molecular level are based on restoration of the CFTR function in the affected tissues.

Several studies show that miRNAs have important roles in maintaining homeostasis in the developed lung, and that their expression profile remains rather constant during aging in the adult lung [217]. Alterations in miRNA expression in inflammatory lung diseases, like CF, shows their implications in regulating innate immunity, inflammation, ion conductance, and endoplasmic reticulum stress related proteins [212]. Particularly, miR-101, miR-145, miR-494, and miR-509-3p play strong roles in regulating CFTR expression [218, 219]. A summary of the miRNAs that are related to CF disease is compiled in the Table 4.1.

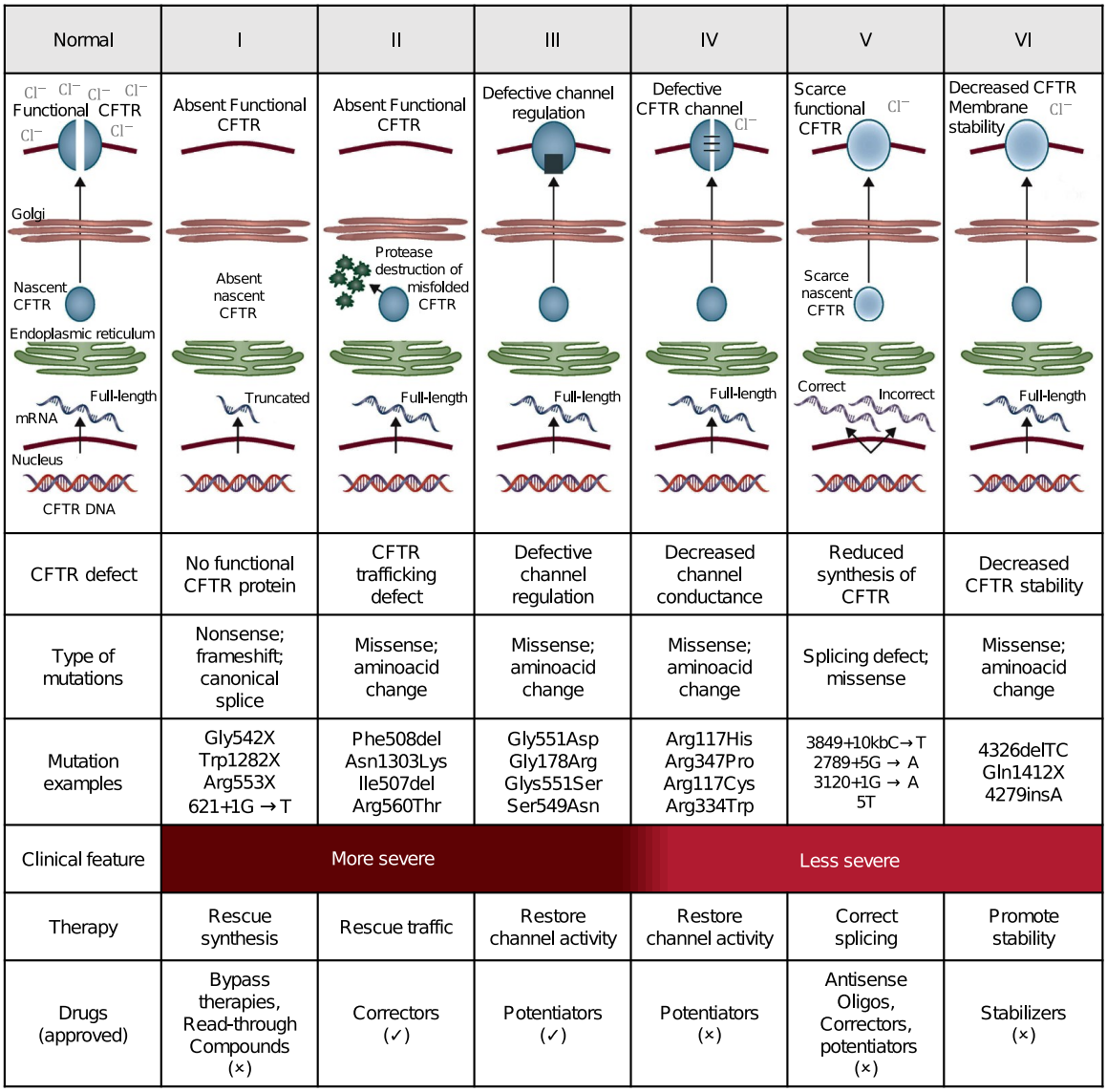


Figure 4.1: An overview of cystic fibrosis. CFTR mutations are categorized into 6 classes (I-IV) according to their functional impact (adopted and modified from [215] and [216]).

4.1.1 Circulating or extracellular miRNAs

Literature on circulating miRNAs profiling in CF is scarce. A recent study [220], used high-throughput microarray technology to extract extracellular miRNA profiles in CF plasma samples compared to healthy controls; they found 10 overexpressed miRNAs (miR-486-5p being the most significant), and only one underexpressed (hsa-miR-598-3p). Interestingly, they did not find three of the miRNAs (miR-122, miR-25, miR-21) previously reported as dysregulated [221], likely because the extracellular miRNA profiles are influenced by the cell or tissue type or time of disease progression.

Another study [28], comparing extracellular miRNA levels from plasma samples from female and male CF pediatric patients (under 6 years) revealed a significant increase in miR-885-5p in females compared to males.

Table 4.1: Summary of dysregulated miRNA reported in the literature. CF, related to cystic fibrosis; COPD, related to Chronic obstructive pulmonary disease (COPD); ↑ upregulated expression; ↓ downregulated expression; ✓related to; - unknown; ×, not related.

miRNA	CF	Cell type	COPD	Mutation	Function(s)
miR-145	↑ [211, 218, 219, 222]	bronchial epithelial [218]	↓ [223, 224]	F508del-CFTR [212, 218, 219]	mediates TGF-β inhibition of CFTR synthesis and function in airway epithelia [225]; regulates expression of ATF6 [219]
miR-138	↓ [226]	airway epithelial [226]	-	F508del-CFTR [226]	regulation of CFTR expression through interacting with the protein SIN3A [226]
miR-31	↓ [227]	bronchial epithelial [227]	✓[228]	F508del-CFTR [227]	its deregulation results in increased IRF-1 and in turn, increased CTSS expression and secretion [227]
miR-223	↑ [218]	airway epithelial [218]	↑ in lung tissue [229]	F508del-CFTR [218]	role in hematopoietic differentiation [230]
miR-494	↑ [219, 222]	airway epithelial [219]	↑ [231]	F508del-CFTR [219]	regulates expression of ATF6 [219]
miR-221	↑ [219]	airway epithelial [219]	↑ [232]	F508del-CFTR [219]	regulates expression of ATF6 [219]
miR-101	↑ [211, 218, 219, 233]	airway [218] and nasal [211] epithelial	↑ [233]	F508del-CFTR [211, 219]	regulates change from strong to low CFTR expression from fetal to after birth [211]

miR-509-3p	↑ [234]	airway epithelial [234]	-	F508del-CFTR [234]	inhibits CFTR expression and then may have a role in CF pathogenesis [234]
miR-9	↑ [235]	bronchial [235]	-	F508del-CFTR [235]	regulates ANO1 [235]
miR-16	basal [236]	primary lung epithelial [236]	× [237]	F508del-CFTR [236]	related to inflammation and CFTR rescue [236]
miR-126	↓ [238]	airway epithelial [238]	-	F508del-CFTR [238]	regulates TOM1 [238]
miR-302a	↓ [236]	primary lung epithelial [236]	-	F508del-CFTR [236]	related to inflammation and CFTR rescue [236]
miR-155	↑ [239, 240]	lung epithelial [240]	✓[241]	F508del-CFTR [240]	promotes hyperexpression of interleukin-8 and inflammation; fibrosis [240]; targets RPTOR [242]
miR-215	↑ [239]	lung epithelial [239]	-	F508del-CFTR [239]	inflammation [239]
miR-144	↑ [233]	airway epithelial [233]	✓[229, 233]	-	regulates expression of the CFTR chloride channel in the lung [233]
miR-17	↓ [243]	bronchial [243]	-	F508del-CFTR [243]	inhibited basal and agonist-induced IL-8 protein production, contributes to inflammation [243]
miR-93	↓ [244]	bronchial epithelial [244]	× [245]	F508del-CFTR [244]	regulates IL-8 protein production [244]

miR-486-5p	↑ [220]	plasma [220]	↑ [246]	F508del-CFTR [220]	role in hematopoietic cell differentiation via regulation of FOXO and AKT expression [247]

4.2 Candidate miRNAs for diagnostics

4.2.1 miR-145 is overexpressed in CF bronchial epithelium

In [218], it was found that increased expression of miR-145 *in vivo* in bronchial epithelium of individuals carrying the F508del-CFTR mutation correlates with decreased CFTR expression. However, it is important to distinguish with the relation of this miRNA with COPD. The miR-145 duplex is composed by miR-145-5p (guide strand) and miR-145-3p (passenger strand).

The role of miR-145-5p in COPD is elusive, however, in a bioinformatics analysis finds it significantly downregulated in COPD patients compared to healthy individuals. It has been shown that miR-145-5p is associated with the ED patterns of lung function growth leading to COPD in children with asthma [223], and additionally increases airway smooth muscle cell proliferation. Thus, the study suggests that reduced expression of miR-145-5p is a risk factor for ED of long-term lung function. Also, miR-145-5p has been shown to be associated with smoke-related COPD [224]; however is downregulated in the lungs from smokers with or without COPD. It was also shown that miR-145-5p overexpression mitigated CSE-induced apoptosis and inflammation response by the regulation of p53-mediated apoptotic signaling and pre-inflammatory factors. Moreover, miR-145 is also related to some other diseases like pulmonary hypertension [248], Atherosclerosis [249], breast cancer [250], among others.

The role of upregulated miRNAs in CFTR regulation *in vivo* in bronchial brushings from individuals homozygous or heterozygous for F508del-CFTR was investigated in [218]. They predicted in silico the miRNA targets, and the expression of miRNA and target genes were measured by quantitative real-time PCR and Western blotting. They further validated results *in vitro*, and determined the impact of defective chloride ion conductance, genotype, and colonization status on miRNA expression. miR-145, miR-223, and miR-494 were upregulated in CF compared with non-CF bronchial brushings and cell lines, in F508del-CFTR homozygotes compared to heterozygotes, in subjects positive for *P. aeruginosa* infection, and in cells treated with a CFTR inhibitor or IL-1β. In addition, their findings suggest that increased expression of miR-145, miR-223, and miR-494 *in vivo* in bronchial epithelium of individuals carrying the F508del-CFTR mutation correlates with decreased CFTR expression.

Another study [222] identified both hsa-miR-145 and hsa-miR-494 (upregulated), as regulators of the CFTR expression, targeting sites in the CFTR 3' UTR directly.

4.2.2 miR-486-5p is differentially expressed in CF human plasma

As reported in [220], miR-486-5p was the most significantly differentially extracellularly elevated miRNA in CF plasma compared to healthy controls. It is associated to hematopoietic cell differentiation via regulation of FOXO and AKT expression [247], and its aberrant expression in plasma has been reported in numerous cancers [251].

Another study [246] investigated the role of this miRNA in COPD progression, and finds it also significantly upregulated in COPD smoker groups compared to controls.

4.2.3 miR-31 is downregulated in CF bronchial epithelia

A study [227] identified decreased expression of miR-31 in the CF airways, and found that this miRNA is a a regulator of CTSS expression via the transcription factor IRF-1 in CF epithelial cells. Dysregulation of this miRNA has also been reported in cancer [252] and psoriasis [253], and COPD.

In [228], miR-31 was found to be dysregulated in the lungs of mice chronically exposed to cigarette smoke and also in smokers with COPD, compared to non-smoking controls, suggesting its potential role in COPD pathogenesis.

4.2.4 miR-221 is upregulated in CF airway epithelia

According to one study [219], this miRNA is upregulated in CF bronchial brushings and CF cell lines compared to non-CF individuals. A significant upregulation of this miRNA was found in blood plasma from acute pulmonary embolism patients compared with healthy controls in [254].

A recent study [232] features this miRNA as a potential biomarker for COPD, as they have found it to be upregulated in plasma from COPD patients, and it negatively correlated to the forced expiratory volume in 1 second (FEV1)/forced vital capacity.

4.2.5 miR-494 levels increased in CF airway epithelia

miRNA-494 was found upregulated in CF bronchial brushings and CF cell lines compared to non-CF [219]. Also another study [218] shows that this miRNA is upregulated in CF bronchial epithelium.

miR-494-3p has been also shown to be directly bound to SIRT3 and elevated in patients with COPD [231].

4.2.6 miR-223 increased expression *in vivo* correlates with decreased F508del-CFTR expression

A study [218] suggests that this miRNA is constitutively upregulated in CF bronchial epithelium of F508del-CFTR CF patients because of a lack of CFTR function. However, this miRNA has also been related to COPD [255].

4.2.7 miR-155 overexpression induces CF inflammatory phenotype

It was shown that this miRNA was at least 5-fold overexpressed in CF lung epithelial cells of diseased individuals compared to controls. Additionally, it is involved in the activation of IL-8-dependent

inflammation in CF [240].

Another study [242] shows that overexpression of miR-155 induces the pro-inflammatory phenotype in CF by hyper-inducing IL-8 levels. They have also identified that regulatory associated protein of mTOR, complex 1 (RPTOR) is a target of this miRNA. Their findings suggest that miR-155 is a potential therapeutic target for CF therapy.

It has also been identified to be involved in cigarette smoke induced inflammation and COPD in mice, by increasing inflammation-related genes, and suggested as a therapeutic target in COPD [241].

4.2.8 miR-215 upregulated in chronic inflammation

In [239], miR-155 is suggested as a biomarker for the absence of infection in CF cells, whereas miR-215 as a biomarker of chronic infection, because it is upregulated in CF cells infected with *P. aeruginosa*.

4.2.9 miR-144 upregulation leads to CFTR suppression

Results from [233] suggest that the upregulation of either or both miR-101 and miR144, could lead to the suppression of CFTR in COPD patients. It has been found to be among the three most upregulated miRNAs in the lung of COPD patients [229], and also has been related to cancer [256].

4.3 Candidate miRNAs for therapeutics

The main challenges miRNA-based therapeutics face are related to *in vivo* stability and organ/tissue specificity delivery. Furthermore, testing has to be performed in appropriate human CF lung models [257]. It is important to note that another challenge for therapeutical use of miRNA is the plethora of mutations affecting the CFTR gene [258]: missense (42%), nonsense (10%), frameshift(15%), splicing (13%), in-frame deletion/insertion (2%), and promoter (0.5%); suggesting that different miRNAs should be considered.

Therapeutics involving delivery (miRNA mimic) and inhibition (antagomiR, peptide nucleic acid) have been performed [211, 218, 225, 235, 259, 260].

4.3.1 miR-145

miR-145 mediates TGF-β inhibition of CFTR synthesis and function in airway epithelia [225], and some miR-145 antagonists discontinue TGF-β signaling and restore the function of F508del CFTR, featuring miR-145 as a novel therapeutic approach for correction of F508del CFTR in CF epithelia. Their experiments show 4-fold increase of miR-145 on CF bronchoalveolar lavage fluid compared with non-CF, and 10-fold increase in CF primary airway epithelial cells. Antagonists of miR-145 were able to interrupt TGF-β signaling and restore F508del CFTR modulation.

Another study [261] reinforces the conclusion of targeting miR-145 for therapy of CF and uses an antisense peptide nucleic acid (PNA) against miR-145-5p (PNA-a145). Measurements of both mRNA (qRT-PCR) and protein (Western blotting) levels, show that the expression of the CFTR gene is enhanced. It is suggested that miR-145 targeting as a therapeutic approach could be used in type IV (low channel activity), V (less protein), and VI (lower protein stability) CFTR mutations; for types

I (no protein), II (no chlorid traffic), and III (loss of function), it could be effective combined with other therapeutic approaches [258].

4.3.2 miR-138

The study [226] presents a global miRNA expression profile in primary human airway epithelia by quantitative PCR. It finds that miR-138 regulates CFTR expression by regulating SIN3A protein. SIN3A in turn, transcriptionally represses CFTR expression. The findings in this work propose miR-138 mimic as a candidate for therapeutic purposes, nonetheless, its specific delivery to the airway epithelia faces still problems.

Evidence of dysregulation or direct relation of miR-138 with COPD has not been found, however it has been reported to be downregulated in asthma patients [262].

4.3.3 miR-9

As suggested in [235], miR-9 is overexpressed in CF cells and it directly regulates anoctamin 1 (ANO1), causing a decrease in both its expression and activity. Since miR-9 targets multiple mRNAs a direct inhibitor is not recommended and hence they designed a specific target site blocker (TSB) that binds to the ANO1 3' UTR (ANO1 TSB) to prevent miR-9 binding. The goal of the therapy approach is to enhance ANO1 chloride channel activity, and propose ANO1 TSB as a candidate drug for CF patients, and it is hypothesized to work in all CFTR mutations. The relation between miR-9 to COPD, has only been implicated on corticosteroid insensitivity in both asthma and COPD [263].

4.3.4 miR-101

In a miRNome profiling study [211], it was found that miR-101 and miR-145 are upregulated in adult lungs. Furthermore, their findings suggest that miR-101 (combined with an overlapping AU-rich element) acts directly on its cognate site in the 3'UTR of CFTR mRNA. Another study [233] also showed an upregulation of miR-101 and miR-144 in human bronchial epithelial cells when exposed to cigarette smoke and cadmium. Furthermore, miR-101 is highly expressed in lung samples from severe COPD smoking patients.

4.3.5 miR-509-3p

One study [234] shows that miR-509–3p and miR-494 (upregulated) have an impact in CFTR expression in human airway epithelia. By transfecting cells with a miR-509–3p mimic the authors observed significantly decreased CFTR mRNA and protein levels in both airway epithelial and Calu-3 cells. In turn, when transfecting with a miR-509–3p anti-miR, the CFTR mRNA and protein expression increased in both cell types. Since the overexpression of F508del CFTR has barely an effect on the Cl^- transport rescue, they do not recommend anti-miRs as therapeutics for F508del-CFTR. Additionally, this study suggests that since infection and inflammation are strongly linked to COPD and asthma, it might influence miR-509–3p and miR-494 concentrations.

In [260], the authors demonstrate the inhibition of miR-509-3p activity with a short (7 bp) peptide nucleic acid (PNA) targeting exclusively the seed region of miR-509-3p miRNA. Also, more

recently [264], PNAs have been shown to protect specific sequences in the 3'UTR of the CFTR mRNA from miR-509–3p, suggesting its therapeutic use.

4.3.6 miR-126

This was the first miRNA associated with CF [238], and is downregulated in CF airway epithelial cells. A target of miR-126 is TOM1 (target of Myb1), which is upregulated *in vivo* in CF bronchial brushings. This miRNA was administered in nanoparticles to investigate the efficient delivery of miRNA mimics into CF airway epithelial cells [259].

Together with miR-9 and miR-21 it has been implicated in corticosteroid insensitivity in asthma and COPD [263].

4.3.7 miR-16

A study [236] finds that miR-16 aids in the functional rescue of the cyclic AMP-activated apical F508del-CFTR chloride channel in primary lung epithelial cells from CF patients. Furthermore, the authors propose this miRNA as a novel approach for CF therapy. Expression of a miR-16 mimic in the CF lung or pancreatic cells increased significantly the expression of the F508del CFTR protein.

This miRNA was also found to be downregulated in primary fibroblast cell lines from lungs of heavy smokers (and COPD) compared to non-smokers [265]. However, a study [237] that analyzed blood samples from COPD patients found no significant difference on miR-16 levels from patients and controls.

4.3.8 miR-302a

It has been shown that its overexpression promoted an eightfold increase in mature CFTR protein expression [236].

4.3.9 miR-17

In [243], miR-17 is suggested as a potential anti-inflammatory therapeutic for CF and chronic inflammatory airway diseases. They found that by overexpressing miR-17, the production of basal and agonist-induced IL-8 protein was inhibited in F508del-CFTR homozygous tracheal and bronchial epithelial cells; therefore, defective CFTR, inflammation, neutrophilia and mucus overproduction could potentially be regulated by miR-17. Decreased levels of this miRNA were observed *in vivo* in CF bronchial brushings from mice with spontaneous airway neutrophilia and mucus obstruction. In epithelial cell lines and primary bronchial cells from pediatric CF samples the levels did not differ, contrary to the decreased levels observed in the adult CF lung.

4.3.10 miR-93

Enhanced expression of miR-93 reduces the expression of the IL-8 gene [244]. Pre–miR-93 mimics can be used as therapeutics to tune down the IL-8–dependent inflammatory response in CF.

A study [245] finds that this mRNA has been related to cancer rather than to COPD, however it was not the most representative biomarker. Additionally, extracellular miR-93 has been suggested as a biomarker for CS-induced emphysema [266].

Chapter 5

Towards semi-autonomous circuits for the diagnosis and mitigation of genetic diseases

This chapter summarizes the efforts to build the first (to the best of our knowledge) prototype of semi—autonomous synthetic RNA circuit for the purpose of non-invasive diagnosis and mitigation of premature termination codon related diseases, exemplified by cystic fibrosis (CF).

5.1 Nonsense mutation-related diseases and RNA-based therapies

Nonsense mutations are single base substitutions within the gene sequence, consequently resulting in the transformation of the involved codon into an in-frame termination codon (TTA, TAG, TGA), termed as premature termination codon (PTC). This type of mutations account for around 10% of the ones related to genetic diseases [22], such as neurological and metabolic pathologies, cancers or rare diseases such as cystic fibrosis (CF), Duchenne muscular dystrophy (DMD), and nephropathic cystinosis. There is currently no cure for around one-third of the genetic diseases that are caused by mutations that produce PTC [267], thus investigation and development of novel therapies targeting these anomalies is of significant importance. For example, many cancers are directly linked to the presence of a PTC in a tumor suppressor gene. PTCs are responsible for early termination of translation, which leads to production of truncated and most of the time, dysfunctional proteins.

Cystic fibrosis is a respiratory disease characterized by inflammation, obstruction and recurring infections that progressively damage the lungs. CF is originated by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, with equal (homozygous) and different (heterozygous) mutations on both alleles. This gene encodes the CFTR protein sequence, which is responsible for controlling the balance of salt and water in lungs and other tissues. CFTR mutations are categorized into five classes regarding how they affect the produced proteins, this study focuses on class I: nonsense mutations, splice mutations or deletions. Approximately 40% of the CFTR mutations belong to this class [268] and, according to the Cystic Fibrosis Foundation, around 22% of the CF patients carry at least one mutation (e.g. G542X, W1282X, R553X) in this class. Approximately 10% of CF cases are caused by PTCs [24], which result in truncated CFTR proteins that are most of the cases dysfunctional.

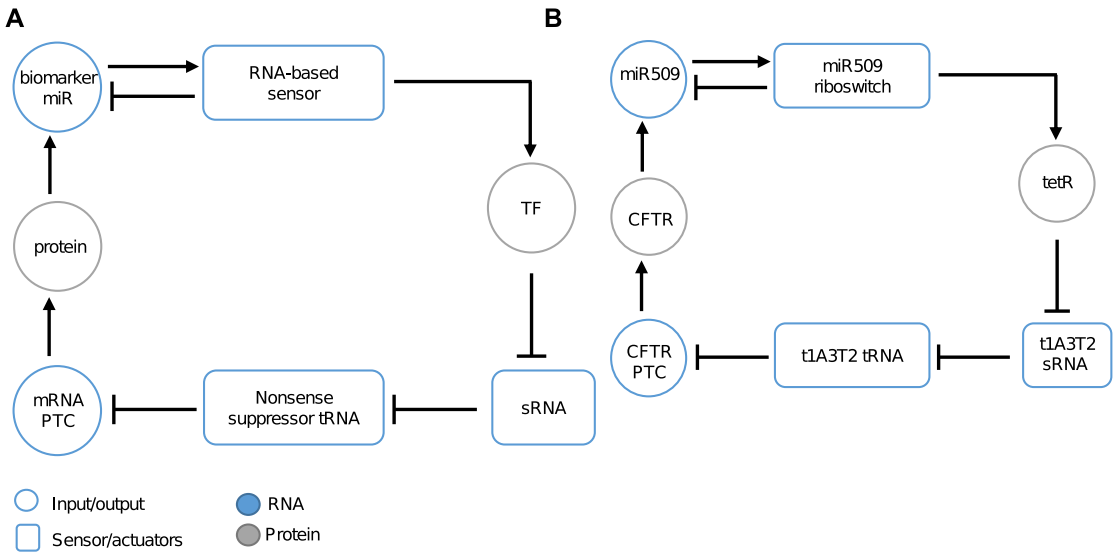


Figure 5.1: Reaction network of the circuit for PTC detection and suppression. (A) generic design for PTC-associated diseases: a biomarker of disease is detected by binding to an RNA-based sensor molecule, this interaction triggers the translation of a TF that in turn triggers the production of a therapeutic tRNA (controlled by synthetic sRNAs). tRNAs enable a read-though of the PTC-mRNA reestablishing the production of the functional protein **(B)** CF design.

Suppression of PTC offers a promising approach as a therapeutic strategy for related diseases, also known as the read-through approach. Some of the existing therapeutic approaches include read-through drugs (e.g. Aminoglycosides) [269], suppressor tRNAs [270], PTC pseudouridylation [271], and inhibition of nonsense-mediated mRNA decay. A promising development for implementing read-through approach is by means of engineered nonsense suppressor tRNAs [33].

Synthetic biological circuits aim to artificially produce cellular functions that otherwise are not present in nature, and thus reprogram organisms [272]. This electronics-inspired approach is achieved by means of interconnected characterized genetic/molecular parts, which in turn assemble components, networks and pathways, regulating targeted gene expression in living systems [273]. Synthetic bio-circuits have been applied for practical applications addressing modern-day concerns, for example de novo systems for gene regulation [1, 2], therapeutics and drug development [3, 4], cellular computation [5, 6], cell reprogramming [7, 8], and bioremediation [9], to name a few.

Recently, synthetic biology is opening new possibilities regarding personalized medicine [272], by engineering cells into therapeutic factories to fight against fatal diseases; some interesting examples are the development of RNA sensors designed to recognize specific disease related molecules, such as part of the genome sequence of Zika virus [156] and Ebola [20].

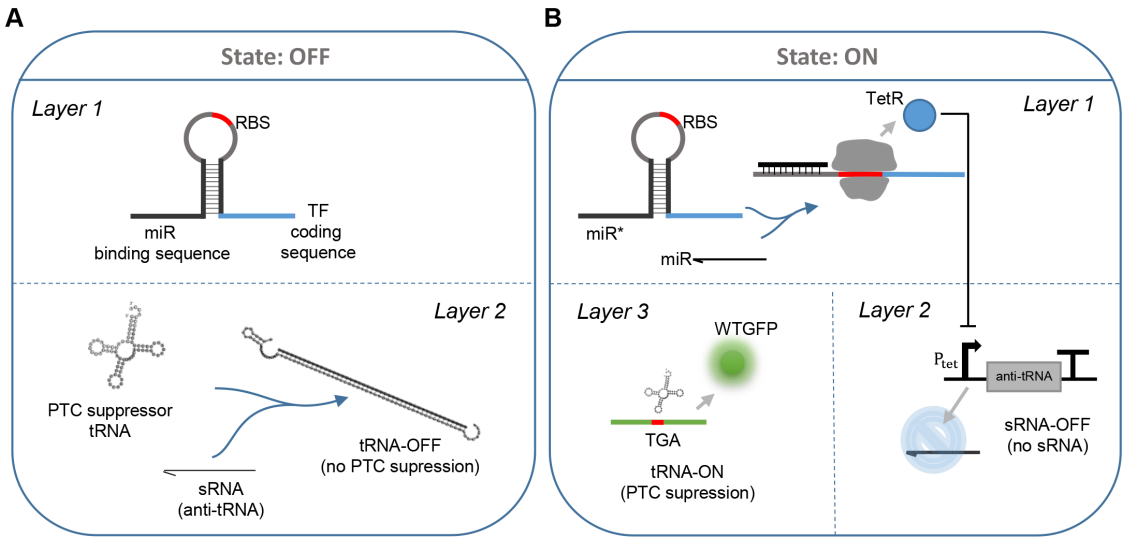


Figure 5.2: Conceptual design of the RNA circuit for PTC detection and suppression. (A) OFF state (default): tRNAs and sRNAs are constitutively expressed. **(B)** ON state (presence of CF biomarker): CF specific miRNA binds to sensor, the transcription factor (TetR) that negatively regulates the promoter of the sRNAs is produced; consequently, sRNA production halts, deriving into functional tRNAs to act upon the mutation on the mRNA.

5.2 Design and description of the circuit and components

A generic synthetic circuit towards implementing semi-autonomous diagnosis and therapeutics approaches against nonsense mutation-associated diseases has been designed (Figure 5.1A). It functions in a negative feedback-loop manner, where the input of the circuit is directly negatively regulated by the output, a therapeutic molecule to mitigate the disease state. The input corresponds to a biomarker (miRNA) indicating the disease state, which is to be recognized by the diagnostic module, which in turn will prompt the production of a therapeutic molecule. The diagnostic module is proposed as an RNA-based sensor, particularly a riboswitch, which in turn controls or switches the state of the therapeutic module. For the therapeutic module, the use of a therapeutic RNA-based molecule is proposed, namely a nonsense mutation suppressor tRNA that acts on the PTC-containing mRNA. CF is chosen as the disease for the design of the proof-of-concept (Figure 5.1B): an RNA-based toehold switch has been designed to sense an miRNA that is an indicator of the disease state, on recognition it allows for the translation of a Transcription factor (TF) that in turn allows for the production of the therapeutic tRNA.

The proposed circuit (Figure 5.2) has two main states: the default OFF state (Figure 5.2A), where tRNAs are expressed from a constitutive promoter, as well as sRNAs targeting the tRNAs, hence turning the tRNAs OFF and non-functional; the ON state (Figure 5.2B) occurs on the recognition of a CF biomarker, which triggers the RNA sensor (toehold-riboswitch) by hybridizing to it, thus allowing the translation of a transcription factor, which in turn regulates negatively the sRNA production and, consequently, functional nonsense suppressor tRNAs are able to act on the corrupted mRNA (a GFP reporter with TGA codon).

The design of the circuit was accomplished using a bottom up approach, were different components that integrate the modules were first selected, designed, adapted and characterized: a

miRNA as biomarker of the disease was selected (input), a toehold switch detecting said miR was designed (diagnosis module), a nonsense suppressor tRNA system [33] was adapted, and sRNAs for its control were designed (therapeutic module). The description of the components is discussed below.

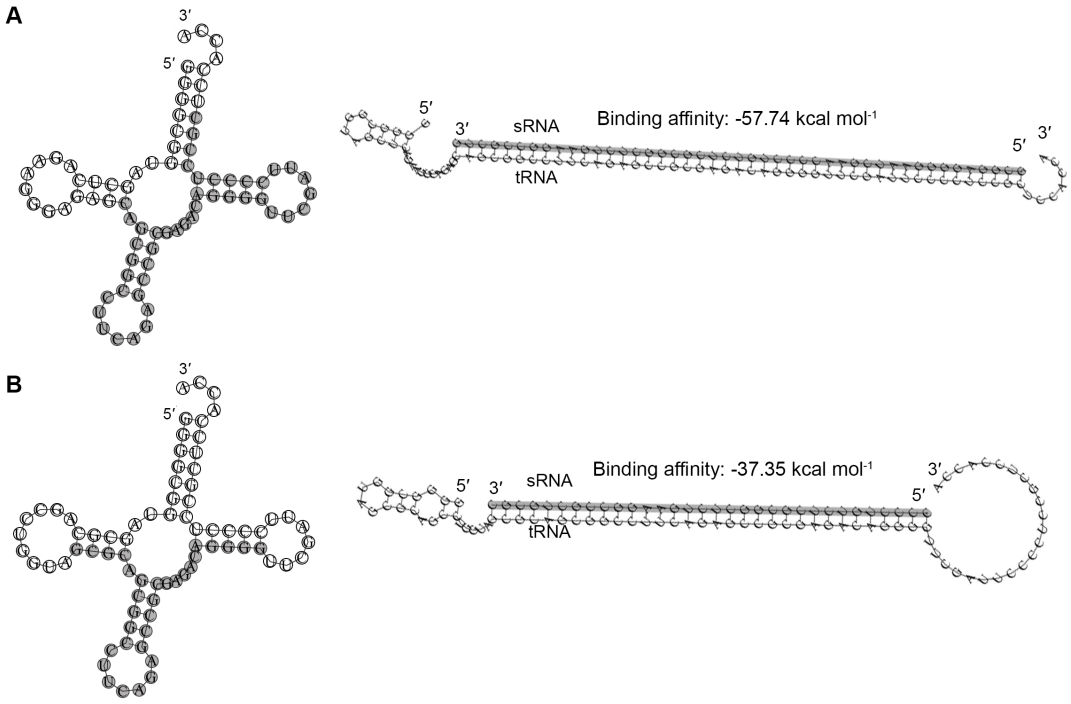
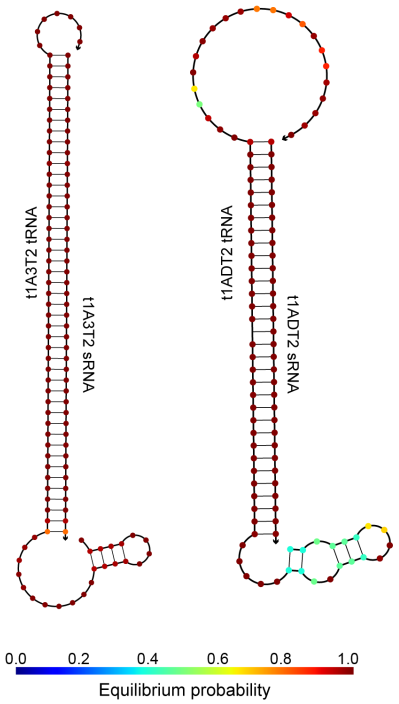


Figure 5.3: Schematics of suppressor tRNAs and duplexes generated by binding with their cognate synthetic sRNAs. Left – schematic of suppressor tRNA 2D structure, depicting the highlighted target binding site of the synthetic sRNA; right – binding energy simulation results and prediction of 2D structure of the tRNA-sRNA duplex for: **(A)** t1A3T2 and **(B)** t1A3DT2 tRNA. The sequences of both suppressors are from [33].

5.2.1 Therapeutic module: a nonsense suppressor tRNA system modulated by synthetic sRNAs

As a proof-of-principle, the design was concentrated on prokaryotic systems. A previously developed nonsense suppressor tRNA system [33] optimized to work in prokaryotes and developed in *E. coli*, was adapted for the circuit design. These *de novo* suppressor tRNAs were designed by structure prediction and logical engineering to enhance the natural $tRNA^{Ala}$ with an isoacceptor that decodes UGA stop codon. Two potential tRNAs were considered for this study, t1A3T2 and t1A3DT2 [33] (Figure 5.3AB - left). They were originally designed to be constitutively expressed (i.e. always ON) under the *lpp* promoter. For their adaptation as part of a module of the circuit, it is necessary to be able to switch and control their production in an OFF/ON state manner. Modulatory synthetic sRNAs (see Chapter 3) were proposed and designed for this endeavor. Transcription of the sRNAs has to be also constitutive to target the sRNAs as default state (Figure 5.2A), however, to be able to stop their transcription on a positive diagnosis (Figure 5.2B), a negatively inducible promoter is needed. A promoter suitable for this purpose is *TetR*, which is a constitutive repressible promoter, in



	t1A3T2		t1ADT2	
Sequence properties				
Free energy of ordered complex ΔG = - kT log Q (kcal/mol)	-105.06		-78.81	
Base	No.	%	No.	%
A	22	18.2	14	12.8
C	38	31.4	41	37.6
G	44	36.4	37	33.9
U	17	14.0	17	15.6
Other	0	0.0	0	0.0
Sequence/structure properties				
Free energy (kcal/mol)	-104.88		-77.58	
Probability	0. 746		0.137	
Ensemble defect (nt)	1.0		9.8	
Normalized ensemble defect (%)	0.8		8.9	
Size (nt)	121		109	

Figure 5.4: Thermodynamic analysis of secondary structures of suppressor tRNAs - synthetic sRNAs complexes. Minimum Free Energy Structure predictions and calculations details.

which by the addition of TetR transcription is repressed; furthermore, repression can be inhibited by the addition of tetracycline or D-anhydrotetracycline (aTc).

Synthetic sRNAs targeting the nonsense suppressor tRNAs (Figure 5.3AB) were rationally designed by calculating their predicted binding energy to their cognate tRNA. A first version of sRNAs was designed (example for t1A3DT2 in Figure 5.3B). The sRNA sequence targeted the sequence corresponding to the full length of the anticodon stem including the anticodon loop of the tRNA. For the t1A3T2 tRNA the sRNA (32 nt length) binding affinity is -36 kcal/mol), falling within the known optimal threshold for synthetic sRNAs [164]. For t1ADT2, the sRNA (32 nt) binding affinity is -37 kcal/mol.

Furthermore, a second version of the sRNA (45 nt) for the t1A3T2 tRNA was designed (Figure 5.3A). The sRNA sequence is complementary to the full anticodon loop, continues through the T-stem and T-loop and finishes with some bases of the acceptor stem of the tRNA. Binding affinity calculation results in -57.74 kcal/mol.

The predictions of secondary structures of the tRNA-sRNA duplexes (Figure 5.4) were further verified. The duplex between the second version sRNA and t1A3T2 tRNA resulted in the most stable structure. In addition, the probability of formation was reasonably higher and at the same time the ensemble defect relatively low, compared with the other duplex.

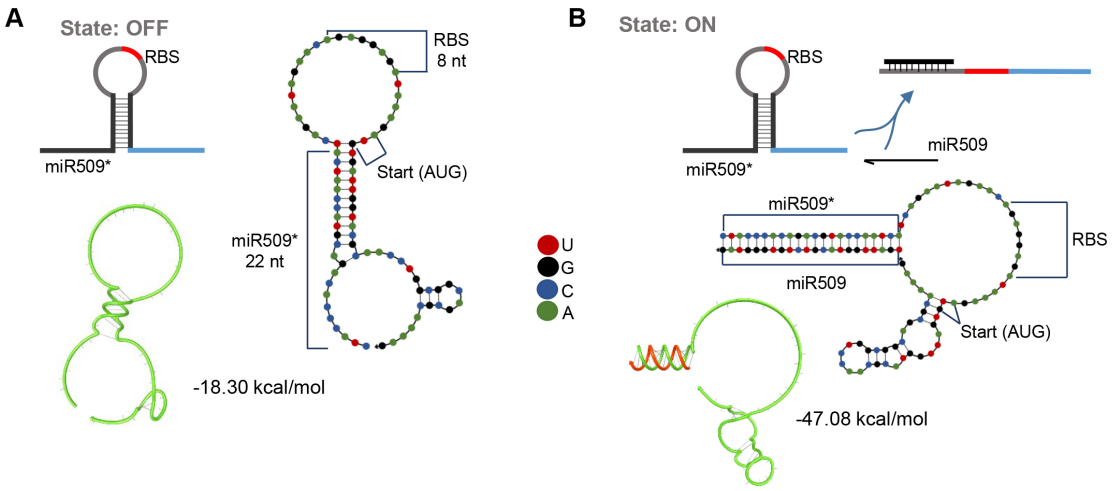


Figure 5.5: Scheme of the design and thermodynamic analysis of the RNA-based miR509 sensor (prokaryotic). Secondary structure evaluation of the RNA-based sensor in both ON and OFF states, identity of the bases in color. Prediction of the ideal helical geometry for both states, mRNA in green; miR509 in orange. **(A)** OFF state (default): while miR509 is not present, the secondary structure of the sensor prevents the ribosome from docking on the RBS and translating the TF that further controls the tRNA production. **(B)** ON state (miR509 detection): when miR509 binds to the sensor, the secondary structure is disrupted and translation of the TF is possible, consequently allowing the next steps for the tRNAs to function.

5.2.2 Diagnosis module: an RNA-based sensor/actuator with a miRNA trigger

Toehold switches [18] are secondary structure-based RNA sensors that are rationally designed to bind and sense specific RNA sequences. The sequences of toehold switches are designed to have a hairpin structure that prevents translation of a reporter gene. Here, the design of a toehold switch that binds to a biomarker of CF and allows for translation of a TF gene that prompts the activation of a therapeutic molecule is proposed.

The CFTR gene is 189 kb in length, the full length spliced mRNA contains 6128 nucleotides, and encodes a protein of 1,480 amino acids [274], thus being virtually impossible to be used as the trigger for a sensor, due to its extensive size. As discussed in Chapter 4, miRNA have been proven as reliable biomarkers for many diseases, without the exclusion of CF. Therefore, a miRNA has been proposed as the trigger for a diagnosis sensor.

5.2.2.1 A specific indicator of cystic fibrosis: miR509

The trigger miRNA for the sensor has to meet the following criteria: *i*) recognizing only/mostly CFTR related, *ii*) function in the lungs and *iii*) not related to Chronic obstructive pulmonary disease (COPD); *iv*) relation to ΔF508 or to R553X mutations. ΔF508 is the most common mutation class II CFTR mutation, which is a deletion of phenylalanine in position 508, and it is as deleterious as R553X (class I CFTR mutation). ΔF508 mutation is the most common CFTR mutation among Caucasians and causes protein misfolding and degradation, and consequently CF [275].

Additionally to the above criteria, it makes sense for the CF sensor miRNA trigger to be upregulated in pathological conditions. As shown in Table 4.1, miR-509-3p, miR-9 and miR-215

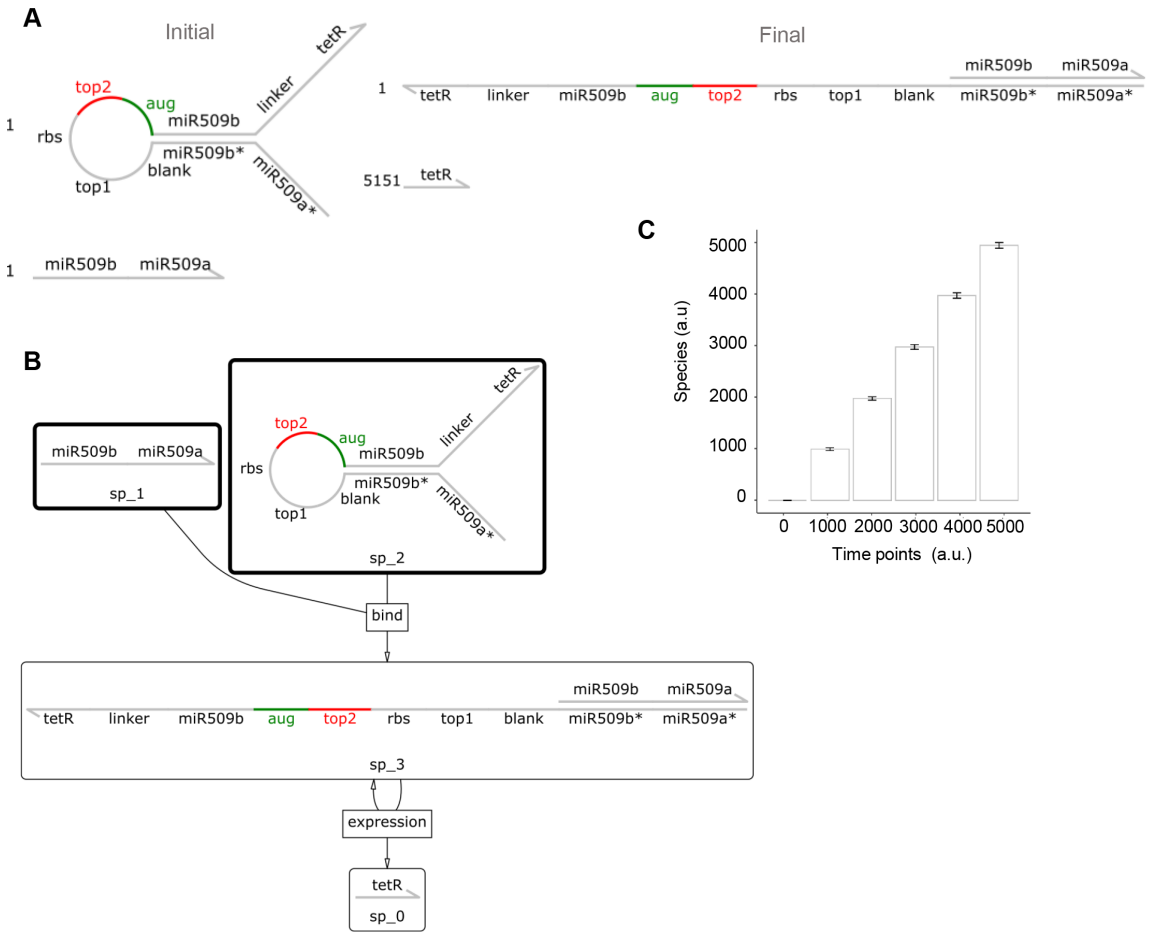


Figure 5.6: Chemical reaction network species and simulation results for miR509 sensor. **(A)** Initial and final (molecule) species of the reaction network; miR509 sensor and miR509 as initial species; the translated tetR gene (TF-protein) and the sensor-miR509 duplex as final species. **(B)** Chemical reaction network. **(C)** Results of the simulations over 5000 time points (n=30, data are means ± s.d.).

are potential candidates that meet several of the above criteria. Since it is known that miR-9 targets multiple mRNAs [235], and its low concentration (reduced by binding to the sensor) might cause undesirable side effects or detection problems it is discarded as a potential trigger. miR-215 is reported to be upregulated in lung epithelia of ΔF508 CF patients, but because of its major role in inflammation, it was suggested primarily as biomarker for chronic infection of the lung [239]. Although, miR-509-3p has been linked to inflammatory conditions in CF, this might suggest its presence in other diseases, like asthma or COPD [234], however it has not been disclosed.

In sum, miR-509-3p fulfills all the above criteria and makes it a suitable candidate for a sensor trigger. Despite some observations that it may be related to other diseases, its relation to COPD is questioned [234], and has been found to be downregulated in patients with pulmonary arterial hypertension with congenital heart disease [276], exactly the contrary conditions as expected in CF.

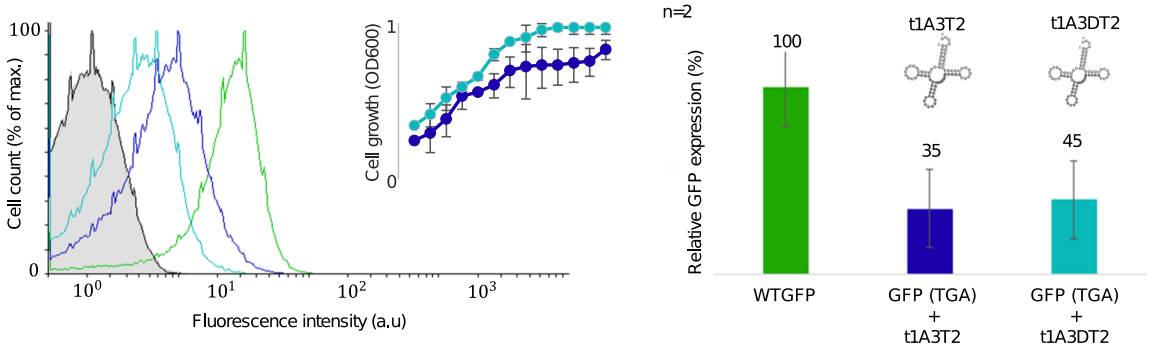


Figure 5.7: Performance verification of nonsense suppressor tRNAs by GFP readthrough assay. Fluorescence signals measured by FACS were normalized to the expression of wildtype GFP whose expression was set to 100%. Data are means ± s.d. (n=2). Confirmation of nonsense suppressor tRNAs performing an endogenous read-through in XL1-blue cells: t1A3T2 tRNA in blue, and t1ADT2 tRNA in light blue; as controls: wtGFP (TGA) in grey, cotransformed wtGFP (TGA) and empty tRNA vector in black, and wtGFP in green.

