ACADEMIC YEAR 2023-2024

UNIVERSITÀ DEGLI STUDI DI PADOVA
Comparative Biomedicine and Food Science

Second Cycle Degree (MSc)
Biotechnologies for Food Science

An Engineered Bacterial Transcription Unit Controlled by RNA:DNA Triplexes

Supervisor
Prof. Fabio Vianello
Co-Supervisor
Dr. Alessandro Cecconello

Submitted by
Nima Fouladi Ghareshiran

Student no.
2005585

ANNO ACCADEMICO 2023-2024

UNIVERSITÀ DEGLI STUDI DI PADOVA
Biomedicina comparata e alimentazione

Laurea specialistica (MSc)
Biotechnologies for Food Science

Un'unità trascrizionale batterica ingegnerizzata e controllata da RNA:DNA Triplex

Relatore
Prof. Fabio Vianello
Correlatore
Dr. Alessandro Cecconello

Laureando
Nima Fouladi Ghareshiran

Matricola
2005585

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ENGLISH ABSTRACT

Organisms can control gene expression at different levels, for example by making some parts of the genome less accessible to RNA polymerase or by changing the affinity of the transcription complex for specific promoter sequences using transcription factors. Recently, special RNA molecules, called long non-coding RNAs, are more and more at the center of scientific interest due to their involvement in gene regulation. Although experimental evidences are still scarce, it was proposed that they can directly interact with regulatory genome sequences via formation of triplex structures.

In this study, a recently proposed model mechanism for gene regulation via triplex formation is further investigated for the production of a molecular gating system reproducing the logic gate XOR, where the triplex forming RNA molecules represent the input, while the transcribed product represents the output. To achieve this goal, a bacterial expression unit was engineered to containing a triplex-controlled promoter upstream of a sequence expressing the fluorogenic aptamer Broccoli. The fluorescence signal of Broccoli was used to estimate the transcription rate while the triplex formation was characterized by gel electrophoresis. The results showed that the input combinations (0-0, 0-1, 1-0, and 1-1) generated outputs in agreement with the designed XOR gate (0, 1, 1, and 0, respectively), demonstrating the feasibility of building triplex-mediated molecular computing units. Furthermore, this molecular gate mechanism could be used to develop a threshold gate, i.e., the fundamental node of neural networks, and thus it can be considered a universal molecular computing mechanism. In conclusion, triplex formation and fluorogenic aptamers can be used to study a new level of gene regulation. This is of central importance in biology, as well as for the development of molecular computing systems.

ABSTRACT IN ITALIANO

Le cellule possono controllare l'espressione genica a diversi livelli, ad esempio rendendo alcune parti del genoma meno accessibili all'RNA polimerasi o utilizzando fattori di trascrizione per modificare l'affinità del complesso di trascrizione per specifiche sequenze del promotore. Recentemente, particolari molecole di RNA, chiamate long non-coding RNA, sono sempre più al centro dell'interesse scientifico per il loro coinvolgimento nella regolazione genica. Sebbene le evidenze sperimentali siano ancora scarse, è stato proposto che tali molecole possano interagire direttamente con sequenze genomiche regolatorie mediante la formazione di strutture triplex.

In questa tesi è stato approfondito lo studio di un nuovo meccanismo per la regolazione genica tramite la formazione di strutture triplex. In particolare, lo studio riguarda la produzione di un sistema di *gating* molecolare che riproduce l'operatore logico XOR, dove le molecole di RNA che formano i triplex rappresentano gli input, mentre il prodotto trascritto rappresenta l'output. A questo scopo, è stata ingegnerizzata un'unità di espressione batterica inserendo un promotore controllato da triplex a monte di una sequenza per l'aptamero fluorogenico Broccoli. L'emissione di fluorescenza di broccoli è stata utilizzataa per stimare la velocità di trascrizione mentre la formazione del triplex è stata caratterizzata per mezzo di elettroforesi su gel. I risultati mostrano che le combinazioni di input 0-0, 0-1, 1-0 e 1-1 hanno generato output in accordo con l'operatore XOR, rispettivamente 0, 1, 1 e 0, dimostrando la fattibilità di progettazione e costruzione di *gate* molecolari mediati da triplex. Inoltre, questo meccanismo di *gate* molecolare potrebbe essere utilizzato per sviluppare un *gate soglia*, ovvero il nodo fondamentale nelle reti neurali, e quindi un meccanismo di calcolo molecolare universale.

In conclusione, la formazione di triplex e la produzione di aptameri fluorogenici possono essere utilizzati congiuntamente per studiare un nuovo livello di regolazione genica. Tale approccio è particolarmente importante per lo studio della biologia del controllo dell'espressione genica, così come per lo sviluppo di sistemi di *gating* molecolare.

1. INTRODUCTION

In this chapter, an overview of the secondary structures of nucleic acids is presented. Specific biological functions of nucleic acid structures, such as the triplex structure, are introduced, along with the description of a family of recently developed RNA structures called fluorogenic aptamers. Furthermore, a short summary of the mechanism of gene expression in E. coli is presented along with dedicated sections presenting the long non-coding RNAs, an artificial switchable genetic element called "genelet", the logic operators AND or XOR and the principles of their molecular implementation, and, lastly, the proposed role of nucleic acid triplexes in gene regulation.

1.1 NUCLEIC ACID SECONDARY STRUCTURES

Nucleic acid molecules, deoxyribonucleic acid DNA and ribonucleic acid RNA, are polyelectrolytes constituted of ribonucleotides and deoxyribonucleotides respectively, and they are the molecular material responsible for transferring information in organisms during reproduction (i.e., genome duplication process) as well as storing information in a single cell for the production of the necessary proteins (i.e., transcription and translation processes). These two nucleic acids, DNA and RNA, are responsible for storing and duplicate information (DNA) and for using the stored information to produce proteins (RNA). Although they share some chemical similarities, their differences reflect their different functions. For example, DNA is mostly found in cells as a double helix comprising two antiparallel linear polymers that exclude the inner part from the solvent, thus having a good stability, while RNA is produced as a single linear polymer that can chemically interact more easily with other cell components. In addition, their composition is slightly different: DNA monomers consist of a phosphate group, one monosaccharide deoxyribose unit, and four nitrogen bases called adenine (A), thymine (T), guanin (G), cytosine (C) while for RNA, deoxyribose is replaced by ribose and uracil (U) replaces thymine (T).

Watson-Crick hydrogen bonds between nitrogen bases stabilize the DNA double-helix geometry, but other nucleic acid architectures are possible, for example intramolecular duplexes or hybrid DNA-RNA duplexes. These two- or three-dimensional structures are stabilized via the same

hydrogen bonds stabilizing the two complementary hydrogen-bonded DNA nitrogen bases in the classic duplex called a base-pair, bp, Figure 1.
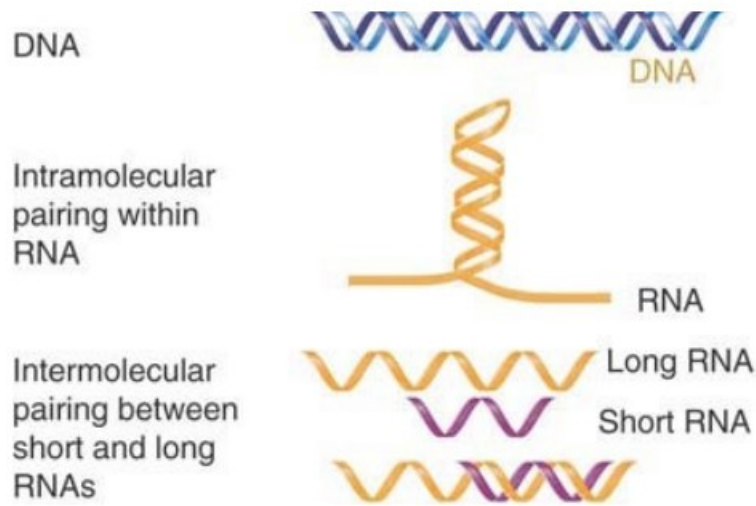
Figure 1. Schematic representation of base-pairing that takes place in: Duplex DNA, single-stranded RNA, and intermolecular interactions between two single-strand RNAs. [1]

Rules dictating the interactions between opposite nitrogen bases of duplex RNA or DNA (i.e., intramolecular interactions) or between nitrogen bases belonging to different polymers (i.e., intermolecular interactions) were investigated in the past seventy years of scientific research pioneered by the discovery of the B-DNA double helix. [2]

Besides the famous B-DNA geometry, additional arrangements, such as A-DNA and Z-DNA, Figure 2, represent a portion of these "alternative" nucleic acid secondary structures.
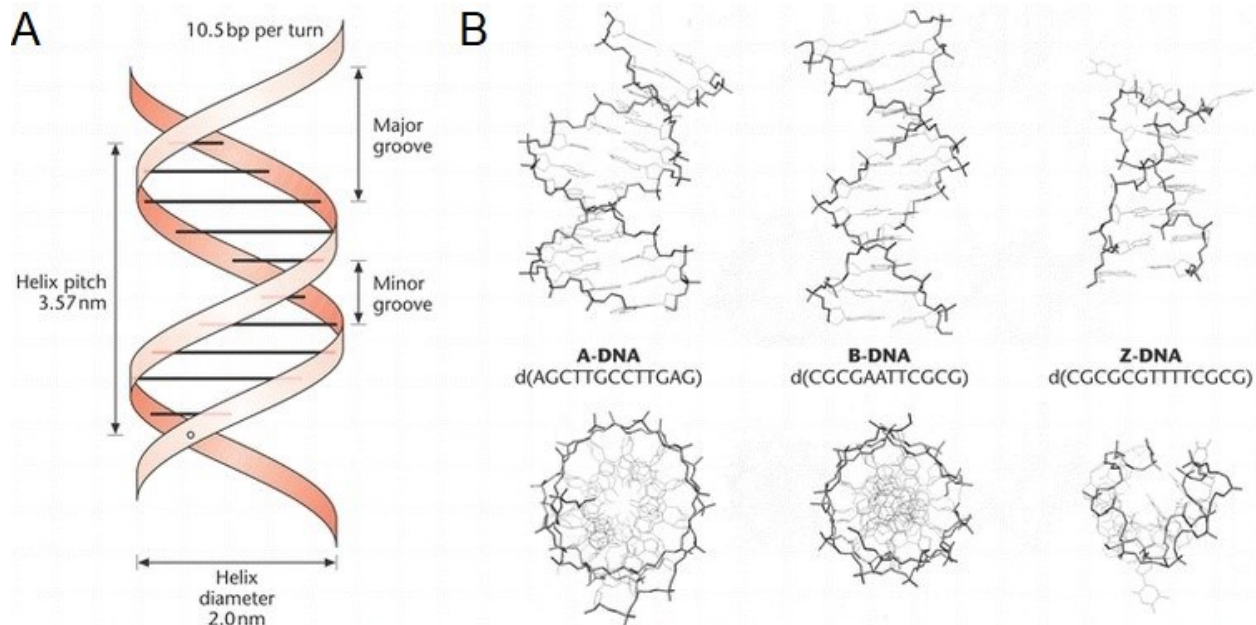
Figure 2. A- Graphical representation of a 12-bp DNA helix and B- models of A-DNA, B-DNA, and Z-DNA. [3]

Similar to B-DNA, A-DNA is a right-handed double-helix, thicker than the B-form and it is detected when DNA is dehydrated. A third type of DNA structure, called Z-DNA is a left-handed helix that is thinner than the B-form. [1]Table 1 reports the geometrical parameters associated with the double helices called A-, B-, and Z-DNA.

Helix parameter/DNA conformation	A-DNA	B-DNA	Z-DNA
Helix sense	Right handed	Right handed	Left handed
Repeating unit/bp	1	1	2
Rotation/bp	33.6°	35.9°	−60°/2
Rise/bp along axis (nm)	0.23	0.33	0.38
Inclination of bp to axis	+19°	−1.2°	−9°
Bp slide (nm)	−0.16	0	−0.11

Table 1. Structural parameters of different DNA geometries (i.e., A-B-Z). [38,39]

Additional secondary structures, such as double-strand hairpins, quadruplexes, and triplexes are dictated by the nucleic acid sequence of both DNA and RNA and by other parameters, such as temperature, ionic strength of the solution, and specific chemical ligands availability. [1]

14

Quadruplexes, as an example, represent secondary structures of nucleic acids where sets of four non-consecutive nitrogen bases interacts via hydrogen bonds to form a tetrad, Figure 3A. These tetrads can comprise four guanines, such as in the G-quadruplex (i.e., G4) DNA and were first identified as a component of telomeres. [4]

An additional example of quadruplex structure is the cytosine-rich structure called i-motif, Figure 3B. [5] This quadruplex is referred to as the complementary sequence of the G-quadruplex and is characterized by hydrogen bonds that form peculiar intercalated geometries, from which its name comes from.
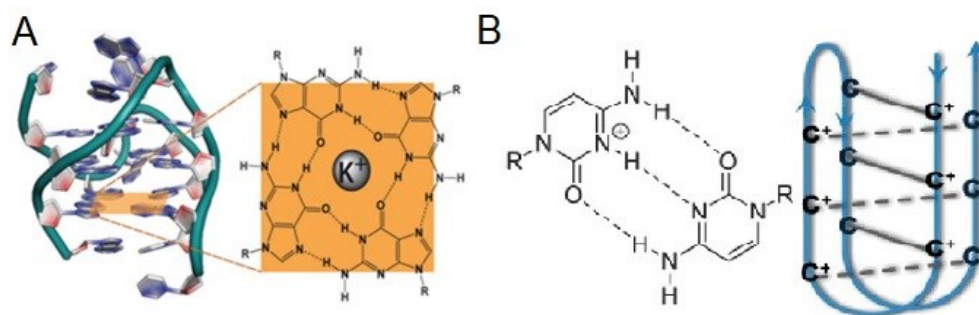
Figure 3. Schematic representation of: A- a G-quadruplex structure (i.e., a G-rich sequence); and B- an i-motif (i.e., a C-rich sequence).

