Technische Universität München

Fakultät für Physik

Physik synthetischer Biosysteme – E14

# Application of strand displacement reactions in DNA nanotechnology and synthetic biology

Mario Franz Pierre Teichmann

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

| | |
|---|---|
| Vorsitzender: | Prof. Dr. Ulrich Gerland |
| Prüfer der Dissertation: | |
| | 1. Prof. Dr. Friedrich C. Simmel |
| | 2. Prof. Dr. Joachim Rädler |

Die Dissertation wurde am 23.03.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 15.06.2018 angenommen.

**“The sea, once it casts its spell, holds one in its net of wonder forever.”**

Jacques Yves Cousteau

# Table of contents

# I. List of Figures

# II. List of Tables

# III. Zusammenfassung

Sowohl das auf Nukleinsäuren basierende Design von Strukturen, als auch die Kontrolle von Reaktionsmechanismen stellen Kernpunkte der DNA Nanotechnologie und Synthetischen Biologie dar. Beide gehören zu den innovativsten Disziplinen im Life Science Bereich. Im ersten Teil dieser Arbeit wurde die DNA Nanotechnologie angewendet um Kolokalisierung auf Basis einer zweistufigen DNA Strangverdrängungskaskade zu untersuchen. Diese besteht aus einem Sender- und Empfänger-Gate, welche jeweils auf einer DNA Origami Plattform immobilisiert wurden. Die Leistungsfähigkeit dieser Kaskade variiert beträchtlich mit dem Abstand zwischen den Gates und wies im Hinblick auf eine Konkurrenzreaktion eine hohe Robustheit bei einem mittleren Abstand von 20 nm auf. Die Ergebnisse legen einen in Teilen lokalen Transfer von diffundierenden Signalsträngen zu Empfänger-Gates nahe, die sich auf der gleichen Plattform befinden. Dieser Effekt beruht wahrscheinlich auf einen direkten physikalischen Kontakt zwischen den Gates. Die Implementierung von mehreren Empfänger-Gates hat die Leistungsfähigkeit der Kaskade und ihre Robustheit noch weiter verbessert.

Das zweite Projekt greift Konzepte der Synthetischen Biologie auf und zielt auf die Kontrolle der posttranskriptionalen Genexpression ab, indem es an das Prinzip der *de-novo* konstruierten „Toehold Switch“ Riboregulatoren angeknüpft hat. Toehold Switches wurden in Bakterien als Repressoren der Translation eines Output-Gens eingesetzt und als „Toehold Repressor“ bezeichnet. Die Repression wurde durch RNA-RNA Strangverdrängungsreaktionen gesteuert, die auf der Hybridisierung einer passenden Trigger RNA mit einer “Toehold“ (zu Deutsch: Ansatz) Domäne eines RNA Schalters beruht, der aus einer Haarnadel-Struktur besteht. Dies hatte zur Folge, dass die Ribosomen nicht mehr effizient an die die ribosomale Bindestelle und das Startcodon AUG binden konnten. Wir konnten zeigen, dass die Translation in *E. coli* sowohl durch Verwendung von optimierten synthetischen Trigger RNA Sequenzen, als auch durch endogene mRNA effizient blockiert wird. Dies hat das Potential der Toehold Switches als mRNA Sensoren erweitert.
Darüber hinaus wurden die Toehold Switches in Säugetierzellen implementiert und untersucht. Charakterisierungsversuche in HeLa und HEK293 Zellen deuten auf eine starke Korrelation zwischen der Stärke der Haarnadelstruktur und der Translationseffizienz hin. Allerdings hat sich herausgestellt, dass die Aktivierung der Translation durch Koexpression des RNA Schalters mit einer passenden Trigger RNA sehr anspruchsvoll ist und daher einer eingehenden Untersuchung bedarf.

# IV. Abstract

Rational design of nucleic acid-based structures as well as control of reaction mechanisms are key aspects of DNA nanotechnology and synthetic biology. Both can be considered among the most innovative disciplines in life science. In the first part, DNA nanotechnology was employed to study the effect of colocalization on the performance of a two-stage DNA strand displacement cascade comprised of a sender and a receiver gate immobilized onto a DNA origami platform. The performance of this cascade is found to vary strongly with the distance between the gates and displayed a high robustness with respect to a competing reaction for an intermediate distance of 20 nm. The results suggest a local transfer of a fraction of diffusing signal strands to a receiver gate on the same platform, probably mediated by direct physical contact between the gates. Implementation of multiple receiver gates further increased performance of the cascade and its robustness.

The second project captures synthetic biology concepts and aims to control gene regulation on the post-transcriptional level, taking up the principle of *de-novo* designed toehold switch riboregulators. In bacteria, toehold switches were utilized to repress translation of an output gene, termed toehold repressors. Repression was mediated by RNA-RNA strand displacement reactions, involving hybridization of a cognate trigger RNA to a toehold domain of a switch RNA hairpin module. Consequently, the ribosomal binding site and AUG start codon becomes inaccessible for efficient ribosome binding. We have shown that translation can be efficiently inhibited in *E. coli* by employing optimized synthetic trigger RNA sequences as well as endogenous mRNAs, hence expanding the potential of toehold repressors as mRNA sensors. Furthermore, toehold switches were implemented and studied in mammalian cells. Preliminary characterization in HeLa and HEK293 cells indicate a strong correlation between hairpin strengths and translation efficiency. However, activation of translation via co-expression of the switch RNA together with a cognate trigger RNA turned out to be challenging and thus requires an in-depth investigation.

# 1 Introduction

In our daily routine, we are not necessarily aware of all the processes happening in our body and rather follow them without further reflection. They do not need to be actively controlled as they are just happening, and anyway, questioning of each and every move would surpass the capacities of our brain. Even being aware of simple changeovers, for instance a transition from a sitting position to a standing position requires extensive training of the mind. Maintenance of those moves is passing off in the background, and it seems that our body is executing those commands in a machine-like routine. However, our body represents a highly organized and compartmentalized system, guaranteeing the commands to be regulated by a sophisticated network of control mechanisms and feedback loops. These regulatory systems involve all hierarchies of the human body, starting from the level of limbs and organ systems, composed of tissue systems, which itself are organized in cellular (sub)-compartments. This network finds ultimate reflection in the cell, known as the smallest unit of life that is capable of replicating independently. It involves a vast mix of molecules among deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins are the most prominent macromolecules. Indeed, these levels are highly intertwined, and the macroscopic processes are minutely controlled via an intricate biochemical reaction network down to the micron and nano scale. The capability of DNA to store and replicate genetic information, hence allowing evolution, renders it for all biological processes, enabling what we refer as “life”. Deeper understanding of molecular processes taking place in our body has set its footprints in the middle of last century with the structural determination of DNA in 1953 by Francis Crick and coworkers [1]. Pioneering follow up conclusions include the postulation of the central dogma of biology. It describes the flow of genetic information within a biological system, known as gene expression, starting with DNA, its transcription into RNA, which then serves as template for its translation into proteins. These new insights into fundamental processes of life were accompanied by rapid advances in fields such as genetics, microbiology or biophysics, alongside with technical advances like molecular cloning or discovery of the polymerase chain reaction. Over the course, new life science disciplines arose, taking advantage of progress achieved in fields related to classic biology.

In the 1980’s, researchers amongst crystallographer Nadrian Seeman started considering DNA as building material in a non-biological context rather than focusing on its undoubted biological relevance. First rationally designed nanostructures made out of synthetic DNA laid

the foundation for a field known as structural DNA nanotechnology which in general aims for thermodynamically stable structures. With the possibility to position molecules within a nanometer precise addressability, DNA nanotechnology offered innovative ways to colocalize reaction components on custom designed DNA structures. This opportunity allowed to study processes that follow non-equilibrium dynamics in a chemically enclosed environment, comprising the field of dynamic DNA nanotechnology. Indeed, colocalization and compartmentalization play an important role in cellular reaction networks such as scaffolded protein signaling cascades or multienzyme complexes, enabling streamlined reaction flux, and by spatial proximity, minimization of unwanted side reactions.

The first project in this thesis employs principles of structural as well as dynamic DNA nanotechnology. It is taking up potential benefits of colocalization as it was aiming at the experimental investigation of the influence of spatial organization and local molecular interactions on the dynamics of signaling networks and molecular reactions. Reactants were brought in close proximity onto a DNA scaffold by using a rectangular DNA platform with dimensions of ~90 x 65 nm, fabricated by the DNA origami method [2]. A biochemical signaling cascade was mimicked by immobilizing a two-stage DNA strand displacement (DSD) cascade comprised of a sender and a receiver gate onto a DNA origami platform. Strand displacement reactions are widely used operations in dynamic DNA nanotechnology and take advantage of the sequence programmability and thermodynamic characteristics of nucleic acids. The addition of a DNA or RNA input strand displaces a signal strand from the sender gate, which in turn can transfer to the receiver gate and displace an output strand. As involved cascade elements were modified with fluorophores, strand displacement reactions could be monitored via fluorescence spectroscopy and allowed valuable insights on how colocalization and spatial proximity as well as external factors influence reaction kinetics.

Even though strand displacement reactions originate from dynamic DNA nanotechnology, permitting to control biomolecular processes in *in vitro* settings, this type of operation has been adopted by another quite recently evolved discipline, known as synthetic biology, to which the second part of this thesis is devoted. Inspired by the mindset and terms commonly found in engineering, synthetic biology seeks to design and create programmable genetic parts and artificial circuits, exhibiting *de-novo* functionalities, in order to control reactions in (non)-living systems.

The projects presented in this thesis are based on *de-novo*-designed toehold switches, a strategy to control gene expression in bacteria on the post-transcriptional level [3]. Here, a RNA switch hairpin sequestering the region around the AUG start codon and leaving the ribosomal binding site (RBS) in a single-stranded loop was inserted upstream of a gene of interest, resulting in repression of translation. The innovative concept of translational activation of this work involves a trigger RNA strand. Its binding to a linear toehold domain at the 5' end of the hairpin initiates strand displacement reactions, making RBS and start codon now accessible for ribosome binding and initiation of translation. Interaction of switch and trigger strands solely relied on synthetic and rationally designed RNA species, hence expanding the design flexibility and therefore resulted in the creation of a sequence library of highly orthogonal parts.

In bacteria, the design principle of toehold switches was utilized to create riboregulators that repress protein translation upon interaction with the trigger RNA. Similar to the immobilized DSD cascade, fluorescence provided optical readout. This time, flow cytometry measurements served as quantitative method of choice. Moreover, the concept of toehold switches was implemented into mammalian cells, aiming for control of translation in eukaryotes. However, mammalian synthetic biology still lacks the theoretical knowledge and the repertoire of standardized genetic parts compared to bacterial synthetic biology. The explanation becomes obvious as the eukaryotic gene expression machinery is far more complex than the one prevalent in bacteria. Preliminary results demonstrate the influence of RNA secondary structures on translation and prove high translational repression for strong hairpin structures. Subsequent attempts to activate gene expression by employing a similar strategy as for the original toehold switches remain challenging though.

Mimicking and controlling cascaded reaction pathways using the advantages of DNA nanotechnology as well as regulating translation in bacteria and mammalian cells feeding in synthetic biology represent two *de-novo* designed examples of how reactions on the nanoscale can actively be controlled by us humans. This is fascinating as the current understanding of our own functionality is still in the early stages of development.

# 2 Basics

## 2.1 Nucleic acids

*"The double helix is indeed a remarkable molecule. Modern man is perhaps 50,000 years old, civilization has existed for scarcely 10,000 years and the United States for only just over 200 years; but DNA and RNA have been around for at least several billion years. All that time the double helix has been there, and active, and yet we are the first creatures on Earth to become aware of its existence."*
Francis Crick, unknown date

Nucleic acids, DNA and its related molecule RNA, are key molecules enabling existence and progression of evolution of life on earth due to their ability for storing, replicating and processing genetic information and thus allow and guarantee what we define as "life". Since its discovery and early description, it has found great attention leading to molecular biology tools and rapid progress in genetic engineering. In particular, the ability to synthesize custom designed DNA, and more recently RNA molecules, makes nucleic acids suitable for addressing advanced cell biology questions as well as for bionanotechnology. The well-defined chemical and structural properties render this type of biopolymer applicable for the fabrication of building blocks on the nanoscale (chapter 2.1.2). In recent years, nucleic acids served as a material to build structures and to modulate reaction pathways. For the first time this molecular addressability allowed bottom up assembly of structures made out of DNA. This field is commonly known as DNA nanotechnology. Another discipline taking advantage of DNA related technology is synthetic biology even though it arises from a different motivation and scientific background (chapter 2.3). Synthetic biology aims to create genetic circuits and biological systems inspired by classic engineering approaches. For this, the knowledge and understanding of nucleic acids' dimensions as well as chemical and structural properties is necessary. In the following passages, chemical and physical properties of nucleic acids will be highlighted, with attention specifically drawn to the projects presented in this work. A more in-depth description of those biochemical and physical properties can be found in well-known literature [4,5].

### 2.1.1 Hard facts – Biophysical properties of nucleic acids

DNA as the most recognized nucleic acid is famous to the broad public for its double-stranded helical shape. Searches in the web for the term "DNA" result in images and descriptions presenting a double helical structure whereas the term "RNA" will mostly be depicted as a single-stranded molecule. Nevertheless, both DNA, and RNA can exist as a linear polymer, of which the primary structure is consisting of repetitive elements with well-defined dimensions. The constituent monomers of this biopolymer, termed nucleotides are themselves composed of three components which are chemically linked together: An aromatic nitrogenous nucleobase (Figure 1A), a furanoside ribose sugar entity – either deoxyribose for DNA or ribose for RNA (Figure 1B) and a phosphate group (Figure 1C) which is linked to the ribose via a covalent phosphodiester bond thus forming a very stable unit. The phosphate group is responsible for a negative charge at physiological conditions. It is linked to the 5' carbon atom of the furanoside sugar resulting in an alternating and repetitive polymer forming the backbone.

Nucleobases are connected with the sugars' first carbon atom at position C1 by N-glycosidic bonds. These nucleobases can be discriminated into two different subgroups which are classified as purines and pyrimidines. They are the basis of a molecule family known as nucleosides which comprises nucleobases connected to a sugar moiety – deoxyribonucleosides in DNA or ribonucleosides in RNA. Purines represent heterocyclic, aromatic molecules composed of a pyrimidine ring fused to an imidazole ring. Within this structure, the nitrogen atom at position N9 connects with the adjacent sugar moiety. Pyrimidines by contrast represent heterocyclic, aromatic, organic molecules possessing two nitrogen atoms that connect to the sugar at position N1. In here, the canonical purine bases are adenine (A) and guanine (G), whereas pyrimidines make up cytosine (C) and thymine (T) for DNA.

Differences in biological function of DNA and RNA which are evident will not be elaborated in this thesis as it is outside its scope. From a structural point of view, as already indicated by its name, the sugar entity differs for DNA and RNA by the presence or absence of a hydroxyl group at the 2' carbon atom of the furanoside in RNA and DNA, respectively, annotated in blue (Figure 1B). In the same breath, one has to consider that in RNA, thymine is replaced by uracil (U) which is unmethylated at position C5, compared to its methylated counterpart thymine, see nucleobases in Figure 1B. For single-stranded DNA, the asymmetry in bonding

determines the directionality of the phosphate sugar backbone resulting in distinct termini known as 5' and 3' ends.

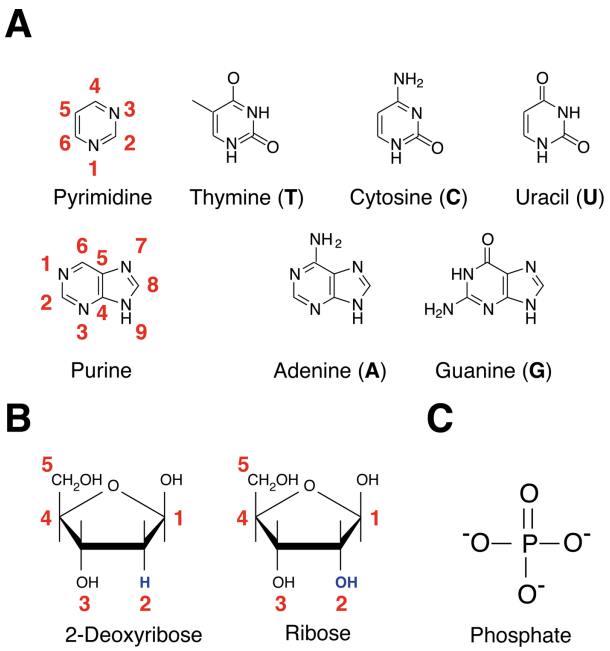


Figure 1. Key components of nucleic acids. Positions of accordant atoms are indicated in red. **A**. Canonical nucleobases and its classification into pyrimidines and purines. Uracil is solely found in RNA and differs from its counterpart thymine in the lack of a methyl group at position N5. **B**. Depiction of the two furanosides, 2-deoxyribose and ribose, found in DNA and RNA, respectively. Main difference of deoxyribose and ribose is the missing of a hydroxyl group in deoxyribose at position C2. **C**. Chemical structure of a phosphate molecule possessing a triple negative charge.