5.2.2.2 Design and simulation of a miR509-sensor

The diagnostic sensor was designed as a proof-of-concept to be tested in a fast growing organism, *E. coli*. The length of miR509 is 22 nt, the toehold or sensing region of the switch corresponds to the first half of the complementary sequence to miR509 (11 nt). The stem of the hairpin is composed of the other half of the complementary sequence to miR509 (remaining 11 nt) paired with its 3′ end. The upper part of the loop contains the ribosome binding sequence (RBS) followed by the start codon (AUG), which is occluded in a secondary hairpin structure and prevents the ribosome binding. Immediately after the sequence corresponding to the hairpin stem, a linker sequence follows, this helps to give stability to the switch. After the linker the gene sequence can be appended, starting from the AUG codon.

The sensor (Figure 5.5) is composed of the reverse complement sequence of miR509 (22 nt), immediately followed by a modified version from the loop conserved sequence of switch 27B from [156], the complementary sequence that forms the stem of the loop (11 nt identical to miR509 sequence) and a linker sequence (22 nt, from switch 27B [156]). The sequence of the repressed gene (tetR TF) is not included in the switch sequence and for the calculations, but added to the plasmid design.

According to GenBank (LM379805.1) [277], the sequence of miR509 is (5′- UGAUUGGUACGUCUGUGGGUAG -3′). The conserved sequences from switch 27B [156] are (5′- GGACTTTAGAACAGAGGAGATAAAGATG -3′) for the conserved loop and the sequence of the linker (5′- CAACCTGGCGGCAGCGCAAAAG -3′). The sequence of the designed miR509 RNA sensor is (5′- CUACCCACAGA CGUACCAAUCAUCAGAAUAGAACAGAGGAGAUAAAGAUGUGAUUGGUACGCAACCU GGCGGCAGCGCAAAAG -3′).

The hairpin of the sensor is expected to be sufficiently stable (Figure 5.5A) on the OFF state, and the free energy change (ΔG) of the miR-sensor duplex on the ON state (Figure 5.5B) is lower (-47.08 kcal/mol) and therefore, with higher probability to occur than the OFF state

(-18.30 kcal/mol). The MFE structures, free energy changes and ideal helical predictions were calculated with NUPACK [92]. The predicted helical structures show that the hairpin formation on the OFF state is achieved (Figure 5.5A) and the disruption of it while binding to miR509 is disrupted in the ON state (Figure 5.5B).

The reaction network of the sensor was verified with simulations with Visual DSD [278]. For this, the sensor sequence was encoded as a collection of long and short domains (AUG and top2) according to the design (Figure 5.6). The simulation was run 30 times, and mean of the results shows that the reactions are as expected.

5.3 Towards *in vivo* implementation of the CF circuit

A first step for the development was replicating the read-though assay of the suppressor tRNAs by a FACS experiment with the same conditions as reported [33]. The suppressor tRNA system [33] comprises two plasmids: one harboring the reporter, the sequence of wtGFP gene with a PTC (TGA) [pBAD33_GFP_TGA], and another plasmid containing tRNA [pBST NAV2_ t1A3T2 or pBST NAV2_ t1A3DT2]. The results from two independent replicates indicated successful PTC suppression function from both tRNAs (Figure 5.7). As can be seen from the FACS histograms, a shift towards higher fluorescence was clearly observed when suppressor tRNAs were present, suggesting successful expression of full-length GFP and consequently, suppression of the TGA premature termination codon. Results were also in accordance with the reported results [33], t1A3T2 yielded a moderate suppression efficiency compared to t1ADT2 tRNA (Figure 5.3). However, since t1A3T2 it yields a better expression, this was selected as one of the components of the system.

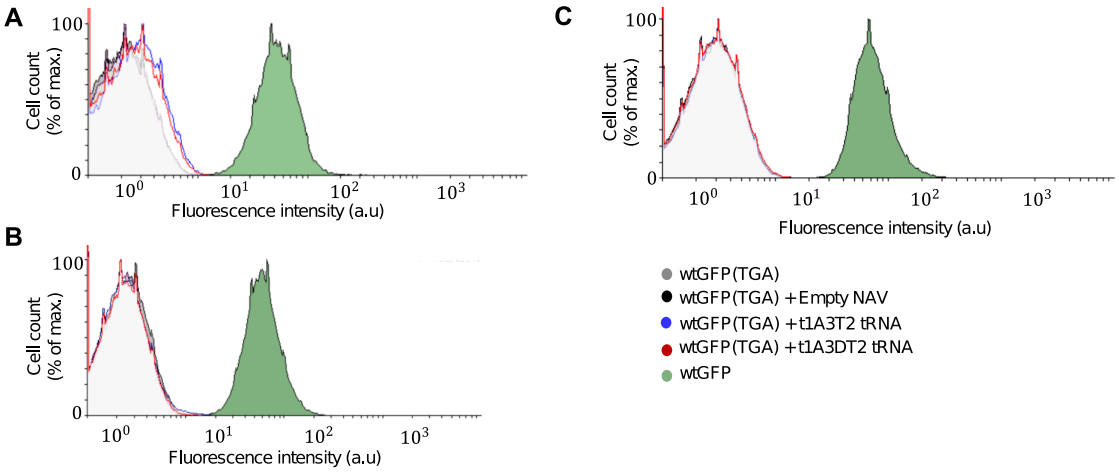


Figure 5.8: Performance verification of nonsense suppressor tRNAs by GFP readthrough assay in BL21. Three assays were performed with different conditions, however they were not successful: **(A)** Original experimental conditions, no effect can be observed. **(B)** Dilutions of the main cultures were changed, tRNAs showed no signal. **(C)** An incubation time of 2 h after induction was introduced, in order to give enough time for the tRNAs to perform rescue, still no fluorescent signal was observed.

5.3.1 Determining the host strain

The suppressor tRNAs were characterized in the *E. coli* XL1-blue [33] and modulating sRNAs were characterized in BL21 (DE) (Chapter 3). To determine if nonsense suppressor tRNAs can function in BL21 strain, three independent assays were performed changing the experimental parameters (Figure 5.8), overall with no positive results. Therefore, this was considered for the design of the expression plasmids, were the originally used promoter for expressing synthetic sRNAs was the T7 promoter, and further exchanged for *Lac* promoter, to adapt it to this strain.

For the first assay (Figure 5.8A), the experimental conditions were as suggested for the tRNA expression in XL1-blue. Nonetheless, BL21 cells grow much faster than XL1-blue, therefore, the main sample cultures reached a high optical density in a very short time. As can be seen from the histograms, the only fluorescence observed was from the GFP control, and no function could be observed from the tRNAs. Next, the inoculation dilution was modified to 1:50 for all samples and not only for the controls, expecting a slower growth of the samples. The cultures grew at a better pace, in theory giving time for the tRNAs to function, yet the results (Figure 5.8B) showed no rescue effect from the tRNAs. Finally, an incubation of two hours after induction before harvesting the cells was attempted in order to give enough time to the tRNAs to function, again with no effect observed (Figure 5.8C). The conclusion was that it is better to use the XL1-blue strain for the *in vivo* implementation of the circuit.

5.3.2 Screening for sRNA possible off-targets

Initially the sRNA sequences were screened for off-targets against Escherichia coli BL21-Gold (DE3) genome (Tax-id:866768), but no potential threatening targets were found. Furthermore, this was performed for Escherichia coli str. K-12 substr. MG1655 (NC_000913), since it is a parent strain from XL1-blue. Analyses were performed with RNApredator [94].

As a last and more stringent verification step, it was important to discard the possibility of the synthetic sRNA interfering with naturally occurring tRNAs, in *E. coli* (Figure 5.9). Sequences from the tRNAs were obtained from GtRNAdb [279]. First, the binding affinity between TS2 sRNA and each one of the tRNA was calculated using the RNAcofold [78] algorithm from Vienna RNA Package Python interface [86]. Furthermore, the free energy and probability of formation of the secondary structure of each dimer were calculated with NUPACK [92]. Although the stability of some duplexes resulted to be high, their probability of formation is relatively low, indicating there are no potential off-targets for *in vivo* implementation.

5.3.3 Assembly design

After verification of tRNAs function in host strain, and discarding possible off targets for synthetic sRNAs, the remaining parts of the circuit were selected and the plasmid for implementation was designed. For building and characterizing the system, four designs of plasmids for the expression of the circuit are presented (Figure 5.10B).

As a first test of the whole system function, a plasmid containing the TetR gene under control of the pLac promoter (without the sensor) is proposed (Figure 5.10A). However, the plasmid design includes the restriction sites *ApaI* and *PvuI* flanking the *tetR* gene, for the cloning of the miR509

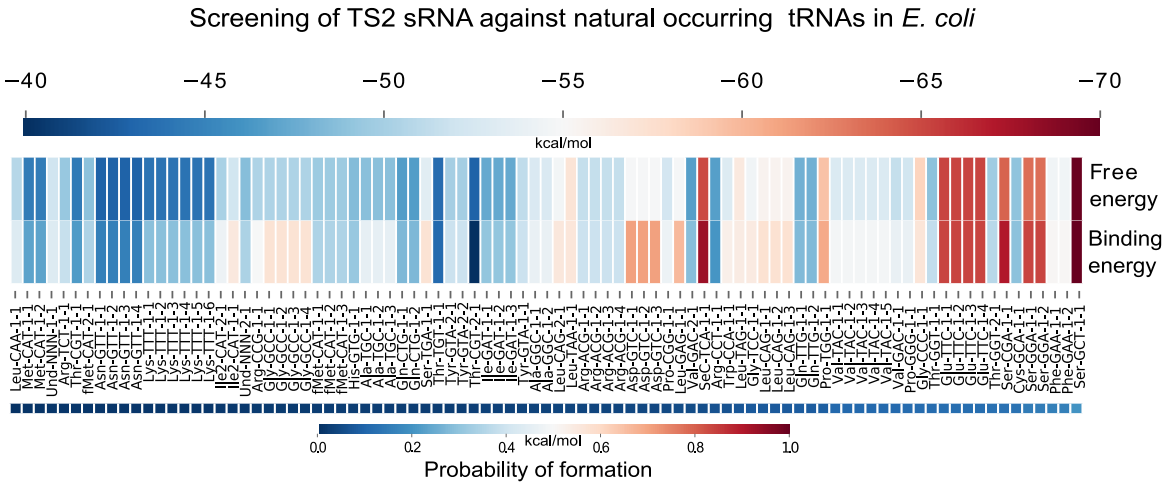


Figure 5.9: Screening sRNA against natural occurring tRNAs in *E. coli* strain K12. The binding affinity between the synthetic sRNA and each of the natural occurring tRNAs was calculated; for further verification, calculations of the free energy of the sRNA-tRNA duplexes were performed, as well as the probability of formation: the most stable structure (-69.88 kcal/mol) would be for the Ser-GCT-1-1 tRNA, however, its probability of formation is low (0.202). Full data set in Table A.5.

sensor after the system characterization (Figure 5.10A). It is hypothesized that by flanking the sRNA with the hammerhead (HHRz) and Hepatitis delta virus (HDVRz) ribozymes will promote better results from the sRNAs, due to exact cleavage of the sRNA sequence without leaving any extra nucleotides from transcription. Since the performance with and without the ribozymes has to be compared, designs of plasmids for this purpose are proposed (v1, v2 in Figure 5.10B).

Synthetic sRNAs that follow the canonical path in prokaryotes (Hfq-recruiting) have been shown effective for gene repression, however it is not clear if it will work for tRNA molecules. It is possible that the tRNAs are targeted by the sRNAs before transcription, but the possibility of using sRNAs for RNA-RNA interactions (sRNA-tRNA) is also desirable to be explored, this can be done with plasmid designs v3 and v4 (Figure 5.10B).

The plasmid designs are meant to be built with the Golden Gate Assembly (GGA) cloning technique, using the BsmBI Type IIS Enzyme in *E. coli* strain DH5-α. The system characterization is intended to be performed in *E. coli* XL1-blue strain grown in LB Miller broth.

5.3.4 Towards implementation in human cells

Since ultimately it is desirable to build a proof-of-concept in human cells instead of prokaryotes, sRNAs were verified for off targets in human tRNAs (Figure 5.11) and the miR509 sensor was designed for eukaryotic cells (Figure 5.12).

Screening of sRNAs against human tRNAs (Figure 5.11) showed no potential targets, suggesting that in theory, their use is also safe to use in human cells. Possible interactions were registered only with $tRNA^{Ala}$ variants (Ala-AGC-3-1) (see Table A.6). Although this probability is much lower than the one expected for a stable duplex.

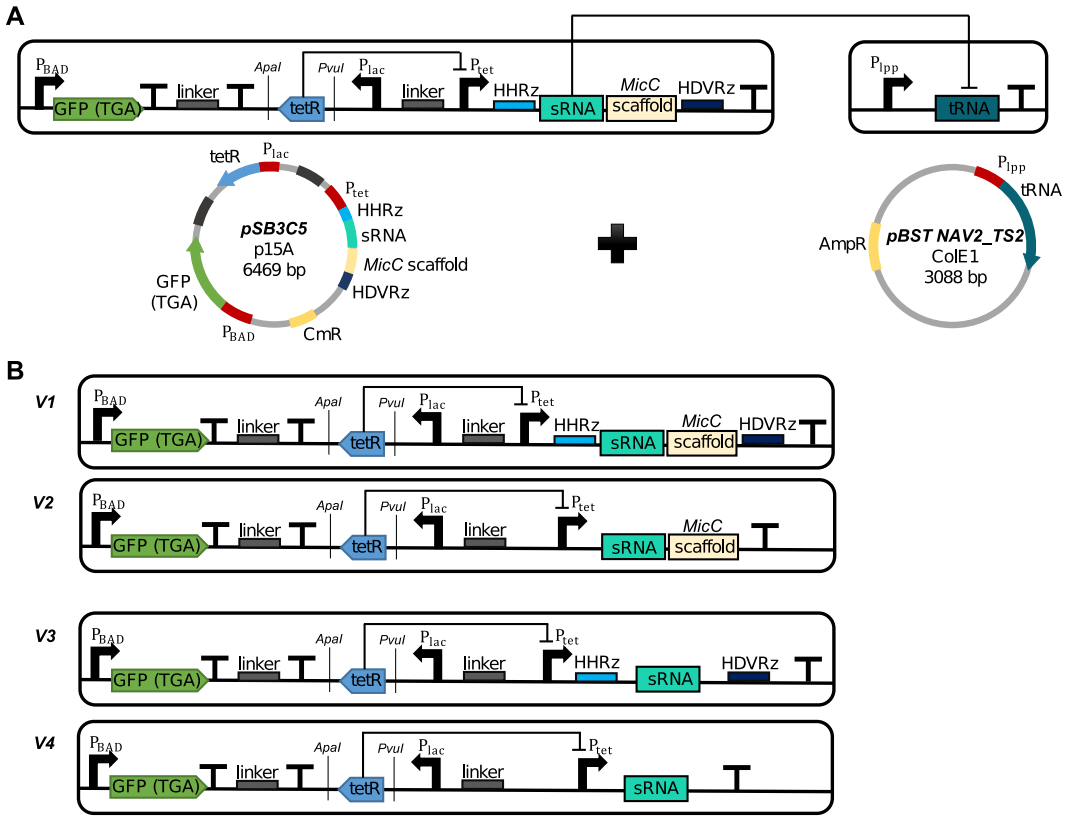


Figure 5.10: Plasmid designs for circuit expression and characterization. (A) Main plasmid for *in vivo* implementation of the circuit. **(B)** Main plasmid can altered into four different versions by using specific pairs of deletion primers. V1 is the main design, V2 can be obtained by deleting the ribozymes from V1. Plasmid V3 can be obtained by deleting the *MicC* scaffold from V1, and V4 by deleting the ribozymes from plasmid V3.

The eukaryotic version of the miR509 sensor (Figure 5.12), contains the Kozak sequence instead of the RBS, and the design follows the principles of the prokaryotic design. The sequence of the eukaryotic miR509 RNA sensor is (5′- CUACCCACAGACGUACCAAUCAUUC AAGAACAAGCCACCAUGUGAUUGGUACGUCAAGAACAGUCAAGAACAG -3′). The sensor is composed of the reverse complement sequence of miR509 (22 nt), immediately followed by the loop sequence, then the complementary sequence that forms the stem of the loop (11 nt identical to miR509 sequence) and a linker sequence. Secondary structure predictions were performed also at 37 °C, since mammalian cell growth is most efficient at this temperature [280].

Results of thermodynamic simulations showed that the structure of the sensor is predicted to be sufficiently stable (Figure 5.12A) in the OFF state, the free energy change (ΔG) of the miR-sensor duplex on the ON state (Figure 5.12B) is lower (-45.88 kcal/mol) and therefore, more probable of occurring in nature than the OFF state (-21.20 kcal/mol). The MFE structures, free energy changes and ideal helical predictions were calculated with NUPACK [92]. The predicted helical structures confirmed the expected formation of correct secondary structures on both the OFF and ON states (Figure 5.5AB).

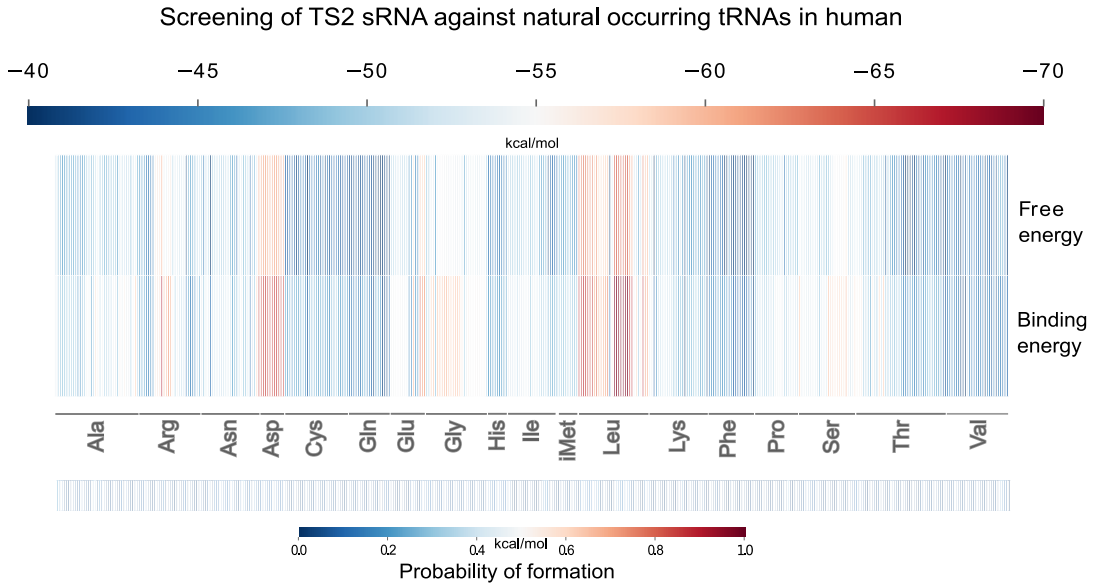


Figure 5.11: Screening sRNA against natural occurring human tRNAs. The binding affinity between the synthetic sRNA was calculated for human tRNAs. For further verification, the free energy of the structure of the sRNA-tRNA duplexes was checked as well as the probability of formation: in summary, the one with the least probable (0.004) intervention would be Ile-TAT-1-1. The one with the highest probability (0.322) is Ala-AGC-3-1, followed by Pro-TGG-2-1 (0.270). Regarding the stability of secondary structures, the most stable would be the Leu variants, with values around −69 kcal/mol. Full data set in Table A.6.

Simulations of the reaction network for the eukaryotic sensor were also performed (Figure 5.13). The sensor was encoded with long domains and only the AUG codon as a short domain. The results corroborated the strand displacement reactions, which should destabilize the hairpin structure and enable translation. Simulation parameters were as described earlier.

5.4 Conclusion

A circuit for diagnosis and therapy of nonsense mutation-associated diseases is proposed. The diagnostic module is comprised by a toehold switch that senses a miRNA which serves as biomarker for a specific disease. The therapy module comprises nonsense suppressor tRNAs controlled by synthetic sRNAs. A proof-of concept is designed for CF. Two RNA-based sensors for miR509, are proposed for the diagnosis of CF, one for implementation in prokaryotic cells and the other for eukaryotic/mammalian cells. Thermodynamic analysis and reaction simulations show that the implementation is viable for both organisms.

Furthermore, synthetic sRNAs are designed and proposed to switch the state (ON/OFF) of therapeutic molecules, namely nonsense suppressor tRNAs. The sRNAs are analysed for their implementation in both *E. coli* and human cells, results indicate that in theory they would represent no threat for their usage in neither organism.

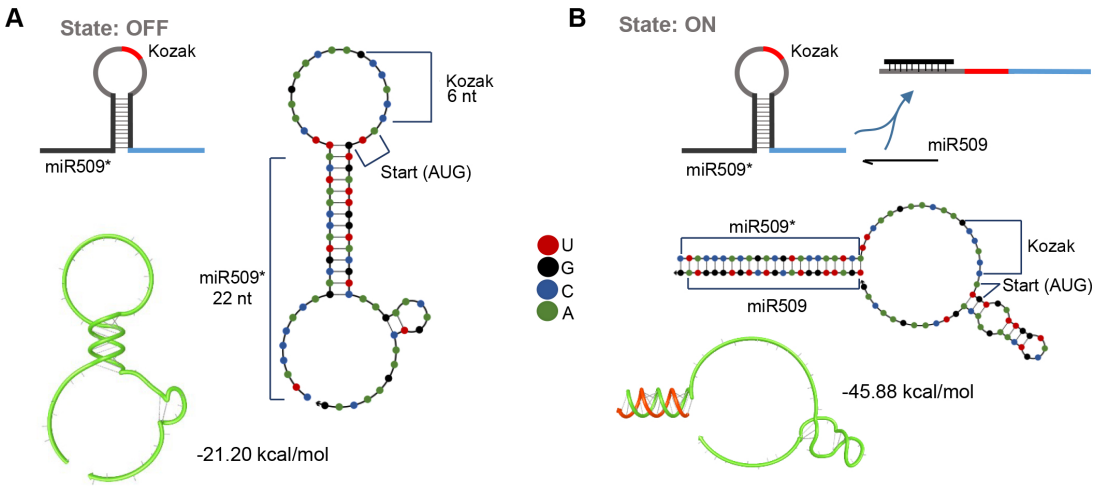


Figure 5.12: Scheme of the design and thermodynamic analysis of the RNA-based miR509 sensor (eukaryotic). Secondary structure evaluation of the RNA-based sensor in both ON and OFF states, identity of the bases in color. Prediction of the ideal helical geometry for both states, mRNA in green; miR509 in orange. **(A)** OFF state (default): while miR509 is not present, the secondary structure of the sensor prevents the ribosome from docking on the Kozak sequence and translating the TF that further controls the tRNA production. **(B)** ON state (miR509 detection): when miR509 binds to the sensor, the secondary structure is disrupted and translation of the gene is possible, consequently allowing the next steps for the tRNAs to function.

The outline for the circuit implementation in *E. coli*, is discussed, plasmids for its expression are designed and proposed, including several designs that will allow for the characterization of the system. Additionally, it is shown that at least with the experimental conditions that have been tested, the therapeutic tRNAs do not function in BL21 *E. coli* strain, while their application in XL1 blue strain is more promising.

In theory, the circuit is functional, however, assembly of the plasmids has to be performed and characterization experiments have to be carried out to elucidate its suitability for *in vivo* applications.

5.5 Materials and Methods

5.5.1 Design of synthetic sRNAs targeting tRNAs

Sequences were obtained using in-house written scripts in Python. Binding energy calculations and 2D structure schemes were produced using RNAcofold from the ViennaRNA package [86] Python bindings. The sRNA sequence was screened for "off-targets" against Escherichia coli BL21-Gold (DE3) genome (Tax-id:866768), using RNAPredator [94], and only selected if no significant results were found.

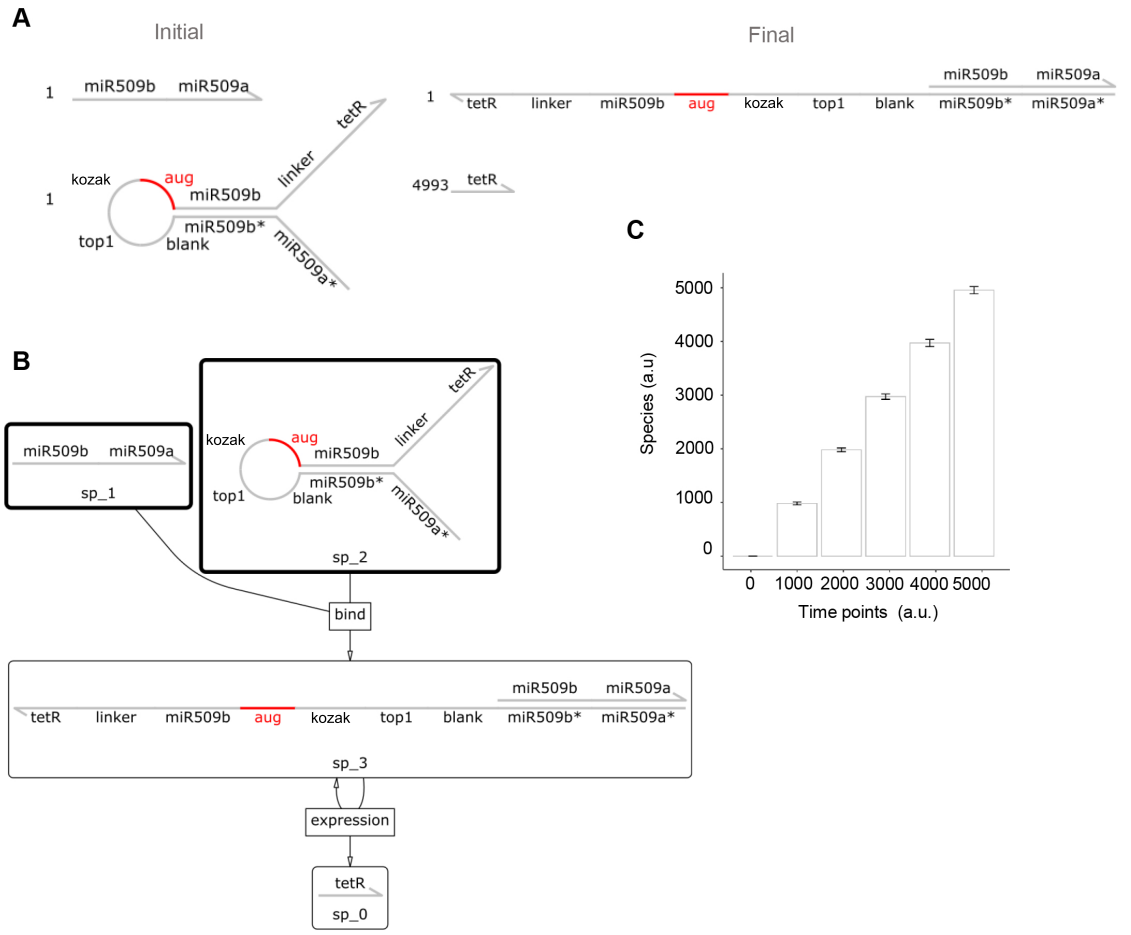


Figure 5.13: Chemical reaction network species and simulation results for miR509 sensor (eukaryotic). (A) Initial and final (molecule) species of the reaction network; miR509 sensor and miR509 as initial species; the translated tetR gene (TF-protein) and the sensor-miR509 duplex as final species. **(B)** Chemical reaction network. **(C)** Results of the simulations over 5000 time points (n=30, data are means ± s.d.).

5.5.2 Thermodynamic simulations

Minimum Free Energy (MFE) of the tRNA-sRNA duplexes secondary structures, as well as their probability of formation were calculated by means of NUPACK [92]. Parameters for the simulations as follows: RNA energy parameter from Serra and Turner (1995); temperature was set to 37° in correspondence to the optimum for *E. coli* in vivo experiments; values for Na^{+} (1.0 M) and Mg^{++} (0.0 M) set to default.

5.5.3 Reaction network simulations

Simulations of the reaction network for the sensor were performed with the Visual DSD language and its web-based compiler [278]. This logic programming language allows for simulating and verifying the behavior of nucleic acid computational devices. To perform the simulations, the strand graph predicates in [278] were taken as a basis for the logic program code, where binding, unbinding, displacing and protein translation predicates were modeled, and modified accordingly to fit the sensor

design. The final time was set to 5000 time points, the toehold binding and unbinding rates 0.0003 and 0.1126 $nM^{-1}s^{-1}$, respectively.

5.5.4 Flow cytometry

Function verification of the nonsense suppressor tRNAs was carried out in recombination (recA) deficient *E. coli* XL1-blue cells. Overnight cultures in LB-medium of co-transformed cells, and antibiotics (Ampicillin and Chloramphenicol) were grown. LB-medium was inoculated with 1:10 overnight culture for co-transformed cells, and 1:50 for single transformations (corresponding to the controls). L-arabinose induction was performed at OD_{600nm} = 0.4 with 0.25% mL from a 20% stock solution. Cultures were harvested at OD_{600nm} = 1.0 and cell pellets washed with 500 µL 1x PBS. Samples were analyzed using a Becton Dickinson FACSCalibur flow cytometer for a total of 100,000 events with the following settings: FSC=E01, SSC=400, FL1=736, in log scale, and the following threshold: FSC=52. The analysis and plots were performed using Flowing Software version 2.5.1.

Chapter 6

Towards theranostic cells for the mitigation of cystic fibrosis

The emerging field of theranostics, especially explored in cancer research, involves the development of systems capable of simultaneously integrating diagnose and therapy [281] (synthesis and delivery of novel therapeutics *in vivo*). In pursuit of escalating the mitigation of diseases and diminishing systemic side effects observed in traditional medicine approaches, and thus allowing for individual targeted medicine [282].

The feasibility of designing and implementing theranostic molecular systems for the mitigation of cystic fibrosis (CF) is explored within this chapter.

6.1 miRNA-based theranostic model for CF

A miRNA-based model for a theranostic approach powered by a molecular circuit against CF was proposed (Figure 6.1). The diagnostic module corresponds to a molecular circuit that solves a Boolean expression (or diagnostic rule) which describes a positive indication of CF. In principle, this rule can be exchanged for the model to be as simplified or a specific as desired: two rules were proposed, a simplified model, and a more specific one, by discarding a COPD diagnosis. Biomarkers (miRNAs) of the disease are the inputs, and the output of the circuit consists of the therapeutic module, the synthesis/administration of either a miRNA mimic or antagonist, in this case.

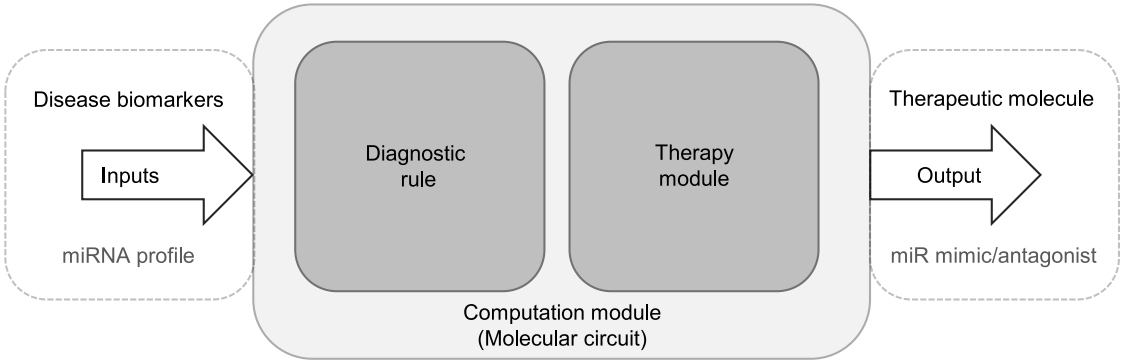


Figure 6.1: Theranostic model for cystic fibrosis. On the detection of specific biomarkers of the disease (miRNA), the diagnostic module (molecular circuit) activates the therapy module (administration of miR-based therapy).

6.1.1 Diagnostic rules for CF

A simplified model (diagnostic rule) of cystic fibrosis disease can be formulated by logically describing (by means of a Boolean expression) the profile of its biomarkers (variables). Boolean expressions representing positive diagnosis for simplified models of prostate [283], breast [284] and small-cell lung [283] cancers have been proposed to be used in combination with molecular devices and circuits. In this chapter, simplified diagnostic rules for CF are proposed.

6.1.1.1 miR-based rule for CF diagnosis

A CF diagnosis is expected on the presence of *i)* a CF highly specific biomarker and, either or both *ii)* an inflammation or *iii)* infection indicators (Figure 6.2):

CF= CF highly specific miRNA `AND` (inflammation indicator miRNA `OR` infection indicator miRNA)

Based on the literature on miRNAs in CF (see Chapter 4), the variables were substituted as:

CF= miR-509 `AND` (miR-155 `OR` miR-215)

miR-509-3p has been shown upregulated in CF and its relation to COPD is questioned [234], suggesting it is a specific indicator of CF. miR-155 is overexpressed in CF and induces the pro-inflammatory phenotype in CF by hyper-inducing IL-8 levels [240, 242]. miR-215 is suggested as a biomarker of chronic infection [239] since it was shown upregulated in CF cells infected with *P. aeruginosa*.

6.1.1.2 Differential diagnosis: CF vs COPD

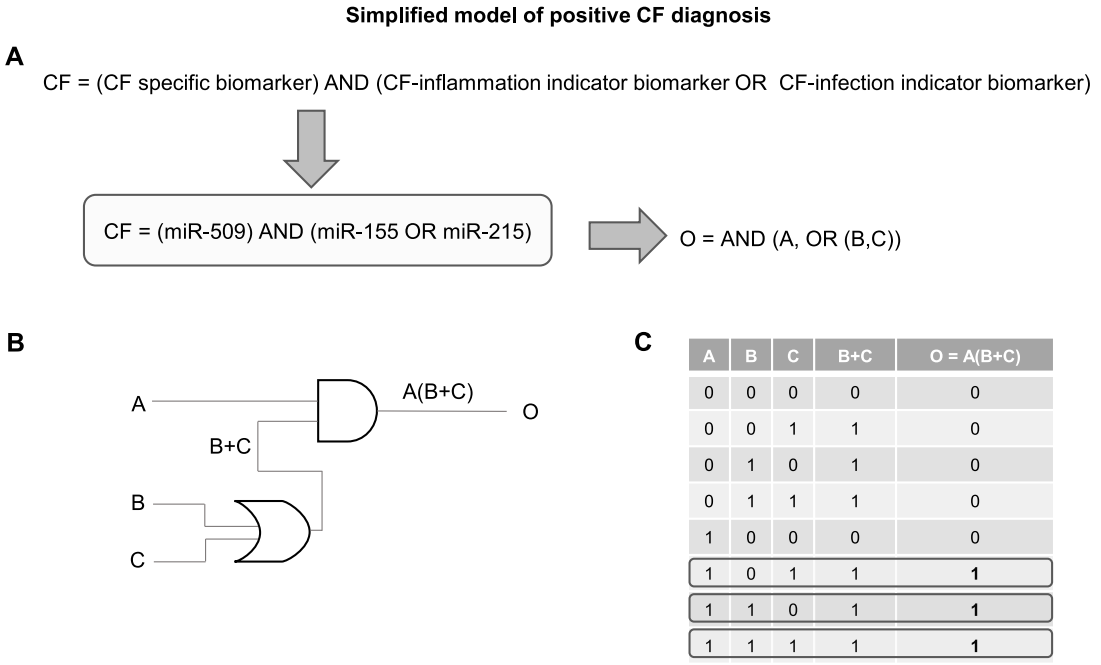
A more specific CF diagnosis differentiating from COPD condition is expected on the presence of *i)* a CF highly specific biomarker and, either or both *ii)* an inflammation or *iii)* infection indicators in CF; and the presence of *iv)* a COPD specific indicator has been ruled out (Figure 6.3):

CF= CF highly specific miRNA `AND` (inflammation indicator miRNA `OR` infection indicator miRNA) `ANDNOT` COPD specific miRNA

Hence, using miRNAs as variables, the rule resulted as follows

CF= miR-509 `AND` (miR-155 `OR` miR-215) `ANDNOT` miR-212

A study found miR-212-5p as the only significantly upregulated (other 11 miRNAs were downregulated) in the miRNA profile of lung tissue samples from COPD patients [285]. Furthermore, to the best of our knowledge, no relation to CF has been reported, therefore suggested as a specific indicator of COPD disease.



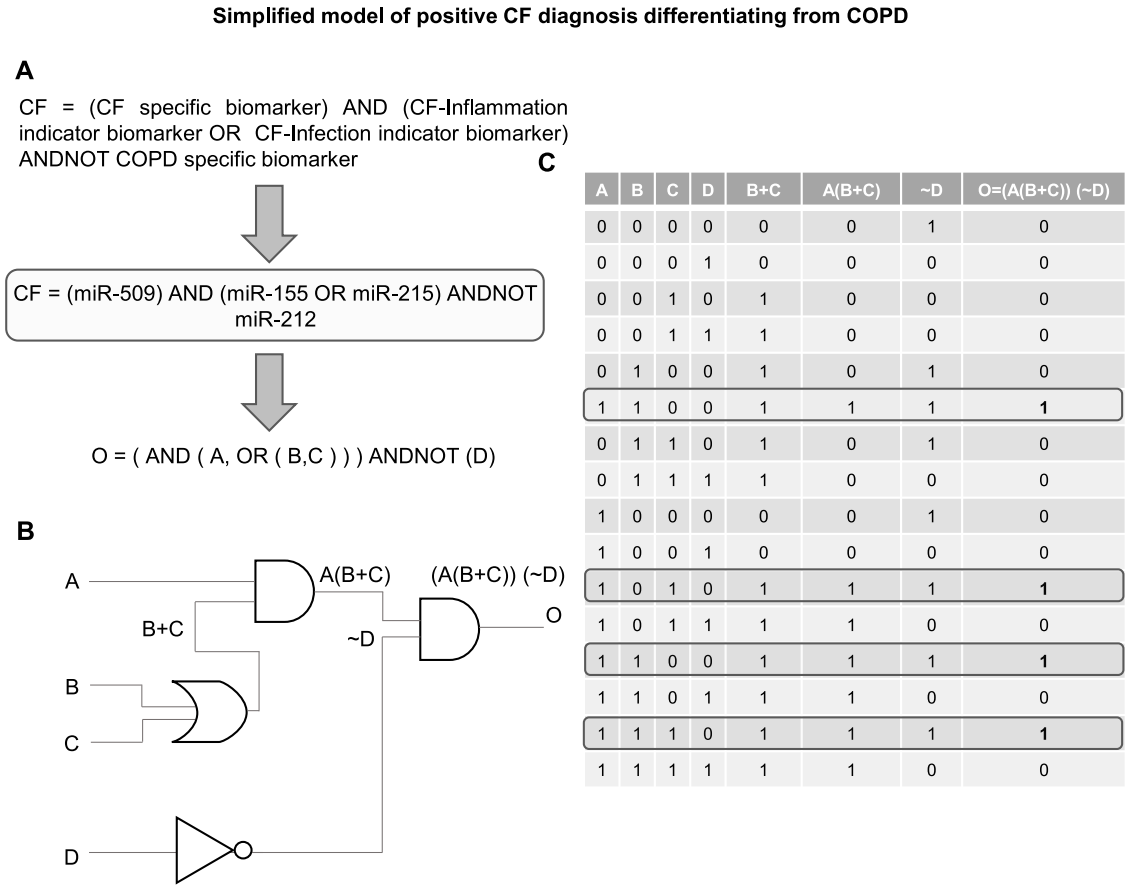
A	B	C	B+C	O = A(B+C)
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	1	0
1	0	0	0	0
1	0	1	1	**1**
1	1	0	1	**1**
1	1	1	1	**1**

Figure 6.2: Simplified model for cystic fibrosis diagnosis: **(A)** diagnostic rule and its corresponding Boolean function; **(B)** digital circuit diagram of logic gates; **(C)** truth table defining the operation.

6.1.2 miR therapeutics for CF

In targeted therapy of diseases, RNAi might be advantageous compared to DNA-based approaches: RNAs (e.g. miRNAs, siRNAs) only need to be delivered into the cytoplasm, whereas DNAs have to be delivered into the nucleus; this might be less prone to trigger cellular innate immune responses [259, 286]. Recently, novel miR-based therapeutics (e.g. miR-126, miR-9, and miR-16) against CF have been proposed.

miR-126 is downregulated in CF airway epithelial cells [238]. A study administered a mimic enclosed in nanoparticles to investigate the efficient delivery of miRNA therapeutics into CF airway epithelial cells [259]. Their results suggest the administration of the miR-126 mimic within PEI nanoparticles helped increased miR-126 levels and reducing the expression of TOM1 (a negative regulator of IL-1β; decrease in IL-1β might correct neutrophil influx into the CF lung [287]). However, the direct effect on CFTR was not measured.



A	B	C	D	B+C	A(B+C)	~D	O=(A(B+C)) (~D)
0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0
0	0	1	0	1	0	0	0
0	0	1	1	1	0	0	0
0	1	0	0	1	0	1	0
1	1	0	0	1	1	1	**1**
0	1	1	0	1	0	1	0
0	1	1	1	1	0	0	0
1	0	0	0	0	0	1	0
1	0	0	1	0	0	0	0
1	0	1	0	1	1	1	**1**
1	0	1	1	1	1	0	0
1	1	0	0	1	1	1	**1**
1	1	0	1	1	1	0	0
1	1	1	0	1	1	1	**1**
1	1	1	1	1	1	0	0

Figure 6.3: Simplified model for cystic fibrosis differential diagnosis: (A) differential diagnosis rule and its corresponding Boolean function; **(B)** digital circuit diagram of logic gates; **(C)** truth table defining the operation.

miR-9 is overexpressed in CF cells, however since it targets multiple miRNAs, an antagonist or inhibitor is not recommended [235]; authors designed a specific target site blocker (TSB) that binds to the ANO1 3' UTR (ANO1 TSB) to prevent miR-9 binding, therefore enhancing ANO1 chloride channel activity. Their findings propose ANO1 TSB as a promising therapy that is hypothesized to work for all CFTR mutation classes.

miR-16 aids in the functional rescue of the cyclic AMP-activated apical F508del-CFTR chloride channel in primary lung epithelial cells from CF patients [236] by expressing a miR-16 mimic in the CF lung or pancreatic cells, the expression of the F508del CFTR protein increased significantly.

6.1.3 Molecular circuits as theranostic modules

In principle, Boolean expression diagnostic rules can be solved by adapting any complete set of modular molecular logic gates (e.g. genetic circuits [288], DNAzymes [145], ribocomputing devices [19], DNA strand displacement [16], RNA-based logic [289]). After the computation, if the diagnostic is positive, a therapeutic miRNA mimic could be synthesized/released as the therapeutic approach (output of circuit/device).

6.2 Towards implementation with strand displacement-based circuits

Our previous modular logic gates [290] based on nucleic acid strand displacement (DSD) were proposed as computing module (Figures 6.4 and 6.5). The design conventions (Figure 6.4AB) of our previous model allowed for the gates to interconnect, and by means of cascading reactions to assemble logic circuits. The variables (inputs and output) had the same representation: two long domains ($\sim$ 20 nt) flanked by toeholds (4-10 nt) on each edge (Figure 6.4A). The presence of the variable denoted true (1), and the absence of signals a false (0) outcome. A logic gate (Figure 6.4B) comprised of two complexes: a computation module and a processing module. The rationale for this is that depending on inputs combination, the processing strand (pre-hybridized in the computation module) is released or not. The processing strand is completely complementary to the processing complex, thus, while hybridization, it displaces the output variable that was pre-hybridized to the complex.

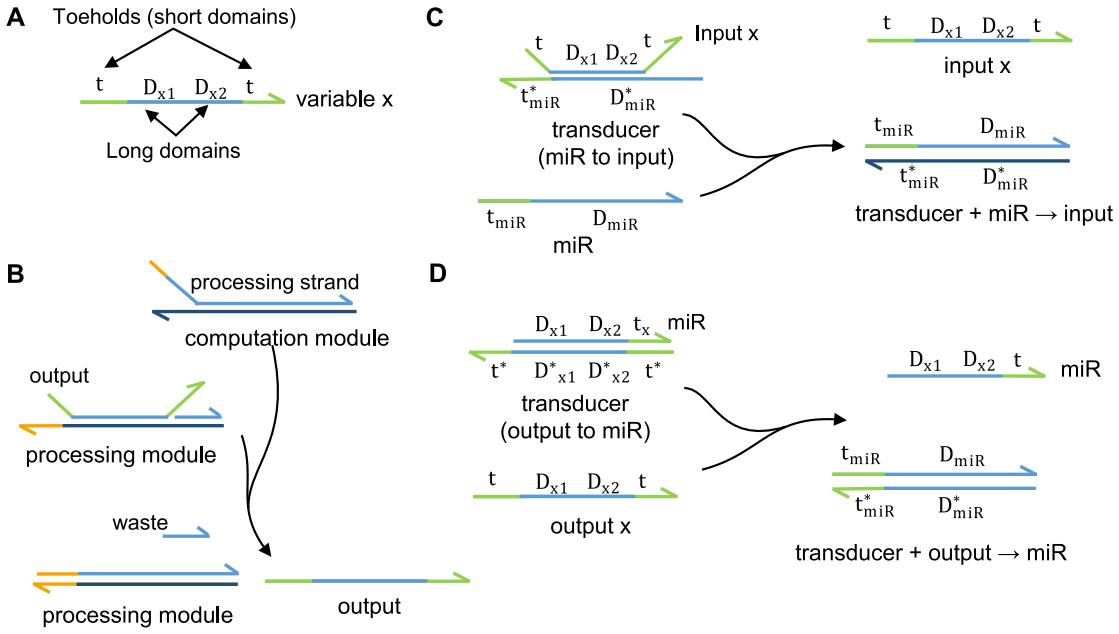


Figure 6.4: Conceptual design of DSD circuits. **(A)** Variables (inputs and outputs) hold the same convention for compatibility in cascading reactions, two long domains flanked by toeholds; **(B)** each logic gate is comprised of two complexes: a computation and a processing module (adapted from [290]). Transducer gate rationale for converting from **(C)** miRNA to variable (input) and from **(D)** (output) variable to miR.