Furthermore, single-stranded DNA or RNA molecules can form aptamers that are molecules with unique secondary structures that allow their highly specific binding to molecular targets, such as proteins and small molecules.

Figure 3 schematically depicts two representative secondary structures of DNA: An intramolecular secondary structure, where paired and unpaired bases form a stem-loop geometry, Figure 4A, commonly referred to as hairpin, [1] and a double helix with a triplex portion, Figure 4B, where the triplets C^+GC and TAT are highlighted.
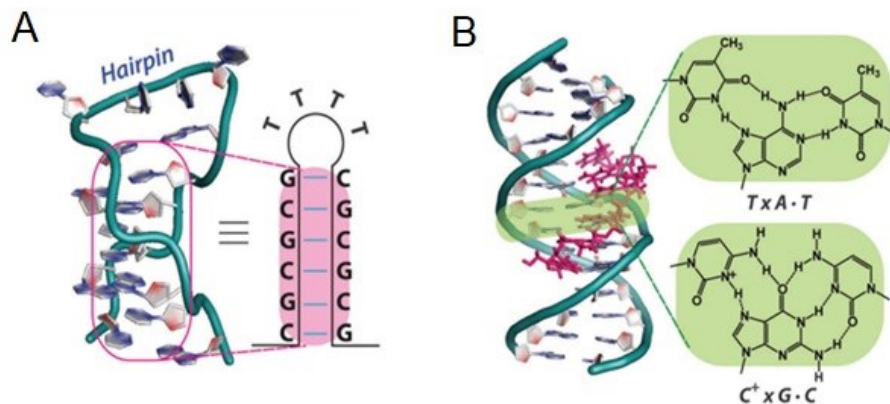
Figure 4. Schematic representation of: A- a hairpin structure with the stem highlighted in pink; and B- a triplex structure which consists of TAT and CGC triads, highlighted in green. Triplex forming oligonucleotide (TFO), bound at the major groove of the DNA duplex, is colored in pink. [4,5]

Notably, a study from 1957 was the first example providing the evidence of the occurrence of triple-helical structures (triplexes) made of DNA. [6] The interactions between a third single-strand nucleic acid molecule, ssDNA or ssRNA, and the double-strand helix, dsDNA, are known as Hoogsteen bonds and are based on different hydrogen interactions with respect to classical Watson-Crick base pairing. [7]

Figure 5A shows a DNA triplex scheme of the Watson-Crick hydrogen bonds and Hoogsteen hydrogen bonds between one purine-rich strand and a third DNA single-strand. The Watson-Crick base pairing together with Hoogsteen interactions stabilize DNA triplexes, one of the distinctive non-canonical DNA conformations. [5] Figure 5B represents the different orientations that the third strand can orient itself in a triplex, depending on triplex composition. Figure 5C is a 3D depiction of the triple helix.
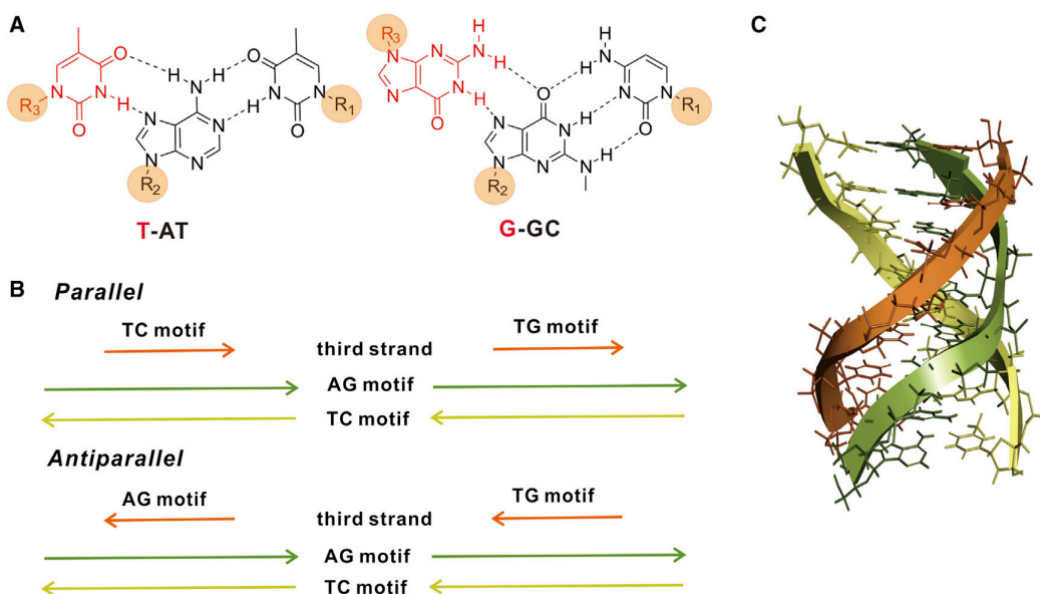
Figure 5. A- Schemes of the Watson-Crick and Hoogsteen three-base interactions in a triplex. B- third strand orientation according to the different motifs. The arrow points in the 5' to 3' direction. C- Three-dimensional representation of the triple-helix. The DNA duplex major groove is where the third-strand RNA (brown) binds. [8]

1.2 FLUOROGENIC APTAMERS

In order to follow the rapidly developing area of RNA biology, additional advances were needed to produce molecular fluorescent probes to observe molecular folding, dynamics, biosynthesis kinetics, and localization. [9]

One important development was the introduction of fluorogenic aptamers. These structures are constituted of nucleic acid sequences that bind to specific fluorescent molecular ligands, producing an aptamer-ligand complex. The complex formation results in a substantial increase of the fluorescence quantum yield of the ligand, which can be measured with a fluorimeter or imaged by fluorescence microscopy. [9]

Fluorogenic aptamers are molecular tools that find their application in studies on RNA interactions, RNA 3D folding research, and research on cellular localization of RNA in cells or tissues.

Numerous fluorogenic RNA aptamers were developed in recent years along with their specific

17

fluorescent ligands. The photophysical and binding characteristics of representative fluorophore-aptamer complexes are shown in Table 2 including their excitation, emission wavelength in nm, extinction coefficient, fluorescence quantum yield (Φ), brightness, and dissociation constant (K_d). [10]

These quantities are compared to green fluorescent protein, GFP, and enhanced green fluorescent protein eGFP.

Flourophore/Aptamer	K_d (nM)	Ex./Em. (nm)	complex Φ^1	Rel. bright.[2]	QY enhan.[3]	Ref.
GFP	/	395/508	0.770	0.60	/	[11]
eGFP	/	490/508	0.680	1.00	/	[11]
DFHBI/Spinach	540	469/501	0.720	0.65	1000X	[12]
DFHBI-1T/Broccoli	560	482/505	0.940	1.10	1000X	[13,14]
TO-1/Mango	3	510/535	0.140	0.40	500X	[15]
DFHO/Corn	70	505/545	0.250	0.27	250X	[16]
Malachite Green/MGA	117	630/650	0.187	1.05	2000X	[17]

Table 2. Representative fluorescent dye/RNA fluorogenic aptamer pairs along with their properties. [1]Fluorescence quantum yield of the complex; [2]Relative brigthness is calculated as (ε x Φ complex)/1000, where ε is the extinction coefficient of the complex. MGA stands for malachite green aptamer. [3]Quantum yield enhacement of the complex in respect to the dye alone.

Figure 6A reports a three-dimensional structure of the aptamer-ligand complex comprising Spinach fluorogenic aptamer and the fluorescent dye (5Z)-5-[(3,5-difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-Imidazol-4-one (DFHBI). Figure 7B contains the chemical structure of the dye.
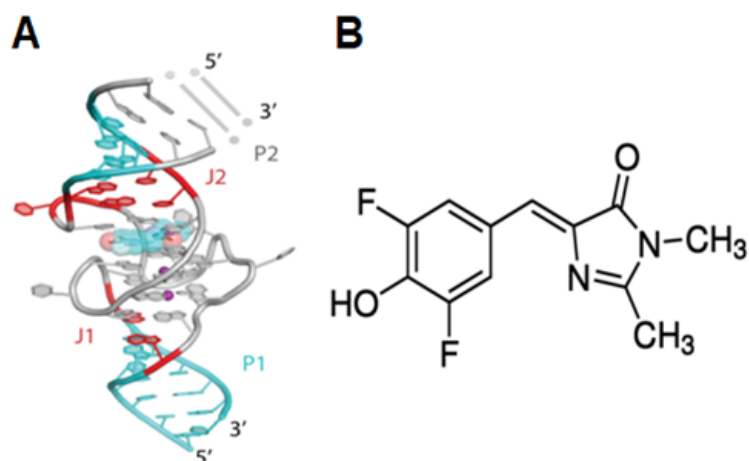
Figure 6. A- Schematic representation of the fluorogenic RNA aptamer Spinach structure linked to its associated dye DFHBI. B- chemical structure of DFHBI fluorescent dye. [9]

One early example of the application of fluorogenic aptamers for metabolite-sensing was the engineering of the fluorogenic Spinach aptamer to produce a structure stabilized by the presence of a molecular target, Figure 7. [18] In order to function as a molecular sensing device, a second aptamer sequence was added to Spinach, allowing the ligand binding detection by fluorescence emission. In the absence of the metabolite target, the fluorescent complex cannot form, resulting in a Spinach molecule that is unable to bind to the fluorophore (DFHBI), Figure 7A. Conversely, the target metabolite stabilizes the three-dimensional structure of the Spinach sequence, which can bind DFHBI, Figure 7B. [18] The resulting binary complex was able to bind DHFBI and thus produce a functional ternary complex target-fluorescent ligand-aptamer complex only in the presence of the target, Figure 7C.
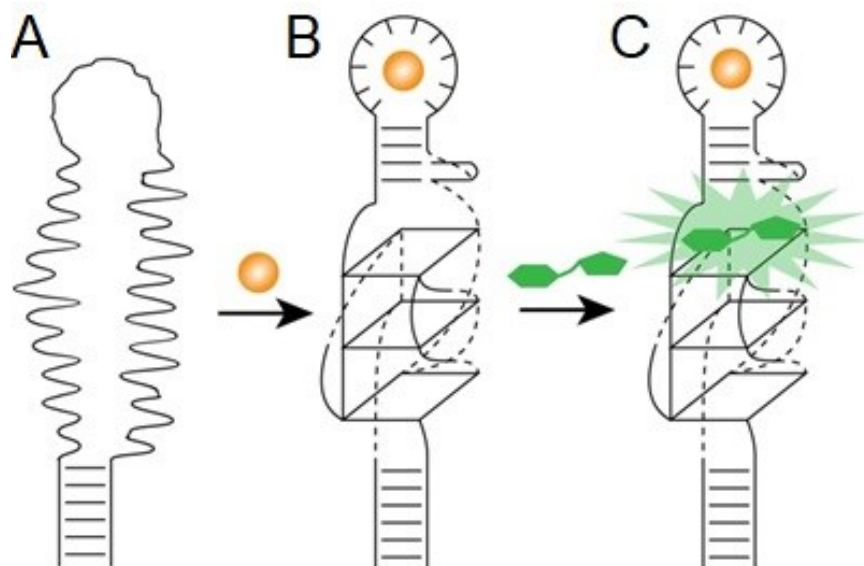
Figure 7. The engineered allosteric light-up aptamer. A- aptamer in its misfolded state cannot bind the dye. B- once the metabolite (orange sphere) binds to the metabolite-binding portion of the aptamer, it promotes and stabilizes the formation of the fluorogenic aptamer portion. C- the presence of the stabilized fluorogenic aptamer structure allows the fluorescent dye (green) to bind to the modified light-up aptamer, resulting in a high-quantum yield emitting fluorescent complex. [18]

Broccoli is a 49-nucleotide long RNA aptamer, which is substantially shorter than the Spinach sequence (84-nucleotide-long) and is able to activate a bright green fluorescence emission upon binding to either DFHBI or its modified version, DFHBI-1T, (5*Z*)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2-methyl-3-(2,2,2-trifluoroethyl)-4*H*-imidazol-4-one, where a methyl group linked to the imidazoline ring was replaced by a trifluoromethyl group, Figure 8. Broccoli aptamer presents improved *in vitro* characteristics and higher folding efficiency compared to the Spinach aptamer. Specifically, Broccoli is significantly less dependent on magnesium concentration for folding and has a higher thermostability. [14]

It is important to note that these aptamers need a high ionic strength to be stabilized, provided for example by high concentrations of sodium (Na^+) or magnesium (Mg^{2+}). Nevertheless, alternative strategies to increase complex stability at lower ionic strength were developed, for example by fusing the aptamer sequence to a longer tRNA. [18]
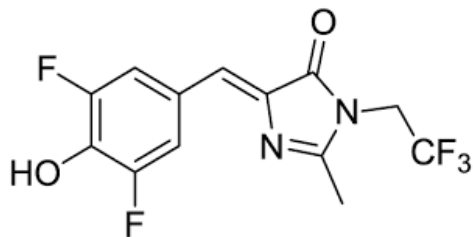
Figure 8. Chemical structure of DFHBI-1T.