## Hybridization

Double stranded DNA is the result of two complementary single-stranded DNA molecules hybridizing to each other, forming a helical antiparallel orientation in relation to their respective backbones. The transition of single-stranded (ss) to double-stranded (ds) DNA is achieved by Watson-Crick base pairing where adjacent bases can pair to each other allowing thymine to form two hydrogen bonds with an opposite adenine and guanine a three-hydrogen bond strong connection with cytosine, see Figure 2 [1]. The predominant, canonical superstructure of ds nucleic acids are the right-handed B-form for DNA and the A-form (also right-handed) for RNA. The forms differ in rise per base pair (bp), thus resulting in different overall compactness in A-form compared to B-form. (see Table 1).

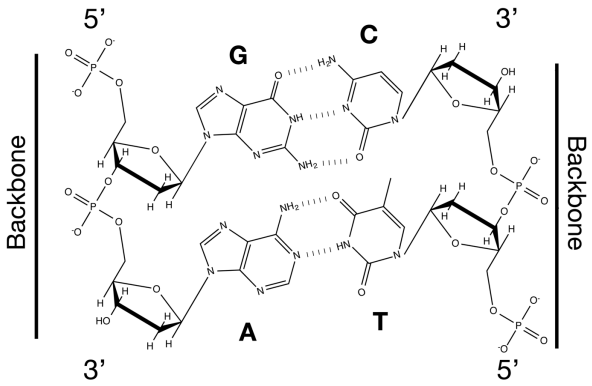


Figure 2. Orientation and overall geometry of an excerpt of double-stranded DNA highlighting the repetitive sugar-phosphate backbone. Watson-Crick base pairing is possible between cytosine (**C**) and thymine (**G**), stabilized by three hydrogen bonds, whereas adenine (**A**) is linked to thymine (**T**) by two hydrogen bonds. The nucleotides are linked to each other by the 3-hydroxyl of one 2-deoxyribose through the phosphate attached to the 5-hydroxyl of another nucleotide 2-deoxyribose. Modified from [6].

Independent of the particular sequence of nucleotides, reverse complementary DNA polymers always form a similar double helical geometrical structure that is held together by Watson-Crick base pairing of respective complementary monomers. The series of successive base pairing is forming the actual sequence pattern and results in a stable helical structure keeping same dimensions. The major contribution to the thermal stability of the double helix nucleobases represents base stacking interactions between vertically adjacent nucleobases. Base stacking indeed contributes to a much higher extent to the thermal stability of the DNA duplex, in comparison to Watson-Crick base pairing of complementary strands [7]. Base stacking effects have also been utilized in various works within the field of DNA nanotechnology [8]. Figure 2 generically underlines the difference between these two forces as it shows base pairing between complementary bases G-C, A-T as well as base stacking interactions between - in this example - G-A and C-T positioned on top of each other.

The right-handed B-form DNA double helix in its anti-parallel nature is presented in a three-dimensional illustration (Figure 3). The antiparallel configuration of DNA is not the only possible configuration. Though there exist other observed ones like parallel and antiparallel configurations due to formation of noncanonical hydrogen bonding arrangements which are stable only under certain conditions [9]. Two examples of such uncommon secondary structures known in DNA nanotechnology are G-quadruplexes and I-motifs which are rich in guanine and cytosine, respectively [10].

For nucleic acid nanotechnology, the dimensions of single nucleic acid strands and double helices are crucial, as they determine the geometrical properties of the resulting constructs. Some of the key parameters are given in the following: For dsDNA, the rise per base pair is 0.34 nm, resulting in a full helical turn of 10.5 nm [5] (Figure 3A). Two grooves provide

substantial influence on the overall shape of the helix, forming a helical twist: The major groove with a width of 1.17 nm and a depth of 0.87 nm and the minor groove possessing a width of 0.57 nm and a depth of 0.75 nm [4,9,11]. Interestingly, interactions of many DNA binding proteins with the double helix mainly occur via the major groove as its accessibility is higher [12]. The diameter of dsDNA in its B- form is reported to be 2 nm under physiological conditions [13] (Figure 3B). DsRNA in its A-form helix is characterized by a slightly more compact nature resulting from a reduced rise per base pairs (0.24-0.28 nm/bp) and a slightly wider diameter of 2.6 nm [14,15].

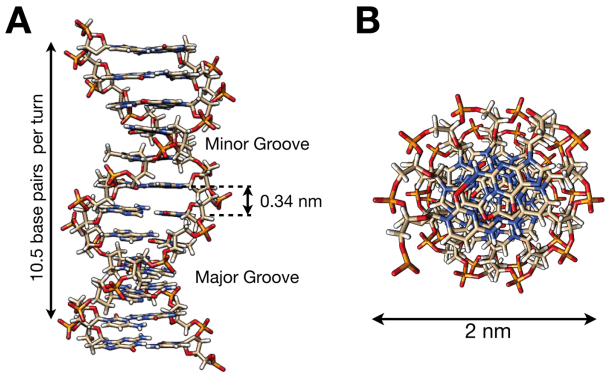


Figure 3. Three-dimensional representation of right-handed B-form double helix. Regions of interest are denoted: **A**. Pitch of 10.5 bp per full helical turn. The structural discrimination between major and minor groove results from the different angle between paired nucleotides. The average separation of each base pair is 0.34 nm. **B**. Cross section of the double helix with its diameter of 2 nm [13]. Respective atoms are colored as follows: Carbon atoms in grey, hydrogen atoms in white, oxygen atoms in red, phosphor atoms in orange, nitrogen atoms in blue.

## Persistence length

Another important mechanical parameter of DNA and RNA represents the persistence length, which defines the stiffness and flexibility of a polymer. It represents an important factor to consider when it comes to design structures and programmed molecular reactions made from nucleic acids. For dsDNA in its B-form the persistence length is reported to be around 50 nm, whereas for single-stranded DNA it is only 1.5 – 3 nm, Table 1 [16]. For dsRNA in its A-form helix, the persistence length is slightly longer due to its increased compactness, noted with 60 nm [17]. SsRNA can be assumed to possess similar dimensions as single-stranded DNA.

| | dsDNA (B-form) | ssDNA | dsRNA (A-form) | ssRNA |
|---|---|---|---|---|
| Rise/bp (nm) | 0.33 | 0.6 | 0.24 – 0.28 | |
| Persistence length (nm) | 50 | 1.5–3 | 60 | 1.5–3 |

Table 1. Key dimensions of double-stranded and single-stranded DNA and RNA [4,9,11,16].

## Nearest neighbor model – Thermodynamics

Thermodynamics of DNA hybridization and the resulting binding energies between nucleobases are crucial for prediction of structure and behavior of DNA molecules. Its biological relevance is undoubtedly important for the conversion between double-stranded and singled stranded structures that occur during major biological processes like replication, recombination or transcription. Apart from classic molecular biology applications like polymerase chain reaction (PCR), exact structure prediction of DNA and RNA as well as hybridization play a crucial role in nanotechnology. In here the so called Nearest Neighbor (NN) model, which is based on Watson-Crick base pairing parameters, will be introduced. Knowledge and key parameters presented in the following is extracted from works of Santa Lucia and Hicks [18,19]. Based on existing literature, the authors created a unified model for predicting thermodynamics and duplex stability of DNA and RNA based on the NN model that Crothers and Zimm initially introduced [20]. The NN model for nucleic acids expects that the analyzed nucleic acids are 5'-3' Watson-Crick base paired. Moreover, the NN model postulates that the identity and orientation of neighboring base pairs are crucial factors for the stability of any given base pair. In general, this parameter for duplex stability is given by the free energy $\Delta G_T^0$ or $\Delta G_{37}^0$, which accounts for physiological conditions (see Equation 2).
Several parameters are involved when initially forming the duplex comprising the initiation parameter ($\Delta G_{37\,\mathrm{initiation}}^0$) which includes differences between terminal and internal nearest neighbors and counter ion condensation. Moreover, differences between duplexes with terminal GC pairs versus AT pairs are introduced in the initiation parameter. Additionally, an entropic penalty for the maintenance of the C2 symmetry of self-complementary duplexes is introduced in the calculation which is built up based on the general formula for calculation of free energy ΔG dependent on temperature (Equation 1)

$$\Delta G_T^0 = \Delta H^0 - T\Delta S^0 \quad (1)$$

in which temperature T is denoted in K, $\Delta H^0$ in kcal/mol and $\Delta S^0$ in cal/kmol, assumed to be independent on temperature.

$$\Delta G^0_{37}(\text{total}) = \Delta G^0_{37\,\text{initiation}} + \Delta G^0_{37\text{symmetry}} + \sum \Delta G^0_{37\,\text{stack}} + \Delta G^0_{\text{AT terminal}} \quad (2)$$

In Table 2, the thermodynamic NN parameters for Watson-Crick base pairs are listed for a salt concentration of 1 M NaCl at 37°C. This dataset represents a compilation of measurements originating from different previous works of which the parameters were extracted from multiple linear regression of 108 sequences solving for 12 unknowns – 10 NN propagation parameters, one initiation parameter, and one correction parameter for terminal AT pairs.

| Propagation sequence | Δ H° (kcal mol$^{-1}$) | Δ S$^0$ (e.u.) | Δ G$^0_{37}$ (kcal mol$^{-1}$) |
|---|---|---|---|
| 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 AT penalty | + 2.2 | + 6.9 | + 0.05 |
| Symmetry correction | 0.0 | − 1.4 | + 0.43 |

Table 2. Nearest-neighbor thermodynamic parameters for DNA [18,19].

Following example applies the unified NN parameters using random duplex 5’ CGTTGA 3’. As this sequence is non-self complementary, symmetry penalty is not applied. Table 2 provides $\Delta H^0$ and $\Delta S^0$ parameters.

$$5'\ \text{CGTTGA}\ 3' = \Delta G^0_{37\,\text{initiation}} + \Delta G^0_{37\text{symmetry}} + \text{CG/GC} + \text{GT/CA} + \text{TT/AA} + \text{TG/AC} + \text{GA/CT} + \text{AT}_{\text{terminal}}$$

$$\Delta G^0_{37}\ (\text{predicted})\ \text{in kcal/mol} = 1.96 + 0 - 2 - 17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.05$$

$$\Delta G^0_{37}\ (\text{predicted}) = -5.35\ \text{kcal/mol}$$

These thermodynamic aspects and parameters were also considered in works presented below. Specifically, nucleic acid structure and devices were designed using the nucleic acid design and analysis software NUPACK, which performs calculations using these assumptions [21].

### 2.1.2 Structural nucleic acid nanotechnology

**Taking advantage of nucleic acids as programmable building blocks**

Knowledge of chemical, physical and mechanical properties of DNA were fundamental for the development of nucleic acid nanotechnology. Apart from its undoubted relevance in biology in terms of storage and the ability to process genetic information, the programmable characteristic of the biopolymer renders DNA as an appropriate molecule to build up structures and control its motions on the nanoscale.

Given the term "Bionanotechnology", this field of research can be considered as one of the most interdisciplinary disciplines in life science combining knowledge and mindset from physics, chemical, biological, and material sciences. In recent years, nucleic acids' non-biological use as building material as well as tool for biomolecular computation steadily grew – also benefiting from continuously falling costs of DNA synthesis. Herein, especially DNA nanotechnology, as well as recently RNA nanotechnology are prominent subdisciplines. In this work the emphasis will be on DNA nanotechnology which can be divided into structural and dynamic DNA nanotechnology. As already described above, DNA predominantly exists as an antiparallel double helical polymer that can potentially base pair with any other base, including itself via Watson Crick base paring [22]. In nature, more complex structures based on the DNA double helix are known, such as the triply branched replication fork which is forming during semiconservative replication [1]. Another example is the four-arm branched Holliday junction, named after the biologist Robin Holliday who first described the structure in 1964. Those junctions are created as intermediates during genetic (homologous) recombination [23].

These structures inspired researchers in the early 1980s to rethink the DNA molecule as construction material for designing and creation of structures with predefined patterns, shape and function. Nadrian Seeman, who started his scientific career as a crystallographer was inspired by the predictable geometry and programmability of the DNA molecule. By observing events happening during genetic replication Seeman and coworkers questioned the concept of formation of immobile, disjoint junction structures (Figure 4A). He summarized that idea in one statement: "*It is possible to generate sequences of oligomeric nucleic acids which will preferentially associate to form migrationally immobile junctions, rather than linear duplexes, as they usually do.*" [24]. In this conceptual work the formation of two- and three-dimensional lattices out of customized immobile junctions using specific design rules

and DNA as the material of choice turned out to be a promising approach. It was only a matter of time until the first rationally design structure would be created. He proposed to position other macromolecules like proteins between the junctions of these three-dimensional lattices (Figure 4B). This epoch can be considered as time of birth for the field of DNA nanotechnology. It became quite clear that in this case one is rather interested in stability and final equilibrium. Only one year later Seeman *et al.* would show the first experimental demonstration of successful design and construction of immobile nucleic acid junctions [25]. Since the description of the programmability of immobile junctions synthesized from customized DNA strands, the repertoire of structures as well as its complexity rapidly increased over the following years. This led to the first demonstration of a three-dimensional cube-like structure that uses stepwise hybridization and cyclization of single-stranded sticky ends on the immobile junctions [26]. This work can be well regarded as pioneering standard reference for DNA nanotechnology (Figure 4C).

This was succeeded by the successful preparation of DNA double crossover structures of parallel or antiparallel nature [27]. Previous works paved the way for the construction of more complex and stable, rigid, periodic lattices, even for the DNA origami technology which emerged 13 years later. Following these groundbreaking works, several more complex two and three-dimensional structures have been designed and successfully demonstrated. In 1998, the first successful example of two-dimensional periodic crystals was presented that were made out of synthetic DNA strands. The researchers used the double crossover motif and sticky ends to create intermolecular connections which have been shown to result in DNA lattices of up to micrometer size (Figure 4D) [28].

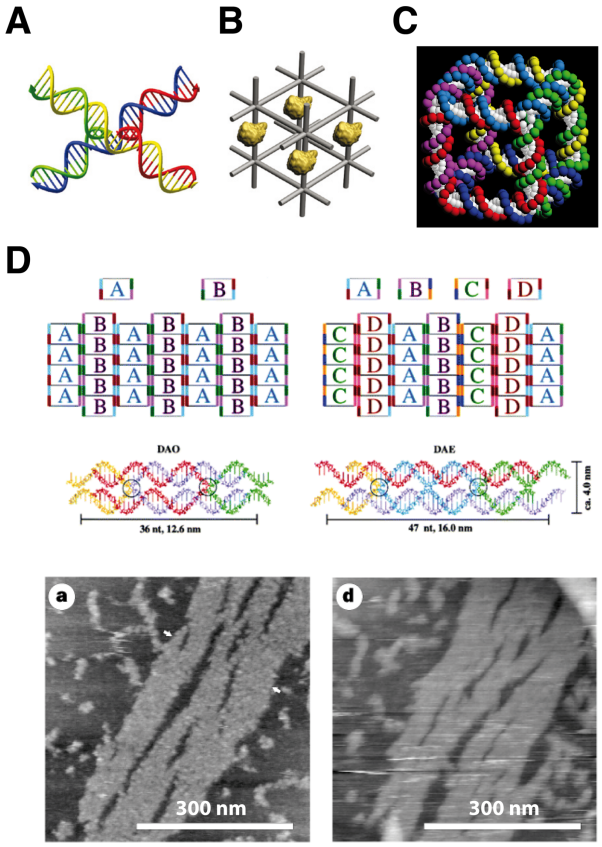


Figure 4. Pioneering structural motifs of early days DNA nanotechnology. **A**. Model of Ned Seemans design of an immobile DNA junction composed of four rationally designed DNA oligonucleotides. **B**. Proposed two- and three-dimensional latticed geometry serving as scaffold to periodically arrange macromolecules such as proteins [24]. **C**. Realization of the first three-dimensional structure made from synthetic DNA strands [26]. **D**. Design of double crossover structural motifs (DAE = double crossover, antiparallel, even spacing, DAO = double crossover, antiparallel, odd spacing) and its organization into two-dimensional crystalline lattices, consisting of two (A, B) and four units (A, B, C, D). Below: Atomic force microscopy (AFM) images are shown for lattices of DAE and DAO units. Scale bars = 300 nm [28].