A conceptual design of a strand-displacement-based circuit for solving the simplified diagnostic rule for CF (Section 6.1.1.2) is outlined (Figure 6.5). For the computation, it is necessary to convert the miRNAs signals into variables compatible with the conventions of the system. We propose using transducer gates (Figure 6.4C). In a similar manner, the output variable after computation can be used to release a miR pre-hybridized with the transducer gate complex (Figure 6.4C). The circuit was composed by two gates: AND and OR; following the simplified diagnostic rule for CF (see Section 6.1.1.2), input A corresponds to miR-509, input B to miR-155, and input C to miR-215.

An example for a positive diagnosis of CF (miR-509=A and miR-215=C present) is illustrated (Figure 6.5). The first computation corresponds to OR (miR-155=B, miR-215=C). Since it is sufficient if one input is true for the output to be true (see Figure 2.12), the output (B+C) of this intermediate calculation is released (true). Input C displaces the processing strand, which in turn hybridizes to the processing module complex, releasing the output strand (input B+C). The next calculation is AND (miR-509=A, input B+C), and for this operation to return a true output both inputs have to be true. In this case, input B+C resulted from the previous cascading reaction, therefore both inputs can react with the processing module of the AND gate, and release the intermediate processing strand. This strand is complementary to the processing module complex to release the final output (true), indicating a positive diagnosis of CF. As discussed earlier, this output signal could be paired with a transducer gate to further release a miRNA mimic or another molecule that triggers a therapeutic approach in response to the diagnostic outcome.

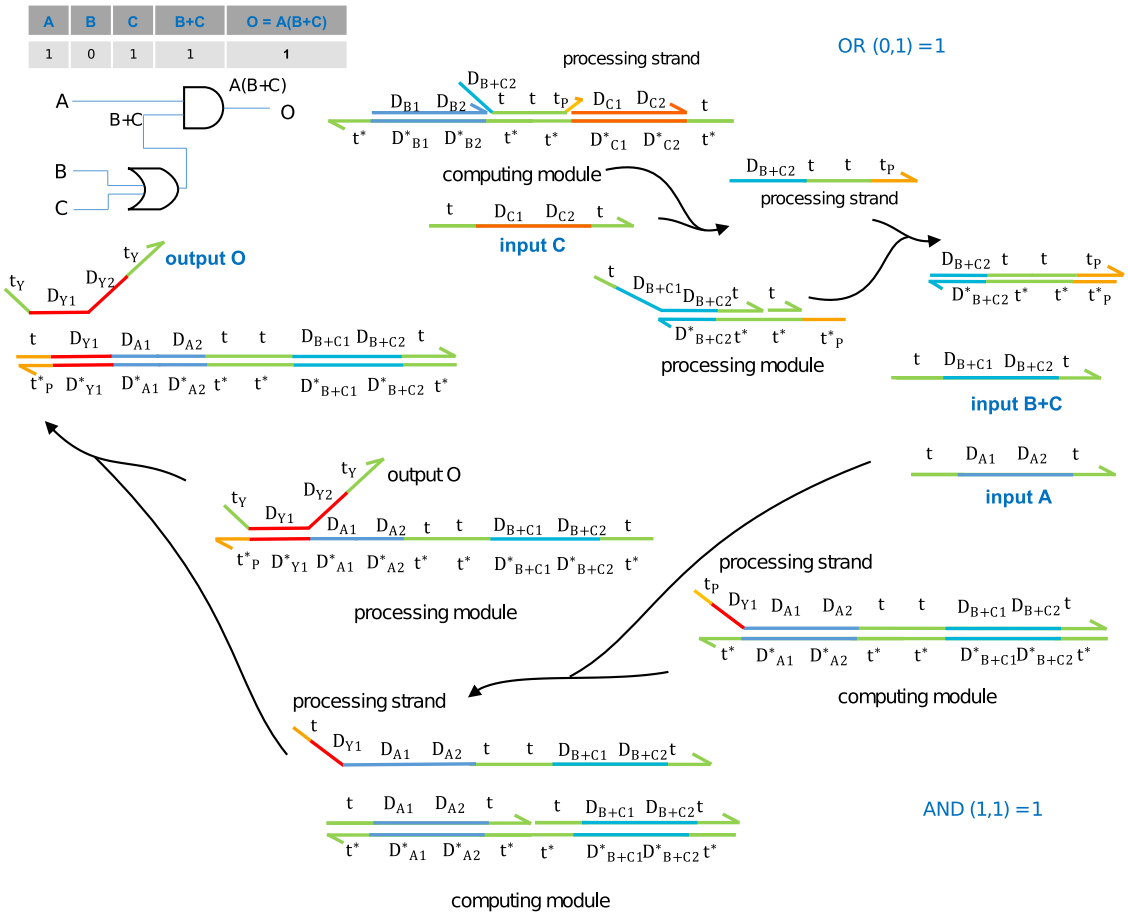


Figure 6.5: Conceptual example of the simplified CF diagnosis model solved by DSD circuits. Reaction pathway of a positive diagnosis (when input A= miR-509 and input C=miR-215 are present) of CF using simplified diagnostic rule (Section 6.1.1.2) solved with strand displacement-based modular logic gates [290].

Chapter 7

Ribosomopathies and the heterogeneity of ribosomes

Ribosomopathies are disorders that result in clinical phenotypes that occur when genetic abnormalities cause defects in ribosome biogenesis and function [291]. They are classified mainly into two broad groups: caused *i)* by single-copy mutations in specific ribosomal proteins, or *ii)* by defects in ribosome biogenesis factors. Ribosomopathies manifest clinically as tissue-specific disorders (e.g. Diamond-Blackfan anemia, Dyskeratosis Congenita, and Treacher Collins syndrome) [292]. These disorders have been linked to ribosome heterogeneity, hypothesized to be given by alterations of ribosomal proteins (RPs) or changes in their composition [25]. The detection of fluctuations of RP composition with possible functional implications remains a challenging task, and this has been addressed in our work [34].

In this chapter we summarize the results obtained from the analysis of data coming from mass spectrometric proteome quantitation of translating ribosomes (separated monoribosomes and polysomes) from four different mouse (with different ages) tissues (cerebellum, hippocampus, cortex and liver) (Figure 7.1A). The mass spectrometry proteomics data can be found in ProteomeXchange Consortium via the PRIDE [293] partner repository with the dataset identifier PXD014138. The main questions that we address are: do ribosomes indeed differ in protein composition, is there a difference depending on the tissue and/or over the life-span? Results along with the analysis are published in [34].

7.1 The ribosome heterogeneity hypothesis debate

The molecular composition of the ribosomes: ribosomal RNA (rRNA) and ribosomal proteins (RPs), varies from bacteria to eukaryotes and among organelles, however the overall architecture is preserved and the functional unit of decoding and peptide bond formation are both conserved [45]. Along the phylogenetic tree, the ribosomes have evolved by idiosyncratic expansion of the RPs and rRNAs without perturbing the conserved functional core [294].

Mammalian cells typically contain $\sim 10^6$ cytoplasmic ribosomes and the question arises as to whether various types of cells maintain homogeneous pools of ribosomes with the exact same stoichiometry of the individual RPs. Expression of functionally related genes consisting functional modules, e.g RPs constituting the ribosome, is conserved at great evolutionary distances by maintaining co-regulation while the regulatory mechanisms controlling single genes can

diverge [295]. Across human tissues, the mRNA levels of RPs span several orders of magnitude [296] and their mRNA levels poorly correlate with the protein levels [297], implying additional regulation at the level of translation of the RP transcripts. In addition to their canonical role in translation, many RPs exhibit extraribosomal function (reviewed in [298]) which also suggests disproportional quantities of RPs. The half-life of single RPs substantially differ for each individual RP [299], but also show variations among tissues [300]. On average, the protein lifetime is three-times shorter in liver than in brain [300]. Various tissues also exhibit different transcript dynamics over age. In brain tissues of *C. elegans* or the short-lived fish *Nothobranchius fuzeri*, the RP mRNA levels increase with age, while in other tissues they decrease or remain constant [301–303]. In addition, individual RPs are exchanged at different rates with the cytoplasmic pool of their free counterparts [304, 305].

In yeast, RP genes are generally duplicated and incorporation of RP paralogs results in compositional heterogeneity of the ribosomes [306]. Genomes of humans and mammals encode multiple copies of the four rRNAs [307, 308], and unlike yeast, most of the 79 RPs are typically encoded by a single gene. From the eight total human paralogs [296, 309] few exhibit tissue-specific patterns of transcription, e.g uL3-like is expressed in striated muscle [310], uL16-like and eL39-like in testis [311], supporting the hypothesis of tissue-specific variations in ribosomal composition. However, the depletion or even complete deletion of some paralogs does not cause phenotypic effects, challenging the role of the paralogs as a unique component of the ribosomes in those tissues [312, 313]. Compositional heterogeneity of ribosomes could have message-specific effects, i.e. subpools of ribosomes might be specialized in translation of specific mRNAs, a hypothesis referred to as specialized ribosomes. Ribosomes deficient of a single RP, such as eL38 or uS2 [314, 315], posttranslational modifications [316, 317], or detachment of ribosome associated-factors [318] provide examples for tissue- and message-specific expression prompting the debate whether a specialized ribosome pool may shape tissue development [319]. This argument is challenged by a recent observation in *C. elegans* that tissue diversification is efficiently maintained during embryonic development by maternal ribosomes and synthesis of new distinct pool of ribosomes is not required [320]. The argument for heterogeneous ribosomal pools conferring message-specific expression are used to explain the tissue specificity of haploinsufficiencies associated with germline mutations disrupting one copy of RP genes in a type of ribosomopathies [306, 319]. Opposing views suggest instead that tissue-specific dosage is shaped by affecting auxiliary factors or ribosome concentration relative to mRNA levels which can alter translation of poorly initiated messages [291, 321, 322].

7.2 Results

7.2.1 Stable RP stoichiometry across various tissues

Measurements of RP composition of mouse embryonic stem cells and yeast suggest differential stoichiometry between monoribosomes (80S) and those engaged in polysomes [323], hence we separated the 80S monoribosomes from the polysomes (Figure B.1) and separately quantified their RPs by tandem mass spectrometry (Figure 7.1A). In total, 1,163 proteins met the requirements for quantitative comparison between samples at false-discovery rate FDR < 0.01. The numbers of ribosomes vary by an order of magnitude across tissues [309]. Thus, it is important to note that our

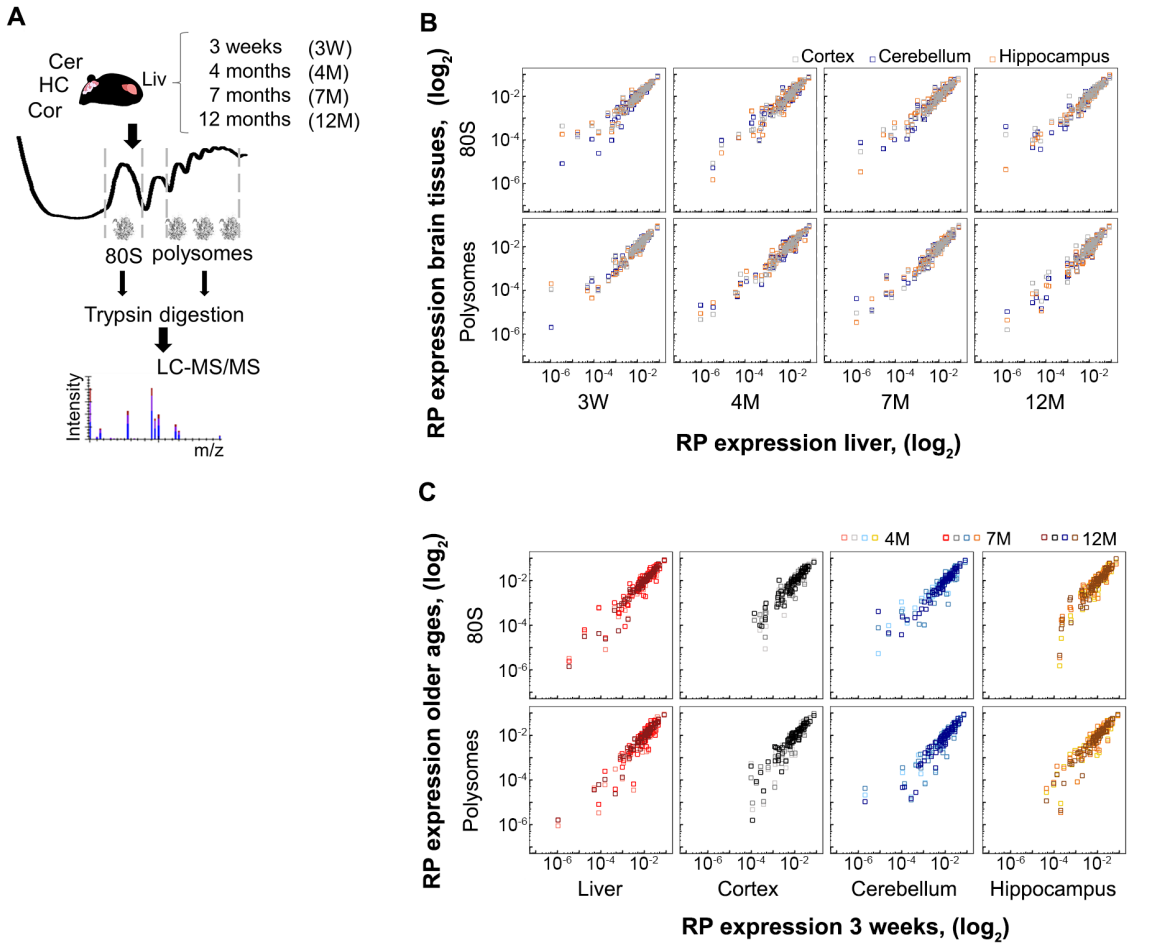


Figure 7.1: RP expression over tissues and ages. (A) Overview of the approach. Cer, cerebellum; HC, hippocampus; Cor, cortex; Liv, liver. Expression levels of RPs in polysomes and 80S of **(B)** brain tissues compared to liver for each age group and **(C)** within each tissue across ages compared to mice at 3 weeks of age. Pearson correlation coefficients are summarized in Table B.3. W, week; M, month (published in [34]).

analysis does not provide any absolute quantitative measures on the ribosome number or protein copy number in tissues. Normalization of the level of each RP to the total amount of detected RPs provides means for assessing the relative amount or fraction of each RP [323] and thus enables exploring quantititative variations among RPs in different tissues.

The mass spectrometric proteome analysis achieved high coverage of the RPs: 72 of 79 core RPs, RACK1 and two of the eight RP paralogs (eL22-like and eS27-like) were faithfully quantified in all samples (Table B.1). Missing RPs (uS19, eS27, eL29, eL34, eL37, eL40) were either localized on the ribosomal surface (hence they might have been lost during isolation) or following trypsinolysis they delivered peptides that were too short (eL41) and thus escaped detection. Strikingly, we detected the identical set of the same 72 core RPs, RACK1 and 2 paralogues (in total 75 proteins) in the polysomal fraction across all tested tissues, e.g three brain tissues and liver, within each age group, and the RPs expression correlated well (Figure 7.1B). Furthermore, the ribosomes of the 80S and polysomal fractions were remarkably similar, and across all tissues we identified the same set of 75 proteins in both the 80S fractions and in the polysomal fractions (Figure B.2). The RP

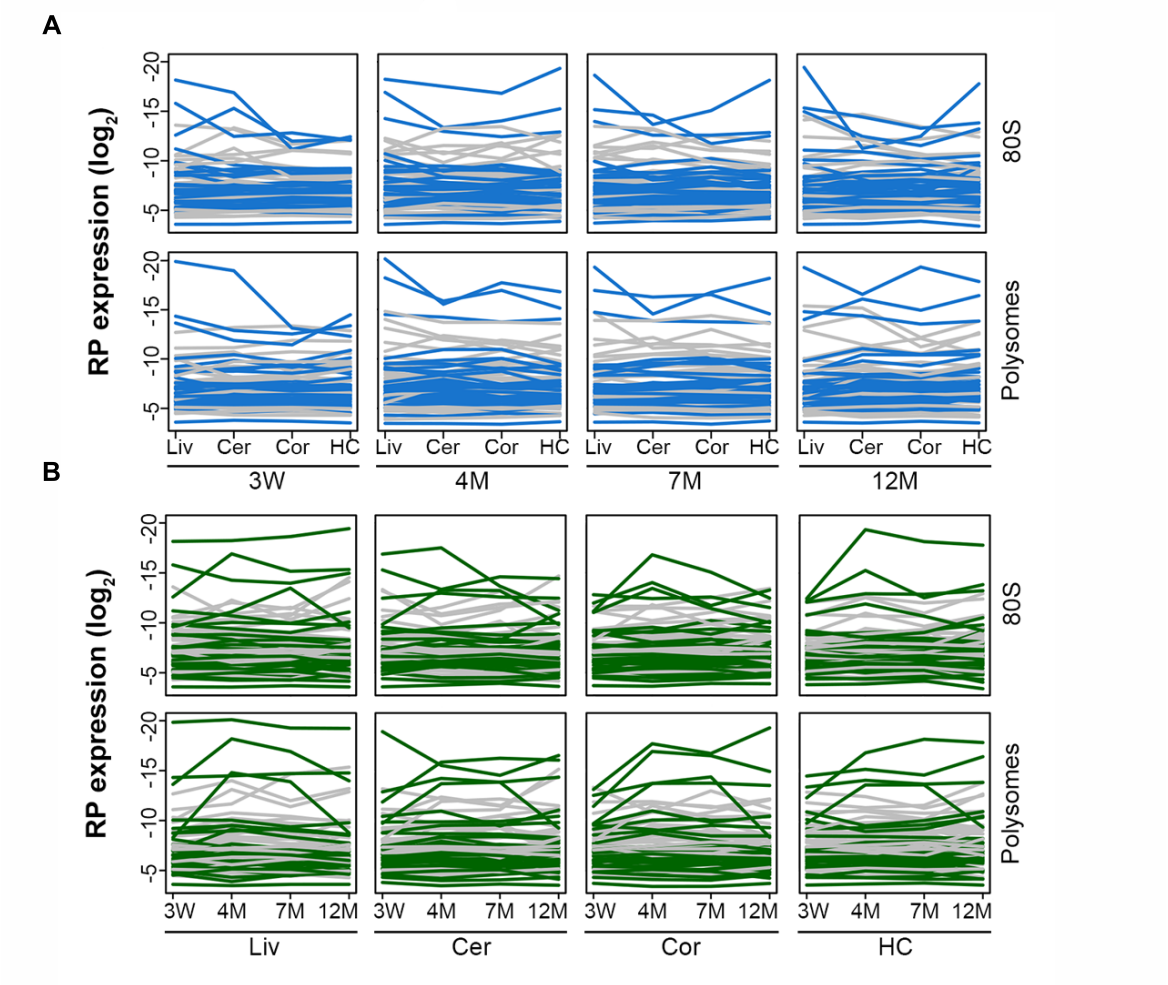


Figure 7.2: Changes in expression of single RPs across tissue and age are not significant. Comparison of the RP levels in tissues **(A)** of the same age **(B)** or across ages within the same tissues. Gray, proteins of 60S; green or blue, proteins of 40S. Despite fluctuating expression, none of the proteins show significant changes in their amounts (FDR < 0.1) in the respective sample or population. Cer, cerebellum; HC, hippocampus; Cor, cortex; Liv, liver; W, week; M,month (published in [34]).

levels spanned an order of magnitude with overall excellent reproducibility between independent biological replicates ($R^2 \geq 0.84$; Figure B.3 and Figure B.4, Table B.4), despite the higher intrinsic heterogeneity of an outbred mouse strain. Low abundance RPs (the lower one-third of abundance levels) had higher contribution to the scattering (Table B.4).

Although globally the RP levels among tissues correlated very well (Figure 7.1B), single proteins may oppose this trend. Thus, we plotted the amount of each single RP across tissues (Figure 7.2A). Some RPs saw much higher fluctuations across tissues, and, using Student's *t*-test, we selected 11 RPs in the 80S and 14 in the polysomal ribosomes with marked changes in at least one tissue. However, they all scored "insignificant" at FDR < 0.1 (Figure B.5). Together, our data suggest that across the four different tissues that we analyzed, the RPs set was identical in both the polysomal and monoribosomal fractions. The RP amounts were remarkably similar among tissues and the fluctuations remained within tight bounds of the measurement noise.

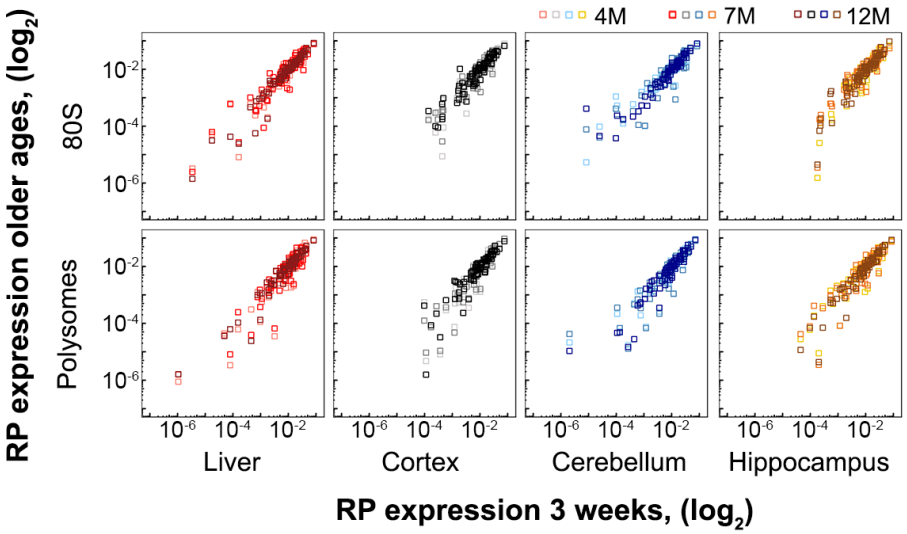


Figure 7.3: RP stoichiometry does not significantly change in adult ages. Expression levels of RPs in polysomes and 80S for each tissue compared to 4 month of age. M, month. Pearson correlation coefficients are summarized in Table B.5 (published in [34]).

7.2.2 Invariable RP stoichiometry with aging

Earlier reports suggest that aging is accompanied by progressive decline in concentration of polysomes in skeletal muscles [324] and translation activity of ribosomes from liver [325] and brain [326] decreases in an age-dependent manner. Also, genetic experiments support the view that loss of particular RPs affects the rate of protein synthesis (reviewed in [327]). To assess changes in RPs amount with aging, we compared four age groups with different age characteristics, i.e. juvenile mice of three weeks which are still in development, young adult mice of four months and two groups of middle-aged adults of seven and twelve months, respectively (Figure 7.1A). Each of these age groups in mice corresponds to different ages in human, e.g < 10 years, $\sim$ 23, 33 and 43 years, respectively [328]. In our analysis we used female mice; under same calorie restrictions the female murine brains undergo age-related changes much earlier (6 to 9 months of age) than male brains (9 to 12 months of age) [329].

Across different age groups we detected the same 72 of 79 core RPs, 2 paralogues and RACK1 in both 80S and polysomal fractions (Figure 7.1C and Table B.1) which we detected across tissues (Figure 7.1B) implying that the core RP composition of the ribosomes does not change with aging. Globally, the RP levels in 80S and polysomal fractions of the tissues from the adult age groups correlated well with those of the three week-old mice (Figure 7.1C). After plotting the level of each single protein at different ages (Figure 7.2B), in some age groups we observed noticeable fluctuations for some proteins. Using Student's *t*-test we selected two RPs in the 80S and three in the polysomal ribosomes, none of which however scored significant even at FDR < 0.1 (Figure B.6), implying that the fluctuations were within the measurement noise. Importantly, we realized that the amount of the RPs we selected with noticeable changes, only punctually fluctuated in a single age group, but did not follow a consistent pattern of decline or increase with age (Figure B.6).

In our analysis, we compared the RP levels in adult mice samples to that of juvenile mice of three weeks of age (Figure 7.1C). Although the brain development in mice is accomplished at

29.7 days after conception (or 12 days after birth), the three weeks-old mice are within the juvenile age group and still undergoing developmental and growth changes [330]. Hence, we next compared the RPs pattern in seven and twelve month-old mice to that in four month-old mice which belongs to the mature adult group past development, yet unaffected by senescence. Gross comparison of the quantitative levels of RPs in the mid-age groups to that of young, four month-old mice also showed invariant RP stoichiometry with age (Figure 7.3). Again, *t*-test-based selection of single RPs from the 80S or polysomal fraction with marked fluctuations in abundance at least at one age (9 and 22 proteins, respectively) did not confirm those changes to be significant at FDR < 0.1.

In sum, our quantitative proteomic analysis shows that in mice the protein composition of the ribosomes is equal across three brain tissues and RP stoichiometry remains stable with aging in both monosomal and polysomal fractions. The quantity of some proteins may punctually vary, but those fluctuations remain within tight bounds of the measurement noise.

7.3 Discussion

The levels of mRNAs encoding RPs reveals high variation across tissues [303]. Together with the identification of tissue-specific RP paralogs [311], tissues with developmental specialization of ribosomes became an appealing place to look for tissue-specific heterogeneity in the ribosomal composition [331]. Here, we report a systematic quantification by differential mass spectrometric proteomics of RPs in three murine brain tissues and liver across four age groups, juvenile (three weeks), young (four months) and middle-aged (seven and twelve months) adults. Across tissues we detect the same RP sets with constant stoichiometry within the monoribosomes and polysomes which remains stable with aging. Our results align well with observations in other systems. Quantitative measurements of RPs show constant composition of ribosomes across other rodent tissues [311], during human hematopoietic lineage commitment [321] and among various cancer cell lines [332].

The role of RP paralogues in higher eukaryotes might be far more subtle than originally proposed [312, 313] and the compensatory drift model may explain the preserved expression of functionally redundant paralogs [333]. In this model, a dosage-dependent balance constrains the total expression of a protein from redundant functional duplicates (paralogs) [333]. Supportive for this is that we detected ubiquitous expression of two RP paralogs (eL22-like and eS27-like) along the major gene product (eL22 and eS27) in all tested tissues. An earlier study, analyzing RP composition in rodent liver, mammary gland and testis also identified a ubiquitously expressed paralog, eL22-like, in all three tissues [311].

Our steady-state measurements are not predictive on synthesis and degradation rates which is regulated in a tissue- and cell type-dependent manner; proteins with similar abundances exhibit a wide range of turnover rates [300]. Difference in abundance of RP mRNAs across tissues [303] cannot be extrapolated to their protein levels, since posttranscriptional processing actively buffers protein concentration [297]. Protein dynamics is shaped by mRNA and protein half-lives, both of which differ among individual transcripts and proteins but also exhibit a tissue-specific pattern of stability with somewhat higher stability in neuronal tissues [300–303]. Recent study in a genetically well-tractable system, *C. elegans* [320] and earlier observations in Xenopus embryos [334] suggest that tissue diversification may not rely on ribosomes with heterogeneous RP pool. Along that line, another study shows that ribosome levels determine the human hematopoietic lineage commitment,

while the RP sets remain unchanged [321]. Thus, ribosome heterogeneity that can potentially reflect functional differences, may rather stem from other processes than changes in RP stoichiometry. Among these processes could be rRNA variations and modifications [308, 335, 336], alteration of ribosome-associated auxiliary [318, 337] or signaling factors [338], or posttranslational modifications of ribosomal proteins [316, 317, 339, 340].

7.4 Materials and methods

7.4.1 Statistical analysis

Two-tailed Student's *t*-test ($\alpha = 0.05$), Wilcoxon sum rank test, Pearson correlation and false discovery rate (FDR) analysis were used for statistical analysis of the data. The statistical analysis were performed using in-house written scripts in R [341].

Chapter 8

Conclusion

This work proposed molecular circuits for the diagnosis and mitigation of genetic diseases, PTC-related diseases in particular. Research efforts to treat these diseases are important, considering that to date there is no cure. The devastating disease cystic fibrosis (CF) was selected as a case study. Circuit designs were presented together with guidelines for their *in vivo* implementation.

8.1 Novelty of the research

Genetic and molecular circuits present a myriad of novel applications with great potential to address contemporary problems. In this work, four themes were explored: *i)* a *de novo* system for fine tuning of genetic circuit regulation was proposed (Chapter 3); the design of a modular molecular circuit capable of performing simultaneous *ii)* RNA-based diagnosis (miRNA biomarkers and riboswitch sensor) and *iii)* therapy (nonsense-suppressor tRNAs) was presented (Chapter 5); in addition, *iv)* models for engineering CF theranostic cells, based on miRNA biomarkers were depicted (Chapter 6).

A comprehensive literature survey (Chapter 4) surrounding miRNA biomarkers for CF was performed to elucidate which miRNAs could serve as a distinct biomarker of the disease and are able to diagnose it in this matter and also to determine miRNA based therapeutics that could be paired with the molecular devices. This allowed for the design of: *i)* an RNA-based sensor for the diagnosis of the disease (Chapter 5), and *ii)* a proposal of simplified models of the disease (Chapter 6) to be used along with molecular circuitry for its diagnosis. Altogether, this effort will facilitate the design of novel RNA-based approaches or molecular devices/circuits to tackle this disease.

In this thesis, different state-of-the-art elements (i.e. synthetic sRNAs, RNA-based sensors, miRNA biomarkers, nucleic acid-based logic, and nonsense suppressor tRNAs) were designed, characterized and/or adapted together in a modular fashion, to explore the feasibility of rationally designing and building *in vivo* molecular circuits to aid mitigation of CF. The rational design of synthetic sRNAs for modulating genetic circuits (Chapter 3) without altering the original circuits, was proposed and further employed to control an existing nonsense suppressor tRNA.

A modular circuit to diagnose and mitigate CF was presented (Chapter 5): synthetic sRNAs were designed to control an existing nonsense suppressor tRNA, to be adapted as a therapeutic module of the circuit. After determining the most suitable miRNA biomarker to diagnose CF (Chapter 4), an RNA-based sensor was designed to diagnose and mitigate it. RNA-based sensors for diagnosis of CF by miRNA biomarkers were designed for their implementation on both prokaryote and eukaryote organisms. The prokaryotic system serves only as a proof-of-principle model organism due to its

favorably fast growth and high degree of controllability.

To the best of our knowledge, this is the first attempt to design theranostic molecular devices (Chapters 5 and 6) for the mitigation of nonsense mutation-related genetic diseases. This will open new possibilities for employing synthetic biology to advance personalized medicine, and be able to harness cells to perform not only *in situ* diagnosis but also simultaneous therapeutics to combat otherwise intractable diseases.

Furthermore, the results from our collaborative work (Chapter 7) are relevant to a group of rare tissue-specific genetic diseases, ribosomopathies. These diseases (also in need of novel therapeutic approaches) are caused by mutations of ribosomal proteins (RPs) or other ribosome biogenesis factors. The specialized-ribosome hypothesis states that ribosomes from different tissues present different RP compositions and posttranslational modifications, and that this heterogeneity may be critical to the translation of specific mRNAs. Ribosome heterogeneity, particularly variations in ribosomal protein composition has been hypothesized to be linked with these diseases. Obtaining evidence of physical ribosome heterogeneity and the detection of fluctuations on ribosomes RP composition with possible functional implications remains a challenging task. This study showed that the protein composition of ribosomes remained invariant across the tissues of mice (cortex, hippocampus, and cerebellum from brain, and liver) and over their lifespan (juvenile, adult, and middle-aged).

8.2 Scope for further research

In principle, the circuit rationale presented in Chapter 5 can be adapted to any PTC-related disease: as a first step, the miRNAs that are dysregulated in the disease of interest have to be identified and differentiated from similar diseases. Secondly, the antisense region of the toehold switch sensor has to be exchanged accordingly to the disease biomarker. Finally, the suppressor tRNA to adapt has to be selected accordingly to the PTC mutation (amber UAG; ochre UAA; opal UGA), since it has to be able to readthrough the related stop codon and incorporate the correct amino acid.

The *in vivo* implementation of the circuits is yet in an inconclusive stage, further experimentation has to be performed, optimized and finally the behavior of the systems has to be characterized. The proof-of-principle experiment was performed in the prokaryotic organism *E. coli*. However, since the models presented here are ultimately intended for human use, the designs have to be further optimized, and their efficacy and functionality in mammalian systems (e.g. CF-related cell culture and CF patient-derived cells) should be tested. A good starting point would be to optimize their function in the most common laboratory used human cell lines: HeLa, Calu-3 or CFBE41o-, conventionally used as laboratory cell models in the CF field. After the modules have been fully assembled and verified to work as intended, the complete circuitry ideally should be characterized in CF patient-derived primary nasal or bronchial epithelial cells.

Since the components of the circuits are RNA-based, the delivery of the circuitry to human cells, particularly related to the potential activation of the innate immune response has to be meticulously studied. It is known that the administration of nucleic acids (e.g siRNA duplexes or dsRNA) to mammalian cells usually activates the innate immune system response. Lastly, sRNAs are higly vulnerable to degradation in the cell environment, therefore, delivery agents or vectors (e.g. viral vectors, liposomes or peptide nanoparticles) have to be investigated for their effective delivery.

References

[1] Ho, J. M., Miller, C. A., Parks, S. E., Mattia, J. R., and Bennett, M. R., 2021. "A suppressor tRNA-mediated feedforward loop eliminates leaky gene expression in bacteria". *Nucleic acids research,* **49**(5), pp. e25–e25.

[2] Bartoli, V., Meaker, G. A., Di Bernardo, M., and Gorochowski, T. E., 2020. "Tunable genetic devices through simultaneous control of transcription and translation". *Nature communications,* **11**(1), pp. 1–11.

[3] Chowdhury, S., Castro, S., Coker, C., Hinchliffe, T. E., Arpaia, N., and Danino, T., 2019. "Programmable bacteria induce durable tumor regression and systemic antitumor immunity". *Nature medicine,* **25**(7), pp. 1057–1063.

[4] Kilcher, S., and Loessner, M. J., 2019. "Engineering bacteriophages as versatile biologics". *Trends in microbiology,* **27**(4), pp. 355–367.

[5] Mogas-Díez, S., Gonzalez-Flo, E., and Macía, J., 2021. "2D printed multicellular devices performing digital and analogue computation". *Nature communications,* **12**(1), pp. 1–10.

[6] Goñi-Moreno, A., and Amos, M., 2012. "A reconfigurable NAND/NOR genetic logic gate". *BMC systems biology,* **6**(1), pp. 1–11.

[7] Treutlein, B., Lee, Q. Y., Camp, J. G., Mall, M., Koh, W., Shariati, S. A. M., Sim, S., Neff, N. F., Skotheim, J. M., Wernig, M., et al., 2016. "Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq". *Nature,* **534**(7607), pp. 391–395.

[8] Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T. S., Cantor, C. R., and Collins, J. J., 2004. "Programmable cells: interfacing natural and engineered gene networks". *Proceedings of the National Academy of Sciences,* **101**(22), pp. 8414–8419.

[9] Tay, P. K. R., Nguyen, P. Q., and Joshi, N. S., 2017. "A synthetic circuit for mercury bioremediation using self-assembling functional amyloids". *ACS synthetic biology,* **6**(10), pp. 1841–1850.

[10] Elowitz, M. B., and Leibler, S., 2000. "A synthetic oscillatory network of transcriptional regulators". *Nature,* **403**(6767), pp. 335–338.

[11] Andrews, L. B., Nielsen, A. A., and Voigt, C. A., 2018. "Cellular checkpoint control using programmable sequential logic". *Science,* **361**(6408).

[12] Seelig, G., Soloveichik, D., Zhang, D. Y., and Winfree, E., 2006. "Enzyme-free nucleic acid logic circuits". *science,* **314**(5805), pp. 1585–1588.

[13] Soloveichik, D., Seelig, G., and Winfree, E., 2010. "DNA as a universal substrate for chemical kinetics". *Proceedings of the National Academy of Sciences,* **107**(12), pp. 5393–5398.

[14] Zhang, D. Y., and Seelig, G., 2011. "Dynamic DNA nanotechnology using strand-displacement reactions". *Nature chemistry,* **3**(2), pp. 103–113.

[15] Qian, L., Winfree, E., and Bruck, J., 2011. "Neural network computation with DNA strand displacement cascades". *Nature,* **475**(7356), pp. 368–372.

[16] Qian, L., and Winfree, E., 2011. "Scaling up digital circuit computation with DNA strand displacement cascades". *Science,* **332**(6034), pp. 1196–1201.

[17] Qian, L., and Winfree, E., 2011. "A simple DNA gate motif for synthesizing large-scale circuits". *Journal of the Royal Society Interface,* **8**(62), pp. 1281–1297.

[18] Green, A. A., Silver, P. A., Collins, J. J., and Yin, P., 2014. "Toehold switches: de-novo-designed regulators of gene expression". *Cell,* **159**(4), pp. 925–939.

[19] Green, A. A., Kim, J., Ma, D., Silver, P. A., Collins, J. J., and Yin, P., 2017. "Complex cellular logic computation using ribocomputing devices". *Nature,* **548**(7665), pp. 117–121.

[20] Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., and Collins, J. J., 2014. "Paper-Based Synthetic Gene Networks". *Cell,* **159**(4), pp. 940–954.

[21] Dahl, N., 2003. "Genetic aspects of diagnosis". In *Handbook of Clinical Neurophysiology*, Vol. 2. Elsevier, pp. 99–112.

[22] Benhabiles, H., Jia, J., and Lejeune, F., 2016. *Pathologies Susceptible to be Targeted for Nonsense Mutation Therapies*. 12, pp. 77–105.

[23] Taguchi, A., Hamada, K., and Hayashi, Y., 2018. "Chemotherapeutics overcoming nonsense mutation-associated genetic diseases: medicinal chemistry of negamycin". *The Journal of antibiotics,* **71**(2), pp. 205–214.

[24] Lueck, J. D., Yoon, J. S., Perales-Puchalt, A., Mackey, A. L., Infield, D. T., Behlke, M. A., Pope, M. R., Weiner, D. B., Skach, W. R., McCray, P. B., et al., 2019. "Engineered transfer RNAs for suppression of premature termination codons". *Nature communications,* **10**(1), pp. 1–11.

[25] Farley-Barnes, K. I., Ogawa, L. M., and Baserga, S. J., 2019. "Ribosomopathies: old concepts, new controversies". *Trends in Genetics,* **35**(10), pp. 754–767.

[26] Schadt, E. E., 2009. "Molecular networks as sensors and drivers of common human diseases". *Nature,* **461**(7261), pp. 218–223.

[27] Rupaimoole, R., and Slack, F. J., 2017. "MicroRNA therapeutics: towards a new era for the management of cancer and other diseases". *Nature reviews Drug discovery,* **16**(3), p. 203.

[28] Mooney, C., McKiernan, P. J., Raoof, R., Henshall, D., Linnane, B., McNally, P., Glasgow, A., and Greene, C., 2020. "Plasma microRNA levels in male and female children with cystic fibrosis". *Scientific reports,* **10**(1), pp. 1–8.

[29] Vishnoi, A., and Rani, S., 2017. "MiRNA biogenesis and regulation of diseases: an overview". *MicroRNA Profiling*, pp. 1–10.

[30] Hayes, J., Peruzzi, P. P., and Lawler, S., 2014. "Micrornas in cancer: biomarkers, functions and therapy". *Trends in molecular medicine,* **20**(8), pp. 460–469.

[31] Banack, S. A., Dunlop, R. A., and Cox, P. A., 2020. "An miRNA fingerprint using neural-enriched extracellular vesicles from blood plasma: towards a biomarker for amyotrophic lateral sclerosis/motor neuron disease". *Open biology,* **10**(6), p. 200116.

[32] Jin, L., Li, M., Wang, H., Yin, Z., Chen, L., Zhou, Y., Han, Y., Cui, Q., Zhou, Y., and Xue, L., 2021. "Transcriptome analysis of arterial and venous circulating miRNAs during hypertension". *Scientific reports,* **11**(1), pp. 1–10.

[33] Albers, S., Beckert, B., Matthies, M. C., Mandava, C. S., Schuster, R., Seuring, C., Riedner, M., Sanyal, S., Torda, A. E., Wilson, D. N., and Ignatova, Z., 2021. "Repurposing tRNAs for nonsense suppression". *Nature Communications,* **12**(1), Jun, p. 3850.

[34] Amirbeigiarab, S., Kiani, P., Velazquez Sanchez, A., Krisp, C., Kazantsev, A., Fester, L., Schlüter, H., and Ignatova, Z., 2019. "Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging". *Proceedings of the National Academy of Sciences,* **116**(45), pp. 22567–22572.

[35] Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P., 2014. *Molecular Biology of the Cell: Sixth International Student Edition*. 500 Tips. Garland Science, Taylor and Francis Group.

[36] Lewin, B., 2008. *Genes IX*. No. 575.12 L48/9.

[37] Li, Y., and Zhao, D., 2013. *Basics of Molecular Biology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 541–601.

[38] Crick, F., 1970. "Central dogma of molecular biology". *Nature,* **227**(5258), pp. 561–563.

[39] Greive, S. J., and Von Hippel, P. H., 2005. "Thinking quantitatively about transcriptional regulation". *Nature Reviews Molecular Cell Biology,* **6**(3), pp. 221–232.

[40] Morita, M., Gravel, S.-P., Chénard, V., Sikström, K., Zheng, L., Alain, T., Gandin, V., Avizonis, D., Arguello, M., Zakaria, C., McLaughlan, S., Nouet, Y., Pause, A., Pollak, M., Gottlieb, E., Larsson, O., St-Pierre, J., Topisirovic, I., and Sonenberg, N., 2013. "mTORC1 controls mitochondrial activity and biogenesis through 4e-bp-dependent translational regulation". *Cell Metabolism,* **18**(5), pp. 698–711.

[41] Kozak, M., 1999. "Initiation of translation in prokaryotes and eukaryotes". *Gene,* **234**(2), pp. 187–208.

[42] Schuller, A. P., and Green, R., 2018. "Roadblocks and resolutions in eukaryotic translation". *Nature Reviews Molecular Cell Biology,* **19**(8), pp. 526–541.

[43] Ehrenberg, M., and Tenson, T., 2002. "A new beginning of the end of translation". *Nature structural biology,* **9**(2), pp. 85–87.

[44] Sun, Q., Zhu, X., Qi, J., An, W., Lan, P., Tan, D., Chen, R., Wang, B., Zheng, S., Zhang, C., et al., 2017. "Molecular architecture of the 90s small subunit pre-ribosome". *Elife,* **6**, p. e22086.

[45] Petrov, A. S., Bernier, C. R., Hsiao, C., Norris, A. M., Kovacs, N. A., Waterbury, C. C., Stepanov, V. G., Harvey, S. C., Fox, G. E., Wartell, R. M., et al., 2014. "Evolution of the ribosome at atomic resolution". *Proceedings of the National Academy of Sciences,* **111**(28), pp. 10251–10256.

[46] Melnikov, S., Ben-Shem, A., De Loubresse, N. G., Jenner, L., Yusupova, G., and Yusupov, M., 2012. "One core, two shells: bacterial and eukaryotic ribosomes". *Nature structural & molecular biology,* **19**(6), p. 560.

[47] Londei, P., 2020. *Archaeal Ribosomes*. American Cancer Society, pp. 1–5.

[48] Ramakrishnan, V., 2002. "Ribosome structure and the mechanism of translation". *Cell,* **108**(4), pp. 557–572.

[49] Mattick, J. S., and Makunin, I. V., 2006. "Non-coding RNA". *Human molecular genetics,* **15**(suppl_1), pp. R17–R29.

[50] Lindsay, M. A., Griffiths-Jones, S., Valadkhan, S., and Gunawardane, L. S., 2013. "Role of small nuclear RNAs in eukaryotic gene expression". *Essays in biochemistry,* **54**, pp. 79–90.

[51] Carthew, R. W., and Sontheimer, E. J., 2009. "Origins and mechanisms of miRNAs and siRNAs". *Cell,* **136**(4), pp. 642–655.

[52] Holmqvist, E., and Wagner, E. G. H., 2017. "Impact of bacterial sRNAs in stress responses". *Biochemical Society Transactions,* **45**(6), pp. 1203–1212.

[53] Hori, H., Tomikawa, C., Hirata, A., Toh, Y., Tomita, K., Ueda, T., and Watanabe, K., 2014. "Transfer RNA synthesis and regulation". *eLS*.

[54] Kirchner, S., and Ignatova, Z., 2015. "Emerging roles of tRNA in adaptive translation, signalling dynamics and disease". *Nature Reviews Genetics,* **16**(2), pp. 98–112.

[55] Modi, S. R., Camacho, D. M., Kohanski, M. A., Walker, G. C., and Collins, J. J., 2011. "Functional characterization of bacterial sRNAs using a network biology approach". *Proceedings of the National Academy of Sciences,* **108**(37), pp. 15522–15527.

[56] Desnoyers, G., Bouchard, M.-P., and Massé, E., 2013. "New insights into small RNA-dependent translational regulation in prokaryotes". *TRENDS in Genetics,* **29**(2), pp. 92–98.

[57] Kavita, K., de Mets, F., and Gottesman, S., 2018. "New aspects of RNA-based regulation by Hfq and its partner sRNAs". *Current opinion in microbiology,* **42**, pp. 53–61.

[58] Raghavan, R., Groisman, E. A., and Ochman, H., 2011. "Genome-wide detection of novel regulatory RNAs in E. coli". *Genome research,* **21**(9), pp. 1487–1497.

[59] Massé, E., Vanderpool, C. K., and Gottesman, S., 2005. "Effect of RyhB small RNA on global iron use in Escherichia coli". *Journal of bacteriology,* **187**(20), pp. 6962–6971.

[60] Morita, T., Maki, K., and Aiba, H., 2005. "RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs". *Genes & development,* **19**(18), pp. 2176–2186.

[61] Møller, T., Franch, T., Højrup, P., Keene, D. R., Bächinger, H. P., Brennan, R. G., and Valentin-Hansen, P., 2002. "Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction". *Molecular cell,* **9**(1), pp. 23–30.

[62] Vogel, J., and Luisi, B. F., 2011. "Hfq and its constellation of RNA". *Nature Reviews Microbiology,* **9**(8), pp. 578–589.

[63] Vytvytska, O., Moll, I., Kaberdin, V. R., von Gabain, A., and Bläsi, U., 2000. "Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding". *Genes & development,* **14**(9), pp. 1109–1118.

[64] Desnoyers, G., and Massé, E., 2012. "Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq". *Genes & development,* **26**(7), pp. 726–739.

[65] Andrade, J. M., Dos Santos, R. F., Chelysheva, I., Ignatova, Z., and Arraiano, C. M., 2018. "The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity". *The EMBO journal,* **37**(11), p. e97631.

[66] Večerek, B., Moll, I., and Bläsi, U., 2005. "Translational autocontrol of the Escherichia coli hfq RNA chaperone gene". *RNA,* **11**(6), pp. 976–984.

[67] Bronze-da Rocha, E., 2014. "MicroRNAs expression profiles in cardiovascular diseases". *BioMed research international,* **2014**.

[68] Wilczynska, A., and Bushell, M., 2015. "The complexity of miRNA-mediated repression". *Cell Death & Differentiation,* **22**(1), pp. 22–33.

[69] Slezak-Prochazka, I., Durmus, S., Kroesen, B.-J., and van den Berg, A., 2010. "MicroRNAs, macrocontrol: regulation of miRNA processing". *RNA,* **16**(6), pp. 1087–1095.