1.3 DNA AND RNA IN BACTERIA

In bacteria, the genetic information is encoded in a DNA macromolecule, which commonly consists of a single circular, double-strand structure (Figure 9). This molecule constitutes the bacterial genome and its replication is required for the inheritance of genetically defined features. Additional circular double-strand DNA molecules, called plasmids, are commonly found in bacteria and they carry unique features, such as antimicrobial resistance, which are beneficial for the survival of the organism in a specific environment. [19]

The process by which the genetic information is used to produce a protein and how this process is regulated is known as gene expression. This phenomenon comprises two fundamental steps: DNA is first converted into RNA via a process called transcription, where the multi-subunit enzyme RNA polymerase (RNAp) catalyzes the polymerization of an RNA single strand starting at a regulatory sequence called promoter. Here, RNAp uses soluble ribonucleosides triphosphate (nucleotides carrying three phosphate groups) found in the cytosol to produce an RNA strand using one of the DNA strands of the duplex as template. The promoter sequence is a short region of DNA (20–100 base pairs) representing the starting position of the transcription process. This process is rapidly followed by conversion of the newly produced RNA into a protein by the ribosome via a process called translation. [19]
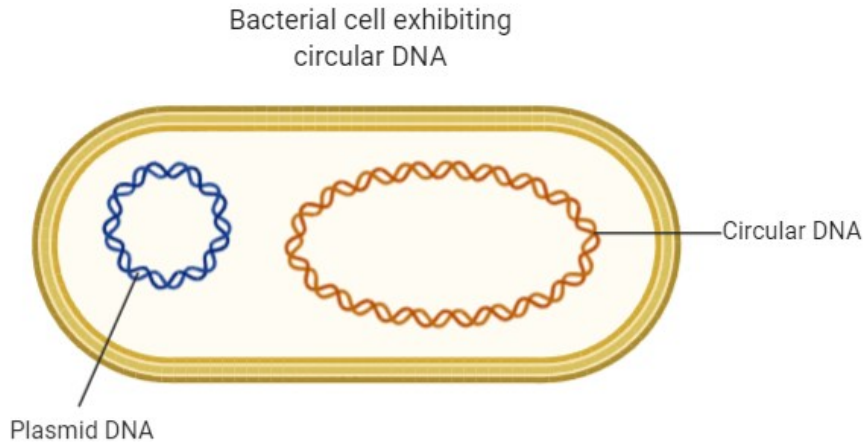
Figure 9. Schematic representation of a circular DNA genome and a DNA plasmid in bacteria.

1.4 TRANSCRIPTION AND GENE REGULATION IN BACTERIA

The first step of gene expression is the transcription process involving the synthesis of new messenger RNA (mRNA) sequences by RNA polymerase enzyme (3' end to 5' end direction) using one of the strands of dsDNA sequence, as only one of the strands of the double helix DNA is transcribed into mRNA. Messenger RNA is then further processed by ribosomes during the translation process using three nucleotide length recognition sites (codons) to build amino acids sequences leading to the final proteins. [20] Not all RNAs are translated into proteins. In fact, according to the simplest description, genes can be divided into two main categories: those encoding a protein (coding RNAs) and those not encoding proteins (non-coding RNAs). These latter genes are transcribed into special RNA molecules, such as tRNA, rRNA, without being translated into proteins. The sequence-element that separates non-coding RNAs from coding RNAs is the presence at the RNA 5'-end of the ribosome binding site, RBS, a sequence that specifically binds to the ribosome and triggers the start of translation. The absence of the RBS results in a non-coding RNA that can function as a modulator, a functional RNA (e.g., rRNA or tRNA), or as a catalytic RNA (e.g., a ribozyme).

In bacteria, expression levels of individual genes differ with time, depending on environmental conditions (e.g., temperature or pH), food availability, and cell cycle. In this case too, genes can be divided into two general groups: Housekeeping genes are constitutively expressed to maintain

22

the basic operations of the cell in any condition and therefore they are not strongly modulated, while inducible genes are activated in specific conditions [20] and their associated RNA or protein vary in concentration during cell cycle or in different extracellular conditions. Gene expression levels can be analyzed at two levels: Transcription, that is evaluated by the intracellular concentration of RNA; and translation, that is estimated by the intracellular amount of protein associated with a specific coding RNA. The primary regulatory process that can modify the expression of many genes in bacteria is associated with the amount of mRNA biosynthesized during metabolism and cell growth and, therefore, assessing the specific RNA sequence concentration at different times and conditions can provide information about its regulation and, indirectly, its function.

Gene expression regulation is necessary for a cell to function. The primary regulatory process that can modify the expression of most genes in bacteria appears to be recognition of the promoter by a protein called transcription factor (also called sigma factor, σ) that recruits the RNA polymerase, to start the transcription process. [21] These sigma factors are divided into seven groups and there are characterized by different molecular weights, indicated by a number following the sigma symbol (e.g. σ^{70}). In addition, each group recognizes a specific consensus sequence of the promoter.

In *E. coli*, the promoter, the regulatory DNA region, is found at the 5'- untranscribed region of a gene, after this regulatory segment, the DNA template sequence for RNA constitutes the gene, where RNAp starts to catalyze the formation of the DNA-templated RNA strand at position +1. A scheme of the genereal E. coli promoter structure, the gene it regulates, and the important positions -35, -10, and +1 are depicted in Figure 10.
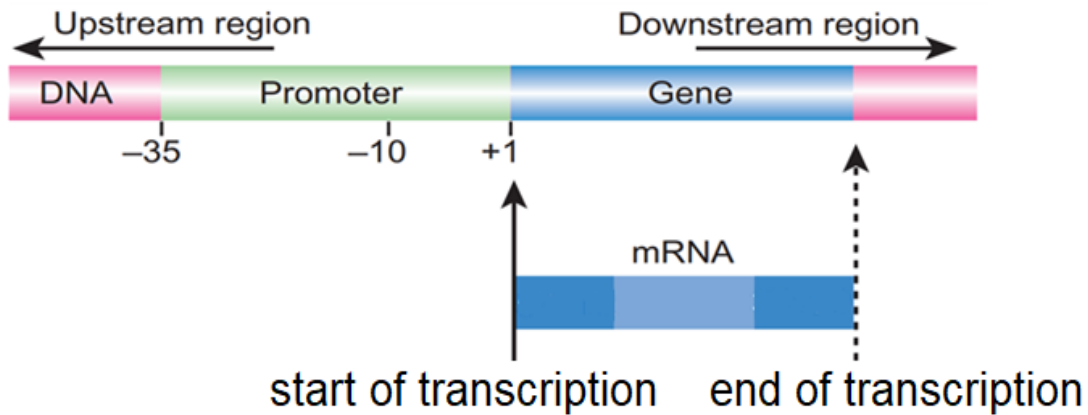
Figure 10. Scheme of a general E. coli transcription unit. Position +1 indicates the starting point of transcription and proceeds downstream until the gene end.

Out of the seven *E. coli* promoter families, six share a common architecture, where two transcription factor-binding sequences are positioned at -35 and -10 bp from transcription starting site +1, Figure 10. In view of this, the scheme shown in Figure 10 represents the general design of transcription unit of E. coli, including the as-described promoter upstream of a DNA template for a specific RNA. It should be noted that since this is a transcription unit, it does not include those sequence-elements required for transcription, such as the RBS element.
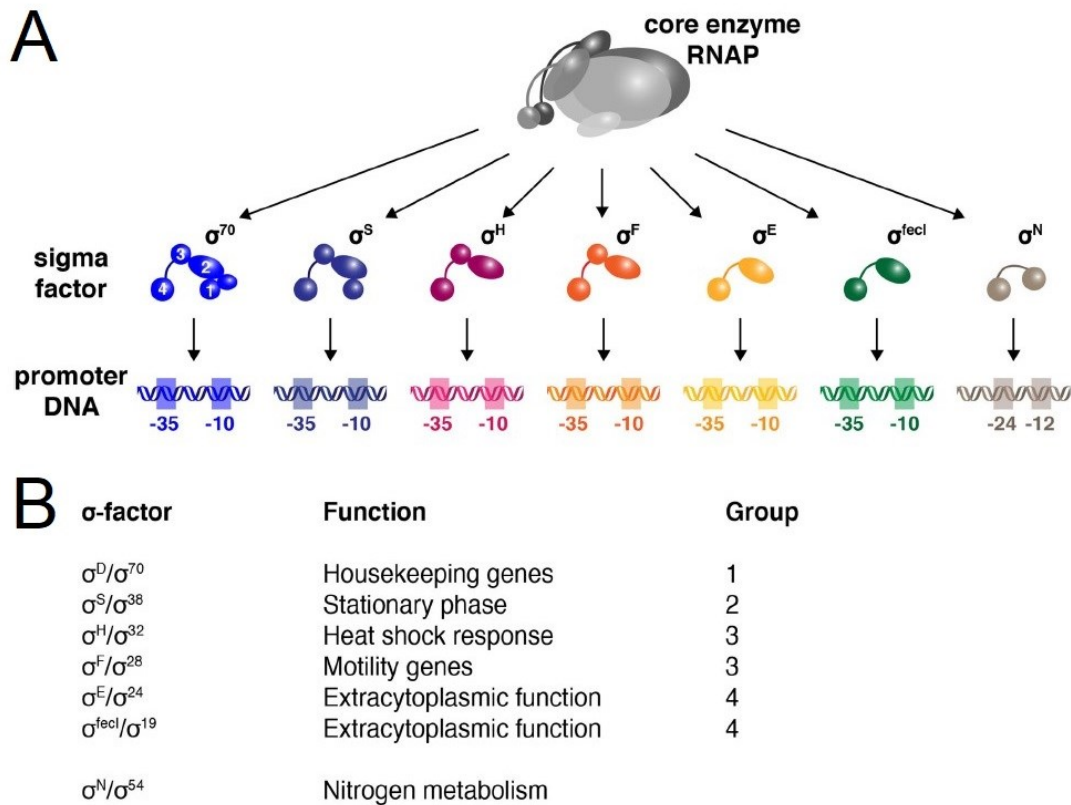
Figure 10. A- E. coli promoters and their transcription factor binding sequences. All promotess present the same binding sequences, except promoter for σ^N. B- E. coli sigma factor classifications. [22]

Commonly, the RNA polymerase and the sigma-factor form a complex called holoenzyme. When the holoenzyme is positioned at the promoter sequence, it induces the unwinding of the double-stranded DNA near the transcription start site. The unwound template strand can then enter into the enzyme active site to create the transcriptionally competent open complex at the 5' end of the RNA transcript. [21]

Bacterial RNA polymerase, the enzymatic complex responsible for DNA-templated RNA biosynthesis, comprises a core enzyme consisting of five subunits: Two α subunits, β, β', and a small ω subunit. The specific interaction of the core enzyme with duplex DNA is mediated by sequence-specific proteins, named sigma (σ) factors. Once RNA polymerase bound to DNA initiated the polymerization of a new strand of RNA, the σ factor is no longer needed, and often detaches from the DNA sequence, leaving the core enzyme. [20]

Anyway, the σ factor subunit, which reversibly interacts with the core of the RNA polymerase complex to create the holoenzyme, is necessary for inducing the process of transcription at the promoter sequence. Figure 11A shows the sigma-factor embedded in the RNAp and interacting with the two -35/-10 consensus sequences (TTGACA, and TATAAT, respectively). Figure 11B shows a scheme of RNAp after release of the sigma-factor and start of transcription.
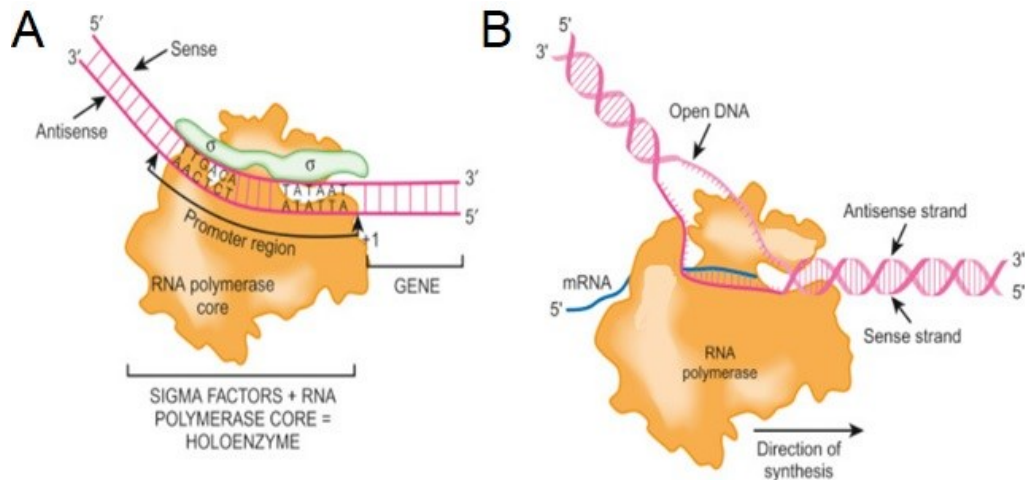
Figure 11. A schematic representation of mRNA production by RNA polymerase. Left: RNA polymerase holoenzyme comprising the sigma factor bound to the promoter region at -10 and -35 positions on the double strand DNA. Right: Synthesis of mRNA by RNA polymerase holoenzyme in the 5' to the 3' end direction of double strand DNA. [20]

Sigma factors recognize specific genomic sequences with high affinity for their binding site, therefore different targets can be studied to construct specific consensus sequences. [23] Examples of promoter sequences recognized by σ70 transcription factor are listed in Table 3 along with their associated -35, -10, and +1-containing consensus.