A breakthrough was achieved in 2006 when Paul Rothemund successfully demonstrated that DNA can be folded into almost arbitrary structures. To achieve this he hybridized a long single-stranded circular M13 viral DNA (termed "scaffold" strand) with a set of short customized oligonucleotides (termed "staple" strands) which he termed "DNA origami" - referring to the Japanese art of folding paper into predefined shapes [2] (Figure 5A-C). This technique enabled the design and fabrication of DNA nanostructures with predetermined size and geometry. The shape in the first place has to be designed *in silico* with the available aid of appropriate design software [21]. This technique was implemented to create two-dimensional structures by applying specific design rules like staple length, crossover positions or helical twist as well as scaffold routing. The result was a great variety of two-dimensional structures, of which the majority seemed to be constructed for demonstration of design features, yet lacking a noticeable purpose. Various DNA origami structures were used as molecular pegboard as they were modified with a broad range of small molecules like gold particles or

fluorescent dyes taking advantage of its chemical addressability. Over the next years this technique was gradually improved and consequently broadened to the creation of three-dimensional shapes from 2009 on (Figure 5D and E). This was possible due to major design adaptions and expansion of design space by e.g. introducing DNA helices into a honeycomb pleat-based-strategy by piling up (stacking) layers of antiparallel DNA helices. Also, more advanced design rules were applied like bending or stretching DNA oligonucleotides into twisted and curved structures along their helical-parallel axes [29-31].

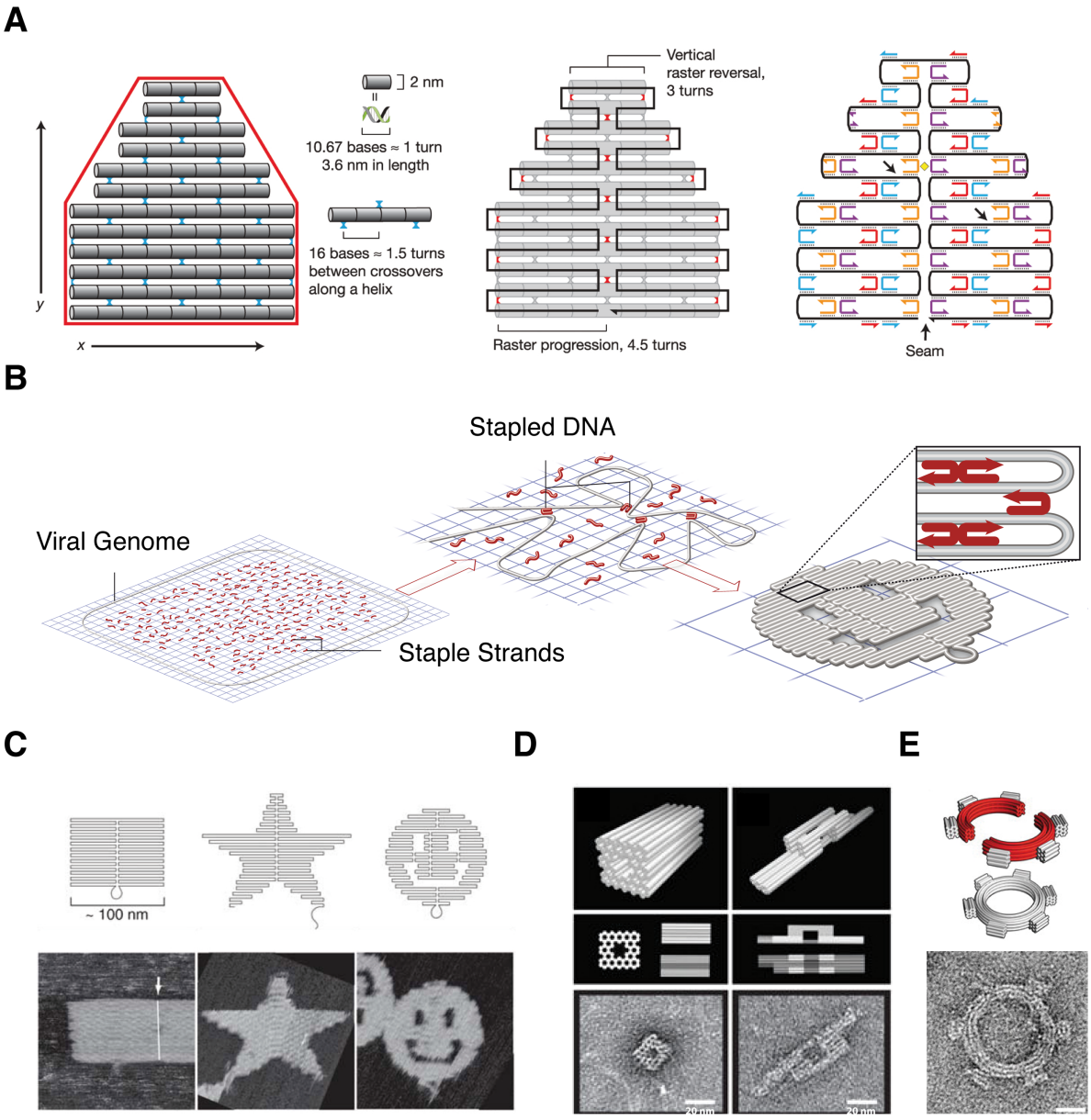


Figure 5. Introduction of the DNA origami method. **A**. Left figure: Design process of a DNA origami structure. Intended shape of structure is predefined (red) which is "filled up" by parallel double helixes which are linked together by crossovers every 1.5. turns (blue). Middle figure: Routing of the scaffold strand (black) resulting in additional crossovers (red). Right figure: Linkage of scaffold strand with 16 nt long staple strands responsible for creation of the desired shape [2]. **B**. Schematic illustration of the DNA origami folding process [32]. **C**. Representation of different two-dimensional DNA origami structures. Top row shows the folding path. Left: Rectangle, middle: Star, right: Disk with three holes ("smiley"). Single stranded sequences occur as dangling curves and loops. Bottom row: Atomic force microscopy (AFM) images of respective shapes

(Left image shows blunt-end stacking, indicated by white lines and arrows). Images are 165 x 165 nm of size [2]. **D**. Expansion of DNA origami technique for the creation of three-dimensional structures. Left: Square nut, right: Railed bridge. Perspective and projection views of structures are shown in top the two rows with accordant transmission electron microscope (TEM) micrographs (bottom row) [29]. **E**. Further development of three-dimensional structures introducing curvature and bending. Top: In scilico design. Bottom: Respective TEM image [30]. Scale bars for all TEM images are 25 nm.

By using appropriate software like CanDO, design of such structures became also more feasible as it enabled its *a priori* simulation and prediction [21,31,33]. The gained ability to design three- dimensional structures obviously offered a much broader range for more sophisticated technical applications. Some examples of functionalization of those two- and three-dimensional structures will be addressed below.

**Excurse: RNA nanotechnology**

Alongside the steady expansion of the knowledge and repertoire in DNA nanotechnology, remarkable advances were achieved in the field of RNA nanotechnology [34,35]. In comparison to DNA, RNA exhibits a more functional potential as it can be produced enzymatically and e.g. displays catalytic activity (ribozymes) or the capability to specifically bind ligands (aptamers). Moreover, it has the potential to trigger cell fate decisive actions inside the cell (miRNA/siRNA), just to name a few advantages aside its undoubted relevance in the fundamental biological processes – and its structural difference to DNA. Still it retains the predictable programmability which is known from its counterpart DNA. This applicability in terms of biological function substantially differs RNA from DNA and consecutively encouraged researchers to produce RNA nanostructures with a defined functionality. Recently Rothemund, Geary and Andersen successfully demonstrated the assembly of various RNA origami structures which provided another breakthrough in nucleic acid self-assembly (Figure 6). Here production of RNA origami structures was performed cotranscriptionally in a one pot assembly via T7 RNA polymerase (RNAP). These steps were followed by imaging on mica surface, or transcription, subsequent gel purification and final annealing on a mica surface [36].

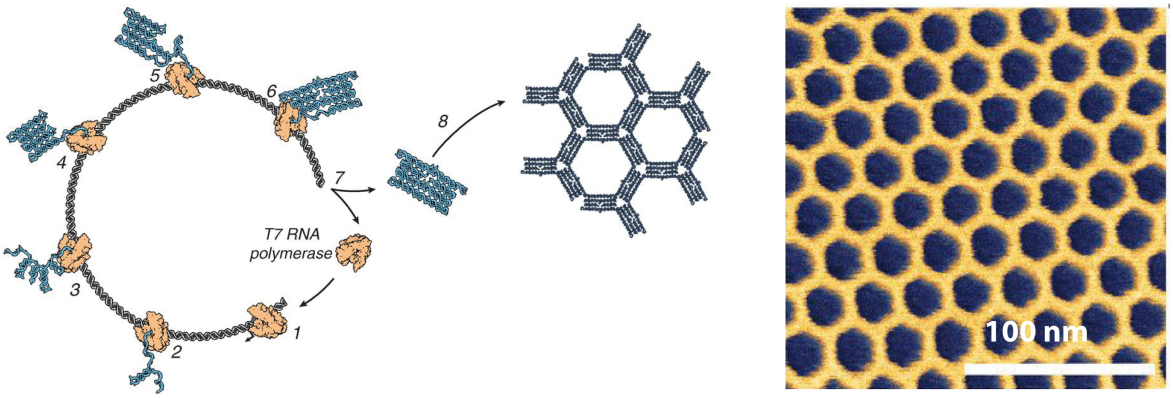


Figure 6. Expansion of the origami method for the creation of RNA structures. Left: Illustration of the co-transcriptional folding pathway initiated by binding of T7 RNAP to template DNA (1), subsequent transcription of RNA and successive folding into its target structure, capable of forming periodic lattices (middle). Right: AFM image of such a lattice [36]. Scale bar = 100 nm.

## Transition towards applications of the DNA origami method

Since it has become possible to successfully design and construct nucleic acid based nanostructures in two and three dimensions, the field was steadily aiming for applications of such devices rather than solely focusing on expanding the structure library. Due to the exact addressability of DNA oligonucleotides within the structure of choice it became possible to position small molecules or nanoparticles like dyes or even proteins, or to modify protruding DNA with a molecule of interest (Figure 7). This allowed a, thus far, unreached accuracy for positioning of chemical compounds [37].

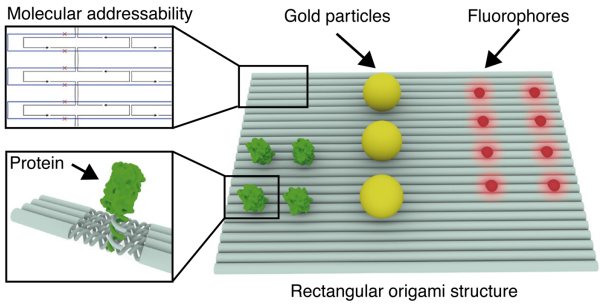


Figure 7. Addressability of synthetic DNA nanostructures and their practicability of nanometer precise positioning of molecular species, e.g. proteins (green), gold nanoparticles (yellow) or fluorophores (red). Top left: Zoom into scaffold and staple strand routing. Bottom left: Hybridization of DNA-conjugated proteins to the DNA origami platform via complementary sequences [38].

Over the last 10 years a broad range of works in structural DNA nanotechnology has taken advantage of the potential of its molecular configurability. One successful example of how DNA nanotechnology can be used to recreate systems with biological function was

demonstrated by an artificial nanopore employing the DNA origami technique. It successfully incorporated into lipid bilayers via DNA strands that had been modified with cholesterol moieties. These synthetic lipid membrane channels enabled electrophysiological measurements and monitoring of translocation of short DNA oligonucleotides [39,40]. Another approach to actually applying DNA nanostructures as pegboards can be found in super resolution microscopy. The possibility to determine specific positions on the DNA origami structure for functionalization and hybridization with oligonucleotides modified with fluorophores offered an innovative approach to study single molecule kinetics on association and dissociation of nucleic acids. Furthermore, it turned out that such described DNA origami structures can serve as *nanoscopic* rulers which now are commercially available [41-46]. Moreover, DNA origami structures were functionalized with linker molecules to which proteins like streptavidin were bound, allowing researchers to monitor reactions via AFM on the single-molecule level [47]. Hence, there has been a consistent shift from solely focusing on structure design towards devices that revealed functionality. Inevitably, once such structures undergo reconfiguration or DNA/RNA-based reactions take place, dynamic processes become increasingly relevant to the community having a great impact on further achievements influencing each other. Especially when reaction components are immobilized onto a DNA nanostructure and intended reactions rely on DNA or RNA strand displacement, the actual structure becomes secondary as it only serves as platform no matter how sophisticated its design is. In the next subchapter, these dynamic processes will be elaborated with regard to project of chapter 3.

The better understanding of thermodynamics of hybridization on the long run helped to develop more sophisticated structures and paved the way for application-based works (see above). A detailed presentation addressing topics like hybridization and nucleic acid chemistry can be found in literature [4,18,19].

### 2.1.3 Dynamic nucleic acid nanotechnology

**From structural to dynamic DNA nanotechnology**

In the previous section, constructions of DNA nanostructures were aimed towards end-time equilibrium states as it was commonly desired to design structures provided with stability and rigidity rather than flexibility and dynamic behavior. Dynamic DNA nanotechnology, in contrast, focuses on non-equilibrium dynamics as one endeavors to construct reconfigurable and autonomous nanodevices. The dynamic properties of such nanodevices are based on rational design of sequence length and sequence specificity, allowing kinetic control of programmable reaction pathways which are characterized by a high degree of orthogonality. This sequence specific addressability distinguishes those devices from previous works that demonstrated switching induced by changes of the environmental conditions like pH, temperature or salt concentration [17,48-51]. The basic operations used in dynamic nucleic acid nanotechnology are strand displacement reactions. This enabled the creation of molecular logic circuits or reprogrammable nanostructures which can be addressed to carry out predetermined tasks similar to reactions taking place in living systems. Description of these devices often uses terms like "*input*" or "*gate*", originating from electrical engineering.
In this thesis, strand displacement reactions play a crucial role as they were used in dynamic nanotechnology as well as in synthetic biology related projects thus underlining its importance within the disciplines. First, they represent the major operational tool for the immobilized DNA strand displacement cascade which was investigated under *in vitro* conditions (chapter 3). Second, strand displacement reactions on the RNA level were shown to enable post-transcriptional control of gene expression in living cells (chapter 4). The performance of molecular devices triggered by strand displacement reactions is designed to be more robust in respect of external conditions like temperature, pH or overall buffer conditions [17]. Moreover, this type of reaction is generally isothermal and does not require enzymes. Nevertheless, the more strands are involved in such reactions, the more complex the task becomes to avoid unwanted cross reactions or to guarantee orthogonality of sequences. Especially, DNA strand displacement reactions became a substantial operation in dynamic DNA nanotechnology and opened up an innovative and steadily growing subdiscipline known as biocomputing. Biocomputing aims to execute operations on DNA, and recently on the RNA level, which, due to its great future promise even became business units of leading software companies like Microsoft research.

Biocomputing aims to control motions of molecular machines and devices based on nucleic acids on the nanoscale by understanding and predicting their specific kinetics [52-55]. Conceptual attempts guiding towards those dynamic aspects of nucleic acid based reactions were achieved already in 1994 by Leonard Adleman who took the PCR reaction as a model and example to study the functionality of nucleic acids and computational processes triggered by DNA, to addressing the directed Hamiltonian path problem [56]. Here the potential of nucleic acid based computation became striking as in theory the limit of possible operations per second may exceed the limitations known from classic computer science. Moreover, the author emphasized the remarkable storage capacity of DNA for an information density of roughly 1 bit per $nm^3$.