[70] Michlewski, G., and Cáceres, J. F., 2019. "Post-transcriptional control of miRNA biogenesis". *RNA,* **25**(1), pp. 1–16.

[71] Lane, A. N., and Jenkins, T. C., 2000. "Thermodynamics of nucleic acids and their interactions with ligands". *Quarterly Reviews of Biophysics,* **33**(3), p. 255–306.

[72] Carreón-Calderón, B., Uribe-Vargas, V., and Aguayo, J. P., 2021. *Fundamental Concepts of Thermodynamics*. Springer International Publishing, Cham, pp. 9–40.

[73] Berg, J. M., Tymoczko, J. L., and Stryer, L., 2002. "Biochemistry, ; w. h". *New York: Freeman and Company: New York*.

[74] Zuker, M., 2000. "Calculating nucleic acid secondary structure". *Current Opinion in Structural Biology,* **10**(3), pp. 303–310.

[75] Nakano, S.-i., Fujimoto, M., Hara, H., and Sugimoto, N., 1999. "Nucleic acid duplex stability: influence of base composition on cation effects". *Nucleic Acids Research,* **27**(14), 07, pp. 2957–2965.

[76] SantaLucia Jr, J., and Hicks, D., 2004. "The thermodynamics of DNA structural motifs". *Annu. Rev. Biophys. Biomol. Struct.,* **33**, pp. 415–440.

[77] Xia, T., SantaLucia Jr, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H., 1998. "Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson- Crick base pairs". *Biochemistry,* **37**(42), pp. 14719–14735.

[78] Zuker, M., and Stiegler, P., 1981. "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information". *Nucleic Acids Research,* **9**(1), 01, pp. 133–148.

[79] Sugimoto, N., Nakano, S.-i., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M., 1995. "Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes.". *Biochemistry,* **34**(35), pp. 11211–11216.

[80] Watkins, N. E., Kennelly, W. J., Tsay, M. J., Tuin, A., Swenson, L., Lee, H.-R., Morosyuk, S., Hicks, D. A., and SantaLucia, J., 2011. "Thermodynamic contributions of single internal rA· dA, rC· dC, rG· dG and rU· dT mismatches in RNA/DNA duplexes". *Nucleic acids research,* **39**(5), pp. 1894–1902.

[81] Mückstein, U., Tafer, H., Hackermüller, J., Bernhart, S. H., Stadler, P. F., and Hofacker, I. L., 2006. "Thermodynamics of RNA–RNA binding". *Bioinformatics,* **22**(10), pp. 1177–1182.

[82] Xia, T., SantaLucia Jr, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H., 1998. "Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson- Crick base pairs". *Biochemistry,* **37**(42), pp. 14719–14735.

[83] Schroeder, S. J., and Turner, D. H., 2000. "Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA". *Biochemistry,* **39**(31), pp. 9257–9274.

[84] Mathews, D. H., and Turner, D. H., 2002. "Experimentally derived nearest-neighbor parameters for the stability of RNA three-and four-way multibranch loops". *Biochemistry,* **41**(3), pp. 869–880.

[85] Bompfünewerer, A. F., Backofen, R., Bernhart, S. H., Hertel, J., Hofacker, I. L., Stadler, P. F., and Will, S., 2008. "Variations on RNA folding and alignment: lessons from Benasque". *Journal of mathematical biology,* **56**(1), pp. 129–144.

[86] Lorenz, R., Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L., 2011. "ViennaRNA Package 2.0". *Algorithms for Molecular Biology,* **6**(1), p. 26.

[87] Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P., 1994. "Fast folding and comparison of RNA secondary structures". *Monatshefte für Chemie/Chemical Monthly,* **125**(2), pp. 167–188.

[88] Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H., 2004. "Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure". *Proceedings of the National Academy of Sciences,* **101**(19), pp. 7287–7292.

[89] Turner, D. H., and Mathews, D. H., 2010. "NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure". *Nucleic acids research,* **38**(suppl_1), pp. D280–D282.

[90] Hofacker, I., Stadler, P., and Lorenz, R. *RNAlib API Reference Manual*. ViennaRNA Package.

[91] Hofacker, I. L., and Stadler, P. F., 2006. "Memory efficient folding algorithms for circular RNA secondary structures". *Bioinformatics,* **22**(10), pp. 1172–1176.

[92] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A., 2011. "NUPACK: analysis and design of nucleic acid systems". *Journal of computational chemistry,* **32**(1), pp. 170–173.

[93] Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al., 2009. "Biopython: freely available Python tools for computational molecular biology and bioinformatics". *Bioinformatics,* **25**(11), pp. 1422–1423.

[94] Eggenhofer, F., Tafer, H., Stadler, P. F., and Hofacker, I. L., 2011. "RNApredator: fast accessibility-based prediction of sRNA targets". *Nucleic acids research,* **39**(suppl_2), pp. W149–W154.

[95] ROSALIND glossary - fasta format. `http://rosalind.info/glossary/fasta-format/`. Accessed: 2021-07-13.

[96] Zhou, D., and Yang, R., 2006. "Global analysis of gene transcription regulation in prokaryotes". *Cellular and Molecular Life Sciences CMLS,* **63**(19), pp. 2260–2290.

[97] Seshasayee, A. S. N., Sivaraman, K., and Luscombe, N. M., 2011. *An Overview of Prokaryotic Transcription Factors*. Springer Netherlands, Dordrecht, pp. 7–23.

[98] Meysman, P., Collado-Vides, J., Morett, E., Viola, R., Engelen, K., and Laukens, K., 2014. "Structural properties of prokaryotic promoter regions correlate with functional features". *PloS one,* **9**(2), p. e88717.

[99] Pérez-Rueda, E., and Collado-Vides, J., 2000. "The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12". *Nucleic acids research,* **28**(8), pp. 1838–1847.

[100] McClure, W. R., 1985. "Mechanism and control of transcription initiation in prokaryotes". *Annual review of biochemistry,* **54**(1), pp. 171–204.

[101] Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S., and Nudler, E., 2007. "An allosteric path to transcription termination". *Molecular cell,* **28**(6), pp. 991–1001.

[102] Evguenieva-Hackenberg, E., and Klug, G., 2011. "New aspects of RNA processing in prokaryotes". *Current opinion in microbiology,* **14**(5), pp. 587–592.

[103] Boeke, J. D., 2003. "The unusual phylogenetic distribution of retrotransposons: a hypothesis". *Genome research,* **13**(9), pp. 1975–1983.

[104] McCarthy, J. E., and Gualerzi, C., 1990. "Translational control of prokaryotic gene expression". *Trends in Genetics,* **6**, pp. 78–85.

[105] Nakagawa, S., Niimura, Y., Miura, K.-i., and Gojobori, T., 2010. "Dynamic evolution of translation initiation mechanisms in prokaryotes". *Proceedings of the National Academy of Sciences,* **107**(14), pp. 6382–6387.

[106] Rodnina, M. V., 2018. "Translation in prokaryotes". *Cold Spring Harbor perspectives in biology,* **10**(9), p. a032664.

[107] Gao, N., Zavialov, A. V., Li, W., Sengupta, J., Valle, M., Gursky, R. P., Ehrenberg, M., and Frank, J., 2005. "Mechanism for the disassembly of the posttermination complex inferred from cryo-EM studies". *Molecular cell,* **18**(6), pp. 663–674.

[108] Phillips, T., 2008. "Regulation of transcription and gene expression in eukaryotes". *Nature Education,* **1**(1), p. 199.

[109] Vannini, A., and Cramer, P., 2012. "Conservation between the RNA polymerase I, II, and III transcription initiation machineries". *Molecular cell,* **45**(4), pp. 439–446.

[110] Gill, G., 2001. "Regulation of the initiation of eukaryotic transcription". *Essays in biochemistry,* **37**, pp. 33–44.

[111] Kornberg, R. D., 2007. "The molecular basis of eukaryotic transcription". *Proceedings of the National Academy of Sciences,* **104**(32), pp. 12955–12961.

[112] Birnstiel, M. L., Busslinger, M., and Strub, K., 1985. "Transcription termination and 3′ processing: the end is in site!". *Cell,* **41**(2), pp. 349–359.

[113] Brown, K. M., and Gilmartin, G. M., 2003. "A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor Im". *Molecular cell,* **12**(6), pp. 1467–1476.

[114] Wells, S. E., Hillner, P. E., Vale, R. D., and Sachs, A. B., 1998. "Circularization of mrna by eukaryotic translation initiation factors". *Molecular cell,* **2**(1), pp. 135–140.

[115] Lee, T. I., and Young, R. A., 2000. "Transcription of eukaryotic protein-coding genes". *Annual review of genetics,* **34**(1), pp. 77–137.

[116] Sokabe, M., and Fraser, C. S., 2019. "Toward a kinetic understanding of eukaryotic translation". *Cold Spring Harbor perspectives in biology,* **11**(2), p. a032706.

[117] Kapp, L. D., and Lorsch, J. R., 2004. "The molecular mechanics of eukaryotic translation". *Annual review of biochemistry,* **73**(1), pp. 657–704.

[118] Aitken, C. E., and Lorsch, J. R., 2012. "A mechanistic overview of translation initiation in eukaryotes". *Nature structural & molecular biology,* **19**(6), p. 568.

[119] Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U., 2001. "Molecular mechanisms of translation initiation in eukaryotes". *Proceedings of the National Academy of Sciences,* **98**(13), pp. 7029–7036.

[120] Dever, T. E., and Green, R., 2012. "The elongation, termination, and recycling phases of translation in eukaryotes". *Cold Spring Harbor perspectives in biology,* **4**(7), p. a013706.

[121] Jain, K. K., 2017. *Biomarkers of Genetic Disorders*. Springer New York, New York, NY, pp. 239–247.

[122] Li, W. J., Wang, Y., Liu, R., Kasinski, A. L., Shen, H., Slack, F. J., and Tang, D. G., 2021. "MicroRNA-34a: Potent Tumor Suppressor, Cancer Stem Cell Inhibitor, and Potential Anticancer Therapeutic". *Frontiers in cell and developmental biology,* **9**, p. 322.

[123] Cheng, L., Sharples, R. A., Scicluna, B. J., and Hill, A. F., 2014. "Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood". *Journal of extracellular vesicles,* **3**(1), p. 23743.

[124] Geekiyanage, H., Rayatpisheh, S., Wohlschlegel, J. A., Brown, R., and Ambros, V., 2020. "Extracellular microRNAs in human circulation are associated with miRISC complexes that are accessible to anti-AGO2 antibody and can bind target mimic oligonucleotides". *Proceedings of the National Academy of Sciences,* **117**(39), pp. 24213–24223.

[125] Pegtel, D. M., and Gould, S. J., 2019. "Exosomes". *Annual review of biochemistry,* **88**, pp. 487–514.

[126] Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S., 2018. "miRBase: from microRNA sequences to function". *Nucleic Acids Research,* **47**(D1), 11, pp. D155–D162.

[127] Gregg, J., 1998. "Ones and zeros understanding Boolean algebra, digital circuits, and the logic of sets".

[128] Boole, G., 1847. "The mathematical analysis of logic, being an essay towards a calculus of deductive reasoning, originally published in cambridge by macmillan, barclay, & macmillan". *Reprinted in Oxford by Basil Blackwell*, p. 175.

[129] Fischer-Cripps, A., 2002. "2.1 - number systems". In *Newnes Interfacing Companion*, A. Fischer-Cripps, ed. Newnes, Oxford, pp. 73–87.

[130] Sinclair, I., 2011. "Chapter 10 - Gating and Logic Circuits". In *Electronics Simplified (Third Edition)*, I. Sinclair, ed., third edition ed. Newnes, Oxford, pp. 185–201.

[131] Feynman, R., 1961. There is Plenty of Room at the Bottom, edited by HD Gilbert, H.

[132] Head, T., 1987. "Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors". *Bulletin of Mathematical Biology,* **49**(6), pp. 737–759.

[133] Adleman, L., 1994. "Molecular computation of solutions to combinatorial problems". *Science,* **266**(5187), pp. 1021–1024.

[134] Ignatova, Z., Martínez-Pérez, I., and Zimmermann, K., 2008. *DNA Computing Models*. Advances in information security. Springer.

[135] Parker, J., 2003. "Computing with DNA: Although DNA clearly outclasses any silicon-based computer when it comes to information storage and processing speed, a DNA-based PC is still a long way off". *EMBO reports,* **4**(1), pp. 7–10.

[136] Liu, X., Liu, H., and Zheng, Y., 2009. "Creative conceptual design based on evolutionary DNA computing technique". In Growth and Development of Computer-Aided Innovation, R. Tan, G. Cao, and N. León, eds., Springer Berlin Heidelberg, pp. 231–241.

[137] Pasotti, L., Zucca, S., and Magni, P., 2013. "Modelling for synthetic biology". *Modeling Methodology for Physiology and Medicine: Second Edition. Published by Elsevier*, pp. 545–564.

[138] Hanczyc, M. M., 2020. "Engineering Life: A Review of Synthetic Biology". *Artificial life,* **26**(2), pp. 260–273.

[139] Monod, J., and Jacob, F., 1961. "General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation". In Cold Spring Harbor symposia on quantitative biology, Vol. 26, Cold Spring Harbor Laboratory Press, pp. 389–401.

[140] Cameron, D. E., Bashor, C. J., and Collins, J. J., 2014. "A brief history of synthetic biology". *Nature Reviews Microbiology,* **12**(5), pp. 381–390.

[141] De Silva, A. P., and Uchiyama, S., 2007. "Molecular logic and computing". *Nature nanotechnology,* **2**(7), pp. 399–410.

[142] Martínez-Pérez, I. M., Zhang, G., Ignatova, Z., and Zimmermann, K.-H., 2007. "Computational genes: a tool for molecular diagnosis and therapy of aberrant mutational phenotype". *BMC bioinformatics,* **8**(1), pp. 1–9.

[143] Stojanovic, M. N., and Stefanovic, D., 2003. "A deoxyribozyme-based molecular automaton.". *Nature biotechnology,* **21**(9), Sept., pp. 1069–74.

[144] Macdonald, J., Li, Y., Sutovic, M., Lederman, H., Pendri, K., Lu, W., Andrews, B. L., Stefanovic, D., and Stojanovic, M. N., 2006. "Medium scale integration of molecular logic gates in an automaton". *Nano letters,* **6**(11), pp. 2598–2603.

[145] Kahan-Hanum, M., Douek, Y., Adar, R., and Shapiro, E., 2013. "A library of programmable DNAzymes that operate in a cellular environment.". *Scientific reports,* **3**, Jan., p. 1535.

[146] Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Livneh, Z., and Shapiro, E., 2001. "Programmable and autonomous computing machine made of biomolecules". *Nature,* **414**(6862), pp. 430–434.

[147] Prokup, A., Hemphill, J., and Deiters, A., 2012. "DNA computation: a photochemically controlled and gate". *Journal of the American Chemical Society,* **134**(8), pp. 3810–3815.

[148] Rothemund, P. W., Papadakis, N., and Winfree, E., 2004. "Algorithmic self-assembly of DNA Sierpinski triangles". *PLoS biology,* **2**(12), p. e424.

[149] Win, M. N., and Smolke, C. D., 2008. "Higher-order cellular information processing with synthetic RNA devices". *Science,* **322**(5900), pp. 456–460.

[150] Brophy, J. A., and Voigt, C. A., 2014. "Principles of genetic circuit design". *Nature methods,* **11**(5), pp. 508–520.

[151] Doudna, J. A., and Charpentier, E., 2014. "The new frontier of genome engineering with CRISPR-Cas9". *Science,* **346**(6213).

[152] Green, A. A., 2019. "Synthetic bionanotechnology: synthetic biology finds a toehold in nanotechnology". *Emerging Topics in Life Sciences,* **3**(5), 10, pp. 507–516.

[153] Ausländer, S., and Fussenegger, M., 2014. "Toehold gene switches make big footprints". *Nature,* **516**(7531), pp. 333–334.

[154] Schmidt, C. M., and Smolke, C. D., 2019. "RNA switches for synthetic biology". *Cold Spring Harbor perspectives in biology,* **11**(1), p. a032532.

[155] Kim, J., Yin, P., and Green, A. A., 2018. Ribocomputing: cellular logic computation using RNA devices.

[156] Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., Daringer, N. M., Bosch, I., Dudley, D. M., O'Connor, D. H., Gehrke, L., and Collins, J. J., 2016. "Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components". *Cell,* **165**(5), pp. 1255–1266.

[157] Boettner, B., 2020. Harvard Wyss Institute's nasal swab and toehold switch technologies licensed to Agile Biodetection to facilitate SARS-CoV-2 diagnostic efforts. https://wyss.harvard.edu/news/, October.

[158] 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**(1), pp. 25–33.

[159] Wiedenheft, B., Sternberg, S. H., and Doudna, J. A., 2012. "RNA-guided genetic silencing systems in bacteria and archaea". *Nature,* **482**(7385), pp. 331–338.

[160] Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., et al., 2007. "A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila". *Nature,* **448**(7150), pp. 151–156.

[161] Sui, G., Soohoo, C., Affar, E. B., Gay, F., Shi, Y., Forrester, W. C., and Shi, Y., 2002. "A DNA vector-based RNAi technology to suppress gene expression in mammalian cells". *Proceedings of the National Academy of Sciences,* **99**(8), pp. 5515–5520.

[162] Gottesman, S., 2004. "The small RNA regulators of Escherichia coli: roles and mechanisms". *Annu. Rev. Microbiol.,* **58**, pp. 303–328.

[163] Liang, J. C., Bloom, R. J., and Smolke, C. D., 2011. "Engineering biological systems with synthetic RNA molecules". *Molecular cell,* **43**(6), pp. 915–926.

[164] Yoo, S. M., Na, D., and Lee, S. Y., 2013. "Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli". *Nature protocols,* **8**(9), pp. 1694–1707.

[165] Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., and Breaker, R. R., 2004. "A glycine-dependent riboswitch that uses cooperative binding to control gene expression". *Science,* **306**(5694), pp. 275–279.

[166] Taxman, D. J., Moore, C. B., Guthrie, E. H., and Huang, M. T.-H., 2010. "Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown". In *RNA therapeutics*. Springer, pp. 139–156.

[167] Coleman, J., Green, P. J., and Inouye, M., 1984. "The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes". *Cell,* **37**(2), pp. 429–436.

[168] Urban, J. H., and Vogel, J., 2007. "Translational control and target recognition by Escherichia coli small RNAs in vivo". *Nucleic acids research,* **35**(3), pp. 1018–1037.

[169] Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., and Lee, S. Y., 2013. "Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs". *Nature biotechnology,* **31**(2), pp. 170–174.

[170] Noh, M., Yoo, S. M., Yang, D., and Lee, S. Y., 2019. "Broad-spectrum gene repression using scaffold engineering of synthetic sRNAs". *ACS synthetic biology,* **8**(6), pp. 1452–1461.

[171] Hu, W., Liu, S., Wang, Z., and Chen, T., 2021. "Improving riboflavin production by knocking down ribF, purA and guaC genes using synthetic regulatory small RNA". *Journal of Biotechnology*.

[172] Fujita, S., Tsumori, Y., Makino, Y., Saito, M., and Kawano, M., 2021. "Development of multiplexing gene silencing system using conditionally induced polycistronic synthetic antisense RNAs in Escherichia coli". *Biochemical and Biophysical Research Communications,* **556**, pp. 163–170.

[173] Cao, Y., Li, X., Li, F., and Song, H., 2017. "CRISPRi–sRNA: transcriptional–translational regulation of extracellular electron transfer in Shewanella oneidensis". *ACS synthetic biology,* **6**(9), pp. 1679–1690.

[174] Kelly, C. L., Harris, A. W. K., Steel, H., Hancock, E. J., Heap, J. T., and Papachristodoulou, A., 2018. "Synthetic negative feedback circuits using engineered small RNAs". *Nucleic acids research,* **46**(18), pp. 9875–9889.

[175] Lammens, E.-M., Nikel, P. I., and Lavigne, R., 2020. "Exploring the synthetic biology potential of bacteriophages for engineering non-model bacteria". *Nature communications,* **11**(1), pp. 1–14.

[176] Rudge, T. J., Brown, J. R., Federici, F., Dalchau, N., Phillips, A., Ajioka, J. W., and Haseloff, J., 2016. "Characterization of intrinsic properties of promoters". *ACS synthetic biology,* **5**(1), pp. 89–98.

[177] Hooshangi, S., Thiberge, S., and Weiss, R., 2005. "Ultrasensitivity and noise propagation in a synthetic transcriptional cascade". *Proceedings of the National Academy of Sciences,* **102**(10), pp. 3581–3586.

[178] Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., Awan, A. R., Gilbert, C., Stan, G.-B., and Ellis, T., 2018. "Burden-driven feedback control of gene expression". *Nature methods,* **15**(5), pp. 387–393.

[179] Quarton, T., Kang, T., Papakis, V., Nguyen, K., Nowak, C., Li, Y., and Bleris, L., 2020. "Uncoupling gene expression noise along the central dogma using genome engineered human cell lines". *Nucleic acids research,* **48**(16), pp. 9406–9413.

[180] Taton, A., Ma, A. T., Ota, M., Golden, S. S., and Golden, J. W., 2017. "NOT gate genetic circuits to control gene expression in cyanobacteria". *ACS synthetic biology,* **6**(12), pp. 2175–2182.

[181] Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S. A., 2010. "Positive regulation by small RNAs and the role of Hfq". *Proceedings of the National Academy of Sciences,* **107**(21), pp. 9602–9607.

[182] Bandyra, K. J., Said, N., Pfeiffer, V., Górna, M. W., Vogel, J., and Luisi, B. F., 2012. "The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E". *Molecular cell,* **47**(6), pp. 943–953.

[183] Brennan, R. G., and Link, T. M., 2007. "Hfq structure, function and ligand binding". *Current opinion in microbiology,* **10**(2), pp. 125–133.

[184] Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M., and Vogel, J., 2010. "Evidence for an autonomous 5′ target recognition domain in an Hfq-associated small RNA". *Proceedings of the National Academy of Sciences,* **107**(47), pp. 20435–20440.

[185] Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H., and Vogel, J., 2008. "Small RNA binding to 5′ mRNA coding region inhibits translational initiation". *Molecular cell,* **32**(6), pp. 827–837.

[186] Vogel, J., and Wagner, E. G. H., 2007. "Target identification of small noncoding RNAs in bacteria". *Current opinion in microbiology,* **10**(3), pp. 262–270.

[187] Storz, G., Vogel, J., and Wassarman, K. M., 2011. "Regulation by small RNAs in bacteria: expanding frontiers". *Molecular cell,* **43**(6), pp. 880–891.

[188] Rehmsmeier, M., Steffen, P., Höchsmann, M., and Giegerich, R., 2004. "Fast and effective prediction of microRNA/target duplexes". *Rna,* **10**(10), pp. 1507–1517.

[189] Tjaden, B., Goodwin, S. S., Opdyke, J. A., Guillier, M., Fu, D. X., Gottesman, S., and Storz, G., 2006. "Target prediction for small, noncoding RNAs in bacteria". *Nucleic acids research,* **34**(9), pp. 2791–2802.

[190] Sharma, V., Yamamura, A., and Yokobayashi, Y., 2012. "Engineering artificial small RNAs for conditional gene silencing in Escherichia coli". *ACS synthetic biology,* **1**(1), pp. 6–13.

[191] Noro, E., Mori, M., Makino, G., Takai, Y., Ohnuma, S., Sato, A., Tomita, M., Nakahigashi, K., and Kanai, A., 2017. "Systematic characterization of artificial small RNA-mediated inhibition of Escherichia coli growth". *RNA biology,* **14**(2), pp. 206–218.

[192] Sung, M., Yoo, S. M., Jun, R., Lee, J. E., Lee, S. Y., and Na, D., 2016. "Optimization of phage λ promoter strength for synthetic small regulatory RNA-based metabolic engineering". *Biotechnology and Bioprocess Engineering,* **21**(4), pp. 483–490.

[193] Yoo, S. M., Na, D., and Lee, S. Y., 2013. "Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli". *Nature Protocols,* **8**(9), pp. 1694–1707.

[194] Lease, R. A., and Belfort, M., 2000. "A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures". *Proceedings of the National Academy of Sciences,* **97**(18), pp. 9919–9924.

[195] Chen, S., Zhang, A., Blyn, L. B., and Storz, G., 2004. "MicC, a second small-RNA regulator of Omp protein expression in Escherichia coli". *Journal of bacteriology,* **186**(20), pp. 6689–6697.

[196] Slettan, A., Gebhardt, K., Kristiansen, E., Birkeland, N., and Lindqvist, B., 1992. "Escherichia coli K-12 and B contain functional bacteriophage P2 ogr genes". *Journal of bacteriology,* **174**(12), pp. 4094–4100.

[197] Julien, B., and Calendar, R., 1996. "Bacteriophage PSP3 and phiR73 activator proteins: analysis of promoter specificities". *Journal of bacteriology,* **178**(19), pp. 5668–5675.

[198] Birkeland, N., Lindqvist, B., and Christie, G., 1991. "Control of bacteriophage P2 gene expression: analysis of transcription of the ogr gene". *Journal of bacteriology,* **173**(21), pp. 6927–6934.

[199] Sadotra, S., Lou, Y.-C., Tang, H.-C., Chiu, Y.-C., Hsu, C.-H., and Chen, C., 2021. "Structural basis for promoter DNA recognition by the response regulator OmpR". *Journal of Structural Biology,* **213**(1), p. 107638.

[200] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J., 1990. "Basic local alignment search tool". *Journal of molecular biology,* **215**(3), pp. 403–410.

[201] Thomas, C. M., and Smith, C. A., 1987. "Incompatibility group P plasmids: genetics, evolution, and use in genetic manipulation". *Annual Reviews in Microbiology,* **41**(1), pp. 77–101.

[202] Novick, R. P., and Hoppensteadt, F., 1978. "On plasmid incompatibility". *Plasmid,* **1**(4), pp. 421–434.

[203] Pirman, N. L., Barber, K. W., Aerni, H. R., Ma, N. J., Haimovich, A. D., Rogulina, S., Isaacs, F. J., and Rinehart, J., 2015. "A flexible codon in genomically recoded Escherichia coli permits programmable protein phosphorylation". *Nature communications,* **6**(1), pp. 1–6.

[204] Fricke, P. M., Link, T., Gätgens, J., Sonntag, C., Otto, M., Bott, M., and Polen, T., 2020. "A tunable L-arabinose-inducible expression plasmid for the acetic acid bacterium Gluconobacter oxydans". *Applied microbiology and biotechnology,* **104**(21), pp. 9267–9282.

[205] Guzman, L.-M., Belin, D., Carson, M. J., and Beckwith, J., 1995. "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter". *Journal of bacteriology,* **177**(14), pp. 4121–4130.

[206] MATLAB, S., 2020. *version 9.8.0.1417392 (R2020a) Update 4*. The MathWorks Inc., Natick, Massachusetts.

[207] Kain, S. R., 1999. "Green fluorescent protein (GFP): applications in cell-based assays for drug discovery". *Drug discovery today,* **4**(7), pp. 304–312.

[208] Vincze, T., Posfai, J., and Roberts, R. J., 2003. "NEBcutter: a program to cleave DNA with restriction enzymes". *Nucleic acids research,* **31**(13), pp. 3688–3691.

[209] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W., 2012. "NIH Image to ImageJ: 25 years of image analysis". *Nature methods,* **9**(7), pp. 671–675.

[210] Bader, A., Brown, D., Stoudemire, J., and Lammers, P., 2011. "Developing therapeutic microRNAs for cancer". *Gene therapy,* **18**(12), pp. 1121–1126.

[211] Viart, V., Bergougnoux, A., Bonini, J., Varilh, J., Chiron, R., Tabary, O., Molinari, N., Claustres, M., and Taulan-Cadars, M., 2015. "Transcription factors and miRNAs that regulate fetal to adult CFTR expression change are new targets for cystic fibrosis". *European Respiratory Journal,* **45**(1), pp. 116–128.

[212] McKiernan, P. J., and Greene, C. M., 2015. "MicroRNA dysregulation in cystic fibrosis". *Mediators of inflammation,* **2015**.

[213] Cutting, G. R., 2015. "Cystic fibrosis genetics: from molecular understanding to clinical application". *Nature Reviews Genetics,* **16**(1), pp. 45–56.

[214] Brennan, M.-L., and Schrijver, I., 2016. "Cystic fibrosis: a review of associated phenotypes, use of molecular diagnostic approaches, genetic characteristics, progress, and dilemmas". *The Journal of Molecular Diagnostics,* **18**(1), pp. 3–14.

[215] Elborn, S., 2016. "Cystic fibrosis". *The Lancet,* **388**(10059), pp. 2519–2531.

[216] Marson, F. A. L., Bertuzzo, C. S., and Ribeiro, J. D., 2016. "Classification of CFTR mutation classes". *The Lancet Respiratory Medicine,* **4**(8), pp. e37–e38.

[217] Williams, A. E., Perry, M. M., Moschos, S. A., and Lindsay, M. A., 2007. "microRNA expression in the aging mouse lung". *BMC genomics,* **8**(1), pp. 1–6.

[218] Oglesby, I. K., Chotirmall, S. H., McElvaney, N. G., and Greene, C. M., 2013. "Regulation of cystic fibrosis transmembrane conductance regulator by microRNA-145,-223, and-494 is altered in ΔF508 cystic fibrosis airway epithelium". *The Journal of Immunology,* **190**(7), pp. 3354–3362.

[219] Oglesby, I. K., Agrawal, R., Mall, M. A., McElvaney, N. G., and Greene, C. M., 2015. "miRNA-221 is elevated in cystic fibrosis airway epithelial cells and regulates expression of ATF6". *Molecular and cellular pediatrics,* **2**(1), pp. 1–8.

[220] Ideozu, J. E., Zhang, X., Rangaraj, V., McColley, S., and Levy, H., 2019. "Microarray profiling identifies extracellular circulating miRNAs dysregulated in cystic fibrosis". *Scientific reports,* **9**(1), pp. 1–11.

[221] Cook, N. L., Pereira, T. N., Lewindon, P. J., Shepherd, R. W., and Ramm, G. A., 2015. "Circulating microRNAs as noninvasive diagnostic biomarkers of liver disease in children with cystic fibrosis". *Journal of pediatric gastroenterology and nutrition,* **60**(2), pp. 247–254.

[222] Gillen, A. E., Gosalia, N., Leir, S.-H., and Harris, A., 2011. "MicroRNA regulation of expression of the cystic fibrosis transmembrane conductance regulator gene". *Biochemical Journal,* **438**(1), pp. 25–32.

[223] Tiwari, A., Li, J., Kho, A. T., Sun, M., Lu, Q., Weiss, S. T., Tantisira, K. G., and McGeachie, M. J., 2020. "COPD-Associated miR-145-5p is Downregulated in Early Decline FEV1 Trajectories in Childhood Asthma". *Journal of Allergy and Clinical Immunology*.

[224] Dang, X., Yang, L., Guo, J., Hu, H., Li, F., Liu, Y., and Pang, Y., 2019. "miR-145-5p is associated with smoke-related chronic obstructive pulmonary disease via targeting KLF5". *Chemico-biological interactions,* **300**, pp. 82–90.

[225] Lutful Kabir, F., Ambalavanan, N., Liu, G., Li, P., Solomon, G. M., Lal, C. V., Mazur, M., Halloran, B., Szul, T., Gerthoffer, W. T., et al., 2018. "MicroRNA-145 antagonism reverses TGF-β inhibition of F508del CFTR correction in airway epithelia". *American journal of respiratory and critical care medicine,* **197**(5), pp. 632–643.

[226] Ramachandran, S., Karp, P. H., Jiang, P., Ostedgaard, L. S., Walz, A. E., Fisher, J. T., Keshavjee, S., Lennox, K. A., Jacobi, A. M., Rose, S. D., et al., 2012. "A microRNA network regulates expression and biosynthesis of wild-type and ΔF508 mutant cystic fibrosis transmembrane conductance regulator". *Proceedings of the National Academy of Sciences,* **109**(33), pp. 13362–13367.

[227] Weldon, S., McNally, P., McAuley, D. F., Oglesby, I. K., Wohlford-Lenane, C. L., Bartlett, J. A., Scott, C. J., McElvaney, N. G., Greene, C. M., McCray Jr, P. B., et al., 2014. "miR-31 dysregulation in cystic fibrosis airways contributes to increased pulmonary cathepsin S production". *American journal of respiratory and critical care medicine,* **190**(2), pp. 165–174.

[228] Conickx, G., Cobos, F. A., van den Berge, M., Faiz, A., Timens, W., Hiemstra, P. S., Joos, G. F., Brusselle, G. G., Mestdagh, P., and Bracke, K. R., 2017. "microRNA profiling in lung tissue and bronchoalveolar lavage of cigarette smoke-exposed mice and in COPD patients: a translational approach". *Scientific reports,* **7**(1), pp. 1–14.

[229] Ezzie, M. E., Crawford, M., Cho, J.-H., Orellana, R., Zhang, S., Gelinas, R., Batte, K., Yu, L., Nuovo, G., Galas, D., et al., 2012. "Gene expression networks in COPD: microRNA and mRNA regulation". *Thorax,* **67**(2), pp. 122–131.

[230] Johnnidis, J. B., Harris, M. H., Wheeler, R. T., Stehling-Sun, S., Lam, M. H., Kirak, O., Brummelkamp, T. R., Fleming, M. D., and Camargo, F. D., 2008. "Regulation of progenitor cell proliferation and granulocyte function by microRNA-223". *Nature,* **451**(7182), pp. 1125–1129.

[231] Zeng, Q., and Zeng, J., 2021. "Inhibition of miR-494-3p alleviates oxidative stress-induced cell senescence and inflammation in the primary epithelial cells of COPD patients". *International Immunopharmacology,* **92**, p. 107044.

[232] Shen, Y., Wang, L., Wu, Y., Ou, Y., Lu, H., and Yao, X., 2021. "A novel diagnostic signature based on three circulating exosomal mircoRNAs for chronic obstructive pulmonary disease". *Experimental and Therapeutic Medicine,* **22**(1), pp. 1–12.

[233] Hassan, F., Nuovo, G. J., Crawford, M., Boyaka, P. N., Kirkby, S., Nana-Sinkam, S. P., and Cormet-Boyaka, E., 2012. "MiR-101 and miR-144 regulate the expression of the CFTR chloride channel in the lung". *PloS one,* **7**(11), p. e50837.

[234] Ramachandran, S., Karp, P. H., Osterhaus, S. R., Jiang, P., Wohlford-Lenane, C., Lennox, K. A., Jacobi, A. M., Praekh, K., Rose, S. D., Behlke, M. A., et al., 2013. "Post-transcriptional regulation of cystic fibrosis transmembrane conductance regulator expression and function by microRNAs". *American journal of respiratory cell and molecular biology,* **49**(4), pp. 544–551.

[235] Sonneville, F., Ruffin, M., Coraux, C., Rousselet, N., Le Rouzic, P., Blouquit-Laye, S., Corvol, H., and Tabary, O., 2017. "MicroRNA-9 downregulates the ANO1 chloride channel and contributes to cystic fibrosis lung pathology". *Nature communications,* **8**(1), pp. 1–11.

[236] Kumar, P., Bhattacharyya, S., Peters, K., Glover, M., Sen, A., Cox, R., Kundu, S., Caohuy, H., Frizzell, R., Pollard, H., et al., 2015. "miR-16 rescues F508del-CFTR function in native cystic fibrosis epithelial cells". *Gene therapy,* **22**(11), pp. 908–916.

[237] Donaldson, A., Natanek, S. A., Lewis, A., Man, W. D., Hopkinson, N. S., Polkey, M. I., and Kemp, P. R., 2013. "Increased skeletal muscle-specific microRNA in the blood of patients with COPD". *Thorax,* **68**(12), pp. 1140–1149.

[238] Oglesby, I. K., Bray, I. M., Chotirmall, S. H., Stallings, R. L., O'Neill, S. J., McElvaney, N. G., and Greene, C. M., 2010. "miR-126 is downregulated in cystic fibrosis airway epithelial cells and regulates TOM1 expression". *The journal of immunology,* **184**(4), pp. 1702–1709.

[239] Tsuchiya, M., Kumar, P., Bhattacharyya, S., Chattoraj, S., Srivastava, M., Pollard, H. B., and Biswas, R., 2013. "Differential regulation of inflammation by inflammatory mediators in cystic fibrosis lung epithelial cells". *Journal of Interferon & Cytokine Research,* **33**(3), pp. 121–129.

[240] Bhattacharyya, S., Balakathiresan, N. S., Dalgard, C., Gutti, U., Armistead, D., Jozwik, C., Srivastava, M., Pollard, H. B., and Biswas, R., 2011. "Elevated miR-155 promotes inflammation in cystic fibrosis by driving hyperexpression of interleukin-8". *Journal of Biological Chemistry,* **286**(13), pp. 11604–11615.

[241] De Smet, E., Van Eeckhoutte, H., Cobos, F. A., Blomme, E., Verhamme, F., Provoost, S., Verleden, S., Venken, K., Maes, T., Joos, G., et al., 2020. "The role of miR-155 in cigarette smoke-induced pulmonary inflammation and COPD". *Mucosal immunology,* **13**(3), pp. 423–436.

[242] Tsuchiya, M., Kalurupalle, S., Kumar, P., Ghoshal, S., Zhang, Y., Lehrmann, E., Becker, K. G., Gorospe, M., and Biswas, R., 2016. "RPTOR, a novel target of miR-155, elicits a fibrotic phenotype of cystic fibrosis lung epithelium by upregulating CTGF". *RNA biology,* **13**(9), pp. 837–847.

[243] Oglesby, I. K., Vencken, S. F., Agrawal, R., Gaughan, K., Molloy, K., Higgins, G., McNally, P., McElvaney, N. G., Mall, M. A., and Greene, C. M., 2015. "miR-17 overexpression in cystic fibrosis airway epithelial cells decreases interleukin-8 production". *European Respiratory Journal,* **46**(5), pp. 1350–1360.

[244] Fabbri, E., Borgatti, M., Montagner, G., Bianchi, N., Finotti, A., Lampronti, I., Bezzerri, V., Dechecchi, M. C., Cabrini, G., and Gambari, R., 2014. "Expression of microRNA-93 and interleukin-8 during pseudomonas aeruginosa–mediated induction of proinflammatory responses". *American journal of respiratory cell and molecular biology,* **50**(6), pp. 1144–1155.

[245] Leidinger, P., Keller, A., Borries, A., Huwer, H., Rohling, M., Huebers, J., Lenhof, H.-P., and Meese, E., 2011. "Specific peripheral miRNA profiles for distinguishing lung cancer from COPD". *Lung cancer,* **74**(1), pp. 41–47.

[246] Zhang, J., Xu, Z., Kong, L., Gao, H., Zhang, Y., Zheng, Y., and Wan, Y., 2020. "miRNA-486-5p Promotes COPD Progression by Targeting HAT1 to Regulate the TLR4-Triggered Inflammatory Response of Alveolar Macrophages". *International Journal of Chronic Obstructive Pulmonary Disease,* **15**, p. 2991.

[247] Wang, L.-S., Li, L., Li, L., Chu, S., Shiang, K.-D., Li, M., Sun, H.-Y., Xu, J., Xiao, F.-J., Sun, G., et al., 2015. "MicroRNA-486 regulates normal erythropoiesis and enhances growth and modulates drug response in CML progenitors". *Blood, The Journal of the American Society of Hematology,* **125**(8), pp. 1302–1313.

[248] McLendon, J. M., Joshi, S. R., Sparks, J., Matar, M., Fewell, J. G., Abe, K., Oka, M., McMurtry, I. F., and Gerthoffer, W. T., 2015. "Lipid nanoparticle delivery of a microRNA-145 inhibitor improves experimental pulmonary hypertension". *Journal of Controlled Release,* **210**, pp. 67–75.

[249] Lovren, F., Pan, Y., Quan, A., Singh, K. K., Shukla, P. C., Gupta, N., Steer, B. M., Ingram, A. J., Gupta, M., Al-Omran, M., et al., 2012. "MicroRNA-145 targeted therapy reduces atherosclerosis". *Circulation,* **126**(11_suppl_1), pp. S81–S90.

[250] Kim, S.-J., Oh, J.-S., Shin, J.-Y., Lee, K.-D., Sung, K. W., Nam, S. J., and Chun, K.-H., 2011. "Development of microRNA-145 for therapeutic application in breast cancer". *Journal of controlled release,* **155**(3), pp. 427–434.

[251] Yang, Y., Ji, C., Guo, S., Su, X., Zhao, X., Zhang, S., Liu, G., Qiu, X., Zhang, Q., Guo, H., et al., 2017. "The miR-486-5p plays a causative role in prostate cancer through negative regulation of multiple tumor suppressor pathways". *Oncotarget,* **8**(42), p. 72835.

[252] Laurila, E. M., and Kallioniemi, A., 2013. "The diverse role of miR-31 in regulating cancer associated phenotypes". *Genes, Chromosomes and Cancer,* **52**(12), pp. 1103–1113.

[253] Wang, Q., Chang, W., Yang, X., Cheng, Y., Zhao, X., Zhou, L., Li, J., Li, J., and Zhang, K., 2019. "Levels of miR-31 and its target genes in dermal mesenchymal cells of patients with psoriasis". *International journal of dermatology,* **58**(2), pp. 198–204.

[254] Liu, T., Kang, J., and Liu, F., 2018. "Plasma levels of microRNA-221 (miR-221) are increased in patients with acute pulmonary embolism". *Medical science monitor: international medical journal of experimental and clinical research,* **24**, p. 8621.

[255] Roffel, M. P., Bracke, K. R., Heijink, I. H., and Maes, T., 2020. "miR-223: a key regulator in the innate immune response in asthma and COPD". *Frontiers in medicine,* **7**, p. 196.

[256] Wang, W., Peng, B., Wang, D., Ma, X., Jiang, D., Zhao, J., and Yu, L., 2011. "Human tumor microRNA signatures derived from large-scale oligonucleotide microarray datasets". *International journal of cancer,* **129**(7), pp. 1624–1634.

[257] Santi, C. D., and Greene, C. M., 2020. Challenges facing microRNA therapeutics for cystic fibrosis lung disease.

[258] Finotti, A., Gasparello, J., Fabbri, E., Tamanini, A., Corradini, R., Dechecchi, M. C., Cabrini, G., and Gambari, R., 2019. "Enhancing the expression of CFTR using antisense molecules against microRNA miR-145-5p". *American journal of respiratory and critical care medicine,* **199**(11), pp. 1443–1444.

[259] McKiernan, P. J., Cunningham, O., Greene, C. M., and Cryan, S.-A., 2013. "Targeting miRNA-based medicines to cystic fibrosis airway epithelial cells using nanotechnology". *International journal of nanomedicine,* **8**, p. 3907.

[260] Amato, F., Tomaiuolo, R., Nici, F., Borbone, N., Elce, A., Catalanotti, B., D'Errico, S., Morgillo, C. M., De Rosa, G., Mayol, L., et al., 2014. "Exploitation of a very small peptide nucleic acid as a new inhibitor of miR-509-3p involved in the regulation of cystic fibrosis disease-gene expression". *BioMed research international,* **2014**.

[261] Fabbri, E., Tamanini, A., Jakova, T., Gasparello, J., Manicardi, A., Corradini, R., Sabbioni, G., Finotti, A., Borgatti, M., Lampronti, I., et al., 2018. "A peptide nucleic acid against microRNA miR-145-5p enhances the expression of the cystic fibrosis transmembrane conductance regulator (CFTR) in Calu-3 cells". *Molecules,* **23**(1), p. 71.

[262] Mousavi, S. R., Ahmadi, A., Jamalkandi, S. A., and Salimian, J., 2019. "Involvement of microRNAs in physiological and pathological processes in asthma". *Journal of cellular physiology,* **234**(12), pp. 21547–21559.

[263] Mei, D., Tan, W. S. D., and Wong, W. S. F., 2019. "Pharmacological strategies to regain steroid sensitivity in severe asthma and COPD". *Current opinion in pharmacology,* **46**, pp. 73–81.

[264] Zarrilli, F., Amato, F., Morgillo, C. M., Pinto, B., Santarpia, G., Borbone, N., D'Errico, S., Catalanotti, B., Piccialli, G., Castaldo, G., et al., 2017. "Peptide nucleic acids as miRNA target protectors for the treatment of cystic fibrosis". *Molecules,* **22**(7), p. 1144.

[265] Andriani, F., Majorini, M. T., Mano, M., Landoni, E., Miceli, R., Facchinetti, F., Mensah, M., Fontanella, E., Dugo, M., Giacca, M., et al., 2018. "MiR-16 regulates the pro-tumorigenic potential of lung fibroblasts through the inhibition of HGF production in an FGFR-1-and MEK1-dependent manner". *Journal of hematology & oncology,* **11**(1), pp. 1–17.

[266] Xia, H., Wu, Y., Zhao, J., Li, W., Lu, L., Ma, H., Cheng, C., Sun, J., Xiang, Q., Bian, T., et al., 2021. "The aberrant cross-talk of epithelium–macrophages via METTL3-regulated extracellular vesicle miR-93 in smoking-induced emphysema". *Cell Biology and Toxicology*, pp. 1–17.

[267] Karijolich, J., and Yu, Y.-T., 2014. "Therapeutic suppression of premature termination codons: mechanisms and clinical considerations". *International journal of molecular medicine,* **34**(2), pp. 355–362.

[268] Gambari, R., Breveglieri, G., Salvatori, F., Finotti, A., Borgatti, M., et al., 2015. "Therapy for Cystic Fibrosis Caused by Nonsense Mutations". *Cystic Fibrosis in the Light of New Research; Wat, D., Ed*, pp. 309–326.

[269] Kuang, L., Hashimoto, K., Huang, E. J., Gentry, M. S., and Zhu, H., 2020. "Frontotemporal dementia nonsense mutation of progranulin rescued by aminoglycosides". *Human Molecular Genetics*.

[270] Leonard, S. P., Powell, J. E., Perutka, J., Geng, P., Heckmann, L. C., Horak, R. D., Davies, B. W., Ellington, A. D., Barrick, J. E., and Moran, N. A., 2020. "Engineered symbionts activate honey bee immunity and limit pathogens". *Science,* **367**(6477), pp. 573–576.

[271] Trzaska, C., Amand, S., Bailly, C., Leroy, C., Marchand, V., Duvernois-Berthet, E., Saliou, J.-M., Benhabiles, H., Werkmeister, E., Chassat, T., et al., 2020. "2, 6-Diaminopurine as a highly potent corrector of UGA nonsense mutations". *Nature communications,* **11**(1), pp. 1–12.

[272] Xie, M., and Fussenegger, M., 2018. "Designing cell function: assembly of synthetic gene circuits for cell biology applications". *Nature Reviews Molecular Cell Biology,* **19**(8), pp. 507–525.

[273] Khalil, A. S., and Collins, J. J., 2010. "Synthetic biology: applications come of age". *Nature Reviews Genetics,* **11**(5), pp. 367–379.

[274] Tian, X., Liu, Y., Yang, J., Wang, H., Liu, T., Xu, W., Li, X., Zhu, Y., Xu, K.-F., and Zhang, X., 2016. "p. G970D is the most frequent CFTR mutation in Chinese patients with cystic fibrosis". *Human genome variation,* **3**(1), pp. 1–6.