	-35 element				-10 element		+1
	TTGACA				trTGnT ATAAT	G	A
LacUV5	TTTACA	CTTTATGCTTCCG	GCTCGT	ATAATGT	GTGGAATTGTGAG		
tac	TTGACA	ATTAATCATCCG	GCTCGT	ATAATGT	GTGGAATTGTGAG		
λpR	TTGACT	ATTTTACCTCTG	GCGGTG	ATAATGG	TTGCATGTACTAA		
T7AI	TTGACT	TAAAGTCTAACC	TATAGG	ATACTTA	CAGCCATCGAGAG		
rrbB P1	TTGTCA	GGCCGGAATAA	CTCCCT	ATAATGC	GCCACCACTGACA		

Table 3. Different sigma 70-dependent promoter sequences. In the left column, the names of the genomic sequence are reported. DNA element consensus sequences are displayed in red. In this list, the non-preferred base for the -10 element is a C (shown in blue). [24]

Based on the sequence similarities among proteins of the σ^{70} family, their primary structure has been divided into four regions, see Figure 11. Only the second and fourth regions are largely conserved and contain important sub-regions for the binding to the core RNA polymerase complex, and able to recognize the -10 and -35 positions of the promoters (regions 2.4 and 4.2, respectively). [23]
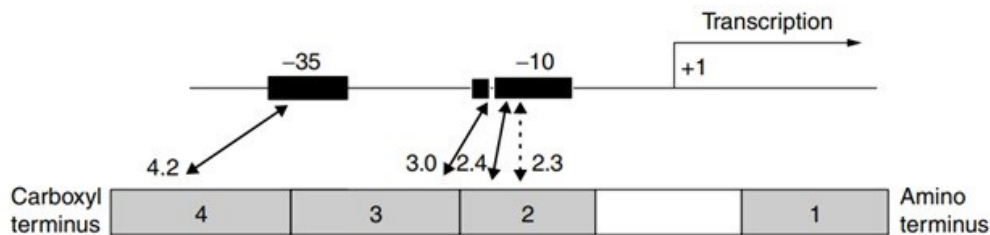
Figure 12. Features of E. coli σ^{70} structure. The interaction with the -10 region of the promoter is mediated by aminoacidic residues from conserved regions 2 and 3. A helix-turn-helix motif in the carboxy-terminal portion of region 4 (subregion 4.2) interacts with the promoter -35 position. [23]

Once RNA polymerase is associated to a σ^{70} factor, it recognizes the promoter and triggers the unwinding of the DNA duplex. In the presence of ribonucleoside triphosphates (i.e., rNTPs), the synthesis of RNA occurs. Then, RNA polymerase continues the transcription process until it comes into contact with a DNA sequence called transcriptional terminator, which weakens the binding with RNA polymerase, leading to the release of the RNA transcript. Thus, RNA polymerase separates from the DNA template to restart the cycle with a new sigma factor.

It is worth to mention that the activity of the bacterial RNA polymerase is influenced by a variety of aspects, including sigma factors, as well as a number of other proteins and ligands controlling RNA synthesis rate or promoter preferences. [21]

As a final remark of this section, the following sections will only mention RNAp instead of mentioning the holoenzyme, unless specified differently. This is due to the fact that most experiments in this work and the cited literature used σ^{70}-saturated RNAp, as this is the commercial holoenzyme.

1.5 SYNTHETIC BIOLOGY APPROACHES FOR GENE REGULATION: GENELETS

Synthetic biology is a recently introduced scientific field of study that focuses on the investigation of the biomolecular mechanisms used by organisms using a set of engineering approaches. One of the aims of synthetic biology is the attempt to artificially produce biological components for controlling cellular events such as the transcription process. In past years, these artificial components were proposed to control genes and to regulate gene expression. [25]

As an example, an artificial DNA design comprising a switchable promoter for controlling a template for RNA polymerase, called "genelet", was developed. The structure is an artificial transcription unit that contains a double-stranded (ds) DNA along with a ssDNA portion localized at the promoter sequence, Figure 13. Sequence-specific ssDNA input A can be introduced to bind to the single-strand portion of the promoter region, resulting in a complete duplex, which can be recognized by RNA polymerase, thus realizing a complete transcription unit that produces the DNA-templated (T-sequence) RNA in the presence of RNAp, Figure 13, top part. [25] In this state, transcription is turned ON. If DNA sequence A is not added to the mixture, the ssDNA portion of the transcription unit generates an incomplete promoter that cannot interact effectively with RNAp resulting in a low-efficiency transcription process. In this state, transcription if turned OFF, Figure 13, bottom part. The ON or OFF state of the system can be monitored by RNA quantification, for example using gel electrophoresis and band intensity comparison.
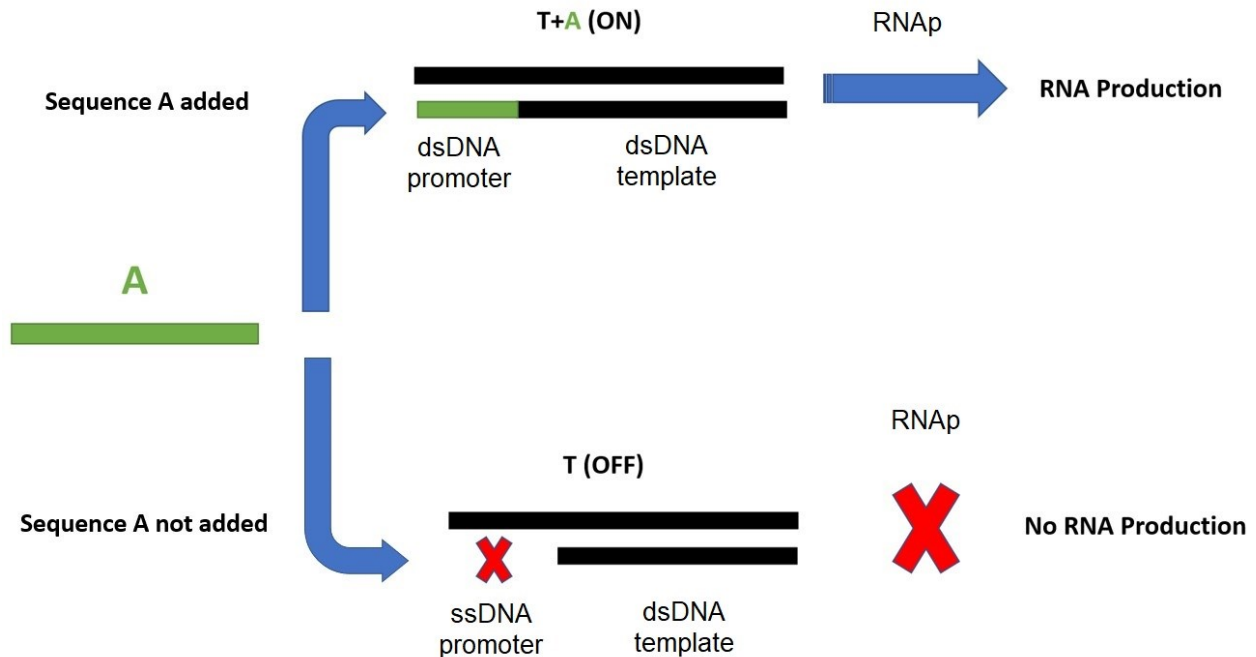
Figure 13. Schematic representation and operation mechanism of a genelet switch. T is the transcription unit comprising the template sequence and the single-strand DNA (ssDNA) promoter sequence. When single-strand DNA sequence A is added, it binds to T, leading to a dsDNA promoter recognized by RNA polymerase, and followed by RNA production, top part, transcription ON. In the absence of A, RNA polymerase cannot recognize the promoter site, and RNA biosynthesis will not occur, bottom part, transcription OFF. [25]

The application of the genelet system *in vivo* has serious limitations, since single-strand DNA portions are highly unstable in a biological environment. This is mainly attributed to the enzymatic degradation of single- or double-strand brakes of duplex DNA by nucleases resulting in the complete degradation of the transcription unit. Alternatively, single-strand portions of DNA are readily recognized as damaged duplexes by the cell biomolecular machinery and repaired, effectively resulting in the loss of the switchable portion of the transcription unit. For these reasons, there is an active search for alternatives to the genelets that could be integrated in cells.

1.6 LONG NON-CODING RNAs

As already mentioned, different categories of RNAs can be found inside cells, these include but are not limited to messenger RNAs (mRNAs) that are the products of DNA transcription and lead to protein production, microRNAs that are non-coding regulatory RNAs usually associated with degradation of specific mRNA targets, and long non-coding RNAs (lncRNAs) that act as regulators of mRNA production via a number of proposed mechanisms.

Generally speaking, non-coding RNA sequences longer than 200 nucleotides are considered lncRNAs. Several examples of the role of lncRNAs on a range of cellular functions were suggested, including their regulation of cell cycle and the formation of ribonucleic protein complexes, which in turn affect gene expression. [26] Some of these mechanisms of lncRNAs are illustrated in Figure 14. Notably, a research carried out using the DNA microarray technique suggested that in a cell lncRNA sequences could be as numerous as protein-coding mRNAs. [27] Although experimental evidence regarding the functions of lncRNAs is increasing, a large picture of their role in cells is still missing.
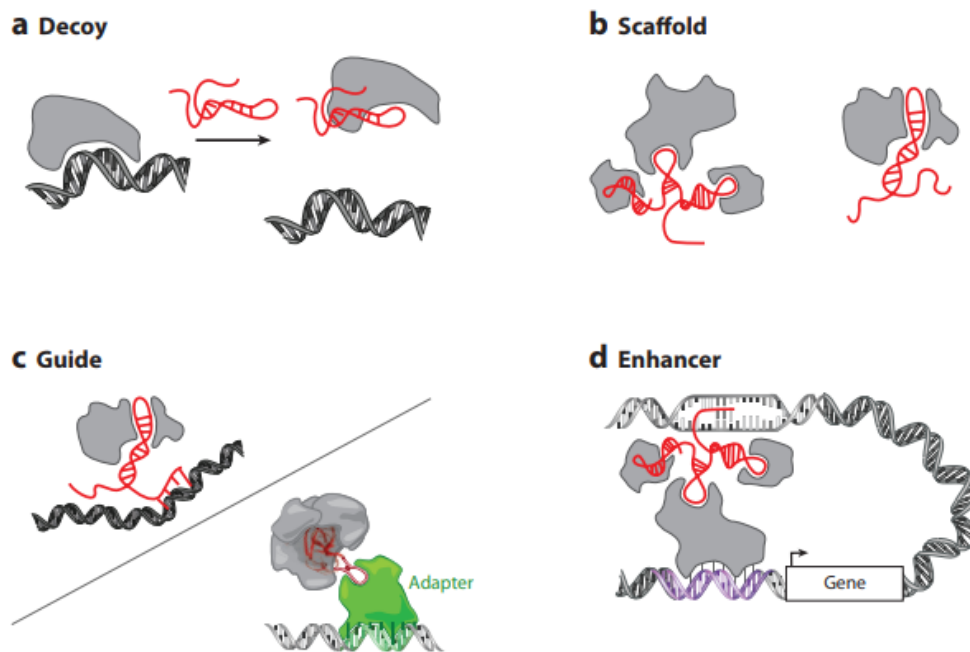

Figure 14. Cellular processes involving long noncoding RNAs (lncRNAs). (A) Competition between DNA and lncRNAs to bind transcription factors. (B) lncRNAs acting as scaffolds for the formation of proteins complexes. (C) lncRNAs as guides for protein interaction, such as enzymes for chromatin modification. (D) lncRNAs acting as enhancer-like systems genome long-distance interactions. [26]

Some literatures suggest that DNA-RNA triplex structures might be involved in the regulatory function of lncRNAs. Indeed, ssRNA can form triple helixes binding to a homopurine sequence (repeated purine sequence) in a duplex DNA via a sequence-specific interaction. These structures were shown to compete with site-specific DNA binding proteins resulting in the inhibition of the protein activity. [28]

1.7 MOLECULAR LOGIC GATES

Logic functions, or Boolean functions, associate two independent variables to a dependent variable where the variables can assume only binary values. This set of values is commonly referred to the set (0, 1), zero and one, but they can be represented as "false" and "true" or "low signal" and "high signal" depending on the context. In addition, alternative names are used, for example the independent and dependent variables are also called inputs and outputs, respectively, while the specific operating function is also called operator or gate. These logic functions can be physically implemented by means of electric signals and are at the basis of conventional electronic circuits.

A specific operator of a logic function, or logic gate, represents the fundamental component of all electronic circuits. It is used to carry out individual logic operations and when placed in parallel and in series with other logic gates, it is used to build electronic circuits operating complex functions. Some of the most common logic gates are AND and OR that associate a pair of inputs, assuming values such as high or low voltages, and generate a single high or low intensity output, that is, a voltage signal, according to the type of logic operator and to the combination of inputs. For each logic operator, a truth table can be produced reporting all input combinations and the corresponding outputs. For example, an AND gate has an output equal to 1 only when both inputs, I_1 and I_2, are equal to 1, and in all other cases the output is 0, as depicted in Figure 15.
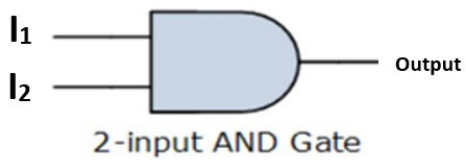
Symbol	Truth Table		
	I_1	I_2	**Output**
I_1 ———⊐ Output I_2 ———⊐ 2-input AND Gate	0	0	0
	0	1	0
	1	0	0
	1	1	1

Figure 15. Schematic representation of an AND gate and its corresponding truth table. [29]

The less well-known XOR logic gate (i.e., exclusive OR gate, sometimes called Ex-OR gate) is a logic gate giving an output equal to 1 (or high) when only 1 of the inputs (I_1, I_2) has a value equal to 1 (or high) and gives an output of 0 when both inputs have the same value (i.e., two zeros or two ones). The truth table of the XOR operator is shown in Figure 16. [30]

Instead of using electronic circuits and voltage or current as inputs and outputs, molecular logics uses quantities associated to chemical reactions. As an example, in molecular logics, high or low concentrations of molecules can be used as input and output signals. In general, rather than high/low voltage or high/low current signals, molecular logics uses substrate concentrations as inputs, reaction rates or chemical yields as outputs, molecular interactions as logic operators, and optical or electrochemical signals as output readouts. Specifically, for enzyme-catalyzed reactions, substrate concentrations are used as inputs, high or low catalytic rates are used as outputs, and time-dependent product quantification is used as readout.
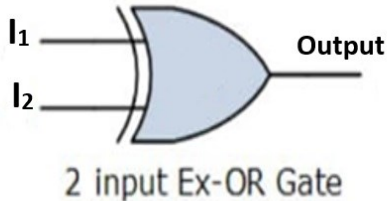
Symbol	Truth Table		
	I_1	I_2	Output
	0	0	0
	0	1	1
I_1 —— Output, I_2 ——, 2 input Ex-OR Gate	1	0	1
	1	1	0