Experimental breakthrough developments on the controlled actuation of DNA devices can be traced back to the early 2000's when first functional DNA-based motors and molecular tweezers were built up. These devices were driven by multiple toehold-mediated strand displacement and hybridization reactions. Please refer to Figure 13 on page 30) [57,58]. For the first time, structures constructed out of synthetic DNA oligonucleotides were able to perform continuous association and dissociation cycles by taking advantage of a recently established structural-functional feature termed 'toehold'. A toehold comprises a rationally designed single-stranded nucleic acid moiety which can interact with a set of nucleic acids and induce strand displacement reactions. By incorporating toeholds into the design of nanodevices reaction kinetics of intended operations are considerably sped up, resulting in a thermodynamically favorable configuration. Now it became possible to induce well-defined structural changes on the nanoscale thus opening up new opportunities to the field of dynamic nucleic acid nanotechnology. Since then strand displacement reactions have been widely used in this field of study. Typically, such described reactions are performed in carefully defined reaction conditions and vessels at nanomolar to low micromolar concentrations resulting in reaction speeds from seconds to minutes, whereas in some cases reactions take hours to complete [17]. In following passages, the focus will be on the working principle of the reactions used in dynamic DNA nanotechnology, which in the next section will be illustrated by examples taking up the theoretical concept.

## Theory of strand displacement reactions

A schematic illustration of a typical strand displacement process is given in Figure 9. Before going into details about this reaction, the different ways of illustrating this type of reaction will be highlighted in Figure 8. In most cases the exact sequence is not annotated as it is assumed that the overall reaction potentially should work for a vast combination of nucleotides. One way of illustrating two interacting nucleic acid strands is by presenting its three-dimensional helicity. However, this representation is chiefly used for demonstration reasons (Figure 8A). A more common way to present interacting DNA strands is shown in Figure 8B. Typically, sequences are drawn in directional lines indicating the 5' and 3' ends by signs indicated as arrows, seldom with exact sequence included, see Figure 8C. Here, defined sequences of certain function e.g. hybridization or dissociation within the DNA device are depicted as domains which are given numbers, letters or different colors for differentiation. These domains which are either fully or partially complementary to each other are able to interact as units, and their complementary sequences and domains are either marked in stars or dashes.

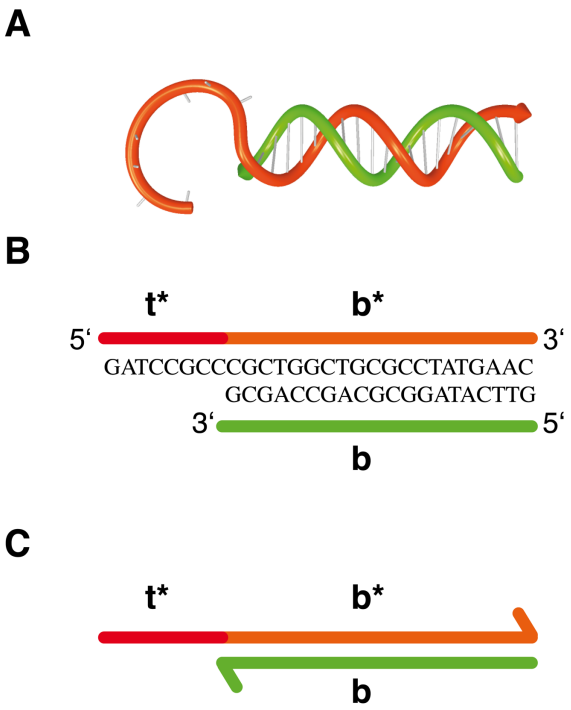


Figure 8. Different forms of duplex strand representation, including one strand possessing a protruding single-stranded toehold domain h* at its 5' end. **A**. Three-dimensional illustration depicting helicity. **B**. Representation by lines with domain names, 5' and 3' ends with exact sequence included. **C**. Most widely used domain representation leaving out sequence. Sequence direction is indicated via arrows.

Each step of a typical strand displacement reaction will be elaborated in the following (Figure 9). Strand displacement reactions as basic logic operations rely at least on three DNA strands which commonly comprise an input strand I, an output strand O and a "gate" strand G to which input I and output O possess a partially complementary sequence domain b (branch migration domain) [53,58]. In the example given below, the initial state represents an incumbent output strand O hybridized to the gate strand building up the initial multistranded "gate" complex C1. In this example the gate strand G possesses a single-stranded protruding overhang at its 5' end, referred as toehold domain t. Present configuration is stable under chosen conditions unless an invading input strand I is added to the system. This invading input strand carries a sequence fully complementary to the gate strand including its toehold domain t. Within this process a binding and unbinding of the input strand to the gate will take place, using the toehold domain t. These binding events are reversible as the newly formed base pairs at the end of the duplex may be disrupted due to thermal fluctuation resulting in its dissociation. Once the toehold has fully bound to its complementary domain h*, input strand I will start initiating branch migration with domain b and will compete with the output strand O for binding to gate G. At this point a three-way branch migration complex exists with invading input strand I and incumbent output strand O both competing for base pairs with the gate strand G. This results in back and forth movement between exchanging base pairs. This process has been characterized as an unbiased random walk as each step causes no net change in base pairing [53,59,60]. Eventually the invading input strand I will have displaced the incumbent output strand O thus completing the strand displacement reaction by the formation the final product denoted gate complex C2. This fully double-stranded duplex represents a thermodynamically favorable state with an overall net gain in base pairs due to the toehold and consequently an enhanced stability compared to the initial configuration.

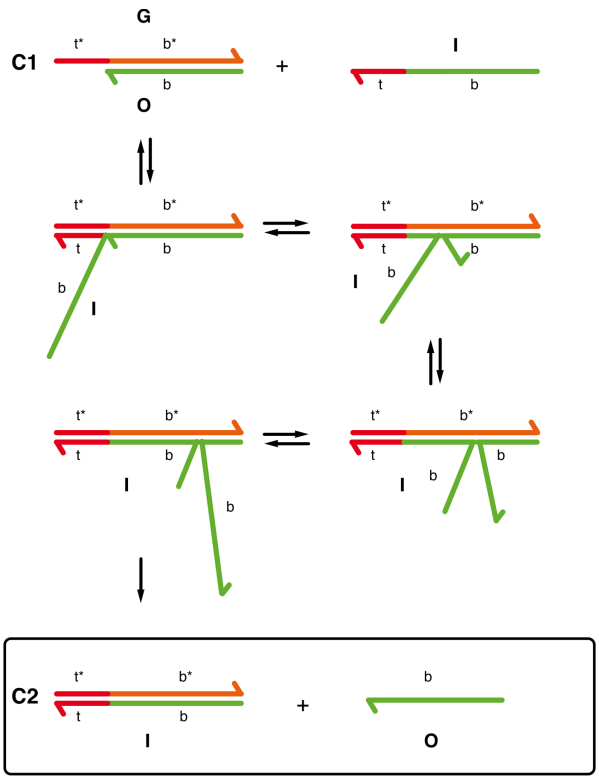


Figure 9. Description of a typical strand displacement reaction. An invading input strand I with toehold t and a branch migration domain b is added to a solution containing a substrate strand to which an incumbent output strand has been hybridized. This complex is termed C1. In this case the invading input strand I binds with its toehold domain t at the 5' end to the gate strand and initiates the branch migration process in a zipper-like mechanism. The final state represents a status in which the invading input strand has completely displaced the incumbent output strand and forms a stable duplex C2 with the input. Figure adopted and redrawn from [61].

### Prediction of $k_{eff}$ by a phenomenological model from the toehold sequence

In this context, it has to be added that strand displacement reactions do also occur within DNA duplexes that to not possess an extended single-stranded toehold domain (known as "toehold free strand displacement"), of course at very low rates though [62]. Taking this under consideration, it is possible to design the kinetics of synthetic DNA reaction networks.

In the simplest case, strand displacement cascades (involving a toehold) will release at least one single-stranded nucleotide product and result in a duplex configuration with the maximal number of hybridized base pairs. These actions are driven by the overall decrease of free energy that is derived from the potential of forming additional base pairs resulting in an overall enthalpy gain and stability. Thus, it becomes obvious that the strength of the toehold highly influences the performance of such described reactions (Figure 10A). This strength is depending on the length (exponentially speeding up the reaction), followed by saturation at a length of 8-9 nucleotides and sequence composition (GC content) [53]. This results in rate constants that vary over a factor of $10^6$ from 1-10 $/M^{-1}$ $s^{-1}$ up to ~ 1-10 x $10^6$ $M^{-1}$ $s^{-1}$.

Experiments investigating kinetics of toehold-free strand displacement reactions resulted in rates of 1.4 $M^{-1}$ $s^{-1}$ at 25°C and 3.6 $M^{-1}$ $s^{-1}$ at 30°C [53,63].

These estimations are of course only valid in the presence of stabilizing counterions like $Na^+$ (high concentration for instance 1M) or $Mg^{2+}$ ions (typically 10-20 mM). Once an input strand possessing a toehold has bound to a gate strand there are two possibilities. Either the toehold pair can dissociate causing the dissociation of the input strand from the gate strand (1) or the nearest base pair of the gate-output complex can fray, allowing for subsequent replacement of the base pair by the input strand and completion of the first step of branch migration (2). Here the authors of this scenario assume a similar rate at which either base pair frays. This process (2) should be half as fast as process (1), as once the gate-output complex frays, there is a 50% chance that the displaced base pair is replaced by the input, and a 50% chance of returning to the initial step [61]. Subsequent forward and reverse steps after completion of the first step of branch migration are assumed to be at the same rate. This aspect was also incorporated into the Intuitive Energy Level, see Figure 11. The life time of the three-stranded intermediate complex comprised of the input and output partially bound to the gate strand is supposed to be independent of concentration. At sufficiently low concentrations, the general reaction can be described as second order process. Thus, the overall effective rate constant ($k_{eff}$) is given by a hybridization rate constant for the toehold, multiplied by a success probability of displacement once the toehold has bound, influenced by the time spent in the tree-stranded intermediate complex.

As the trend of overall exponential acceleration of reaction speed with increasing the toehold up to a certain length has been shown in several studies it strongly suggests that this observation is not depending on a particular sequence.

**Factors influencing reaction kinetics**

However, individual experimental outputs using strand displacement reactions are very likely to fluctuate due to various factors. First, external buffer conditions like pH, salt concentration as well as external factors like temperature have an influence, refer to above mentioned different $k_{eff}$ obtained by Reynaldo and Winfree for toehold-free strand displacement reactions at different temperatures. Recent theoretical considerations on RNA strand displacement reactions also reveal differences for DNA or RNA. Šulc *et al*. have shown that in contrast to DNA, the displacement rate is different depending on whether the toehold is designed at the 3' or 5' end. The 5' end turns out to result in faster reactions. Also the fact that

RNA duplexes are generally more stable than DNA duplexes might have an influence on reaction performance. In those simulations the reaction rate was shown to be sped up by six orders of magnitude when toehold length was increased from one to six bases at 37°C [64]. Moreover the actual nucleic acid species will influence reaction kinetics as thermodynamic stability is different for DNA-DNA, RNA-RNA and its hybdrid complexes [65]. To a certain extent, kinetics also depends on the length of the branch migration domain. The specific sequence composition (different binding energies based on GC content) of the toehold and eventual sequence mismatches make a contribution to reaction kinetics. All these aspects influence the rate constants and have to be considered when designing nano-devices with well-defined properties and functions.

Based on previous findings, a phenomenological model was implemented in order to predict $k_{eff}$ from the toehold sequence [53]. Here, the branch migration process was defined by assuming two macro states, which each represent the first and second half of the isoenergetic branch migration intermediates, associated by rate constant $k_b$ (Figure 10B). Branch migration can be considered as a random walk characterized by many isoenergetic microstates rather than just a single reversible transition on the basis of a first order process [61].

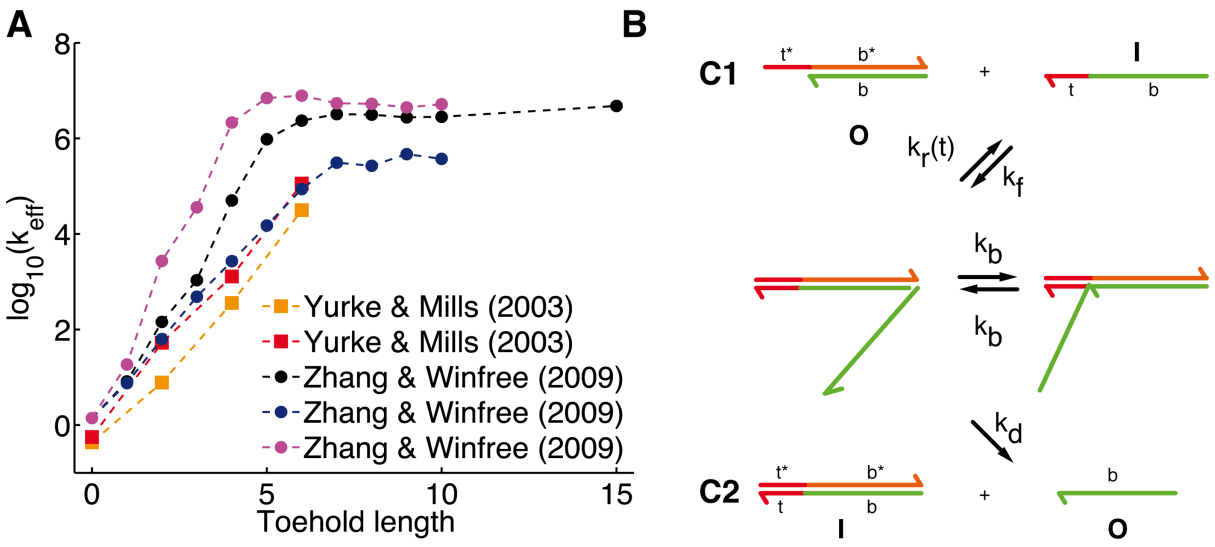


Figure 10. **A**. Effective rate constants ($k_{eff}$) for different toehold lengths based on experimental results ([53], [66]). **B**. Phenomenological model of a typical strand displacement reaction adapted from [53].

**Intuitive Energy Level as a model for toehold-mediated strand displacement**

In their review, Winfree *et al.* introduced the idea of an Intuitive Energy Landscape (IEL) model which mimics the state of a gated multistranded complex with an output strand bound

to a gate to which an input strand is added in to a virtual volume V at defined concentrations [61]. The IEL models the free energy of the virtual box $\Delta G_{box}$. In this particular case (as shown in Figure 11 below), the free energy landscape for a toehold length of six was modeled with given effective rate constants $k_{eff}$. In here, initial state A represents the starting point of the reaction in which gate complex and invading input strand are initially separated from each other and thus are not able to interact until they are both present in solution. The subsequent interaction involves the formation of the initial base pair within the toehold (step B). This zipping up of toehold base pairs, in which each newly formed base pair is denoted as a new state, progresses until the toehold domain has fully become double-stranded (state C). Subsequent branch migration process describes the replacement of an output strand base bound to the gate strand base with an input-gate base pair. This process can be considered a vague reversible stepping between these intermediates which the system requires to pass through a single effective transition state of raised free energy. Each of this branch migration step is represented by a single tooth of the "*sawtooth*" pattern between states C and D, termed as sawtooth amplitude $\Delta G_s$. Herein $\Delta G_p$ describes a plateau height that depicts the variation of free energy of branch migration intermediates within the structure of the branch migration junction. Upon completion of the branch migration process the output strand is fully displaced from the gate strand (state E), resulting in the formation of the final base pair between input strand and gate strand (state F).

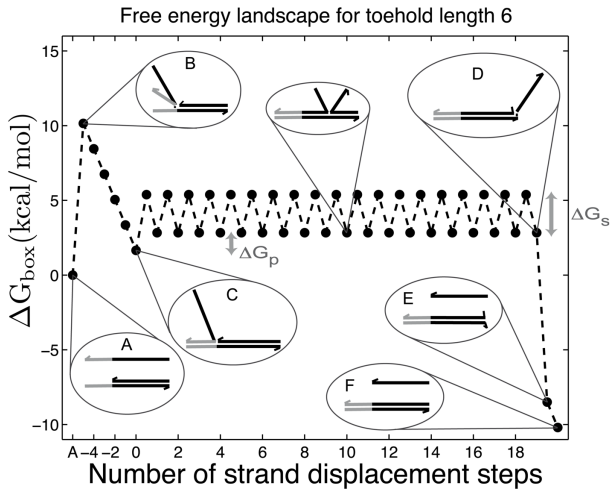


Figure 11. Intuitive Energy Landscape model (IEL). Modeling the strand displacement process step by step starting with an incumbent strand bound to a substrate and an invading strand being separated from each other (state A). Subsequent interaction of the invading strand with the incumbent-substrate complex initiates branch migration (states B – E) and finally results in dissociation of the incumbent strand (state F) [61].