[275] Sondo, E., Falchi, F., Caci, E., Ferrera, L., Giacomini, E., Pesce, E., Tomati, V., Bertozzi, S. M., Goldoni, L., Armirotti, A., et al., 2018. "Pharmacological inhibition of the ubiquitin ligase RNF5 rescues F508del-CFTR in cystic fibrosis airway epithelia". *Cell Chemical Biology,* **25**(7), pp. 891–905.

[276] Tang, P., 2020. "Clinical diagnostic value of circulating serum miR-509-3p in pulmonary arterial hypertension with congenital heart disease". *Hellenic Journal of Cardiology,* **61**(1), pp. 26–30.

[277] Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T., et al., 2019. "Database resources of the national center for biotechnology information". *Nucleic acids research,* **47**(Database issue), p. D23.

[278] Spaccasassi, C., Lakin, M. R., and Phillips, A., 2018. "A logic programming language for computational nucleic acid devices". *ACS synthetic biology,* **8**(7), pp. 1530–1547.

[279] Chan, P. P., and Lowe, T. M., 2009. "GtRNAdb: a database of transfer RNA genes detected in genomic sequence". *Nucleic acids research,* **37**(suppl_1), pp. D93–D97.

[280] Watanabe, I., and Okada, S., 1967. "Effects of temperature on growth rate of cultured mammalian cells (L5178Y)". *The Journal of cell biology,* **32**(2), pp. 309–323.

[281] Lim, E.-K., Kim, T., Paik, S., Haam, S., Huh, Y.-M., and Lee, K., 2015. "Nanomaterials for theranostics: recent advances and future challenges". *Chemical reviews,* **115**(1), pp. 327–394.

[282] Kojima, R., Aubel, D., and Fussenegger, M., 2016. "Toward a world of theranostic medication: programming biological sentinel systems for therapeutic intervention". *Advanced drug delivery reviews,* **105**, pp. 66–76.

[283] Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., and Shapiro, E., 2004. "An autonomous molecular computer for logical control of gene expression". *Nature,* **429**(6990), pp. 423–429.

[284] Kahan-Hanum, M., Douek, Y., Adar, R., and Shapiro, E., 2013. "A library of programmable DNAzymes that operate in a cellular environment". *Scientific reports,* **3**(1), pp. 1–6.

[285] Kim, W. J., Lim, J. H., Hong, Y., Hong, S.-H., Bang, C. Y., Lee, J. S., Oh, Y.-M., and Kim, J. H., 2017. "Altered mirna expression in lung tissues of patients with chronic obstructive pulmonary disease". *Molecular & Cellular Toxicology,* **13**(2), pp. 207–212.

[286] Durcan, N., Murphy, C., and Cryan, S.-A., 2008. "Inhalable siRNA: potential as a therapeutic agent in the lungs". *Molecular pharmaceutics,* **5**(4), pp. 559–566.

[287] Greene, C. M., 2013. "MicroRNA expression in cystic fibrosis airway epithelium". *Biomolecules,* **3**(1), pp. 157–167.

[288] Weinberg, B. H., Pham, N. H., Caraballo, L. D., Lozanoski, T., Engel, A., Bhatia, S., and Wong, W. W., 2017. "Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells". *Nature biotechnology,* **35**(5), pp. 453–462.

[289] Matsuura, S., Ono, H., Kawasaki, S., Kuang, Y., Fujita, Y., and Saito, H., 2018. "Synthetic RNA-based logic computation in mammalian cells". *Nature communications,* **9**(1), pp. 1–8.

[290] Velazquez Sanchez, A. K., 2014. "Logic gates design based on DNA strand displacement". Master's thesis, Center for Scientific Research and Higher Education at Ensenada, Mexico, Ensenada, Baja California, Mexico.

[291] Mills, E. W., and Green, R., 2017. "Ribosomopathies: There's strength in numbers". *Science,* **358**(6363).

[292] Narla, A., and Ebert, B. L., 2010. "Ribosomopathies: human disorders of ribosome dysfunction". *Blood, The Journal of the American Society of Hematology,* **115**(16), pp. 3196–3205.

[293] Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al., 2019. "The PRIDE database and related tools and resources in 2019: improving support for quantification data". *Nucleic acids research,* **47**(D1), pp. D442–D450.

[294] Roberts, E., Sethi, A., Montoya, J., Woese, C. R., and Luthey-Schulten, Z., 2008. "Molecular signatures of ribosomal evolution". *Proceedings of the National Academy of Sciences,* **105**(37), pp. 13953–13958.

[295] Wohlbach, D. J., Thompson, D. A., Gasch, A. P., and Regev, A., 2009. "From elements to modules: regulatory evolution in Ascomycota fungi". *Current opinion in genetics & development,* **19**(6), pp. 571–578.

[296] Guimaraes, J. C., and Zavolan, M., 2016. "Patterns of ribosomal protein expression specify normal and malignant human cells". *Genome biology,* **17**(1), pp. 1–13.

[297] Franks, A., Airoldi, E., and Slavov, N., 2017. "Post-transcriptional regulation across human tissues". *PLoS computational biology,* **13**(5), p. e1005535.

[298] Bhavsar, R. B., Makley, L. N., and Tsonis, P. A., 2010. "The other lives of ribosomal proteins". *Human genomics,* **4**(5), pp. 1–18.

[299] Dörrbaum, A. R., Kochen, L., Langer, J. D., and Schuman, E. M., 2018. "Local and global influences on protein turnover in neurons and glia". *Elife,* **7**, p. e34202.

[300] Price, J. C., Guan, S., Burlingame, A., Prusiner, S. B., and Ghaemmaghami, S., 2010. "Analysis of proteome dynamics in the mouse brain". *Proceedings of the National Academy of Sciences,* **107**(32), pp. 14508–14513.

[301] Adamla, F., and Ignatova, Z., 2015. "Somatic expression of unc-54 and vha-6 mRNAs declines but not pan-neuronal rgef-1 and unc-119 expression in aging Caenorhabditis elegans". *Scientific reports,* **5**(1), pp. 1–10.

[302] Baumgart, M., Groth, M., Priebe, S., Savino, A., Testa, G., Dix, A., Ripa, R., Spallotta, F., Gaetano, C., Ori, M., et al., 2014. "RNA-seq of the aging brain in the short-lived fish N. furzeri–conserved pathways and novel genes associated with neurogenesis". *Aging cell,* **13**(6), pp. 965–974.

[303] Zahn, J. M., Poosala, S., Owen, A. B., Ingram, D. K., Lustig, A., Carter, A., Weeraratna, A. T., Taub, D. D., Gorospe, M., Mazan-Mamczarz, K., et al., 2007. "AGEMAP: a gene expression database for aging in mice". *PLoS Genet,* **3**(11), p. e201.

[304] Dice, J. F., and Schimke, R. T., 1972. "Turnover and exchange of ribosomal proteins from rat liver". *Journal of Biological Chemistry,* **247**(1), pp. 98–111.

[305] Pulk, A., Liiv, A., Peil, L., Maiväli, Ü., Nierhaus, K., and Remme, J., 2010. "Ribosome reactivation by replacement of damaged proteins". *Molecular microbiology,* **75**(4), pp. 801–814.

[306] Komili, S., Farny, N. G., Roth, F. P., and Silver, P. A., 2007. "Functional specificity among ribosomal proteins regulates gene expression". *Cell,* **131**(3), pp. 557–571.

[307] Malinovskaya, E. M., Ershova, E. S., Golimbet, V. E., Porokhovnik, L. N., Lyapunova, N. A., Kutsev, S. I., Veiko, N. N., and Kostyuk, S. V., 2018. "Copy number of human ribosomal genes with aging: unchanged mean, but narrowed range and decreased variance in elderly group". *Frontiers in genetics,* **9**, p. 306.

[308] Parks, M. M., Kurylo, C. M., Dass, R. A., Bojmar, L., Lyden, D., Vincent, C. T., and Blanchard, S. C., 2018. "Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression". *Science advances,* **4**(2), p. eaao0665.

[309] Gupta, V., and Warner, J. R., 2014. "Ribosome-omics of the human ribosome". *Rna,* **20**(7), pp. 1004–1013.

[310] Chaillou, T., Zhang, X., and McCarthy, J. J., 2016. "Expression of Muscle-Specific Ribosomal Protein L3-Like Impairs Myotube Growth". *Journal of cellular physiology,* **231**(9), pp. 1894–1902.

[311] Sugihara, Y., Honda, H., Iida, T., Morinaga, T., Hino, S., Okajima, T., Matsuda, T., and Nadano, D., 2010. "Proteomic analysis of rodent ribosomes revealed heterogeneity including ribosomal proteins L10-like, L22-like 1, and L39-like". *Journal of proteome research,* **9**(3), pp. 1351–1366.

[312] O'Leary, M. N., Schreiber, K. H., Zhang, Y., Duc, A.-C. E., Rao, S., Hale, J. S., Academia, E. C., Shah, S. R., Morton, J. F., Holstein, C. A., et al., 2013. "The ribosomal protein Rpl22 controls ribosome composition by directly repressing expression of its own paralog, Rpl22l1". *PLoS Genet,* **9**(8), p. e1003708.

[313] Wong, Q. W.-L., Li, J., Ng, S. R., Lim, S. G., Yang, H., and Vardy, L. A., 2014. "RPL39L is an example of a recently evolved ribosomal protein paralog that shows highly specific tissue expression patterns and is upregulated in ESCs and HCC tumors". *RNA biology,* **11**(1), pp. 33–41.

[314] Kondrashov, N., Pusic, A., Stumpf, C. R., Shimizu, K., Hsieh, A. C., Xue, S., Ishijima, J., Shiroishi, T., and Barna, M., 2011. "Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning". *Cell,* **145**(3), pp. 383–397.

[315] Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., and Barna, M., 2015. "RNA regulons in Hox 5′ UTRs confer ribosome specificity to gene regulation". *Nature,* **517**(7532), pp. 33–38.

[316] Werner, A., Iwasaki, S., McGourty, C. A., Medina-Ruiz, S., Teerikorpi, N., Fedrigo, I., Ingolia, N. T., and Rape, M., 2015. "Cell-fate determination by ubiquitin-dependent regulation of translation". *Nature,* **525**(7570), pp. 523–527.

[317] Zeidan, Q., Wang, Z., De Maio, A., and Hart, G. W., 2010. "O-GlcNAc cycling enzymes associate with the translational machinery and modify core ribosomal proteins". *Molecular biology of the cell,* **21**(12), pp. 1922–1936.

[318] Yang, H. W., Kim, H. D., Kim, T.-S., and Kim, J., 2019. "Senescent cells differentially translate senescence-related mRNAs via ribosome heterogeneity". *The Journals of Gerontology: Series A,* **74**(7), pp. 1015–1024.

[319] Xue, S., and Barna, M., 2012. "Specialized ribosomes: a new frontier in gene regulation and organismal biology". *Nature reviews Molecular cell biology,* **13**(6), pp. 355–369.

[320] Cenik, E. S., Meng, X., Tang, N. H., Hall, R. N., Arribere, J. A., Cenik, C., Jin, Y., and Fire, A., 2019. "Maternal ribosomes are sufficient for tissue diversification during embryonic development in C. elegans". *Developmental cell,* **48**(6), pp. 811–826.

[321] Khajuria, R. K., Munschauer, M., Ulirsch, J. C., Fiorini, C., Ludwig, L. S., McFarland, S. K., Abdulhay, N. J., Specht, H., Keshishian, H., Mani, D. R., et al., 2018. "Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis". *Cell,* **173**(1), pp. 90–103.

[322] Lodish, H. F., 1974. "Model for the regulation of mRNA translation applied to haemoglobin synthesis". *Nature,* **251**(5474), pp. 385–388.

[323] Slavov, N., Semrau, S., Airoldi, E., Budnik, B., and van Oudenaarden, A., 2015. "Differential stoichiometry among core ribosomal proteins". *Cell reports,* **13**(5), pp. 865–873.

[324] Nakano, K., and Sidransky, H., 1978. "Age-related changes in ribosomal profiles and in vitro protein synthesis in skeletal muscle during fasting and subsequent refeeding of rats". *The Journal of nutrition,* **108**(3), pp. 399–409.

[325] Mori, N., Mizuno, D., and Goto, S., 1979. "Conservation of ribosomal fidelity during ageing". *Mechanisms of ageing and development,* **10**(6), pp. 379–398.

[326] Rattan, S. I., 1996. "Synthesis, modifications, and turnover of proteins during aging". *Experimental gerontology,* **31**(1-2), pp. 33–47.

[327] Gonskikh, Y., and Polacek, N., 2017. "Alterations of the translation apparatus during aging and stress response". *Mechanisms of ageing and development,* **168**, pp. 30–36.

[328] Fox, J. G., Barthold, S., Davisson, M., Newcomer, C. E., Quimby, F. W., and Smith, A., 2006. *The mouse in biomedical research: normative biology, husbandry, and models*, Vol. 3. Elsevier.

[329] Zhao, L., Mao, Z., Woody, S. K., and Brinton, R. D., 2016. "Sex differences in metabolic aging of the brain: insights into female susceptibility to Alzheimer's disease". *Neurobiology of aging,* **42**, pp. 69–79.

[330] Finlay, B. L., and Darlington, R. B., 1995. "Linked regularities in the development and evolution of mammalian brains". *Science,* **268**(5217), pp. 1578–1584.

[331] Emmott, E., Jovanovic, M., and Slavov, N., 2019. "Ribosome stoichiometry: from form to function". *Trends in biochemical sciences,* **44**(2), pp. 95–109.

[332] Reschke, M., Clohessy, J. G., Seitzer, N., Goldstein, D. P., Breitkopf, S. B., Schmolze, D. B., Ala, U., Asara, J. M., Beck, A. H., and Pandolfi, P. P., 2013. "Characterization and analysis of the composition and dynamics of the mammalian riboproteome". *Cell reports,* **4**(6), pp. 1276–1287.

[333] Thompson, A., Zakon, H. H., and Kirkpatrick, M., 2016. "Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes". *Genetics,* **202**(2), pp. 765–774.

[334] Steele, R. E., Thomas, P. S., and Reeder, R. H., 1984. "Anucleolate frog embryos contain ribosomal DNA sequences and a nucleolar antigen". *Developmental biology,* **102**(2), pp. 409–416.

[335] Kurylo, C. M., Parks, M. M., Juette, M. F., Zinshteyn, B., Altman, R. B., Thibado, J. K., Vincent, C. T., and Blanchard, S. C., 2018. "Endogenous rRNA sequence variation can regulate stress response gene expression and phenotype". *Cell reports,* **25**(1), pp. 236–248.

[336] Song, W., Joo, M., Yeom, J.-H., Shin, E., Lee, M., Choi, H.-K., Hwang, J., Kim, Y.-I., Seo, R., Lee, J. E., et al., 2019. "Divergent rRNAs as regulators of gene expression at the ribosome level". *Nature microbiology,* **4**(3), pp. 515–526.

[337] Wang, Y. J., Vaidyanathan, P. P., Rojas-Duran, M. F., Udeshi, N. D., Bartoli, K. M., Carr, S. A., and Gilbert, W. V., 2018. "Lso2 is a conserved ribosome-bound protein required for translational recovery in yeast". *PLoS biology,* **16**(9), p. e2005903.

[338] Lee, C.-H., Kiparaki, M., Blanco, J., Folgado, V., Ji, Z., Kumar, A., Rimesso, G., and Baker, N. E., 2018. "A regulatory response to ribosomal protein mutations controls translation, growth, and cell competition". *Developmental cell,* **46**(4), pp. 456–469.

[339] Imami, K., Milek, M., Bogdanow, B., Yasuda, T., Kastelic, N., Zauber, H., Ishihama, Y., Landthaler, M., and Selbach, M., 2018. "Phosphorylation of the ribosomal protein RPL12/uL11 affects translation during mitosis". *Molecular cell,* **72**(1), pp. 84–98.

[340] Mazumder, B., Sampath, P., Seshadri, V., Maitra, R. K., DiCorleto, P. E., and Fox, P. L., 2003. "Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control". *Cell,* **115**(2), pp. 187–198.

[341] R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[342] Ban, N., Beckmann, R., Cate, J. H., Dinman, J. D., Dragon, F., Ellis, S. R., Lafontaine, D. L., Lindahl, L., Liljas, A., Lipton, J. M., et al., 2014. "A new system for naming ribosomal proteins". *Current opinion in structural biology,* **24**, pp. 165–169.

Appendix A

Supplementary information

Table A.1: Plasmids used in this study.

Plasmid name	Description
Genetic circuit plasmids	
psB3C5_araC_delta	Generation 3 – ORI exchanged, araC included
psB3C5_araC_pag	Generation 3 – ORI exchanged, araC included
psB3C5_araC_ogr	Generation 3 – ORI exchanged, araC included
psB3C5_delta	Generation 2 – vector exchange pSB3C5, p15A ORI
psB3C5_pag	Generation 2 – vector exchange pSB3C5, p15A ORI
psB3C5_ogr	Generation 2 – vector exchange pSB3C5, p15A ORI
pGGA_delta	Generation 1 – vector pGGA pUC13-derived pMB1 ORI
pGGA_pag	Generation 1 – vector pGGA pUC13-derived pMB1 ORI
pGGA_ogr	Generation 1 – vector pGGA pUC13-derived pMB1 ORI
sRNA plasmids	
delta_sRNA_H	sRNA high repression against delta TF
delta_sRNA_M	sRNA medium repression against delta TF
delta_sRNA_L	sRNA low repression against delta TF
pag_sRNA_H	sRNA high repression against pag TF
pag_sRNA_M	sRNA medium repression against pag TF
pag_sRNA_L	sRNA low repression against pag TF
ogr_sRNA_H	sRNA high repression against ogr TF
ogr_sRNA_M	sRNA medium repression against ogr TF
ogr_sRNA_L	sRNA low repression against ogr TF
scr_sRNA	sRNA dysfunctional (control)

Table A.2: Oligonucleotide primers used in this study. fw, forward primer, rev, reverse primer. Oligonucleotides were purchased from Microsynth in desalted quality. Fluorescent labelled oligonucleotides were ordered in PAGE-purified quality. Lyophilized oligonucleotides were resuspended in water to reach a concentration of 100 µM and subsequently stored at -20°C.

Name	Sequence (5′ - 3′)
Primers for sequencing	
VF2	TGCCACCTGACGTCTAAGAA
VR	ATTACCGCCTTTGAGTGAGC

GGA_cPCR_fw	CTGCAGGAAGGTTTAAACGCATTTAGG
analysis_fw	CTGCAGGAAGGTTTAAACGCATTTAGG
analysis_rev	TAATACGACTCACTATAGGGAGACGC

Fluorescent oligonucleotides for sRNA probing, 5'-Cy3 modified

delta_sRNA_probe	ATGATGCGCTGCCCTTTCTGTCGT
pag_sRNA_probe	ATGATGCACTGCCCGTTATGCCAA
ogr_sRNA_probe	GAAAATGTTTCATTGTCCTTTATGCCAGC

Primers for RT PCR

ogr 2_fw	CCTTTATGCCAGCATGCCGCAC
ogr 2_rev	GCAACGGGTGCGGCCTTAC
ogr_sRNA_fw	GCTGGCATAAAGGACAATGAAAC
ogr_sRNA_rev	ACGACTGTTCGGGCTTGTC
pag_sRNA_fw	GGCATAACGGGCAGTGCATC
pag_sRNA_rev	CGGACGACTGTTCGGGCTTG
delta_sRNA_fw	CGACAGAAAGGGCAGCGCATC
delta_sRNA_rev	AGCCCGGACGACTGTTCG

Primers for TF-GFP circuit 3nd Generation cloning
(araC subcloned into 2nd Generation plasmids)

BsmBI_AraC_fw	GAGGTTCGTCTCCGCGGGGTTTTTTGCGTTATGACAACT TGACGGCTACATCATTCAC
BsmBI_pBAD_rev	GAGGTTCGTCTCGGCTAGCCCAAAAAAACGGGTATGGAG
BsmBI_TF_fw	CAAATTCGTCTCCTAGCTCTAGAAAAGAGGAGAAAAAGCT TATG
BsmBI_PF_rev	GAGGTTGGTCTCGTGCCCGCCTTTTCTTTACCGG

Primers for TF-GFP circuit 2nd Generation cloning
(TF-GFP cassetes subcloned into pSB3C5 vector)

cassette_GGA_fw	GAATTCGTCTCTCTAGAGCGAAGACTTGGTACGGAGTATA AACGCA
cassette_GGA_rev	GATATCGTCTCTTAGTAAAGAAGACTACAGGAATGGTAA ACGCAG
Pre_Backbone_GGA_rev	CAATTCGTCTCTCTAGAAGCGGCCGCGAATTC
Suf_Backbone_GGA_fw	CATATCGTCTCTACTAGTAGCGGCCGCTGCAG

Primers for 1st Generation TF-GFP circuit cloning

BsaI_GFP_rev	GAGGTTGGTCTCGGGAGTATAAACGCAGAAAGGCCCACCC
BsaI_ogr_rev	GAGGTTGGTCTCGATGGTAAACGCAGAAAGGCCCACCC
BsaI_GFP_fww	GAGGTTGGTCTCGCACTATTAAAGAGGAGAAATACTAGAT GCGTAAAGG
BsaI_PF_fw	GAGGTTGGTCTCGAGTGGTATTTTCAGTTGCGCCGG
BsaI_PF_rev	GAGGTTGGTCTCGTGCCCGCCTTTTCTTTACCGG
BsaI_pBAD_fw	GAGGTTGGTCTCGGGCAAACATTGATTATTTGCACGGCG TCAC
BsaI_pBAD_rev	GAGGTTGGTCTCGGCTAGCCCAAAAAAACGG

BsaI_ogr_fw	GAGGTTGGTCTCGGCTTATGTTTCATTGTCCTTTATGCCA GCATGC
BsaI_PSP3_fw	GAGGTTGGTCTCGGCTTATGATGCACTGCCCGTTATGCC
BsaI_PhiR73_fw	GAGGTTGGTCTCGGCTTATGATGCGCTGCCCTTTCTGTC
XbaI-RBS-HindIII_fw	TAGCTCTAGAAAAGAGGAGAAAAA
XbaI_RBS_HindIII_rev	AAGCTTTTTCTCCTCTTTTCTAGA
Primers for sRNA cloning (into pSB1A3)	
Bsa1_micC_Suf_fw	CAATTGGTCTCGTTTCTGTTGGGCCATTGCATTGCCACTG ATTTTCCAACATATAAAAAGACAAGCCCGAACAGTCGTCC GGGCTTTTTTTTACTAGTAGCGGCCGCTGCAGTCC
pag_sRNA_h_Pre_rev	CAATTGGTCTCGGAAAATGATGCACTGCCCGTTATGCCAA TATAGTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCC AG
pag_sRNA_m_Pre_rev	CAATTGGTCTCGGAAAATGATGCACTGCCCGTTATGTATA GTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCCAG
pag_sRNA_l_Pre_rev	CAATTGGTCTCGGAAAATGATGCACTGTCCGATATGTATA GTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCCAG
ogr_sRNA_h_Pre_rev	CAATTGGTCTCGGAAAATGTTTCATTGTCCTTTATGCCAG CTATAGTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTC CAG
ogr_sRNA_m_Pre_rev	CAATTGGTCTCGGAAAGTTTCATTGTCTTTATGCCAGCTA TAGTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCCAG
ogr_sRNA_l_Pre_rev	CAATTGGTCTCGGAAAGTTCATTGTCTTATGCCAGCTATA GTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCCAG
delta_sRNA_h_Pre_rev	CAATTGGTCTCGGAAAATGATGCGCTGCCCTTTCTGTCGT TATAGTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCC AG
delta_sRNA_m_Pre_rev	CAATTGGTCTCGGAAAATGATGCGCTGCTCTTTCTGTTAT AGTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCCAG
delta_sRNA_l_Pre_rev	CAATTGGTCTCGGAAAATGATTCGCTGCTCTTTCTGTATA GTGAGTCGTATTACTCTAGAAGCGGCCGCGAATTCCAG
scr_sRNA_rev	CAATTGGTCTCGGAAATACCGGCAGATTCAGTTCTGTATA GTGAGTCGTATTACTCTAGAAGCGGCCGCG
OmpR_sRNA_h_rev	CAATTGGTCTCGGAAAATGCAAGAGAACTACAAGATTCTG GTATAGTGAGTCGTATTACTCTAGAAGCGGCCGCG
OmpR_sRNA_m_rev	CAATTGGTCTCGGAAAATGCAAGAGAACTACAAGATTCGT TATAGTGAGTCGTATTACTCTAGAAGCGGCCGCG
OmpR_sRNA_l _rev	CAATTGGTCTCGGAAAATGCAAGAGAACTATAAAATTCGT TATAGTGAGTCGTATTACTCTAGAAGCGGCCGCG

Table A.3: Oligonucleotide primers used for the GGA of sRNA against tRNA construct. fw, forward primer, rev, reverse primer. Oligonucleotides were purchased from Microsynth in desalted quality. Lyophilized oligonucleotides were resuspended in water to reach a concentration of 100 µM and subsequently stored at -20°C.

Name	Sequence (5′ - 3′)
Primers for sequencing	
VF2	TGCCACCTGACGTCTAAGAA
VR	ATTACCGCCTTTGAGTGAGC

pBAD33_Seq1_fw	TCGTTCCACTGAGCGTCAGACC
rrnB_T1_Seq_fw	CAAATAAAACGAAAGGCTCAGTCGAAAGAC
tetR_seq_fw	CAGCTTTCCCCTTCTAAAGG
Oligonucleotides for GGA	
pSB3C5_rev	GCTCTTCGTCTCCCTCTAGAAGCGGCCGCGAAT
BsmB1_AraC_fw	GAGGTTCGTCTCCGCGGGGTTTTTTGCGTTATGACAACTTG ACGGCTACATCATTCAC
pBAD33_GFP_TGA fw	GCTCTTCGTCTCCAGAGTTATGACAACTTGACGGCTACATC
BsmB1_pBAD_rev	GAGGTTCGTCTCGGCTAGCCCAAAAAAACGGGTATGGAG
pBAD33_GFP_TGA rev	GCTCTTCGTCTCCACGCAAAAAGGCCATCCGTC
BBa_P0440_fw	GCTCTTCGTCTCCGCGTTTATAAACGCAGAAAGGCCCACC
BBa_P0440_rev	GCTCTTCGTCTCCCAAAAGAGGAGAAATACTAGATGTCCA GATT
BBa_R0011_fw	GCTCTTCGTCTCGTTTGTGCTCAGTATCTTGTTATCCGC
BBa_R0011_rev	GCTCTTCGTCTCGTGGGCTTAATTGTGAGCGGATAACAAT TGACA
BBa_R0040 p(tetR)_fw	GCTCTTCGTCTCGCCCATTTCTCGCTCCCTATCAGTGATA GAGATTGACATC
BBa_R0040 p(tetR) rev	GCTCTTCGTCTCGGACAGGGGTTCGATTCCCCTCCGCGTG CTCAGTATCTCTATCACTGATAGG
sRNA_TS2_fw	GCTCTTCGTCTCGTGTCTCGCGGCTCTGAAGGCCGCTGTT TCTGTTGGGCCATTGCATT
micC_seq_rev	GACGACTGTTCGGGCTTGTC
sRNA_TS2_rev	GCTCTTCGTCTCGGTAAAAAAAAGCCCGGACGACTGTTC
pSB3C5_fw	GCTCTTCGTCTCCTTACTAGTAGCGGCCGCTGCA

Table A.4: Materials used in this study. Details on supplier or purpose.

Name	Details
Reagents	
RedSafe™ Nucleic Acid Staining Solution (20,000 x)	iNtRON biotechnology
LE Agarose for gel electrophoresis	Biozym
TRIzol reagent	Invitrogen
Acid-Phenol:Chloroform:/Isoamylalcohol (125:24:1)	Ambion
Ribolock RNase Inhibitor (40 U/µL)	Thermo Fisher Scientific
Kits	
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific
GeneJET PCR purification Kit	Thermo Fisher Scientific
GeneJET Gel Extraction Kit	Thermo Fisher Scientific
Enzymes	
EcoRI	New England Biolabs
BsmBI	New England Biolabs
DpnI (10 U/µ)	Thermo Fisher Scientific

T4 DNA Ligase (2 U/µL)	Thermo Fisher Scientific
Phusion High-Fidelity DNA polymerase (2 U/µ)	Thermo Fisher Scientific
Fast digest BsaI	Thermo Fisher Scientific
Fast digest PstI	Thermo Fisher Scientific
DNase I, RNase free (1 U/µL)	Thermo Fisher Scientific
RevertAid Reverse Transcriptase (200 U/µL)	Thermo Fisher Scientific

Other

Amersham Hybond™-N (nylon membrane for optimized nucleic transfer)	GE Healthcare
LB-Medium (Luria/Miller)	Carl Roth
LB-Agar (Luria/Miller)	Carl Roth
Costar™ 96-well assay plate (black, clear bottom)	Corning

Vectors

pGGA	expression of Generation 1 - genetic circuits
pSB3C5	expression of Generation 2 and 3 - genetic circuits
pSB1A3	sRNA expression

***E. coli* cell strains**

DH5α	Cloning
BL21 (DE3)	Tandem expression of genetic circuits and sRNAs, *in vivo* verification of sRNA-modulating gene, expression, verification of nonsense suppressor function by *in vivo* GFP read-through assay
XL1-blue	Replication of nonsense suppressor tRNAs expression, *in vivo* GFP read-through assay

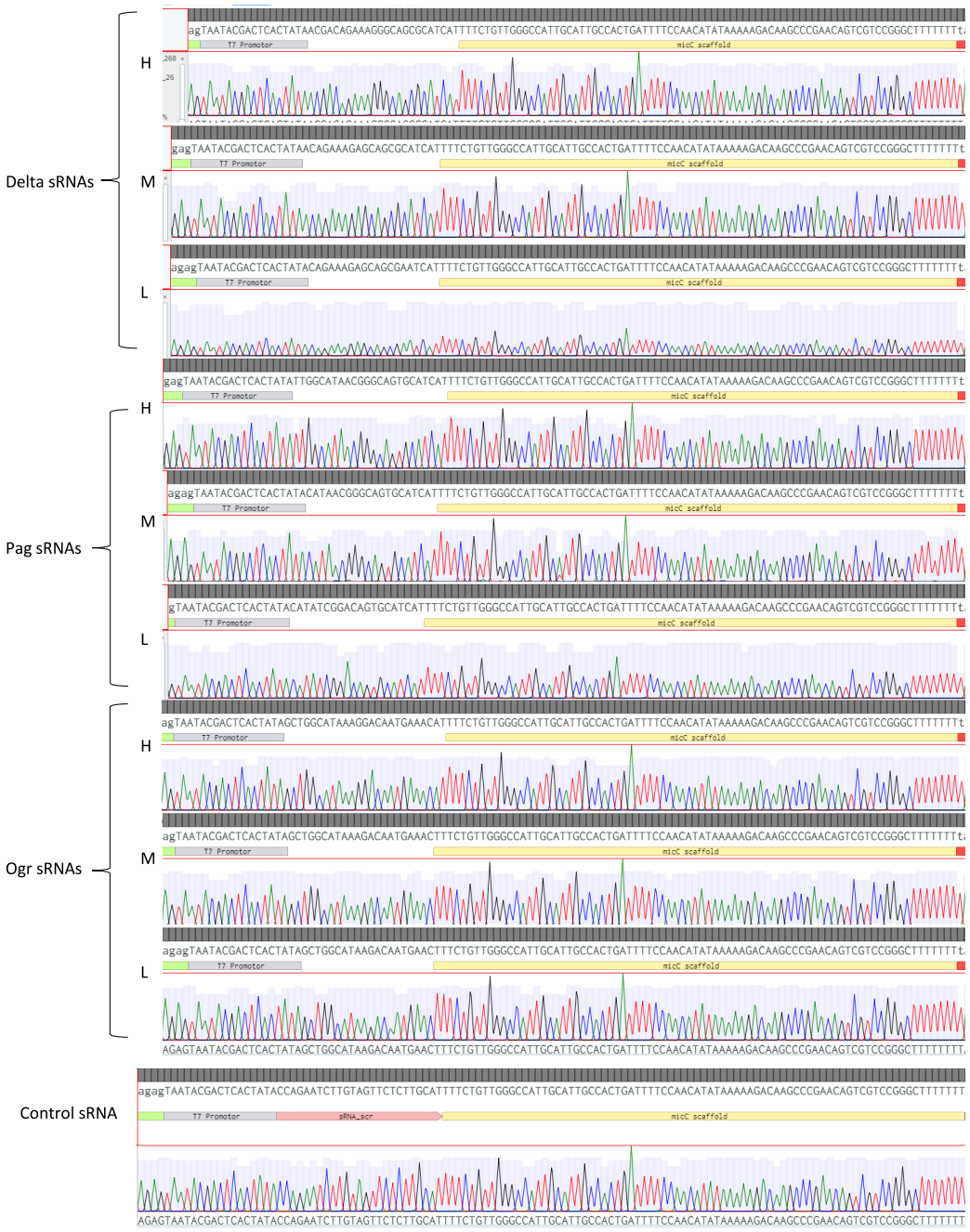


Figure A.1: Screenshot of sequencing confirmation of sRNA constructs. No mismatches; promoter (grey), target-binding sequence (clear) followed by, MicC scaffold sequence (yellow). Alignments were performed with Clustal Omega algorithm (it is precise, it does not reverse sequences) using Benchling.

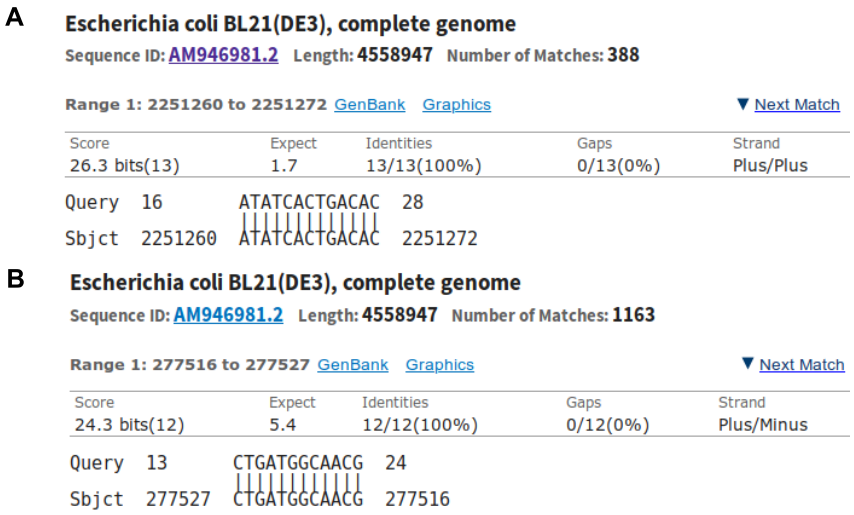


Figure A.2: Screenshot of Blast alignments of primer sequences with *E.coli* genome. (A) *ogr* transcript forward and **(B)** reverse primers.

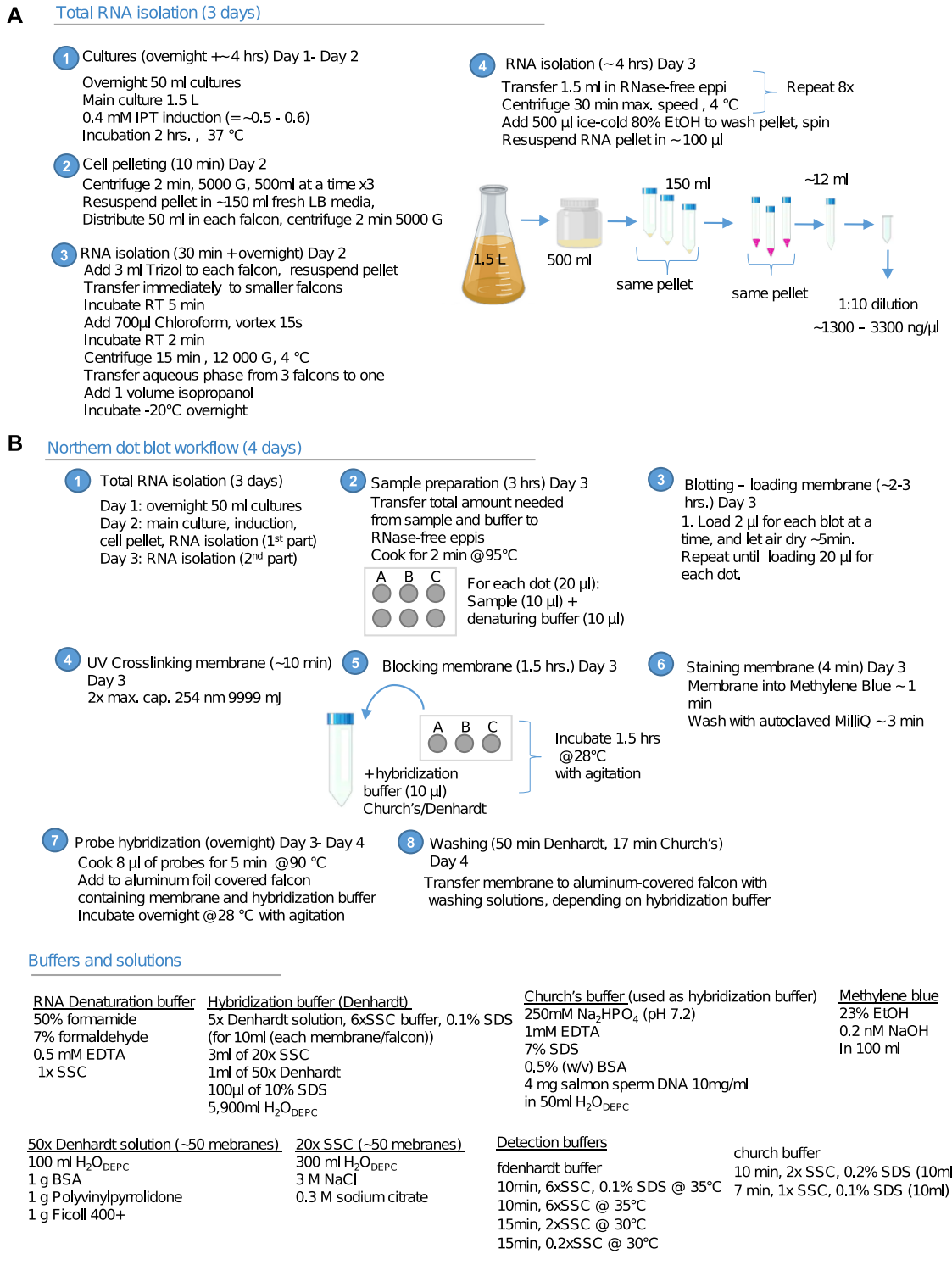
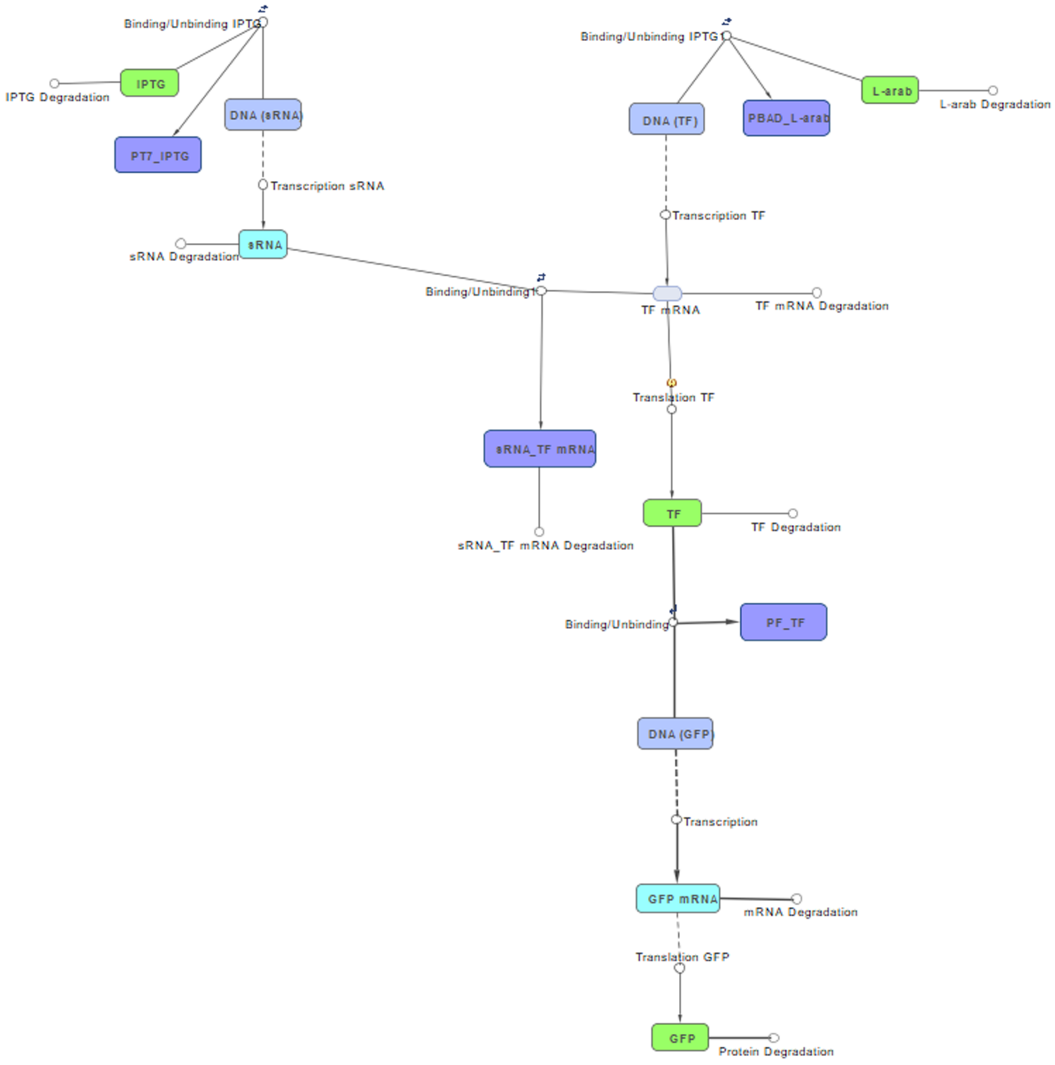


Figure A.3: Protocols of debugging strategies for tandem expression optimization. (A) Protocol for customized RNA isolation that allows for obtaining total RNA concentrations in the kilo order. **(B)** Protocol for northern dot blot for sRNA production verification.



1	Reactions		
2	Transcription	▸	[TF-GFP synthesis].[DNA (GFP)] -> [TF-GFP synthesis].[DNA (GFP)] + [TF-GFP synthesis].[GFP mRNA]
3	Binding/Unbinding	▸	[TF-GFP synthesis].[DNA (GFP)] + [TF-GFP synthesis].TF <-> [TF-GFP synthesis].PF_TF
4	mRNA Degradation	▸	[TF-GFP synthesis].[GFP mRNA] -> *null*
5	Protein Degradation	▸	[TF-GFP synthesis].GFP -> *null*
6	IPTG Degradation	▸	[TF-GFP synthesis].IPTG -> *null*
7	sRNA Degradation	▸	[TF-GFP synthesis].sRNA -> *null*
8	Transcription sRNA	▸	[TF-GFP synthesis].[DNA (sRNA)] -> [TF-GFP synthesis].sRNA + [TF-GFP synthesis].[DNA (sRNA)]
9	Binding/Unbinding IPTG	▸	[TF-GFP synthesis].[DNA (sRNA)] + [TF-GFP synthesis].IPTG <-> [TF-GFP synthesis].PT7_IPTG
10	Binding/Unbinding IPTG1	▸	[TF-GFP synthesis].[L-arab] + [TF-GFP synthesis].[DNA (TF)] <-> [TF-GFP synthesis].[PBAD_L-arab]
11	Transcription TF	▸	[TF-GFP synthesis].[DNA (TF)] -> [TF-GFP synthesis].[DNA (TF)] + [TF-GFP synthesis].[TF mRNA]
12	TF mRNA Degradation	▸	[TF-GFP synthesis].[TF mRNA] -> *null*
13	L-arab Degradation	▸	[TF-GFP synthesis].[L-arab] -> *null*
14	TF Degradation	▸	[TF-GFP synthesis].TF -> *null*
15	Binding/Unbinding1	▸	[TF-GFP synthesis].sRNA + [TF-GFP synthesis].[TF mRNA] <-> [TF-GFP synthesis].[sRNA_TF mRNA]
16	sRNA_TF mRNA Degradation	▸	[TF-GFP synthesis].[sRNA_TF mRNA] -> *null*
17	Translation TF	▸	[TF-GFP synthesis].[TF mRNA] -> [TF-GFP synthesis].TF
18	Translation GFP	▸	[TF-GFP synthesis].[GFP mRNA] -> [TF-GFP synthesis].GFP + [TF-GFP synthesis].[GFP mRNA]

Figure A.4: Screenshot of reaction pathway of the gene regulation simulation. Modeled with MATLAB SimBiology [206].

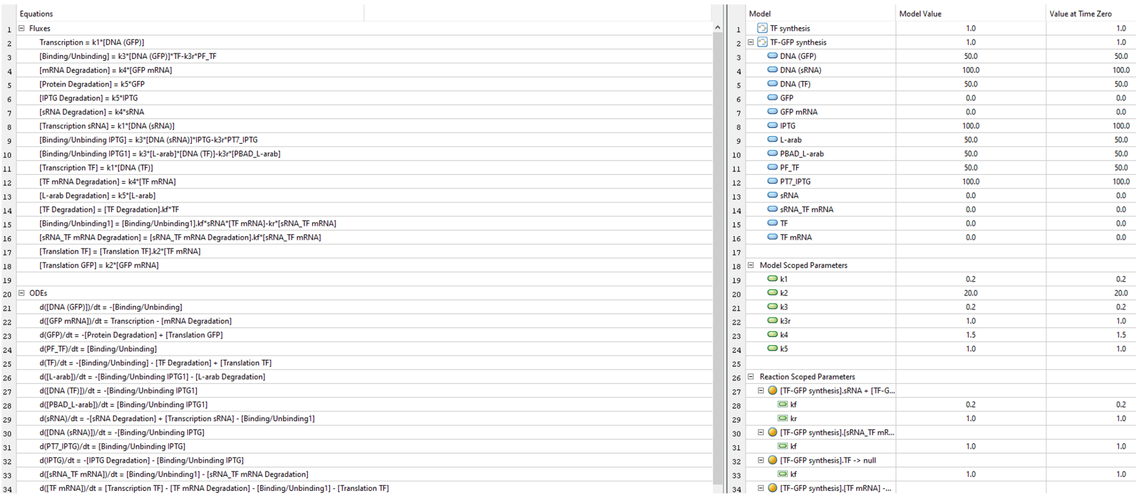


Figure A.5: Screenshot of details on the simulation of the gene regulation model. Specifications: reactions, rates, ODE's of reactions, and parameters for the simulation, modeled and performed with MATLAB SimBiology [206].