Figure 16. Schematic representation of a XOR logic gate and its corresponding truth table. [30]

It should not come as a surprise that the use of nucleic acids as components of molecular logic systems was already pioneered in 1994 by Adelman with his publication in *Science*, where he demonstrated that a classic computational problem (i.e., the travelling salesman, where the shortest path through a set of cities must be calculated) could be solved using DNA self-assembly reactions. [31]

Molecular logics is a recently introduced field of scientific research investigating the possibility of building logic gates with molecules in a way that resembles electronic components. Indeed, from the point of view of information technology, many biochemical reactions and different biological processes strongly rely on information processing and on complex feedback loops dictated by logic operations. [32] In fact, the multi-parameter nature of bioprocesses (including those involving diseases), taking place in a biological soup of interacting components via simultaneous reactions and feedback loops can represent challenging case-studies for the application of molecular logics. Thus, the goals of molecular logic studies are: i) To understand biological processes from a computational point of view and to apply the acquired knowledge to artificially regulate faulty biomechanisms, or ii) to identify molecular logic gates associated with a specific stage a disease as targets for new therapeutic approaches. In this context, a molecular logic gate or a network of gates can take multiple inputs (e.g., different concentrations of DNA, RNA, proteins, or other biomolecules) and produce programmed outputs, such as the release of a specific drug required to treat the condition related to the set of inputs. [32]

The advantages of using molecules instead of integrated electronic circuits are: (i) the large number of individual gates involved in a reaction in solution results in the high redundancy of computing units. This means that, even if a large fraction of molecules is damaged, the molecular logic unit will still work; (ii) molecular logic gates can be designed to be bio-compatible and then used as triggering mechanisms for the activation of smart drugs or smart delivery systems performing a specific task (output) only when a designed set of inputs is present (i.e. specific inputs are equal to 1, while all the other inputs are equal to 0). [33]

In the scientific literature, molecular logic gates were already demonstrated to have a variety of applications, including molecular calculators performing arithmetic operations, such as additions and subtractions, and were applied for the development of fluorescence sensors for the multiplex detection of metal ions. [32] In comparison to their conventional electronic counterparts, extra features, such as parallel processing, referring to the use of two or more processors to handle different components of a larger task, were demonstrated for molecular logic gates. [34]

1.8 RNA-DNA TRIPLEXES AS TRANSCRIPTION REGULATORS

As mentioned earlier, the formation of hybrid DNA-RNA triplexes was already demonstrated *in vitro* and their formation on a DNA duplex inhibited the formation of a competing dsDNA-protein complex by competition. Furthermore, triplex structures were suggested to mediate the function of lncRNAs toward the regulation of gene expression. Although evidences are still scarce, there is an active community investigating these structures. Recently, hybrid triplexes were proposed to function as inhibitors or enhancers of transcription via a mechanism of competition or distortion of the duplex DNA that activates transcription. Figure 17 shows schematically how triplex structures might produce activation or repression of transcription, via three different mechanisms.
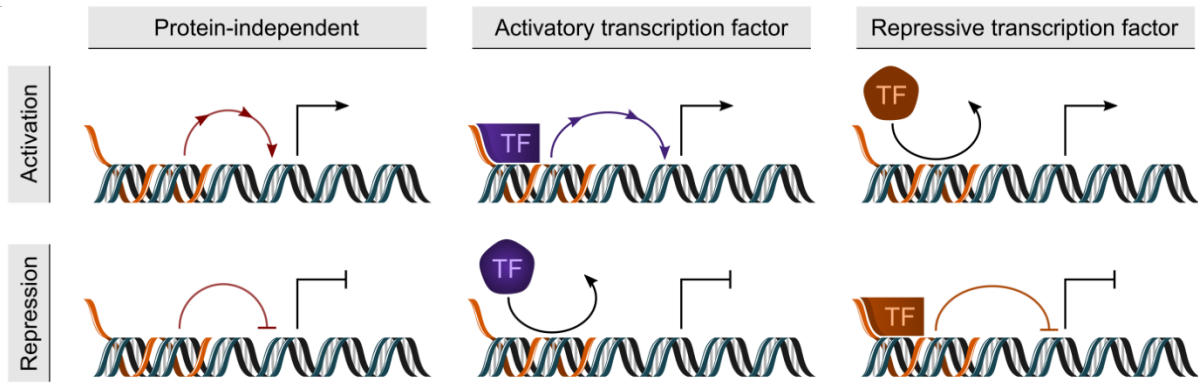
Figure 17. RNA-DNA hybrid triplex mechanisms for transcription regulation. Left- a triplex can act without the aid of protein factors and inhibit or enhance transcription. Center- a triplex can recruit or compete with an activator transcription factor. Right- a triplex can recruit or compete with a repressive transcription factor.

2. MATERIALS AND METHODS

In this section, chemicals and instruments used to perform the experiments, methods used for the production of triplex structures and protocols used to study their effect on transcription are presented.

2.1 MATERIALS

This section lists reagents, enzymes, and nucleic acids used in this study. Unless otherwise specified, all materials were purchased at the highest purity and used without any further purification.

2.1.1 REAGENTS AND PLASTICWARE

DFHBI-1T (i.e., (Z)-4-(3,5-difuoro-4-hydroxybenzylidene)-2-methyl-1-(2,2,2-trifluoroethyl)-1H-imidazol-5(4H)-one) was purchased from Lucerna (Avenue Brooklyn, USA). DFHBI-1T is a fluorescent molecule that binds to Broccoli RNA fluorogenic aptamer. It absorbs light at 482 nm excitation wavelength and emits fluorescence at 505 nm.

In vitro transcription was performed using E. coli RNA polymerase holoenzyme (σ^{70}-saturated RNA polymerase), an rNTP mixture (i.e., ribonucleotide solution mix), containing all the ribonucleotide triphosphates (rATP, rCTP, rGTP and rUTP), a duplex DNA template, and the appropriate buffer. The sigma factor-saturated RNA polymerase enzyme and the ribonucleotide mixture were purchased from New England BioLabs Inc (NEB, Massachusetts, USA). The buffer for the polymerization was produced in the laboratory following a modified receipt of the NEB buffer (*vide infra*).

For gel electrophoresis analysis, acrylamide/bis-acrylamide (19:1) 30% solution, ammonium persulfate, and N,N,N',N'-tetramethylethylenediamine (TEMED) were purchased from Sigma (Germany). The tris-borate-EDTA (TBE) buffer was purchased from Millipore (MA, USA), $MgCl_2$ was purchased from Honeywell Int. Inc. SDS (sodium dodecyl sulphate), glycerol, and bromophenol blue were purchased from Sigma-Aldrich, (USA). The gel stain oxazole gold was purchased from Biotium (USA). Tris hydrochloride (Tris- HCl) was purchased from Thermo Fisher

Scientific Inc (UK), and Triton x-100, potassium chloride (KCl), and 1,4-dithiothreitol (DTT) were purchased from (Sigma-Aldrich®, USA).

In all experiments ultra-pure water (resistivity at least 18.0 MΩ·cm) from a Genie machine manufactured by RephiLe (Florida, USA) was used.

The oligonucleotides used in this study (i.e., RNA and DNA sequences) were purchased as lyophilized materials from Integrated DNA technologies, Inc. (Iowa, USA) and their sequences are reported in Table 4.

Multiwell plates used for *in vitro* RNA polymerization experiments were purchased from Sarstedt Ag & Co. (Germany). The seal used for the multi-well plates was the AB-0580 Adhesive Plate Seals from Thermo scientific (Massachusetts, USA).

Name	Sequence
TFO1	5' - UCC UCC UCU UCU CCU -3'
TFO2	5'- AGG AGA AGA GGA GGA -3'
TU sense	5'- TTG ACA TCC TCT TCT CCT CCT ATA ATA GGA GGA GAA GAG GAG GAA CGA GAC GGT CGG GTC CAG ATA TTC GTA TCT GTC GAG TAG AGT GTG GGC TCG TTC C -3'
TU antisense	5'- GGA ACG AGC CCA CAC TCT ACT CGA CAG ATA CGA ATA TCT GGA CCC GAC CGT CTC GTT CCT CCT CTT CTC CTC CTA TTA TAG GAG GAG AAG AGG ATG TCA A -3'
TTS sense	5'- TCC TCT TCT CCT CCT -3'
TTS antisense	5'- AGG AGG AGA AGA GGA -3'

Table 4. Oligonucleotide names and sequences. DNA is marked in black color while RNA is marked in blue. TU is an artificial transcription unit capable of producing the Broccoli RNA sequence in the presence of RNAp holoenzyme and ribonucleotides.

2.1.2 INSTRUMENTATION

Samples were centrifuged with an Eppendorf centrifuge 5810 R purchased from Eppendorf (Germany). Single strand DNA (ssDNA) sequences were treated using a thermal cycler mod. 2720 purchased by Applied Biosystems®, Germany, to form a double helix. The pH was corrected using a Crison pH-meter Basic 20+ manufactured by Crison Riera Principal (Spain). Electrophoresis experiments were carried out using a Mini-Protean tetra cell from Bio-Rad Laboratories (California, USA). Fluorescence images of gels were acquired using an iBright 1500 from Invitrogen (Thermo Fisher Scientific, USA). The *in vitro* RNA polymerization experiments were conducted using a Viktor-X4 fluorescence emission plate reader by Perkin-Elmer (CA, USA) with excitation/emission filters: 482 nm and 535 nm.

2.2 METHODS

In this section protocols, and procedures to evaluate the effect of triplex structures on the transcription of Broccoli RNA are described.

2.2.1 KINETIC ANALYSIS OF *IN VITRO* RNA POLYMERIZATION

In order to form double-strand DNA TU (Transcription unit), the following process was carried out: Sense and antisense sequences (10 µM) were collected into 0.2 mL PCR tubes and mixed in a solution containing 100 mM KCl. The mixture was then placed into a thermal cycler and a temperature ramp was set from 97°C to 10°C, at a rate of -1°C per minute. The process allowed the formation of double strand DNA (dsDNA) from complementary single strand DNA (ssDNA) with high yield, while reducing unwanted secondary structures. Newly formed duplexes (dsDNA) were stored at -20°C.

Transcription units TU was capable of producing the fluorogenic RNA aptamer Broccoli *in vitro* in the presence of RNA polymerase and a ribonucleotide (rNTPS) solution mix. The buffer solution for *in vitro* RNA polymerization was prepared in advance at 10 X concentration (1X concentration: 150 mM KCl, 40 mM Tris-HCl, 30 mM $MgCl_2$, 1 mM dithiothreitol, and 0.01% v/v Triton x-100) and the pH was corrected to pH 6.9 using concentrated HCl or NaOH solutions. This solution was then stored at 4°C.

In vitro RNA polymerization samples for the plate reader were prepared in a final volume of 25 µL containing: 1X RNA polymerization buffer (described in the previous paragraph), 40 nM dsDNA template (i.e., TU), and different concentrations of ssRNA sequences (TFO1 or TFO2) in the interval 0.1-1.0 µM. Samples were then incubated at 4°C for 30 minutes. This time interval allowed the binding of ssRNA to the dsDNA template to form the hybrid triplex. Following this incubation period, 2 mM DFHBI-1T fluorescent dye, 0.016 (U/ µL) σ^{70} saturated RNA polymerase (holoenzyme), and 4 mM rNTPS solution mix were added to each sample. Samples were then transferred to microplate wells and then a transparent adhesive seal was firmly applied on the microplate top to avoid sample evaporation. Microplates were centrifuged at 400 RPM for 1 min at room temperature and immediately placed into the plate reader for a 5-hour time period. To ensure that all polymerization experiments were subjected to the same conditions, only fluorescence data collected from the same microplate experiment were used for the rate-comparison analyses.

Data collections were performed using the following parameter set: 10-second signal accumulation time, 50000 V lamp energy, and temperature set at 29 °C.

Data were processed using Origin software in the following way: First, time-dependent fluorescence emission values were linearly fitted in a period time of 2 hours discarding the first 60 minutes of the 5-hour data collection. Then, slope values (first-order kinetic rates) were used as estimates of the polymerization rates of Broccoli RNA and normalized to allow the comparison of different experiments.

2.2.2 DOSE-RESPONSE STUDIES OF BROCCOLI POLYMERIZATION CONTROLLED BY RNA-DNA HYBRID TRIPLEXES

To quantify the effect of the TFO, and ultimately the effect of the triplex formation, on RNA polymerization, data collected from the *in vitro* polymerization experiments were processed using Origin software in the following way: First, time-dependent fluorescence emission values were linearly fitted in a time period of 1 hour between the second and the third hour of the 5-hour data collection. Then, slope values (first-order kinetics) were used to calculate the estimates of the RNA polymerization rates and normalized to allow comparison of sample-associated rate estimates. The resulting plots were fitted with the following dose-response equation (eq. 1) to

calculate maximal effect (i.e., right asymptote of the resulting sigmoid curve) and EC_{50} (i.e., the TFO concentration at which half-maximal effect is observed, i.e., the TFO concentration corresponding to the symmetry/inflection point of the dose-response sigmoid function).