## Toehold exchange reaction

A variation to the above described strand displacement reaction also taking advantage from nonequilibrium DNA reactions is the toehold exchange reaction [53,67]. This reaction allows for better control of strand displacement kinetics, see Figure 12. Similar to original strand displacement, an invading input strand I hybridizes to a toehold of a previously formed duplex C1 resulting in initiation of branch migration process. Both, input and output strand share the branch migration domain b. But in contrast to previously described strand displacement reaction, the incumbent output strand O designed here contains another unique toehold domain t2 that is required to dissociate from gate G until the reaction can be completed. As a result, the originally active toehold t* is sequestered while the formerly inactive toehold t2* is switched to an active state. One main advantage of this mechanism is that this type of reaction is suitable for multiple consecutive toehold-mediated reactions as the activated toehold provides a "new" complex (C2) for another subsequent reaction thus yielding in two reactive products rather than just one. On the other hand, this reaction can occur in reverse direction as the output strand O contains toehold domain t which hybridize to t* input strand I.

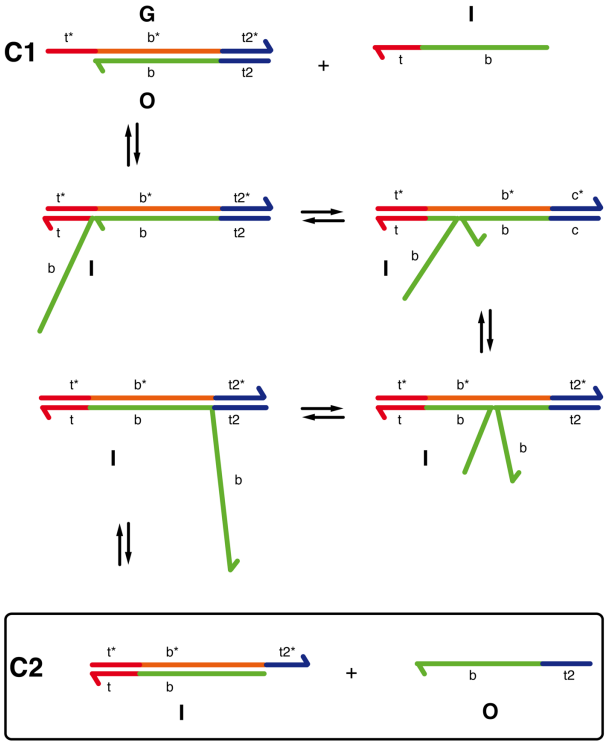


Figure 12. Schematic illustration of a toehold exchange reaction. Binding of the invading input strand I to gate strand G results in subsequent branch migration process and toehold-mediated strand displacement of incumbent output strand O. After dissociation of output strand O, a new single-stranded domain t2* becomes available. In this configuration, C2 can serve as "new" gate for subsequent reactions. Adopted from [53,67].

**Excursion: Strand displacement reactions in nature**

Strand displacement reactions are not only related to DNA nanotechnology. In nature, molecular biology uses strand displacement reactions for instance in DNA repair mechanisms as it is the case for the base excision repair (BER) pathway. Here, replicative DNA polymerases and cofactors perform strand-displacement DNA synthesis reactions in which damaged DNA strands of 2-12 nt are displaced, and replaced by synthesized DNA patches with the correct sequence pattern [68,69]. Furthermore, strand displacement reactions take place during transcription. Here, replicative T7 RNA polymerases are involved in the displacement of the synthesized RNA transcript and subsequent restoration of the DNA duplex [70]. Such strand reactions show distinct different behavior in reaction kinetics *in vivo* conditions, compared to optimized *in vitro* settings [71]. Here, macromolecular crowding results e.g. in an increase of the effective local concentration of reaction partners. Compartmentalization, due to the presence of various organelles possessing different chemical conditions, has a major effect on nucleic acid interactions such as strand hybridization. On the other hand, it is obvious that DNA and also RNA interacting proteins will either speed up or slow down such discussed reactions in either specific or unspecific manner.

### 2.1.4 Applications of DSD reactions in dynamic DNA nanotechnology

In the previous chapter, the concept of predominant reactions applied in dynamic DNA nanotechnology has been covered with emphasis on strand displacement. In the following a selection of some key publications will be presented which predominantly use strand displacement reactions to give nucleic acid structures a defined functionality, or to induce specific change in conformation. This can be achieved by the interplay between single- and double-stranded domains. Some of them apply strand displacement reactions to rationally drive reconfiguration of DNA devices. Others trigger molecular circuits and logic gates, feeding into the subject of biocomputing in which sensing and regulation plays a role. By applying the benefits of structural and dynamical DNA nanotechnology, each taking advantage of the sequence-programmability of nucleic acids, it is also possible to investigate cascaded reaction pathways thus exploring the potential benefits of colocalization. Even though past years produced many devices of which its motion is triggered by change of buffer conditions, such reactions affected all present structures and made a controlled actuation of sequence specific movements impossible. In particular, this applies to multi-gate systems that require a high degree of (sequence-specific) orthogonality. The introduction of strand reactions for triggering the motion of nanodevices successfully addressed this problem. Performance of these types of devices entirely relies on the interaction of specific nucleic acid sequences. Control of its intended motion and function is independent of external conditions such as salt, pH or temperature. When introducing this subchapter, Adleman's pioneering work that considered DNA as suitable 'material' for doing sophisticated computation on the molecular level is again worth to be mentioned. For this he was using the well-established PCR reaction and subsequent gel electrophoresis [56]. His work on recognizing DNA molecules as suitable tool for storing and processing information can be considered as precursor in DNA computing and subsequent advances in the field of dynamic DNA nanotechnology.

**Switchable DNA devices**

Similar impact to the community in the field of dynamic DNA nanotechnology had the design and construction of the first reconfigurable DNA machines by Yurke and coworkers, published in the early 2000's which at this point shall be highlighted, see Figure 13 [58,72,73]. The implementation of what they called "molecular tweezers" can be considered as pioneering work in the development of reconfigurable DNA machines. The original design of

the tweezer encompasses three hybridized DNA strands (A, B, C) (Figure 13A). This tweezer changes its conformation from an open state to a closed state upon addition of so called fuel strands F. These interact and hybridize to the structure via toehold domains located at its 5' and 3' ends, respectively. This process is reversible and reopening of the tweezer is achieved by the addition of antifuel strands $\bar{F}$ that initiate toehold-mediated strand displacement reactions resulting in the return of the tweezer into its open configuration underlining its dynamic capabilities (Figure 13B). This switching process has been monitored in fluorescence spectroscopic measurements (Figure 13C).

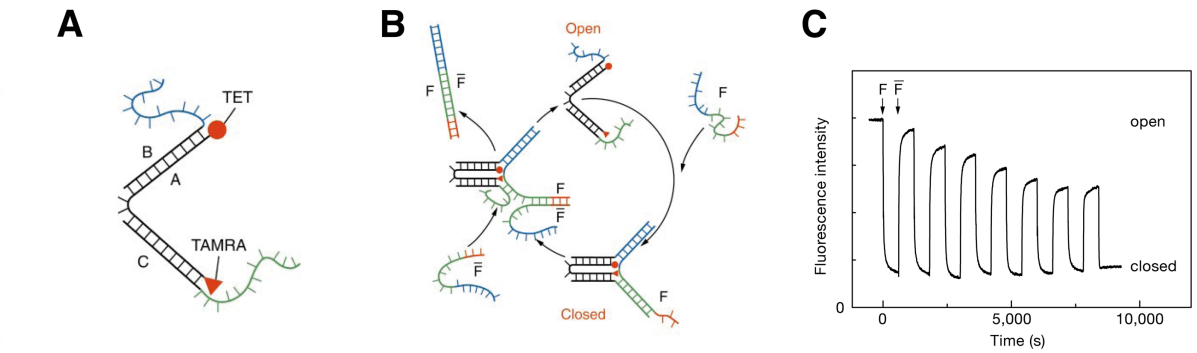


Figure 13. Construction and operation of the molecular tweezers. **A**. Molecular tweezer structure formed by hybridization of oligonucleotide strands A, B and C. **B**. Closing and opening mechanism of the molecular tweezers. Closing strand F hybridizes with the dangling ends of strands B and C (shown in blue and green) to pull the tweezers into a closed state. Toehold-mediated strand displacement of F (red) allows antifuel strand $\bar{F}$ to remove F from the tweezers, forming a double-stranded waste product F$\bar{F}$. This allows the tweezers to open up again. **C**. Repeated cycling of the tweezers by adding stoichiometric quantities of closing and removal strands F and $\bar{F}$ in sequence, monitored via fluorescence spectroscopy [58].

In the same era, Seeman and coworkers demonstrated a nano-mechanical DNA-based system that allowed to interconvert and fuel a paranemic crossover (PX) DNA motif into its topoisomer JX2 DNA (characterized by two adjacent sites where backbones juxtapose without crossing over), where one strand end is rotated relative to the other by 180° upon addition of rationally designed DNA strands resulting in structural-dynamic reconfiguration [74], see Figure 14. The authors termed such reaction "strand replacement reactions". In following years this approach served as model for more advanced devices [75].

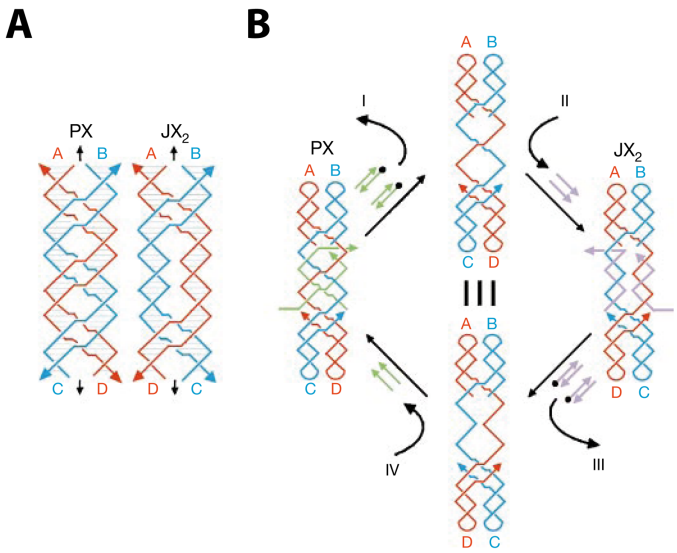


Figure 14. Representation of PX and $JX_2$ structures and the switching process of interconversion. **A**. Schematics of PX and $JX_2$ structures. The PX motif is composed of two helical domains and four individual single strands held together by the maximum of crossovers, whereas the $JX_2$ motif lacks two crossovers in the middle of the structure. In its end states, the bottom of the $JX_2$ structure is rotated by 180° in relation to the PX motif. **B**. Reaction pathway of the interconversion between PX and $JX_2$ is mediated by cascaded strand replacement reactions. Initially present green strands in the center of the PX motif are removed by addition of its complementary counterparts (state I). Conversion of resulting intermediate into $JX_2$ is initiated by subsequent introduction of purple set strands (state II). Switching back to PX motif is achieved by first removing purple strand resulting in an again flexible intermediate (state III) and subsequent addition in of green set strands (state IV). Modified from [74].

Inspired by these groundbreaking works, following years generated quite a few dynamic nano-mechanical devices being triggered by such DNA strand displacement reactions [17]. In parallel, also in the field of DNA computation such operations contributed to substantial progress and recognition within the community of which in the following some subtopics featuring selected illustrating examples will be brought up.

**Dynamic three-dimensional DNA devices controlled via displacement reactions**

One approach taking advantage of the synergistic aspects of structural and dynamic DNA nanotechnology with its implementation of displacement reactions is aiming towards real-world applications like controlled reactions or therapeutic functionality. A well-known example represents a DNA nanorobot. It was designed to be able to transport molecular payloads, in this case gold nanoparticles as well as fab antigen fragments, into cells via a key lock mechanism. The payloads are released using a logic gate that is conditionally triggered via an aptamer lock mechanism (Figure 15B). Loading this robotic nanodevice with specific antibody fragments and its subsequent release and stimulation after target binding even induced signaling pathway activation [76]. This idea widened the field of dynamic DNA devices

into potential “theranostic” applications. Certainly, this work was inspired by a previous publication presenting a DNA origami cube of which its lid could be opened upon addition of DNA strands binding to sticky end extension protruding the cube thus being able to open the structure in a controlled fashion via DNA strand displacement reactions (Figure 15A) [77]. Recently, Andersen and coworkers demonstrated a reconfigurable DNA vault that relies on the interface of structural and dynamic DNA nanotechnology and is able to be reversibly opened or closed [78] (Figure 15C and D). This switchable three-dimensional DNA container was designed using the DNA origami method and shows characteristics of previously described DNA containers [33,76]. In the inside of this DNA vault a protease (in this case bovine alpha-chymotrypsin) is inactive as long the vault remains in its closed state. Closing of the structure is achieved via addition of closing key strands which upon addition of complementary opening key strands can open via toehold-mediated strand displacement reactions and initiate enzyme activity. The dynamic behavior of this device becomes evident as the enzyme substrate reaction is controlled by a reversible multi-lock system allowing its rational opening or closing, induced by conformational changes. Moreover, it has been shown that enzyme activity is efficiently shielded in its closed state thus leading towards efforts to compartmentalize reactions and control them in a directed way. Devices such as those described above have to unify aspects of structural as well as dynamical DNA nanotechnology, resulting in well-defined functions, especially target specific release.

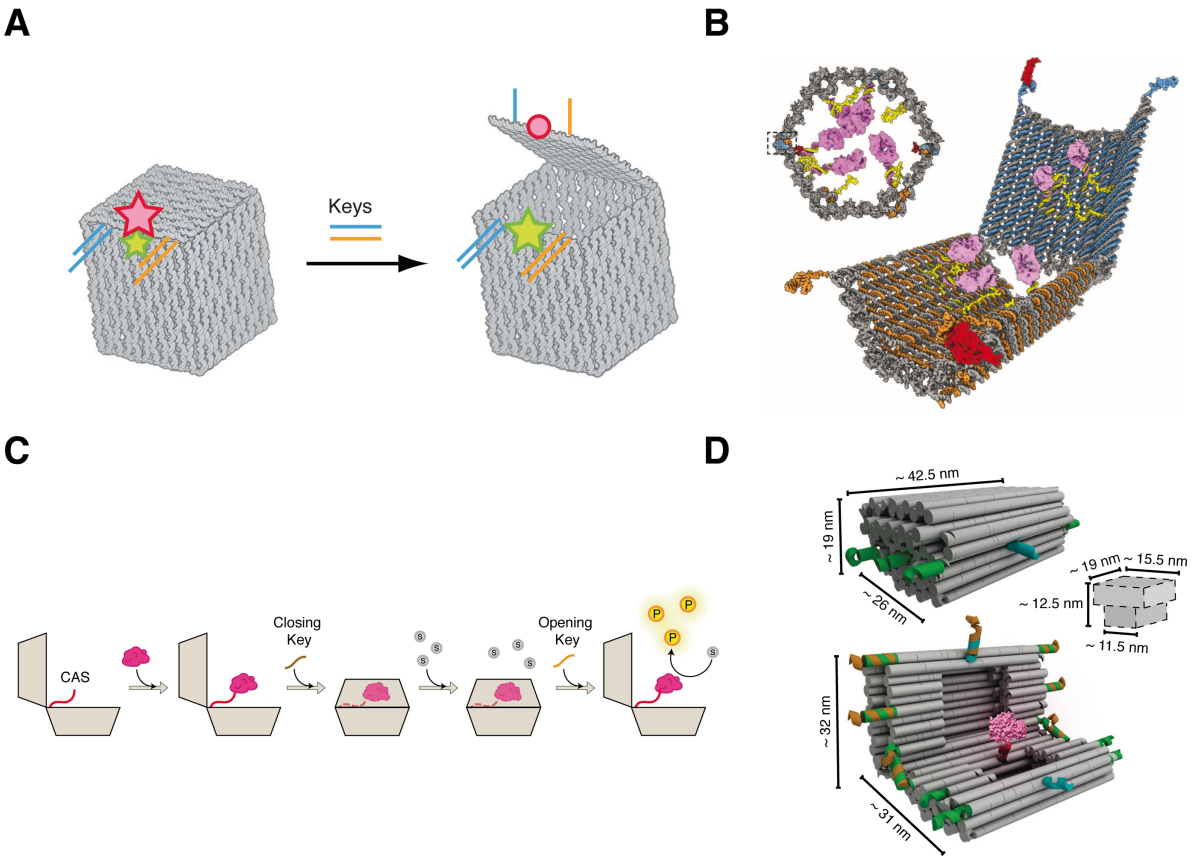


Figure 15. Dynamic, switchable three-dimensional DNA nanorobots fabricated by the DNA origami technique and triggered by DNA strand and protein displacement reactions. **A**. Nanoscale box with a controllable lid. Opening of the box occurs via adding key strands which displace previously bound short oligonucleotides from the key-lock mechanism (blue and orange lines). The opening mechanism is observed by a FRET pair (red and green symbols, Cy5-Cy3) [77]. **B**. Perspective illustration of a logic-gated nanorobot. Opening and closing is controlled by an aptamer lock mechanism. Protruding DNA aptamers are initially hybridized to a partially complementary strand (depicted in blue and orange, respectively). The nanorobot opens upon presentation of specific antigen keys (red) being recognized by the aptamers via protein displacement, resulting in release of molecular payloads (purple) [76]. **C** and **D**. Function and design of the DNA vault. In its open state the structure is loaded with an enzyme (purple) at the cargo-anchoring site (CAS) and subsequently closed by addition of a sequence-specific closing key (brown). Reversible opening of the vault and subsequent initiation of enzyme activity (represented in yellow products) occurs via addition of opening keys (orange) and toehold-mediated strand displacement reactions. Three-dimensional model of the DNA vault in its closed and open configuration with dimensions of the cavity included. Modified from[78].