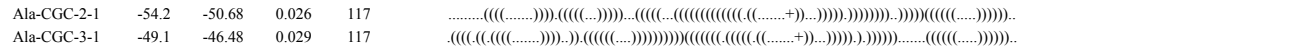
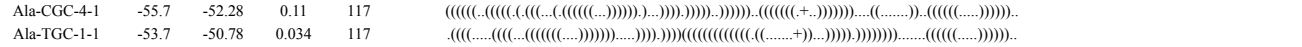
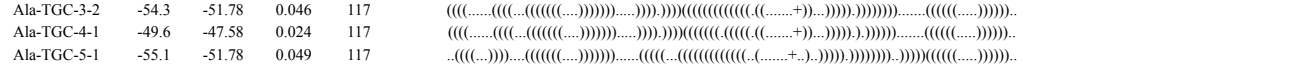
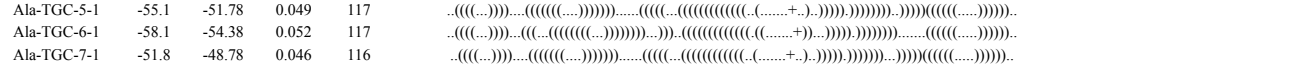
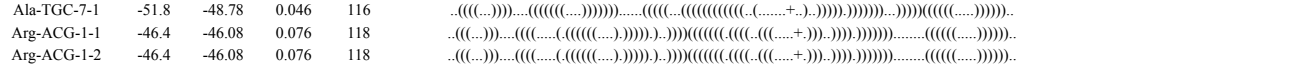
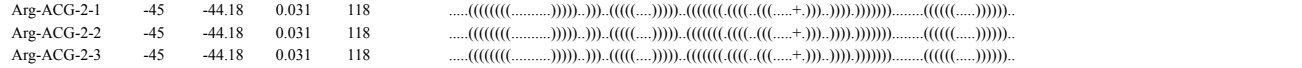
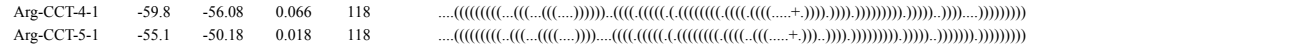
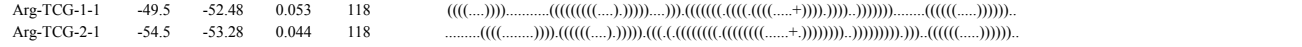
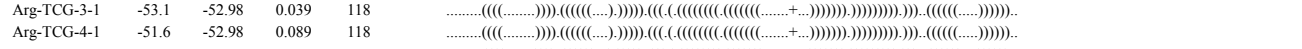
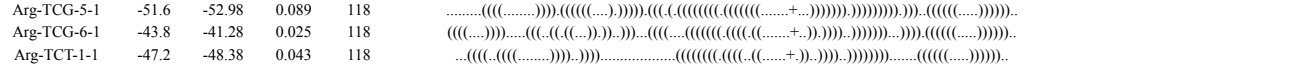
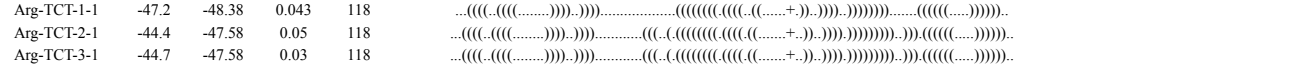
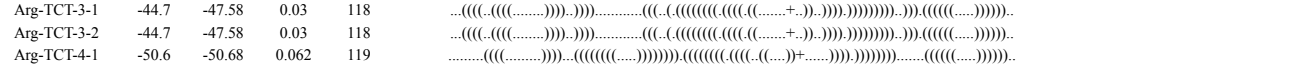
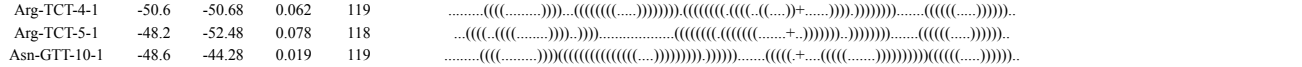
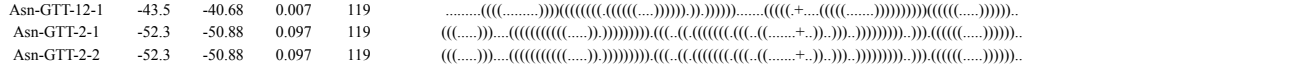
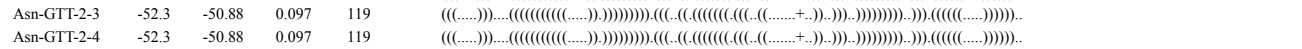
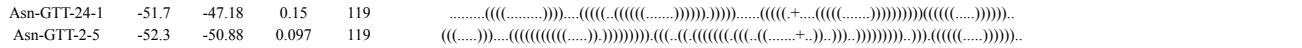
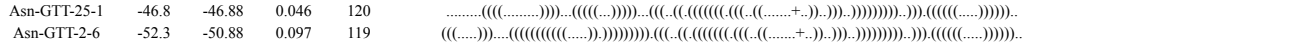
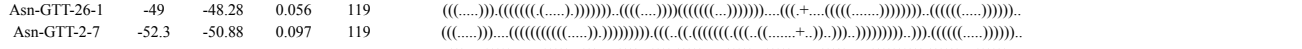
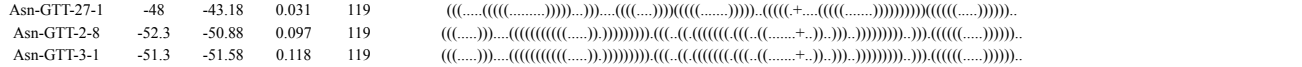
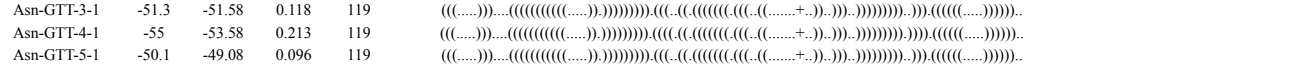
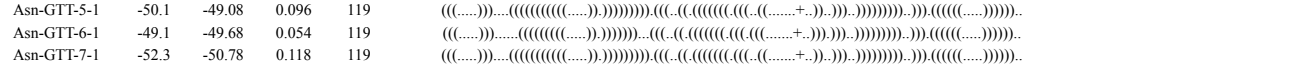
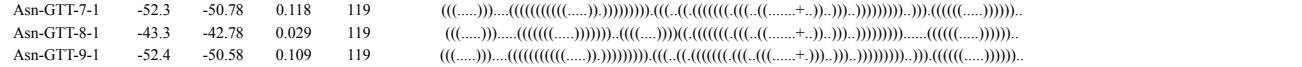
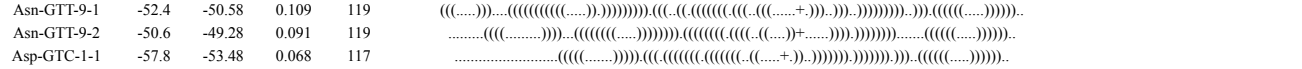
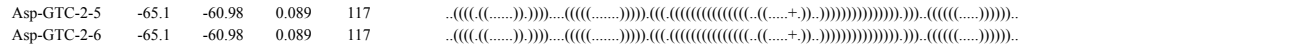
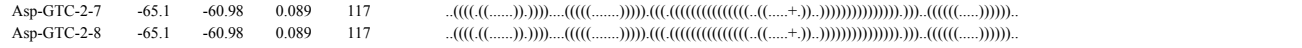
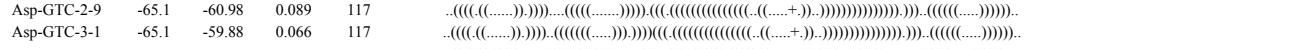
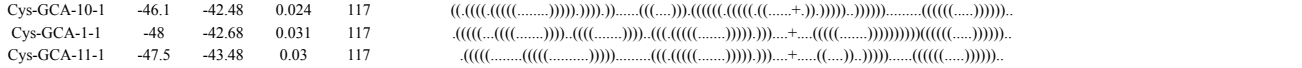
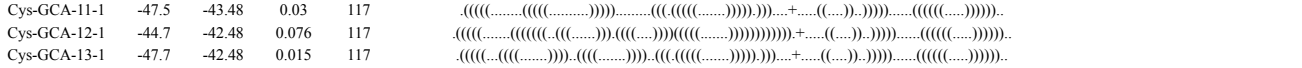
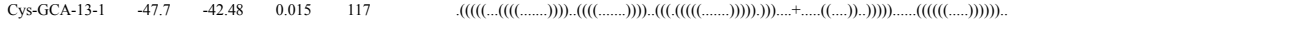
Table A.5: Screening synthetic t1A3T2 sRNA against natural occurring tRNAs in *E.coli* strain K12 MG1655. BE - binding energy of sRNA and ΔG - free energy of the secondary structure in kcal/mol; P - probability of formation of the secondary structure; MFE - minimum free energy structure.

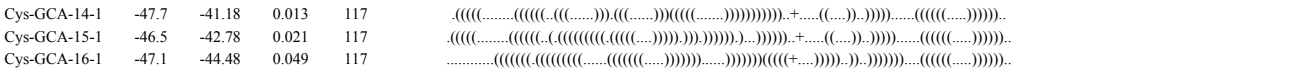
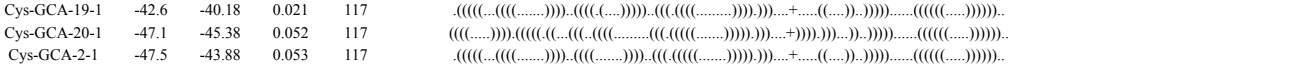
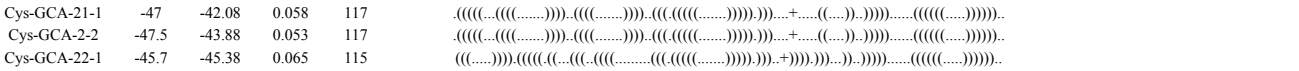
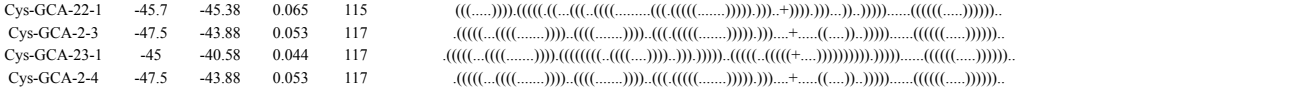
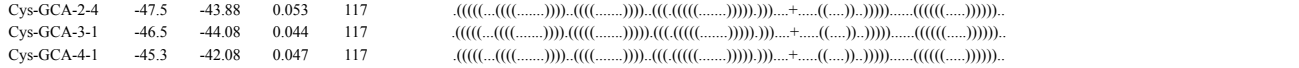
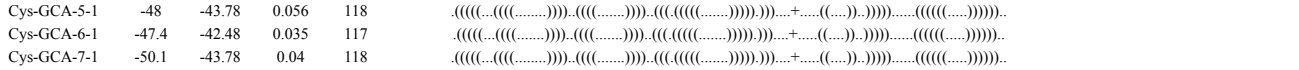
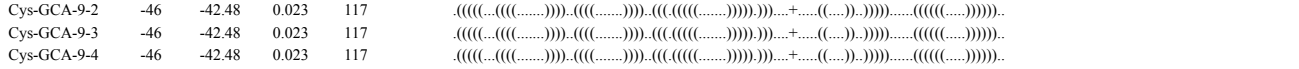
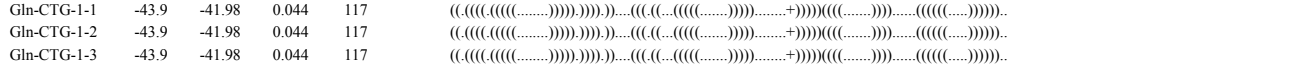
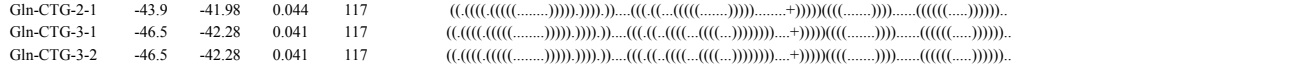
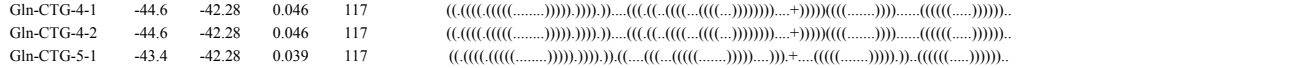
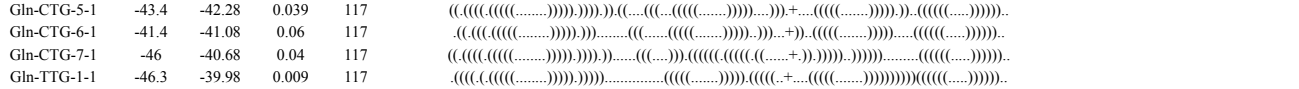
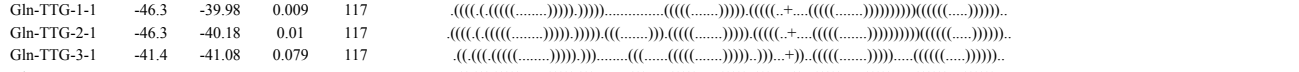
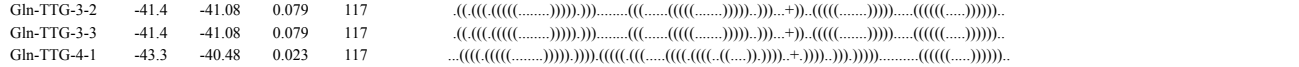
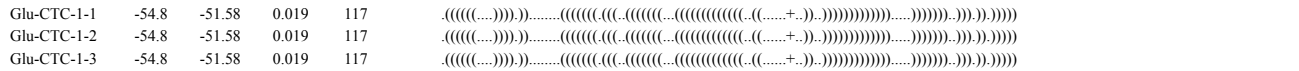
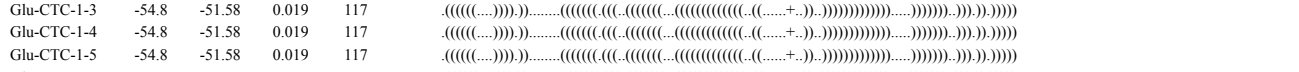
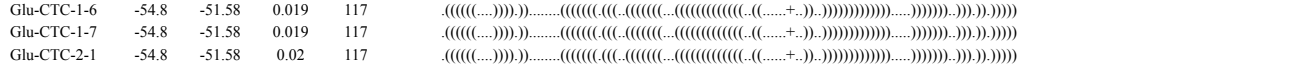
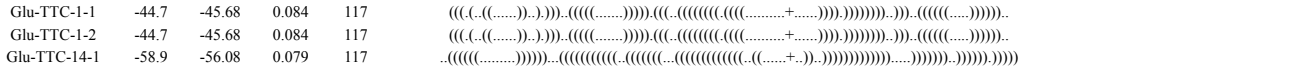
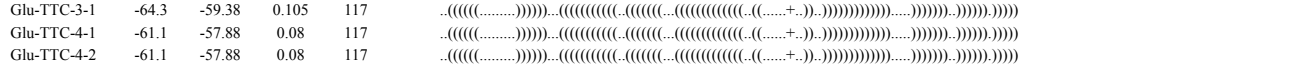
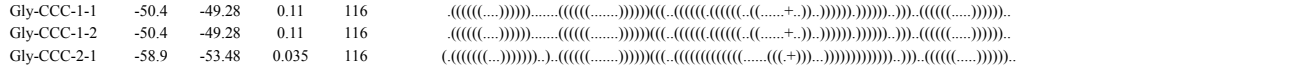
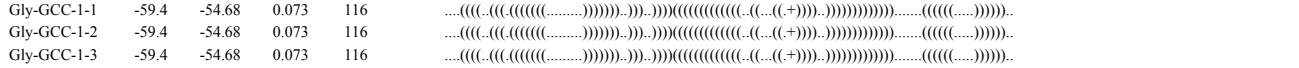
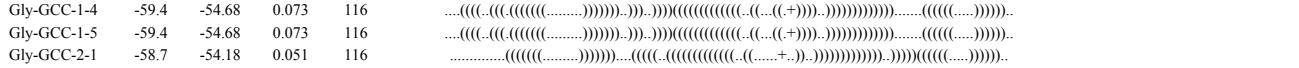
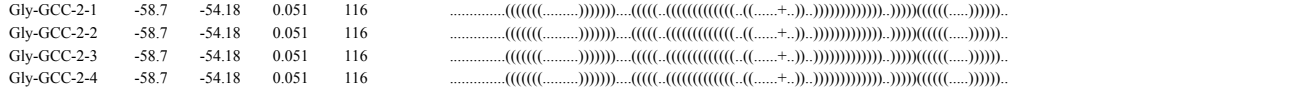
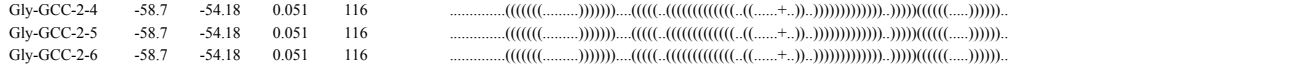
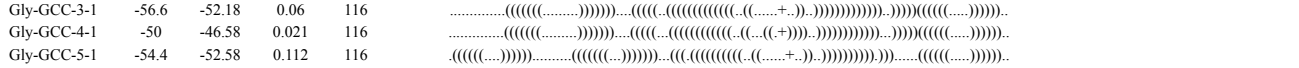
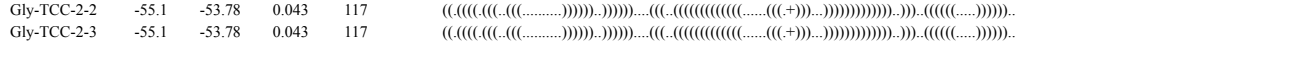
tRNA	BE	ΔG	P	(nt)	MFE
Ala-GGC-1-1	-53.9	-52.68	0.037	121	.((((....)))).(((.....))).((((((...)))))).((.(((.(((((((((.(((.((.......+)).))))))).))))))).))))).((((((.....))))))..
Ala-GGC-1-2	-53.9	-52.68	0.037	121	.((((....)))).(((.....))).((((((...)))))).((.(((.(((((((((.(((.((.......+)).))))))).))))))).))))).((((((.....))))))..
Ala-TGC-1-1	-53.6	-48.68	0.027	121	.((((....))))......(((((....)))))...(((.((((.(.(((((.......))))).).))))....+))).(((((.......)))))....((((((.....))))))..
Ala-TGC-1-2	-53.6	-48.68	0.027	121	.((((....))))......(((((....)))))...(((.((((.(.(((((.......))))).).))))....+))).(((((.......)))))....((((((.....))))))..
Ala-TGC-1-3	-53.6	-48.68	0.027	121	.((((....))))......(((((....)))))...(((.((((.(.(((((.......))))).).))))....+))).(((((.......)))))....((((((.....))))))..
Arg-ACG-1-1	-51.8	-51.48	0.042	122	...((((..(((((((((.((...)).))))))).......))))...))))((((((.(((((((.........+.))).))))).)))))))........((((((.....))))))..
Arg-ACG-1-2	-51.8	-51.48	0.042	122	...((((..(((((((((.((...)).))))))).......))))...))))((((((.(((((((.........+.))).))))).)))))))........((((((.....))))))..
Arg-ACG-1-3	-51.8	-51.48	0.042	122	...((((..(((((((((.((...)).))))))).......))))...))))((((((.(((((((.........+.))).))))).)))))))........((((((.....))))))..
Arg-ACG-1-4	-51.8	-51.48	0.042	122	...((((..(((((((((.((...)).))))))).......))))...))))((((((.(((((((.........+.))).))))).)))))))........((((((.....))))))..
Arg-CCG-1-1	-55	-50.38	0.021	122	...((((..(((((((((.((...)).))))))).......))))...))))((((((.(((((((.........+.))).))))).)))))))........((((((.....))))))..
Arg-CCT-1-1	-46.1	-46.68	0.063	120	.(((............)))....(((.(((((((...(((((..(((((((.......))))))).)).).)).+..)))))))..)))...........((((((.....))))))..
Arg-TCT-1-1	-51.6	-49.28	0.013	122	((((((...((((.........)))).(((((.......))))))))).))).(((((((.((((......(((....+)))...)))).)))))))........((((((.....))))))..
Asn-GTT-1-1	-44.8	-42.88	0.014	121	((((((..(..((..(((...(((((....)))))...)))..)).)(((((.......)))))))))))((..+))..(((((.......)))))....((((((.....))))))..
Asn-GTT-1-2	-44.8	-42.88	0.014	121	((((((..(..((..(((...(((((....)))))...)))..)).)(((((.......)))))))))))((..+))..(((((.......)))))....((((((.....))))))..
Asn-GTT-1-3	-44.8	-42.88	0.014	121	((((((..(..((..(((...(((((....)))))...)))..)).)(((((.......)))))))))))((..+))..(((((.......)))))....((((((.....))))))..
Asn-GTT-1-4	-44.8	-42.88	0.014	121	((((((..(..((..(((...(((((....)))))...)))..)).)(((((.......)))))))))))((..+))..(((((.......)))))....((((((.....))))))..
Asp-GTC-1-1	-61.2	-54.88	0.046	122	.((((....))))..((.....))..((((.(((((...))))))))))...(((((((.((((.((((........+.)))).)))).))))))).......((((((.....))))))..
Asp-GTC-1-2	-61.2	-54.88	0.046	122	.((((....))))..((.....))..((((.(((((...))))))))))...(((((((.((((.((((........+.)))).)))).))))))).......((((((.....))))))..
Asp-GTC-1-3	-61.2	-54.88	0.046	122	.((((....))))..((.....))..((((.(((((...))))))))))...(((((((.((((.((((........+.)))).)))).))))))).......((((((.....))))))..
Cys-GCA-1-1	-48.8	-48.98	0.129	119	(((((((((((......)))))..((((((((...)))))).))).(((((.......)))))))))))).((.+..))(((((.......)))))....((((((.....))))))..
Gln-CTG-1-1	-48.3	-46.58	0.027	120	((((((..(((.........)))...((((((.((((.(((....))).)))).))))))...)))))))((..+))..(((((.......)))))....((((((.....))))))..
Gln-CTG-1-2	-48.3	-46.58	0.027	120	((((((..(((.........)))...((((((.((((.(((....))).)))).))))))...)))))))((..+))..(((((.......)))))....((((((.....))))))..
Gln-TTG-1-1	-48.4	-47.78	0.085	120	((((((((((((.........)))...((..(((((.(((((......)))))))))).)).)))))))))((..+))..(((((.......)))))....((((((.....))))))..
Gln-TTG-1-2	-48.4	-47.78	0.085	120	((((((((((((.........)))...((..(((((.(((((......)))))))))).)).)))))))))((..+))..(((((.......)))))....((((((.....))))))..
Glu-TTC-1-1	-65.1	-65.18	0.119	121	((((.((((.....))))....))))((((((.......)))))).((((((((((((.((((((.((.....)).+....)))))).)))))))))))...((((((.....))))))..
Glu-TTC-1-2	-65.1	-65.18	0.119	121	((((.((((.....))))....))))((((((.......)))))).((((((((((((.((((((.((.....)).+....)))))).)))))))))))...((((((.....))))))..
Glu-TTC-1-3	-65.1	-65.18	0.119	121	((((.((((.....))))....))))((((((.......)))))).((((((((((((.((((((.((.....)).+....)))))).)))))))))))...((((((.....))))))..
Glu-TTC-1-4	-65.1	-65.18	0.119	121	((((.((((.....))))....))))((((((.......)))))).((((((((((((.((((((.((.....)).+....)))))).)))))))))))...((((((.....))))))..
Gly-CCC-1-1	-56.9	-58.38	0.116	119	.........((((.......)))).(((((((...)))))))..(((.((((((((((((((...((((....+))))))))))))))))))).)))...((((((.....))))))..
Gly-GCC-1-1	-57.2	-50.38	0.021	121	((((...(.(((((........)))).((((((...)))))))(((((((((..((.((.(((.(((((.((((.+..)))))))).)))))...)).))))))))))...)..))))..
Gly-GCC-1-2	-57.2	-50.38	0.021	121	((((...(.(((((........)))).((((((...)))))))(((((((((..((.((.(((.(((((.((((.+..)))))))).)))))...)).))))))))))...)..))))..
Gly-GCC-1-3	-57.2	-50.38	0.021	121	((((...(.(((((........)))).((((((...)))))))(((((((((..((.((.(((.(((((.((((.+..)))))))).)))))...)).))))))))))...)..))))..
Gly-GCC-1-4	-57.2	-50.38	0.021	121	((((...(.(((((........)))).((((((...)))))))(((((((((..((.((.(((.(((((.((((.+..)))))))).)))))...)).))))))))))...)..))))..
Gly-TCC-1-1	-56.4	-54.48	0.079	120	(((..(((((......((((.........))))........))))).)))((((((((((((.((..((((....+))))))))))))))))))).......((((((.....))))))..
His-GTG-1-1	-49.3	-48.48	0.026	121	..((((...(((.....)))..))))(((((.......)))))....((.((((((((.((((.((.((.......+)).)).)))).)))))))).))....((((((.....))))))..
Ile-GAT-1-1	-49.1	-47.28	0.035	122	..(((...))).((((.((((.((.....(((((......(((.((((((((((.......)))))..))))))))).+....)))))..)).))))))))....((((((.....))))))..
Ile-GAT-1-2	-49.1	-47.28	0.035	122	..(((...))).((((.((((.((.....(((((......(((.((((((((((.......)))))..))))))))).+....)))))..)).))))))))....((((((.....))))))..
Ile-GAT-1-3	-49.1	-47.28	0.035	122	..(((...))).((((.((((.((.....(((((......(((.((((((((((.......)))))..))))))))).+....)))))..)).))))))))....((((((.....))))))..
Ile2-CAT-1-1	-56.6	-52.18	0.019	121	((((((..((....(((((.((((.(((((((...........))))))).)))).)).))))))))))).((.+..))(((((.......)))))....((((((.....))))))..
Ile2-CAT-2-1	-54.2	-50.18	0.017	121	((((((..((....(((((.((((.(((((((...........))))))).)))).)).))))))))))).((.+..))(((((.......)))))....((((((.....))))))..
Leu-CAA-1-1	-53	-50.78	0.007	130	(((.....)))......((.....))..(((((((...)))))))((((....))))....(((((((.(((..((((((......+.)))))))).))))))).........((((((.....))))))..
Leu-CAG-1-1	-57.3	-55.48	0.082	132	..((((.((((((............)))))).))))..........((((....))))...(((((((.(.((((((....((((....+)))).))))).).)))))))......((((((.....))))))..
Leu-CAG-1-2	-57.3	-55.48	0.082	132	..((((.((((((............)))))).))))..........((((....))))...(((((((.(.((((((....((((....+)))).))))).).)))))))......((((((.....))))))..
Leu-CAG-1-3	-57.3	-55.48	0.082	132	..((((.((((((............)))))).))))..........((((....))))...(((((((.(.((((((....((((....+)))).))))).).)))))))......((((((.....))))))..

Leu-CAG-2-1	-56.7	-54.58	0.041	132	((.......)).......((((..((((((((.......))))))))..)))).(((..((((((((.(.(((((....((((....+)))).))))).).))))))))))..((((((.....))))))..
Leu-GAG-1-1	-59.9	-54.78	0.051	132	..((((((((((.............))))))))))).....((((.((((....))))))))((((((.(.((((.(((.........+..))).)))).).))))))........((((((.....))))))..
Leu-TAA-1-1	-56.8	-57.08	0.041	132	..((((.(.....).)))).(((((..(((((((...)))))))(((((....)))))...((((((.(.((((.(((((.......+.))))))))).).)))))).))))).((((((.....))))))..
Leu-TAG-1-1	-56.2	-56.48	0.073	130	((((((..(((((..((((....((((((((((........)))))((((....))))))))...))))...)))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Lys-TTT-1-1	-48.7	-43.88	0.015	121	((((((..(.((((((((((((((....).)))).))))))))).)(((((.......))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Lys-TTT-1-2	-48.7	-43.88	0.015	121	((((((..(.((((((((((((((....).)))).))))))))).)(((((.......))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Lys-TTT-1-3	-48.7	-43.88	0.015	121	((((((..(.((((((((((((((....).)))).))))))))).)(((((.......))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Lys-TTT-1-4	-48.7	-43.88	0.015	121	((((((..(.((((((((((((((....).)))).))))))))).)(((((.......))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Lys-TTT-1-5	-48.7	-43.88	0.015	121	((((((..(.((((((((((((((....).)))).))))))))).)(((((.......))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Lys-TTT-1-6	-48.7	-43.88	0.015	121	((((((..(.((((((((((((((....).)))).))))))))).)(((((.......))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Met-CAT-1-1	-47	-44.48	0.01	122	......((.((((.........))))))....(((((....)))))...(((((((((.((((.((...((.....+)).)).)))).)))))))..)))...((((((.....))))))..
Met-CAT-1-2	-47	-44.48	0.01	122	......((.((((.........))))))....(((((....)))))...(((((((((.((((.((...((.....+)).)).)))).)))))))..)))...((((((.....))))))..
Phe-GAA-1-1	-55.2	-54.48	0.158	121	....(((((((((........)))).(((((((...))))))).)))))..((((((((((...((((.......+.))))..))))))))))........((((((.....))))))..
Phe-GAA-1-2	-55.2	-54.48	0.158	121	....(((((((((........)))).(((((((...))))))).)))))..((((((((((...((((.......+.))))..))))))))))........((((((.....))))))..
Pro-CGG-1-1	-54.1	-52.28	0.049	122	.........((((.((...)).)))).(((((((...)))))))......((.(((((((.((((((....(((....+.))))))))).)))))))))......((((((.....))))))..
Pro-GGG-1-1	-53.9	-52.08	0.115	122	((((((...((((.((...)).)))).((.....)).)))))).(((.(((.(((((.(.(((((((.(((......+))))))))))).).))))).))).)))((((((.....))))))..
Pro-TGG-1-1	-60.7	-59.38	0.087	122	.........((((.((...)).)))).(((((((...)))))))(((.(((.((((((.((((((.(((......+))))))))))).)))))).))).)))((((((.....))))))..
SeC-TCA-1-1	-67.8	-64.88	0.055	140	.((((.((((((((((....)))))))).)).)))).........(((((((....))))))).((((((((((.((((......((((((..+))))))))))).))))))..)))).((((((.....))))))..
Ser-CGA-1-1	-67	-63.88	0.128	135	........((((...))))........(((((.......))))).((((((....))))))((((((((.(.(((((.....(((((..+)))))))))).).))))))))....((((((.....))))))..
Ser-GCT-1-1	-71.7	-69.88	0.202	138	.(((...(((((....)))))....)))..((((....))))...(((((((.....)))))))..((((((((.(((((......((((..+)))).))))).)))))))))......((((((.....))))))..
Ser-GGA-1-1	-65.1	-63.38	0.15	133	.........((((.((((((.....))).))).))))........(((((....)))))((((((((((.(((((......((((..+)))).))))).))))))))))....((((((.....))))))..
Ser-GGA-1-2	-65.1	-63.38	0.15	133	.........((((.((((((.....))).))).))))........(((((....)))))((((((((((.(((((......((((..+)))).))))).))))))))))....((((((.....))))))..
Ser-TGA-1-1	-57.2	-53.28	0.027	133	.........(((...(....)..))).(((((.......)))))..((((....))))..(((.((((((.(((((.....(((((..+)))))))))).))))))..)))....((((((.....))))))..
Thr-CGT-1-1	-46.5	-44.38	0.013	121	((((((((((...((((((..((.(..((((((...))))))..).))......)))))).))).))))))).((.+..))(((((.......))))).....((((((.....))))))..
Thr-CGT-2-1	-39.5	-42.88	0.031	116	...(((((.......))))).(((((((...)))))))........(((((((.((((....((.((....+)).)).)))).)))))))........((((((.....))))))..
Thr-GGT-1-1	-51.3	-48.88	0.116	121	((((((.(.((((........)))).((((((((...)))))))).)))))).((((((.(((.(((....((....+))..)))))).))))))........((((((.....))))))..
Thr-GGT-2-1	-50.8	-49.68	0.126	121	((((((...((((((..(((.((((.(((((((...))))))).((...)).)))).)))))))))))))).((.+..))(((((.......))))).....((((((.....))))))..
Thr-TGT-1-1	-43.3	-43.48	0.027	121	...(((((.((((........)))).(((((.......))))))))))....((((((((((.((.....))....+.......))))))))))........((((((.....))))))..
Trp-CCA-1-1	-55.9	-52.98	0.063	121	..((((...))))........((((((((((.......))).))))))(((.((((((.(((((((((........+)))))))))..)))))))))......((((((.....))))))..
Tyr-GTA-1-1	-51.8	-51.18	0.036	130	...(((...))).(((((((..((((.((.(((.(((....))).)))..((((......((((((....((((((..((....+..)).))))))))))))...)))).))..))))...))))))).
Tyr-GTA-2-1	-51.8	-49.88	0.027	130	...(((...))).(((((((..((((.((.(((.(((....))).)))...(((......((((((....((((((..((....+..)).))))))))))))...)))..))..))))...))))))).
Tyr-GTA-2-2	-51.8	-49.88	0.027	130	...(((...))).(((((((..((((.((.(((.(((....))).)))...(((......((((((....((((((..((....+..)).))))))))))))...)))..))..))))...))))))).
Und-NNN-1-1	-53	-51.98	0.012	140	.((((.......))))..((((((...))))))............((..((((..(((((((((.((..(((((((.((((......(((....+)))...)))).))))))))..)).)))))))))..)))).))....
Und-NNN-2-1	-50.3	-48.58	0.019	142	.(((((.(((....)))....((....))..((((......)))).((((((.(.....).))))))..(((((.......)))))...........+)))))((((.......))))......((((((.....))))))..
Val-GAC-1-1	-53.9	-51.38	0.111	122	((.......((((.........)))).(((((.......))))).....((((((((((((....)))))))).)))).+))..(((((.......))))).....((((((.....))))))..
Val-GAC-2-1	-48.6	-46.68	0.054	122	..((((.((((((.....))))))))).(((((.......))))).........(((((((.(((.......(((....+)))....))).)))))))........((((((.....))))))..
Val-TAC-1-1	-54.7	-52.88	0.107	121	((((.....)))).(((.....)))..(((((((.....)))))))..((.((((((((((.((...........+...)).)))).))))))))).......((((((.....))))))..
Val-TAC-1-2	-54.7	-52.88	0.107	121	((((.....)))).(((.....)))..(((((((.....)))))))..((.((((((((((.((...........+...)).)))).))))))))).......((((((.....))))))..
Val-TAC-1-3	-54.7	-52.88	0.107	121	((((.....)))).(((.....)))..(((((((.....)))))))..((.((((((((((.((...........+...)).)))).))))))))).......((((((.....))))))..
Val-TAC-1-4	-54.7	-52.88	0.107	121	((((.....)))).(((.....)))..(((((((.....)))))))..((.((((((((((.((...........+...)).)))).))))))))).......((((((.....))))))..
Val-TAC-1-5	-54.7	-52.88	0.107	121	((((.....)))).(((.....)))..(((((((.....)))))))..((.((((((((((.((...........+...)).)))).))))))))).......((((((.....))))))..
fMet-CAT-1-1	-50.4	-49.58	0.025	122	.(((((((((..((((.(((.(((..(((((.......))))).))).)))))))....))))..)))))..((.+..))(((((.......))))).....((((((.....))))))..
fMet-CAT-1-2	-50.4	-49.58	0.025	122	.(((((((((..((((.(((.(((..(((((.......))))).))).)))))))....))))..)))))..((.+..))(((((.......))))).....((((((.....))))))..
fMet-CAT-1-3	-50.4	-49.58	0.025	122	.(((((((((..((((.(((.(((..(((((.......))))).))).)))))))....))))..)))))..((.+..))(((((.......))))).....((((((.....))))))..
fMet-CAT-2-1	-50.4	-48.88	0.013	122	.(((((((((..((((.....(.((.(((((.......))))).)).)...))))....))))..)))))..((.+..))(((((.......))))).....((((((.....))))))..

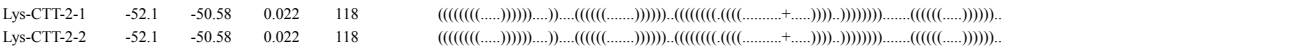
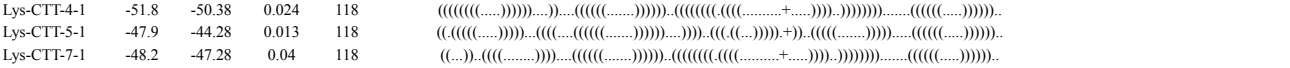
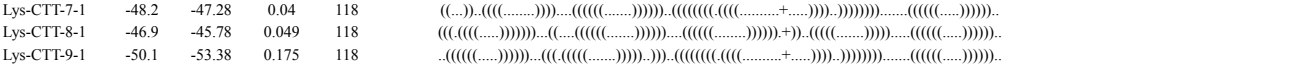
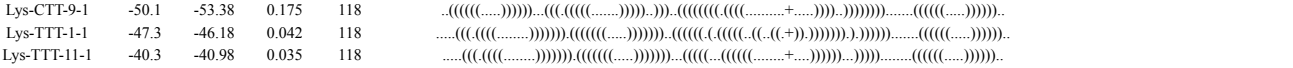
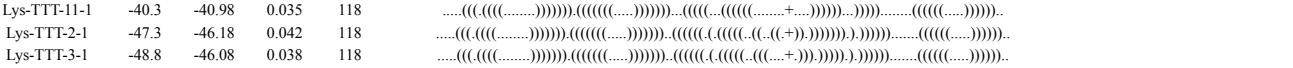
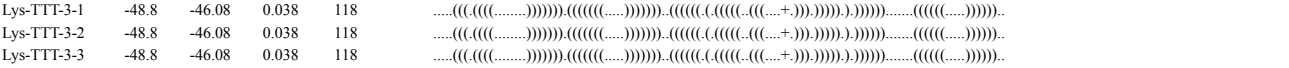
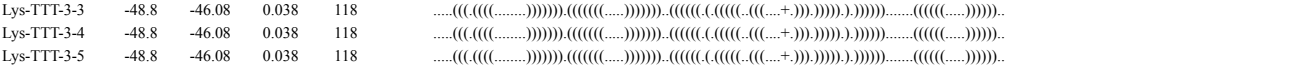
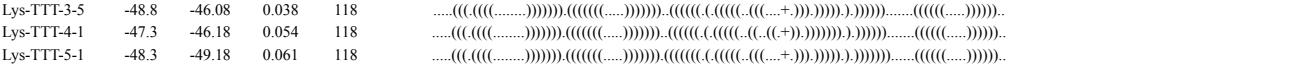
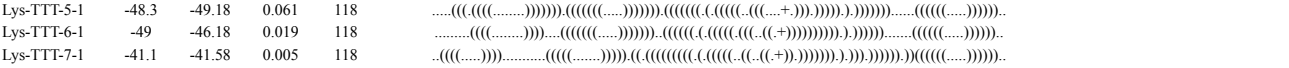
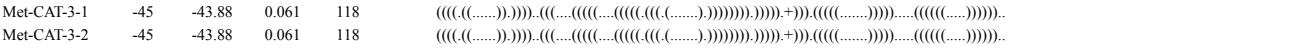
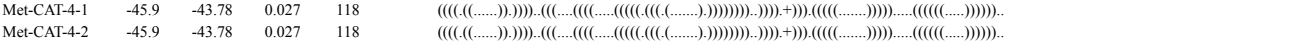
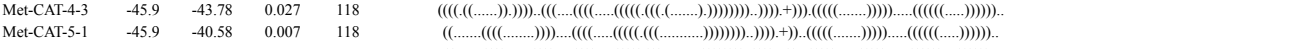
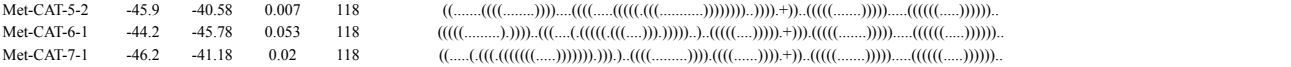
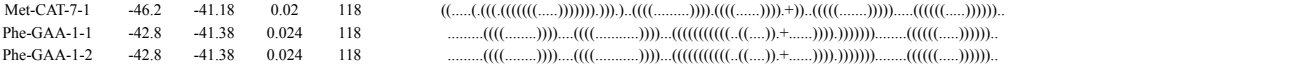
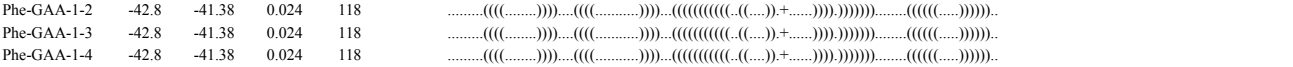
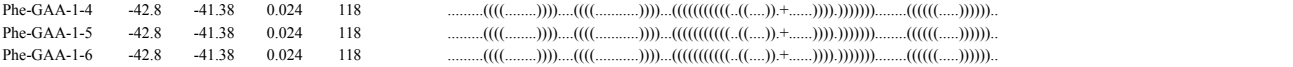
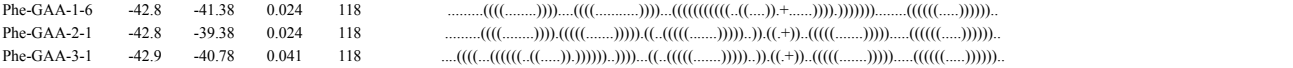
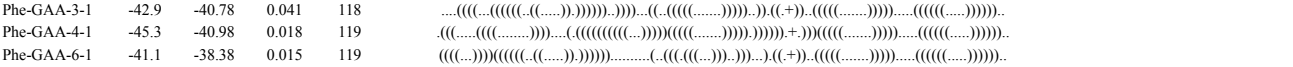
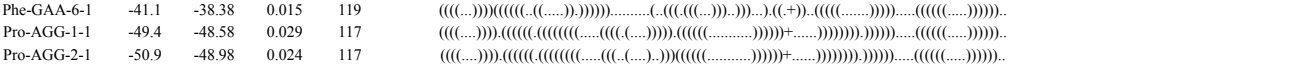
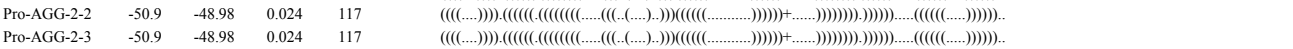
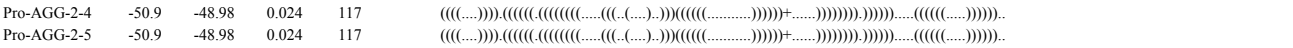
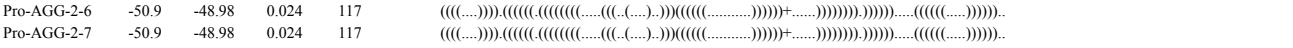
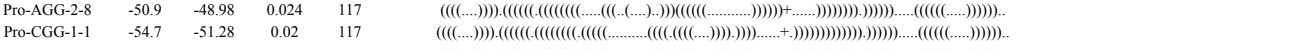
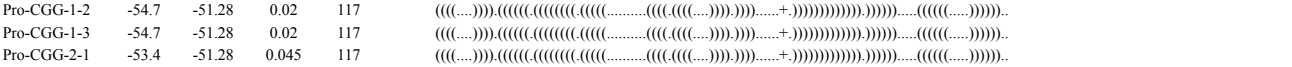
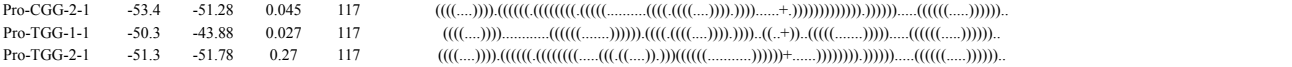
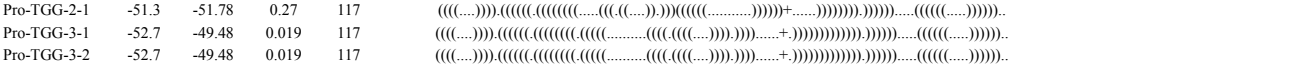
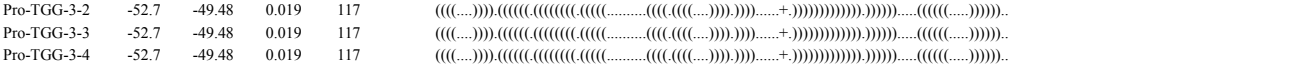
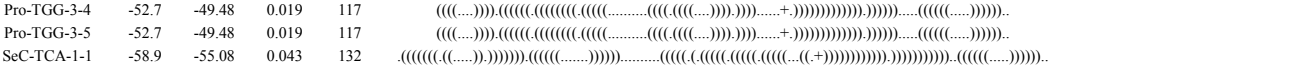
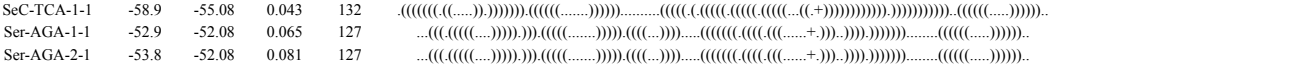
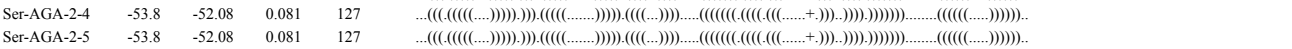
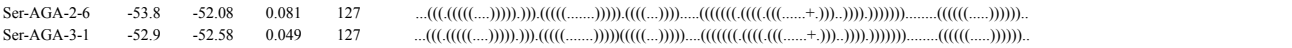
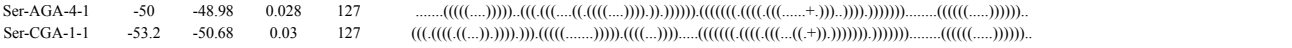
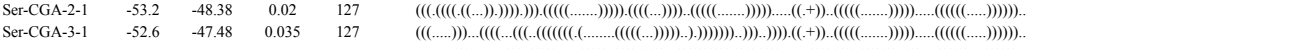
Table A.6: Screening synthetic t1A3T2 sRNA against natural occurring tRNAs in human. BE - binding energy of sRNA and ΔG - free energy of the secondary structure in kcal/mol; P - probability of formation of the secondary structure; MFE - minimum free energy structure.