(eq. 1)
$$y = A1 + \frac{A2-A1}{1+10^{(logx_0-x)p}}$$

Where x is the TFO concentration, y is the fluorescence-rate (first-order kinetics), A1 and A2 are the left and right asymptotes, respectively, x_0 is the TFO concentration at the inflection point (i.e., E_{50}), and p is the slope of the sigmoid at the inflection point. Figure 18 shows a representative plot of eq. 1 where the relevant coordinates are marked.
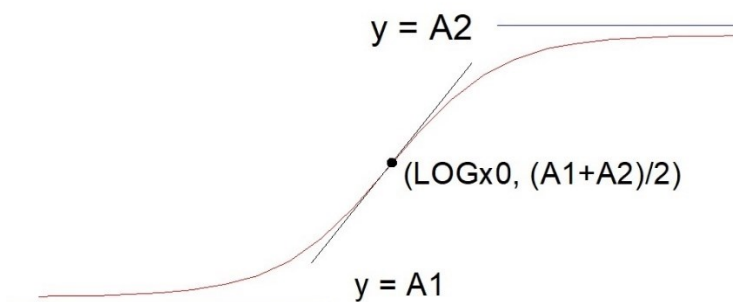
Figure 18. Representative general plot of eq. 1. A1 and A2 are the y coordinates of the left and right asymptotes, respectively. (LOGx0, (A1+A2)/2) are the x, y coordinates of the symmetry/inflection point of the sigmoid.

2.2.3   TRIPLEX-BASED XOR GATE DESIGN

Engineered transcription units were designed based on the σ70 promoter, containing the -35 and -10 consensus, two triplex target-sequences (Transcription Target Sequences, TTS), and the template for Broccoli fluorogenic RNA aptamer. The transcription unit can form enhancement or triplexes in the presence of a TFO available to form a triplex. In the present study, two 15 base-long RNA TFO sequences, were called TFO1 and TFO2, both generating enhancement triplexes. They were designed to be complementary to each other and to specifically bind independently to their target site (i.e., the two target sites were positioned at different locations on the transcription unit to avoid competitive binding).

40

To encode the XOR gate mechanism into molecular interactions, the two TFO RNA sequences were considered as inputs, and the Broccoli RNA polymerization rate was considered as the output. In this system, the transcription unit represents the molecular analogue of a XOR gate, where the polymerization reaction catalyzed by the template-dependent RNA polymerase is positively perturbed (i.e., the rate increases) by the presence of one of the two TFO (i.e., the TFO promotes triplex formation resulting in a modification of the polymerization rate), but no effect con be detected when no TFO or both TFOs were added to the mixture in equimolar concentrations (i.e., the absence of TFOs cannot promote triplex formation or the two TFOs form a duplex that cannot participate to the hybrid triplex structure, resulting in an unperturbed polymerization rate). As already described, since the direct analysis of RNA polymerization rate is difficult, the time-dependent fluorescence emission from Broccoli RNA was used to estimate the RNA polymerization rate. Therefore, in this study the fluorescence rate was used as readout of the molecular gate.

The molecular implementation of the logic gate was represented using the truth table of the XOR operator (see Figure 16), where TFO1 is the input 1 and TFO2 is the input 2, and their values (0 or 1) are encoded by their concentrations within the experimental parameters. In order to decide what "low" or "high" concentration refer to, an arbitrary threshold was chosen. As already described, the rate of broccoli polymerization, measured by the time-dependent fluorescence of Broccoli, was encoded as output (i.e., the readout). Similar to the inputs, to attribute a digital value to the output, an arbitrary threshold was chosen for the polymerization rate, too.

2.2.4 ELECTROPHORETIC CHARACTERIZATION OF THE XOR GATE

In order to quantify the yield of the triplex structures comprising TFO1 or TFO2 sequences and the transcription unit, an electrophoretic mobility shift assay (EMSA) was carried out. The technique is based upon gel electrophoretic and evaluates how molecules migrate differently in different conditions. Since hybrid triplexes migrate at a different rate compared to their constituent single strands and duplexes, this allows the discrimination of the different species by analyzing the gel bands. Moreover, to increase the resolution of the electrophoretic separation,

the TU was reduced to the part comprising the two triplex target sites only, and for this reason was called TTS.

The following protocol was used to prepare mixtures containing different concentrations of TFO1 and TFO2, in the presence of a fixed concentration of TTS: Samples were prepared in a buffer solution containing 40 mM Tris-HCl, 30 mM $MgCl_2$, 1 mM DTT, and 0.01% Triton X-100 at pH 6.9. A series of solutions were prepared to contain two opposite gradients of concentrations of TFO1 and TFO2 (from 1 nM to 10 µM) in the presence of 200 nM dsTTS. TFO1/TFO2 ratios were: 1:10000, 1:1000, 1:100, 1:10; 1:1, 10:1, 100:1, 1000:1, and 10000:1.

The gel was loaded on a Mini-PROTEAN Tetra System connected to a Power Ease Touch 350W power supply. The composition of a typical 10 mL volume electrophoresis gel was 12% acrylamide/bis-acrylamide (19:1), 1x TBE buffer at pH 6.9, 10 mM $MgCl_2$, 5 mg ammonium persulfate, and 5 µL N,N,N',N'-tetramethylethylenediamine (TEMED). The running buffer was 0.5x TBE supplemented with 10 mM $MgCl_2$ at pH 6.9. Each well was filled with 13 µL of sample (i.e., 10 µL of the RNA/DNA mixture and 3 µL of loading dye, which contained glycerol and bromophenol blue). The electrophoresis run was carried out at 75 V for 2 hours in an ice bath.

At the end of the 2-hour time period, the gel was incubated for 10 minutes in a freshly prepared staining solution containing oxazole gold 1X under gentle agitation. After that, the gel was collected from the staining solution and fluorescence images were acquired using an iBright1500 imaging system. All the data were processed using Origin data analysis software (Origin Lab Corporation, MA, USA).

2.2.5 KINETIC ANALYSIS OF THE XOR GATE

In order to confirm the correct molecular implementation of the XOR gating mechanism, the effect of the different TFO1/TFO1 ratios described in the previous section was analyzed during *in vitro* RNA polymerization. The formation of a triplex structure comprising TFO1 or TFO2 controls RNA transcription, determining a rate enhancement, while the absence of TFOs or the presence of both does not affect transcription, as detailed by the truth table of the XOR gate in Figure 16. Similar to samples described in chapter 2.2.4 for the EMSA analysis, new samples were prepared for the kinetic analysis. In this case, the TTS duplex was replaced by the actual TU duplex for of the synthesis of Broccoli RNA in the presence of different concentrations of TFO1 and TFO2.

Experiments in the presence of TU2 dsDNA and ssRNA TFO1/TFO2 ratio mixtures were conducted using the following protocol: Samples were prepared in a final volume of 25 µL containing 1X RNA polymerization buffer, 40 nM dsDNA template (i.e., TU2), and different concentrations of TFO1 and TFO2 in the 1 nM -10 µM concentration range. Nine samples were prepared containing TFO1/TFO2 ratios as described for the EMSA analysis in the previous section.

Samples were then incubated at 4°C for 30 minutes. This time interval allowed the binding of the ssRNA to the dsDNA TU and the formation of the triplex hybrid. Following this incubation period, 2mM DFHBI-1T fluorescent dye, 0.016 (U/ µL) σ^{70} saturated RNA polymerase (holoenzyme), and 4 mM rNTPS mix solution were added to each sample. Samples were then transferred to the microplate wells and then a transparent adhesive film plate seal was firmly placed on the top to avoid sample evaporation. Microplates were centrifuged at 400 RPM for 1 min at room temperature and immediately placed into a Viktor-X4 fluorescence emission plate reader for a 5-hour time period for data collection. To ensure that all polymerization experiments were subjected to the same conditions, only fluorescence data collected from the same microplate experiment were used in the rate-comparison analyses. Data were processed using Origin software in the following way: First, time-dependent fluorescence emission values were linearly fitted in a period time of 2 hours between the second and the third hour of the 5-hour data collection. Then, the slope values (first-order kinetics) were used to estimate the RNA polymerization rates and normalized to allow the comparison across different samples. In order to demonstrate the effective implementation of the molecular XOR gate, rates associated with no added TFO, equimolar concentrations of TFO1 and TFO2, or samples containing only one the two TFO (i.e., inputs 0,0; 1,1; or 0,1 and 1,0) were plotted in a bar graph and an arbitrary threshold was applied to show the respective outputs 0, 0, 1, and 1, respectively.

3. RESULTS

In this chapter, results of the experiments described in chapter 2 (i.e., kinetic analysis of Broccoli polymerization from engineered promoters and in the presence of different triplex-forming systems, characterization of the triplex structures, and characterization of the molecular XOR gate) are reported. Additional details emerging from the triplex-modulated transcription rate analysis as described in the Methods section are shown.

3.1 KINETIC ANALYSIS OF BROCCOLI IN VITRO TRANSCRIPTION

Fluorescence emission over time, arising from the *in vitro* synthesis of Broccoli RNA in the presence of an artificial transcription unit TU containing an engineered promoter is shown in Figure 19. The detailed protocol followed to obtain this data is described in section 2.2.1. The fluorescence signal increases over time indicating that a ssRNA (Broccoli) was synthesized and correctly formed a complex with DFHBI-1T. The complex has a quantum yield 1000-fold higher than the dye alone, producing a fluorescence emission dependent on Broccoli RNA concentration. The kinetic analysis showed the successful design of the artificial transcription unit (TU) that is able to produce the correct RNA in the presence of RNAp and ribonucleotides. To compare linear rates, a linear fitting was operated in a time interval of two hours, discarding the first 60 minutes of the reaction. Figure 19 shows with a grey box this time interval.
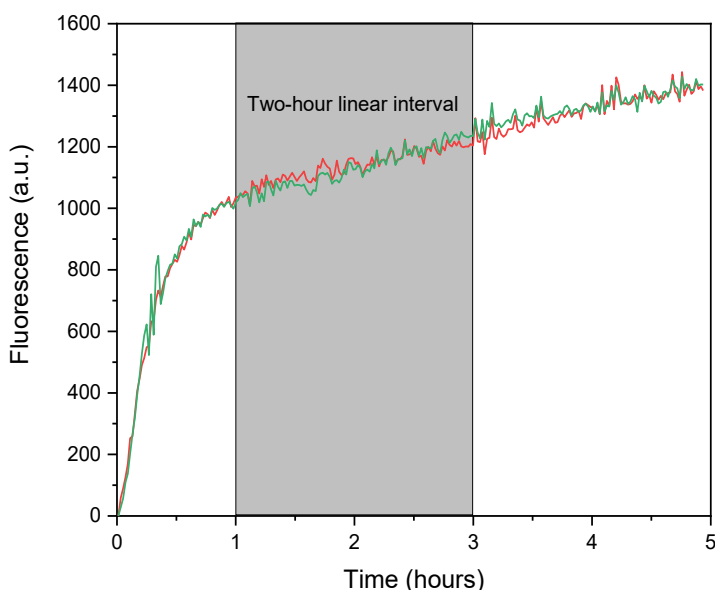
Figure 19. Fluorescence emission changes from replicas of samples for *in vitro* polymerization of Broccoli RNA from transcription unit TU. In grey color, the time interval used for the determination of the linear rate of RNA synthesis.

3.2 KINETIC ANALYSIS IN THE PRESENCE OF TFO1 OR TFO2

In order to analyze the effects of the different TFOs, and of their associated triplex structures, on RNA synthesis, rates of polymerization from engineered the promoter were estimated using the method detailed in chapter 2.2.2. Average values and standard deviations were calculated using data from three independent experiments.

The effect of TFO1 and its associated triplex structure on RNA synthesis from TU transcription unit was assessed with experiments described in the method section (see section 2.2.2) while data were processed using Origin data analysis software to extract the maximal effect and EC_{50} values. The mean values of three experiments were used to calculate the standard deviation.

Figure 20 shows representative fluorescence changes of samples containing a constant TU concentration and different TFO1 concentrations (0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, and 2.00 μM). Indeed, at higher TFO concentrations, the enhancement of the transcription rate becomes clearly visible.
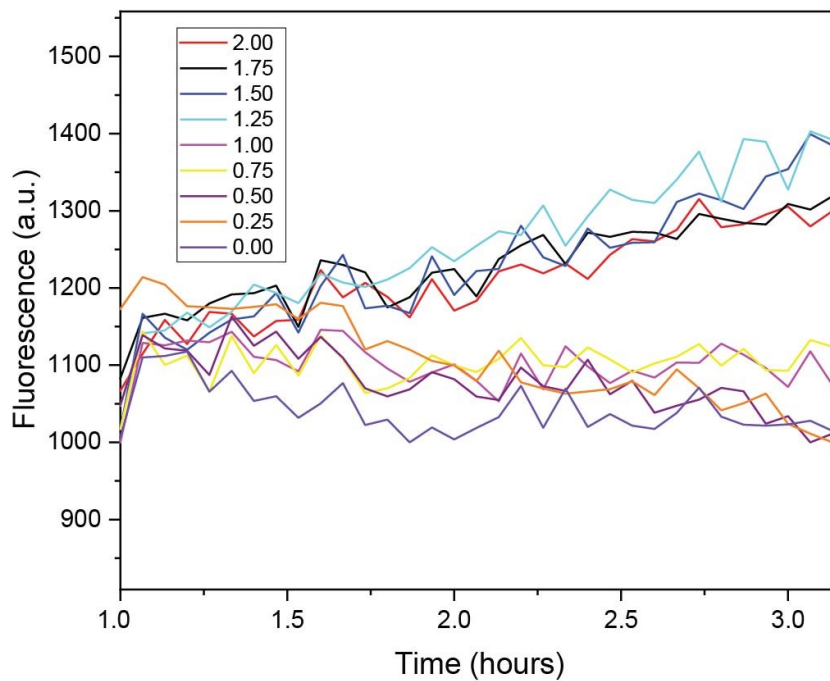
Figure 20. Fluorescence changes during in vitro RNA polymerization in the presence of TU and increasing concentrations of TFO1. Inset: TFO1 concentrations in micro molarity.