## Logic gates and circuits

In analogy to electrical engineering, logic elements and signal processing units were increasingly implemented into dynamic DNA nanotechnology. In 2006, Seelig *et al.* designed *in vitro* DNA-based logic gates and circuits composed of rationally designed single strands interacting as input or output via hybridization and toehold-mediated strand displacement, see Figure 16 [79]. They implemented AND, OR, and NOT gates including signal amplification, restoration and cascaded reactions. Gates and respective sequences were designed such way that disease related sequences like microRNAs could serve as input which upon strand

displacement reactions release an output strand. The output in turn can be used as an input to feed a cascaded downstream gate thus allowing a system of high orthogonality. This type of biocomputing opened the potential for diagnostic applications like sensing of multiple therapeutic relevant DNA or RNA species in parallel.

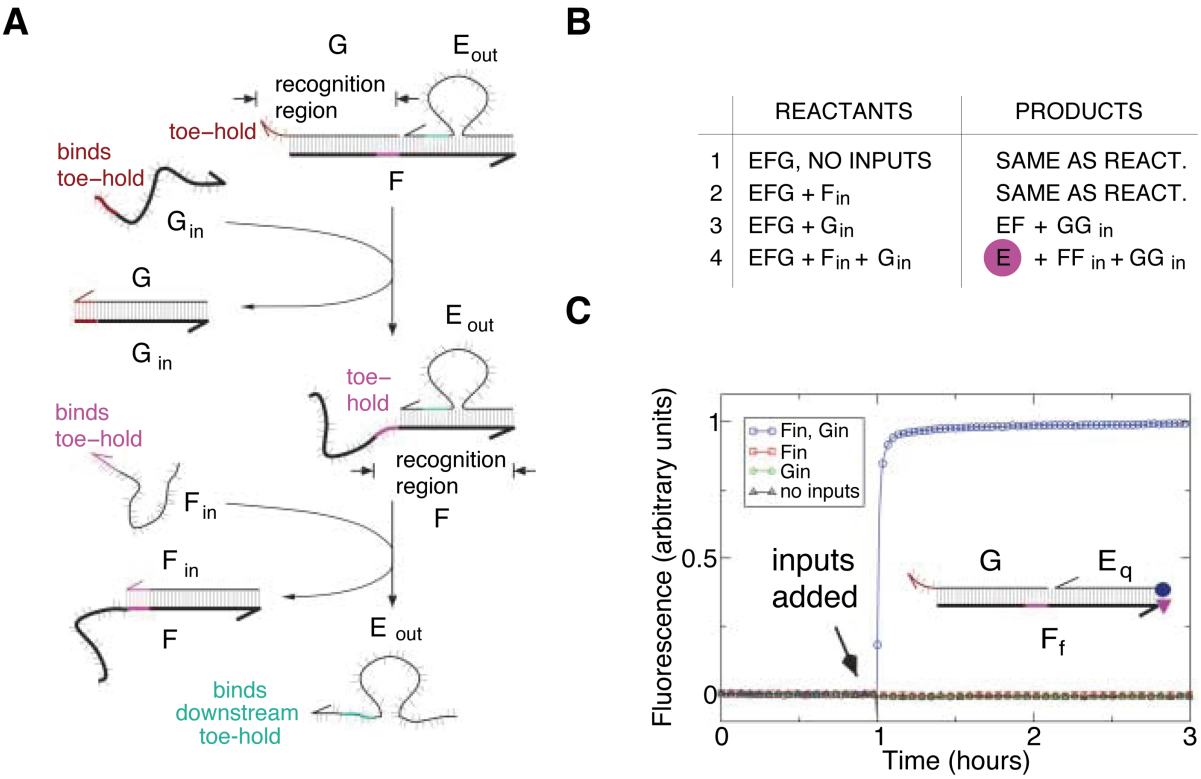


Figure 16. Implementation of a two-input AND gate. **A**. The initial gate is composed of three DNA strands, G, F, $E_{out}$ which initially form a stable complex. Toeholds (colored) enable subsequent strand displacement reactions of strands G and F leaving single strand $E_{out}$ accessible for potential cascaded downstream reactions. **B**. Truth table for the two input AND gate. **C**. Typical fluorescence spectroscopy experiments with respective strands modified with fluorophore and quencher molecules for readout. Modified from [79].

A similar approach using short DNA nucleotides as material for the creation of digital multilayer molecular circuits based on strand displacement cascade was published five years later by Qian and Winfree. They designed an expandable DNA circuit system based on a simple DNA gate motif nowadays known as seesaw gate. It is triggered by toehold-mediated strand displacement reactions enabling a functional scalable system. The result was a four-bit square-root circuit comprising as much as 130 DNA strands. Their design includes fundamental elements for the design of large-scale circuits opening the potential for advanced biosensing purposes. Qian, Winfree *et al.* applied this approach to design a chemical Hopfield network aiming to use DNA strand displacement cascades for neural network computation [80].

**Excursion: Compartmentalization and colocalization of reaction components in nature**

In higher ordered organisms, compartmentalization can be traced down from the whole body to organ systems which are composed of individual organs, that in turn are built up by different tissues. On this level one can further distinguish between different cell types being responsible for conducting clear-cut tasks distributed onto individual cells which itself are comprised and characterized by a complex intracellular network. The interested reader will find dozens of works and in depth-literature addressing each level of complexity. In a molecular biological context on subcellular scale, spatial colocalization and compartmentalization represent crucial aspects and were crucial for the development of higher order organisms as it allowed to conduct complex reactions in well-ordered ways within distinct chemical and physical environments and conditions. Undoubtedly major metabolic processes also fall into this regime. Herein, especially separation of intra and extracellular reactions as well as spatial organization in organelles found in eukaryotic organisms and plants provides the opportunity to perform tasks in a chemically enclosed atmosphere (Figure 17A) [81,82]. Compartmentalization is also accompanied with enclosed environments of different pH and distinct osmolarity which applies for specific organelles like mitochondria. This allows for performing chemical reactions that would be unfavorable in one subcellular compartment whereas in other compartments such reactions turn out to be crucial for proper cellular function.

In contrast to the well-defined and controllable environment of buffers or reaction mixes in a test tube, the cell represents a crowded environment packed with a vast number of macromolecules calling for strategies that allow multiplexed processing of information. Indeed, macromolecular crowding, which is also known as the excluded volume effect, significantly contributes to kinetic as well as thermodynamic properties of respective macromolecules and has a big impact on its effective concentration. The effective concentration can be much higher than its respective actual concentration, resulting in enhanced reaction speeds [83]. A similar aspect addresses coordinated spatial organization of reaction components and its arrangement within a scaffold. This especially comprises scaffolded protein complexes (Figure 17B). Protein signaling and phosphorylation pathways like the PI3K/AKT pathway which are e.g. involved in proliferation and replication are structurally connected to each other. Signaling cascades follow as kinase cascades or multienzyme complexes minutely control (de-)phosphorylation steps. Mutations of involved components can result in dysfunctionality such as uncontrolled phosphorylation events which

are attributed to diseases like cancer [84,85]. In many cases these complex signaling cascades are associated with the cell membrane as the input originating from outside the cell will trigger downstream processes producing an output that causes major changes in the cytoplasm. Another example of the relevance of colocalization in nature can be found in assembly-lined processes (Figure 17C). The example shown here highlights the relevance of the Hsp70-Hsp90 Organizing Protein (HOP) which plays a crucial role within the folding process of proteins by spatially bringing together Hsp70 (Heat shock protein 70) and Hsp90 (Heat shock protein 90). Often, proteins are composed of multiple subunits or represent multifunctional enzyme systems being only functional as an ensemble, e.g. the tryptophan synthase complex or the AROM complex which is a pentafunctional polypeptide [86,87]. Spatially arranging reaction components that are only functional when connected, and channeling of intermediate products between each of the active sites has been shown to speed up interactions. This is thought for above mentioned tryptophan synthase complex as it channels the metabolic intermediate indole in the inner part of the enzyme complex. These macromolecular complexes are highly organized in which colocalization plays a major role in order to guarantee streamlined reactions. As a consequence, colocalization of reaction compounds and its intermediates can be a model for biologically inspired engineering either for optimizing and upscaling of biocompounds for industrial purposes or in the field of nucleic acid nanotechnology.

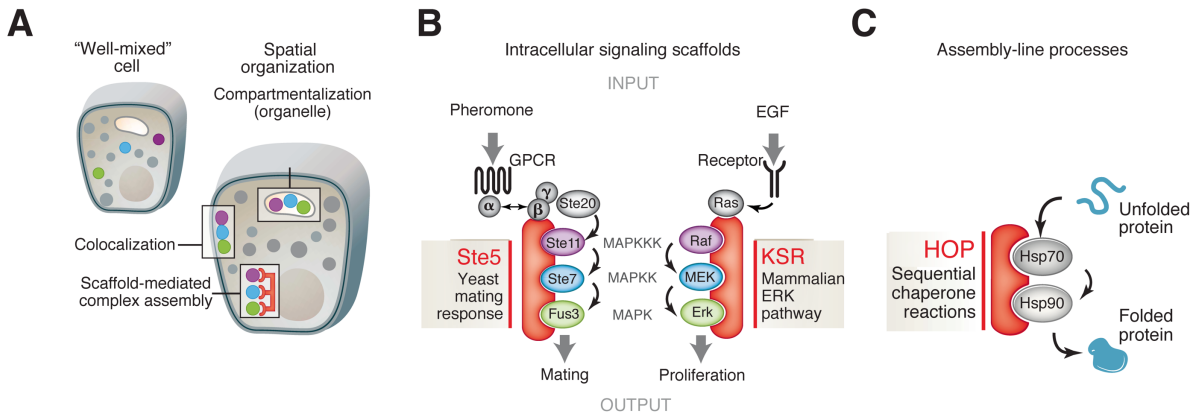


Figure 17. Cellular compartmentalization and well-defined spatial organization of reaction component flux. **A**. In the crowded environment of a "well-mixed cell", spatial organization is fundamental for efficient processing of intracellular signals. Colocalization is also related to assemble proteins into induvial complexes e.g. either by colocalization or scaffold-mediated complex-assembly. **B**. Left: Streamlining of intracellular signaling by protein scaffolds with an input and output "signal" signal. Two canonical examples are shown. Within the yeast MAP (Mitogen-activated protein) kinase pathway, Ste5 protein plays a crucial role in holding together associated kinases Ste11, Ste7 and Fus3, and consequently in the regulation of downstream (de)phosphorylation events of MAPK cascade. This cascade is activated upon binding of an adequate pheromone to a G protein-coupled receptor (GPCR). Right: KSR protein represents a crucial element in the mammalian Ras-Raf-MEK-MAPK pathway. This scaffold protein recruits and binds kinases Raf, Mek and Erk and positively regulates proliferation, with coupled phosphorylation events upon binding of an external growth factor (EGF) to accordant receptor. **C**. Involvement of scaffold proteins in protein folding. HOP protein (Hsp70-Hsp90 Organizing Protein) functions as co-

chaperone by linking together Hsp70 (Heat shock protein 70) and Hsp90 (Heat shock protein 90) which both are essential molecules in protein folding. Adopted from [84].

## Theoretical considerations

Of course, comprehensive theoretical studies were carried out on how colocalization as well as clustering of reaction components like proteins impact overall reaction flux and activity [86]. For instance, Gerland and coworkers studied the reaction diffusion model of a collaborating two-enzyme system with a sender enzyme converting a substrate into an intermediate which in turn serves as substrate for an acceptor enzyme [88,89]. Taking under consideration the ratio between the diffusive time-scale $R^2/D$ and reactive time-scale $\sim K_M/k_{cat}$ of the acceptor, clustering of enzymes turns out to be beneficial for fast enzyme turnover rates. Well-ordered, tightly clustered enzyme arrangement favors settings when diffusion is faster than the second enzyme reaction as broadly distributed backup enzymes can "catch" intermediates which are able to diffuse away. Thus, timing of reaction and diffusion events determines the optimal enzyme distribution. Taking nature as a model has usually been a good choice when it comes to optimizing processes in scientific questions. Some studies mentioned later take advantage of the circumstance that DNA scaffolds offer an elegant way for the precise arrangement and colocalization of molecules. For example, in DNA-templated synthesis unfavorable reactions can be sped up by placing reactants into close proximity onto a DNA scaffold [90].

As delineated above, colocalization represents a crucial mechanism in (cell)-biology. It seemed natural to use such described scientific achievements to study well-defined positioning of reaction components on a DNA origami surface. In 2009, Wilner and coworkers reported the first study using rationally designed DNA strands as scaffold to position the well-known glucose oxidase (GOx)/horseradish peroxidase (HRP) enzyme pairs [91]. Three years later, Yan *et al.* expanded this field by using DNA origami rectangles to immobilize the same enzyme system, in order to analyze their activity in dependence of their relative distance from 10 – 65 nm [92], see Figure 18A. It turned out that enzyme activity is strongly depending on colocalization with high activity for a spatially packed configuration, and low activity for reaction partners placed far away from each other. The same multienzyme system was encapsulated in a modular DNA origami tube. The result was an enhanced enzyme activity within this channel once both enzyme units were confined by linkage of the two structural modules, each containing either GoX or HRP, see Figure 18B [93]. Another example represents an enzyme cascade for which the reaction activity of individual enzymes could be triggered by changing the pH of the solution [94].

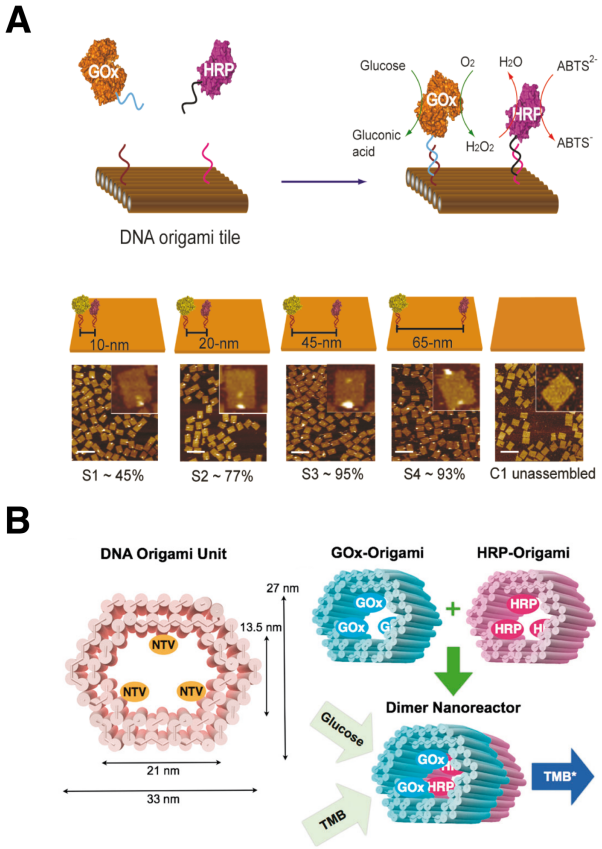


Figure 18. Immobilization of the GOx/HRP enzyme cascade system onto synthetic DNA scaffolds **A**. Immobilization of the enzyme system on a rectangular DNA origami structure with different distances between sender and receiver enzymes [92]. Scale bar = 200 nm. **B**. The same enzyme system was investigated by encapsulating it into three-dimensional DNA origami structures. Each enzyme was immobilized in a separate DNA origami container which could be connected by hybridization of protruding DNA strands [93].