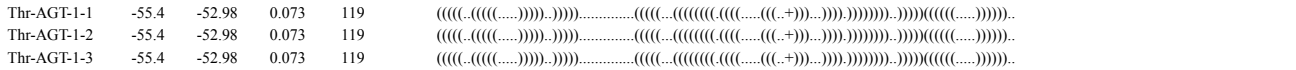
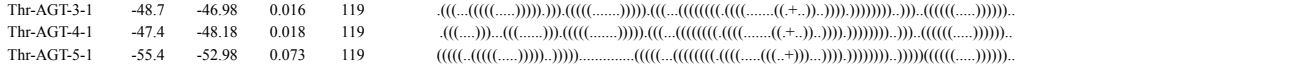
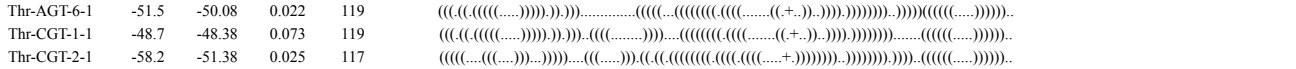
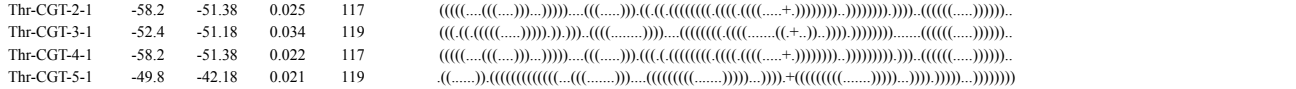
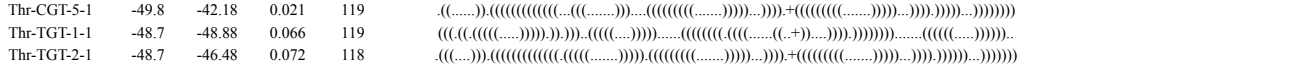
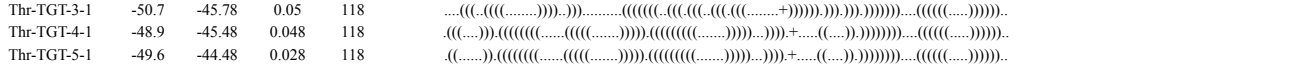
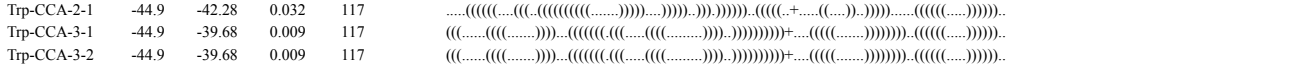
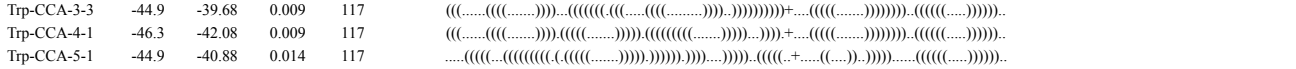
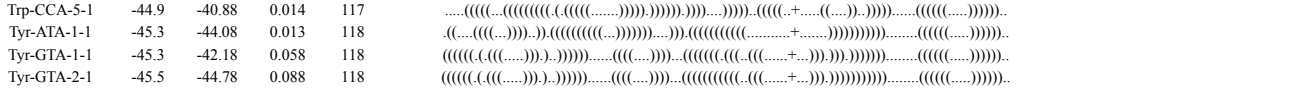
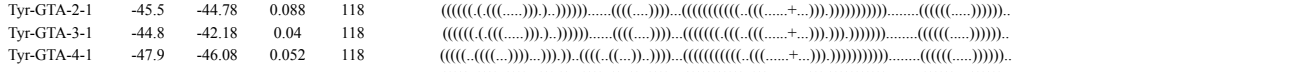
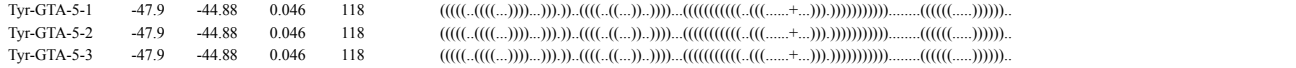
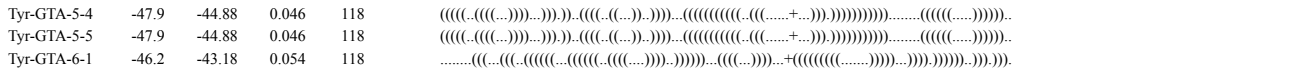
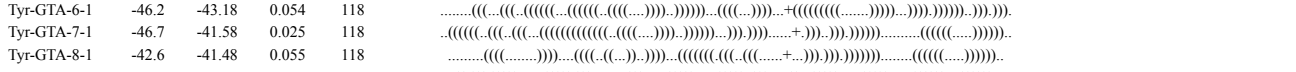
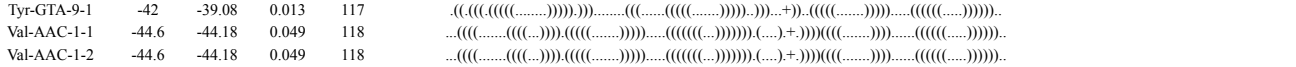
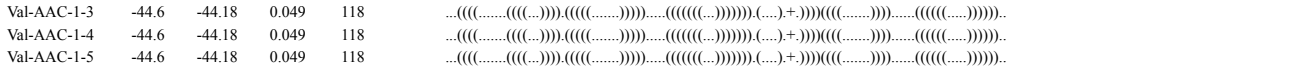
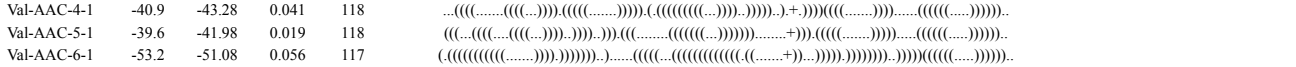
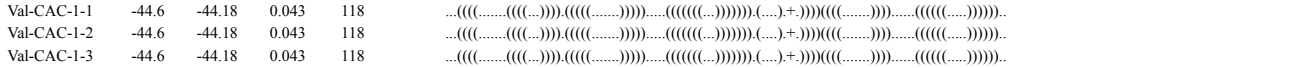
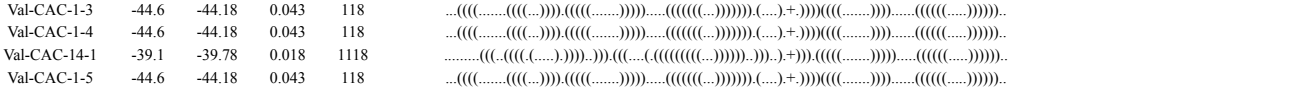
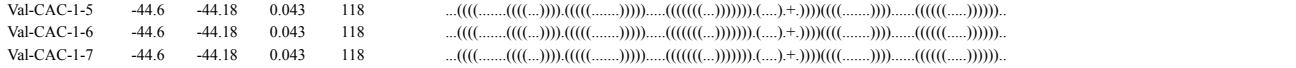
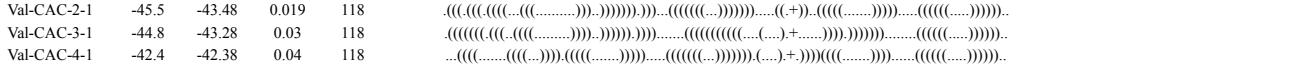
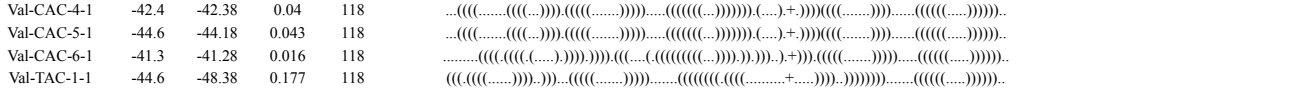
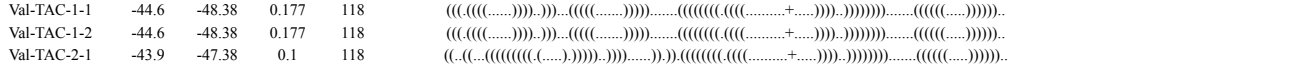
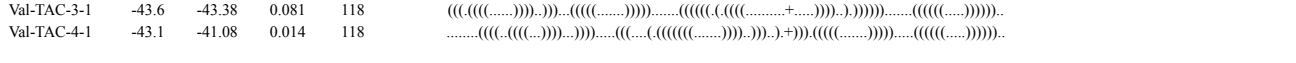
tRNA	BE	ΔG	P	(nt)	MFE
Ala-AGC-10-1	-49.2	-47.88	0.125	118	.(((......)))..(((((.((((....)))).)).)))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-10-2	-49.2	-47.88	0.125	118	.(((......)))..(((((.((((....)))).)).)))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-1-1	-53.7	-51.68	0.203	117	.(((((...((((.......))))((((((...))))))..((((.(((((.......))))).)))))..+.....((....))..)))))......((((((.....))))))..
Ala-AGC-11-1	-49.6	-46.78	0.029	118	.((((....((((........))))..((((.......))))((((..(((((.......)))))..))))..+....(((((.......))))).))))((((((.....))))))..
Ala-AGC-12-1	-49	-46.18	0.042	118	.........((((........))))((.(((...))).))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-12-2	-49	-46.18	0.042	118	.........((((........))))((.(((...))).))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-12-3	-49	-46.18	0.042	118	.........((((........))))((.(((...))).))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-13-1	-50.5	-48.18	0.062	118	.(((......)))..(((((.((((....)))).)).)))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-13-2	-50.5	-48.18	0.062	118	.(((......)))..(((((.((((....)))).)).)))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-13-3	-50.5	-48.18	0.062	118	.(((......)))..(((((.((((....)))).)).)))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-14-1	-45.6	-45.38	0.082	118	.(((......))).((((((((((((.(((...))).))..((((..(((((.......)))))..))))..+((((((((.......)))))...)))).)))))....)))))).
Ala-AGC-14-2	-45.6	-45.38	0.082	118	.(((......))).((((((((((((.(((...))).))..((((..(((((.......)))))..))))..+((((((((.......)))))...)))).)))))....)))))).
Ala-AGC-15-1	-49.9	-46.48	0.03	118	.(((......))).((((((((((((((.....)))..((((((..(((.(((.(((((.(((.....((((.+..))))))).))))).))))))))))).)))))....)))))).
Ala-AGC-16-1	-51.1	-48.38	0.06	118	.(((......)))..(((((.((((....)))).)).)))..((((((.(((.(((((.(((.....((((.+..))))))).))))).)))))))))((((((.....))))))..
Ala-AGC-2-1	-54.6	-49.88	0.044	117	.(((((...((((.......))))((((((...)))))).(((..(((((((.......)))))..(((((.+..)))))))..))).)))))......((((((.....))))))..
Ala-AGC-2-2	-54.6	-49.88	0.044	117	.(((((...((((.......))))((((((...)))))).(((..(((((((.......)))))..(((((.+..)))))))..))).)))))......((((((.....))))))..
Ala-AGC-24-1	-46.4	-47.58	0.16	118	.(((......))).((((((((((((.(((...))).))..((((..(((((.......)))))..))))..+((((((((.......)))))...)))).)))))....)))))).
Ala-AGC-3-1	-51.7	-50.28	0.322	117	.(((((...((((.......)))).(((((...)))))...(((((.(((((.......))))).)))))..+.....((....))..)))))......((((((.....))))))..
Ala-AGC-4-1	-56.5	-52.38	0.046	117	.(((((.(.((((.......))))((((((...))))))...).)))))((((((((((((.((.......+))...))))).))))))))).......((((((.....))))))..
Ala-AGC-5-1	-55.4	-53.98	0.188	117	((((((..(((((((.(((...(.((((((...)))))).)...)))))))))))..)))))).((.......+))..(((((.......))))).....((((((.....))))))..
Ala-AGC-6-1	-51	-48.38	0.076	117	..((((...)))).((((((((((((((((...))))))..(((((..((((.......))))..)))))..+((((((((.......)))))...)))).)))))....)))))..
Ala-AGC-7-1	-54.6	-49.88	0.044	117	.(((((...((((.......))))((((((...)))))).(((..(((((((.......)))))..(((((.+..)))))))..))).)))))......((((((.....))))))..
Ala-AGC-8-1	-51.3	-48.38	0.035	118	..(((.....))).....((.((((....)))).))..((..((((..(((((.......)))))..))))..+))..(((((.......))))).....((((((.....))))))..
Ala-AGC-8-2	-51.3	-48.38	0.035	118	..(((.....))).....((.((((....)))).))..((..((((..(((((.......)))))..))))..+))..(((((.......))))).....((((((.....))))))..
Ala-AGC-9-1	-50.8	-48.18	0.047	118	.........(((.(((((((.....)).))))))))..((..((((..(((((.......)))))..))))..+))..(((((.......))))).....((((((.....))))))..
Ala-AGC-9-2	-50.8	-48.18	0.047	118	.........(((.(((((((.....)).))))))))..((..((((..(((((.......)))))..))))..+))..(((((.......))))).....((((((.....))))))..
Ala-CGC-1-1	-54.4	-50.28	0.062	117	.(((((....((((...((.((((....)))).))......)))))))))((((((((((((.((.......+))...))))).))))))))).......((((((.....))))))..

Ala-CGC-2-1	-54.2	-50.68	0.026	117	
Ala-CGC-3-1	-49.1	-46.48	0.029	117	
Ala-CGC-4-1	-55.7	-52.28	0.11	117	
Ala-TGC-1-1	-53.7	-50.78	0.034	117	
Ala-TGC-2-1	-55.7	-52.38	0.09	117	
Ala-TGC-3-1	-54.3	-51.78	0.046	117	
Ala-TGC-3-2	-54.3	-51.78	0.046	117	
Ala-TGC-4-1	-49.6	-47.58	0.024	117	
Ala-TGC-5-1	-55.1	-51.78	0.049	117	
Ala-TGC-6-1	-58.1	-54.38	0.052	117	
Ala-TGC-7-1	-51.8	-48.78	0.046	116	
Arg-ACG-1-1	-46.4	-46.08	0.076	118	
Arg-ACG-1-2	-46.4	-46.08	0.076	118	
Arg-ACG-1-3	-46.4	-46.08	0.076	118	
Arg-ACG-2-1	-45	-44.18	0.031	118	
Arg-ACG-2-2	-45	-44.18	0.031	118	
Arg-ACG-2-3	-45	-44.18	0.031	118	
Arg-ACG-2-4	-45	-44.18	0.031	118	
Arg-CCG-1-1	-56.2	-56.18	0.057	118	
Arg-CCG-1-2	-56.2	-56.18	0.057	118	
Arg-CCG-1-3	-56.2	-56.18	0.057	118	
Arg-CCG-2-1	-67	-60.58	0.146	118	
Arg-CCT-1-1	-59.6	-53.88	0.03	118	
Arg-CCT-2-1	-59.6	-53.88	0.03	118	
Arg-CCT-3-1	-61.1	-55.28	0.053	118	
Arg-CCT-4-1	-59.8	-56.08	0.066	118	
Arg-CCT-5-1	-55.1	-50.18	0.018	118	
Arg-TCG-1-1	-49.5	-52.48	0.053	118	
Arg-TCG-2-1	-54.5	-53.28	0.044	118	
Arg-TCG-3-1	-53.1	-52.98	0.039	118	
Arg-TCG-4-1	-51.6	-52.98	0.089	118	
Arg-TCG-5-1	-51.6	-52.98	0.089	118	
Arg-TCG-6-1	-43.8	-41.28	0.025	118	
Arg-TCT-1-1	-47.2	-48.38	0.043	118	
Arg-TCT-2-1	-44.4	-47.58	0.05	118	
Arg-TCT-3-1	-44.7	-47.58	0.03	118	
Arg-TCT-3-2	-44.7	-47.58	0.03	118	
Arg-TCT-4-1	-50.6	-50.68	0.062	119	
Arg-TCT-5-1	-48.2	-52.48	0.078	118	
Asn-GTT-10-1	-48.6	-44.28	0.019	119	
Asn-GTT-1-1	-53.9	-50.68	0.046	119	
Asn-GTT-11-1	-53.2	-50.58	0.078	119	
Asn-GTT-11-2	-53.2	-50.58	0.078	119	
Asn-GTT-12-1	-43.5	-40.68	0.007	119	
Asn-GTT-2-1	-52.3	-50.88	0.097	119	
Asn-GTT-2-2	-52.3	-50.88	0.097	119	
Asn-GTT-2-3	-52.3	-50.88	0.097	119	
Asn-GTT-2-4	-52.3	-50.88	0.097	119	
Asn-GTT-24-1	-51.7	-47.18	0.15	119	
Asn-GTT-2-5	-52.3	-50.88	0.097	119	
Asn-GTT-25-1	-46.8	-46.88	0.046	120	
Asn-GTT-2-6	-52.3	-50.88	0.097	119	
Asn-GTT-26-1	-49	-48.28	0.056	119	
Asn-GTT-2-7	-52.3	-50.88	0.097	119	
Asn-GTT-27-1	-48	-43.18	0.031	119	
Asn-GTT-2-8	-52.3	-50.88	0.097	119	
Asn-GTT-3-1	-51.3	-51.58	0.118	119	
Asn-GTT-4-1	-55	-53.58	0.213	119	
Asn-GTT-5-1	-50.1	-49.08	0.096	119	
Asn-GTT-6-1	-49.1	-49.68	0.054	119	
Asn-GTT-7-1	-52.3	-50.78	0.118	119	
Asn-GTT-8-1	-43.3	-42.78	0.029	119	
Asn-GTT-9-1	-52.4	-50.58	0.109	119	
Asn-GTT-9-2	-50.6	-49.28	0.091	119	
Asp-GTC-1-1	-57.8	-53.48	0.068	117	
Asp-GTC-2-1	-65.1	-60.98	0.089	117	
Asp-GTC-2-10	-65.1	-60.98	0.089	117	
Asp-GTC-2-11	-65.1	-60.98	0.089	117	
Asp-GTC-2-2	-65.1	-60.98	0.089	117	
Asp-GTC-2-3	-65.1	-60.98	0.089	117	
Asp-GTC-2-4	-65.1	-60.98	0.089	117	
Asp-GTC-2-5	-65.1	-60.98	0.089	117	
Asp-GTC-2-6	-65.1	-60.98	0.089	117	
Asp-GTC-2-7	-65.1	-60.98	0.089	117	
Asp-GTC-2-8	-65.1	-60.98	0.089	117	
Asp-GTC-2-9	-65.1	-60.98	0.089	117	
Asp-GTC-3-1	-65.1	-59.88	0.066	117	
Cys-GCA-10-1	-46.1	-42.48	0.024	117	
Cys-GCA-1-1	-48	-42.68	0.031	117	
Cys-GCA-11-1	-47.5	-43.48	0.03	117	
Cys-GCA-12-1	-44.7	-42.48	0.076	117	
Cys-GCA-13-1	-47.7	-42.48	0.015	117	

Cys-GCA-14-1	-47.7	-41.18	0.013	117	
Cys-GCA-15-1	-46.5	-42.78	0.021	117	
Cys-GCA-16-1	-47.1	-44.48	0.049	117	
Cys-GCA-17-1	-48.5	-43.68	0.03	117	
Cys-GCA-18-1	-45.2	-43.08	0.058	117	
Cys-GCA-19-1	-42.6	-40.18	0.021	117	
Cys-GCA-20-1	-47.1	-45.38	0.052	117	
Cys-GCA-2-1	-47.5	-43.88	0.053	117	
Cys-GCA-21-1	-47	-42.08	0.058	117	
Cys-GCA-2-2	-47.5	-43.88	0.053	117	
Cys-GCA-22-1	-45.7	-45.38	0.065	115	
Cys-GCA-2-3	-47.5	-43.88	0.053	117	
Cys-GCA-23-1	-45	-40.58	0.044	117	
Cys-GCA-2-4	-47.5	-43.88	0.053	117	
Cys-GCA-3-1	-46.5	-44.08	0.044	117	
Cys-GCA-4-1	-45.3	-42.08	0.047	117	
Cys-GCA-5-1	-48	-43.78	0.056	118	
Cys-GCA-6-1	-47.4	-42.48	0.035	117	
Cys-GCA-7-1	-50.1	-43.78	0.04	118	
Cys-GCA-8-1	-47.5	-43.88	0.023	117	
Cys-GCA-9-1	-46	-42.48	0.023	117	
Cys-GCA-9-2	-46	-42.48	0.023	117	
Cys-GCA-9-3	-46	-42.48	0.023	117	
Cys-GCA-9-4	-46	-42.48	0.023	117	
Gln-CTG-1-1	-43.9	-41.98	0.044	117	
Gln-CTG-1-2	-43.9	-41.98	0.044	117	
Gln-CTG-1-3	-43.9	-41.98	0.044	117	
Gln-CTG-1-4	-43.9	-41.98	0.044	117	
Gln-CTG-1-5	-43.9	-41.98	0.044	117	
Gln-CTG-2-1	-43.9	-41.98	0.044	117	
Gln-CTG-3-1	-46.5	-42.28	0.041	117	
Gln-CTG-3-2	-46.5	-42.28	0.041	117	
Gln-CTG-4-1	-44.6	-42.28	0.046	117	
Gln-CTG-4-2	-44.6	-42.28	0.046	117	
Gln-CTG-5-1	-43.4	-42.28	0.039	117	
Gln-CTG-6-1	-41.4	-41.08	0.06	117	
Gln-CTG-7-1	-46	-40.68	0.04	117	
Gln-TTG-1-1	-46.3	-39.98	0.009	117	
Gln-TTG-2-1	-46.3	-40.18	0.01	117	
Gln-TTG-3-1	-41.4	-41.08	0.079	117	
Gln-TTG-3-2	-41.4	-41.08	0.079	117	
Gln-TTG-3-3	-41.4	-41.08	0.079	117	
Gln-TTG-4-1	-43.3	-40.48	0.023	117	
Glu-CTC-1-1	-54.8	-51.58	0.019	117	
Glu-CTC-1-2	-54.8	-51.58	0.019	117	
Glu-CTC-1-3	-54.8	-51.58	0.019	117	
Glu-CTC-1-4	-54.8	-51.58	0.019	117	
Glu-CTC-1-5	-54.8	-51.58	0.019	117	
Glu-CTC-1-6	-54.8	-51.58	0.019	117	
Glu-CTC-1-7	-54.8	-51.58	0.019	117	
Glu-CTC-2-1	-54.8	-51.58	0.02	117	
Glu-TTC-1-1	-44.7	-45.68	0.084	117	
Glu-TTC-1-2	-44.7	-45.68	0.084	117	
Glu-TTC-14-1	-58.9	-56.08	0.079	117	
Glu-TTC-2-1	-45.9	-44.78	0.027	117	
Glu-TTC-2-2	-45.9	-44.78	0.027	117	
Glu-TTC-3-1	-64.3	-59.38	0.105	117	
Glu-TTC-4-1	-61.1	-57.88	0.08	117	
Glu-TTC-4-2	-61.1	-57.88	0.08	117	
Gly-CCC-1-1	-50.4	-49.28	0.11	116	
Gly-CCC-1-2	-50.4	-49.28	0.11	116	
Gly-CCC-2-1	-58.9	-53.48	0.035	116	
Gly-CCC-2-2	-58.9	-53.48	0.035	116	
Gly-CCC-3-1	-49.8	-47.08	0.031	116	
Gly-GCC-1-1	-59.4	-54.68	0.073	116	
Gly-GCC-1-2	-59.4	-54.68	0.073	116	
Gly-GCC-1-3	-59.4	-54.68	0.073	116	
Gly-GCC-1-4	-59.4	-54.68	0.073	116	
Gly-GCC-1-5	-59.4	-54.68	0.073	116	
Gly-GCC-2-1	-58.7	-54.18	0.051	116	
Gly-GCC-2-2	-58.7	-54.18	0.051	116	
Gly-GCC-2-3	-58.7	-54.18	0.051	116	
Gly-GCC-2-4	-58.7	-54.18	0.051	116	
Gly-GCC-2-5	-58.7	-54.18	0.051	116	
Gly-GCC-2-6	-58.7	-54.18	0.051	116	
Gly-GCC-3-1	-56.6	-52.18	0.06	116	
Gly-GCC-4-1	-50	-46.58	0.021	116	
Gly-GCC-5-1	-54.4	-52.58	0.112	116	
Gly-TCC-1-1	-55.5	-53.78	0.038	117	
Gly-TCC-2-1	-55.1	-53.78	0.043	117	
Gly-TCC-2-2	-55.1	-53.78	0.043	117	
Gly-TCC-2-3	-55.1	-53.78	0.043	117	

Gly-TCC-2-4	-55.1	-53.78	0.043	117	((.((((.(((..(((..........))))))..))))))....(((..((((((((((((......(((.+)))...))))))))))))..)))..((((((.....))))))..
Gly-TCC-2-5	-55.1	-53.78	0.043	117	((.((((.(((..(((..........))))))..))))))....(((..((((((((((((......(((.+)))...))))))))))))..)))..((((((.....))))))..
Gly-TCC-2-6	-55.1	-53.78	0.043	117	((.((((.(((..(((..........))))))..))))))....(((..((((((((((((......(((.+)))...))))))))))))..)))..((((((.....))))))..
Gly-TCC-3-1	-55.4	-53.78	0.045	117	((.((((.(((..(((..........))))))..))))))....(((..((((((((((((......(((.+)))...))))))))))))..)))..((((((.....))))))..
Gly-TCC-4-1	-51.3	-50.58	0.108	117	((.((((.((((..(((.....)))...))))..))))))....(((..(((.((((((((......(((.+)))...)))))))).)))..)))..((((((.....))))))..
His-GTG-1-1	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-2	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-3	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-4	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-5	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-6	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-7	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-8	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
His-GTG-1-9	-46.1	-44.18	0.094	117	((((.((.((((.((((.....)))).)))))).))))(((.((...(((((.......)))))......)).+))).(((((.......)))))......((((((.....))))))..
Ile-AAT-1-1	-53.6	-52.38	0.08	119	((((.....((((.........)))).(((((.......))))).))))..((((((((((((...((.....+.))..))))).))))))))).......((((((.....))))))..
Ile-AAT-12-1	-53.5	-51.18	0.008	119	(((..(((((((((.((.....))...))).))))))...))).(((.((.((((((((((((.(((......+)))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-2-1	-52.7	-49.48	0.009	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...(((....+.))).))))).))))))))).))))).((((((.....))))))..
Ile-AAT-3-1	-52.7	-49.48	0.007	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...(((....+.))).))))).))))))))).))))).((((((.....))))))..
Ile-AAT-4-1	-50.9	-48.48	0.042	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((.((((...((.....+.))..))))..))))))))).))))).((((((.....))))))..
Ile-AAT-5-1	-53.6	-50.68	0.022	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...((.....+.))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-5-2	-53.6	-50.68	0.022	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...((.....+.))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-5-3	-53.6	-50.68	0.022	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...((.....+.))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-5-4	-53.6	-50.68	0.022	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...((.....+.))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-5-5	-53.6	-50.68	0.022	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...((.....+.))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-6-1	-50	-49.68	0.026	119	((((..(((((...((((.....(((.....)))..))))))))))))..((((((((((((...(((....+.))).))))).))))))))).......((((((.....))))))..
Ile-AAT-7-1	-53.6	-50.68	0.014	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...((.....+.))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-7-2	-53.6	-50.68	0.014	119	(((..((((((((((((......))).))).))))))...))).(((.((.((((((((((((...((.....+.))..))))).))))))))).))))).((((((.....))))))..
Ile-AAT-8-1	-52.7	-51.68	0.016	119	((((.(((((((.....)))))..)).(((((.......))))).))))..((((((((((((...((.....+.))..))))).))))))))).......((((((.....))))))..
Ile-AAT-9-1	-51.9	-47.58	0.005	119	(((..((((((...(((......)))......))))))...))).(((.((.((((((((((((...(((....+.))).))))).))))))))).))))).((((((.....))))))..
Ile-GAT-1-1	-51.1	-47.68	0.151	119	((((.....((((.........)))).(((((.......))))).))))..(((.((((.((((.(.(((....+.))).))))).)))).))).......((((((.....))))))..
Ile-GAT-1-2	-51.1	-47.68	0.151	119	((((.....((((.........)))).(((((.......))))).))))..(((.((((.((((.(.(((....+.))).))))).)))).))).......((((((.....))))))..
Ile-GAT-1-3	-51.1	-47.68	0.151	119	((((.....((((.........)))).(((((.......))))).))))..(((.((((.((((.(.(((....+.))).))))).)))).))).......((((((.....))))))..
Ile-TAT-1-1	-47.3	-40.98	0.004	119	.........((((.........)))).(((((.......))))).........((((((((((.((.......+..)).))))).))))))..........((((((.....))))))..
Ile-TAT-2-1	-48.1	-43.48	0.028	119	(((.....)))......(((.((.((.(((.(((((((((......)))))))))))).)))).)))....((.+))..(((((.......))))).....((((((.....))))))..
Ile-TAT-2-2	-48.1	-43.48	0.028	119	(((.....)))......(((.((.((.(((.(((((((((......)))))))))))).)))).)))....((.+))..(((((.......))))).....((((((.....))))))..
Ile-TAT-2-3	-48.1	-43.48	0.028	119	(((.....)))......(((.((.((.(((.(((((((((......)))))))))))).)))).)))....((.+))..(((((.......))))).....((((((.....))))))..
Ile-TAT-3-1	-47.5	-42.98	0.024	119	..(((((..((((.........))))..((((.......))))((((((.........))))))..)))))((.+))..(((((.......))))).....((((((.....))))))..
iMet-CAT-1-1	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-1-2	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-1-3	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-1-4	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-1-5	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-1-6	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-1-7	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-1-8	-49.7	-46.98	0.054	117	.((......))...((....))...(((((.......))))).....(((((.((((.....((((((((..+))))))))...)))).))...)))..((((((.....))))))..
iMet-CAT-2-1	-49.5	-44.98	0.032	117	.((......))...((....))...(((((.......))))).....(((((.......))))).(((((..+)))))((((.......))))......((((((.....))))))..
Leu-AAG-1-1	-66.1	-63.08	0.119	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-AAG-1-2	-66.1	-63.08	0.119	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-AAG-1-3	-66.1	-63.08	0.119	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-AAG-2-1	-65.2	-62.38	0.121	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-AAG-2-2	-65.2	-62.38	0.121	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-AAG-2-3	-65.2	-62.38	0.121	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-AAG-2-4	-65.2	-62.38	0.121	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-AAG-3-1	-59.6	-57.28	0.108	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((.(.((((.((((.....+))))..)))).).)))))).))))))((((((.....))))))..
Leu-AAG-4-1	-63	-61.28	0.186	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-CAA-1-1	-60.2	-57.98	0.039	128	........((((....))))...((.((((((.......))))))..))....((((((.((((((((.((((.((((.....+..)))))))).)))))))).))))))((((((.....))))))..
Leu-CAA-1-2	-60.2	-57.98	0.039	128	........((((....))))...((.((((((.......))))))..))....((((((.((((((((.((((.((((.....+..)))))))).)))))))).))))))((((((.....))))))..
Leu-CAA-2-1	-60.1	-57.98	0.035	129	........((((....))))...((.((((((.......))))))..)).....((((((.((((((((.((((.((((.....+..)))))))).)))))))).))))))((((((.....))))))..
Leu-CAA-3-1	-60.1	-57.98	0.022	129	........((((....))))...((.((((((.......))))))..)).....((((((.((((((((.((((.((((.....+..)))))))).)))))))).))))))((((((.....))))))..
Leu-CAA-4-1	-62.2	-58.68	0.033	129	(((.((..((((....)))).......)).)))......(((((...)))))..((((((.((((((((.((((.((((.....+..)))))))).)))))))).))))))((((((.....))))))..
Leu-CAA-5-1	-43.2	-40.68	0.016	119	((.......((((.........))))....(.((((..(((.....)))..))))..)..(((((....))))).+))..(((((.......))))).....((((((.....))))))..
Leu-CAA-6-1	-56.1	-52.68	0.005	129	....(((.((.((((((((.........)))))........)))..)).)))....((((((.((((((((.((((..........+......)))).)))))))).))))))((((((.....))))))..
Leu-CAG-1-1	-69	-63.98	0.138	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-1-2	-69	-63.98	0.138	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-1-3	-69	-63.98	0.138	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-1-4	-69	-63.98	0.138	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-1-5	-69	-63.98	0.138	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-1-6	-69	-63.98	0.138	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-1-7	-69	-63.98	0.138	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-2-1	-66.1	-61.28	0.052	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-CAG-2-2	-66.1	-61.28	0.052	128	((((...))))..((((((((..((.((((((..((((.(((.((((((...))))))))))((((((.((((.((((.....+..)))))))).))))))))))))...)))))).)).))))))))).
Leu-TAA-1-1	-55.1	-52.18	0.037	128	.........(((..........))).(((((.......))))).((((....)))).((.((((((((..(((.((((.....+..)))))))..)))))))).))....((((((.....))))))..
Leu-TAA-2-1	-57.3	-52.08	0.026	128	..(((.....))).((((.((.((((((((((.......))))))..)))..)).))))..((((((((..(((.(((.((...+.))))))))..))))))))).......((((((.....))))))..
Leu-TAA-3-1	-50.5	-48.18	0.014	128	.(((...)))....((((.((.((((((((((.......)))))))...)).)).))))..((((((((.((((...((((...+.))))))))..)))))))).......((((((.....))))))..
Leu-TAA-4-1	-55.4	-51.88	0.009	128	.........(((..........))).(((((.......))))).((((....)))).((.((((((((..(((.((((.....+..)))))))..)))))))).))....((((((.....))))))..
Leu-TAG-1-1	-66.1	-63.08	0.118	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((.((((.....+))))..)))).)))))))).))))))((((((.....))))))..
Leu-TAG-2-1	-61.3	-59.28	0.104	127	........((((....)))).((((.((((((.......)))))).))))..((((((.((((((((.((((....((((..+))))..)))).)))))))).))))))((((((.....))))))..
Leu-TAG-3-1	-59.9	-59.18	0.117	127	(((......)))......((.(((..((((((.......))))))..))).)).((((.((((((((.((((.((((.....+))))..)))).)))))))).))))..((((((.....))))))..
Lys-CTT-10-1	-42.5	-40.28	0.025	118	((...))..((((........))))..(.((((((.......)))))).)((.((((..((((.(((......+)))..))))...)))).))......((((((.....))))))..
Lys-CTT-1-1	-54.2	-51.88	0.021	118	..((((((.....))))))..((.(.(((((.......))))).).))..((((((((.((((..........+.....))))..))))))))).......((((((.....))))))..
Lys-CTT-11-1	-45.4	-47.38	0.045	119	.......((((........))))....((((((.......))))))..(((((((..((((....(((...)))+.....))))...))))))).......((((((.....))))))..
Lys-CTT-1-2	-54.2	-51.88	0.021	118	..((((((.....))))))..((.(.(((((.......))))).).))..((((((((.((((..........+.....))))..))))))))).......((((((.....))))))..

Lys-CTT-2-1	-52.1	-50.58	0.022	118	
Lys-CTT-2-2	-52.1	-50.58	0.022	118	
Lys-CTT-2-3	-52.1	-50.58	0.022	118	
Lys-CTT-2-4	-52.1	-50.58	0.022	118	
Lys-CTT-2-5	-52.1	-50.58	0.022	118	
Lys-CTT-3-1	-53.4	-53.08	0.04	118	
Lys-CTT-4-1	-51.8	-50.38	0.024	118	
Lys-CTT-5-1	-47.9	-44.28	0.013	118	
Lys-CTT-7-1	-48.2	-47.28	0.04	118	
Lys-CTT-8-1	-46.9	-45.78	0.049	118	
Lys-CTT-9-1	-50.1	-53.38	0.175	118	
Lys-TTT-1-1	-47.3	-46.18	0.042	118	
Lys-TTT-11-1	-40.3	-40.98	0.035	118	
Lys-TTT-2-1	-47.3	-46.18	0.042	118	
Lys-TTT-3-1	-48.8	-46.08	0.038	118	
Lys-TTT-3-2	-48.8	-46.08	0.038	118	
Lys-TTT-3-3	-48.8	-46.08	0.038	118	
Lys-TTT-3-4	-48.8	-46.08	0.038	118	
Lys-TTT-3-5	-48.8	-46.08	0.038	118	
Lys-TTT-4-1	-47.3	-46.18	0.054	118	
Lys-TTT-5-1	-48.3	-49.18	0.061	118	
Lys-TTT-6-1	-49	-46.18	0.019	118	
Lys-TTT-7-1	-41.1	-41.58	0.005	118	
Met-CAT-1-1	-44.2	-40.68	0.033	118	
Met-CAT-2-1	-45.9	-43.78	0.027	118	
Met-CAT-3-1	-45	-43.88	0.061	118	
Met-CAT-3-2	-45	-43.88	0.061	118	
Met-CAT-4-1	-45.9	-43.78	0.027	118	
Met-CAT-4-2	-45.9	-43.78	0.027	118	
Met-CAT-4-3	-45.9	-43.78	0.027	118	
Met-CAT-5-1	-45.9	-40.58	0.007	118	
Met-CAT-5-2	-45.9	-40.58	0.007	118	
Met-CAT-6-1	-44.2	-45.78	0.053	118	
Met-CAT-7-1	-46.2	-41.18	0.02	118	
Phe-GAA-1-1	-42.8	-41.38	0.024	118	
Phe-GAA-1-2	-42.8	-41.38	0.024	118	
Phe-GAA-1-3	-42.8	-41.38	0.024	118	
Phe-GAA-1-4	-42.8	-41.38	0.024	118	
Phe-GAA-1-5	-42.8	-41.38	0.024	118	
Phe-GAA-1-6	-42.8	-41.38	0.024	118	
Phe-GAA-2-1	-42.8	-39.38	0.024	118	
Phe-GAA-3-1	-42.9	-40.78	0.041	118	
Phe-GAA-4-1	-45.3	-40.98	0.018	119	
Phe-GAA-6-1	-41.1	-38.38	0.015	119	
Pro-AGG-1-1	-49.4	-48.58	0.029	117	
Pro-AGG-2-1	-50.9	-48.98	0.024	117	
Pro-AGG-2-2	-50.9	-48.98	0.024	117	
Pro-AGG-2-3	-50.9	-48.98	0.024	117	
Pro-AGG-2-4	-50.9	-48.98	0.024	117	
Pro-AGG-2-5	-50.9	-48.98	0.024	117	
Pro-AGG-2-6	-50.9	-48.98	0.024	117	
Pro-AGG-2-7	-50.9	-48.98	0.024	117	
Pro-AGG-2-8	-50.9	-48.98	0.024	117	
Pro-CGG-1-1	-54.7	-51.28	0.02	117	
Pro-CGG-1-2	-54.7	-51.28	0.02	117	
Pro-CGG-1-3	-54.7	-51.28	0.02	117	
Pro-CGG-2-1	-53.4	-51.28	0.045	117	
Pro-TGG-1-1	-50.3	-43.88	0.027	117	
Pro-TGG-2-1	-51.3	-51.78	0.27	117	
Pro-TGG-3-1	-52.7	-49.48	0.019	117	
Pro-TGG-3-2	-52.7	-49.48	0.019	117	
Pro-TGG-3-3	-52.7	-49.48	0.019	117	
Pro-TGG-3-4	-52.7	-49.48	0.019	117	
Pro-TGG-3-5	-52.7	-49.48	0.019	117	
SeC-TCA-1-1	-58.9	-55.08	0.043	132	
Ser-AGA-1-1	-52.9	-52.08	0.065	127	
Ser-AGA-2-1	-53.8	-52.08	0.081	127	
Ser-AGA-2-2	-53.8	-52.08	0.081	127	
Ser-AGA-2-3	-53.8	-52.08	0.081	127	
Ser-AGA-2-4	-53.8	-52.08	0.081	127	
Ser-AGA-2-5	-53.8	-52.08	0.081	127	
Ser-AGA-2-6	-53.8	-52.08	0.081	127	
Ser-AGA-3-1	-52.9	-52.58	0.049	127	
Ser-AGA-4-1	-50	-48.98	0.028	127	
Ser-CGA-1-1	-53.2	-50.68	0.03	127	
Ser-CGA-2-1	-53.2	-48.38	0.02	127	
Ser-CGA-3-1	-52.6	-47.48	0.035	127	
Ser-CGA-4-1	-58.9	-49.78	0.014	127	
Ser-GCT-1-1	-56.7	-52.68	0.023	127	
Ser-GCT-2-1	-57.3	-53.28	0.014	127	
Ser-GCT-3-1	-56.9	-54.68	0.033	127	
Ser-GCT-4-1	-56.9	-54.88	0.041	127	

Ser-GCT-4-2	-56.9	-54.88	0.041	127	
Ser-GCT-4-3	-56.9	-54.88	0.041	127	
Ser-GCT-5-1	-56.9	-55.48	0.039	127	
Ser-GCT-6-1	-60.4	-57.18	0.023	129	
Ser-TGA-1-1	-54.1	-51.18	0.037	127	
Ser-TGA-2-1	-52.9	-52.08	0.069	127	
Ser-TGA-3-1	-53.8	-52.08	0.081	127	
Ser-TGA-4-1	-56.9	-52.88	0.108	127	
Thr-AGT-1-1	-55.4	-52.98	0.073	119	
Thr-AGT-1-2	-55.4	-52.98	0.073	119	
Thr-AGT-1-3	-55.4	-52.98	0.073	119	
Thr-AGT-2-1	-50	-47.98	0.017	119	
Thr-AGT-2-2	-50	-47.98	0.017	119	
Thr-AGT-3-1	-48.7	-46.98	0.016	119	
Thr-AGT-4-1	-47.4	-48.18	0.018	119	
Thr-AGT-5-1	-55.4	-52.98	0.073	119	
Thr-AGT-6-1	-51.5	-50.08	0.022	119	
Thr-CGT-1-1	-48.7	-48.38	0.073	119	
Thr-CGT-2-1	-58.2	-51.38	0.025	117	
Thr-CGT-3-1	-52.4	-51.18	0.034	119	
Thr-CGT-4-1	-58.2	-51.38	0.022	117	
Thr-CGT-5-1	-49.8	-42.18	0.021	119	
Thr-TGT-1-1	-48.7	-48.88	0.066	119	
Thr-TGT-2-1	-48.7	-46.48	0.072	118	
Thr-TGT-3-1	-50.7	-45.78	0.05	118	
Thr-TGT-4-1	-48.9	-45.48	0.048	118	
Thr-TGT-5-1	-49.6	-44.48	0.028	118	
Thr-TGT-6-1	-48.9	-44.18	0.031	117	
Trp-CCA-1-1	-46.8	-43.08	0.029	117	
Trp-CCA-2-1	-44.9	-42.28	0.032	117	
Trp-CCA-3-1	-44.9	-39.68	0.009	117	
Trp-CCA-3-2	-44.9	-39.68	0.009	117	
Trp-CCA-3-3	-44.9	-39.68	0.009	117	
Trp-CCA-4-1	-46.3	-42.08	0.009	117	
Trp-CCA-5-1	-44.9	-40.88	0.014	117	
Tyr-ATA-1-1	-45.3	-44.08	0.013	118	
Tyr-GTA-1-1	-45.3	-42.18	0.058	118	
Tyr-GTA-2-1	-45.5	-44.78	0.088	118	
Tyr-GTA-3-1	-44.8	-42.18	0.04	118	
Tyr-GTA-4-1	-47.9	-46.08	0.052	118	
Tyr-GTA-5-1	-47.9	-44.88	0.046	118	
Tyr-GTA-5-2	-47.9	-44.88	0.046	118	
Tyr-GTA-5-3	-47.9	-44.88	0.046	118	
Tyr-GTA-5-4	-47.9	-44.88	0.046	118	
Tyr-GTA-5-5	-47.9	-44.88	0.046	118	
Tyr-GTA-6-1	-46.2	-43.18	0.054	118	
Tyr-GTA-7-1	-46.7	-41.58	0.025	118	
Tyr-GTA-8-1	-42.6	-41.48	0.055	118	
Tyr-GTA-9-1	-42	-39.08	0.013	117	
Val-AAC-1-1	-44.6	-44.18	0.049	118	
Val-AAC-1-2	-44.6	-44.18	0.049	118	
Val-AAC-1-3	-44.6	-44.18	0.049	118	
Val-AAC-1-4	-44.6	-44.18	0.049	118	
Val-AAC-1-5	-44.6	-44.18	0.049	118	
Val-AAC-2-1	-44.6	-44.18	0.047	118	
Val-AAC-3-1	-41	-39.98	0.046	118	
Val-AAC-4-1	-40.9	-43.28	0.041	118	
Val-AAC-5-1	-39.6	-41.98	0.019	118	
Val-AAC-6-1	-53.2	-51.08	0.056	117	
Val-CAC-1-1	-44.6	-44.18	0.043	118	
Val-CAC-1-2	-44.6	-44.18	0.043	118	
Val-CAC-1-3	-44.6	-44.18	0.043	118	
Val-CAC-1-4	-44.6	-44.18	0.043	118	
Val-CAC-14-1	-39.1	-39.78	0.018	1118	
Val-CAC-1-5	-44.6	-44.18	0.043	118	
Val-CAC-1-6	-44.6	-44.18	0.043	118	
Val-CAC-1-7	-44.6	-44.18	0.043	118	
Val-CAC-2-1	-45.5	-43.48	0.019	118	
Val-CAC-3-1	-44.8	-43.28	0.03	118	
Val-CAC-4-1	-42.4	-42.38	0.04	118	
Val-CAC-5-1	-44.6	-44.18	0.043	118	
Val-CAC-6-1	-41.3	-41.28	0.016	118	
Val-TAC-1-1	-44.6	-48.38	0.177	118	
Val-TAC-1-2	-44.6	-48.38	0.177	118	
Val-TAC-2-1	-43.9	-47.38	0.1	118	
Val-TAC-3-1	-43.6	-43.38	0.081	118	
Val-TAC-4-1	-43.1	-41.08	0.014	118	

Appendix B

Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging: Supplementary information

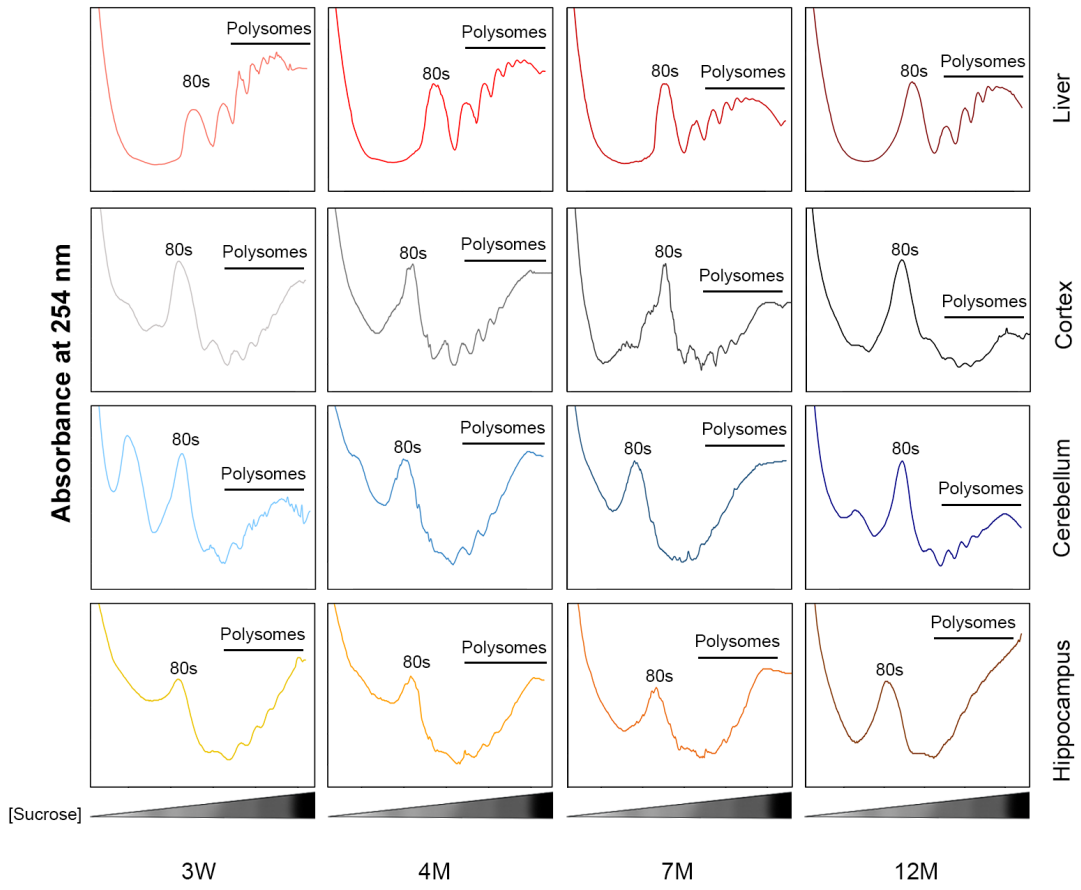


Figure B.1: Polysome profiles of brain and liver tissues. 80S and polysomal fractions were separated on sucrose gradients. For clear separation between 80S and polysomes, the fraction of the disome peak was omitted. W, week; M, month (published in [34]).