To build a curve that could be fitted with eq. 1, the interval of TFO concentrations was extended to the range 1 nM – 10 μM. The *in vitro* RNA synthesis experiments were performed according to the protocol described in the methods section. Representative values of the estimated rates of RNA polymerization plotted against TFO1 RNA concentrations are reported in Figure 21. This series of experiments showed an increase of the transcription rate of RNA in the presence of increasing concentrations of the TFO1 RNA sequence. On average, the calculated EC_{50} was 350 ± 20 nM and maximal enhancement effect was +210 ± 20%.
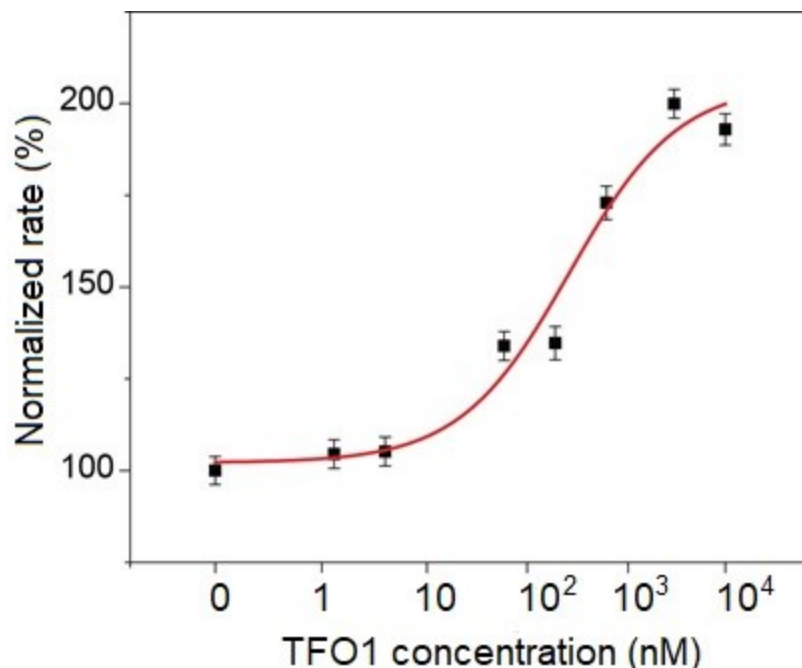
Figure 21. RNA polymerization rates are estimated using Broccoli RNA fluorescence (black squares). The effect of increasing concentrations of TFO1 RNA sequence were estimated using eq. 1 fitting (red line) to calculate EC_{50} and maximal effect (right asymptote). Bars indicate standard deviation.

In the second step, TFO2 was tested against the same TU. Figure 22 shows representative rates plotted against TFO2 concentrations. Since TFO1 and TFO2 will be used as inputs for the XOR gate, the expected effect was similar to the effect of TFO1 on RNA synthesis rates. Indeed, the plot shows a similar eq. 1 fitting as for TFO1 (Figure 21), with EC_{50}= 1.0 ±0.1 μM and maximal effect +177 ±20%.

Figure 22. TFO2 effect on RNA in vitro polymerization using TU dsDNA. Black squares represent polymerization rates, bars represent standard deviation, and the red line represents the fitting according to eq. 1.
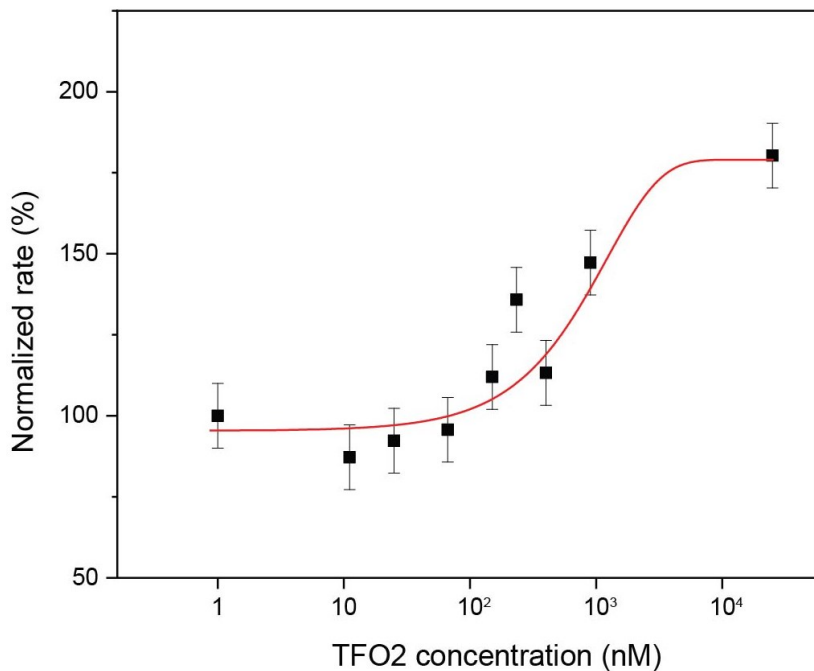
3.3 OUTPUT ANALYSIS OF THE TRIPLEX-BASED XOR BIOMOLECULAR GATE

The polymerization rate-outputs of the proposed XOR gate are reported in Figure 23. To build this plot, results obtained from *in vitro* RNA polymerization experiments were obtained using four specific conditions. The first condition, inputs (0, 0), was realized assembling a system that contained all essential components for transcription (i.e., TU, RNAp, and ribonucleotides) but no TFOs. For the second and third conditions, inputs (0, 1) and (1, 0), the mixture contained either one of the two TFOs in addition to all essential components for transcription. In the fourth condition, TFO1 and TFO2 were added to the mixture in equimolar concentrations (i.e., 500 nM). In this case, the two RNA strands form a Watson-Crick RNA duplex due to their complementary sequences and therefore cannot participate in triplex formation.

Data from the *in vitro* RNA polymerization were linearly fitted as outlined in the method section and rate average values were represented as bars. The four conditions reflect the four input combinations (0,0), (0,1), (1,0), and (1,1) and are reported as labels of the y-axis. A threshold

value for the normalized rates was also set at ca. 40% of the highest average rate associated to input set (1,0). An output value of 1 corresponded to average rate values higher than this threshold, while an output value of 0 corresponded to average rate values lower than the threshold.
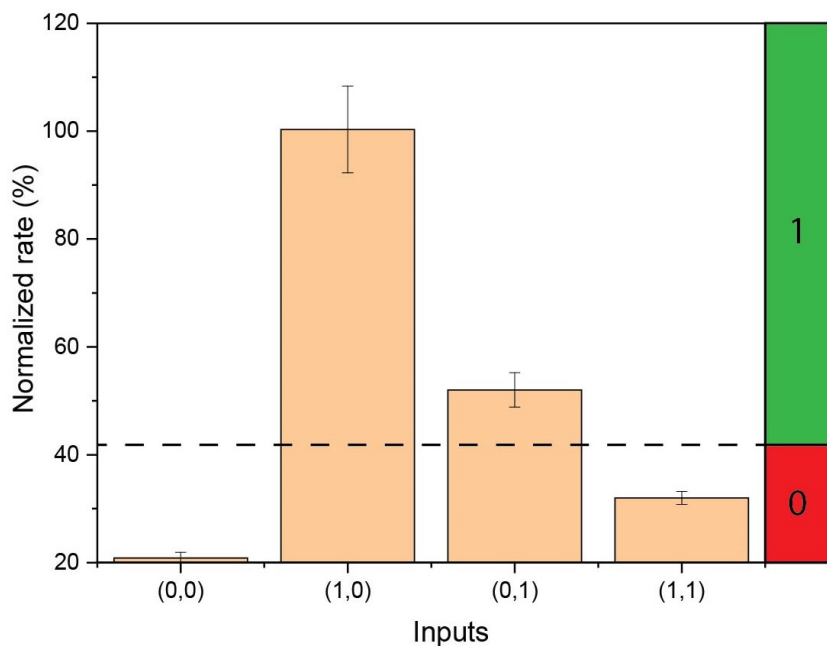

Figure 23. Rate outputs of the proposed XOR gate. Input set (0,0) indicates the *in vitro* RNA polymerization reaction mixture did not contain any TFO RNA; input set (1,0) indicates the reaction mixture contained 500nM TFO1; input set (0,1) indicates the reaction mixture contained 500nM TFO2; and input set (1,1) indicates the reaction mixture contained equimolar concentrations of TFO1 and TFO2 (i.e., 500 nM).

3.4 ELECTROPHORETIC CHARACTERIZATION OF THE TRIPLEX FORMATION

To estimate the amount of DNA:DNA:RNA triplex formed in solution, EMSA experiments were carried out, as described in the methods section. The electrophoretic run, reported in Figure 24, shows individual bands associated with ssRNA (blue arrow), dsDNA (yellow arrow), and the hybrid triplex comprising ssRNA and dsDNA (green arrow). The slowest band was attributed to excess RNA TFO and was not detected with high concentrations of TFO1 (lanes A, B, C, D, and E), while at high concentrations of TFO2 an intense band appeared (lanes F, G, H, and I). The fastest running band (i.e., the lowest band in the gel) was attributed to dsDNA TTS due to its smaller

hydrodynamic volume, and consistent with recent literature. In addition, this band became weaker at higher TFO concentrations, especially in the presence of TFO2 (lanes H and I). The middle band in the gel was attributed to the hybrid triplex structure, due to its larger hydrodynamic volume compared to dsDNA TTS, and it intensified (i.e., the band become darker) at higher TFO concentrations (i.e., rightmost, or leftmost side of the gel). The smear of lanes H and I is the result of the overloading (i.e., saturation) of the gel pores in the presence of excess RNA.

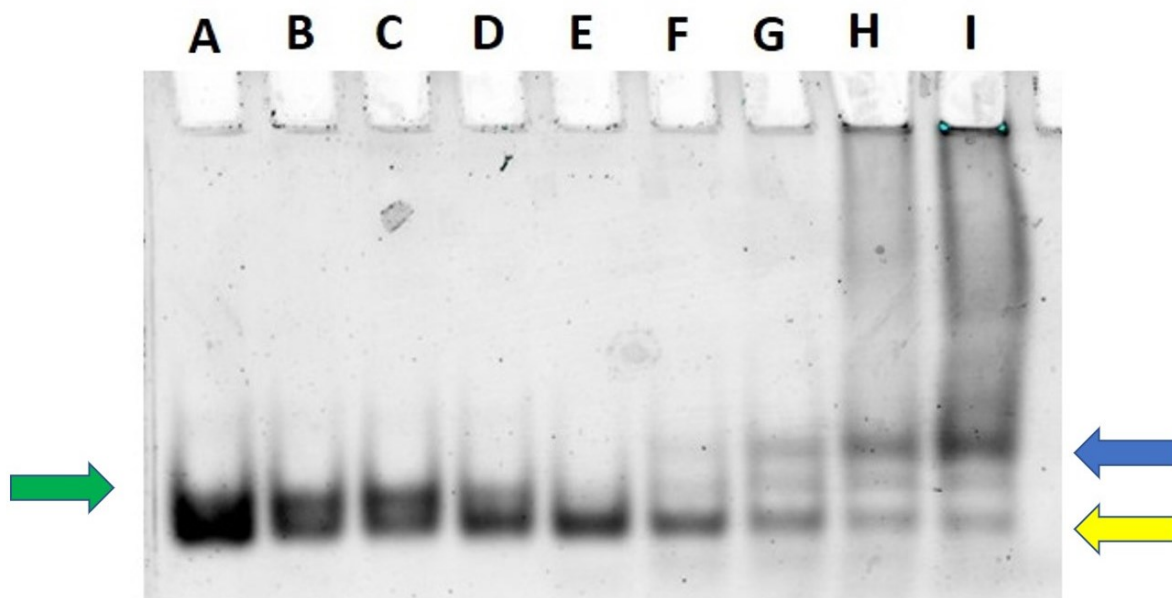**Figure 24.** Polyacrylamide gel electrophoresis based EMSA (Electrophoretic Mobility Shift Assay)_analysis of the duplex TTS DNA in the presence of different concentrations of TFO1 and TFO2. Lanes A to I correspond to TFO1 to TFO2 ratios of 10000:1, 1000:1, 100:1, 10:1, 1:1, 1:10, 1:100, 1:1000, and 1:10000. Arrows indicate: Duplex TTS DNA (yellow), the DNA-RNA hybrid triplex (green), and excess TFO (blue).

4 DISCUSSION

In this chapter, the results of the different experiments are discussed. Future developments and applications of the hybrid triplex structures as a means for the regulation of transcription are presented.

4.1 IN VITRO TRANSCRIPTION FROM ENGINEERED TRANSCRIPTION UNITS

The outcomes presented in section 3.1 demonstrated the successful design of the artificial DNA transcription unit TU. Here, a bacterial promoter consensus (σ70) from *E.coli* was integrated with sequences (TTSs) that can form triplex structures with an additional ssRNA (TFO). This TU component was placed upstream of a template for the fluorogenic aptamer Broccoli and the resulting 90 bp dsDNA was obtained as a dry material from an international manufacturer. The transcription unit, in the presence of σ70-saturated E. coli RNA polymerase, and the four rNTPs (i.e., ribonucleotide mix solution) produced Broccoli RNA that produced the specific fluorescence emission at 505 nm in the presence of its ligand DFHBI-1T.

These results showed that the design approach was correct and that either position of the TTS, downstream of consensus -10 or between consensus -35/-10 did not affect transcription substantially. Although additional characterization of the system might be advisable for the application of this design, for example, to other sigma-factors, in this work the results suggest a possible generalization of the design approach.

4.2 USING TRIPLEXES TO MODULATE TRANSCRIPTION

As already mentioned, previous studies showed that in addition to the inhibitory effect, hybrid triplexes can exert an enhancement effect probably due distortion of the adjacent double helix portions of the dsDNA or via weakening of the DNA duplex. In the first case, the distortion would result in the alignment of the consensus with the sigma factor contact sites, increasing the affinity, while in the second case, the duplex weakening would facilitate its opening by RNAp and the following RNA polymerization. Although the mechanism is still elusive and require further investigation, results shown in section 3.2 showed that the use of this structures to modulate transcription is indeed modular and predictable. The modularity of the system was demonstrated

by the two individual TFOs 1 and 2 targeting the two TTSs downstream of the -10 consensus and the TTS between the -35 and -10 consensus. The formation of one the two triplexes does not affect the formation of the other, resulting in independent triplex formation. On the other hand, the predictability of the system was demonstrated by the triplex effect. This novel TU design comprised TTSs that would generate enhancement triplex, according to a published model. This effect was verified experimentally, validating the model.