In the following, two recently published works will be proceeding with this topic by focusing more on biocomputing and immobilization of logic circuits on DNA scaffolds (Figure 19). Beforehand, tasks were designed and carried out along trajectories. Again, Seelig and coworkers refined the state of art in dynamic DNA nanotechnology by creating modular logic circuits on a DNA origami platform (Figure 19A). They designed a localized signal propagation mechanism by spatially arranging reactive DNA hairpins on a DNA origami platform that was exploiting the DNA domino architecture. Here spatial organization was applied to perform arbitrary logic on the nanoscale. In its most simple configuration, the circuit is composed of a minimal two-hairpin wire consisting of an input and output hairpin each attached to the DNA origami. The circuit cascade is initiated upon addition of an input strand which binds and opens the first hairpin by toehold-mediated strand displacement. In each step one hairpin stem is unwound and by this a previously sequestered toehold within the hairpin loop becomes available to initiate opening of a subsequent hairpin stem. By modifying

the distance and spatial distribution between input and output a strong dependency in circuit performance could be detected. This could be further adjusted by placing intermediate hairpins along wires of different lengths and orientations. Immobilization and confinement of circuit components on a DNA platform led to a drastic reduction of computation time, compared to circuits freely diffusing in bulk solution. In another recent study, a cargo-sorting DNA robot system was presented which is capable to carry out cargo sorting at the molecular level, see Figure 19B [95]. This system is composed of three modular building blocks capable of carrying out autonomous tasks by random walk without external energy supply. Here reaction partners were also immobilized on a DNA origami substrate and operation of the robot were relying on DNA strand displacement reactions. Reaction processes were designed to be run by a simple algorithm.

Apparently these two works were certainly inspired by the work presented in this thesis as they applied similar design features and were motivated by similar ideas. Moreover, rate constants obtained in this publication confirmed the findings presented in this thesis.

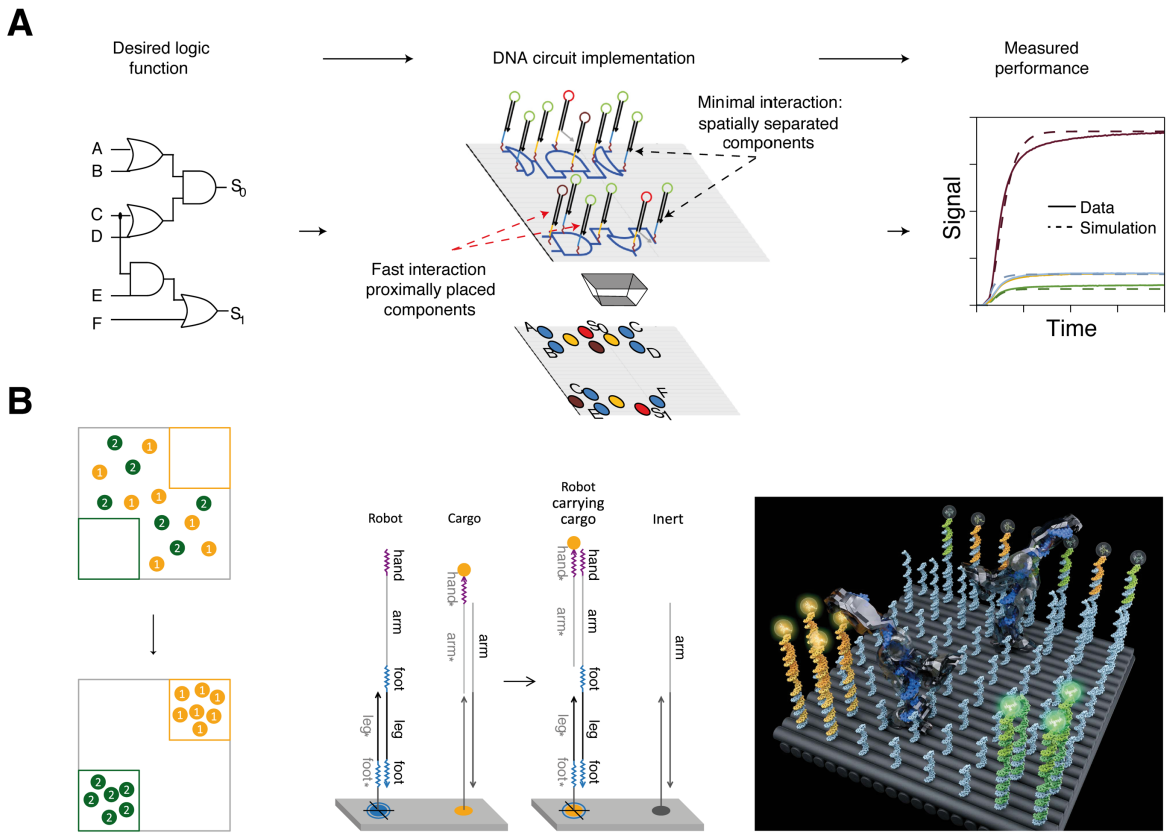


Figure 19. **A**. Immobilization of a logic gate structure on a DNA origami platform. Controlled signal propagation by spatial organization. Left: Schematics of a logic circuit. Middle: Realization of a molecular logic circuit by rationally immobilizing hairpin components on a DNA origami substrate. Circuit component interaction strongly depends on its spatial distribution on its pegboard. Middle bottom: Use of sequence-specific hairpin structures (depicted as different colored circles) for realization of a molecular circuit. Right: Typical experimental output and comparison of the data with simulation. Adopted from [96]. **B**. Immobilization of a DNA robot which is capable to conduct sorting and transportation of cargo from initially arbitrary spots to separated locations. Left: Layout of the sorting mechanism by placing randomly distributed molecules into

well-defined compartments within the structural topology. Middle: Schematics of strands and domains involved in correspondent cargo pick-up process. Right: Three-dimensional model of the cargo sorting molecular robot. Modfied from [95].

## Readout of strand displacement reactions

On the experimental level, results of works presented in this thesis are mostly based on fluorescence readout and analytical methods like polyacrylamide or agarose gel electrophoresis. As quantitative method of choice, fluorescence-based measurements deliver accurate results for precise kinetic readout. Synthetic nucleic acids can be readily synthesized containing fluorophore labels or stained with intercalating molecules or dyes. In here, appropriate fluorophore quencher pairs are used as well as FRET (Fluorescence Resonance Energy Transfer) pairs which can deliver additional information like the relative position of strands or the hybridization state. From these observations, kinetics and relevant rate constants can be derived and serve as benchmarks for present and future projects in (dynamic) nucleic nanotechnology.

## Closing remarks

All these aspects of colocalization combined with analysis and fabrication tools are offered by DNA nanotechnology. This includes the DNA origami technique as well as the knowledge of strand displacement reactions, plus nature's perfectly working role model. These aspects were taken into account for the design of a localized DNA strand displacement cascade on a DNA origami substrate (chapter 3). The clustering of reaction components, especially DNA molecules that interact with each other provided valuable insights and knowledge about the benefits of using DNA nanotechnology to aid the colocalization of biomolecules in *in vitro* conditions.

In my PhD thesis, DNA nanotechnology was, for the first time, employed to experimentally study dynamic processes under consideration of spatial arrangement on DNA origami platforms by designing a two-stage DNA strand displacement cascade. The concept applies benefits and tools from both, structural and dynamic DNA nanotechnology by exploiting the potential of colocalization of reaction components onto a defined geometric entity. This system, which can be considered as a precursor for more recent studies (Figure 19) consists of a "sender" and a "receiver" gate immobilized on a rectangular DNA origami platform. For the first time, this idea enabled us to study the effect of colocalization on the performance of DNA strand displacement and branch migration reactions by changing the distance and

arrangement of the gate complexes. Besides studying strand displacement reactions *in vitro*, this process has also gained importance in *in vivo* experiments within the field of synthetic biology. Here strand displacement was used as a tool to control gene expression on the post-transcriptional level. This topic will thoroughly be discussed in succeeding subchapters.

## 2.2 Molecular biology essentials

### 2.2.1 Revisiting central dogma of molecular biology

Understanding the origin of life represents one of the major intellectual goals of humankind of which its endeavors showed up in manifold directions. Undoubtedly the identification of nucleic acids by Friedrich Miescher in 1871 paved the way towards better understanding of life on a molecular level. He isolated a precipitate originating from the nuclei of white blood cells and called this compound "*nuclein*", referring to the Latin term "*nucleus*" of which its literal translation is "*core*" [97]. In that early time the exact function of nucleic acids was still unknown, but since then a lot of research was dedicated to solve the fundamental processes of life on a molecular level [98]. In 1928, Frederick Griffith, a British bacteriologist successfully demonstrated that genetic material can be transferred between bacteria. He was working with two different pneumococci variants and proved that the exchange of genetic material can turn a previously non-pathogenic strain into a pathogenic strain by a process which he termed transformation [99]. Nowadays, transformation of bacteria is considered as a fundamental working procedure in molecular biology. 16 years later, Oswald T. Avery, Colin MacLeod and Maclyn McCarty provided evidence by further investigation of the transforming capability of pneumococci that DNA must be the molecule responsible for the results obtained by Griffith before, and not proteins or any other compound [100]. In the same era, Erwin Chargaff and coworkers verified that the composition of DNA including distribution of purines and pyrimidines as well as respective sugar moieties remain similar throughout different organs (in this case thymus and spleen) [101,102].

Subsequent infections of bacteria with bacteriophages, performed by Alfred Hershey and Marta Chase proved DNA and not proteins are the molecules responsible for entering a bacterial host cell, consequently being subject for inheritance [103]. In parallel, researchers were also dedicating their efforts on solving the structure of DNA. Finally, this was achieved by James Watson, Maurice Wilkins and Francis Crick in 1953 together with major contribution by Rosalind Franklin and Raymond Gosling when they interpreted X-ray fiber diffraction patterns of DNA molecules – actually produced by Raymond Gosling – as a double helical structure [1].

In this time, Crick already postulated the "Central Dogma" (of gene expression) in which he stated that "*once information has got into a protein it can't get out again*" (Figure 20).

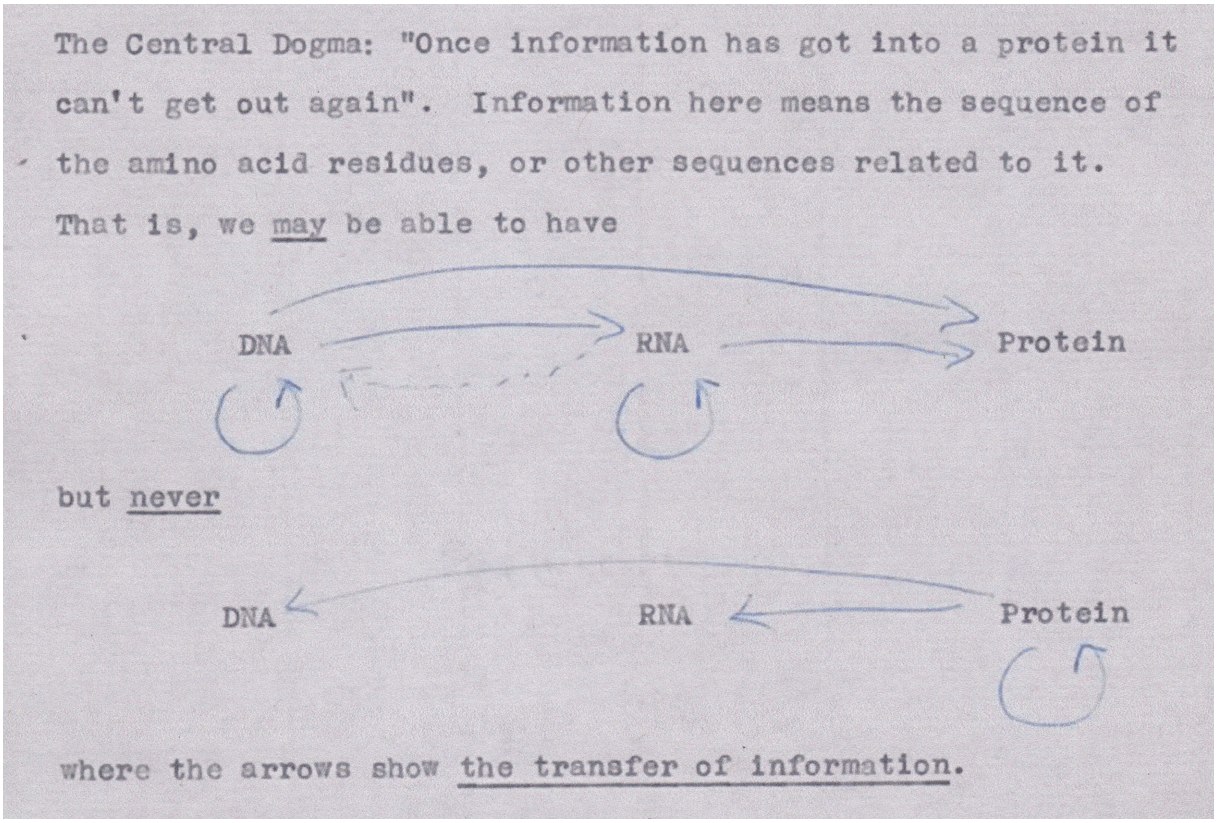


Figure 20. Original drawing from an early draft made by Francis Crick [104,105]

Subsequent successful attempts in understanding protein synthesis and the genetic code were achieved in various decoding experiments such as the ground breaking poly-U experiment by Marshall W. Nirenberg and J. Heinrich Matthaei in which they deciphered the first of total 64 RNA triplet codons (UUU triplet) into an amino acid (phenylalanine), hence doing fundamental work on protein synthesis research [106]. Later, Francis Crick expanded his idea of the "central dogma" and postulated the famous statement that "*the central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to either protein or nucleic acid*" [107]. His early drawing illustrates the flow of genetic information and the pathway of protein synthesis. It is proposing that transfer of genetic information is only possible from one direction starting at the level of DNA, being the template for RNA, which finally is converted into protein (and not in the other direction).

All these pioneering works presented among others deserve special attention and each in an individual manner can be considered as the starting signal for modern molecular biology on the structural and physiological level of DNA, RNA and proteins, of course keeping in mind their connection to each other.

In the following, each step of this process will be addressed starting with DNA as material possessing the capability to be duplicated in a process called DNA replication as well as being transcribed into RNA which finally serves as template for the synthesis of proteins (see "poly U-experiment") in a process known as translation. Transcription and translation can be summarized by the term "gene expression". Herein special attention will be drawn to initiation of translation including major differences in prokaryotic and eukaryotic systems as post-transcriptional control of gene expression attires special attention in chapter 4.
On a cellular basis, life can be classified by the kingdom made up by following domains: archaea, bacteria and eukaryotes. Archaea and bacteria are combined within the class of prokaryotes, which without going into detail differ from the class of eukaryotes by the lack of a nucleus [108]. The nucleus as a special compartment mainly contains the DNA which itself, recalling its double helical character prompts to be the molecule suitable for inheritance of genetic information. The process of inheriting genetic information, which was already proposed by Crick in his drawing (indicated by the loop below "DNA") can be considered as the starting point in evolution and the complex machinery of life. DNA replication happens in all living organisms and will briefly be adumbrated in following paragraph.

**DNA replication**

During cell division, two identical DNA replicas are produced originating from one starting DNA source. The starting point usually represents a double-stranded DNA molecule which within the process of cell division is duplicated in a semiconservative manner. The double helix is unwound by helicase enzymes. These form a structure known as replication fork (or "bubble") providing access to each of the two single strands of the template. Subsequently, DNA polymerases (DNAP) add complementary deoxynucleoside triphosphates in a 5' to 3' direction to previously bound RNA primer [109]. DNA replication always starts at a specific location within the genome, called origin of replication, characterized by a high AT content thus making double strand unwinding more feasible. This process is shown in Figure 21 for replication in eukaryotes where replication starts from distinct origins of replication on each present chromosome. In prokaryotes, the overall process is very similar to the one in eukaryotes, apart from the fact that replication is initiated from a circular double-stranded moiety, known as plasmid which harbors only one origin of replication. It should be noted that during replication not only the DNA is duplicated but also all involved proteins, thus making cell division a fundamental process within the gene expression of a cell [109].