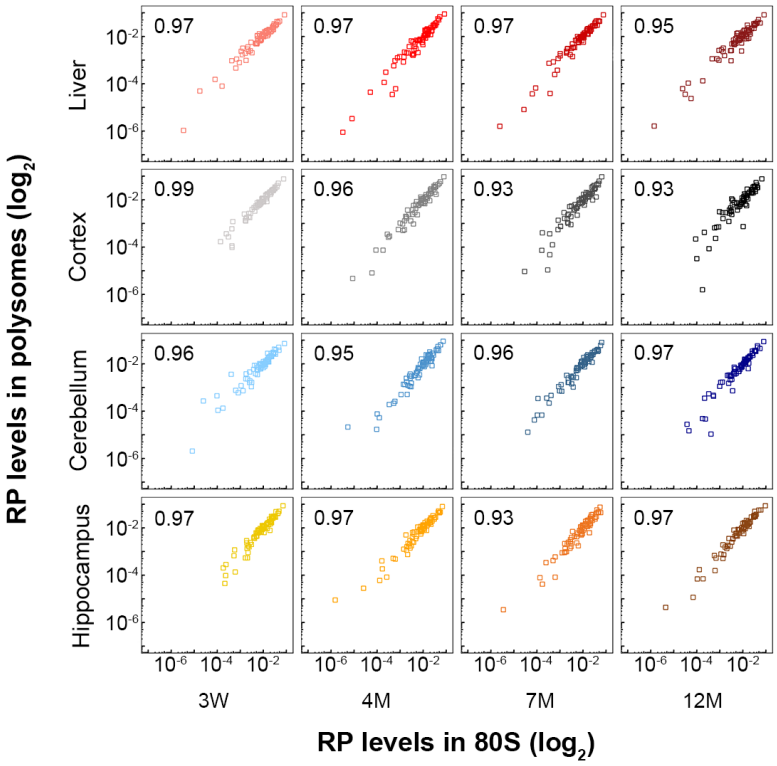


Figure B.2: Correlation between the RP levels in the 80S and polysomal fraction. R2, Pearson correlation coefficient is given in each plot and also summarized in Table S3. W, week; M, month (published in [34]).

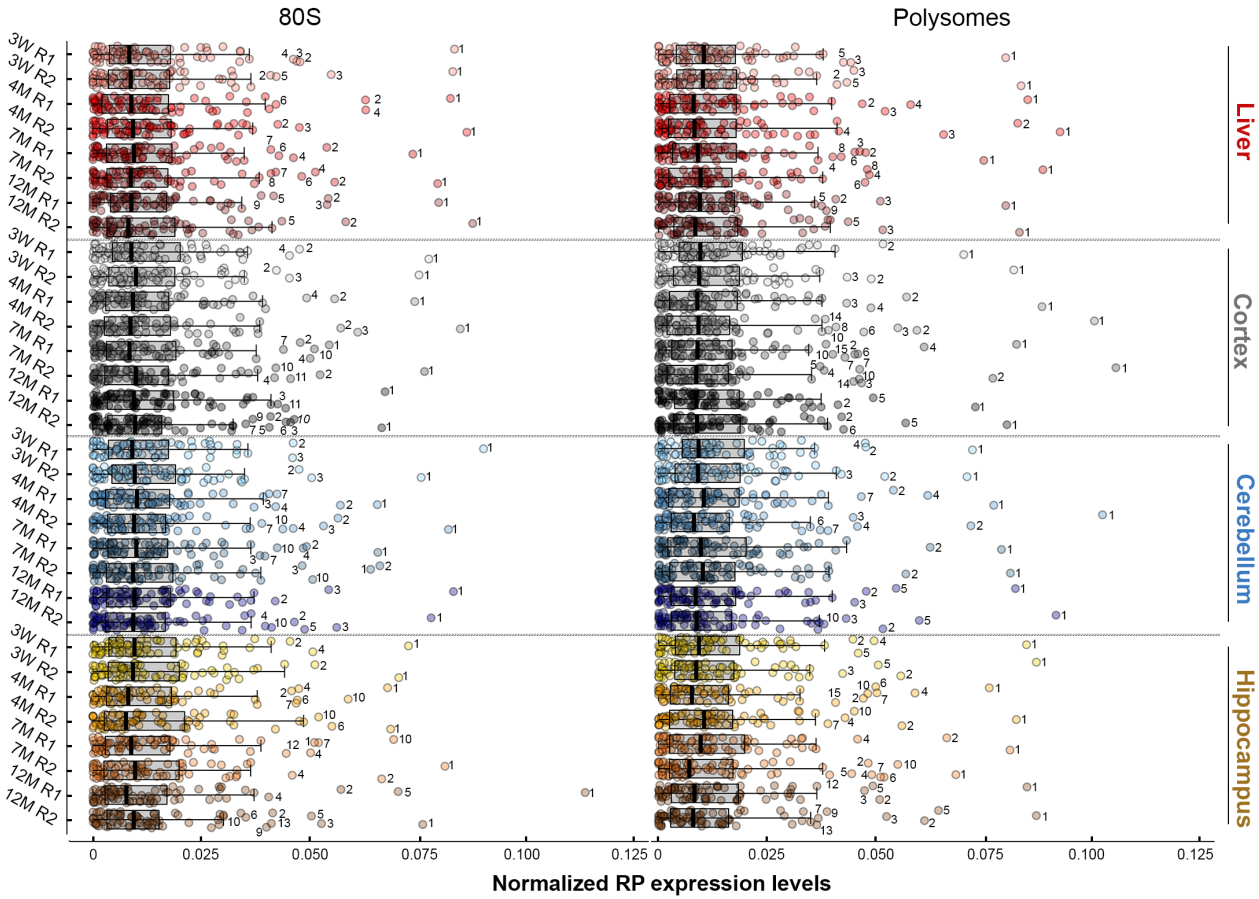


Figure B.3: Expression span of RPs in the replicates. Boxplot of the expression levels of RPs in 80S and polysomes in each analyzed sample. Each dot is a single RP. R1, R2 denote each of the two biological replicate; W, week and M, month. Boxplots are median of expression ± SD (n=2). The outliers are RPs with the highest detected levels and are designated as follows: 1 – uS3; 2 – uS2; 3 – uL4; 4 – uS4; 5 – RACK1; 6 – eL13; 7 – uL30; 8 – eS10; 9 – uL18; 10 – uL15; 11 – eS4; 12 – eL18; 13 – eS8; 14 – uL10; 15 – eS25. Both biological replicates are highly similar (p-values, Wilcoxon rank sum test, or each replicate pair are summarized in Table S3) (published in [34]).

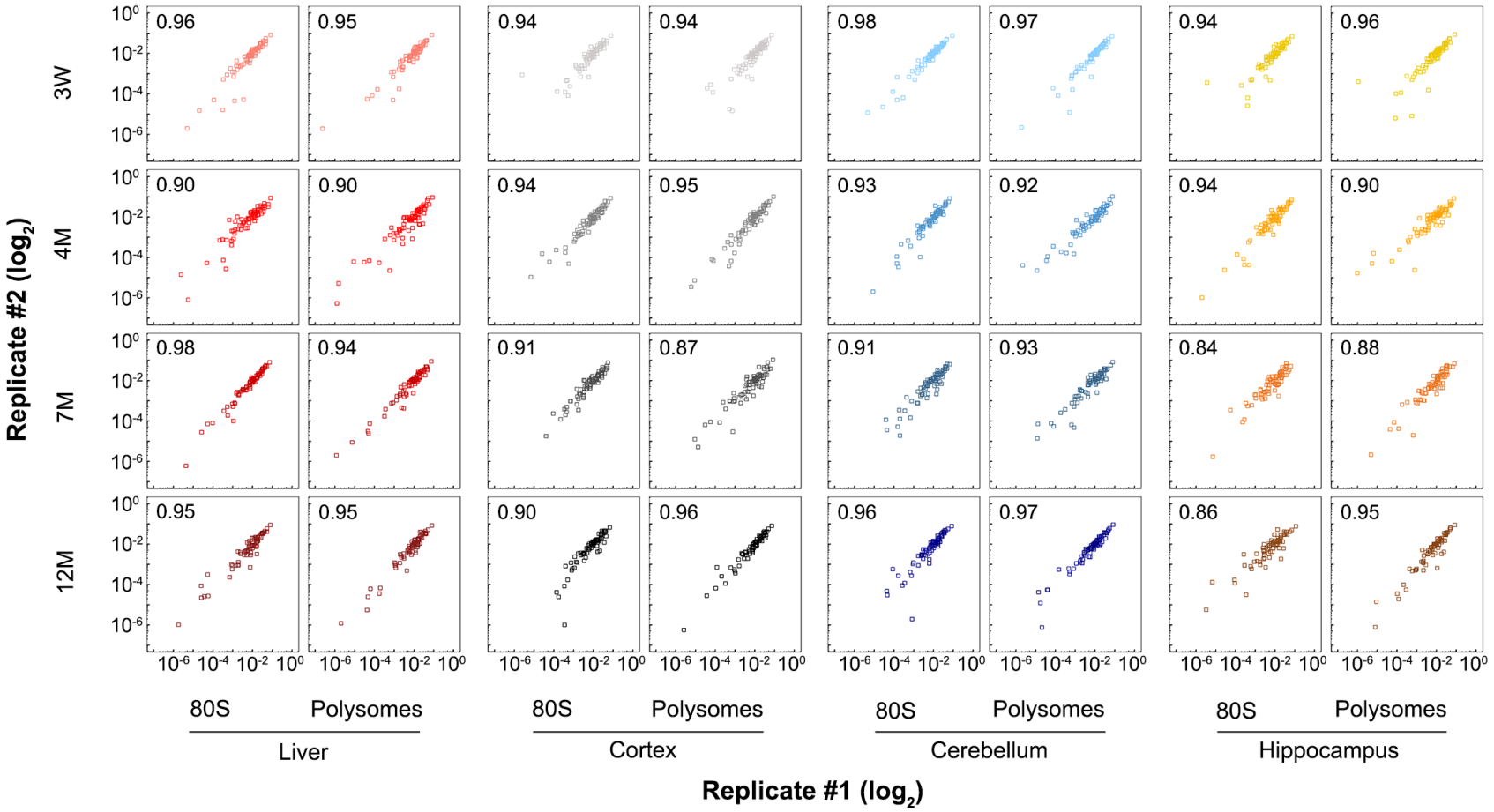


Figure B.4: Correlation between biological replicates. The RPs in each biological replicate consisted of pooled homogenized tissues from three littermates. R2, Pearson correlation coefficient is given in each plot and also summarized in Table S3. W, week; M, month (published in [34]).

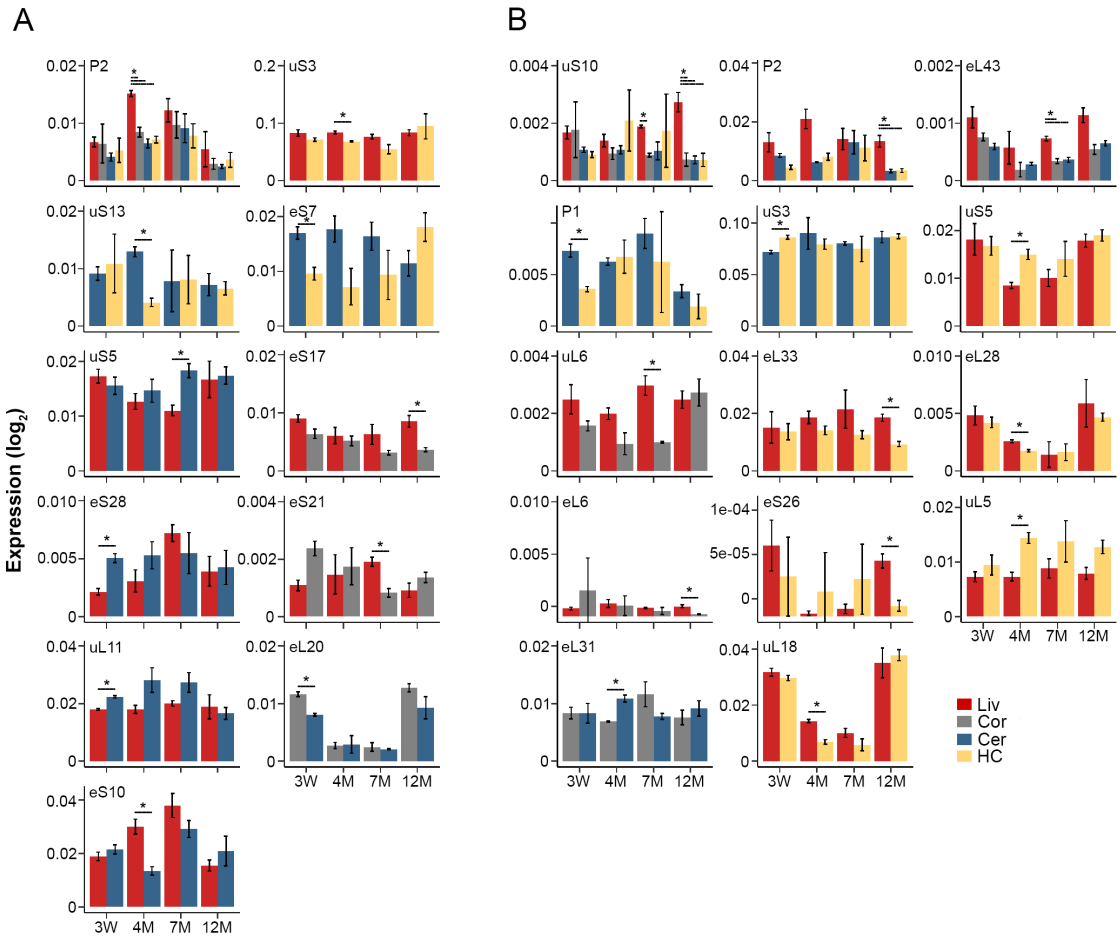


Figure B.5: The expression level of some RPs changes across tissues which changes however score insignificant. *, $P \leq 0.05$ in two tailed Student's t-test, which is insignificant by FDR < 0.1 considering the fluctuations between two biological replicates. W, week; M, month (published in [34]).

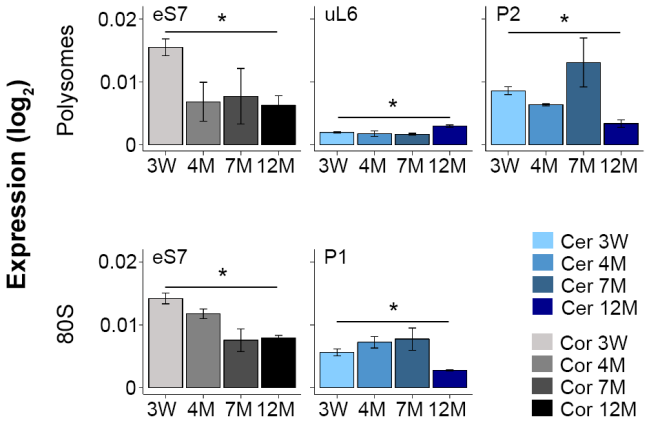


Figure B.6: The expression level of few RPs changes with age which scores insignificant. *, $P \leq 0.05$ in two tailed Student's t-test, which is insignificant by FDR < 0.1 considering the fluctuations between two biological replicates. W, week; M, month (published in [34]).

Table B.1: Number of detected and predicted peptides for each RP. n.d., not detected. The new nomenclature of naming ribosomal protein is considered [342].
[a]Note that in some cases the number of detected peptides can be higher than that of the in silico predicted peptides as peptides with modified amino acids are considered as unique peptide.
[b]All lengths of *in silico* predicted peptides were considered without length restriction. The shortest peptides were composed of three amino acids (published in [34]).

RPs large subunit	# of detected unique peptides in LC-MS/MS	# of in silico predicted[b] unique peptide	RPs small subunit	# of detected unique peptides in LC-MS/MS	# of in silico predicted[b] unique peptides
uL1	5	19	RACK1	16	22
uL2	7	20	eS1	10	26
uL3	12	34	uS2	16	20
uL4	18	33	uS3	18	21
uL5	4	15	eS4	13	22
uL6	4	13	uS4	11	22
eL6	7	26	uS5	10	25
eL8	15	20	eS6	9	22
uL11	6	12	uS7	6	15
uL13	9	22	eS7	8	16
eL13	9	20	uS8	7	12
uL14	6	10	eS8	10	15
eL14	4	18	uS9	7	14
uL15	4	14	uS10	3	8
eL15	7	19	eS10	6	17
uL16	2	20	uS11	6	8
uL18	11	28	uS12	1	11
eL18	6	13	eS12	7	10
eL19	5	17	uS13	11	14
eL20	6	18	uS14	1	8
eL21	5	16	uS15	6	14
uL22	8	16	uS17	8	14
eL22	4	10	eS17	9	10
uL23	6	11	uS19	n.d.	10
uL24	7	14	eS19	6	16
eL24	5	12	eS21	3	8
eL27	7	12	eS24	5	11
eL28	3	16	eS25	4	9
uL29	5	9	eS26	1	9
eL29	n.d.	13	eS27	n.d.	6
uL30	12	23	eS28	3	5
eL30	4	10	eS30	2	5
eL31	2	11	eS31	2	15
eL32	6	12	eS27L	1	6
eL33	3	12			
eL34	n.d.	11			
eL36	2	8			
eL37	n.d.	10			
eL38	5a	4			
eL39	2	3			
eL40	n.d.	17			
eL41	n.d.	0			
eL42	3	10			
eL43	2	6			
uL10	12	24			
P1	1	5			
P2	4	6			
eL22L	1	12			

Table B.2: Peptide coverage of detected RPs. n.d., not detected.
[a]Numbers denote the position of the first and last amino acid of the peptide in the RP sequence (published in [34]).

Rps large subunit	Length, amino acids	Total amino acid coverage, %	Sequence and position of the detected peptides[a]
Large subunit			
uL1	217	23.04	5-VSRDTLYEAVR-15; 79-AVDIPHMDIEALKK-92; 106-KYDAFLASESLIK-118; 107-YDAFLASESLIK-118; 162-VLCLAVAVGHVK-173; 79-AVDIPHMDIEALKK-92; 106-KYDAFLASESLIK-118; 107-YDAFLASESLIK-118; 162-VLCLAVAVGHVK-173
uL2	257	40.86	43-GIVKDIIHDPGR-54; 73-TELFIAAEGIHTGQFVYCGK-92; 93-KAQLNIGNVLPVGTMPEGTIVCCLEEKPGDR-123; 129-ASGNYATVISHNPETK-144; 164-AVVGVVAGGGR-174; 164-AVVGVVAGGGRIDKPILK-181; 234-KVGLIAAR-241
uL3	403	30.27	11-HGSLGFLPR-19; 35-DDASKPVHLTAFLGYK-50; 104-TVFAEHISDECK-115; 145-QLEKDFNSMK-154; 177-KAHLMEIQVNGGTVAEK-193; 178-AHLMEIQVNGGTVAEK-193; 199-ERLEQQVPVNQVFGQDEMIDVIGVTK-224; 201-LEQQVPVNQVFGQDEMIDVIGVTK-224; 250-KVACIGAWHPAR-261; 251-VACIGAWHPAR-261; I287-GQGYLIK-294; 301-NNASTDYDLSDK-312
uL4	419	36.04	1-MACARPLISVYSEK-14; 1-MACARPLISVYSEKGESSGK-20; 30-APIRPDIVNFVHTNLR-45; 30-APIRPDIVNFVHTNLRK-46; 47-NNRQPYAVSELAGHQTSAESWGTGR-71; 50-QPYAVSELAGHQTSAESWGTGR-71; 121-RYAICSALAASALPALVMSK-140; 122-YAICSALAASALPALVMSK-140; 144-IEEVPELPLVVEDK-157; 206-GPCIIYNEDNGIIK-219; 149-FCIWTESAFR-258; 249-FCIWTESAFRK-259; 259-KLDELYGTWR-268; 259-KLDELYGTWRK-269; 260-LDELYGTWR-268; 284-MMNTDLSR-291; 353-KLEAAATALATK-364; 354-LEAAATALATK-364
uL5	178	17.42	19-KLCLNICVGESGDR-32; 20-LCLNICVGESGDR-32; 30-LCLNICVGESGDRLTR-35; 39-VLEQLTGQTPVFSK-52
uL6	192	29.69	3-TILSNQTVDIPENVEITLK-21; 37-DFNHINVELSLLGKK-51; 72-TICSHVQNMIK-82; 130-TGVACSVSQAQK-141
eL6	296	25	123-YYPTEDVPR-131; 175-QLDSGLLLVTGPLVINRVPLR-195; 175-QLDSGLLLVTGPLVINRVPLRR-196; 209-VDISDVKIPK-218; 219-HLTDAYFK-226; 235-HQEGEIFDTEKEK-247; 269-IKAVPQLQGYLR-280
eL8	266	37.22	26-KVVNPLFEK-34; 27-VVNPLFEK-34; 38-NFGIGQDIQPK-48; 38-NFGIGQDIQPKR-49; 76-VPPAINQFTQALDR-89; 90-QTATQLLK-97; 138-AGVNTVTTLVENK-150; 138-AGVNTVTTLVENKK-151; 176-KMGVPYCIIK-185; 177-MGVPYCIIK-185; 197-KTCTTVAFTQVNSEDKGALAK-217; 198-TCTTVAFTQVNSEDK-212; 198-TCTTVAFTQVNSEDKGALAK-217; 235-RHWGGNVLGPK-245; 236-HWGGNVLGPK-245
uL11	165	49.7	32-IGPLGLSPK-40; 78-QAQIEVVPSASALIIK-83; 100-HSGNITFDEIVNIAR-114; 124-ELSGTIKEILGTAQSVGCNVDGR-146; 131-EILGTAQSVGCNVDGR-146; 147-HPHDIIDDINSGAVECPAS-165
uL13	203	34.98	1-MAEGQVLVLDGR-12; 1-MAEGQVLVLDGRGHLLGR-18; 38-CEGINISGNFYR-49; 54-YLAFLRK-60; 102-LKVLDGIPPPYDK-114; 104-VLDGIPPPYDKK-115; 117-RMVVPAALK-125; 118-MVVPAALK-125; 149-YQAVTATLEEK-159
eL13	211	38.39	6-NGMILKPHFHK-16; 22-VDTWFNQPAR-31; 22-VDTWFNQPARK-32; 75-GFSLEELR-82; 93-TIGISVDPR-101; 106-STESLQANVQR-116; 122-SKLILFPR-129; 136-KGDSSAEELKLATQLTGPVMPIR-158; 146-LATQLTGPVMPIR-158
uL14	140	35	14-FRISLGLPVGAVINCADNTGAK-35; 16-ISLGLPVGAVINCADNTGAK-35; 49-LNRLPAAGVGDMVMATVK-66; 52-LPAAGVGDMVMATVK-66; 52-LPAAGVGDMVMATVKK-67; 124-ECADLWPR-131
eL14	217	21.66	12-VAYISFGPHAGK-23; 24-LVAIVDVIDQNR-35; 54-CMQLTDFILK-63; 148-AAIAAAAAAAAAAK-160
uL15	148	23.65	66-NQSFCPTVNLDK-77; 78-LWTLVSEQTR-87; 93-NKTGVAPIIDVVR-105; 95-TGVAPIIDVVR-105
eL15	204	28.92	1-MGAYKYIQELWR-12; 6-YIQELWR-12; 57-QGYVIYR-63; 115-VLNSYWVGEDSTYK-128; 129-FFEVILIDPFHK-140; 144-RNPDTQWITKPVHK-157; 145-NPDTQWITKPVHK-157
uL16	214	6.54	176-FNADEFEDMVAEK-188; 176-FNADEFEDMVAEKR-189
uL18	297	44.44	55-VTNRDIICQIAYAR-68; 59-DIICQIAYAR-68; 69-IEGDMIVCAAYAHELPK-85; 90-VGLTNYAAAYCTGLLLAR-107; 153-TTTGNKVFGALK-164; 165-GAVDGGLSIPHSTKR-179; 179-RFPGYDSESKEFNAEVHR-196; 197-KHIMGQNVADYMR-209; 198-HIMGQNVADYMR-209; 210-YLMEEDEDAYKK-221; 229-NNVTPDMMEEMYKK-242
eL18	188	21.81	38-RTNSTFNQVVLK-49; 39-TNSTFNQVVLK-49; 39-TNSTFNQVVLKR-50; 79-TAVVVGTVTDDVR-91; 120-ILTFDQLALESPK-132; 120-ILTFDQLALESPKGR-134
eL19	196	18.88	20-KKVWLDPNETNEIANANSR-38; 21-KVWLDPNETNEIANANSR-38; 22-VWLDPNETNEIANANSR-38; 44-LIKDGLIIR-52; 154-LLADQAEAR-162
eL20	176	28.98	42-SRFWYFVSQLKK-53; 57-SSGEIVYCGQVFEK-70; 75-VKNFGIWLR-83; 77-NFGIWLR-83; 96-EYRDLTTAGAVTQCYR-111; 99-DLTTAGAVTQCYR-111
eL21	160	37.5	21-KHGVVPLATYMR-32; 22-HGVVPLATYMR-32; 36-KGDIVDIKGMGTVQK-50; 64-VYNVTQHAVGIIVNK-78; 143-TNGKEPELLEPIPYEFMA-160
uL22	184	38.59	4-YSLDPENPTK-13; 75-QWGWTQGR-82; 75-QWGWTQGRWPK-85; 86-KSAEFLLHMLK-96; 87-SAEFLLHMLK-96; 97-NAESNAELKGLDVDSLVIEHIQVNK-121; 106-GLDVDSLVIEHIQVNK-121; 154-EQIVPKPEEEVAQK-167
eL22	128	21.09	53-AGNLGGGVVTIER-65; 68-SKITVTSEVPFSK-80; 68-SKITVTSEVPFSKR-81; 70-ITVTSEVPFSKR-81
uL23	156	28.21	69-NKLDHYAIIK-78; 71-LDHYAIIKFPLTTESAMK-88; 79-FPLTTESAMK-88; 89-KIEDNNTLVFIVDVK-103; 115-KLYDIDVAK-123; 116-LYDIDVAK-123
uL24	145	38.62	1-MKFNPFVTSDR-11; 3-FNPFVTSDR-11; 29-IMSSPLSK-36; 46-SMPIRKDDEVQVVR-59; 76-KKYVIYIER-84; 78-YVIYIER-84; 90-ANGTTVHVGIHPSK-103
eL24	157	29.3	1-MKVELCSFSGYK-12; 3-VELCSFSGYK-12; 24-TDGKVFQFLNAK-35; 48-QINWTVLYR-56; 81-AITGASLADIMAK-93
eL27	136	47.06	10-VVLVLAGR-17; 28-NIDDGTSDRPYSHALVAGIDR-48; 28-NIDDGTSDRPYSHALVAGIDRYPR-51; 74-VYNYNHLMPTR-84; 94-TVVNKDVFRDPALK-107; 99-DVFRDPALK-107; 127-NKWFFQK-133

eL28	137	21.9	1-MSAHLQWMVVR-11; 12-NCSSFLIK-19; 23-QTYSTEPNNLK-33
uL29	123	30.08	20-QLDDLKVELSQLR-32; 57-VLTVINQTQK-66; 57-VLTVINQTQKENLR-70; 77-KYKPLDLRPK-86; 78-YKPLDLRPK-86
eL29	n.d.		
uL30	270	30	10-KVATVPGTLK-19; 31-KKVPAVPETLK-41; 32-KVPAVPETLK-41; 99-KAGNFYVPAEPK-110; 100-AGNFYVPAEPK-110; 129-KVLQLLR-135; 138-QIFNGTFVK-146; 171-SVNELIYK-178; 188-RIALTDNSLIAR-199; 189-IALTDNSLIAR-199; 223-FKEANNFLWPFK-234; 225-EANNFLWPFK-234
eL30	115	51.3	45-LVILANNCPALR-56; 57-KSEIEYYAMLAK-68; 69-TGVHHYSGNNIELGTACGK-87; 91-VCTLAIIDPGDSDIIR-106
eL31	125	18.4	93-NEDEDSPNKLYTLVTYVPVTTFK-115; 102-LYTLVTYVPVTTFK-115
eL32	135	25.93	49-FKGQILMPNIGYGSNK-64; 49-FKGQILMPNIGYGSNKK-65; 51-GQILMPNIGYGSNK-64; 51-GQILMPNIGYGSNKK-65; 68-HMLPSGFR-75; 84-ELEVLLMCNK-93
eL33	110	22.73	30-IEGVYARDETEFYLGKR-46; 37-DETEFYLGKR-46; 103-VMLYPSRI-110
eL34	n.d.		
eL36	105	21.91	1-MALRYPMAVGLNK-13; 46-EVCGFAPYER-55
eL37	n.d.		
eL38	70	52.86	4-KIEEIKDFLLTAR-16; 5-IEEIKDFLLTAR-16; 41-YLYTLVITDKEK-52; 56-LKQSLPPGLAVK-67; 58-QSLPPGLAVK-67
eL39	51	23.53	17-QKQNRPIPQWIR-28; 19-QNRPIPQWIR-28
eL40	n.d.		
eL41	n.d.		
eL42	106	25.47	45-QSGYGGQTKPIFR-57; 65-KIVLRLECVEPNCR-78; 70-LECVEPNCR-78
eL43	92	32.61	51-AVGIWHCGSCMK-62; 63-TVAGGAWTYNTTSAVTVK-80
uL10	317	46.06	17-IIQLLDDYPK-26; 27-CFIVGADNVGSK-38; 84-GNVGFVFTK-92; 93-EDLTEIRDMLLANK-106; 93-EDLTEIRDMLLANKVPAAAR-112; 100-DMLLANKVPAAAR-112; 113-AGAIAPCEVTVPAQNTGLGPEK-134; 113-AGAIAPCEVTVPAQNTGLGPEKTSFFQALGITTK-146; 135-TSFFQALGITTK-146; 150-GTIEILSDVQLIK-162; 248-VLALSVETEYTFPLTEK-264; 267-AFLADPSAFAAAAPAAAATTAAPAAAAAPAK-297
P1	114	14.04	34-AAGVSVEPFWPGLFAK-49
P2	115	41.74	3-YVASYLLAALGGNSSPSAK-21; 25-KILDSVGIEADDDRLNK-41; 26-ILDSVGIEADDDRLNK-41; 50-NIEDVIAQGVGK-61
eL22L	122	9.84	48-TGNLGNVVHIER-59;
Small subunit			
RACK1	317	68.14	13-GHNGWVTQIATTPQFPDMILSASR-36; 37-DKTIIMWK-44; 89-LWDLTTGTTTR-99; 89-LWDLTTGTTTRR-100; 107-DVLSVAFSSDNR-118; 131-LWNTLGVCK-139; 140-YTVQDESHSEWVSCVR-155; 156-FSPNSSNPIIVSCGWDK-172; 184-LKTNHIGHTGYLNTVTVSPDGSLCASGGK-212; 186-TNHIGHTGYLNTVTVSPDGSLCASGGK-212; 213-DGQAMLWDLNEGK-225; 226-HLYTLDGGDIINALCFSPNR-245; 246-YWLCAATGPSIK-257; 258-IWDLEGK-264; 281-AEPPQCTSLAWSADGQTLFAGYTDNLVR-208; 309-VWQVTIGTR-317
eS1	264	40.53	35-APAMFNIR-42; 66-VFEVSLADLQNDEVAFR-82; 66-VFEVSLADLQNDEVAFRK-83; 84-FKLITEDVQGK-94; 95-NCLTNFHGMDLTR-107; 137-LFCVGFTK-144; 168-MMEIMTR-174; 175-EVQTNDLKEVVNK-187; 200-ACQSIYPLHDVFVR-213; 250-VERADGYEPPVQESV-264
uS2	295	65.42	1-MSGALDVLQMK-11; 1-MSGALDVLQMKEEDVLK-17; 18-FLAAGTHLGGTNLDFQMEQYIYK-40; 42-KSDGIYIINLK-52; 42-KSDGIYIINLKR-53; 43-SDGIYIINLK-52; 64-AIVAIENPADVSVISSR-80; 86-AVLKFAAATGATPIAGR-102; 90-FAAATGATPIAGR-102; 103-FTPGTFTNQIQAAFR-117; 103-FTPGTFTNQIQAAFREPR-120; 121-LLVVTDPR-128; 129-ADHQPLTEASYVNLPTIALCNTDSPLR-155; 156-YVDIAIPCNNK-166; 167-GAHSVGLMWWMLAR-180; 192-EHPWEVMPDLYFYRDPEEIEKEEQAAAEK-220
uS3	243	73.66	10-KFVADGIFK-18; 11-FVADGIFK-18; 19-AELNEFLTR-27; 28-ELAEDGYSGVEVR-40; 46-TEIIILATR-54; 67-IRELTAVVQK-76; 67-IRELTAVVQKR-77; 78-FGFPEGSVELYAEK-91; 96-GLCAIAQAESLR-107; 108-YKLLGGLAVR-117; 135-GCEVVVSGK-143; 154-FVDGLMIHSGDPVNYYVDTAVR-175; 181-QGVLGIKVK-189; 188-VKIMLPWDPSGK-199; 190-IMLPWDPSGK-199; 204-KPLPDHVSIVEPKDEILPTTPISEQK-229; 217-DEILPTTPISEQK-229; 230-GGKPEPPAMPQPVPTA-245
eS4	263	47.91	38-LRECLPLIIFLR-49; 52-LKYALTGDEVKK-63; 76-VRTDITYPAGFMDVISIDK-94; 76-VRTDITYPAGFMDVISIDKTGENFR-100; 78-TDITYPAGFMDVISIDK-94; 146-TIRYPDPLIK-155; 149-YPDPLIK-155; 156-VNDTIQIDLETGK-168; 169-ITDFIKFDTGNLCMVTGGANLGR-191; 175-FDTGNLCMVTGGANLGR-191; 201-HPGSFDVVHVK-211; 201-HPGSFDVVHVKDANGNSFATR-221; 231-GNKPWISLPR-240
uS4	194	37.63	23-SRLDQELKLIGEYGLR-38; 31-LIGEYGLR-38; 59-ELLTLDEKDPR-69; 59-ELLTLDEKDPRR-70; 70-RLFEGNALLR-79; 71-LFEGNALLR-79; 92-MKLDYILGLK-101; 102-IEDFLER-108; 139-KQVVNIPSFIVR-150; 140-QVVNIPSFIVR-150; 156-HIDFSLR-162
uS5	293	35.15	55-AEDKEWIPVTK-65; 126-AFVAIGDYNGHVGLGVK-142; 160-LSIVPVRR-167; 201-GTGIVSAPVPK-211; 212-KLLMMAGIDDCYTSAR-227; 213-LLMMAGIDDCYTSAR-227; 228-GCTATLGNFAK-238; 247-TYSYLTPDLWK-257; 247-TYSYLTPDLWKETVFTK-263; 264-SPYQEFTDHLVK-275
eS6	249	25.3	1-MKLNISFPATGCQK-14; 3-LNISFPATGCQK-14; 31-RMATEVAADALGEEWKGYVVR-51; 32-MATEVAADALGEEWK-46; 32-MATEVAADALGEEWKGYVVR-51; 116-KGEKDIPGLTDTTVPR-131; 117-GEKDIPGLTDTTVPR-131; 143-KLFNLSKEDDVR-154; 144-LFNLSKEDDVR-154
uS7	204	31.86	1-MTEWEAATPAVAETPDIK-18; 1-MTEWEAATPAVAETPDIKLFGK-22; 23-WSTDDVQINDISLQDYIAVK-42; 137-QAVDVSPLR-145; 146-RVNQAIWLLCTGAR-159; 147-VNQAIWLLCTGAR-159
eS7	194	51.55	8-IVKPNGEKPDEFESGISQALLELEMNSDLK-37; 58-KAIIIFVPVPQLK-70; 59-AIIIFVPVPQLK-70; 121-TLTAVHDAILEDLVFPSEIVGKR-143; 161-AQQNNVEHKVETFSGVYK-178; 170-VETFSGVYKK-179; 180-LTGKDVNFEFPEFQL-194

uS8	130	46.15	4-MNVLADALK-12; 37-FLTVMMK-43; 44-HGYIGEFEIIDDHR-57; 61-IVVNLTGR-68; 61-IVVNLTGRLNK-71; 79-FDVQLKDLEKWQNNLLPSR-97; 89-WQNNLLPSR-97
eS8	716	54.81	57-ALRLDVGNFSWGSECCTR-74; 60-LDVGNFSWGSECCTR-74; 78-IIDVVYNASNNELVR-92; 99-NCIVLIDSTPYR-110; 111-QWYESHYALPLGR-123; 129-LTPEEEEILNKKR-141; 158-ISSLLEEQFQQGK-170; 171-LLACIASRPGQCGR-184; 185-ADGYVLEGKELEFYLR-200; 194-ELEFYLR-200
uS9	146	39.73	1-MPSKGPLQSVQVFGR-15; 5-GPLQSVQVFGR-15; 34-VNGRPLEMIEPR-45; 74-GGGHVAQIYAIR-85; 107-EIKDILIQYDR-117; 110-DILIQYDR-117; 118-TLLVADPR-125
uS10	119	26.05	1-MAFKDTGKTPVEPEVAIHR-19; 9-TPVEPEVAIHR-19; 88-LIDLHSPSEIVK-99
eS10	165	39.39	32-HPELADKNVPNLHVMK-47; 56-GYVKEQFAWR-65; 81-DYLHLPPEIVPATLR-95; 130-SAVPPGADKKAEAGAGSATEFQFR-153; 139-KAEAGAGSATEFQFR-153; 140-AEAGAGSATEFQFR-153
uS11	151	33.11	62-VKADRDESSPYAAMLAAQDVAQR-84; 64-ADRDESSPYAAMLAAQDVAQR-84; 87-ELGITALHIK-96; 126-IGRIEDVTPIPSDSTR-141; 129-IEDVTPIPSDSTR-141; 129-IEDVTPIPSDSTRR-142
uS12	143	7.69	125-VANVSLLALYK-135
eS12	132	59.09	1-MAEEGIAAGGVMDVNTALQEVLK-23; 46-QAHLCVLASNCDEPMYVK-63; 85-LGEWVGLCK-93; 102-KVVGCSCVVVK-112; 102-KVVGCSCVVVKDYGK-116; 117-ESQAKDVIEEYFK-129; 122-DVIEEYFK-129
uS13	152	49.34	1-MSLVIPEK-8; 1-MSLVIPEKFQHILR-14; 25-KIAFAITAIK-34; 26-IAFAITAIK-34; 55-RAGELTEDEVER-66; 56-AGELTEDEVER-66; 56-AGELTEDEVERVITIMQNPR-75; 67-VITIMQNPR-75; 76-QYKIPDWFLNR-86; 95-YSQVLANGLDNK-106; 95-YSQVLANGLDNKLREDLER-113
uS14	56	21.43	1-MGHQQLYWSHPR-12
uS15	151	37.75	10-GLSQSALPYR-19; 28-LTSDDVKEQIYK-39; 43-KGLTPSQIGVILR-55; 44-GLTPSQIGVILR-55; 79-GLAPDLPEDLYHLIK-93; 115-LILIESR-121
uS17	158	46.2	13-QPTIFQNK-20; 49-EAIEGTYIDKK-59; 59-KCPFTGNVSIR-69; 60-CPFTGNVSIR-69; 72-ILSGVVTK-79; 91-DYLHYIR-97; 108-NMSVHLSPCFR-118; 119-DVQIGDIVTVGECRPLSK-136
eS17	135	53.33	33-RVCEEIAIIPSK-44; 33-RVCEEIAIIPSKK-45; 34-VCEEIAIIPSK-44; 34-VCEEIAIIPSKK-45; 50-IAGYVTHLMK-59; 81-RDNYVPEVSALDQEIIEVDPDTK-103; 81-RDNYVPEVSALDQEIIEVDPDTKEMLK-107; 82-DNYVPEVSALDQEIIEVDPDTK-103; 108-LLDFGSLSNLQVTQPTVGMNFK-129
uS19	n.d.		
eS19	145	45.52	1-MPGVTVKDVNQQEFVR-16; 8-DVNQQEFVR-16; 28-LKVPEWVDTVK-38; 42-HKELAPYDENWFYTR-56; 102-RVLQALEGLK-111; 130-DLDRIAGQVAAANK-143
eS21	83	40.96	1-MQNDAGEFVDLYVPRK-16; 28-DHASIQMNVAEVDR-41; 28-DHASIQMNVAEVDRTTGR-45
eS24	133	27.07	1-MNDTVTIR-8; 21-KQMVIDVLHPGK-32; 22-QMVIDVLHPGK-32; 69-TTGFGMIYDSLDYAK-83; 69-TTGFGMIYDSLDYAKK-84
eS25	125	20.8	42-DKLNNLVLFDK-52; 42-DKLNNLVLFDKATYDK-57; 44-LNNLVLFDK-52; 67-LITPAVVSER-76
eS26	115	10.44	71-LHYCVSCAIHSK-82
eS27	n.d.		
eS28	69	27.54	48-GPVREGDVLTLLESER-63; 52-EGDVLTLLESER-63; 52-EGDVLTLLESEREAR-66
eS30	59	20.34	41-RFVNVVPTFGK-51; 42-FVNVVPTFGKK-52
eS31	156	12.18	120-ECPSDECGAGVFMGSHFDR-138
eS27L	84	14.29	6-DLLHPSLEEEKK-17

Table B.3: Correlation between RP levels in both biological replicates and between different tissues and ages. R^2, Pearson correlation coefficient for expression in tissues (upper panel, data summarized in Figure 7.1B) and ages (lower panel, data summarized in Figure 7.1C). Cer, cerebellum; HC, hippocampus; Cor, cortex, Liv, liver; P, polysomes; W, week; M, month (published in [34]).

Compared to Liv 80S	Cer 80S	Cor 80S	HC 80S	Compared to Liv P	Cer P	Cor P	HC P
3 weeks R^2	0.98	0.97	0.96	3 weeks R^2	0.94	0.95	0.97
4 months R^2	0.93	0.95	0.91	4 months R^2	0.95	0.97	0.91
7 months R^2	0.94	0.93	0.89	7 months R^2	0.95	0.94	0.92
12 months R^2	0.97	0.94	0.96	12 months R^2	0.95	0.92	0.94

Comparisons	Cer 80S	Cer P	Cor 80S	Cor P	HC 80S	HC P	Liv 80S	Liv P
3W vs. 4M R^2	0.91	0.96	0.94	0.95	0.89	0.9	0.95	0.89
3W vs. 7M R^2	0.89	0.94	0.9	0.93	0.88	0.89	0.89	0.86
3W vs. 12M R^2	0.97	0.95	0.94	0.95	0.94	0.97	0.98	0.97

Table B.4: Correlation between the RP expression levels in different samples. Pearson correlation coefficient (R^2) for the 80S and polysomal fractions (upper panel, data summarized in Figure B.2) and the whole RPs set (global) or for RPs within expression values in the lower $1/3^{rd}$ between biological replicates (lower panel, data summarized in Figure B.4). P values for each two biological replicates from Wilcoxon rank sum test (middle panel, data summarized in Figure B.3). Cer, cerebellum; HC, hippocampus; Cor, cortex; Liv, liver; P, polysomes; W, week; M, month (published in [34]).

Ages	Cer 80S vs P	Cor 80S vs P	HC 80S vs P	Liv 80S vs P
3 weeks	0.96	0.99	0.97	0.97
4 months	0.95	0.96	0.97	0.97
7 months	0.96	0.93	0.93	0.97
12 months	0.97	0.93	0.97	0.95

Samples	Cer 80S	Cer P	Cor 80S	Cor P	HC 80S	HC P	Liv 80S	Liv P
3 weeks	0.9	0.97	0.93	0.85	0.92	0.92	0.95	0.94
4 months	0.79	0.98	0.77	0.75	0.89	0.66	0.57	0.85
7 months	0.79	0.86	0.99	0.8	0.78	0.65	0.85	0.89
12 months	0.73	0.89	0.98	0.77	0.67	0.62	0.72	0.64

	Cer 80S	Cer P	Cor 80S	Cor P	HC 80S	HC P	Liv 80S	Liv P
3W								
R^2 global	0.98	0.97	0.94	0.94	0.94	0.96	0.96	0.95
R^2 lower $1/3^{rd}$	0.92	0.94	0.75	0.84	0.69	0.82	0.77	0.83
4M								
R^2 global	0.93	0.92	0.94	0.95	0.94	0.9	0.9	0.9
R^2 lower $1/3^{rd}$	0.88	0.66	0.84	0.87	0.79	0.62	0.6	0.7
7M								
R^2 global	0.91	0.93	0.91	0.87	0.84	0.88	0.98	0.94
R^2 lower $1/3^{rd}$	0.81	0.8	0.69	0.71	0.77	0.85	0.93	0.73
12M								
R^2 global	0.96	0.97	0.9	0.96	0.86	0.95	0.95	0.95
R^2 lower $1/3^{rd}$	0.85	0.93	0.83	0.87	0.67	0.83	0.88	0.87

Table B.5: Correlation between the RP levels in seven and twelve-month samples compared to four month sample. R^2, Pearson correlation coefficient (data are summarized in Figure 7.3). Cer, cerebellum; HC, hippocampus; Cor, cortex; Liv, liver; P, polysomes; W, week; M, month (published in [34]).

	Cer 80S	Cer P	Cor 80S	Cor P	HC 80S	HC P	Liv 80S	Liv P
4M vs 7M	0.96	0.97	0.96	0.97	0.97	0.98	0.97	0.95
4M vs 12M	0.9	0.91	0.9	0.9	0.81	0.84	0.93	0.91

List of publications

Internationally indexed journals [1]

1. Amirbeigiarab, S., Kiani, P., Velazquez Sanchez, A., Krisp, C., Kazantsev, A., Fester, L., Schlüter, H., and Ignatova, Z., 2019. "Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging". Proceedings of the National Academy of Sciences,116(45), pp. 22567–22572. DOI: https://doi.org/10.1073/pnas.1912060116 [Chapter 7]

Articles in preparation [2]

1. Velazquez Sanchez Ana Karen, Klopprogge Bjarne, Zimmermann Karl-Heinz, Ignatova Zoya. "Engineering synthetic sRNAs as dynamical modulators of gene expression". [Chapter 3]
2. Velazquez Sanchez Ana Karen, Albers Suki, Zimmermann Karl-Heinz, Ignatova Zoya. "Towards semi-autonomous circuits for the diagnosis and mitigation of genetic diseases". [Chapter 5]

[1] Articles already published, in press, or formally accepted for publication.
[2] Articles under review, communicated, or to be communicated.