To summarize these results, the sequences used in this work were designed according to previous literature in a new artificial transcription unit TU that could generate enhancement triplexes with the TTSs integrated in it and, indeed, *in vitro* RNA transcription experiments showed active Broccoli production, Figure 19, and TFO concentration-dependent enhancement of transcription for individual TFO1 or TFO2, Figures 21 and 22.

Moreover, the successful construction of a new artificial gene regulation system, which presents the advantage of being more stable than previous genelet-based strategies mentioned in the introduction (section 1.5), as the proposed approach, in principle, can be implemented into cells. Finally, as explained in sections 1.7 and 2.2.3, the described system can be further used for the development of a triplex based XOR gate as computational mechanism.

4.3 CONSTRUCTING MOLECULAR GATES USING ENGINEERED BIOLOGICAL COMPONENTS

The construction of engineered components using biomolecules, such as nucleic acids, offers a high degree of specificity and programmability in molecular gate systems. Some examples of nucleic acid-based molecular gates, based on the specificity of DNA/RNA duplex formation and, in general, on the specific sequence recognition of nucleic acids, commonly via Watson-Crick hydrogen bond formation. RNA aptamers, which are single-stranded RNA sequences that bind to specific target molecules with high affinity, have also been studied for their potential involvement in molecular gating. [35]

In this study, a new approach based on hybrid DNA-RNA triplexes was used to construct a logic gate, and specifically a XOR gate. Using the TFOs and TU already described and whose characterization was discussed in the previous section, a XOR gate was implemented where the

TFOs represented the inputs and the TU represented the gate. In this context, the transcription rate was used as output and the fluorescence emission of Broccoli-DFHBI-1T complex was used as readout.

Results demonstrating the effective implementation of the triplex-operated XOR gate are presented as a bar plot in Figure 23. Here, average transcription rates estimated using Broccoli fluorescence emission changes in time are reported, for each input combination. According to the XOR gate truth table, reported in Figure 16, the output of such gate is "0" when the two inputs are identical (0, 0 or 1, 1) while the output is "1" when the input set contains two inputs of different values (0, 1, or 1, 0). As designed, the rates are high when only one input TFO is added to the reaction solution (central bars of the plot), while the rates are slower either in the presence of both input TFOs or in the absence of both (leftmost and rightmost bars). Although the individual average rates producing a "zero" are different, this is not affecting the outcome of the gate due to the arbitrary threshold set to discriminate between the two output values. Similarly, although being higher than threshold, the individual "one" outputs are different. These small differences were attributed to the different "states" of the transcription units and might be used to build non-logic computing units where inputs and outputs could assume more than two values, for example ternary gates.

In conclusion, the findings of this research demonstrate the feasibility of constructing a logic gate, particularly an XOR gate, capable of generating "0" or a "1" outputs encoded by transcription rates, in accordance with the molecular TFO input set, following a XOR truth table. The system might be used to develop additional gates and its biocompatibility could be used to improve previous systems such as the genelets. In addition, their introduction in cells is foreseen as molecular gates using engineered biological components have shown great potential for controlling molecular transport across cellular membranes and nanoscale compartments. [36]

4.4 FUTURE PERSPECTIVES: TRIPLEX-BASED THRESHOLD GATES AND NEURAL NETWORKS

Triplex-based threshold gates have the potential to be used as promising tools for the advancement of the molecular computing field. As highlighted in the previous sections, in this work a triplex-controlled transcription unit was successfully used to exhibit the properties of a XOR gate. Notably, this molecular gate mechanism holds potential for the development of threshold gates, which serve as fundamental nodes in neural networks, the structure of biological brains. These networks, composed of neurons, rely on threshold gates to determine whether an input set signal will propagate through the network or not. When the neuronal node receives an input set composed of inputs with different values, it operates two operations: First, a sum of the inputs is calculated followed by the thresholding operation. If the sum is higher than the threshold, the neuron fires (output = 1), while if the input sum is lower than the threshold, the neuron does not fire (output = 0). To molecularly implement this mechanism, TFOs could be used as inputs while a complementary TFO could be used as threshold, realizing the molecular threshold gate with a similar mechanism as the XOR gate described in this thesis. This triplex-operated molecular threshold gate would include an engineered transcription unit producing Broccoli that could be used as readout.

This mechanism possesses characteristics that deem it an intriguing candidate for universal molecular computing thanks to the tunability of neural networks that could operate complex functions even without logic gates. However, it is important to acknowledge that further investigation and research are imperative to validate the reliability and functionality of this system.

In summary, the construction of a XOR gate, exemplifies the potential of this molecular gate mechanism. It holds promise as a fundamental component for neural networks and has the prospect of serving as a universal molecular computing mechanism, albeit requiring further scrutiny and exploration to ensure its robustness and efficacy.

This experiment on this subject was included in a recently published paper. [37]

5 REFERENCES

[1] Krebs JE, Goldstein ES, Kilpatrick ST, Bartlett J&, Lewin's Genes Xii, Volume 107-110, 2018.
https://www.yumpu.com/en/document/view/68308335/pdf-lewins-genes-xii-12th-edition

[2] Horiuchi H. Molecular structure of nuclei. European Physical Journal, Volume 131-133, 2002.
https://doi.org/10.1140/epja/i2001-10240-x.

[3] Ussery DW. DNA Structure: A-, B- and Z-DNA Helix Families. Encyclopedia Of Life Sciences, Volume
1, 2002. DOI:10.1038/npg.els.0003122.

[4] Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: Stability and function of G-
quadruplex structures. Nature Reviews Genetics, Volume 13, 2012. DOI:10.1038/nrg3296.

[5] Choi J, Majima T. Advances in DNA-based nanotechnology themed issue. Chemical Society
Reviews, Volume 7, 2011. DOI:10.1039/c1cs15153c.

[6] Felsenfeld G, Davies DR, Rich A. Formation of A Three-Stranded Polynucleotide Molecule. Journal
of American Chemical Society, Volume 1-3, 1957. DOI:.1021/ja01565a074.

[7] Kaufmann B, Willinger O, Kikuchi N, Navon N, Kermas L, Goldberg S, et al. An Oligo-Library-Based
Approach for Mapping DNA-DNA Triplex Interactions in Vitro. Journal of American Chemical
Society- Synthetic Biology, Volume 1-2, 2021. DOI:10.1021/acssynbio.1c00122.

[8] Zhou H, Hintze BJ, Kimsey IJ, Sathyamoorthy B, Yang S, Richardson JS, et al. New insights into
Hoogsteen base pairs in DNA duplexes from a structure-based survey. Nucleic Acids Research,
Volume 43, 2015. DOI:10.1093/nar/gkv241.

[9] Jeng SCY, Trachman RJ, Weissenboeck F, Truong L, Link KA, Jepsen MDE, et al. Fluorogenic
aptamers resolve the flexibility of RNA junctions using orientation-dependent FRET. RNA-Journal,
Volume 435-438, 2021. DOI:10.1261/rna.078220.120.

[10] Li X, Wu J, Jaffrey SR. Engineering Fluorophore Recycling in a Fluorogenic RNA Aptamer.
Angewandte Chemie - International Edition, Volume 19-20, 2021. DOI:10.1002/anie.202108338.

[11] Kilgard R, Heim AB, Tsien RY. Improved green fluorescence. Nature, Volume 663-664, 1995.
https://doi.org/10.1038/373663b0.

[12]   Paige JS, Wu KY, Jaffrey SR. RNA mimics of green fluorescent protein. Science, Volume 13-14, 2011. https://doi.org/10.1126/science.1207339.

[13]   Song W, Strack RL, Svensen N, Jaffrey SR. Plug-and-play fluorophores extend the spectral properties of spinach. Journal of American Chemical Society, Volume 2-5, 2014. DOI:10.1021/ja410819x.

[14]   Filonov GS, Moon JD, Svensen N, Jaffrey SR. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. Journal of American Chemical Society, Volume 136, 2014. DOI:10.1021/ja508478x.

[15]   Dolgosheina E V., Jeng SCY, Panchapakesan SSS, Cojocaru R, Chen PSK, Wilson PD, et al. RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. Journal of American Chemical Society - Chemical Biology, Volume 8-10, 2014. DOI:10.1021/cb500499x.

[16]   Xue L, Karpenko IA, Hiblot J, Johnsson K. Imaging and manipulating proteins in live cells through covalent labeling. Nature Chemical Biology, Volume 11, 2015. DOI:10.1038/nchembio.1959.

[17]   Babendure JR, Adams SR, Tsien RY. Aptamers Switch on Fluorescence of Triphenylmethane Dyes. Journal of American Chemical Society, Volume 14716-14717, 2003. DOI:10.1021/ja037994o.

[18]   Ouellet J. RNA fluorescence with light-Up aptamers. Frontiers, Volume 4, 2016. DOI:10.3389/fchem.2016.00029.

[19]   Anand Kumar S. The structure and mechanism of action of bacterial DNA-dependent RNA polymerase. Progress In Biophysics And Molecular Biology, Volume 38, 1981. DOI:10.1016/0079-6107(81)90013-4.

[20]   Clark DP, Pazdernik NJ, McGehee MR. Cloning Genes for Synthetic Biology. Science Direct, Volume 199-239, 2019. DOI:10.1016/b978-0-12-813288-3.00007-0.

[21]   Browning DF, Busby SJW. Local and global regulation of transcription initiation in bacteria. Nature Reviews Microbiology, Volume 14, 2016. DOI:10.1038/nrmicro.2016.103.

[22]   Bervoets I, Charlier D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: Opportunities and drawbacks for applications in synthetic biology. FEMS Microbiology Reviews, Volume 310, 2019. DOI:10.1093/femsre/fuz001.

[23]   Paget M, Helmann J. Protein family review: The σ70 family of sigma factors. Genome Biology, Volume 4, 2003. DOI:10.1186/gb-2003-4-1-203.

[24]   Hook-Barnard IG, Hinton DM. Transcription Initiation by Mix and Match Elements: Flexibility for Polymerase Binding to Bacterial Promoters. Gene Regulation And Systems Biology, Volume 279, 2007. DOI:10.1177/117762500700100020.

[25]   Jeong D, Klocke M, Agarwal S, Kim J, Choi S, Franco E, et al. Cell-free synthetic biology platform for engineering synthetic biological circuits and systems. Methods And Protocols, Volume 4-7, 2019. DOI:10.3390/mps2020039.

[26]   Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annual Review Of Biochemistry, Volume 2,7,12,24, 2012. DOI:10.1146/annurev-biochem-051410-092902.

[27]   Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, et al. Large-scale transcriptional activity in chromosomes 21 and 22. Science, Volume 296, 2002. DOI:10.1126/science.1068597.

[28]   Maher LJ, Wold B, Dervan PB. Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation. Science, Volume 245, 1989. DOI:10.1126/science.2549631.

[29]   Bhattacharya S. Truth Table Analysis of Logic Circuits using Reversible Gates. International Journal for Research in Applied Science & Engineering Technology, Volume 552, 2020. DOI:10.22214/ijraset.2020.2085.

[30]   Tucaković Z. Technical Diagnosis of Basic Logic Gates. Journal Of MIPRO, Volume 24, 2016. DOI:10.1515/dma-2014-0020.

[31]   Adleman LM. Molecular computation of solutions to combinatorial problems. Science, Volume 266, 1994. DOI:10.1126/science.7973651.

[32]   Erbas-Cakmak S, Kolemen S, Sedgwick AC, Gunnlaugsson T, James TD, Yoon J, et al. Molecular logic gates: The past, present and future. Chemical Society Reviews, Volume 47, 2018. DOI:10.1039/c7cs00491e.

[33]   Liu L, Liu P, Ga L, Ai J. Advances in Applications of Molecular Logic Gates. Journal of American Chemical Society Omega, Volume 6, 2021. DOI:10.1021/acsomega.1c02912.

[34] Credi A, Balzani V, Langford SJ, Stoddart JF. Logic operations at the molecular level. An XOR gate based on a molecular machine. Journal of American Chemical Society, Volume 119, 1997. DOI:10.1021/ja963572l.

[35] Cecconello A, Magro M, Vianello F, Simmel FC. Rational design of hybrid DNA-RNA triplex structures as modulators of transcriptional activity in vitro. Nucleic Acids Research, Volume 50, 2022. DOI:10.1093/nar/gkac1131.

[36] Lanphere C, Arnott PM, Jones SF, Korlova K, Howorka S. A Biomimetic DNA-Based Membrane Gate for Protein-Controlled Transport of Cytotoxic Drugs. Angewandte Chemie - International Edition, Volume 60, 2020. DOI:10.1002/anie.202011583.

[37] Rilievo G, Cecconello A, Ghareshiran NF, Magro M, Simmel F, Vianello F. Integration of DNA–RNA-triplex-based regulation of transcription into molecular logic gates. FEBS Letters 2023. Doi:10.1002/1873-3468.14721.

[38] Dickerson RE, Ng HL. DNA structure from A to B. Proceedings Of The National Academy Of Sciences, Volume 98, 2001. DOI:10.1073/pnas.141238898.

[39] Semsarha F, Raisali G, Goliaei B, Khalafi H. Microdosimetry of DNA conformations: relation between direct effect of 60Co gamma rays and topology of DNA geometrical models in the calculation of A-, B- and Z-DNA radiation-induced damage yields. Radiation And Environmental Biophysics, Volume 55, 2016. DOI:10.1007/s00411-016-0644-7.