In the 1970s an *in vitro* technique was developed by Mullis, Erlich and coworkers, taking DNA replication as model by enabling the amplification of synthetic DNA strands in a controlled manner, termed polymerase chain reaction (PCR) [110]. PCR quickly evolved to be one of the most frequently used techniques in molecular biology and was applied in chapter 4 when it came to design plasmid constructs.

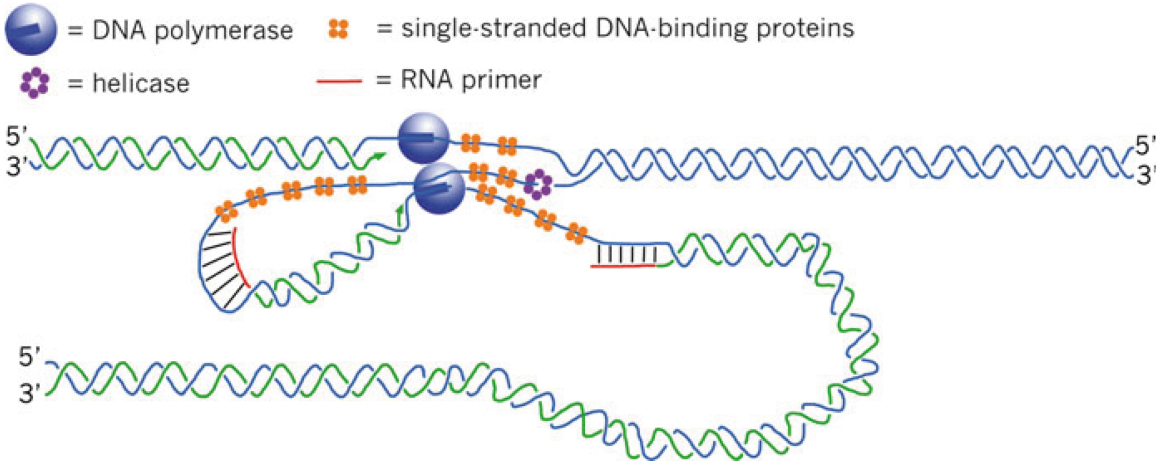


Figure 21. Simplified schematics of DNA replication. Replication of DNA is initiated with unwinding the double-stranded template strand by helicases, stabilization by single-stranded DNA binding proteins and binding of short RNA primers to each of the resulting single strands. Elongation of each single strand is performed by DNA polymerases by adding complementary nucleotides in 5'-3' direction. Taken from [9,111].

Subsequent passages will be dedicated to gene expression, leaving behind DNA replication as fundamental process in evolution of life.

**Gene expression**

Gene expression is characterized by the process in which a physical genetic domain within the DNA is interpreted and finds its physical expression in the phenotype. This process is characterized by a two-step mechanism. First, the DNA template is transcribed into a short-life copy messenger RNA (mRNA) carrying the genetic information. Second, the mRNA is decoded and translated into a functional entity, the protein. This process shares some similarities within prokaryotes and reveals striking differences in eukaryotes though, owed to the much more complex cell machinery, see Figure 22.

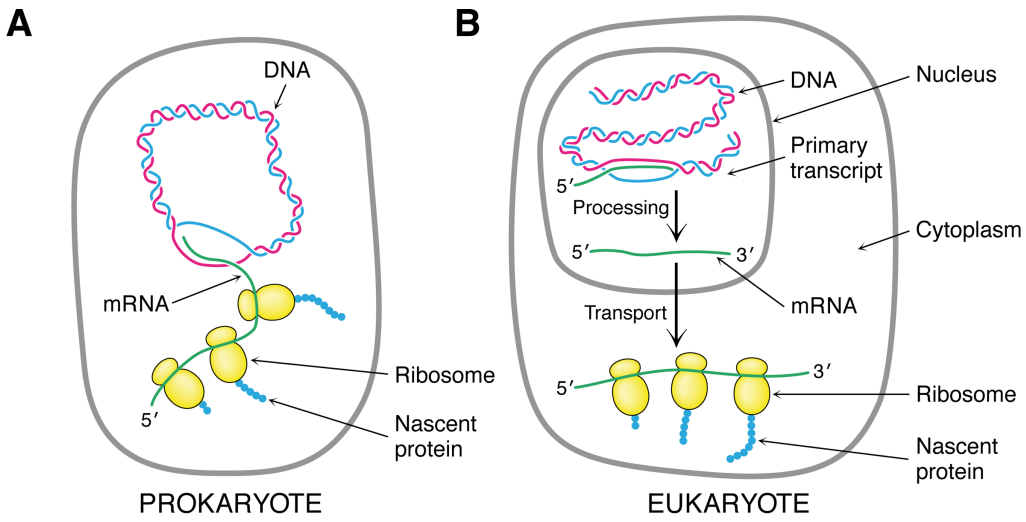


Figure 22. Transcription and translation in prokaryotes (**A**) and eukaryotes (**B**). A double-stranded DNA serves as template for its transcription into messenger RNA (mRNA) which is further processed and translated into proteins which are composed of multiple amino acids linked together by peptide bonds. Major differences within the transcription process are clearly visible. Whereas in prokaryotes the primary mRNA transcript is readily translated into protein, primary mRNA in eukaryotes is processed and spliced in the nucleus. After export from the nucleus into the cytoplasm the actual translation starts. Taken from [108].

## Transcription

Before describing main steps, it is interesting to mention that to some extent similar processes take place in transcription as well as in replication. Comparably to DNA replication that is performed by the DNAP, the RNA polymerase (RNAP) is in charge of producing complementary copies of RNA using DNA molecules as templates. The process can be divided into distinct steps. Transcription is initiated by recruiting RNAP accompanied by a set of transcription factors (possessing either activating or repressing functions) to specific locations within genome, known as promoter regions, which are upstream of the to be transcribed gene within the transcription start site. As it happens during DNA replication, a small portion of the double helix is unwound by RNAP, separating the DNA helix into the coding strand and the template strand, see Figure 23. The RNAP bound to the promoter region of the template strand then catalyzes the formation of covalently linked phosphodiester bonds between the nucleotides to the growing RNA chain in 5'-3' direction until it reaches and recognizes a termination sequence. This induces the displacement of the RNAP. In contrast to DNA replication, the newly formed RNA strand is instantly displaced, followed by reforming of the DNA helix. This allows the production of multiple copies of RNA from one single DNA template [108]. Transcription termination usually occurs at palindromic regions of DNA where strong secondary structures are formed, hence creating inefficient and unstable DNA-RNA hybrids, accompanied by poly polyadenylation of the 3' end that in the finally

results in a “falling off” of the RNAP [112]. Any in depth information can be drawn out of accordant standard literature [108].

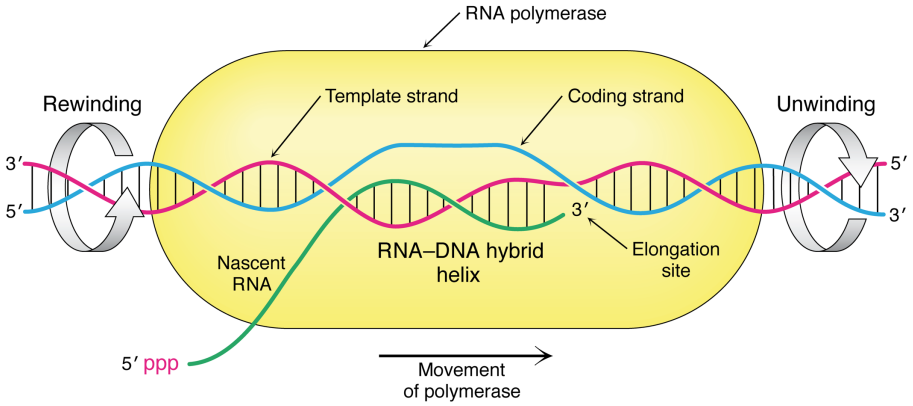


Figure 23. Formation of the “transcription bubble”. RNAP unwinds the DNA double helix, thus dividing it into coding strand and template strand. RNA ribonucleotides are subsequently added in 5’-3’ direction to the growing RNA single strand complementary to the template strand. The resulting messenger RNA serves as information carrier for subsequent translation. Taken from [5,108].

**Differences in prokaryotic and eukaryotic transcription**

There are striking differences between prokaryotic and eukaryotic transcription as gene expression in prokaryotes requires less components and is less complex, compared to eukaryotes. First, there is only one RNAP in prokaryotes, whereas there are three types found in eukaryotes [5]. Second, in prokaryotes, transcription and subsequent translation take place simultaneously. In eukaryotes, mRNAs are transcribed and then processed before being exported to the cytoplasm, followed by translation. The processing of the mRNA involves three distinct steps: intron splicing, modification of the 5’ nucleotide (capping) and 3’ end polyadenylation, resulting in the addition of 200-300 adenosine nucleotides to the 3’ end of the mRNA.

Understanding the basics of these mechanisms is essential for following the framework of this thesis as e.g. for some experiments in chapter 4.2, mammalian RNA has been transcribed *in vitro* and subsequently used for transfection experiments.

## 2.2.2 Highlighting translation in bacteria and eukaryotes

### Translation

Translation describes the mechanism of protein synthesis. Here the four-letter code representing nucleotides is rewritten into the 20-letter code of which proteins are commonly made of.

The process is briefly described in the following section with attention being drawn to the initiation of translation in eukaryotes as this stage is important to understand when it comes to chapter 4.2. For both, prokaryotes and eukaryotes, translation begins at the post-transcriptional stage of the mRNA template. Here, the transferRNAs (tRNA) come into play. tRNAs are RNA structures that function as an adapter and contain a triplet anticodon that binds to its complementary sequence on the mRNA. They deliver the corresponding amino acid that will be incorporated to the growing polypeptide chain. Similar to replication and transcription, the mRNA is read in 5'-3' direction with the result that the polypeptide chain is synthesized from the 5' start (amino terminus) to the 3' end (carboxyl terminus). These steps are carried out by the ribosome, a large cellular machinery consisting of ribosomal RNA and a set of accessory proteins. Ribosomes are divided up into subunits. In prokaryotes: 30S small subunit + 50S large subunit together form the 70S ribosome complex. In eukaryotes: 40S subunit + 60S subunit together form 80S ribosome elongation complex.

In general, protein synthesis is initiated close to the 5' end of mRNA – for prokaryotes more than 25 nucleotides downstream of the 5' end within a purine rich sequence, for eukaryotes close to the 5' cap structure. Initiation of translation is connected with the small ribosomal subunit forming a complex together with initiation factor (IF) proteins and the joint initiator tRNA anticodon binding to complement mRNA. This is accompanied by subsequent scanning for the ribosomal binding site (RBS) including the triplet AUG, representing methionine as start codon, followed by assembly of the respective large subunit, shown in Figure 24. At this stage, the ribosome is ready to initiate the elongation phase which is escorted by recruitment of tRNAs loaded with amino acids and anticodon binding of tRNAs to complementary mRNA. Associated amino acids are incorporated to the growing peptide chain by peptide bond formation and subsequent dissociation of tRNA molecules without attached amino acid. The elongation phase terminates once the ribosome encounters a stop codon on the mRNA that is recognized by release factors, resulting in the release of the polypeptide chain [5,108].

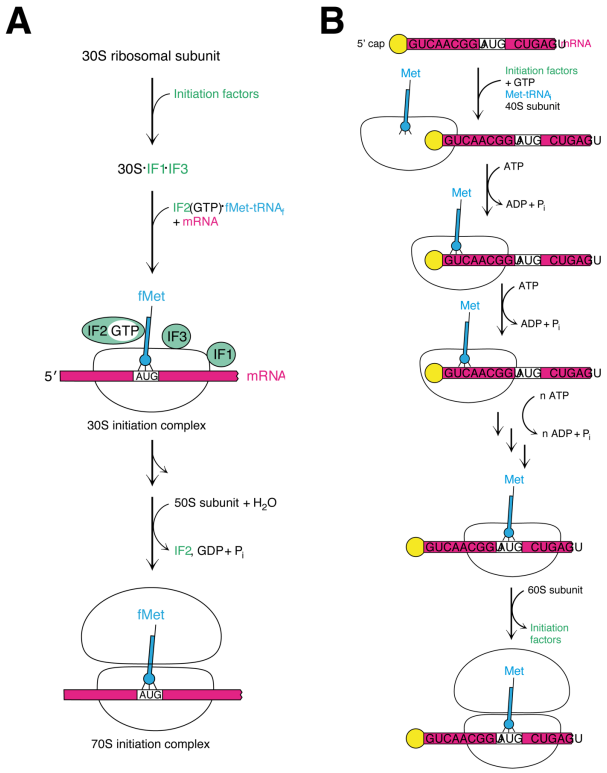


Figure 24. Initiation of translation in prokaryotes (**A**) and eukaryotes (**B**). In prokaryotes, initiation of translation starts with 30S ribosomal subunit forming a complex with IFs, tRNA binding to mRNA at AUG start codon and subsequent recruitment of 50S subunit resulting in 70S initiation complex. In eukaryotes, translation initiation starts at the 5' cap of mRNA and involves mobilization of 40S subunit together with a set of IFs and initiator t-RNA. Forming the 43S complex initiates scanning 5' UTR of mRNA and recognition of AUG start codon. This in turn guides 60S subunit to the 43S complex resulting in the fully assembled 80S complex and subsequent elongation phase. Modified from [108].

This stage can be considered as the final step within the "central dogma" as proteins represent the functional entity whose sequence is stored within the DNA sequence. Recently, with the rise of synthetic biology, controlling gene expression at the transcriptional and post-transcriptional levels became increasingly desirable, see chapter 2.3.2.

In the following sections, some aspects addressing initiation of translation with focus on the eukaryotic mechanism will be elaborated, as understanding of these aspects will contribute to understand motivation and goals of presented works in chapter 4.

### Initiation of translation with special emphasis on eukaryotes

The following section will focus on crucial factors involved in the initiation of translation in eukaryotes and highlights the relevance of mRNA secondary structures for translation efficiency as well as the role of the Kozak consensus sequence.

In prokaryotes, the Shine-Dalgarno sequence represents the RBS for ribosomes and therefore is responsible for the initiation of translation as it distinguishes AUG as initiation codon from

other internal triplets in the mRNA [113]. The Shine-Dalgarno sequence is located within the 5' untranslated region (UTR) and is composed of a purine rich sequence with AGGAGG as a core sequence of about 10 nucleotides upstream of the AUG start codon and in most cases, at least 25 nucleotides from the 5' end. The sequence is complementary to the 16S RNA which is part of the 30S ribosomal subunit. This subunit is recruited and bound to the RBS, and as consequence initiates subunit formation, accompanied by initiation of translation.

In eukaryotes, the mRNA is circularized during translation. The circularization occurs when the 5' cap and polyA tail are spatially brought together and stabilized by accessory proteins such as eukaryotic initiation factors (eIF) and polyadenylate binding proteins (PABPI), see Figure 26 below [5,108]. Initiation of translation in eukaryotes differs from the one in prokaryotes as it starts at the 5' end close to the cap region (Figure 25). Previously recycled 40S ribosome subunit assembles with the initiator GTP-methionine-tRNA and a set of eukaryotic initiation factors (IF), especially eIF4A that possesses helicase activity. Together they form the 43S ribosome complex. Furthermore, a complex of additional eukaryotic IF4 helicase proteins, probably binding to the mRNA $m^7$G-cap structure, is assembled. These two complexes are subsequently interacting with each other, bringing the 43S ribosomal complex and mRNA together to the 5' cap region of mRNA, assisted by eIF4G which serves as a scaffold bridging the ribosome to the mRNA cap. Binding of the 43S ribosomal complex to mRNA via eIF4 proteins is considered as the rate-limiting step of translation initiation [114,115]. This is followed by unwinding the 5' terminal secondary structure, subsequent scanning of the 5' untranslated region (UTR) in 5'-3' direction and unwinding of weak secondary structures until encounter of start codon AUG. Following start codon recognition, the 48S complex is formed and favors recruitment of 60S subunit forming 80S elongation complex including the release of initiation factors [116,117]. In turn, the protein synthesis is initiated and proceeds to elongation phase.

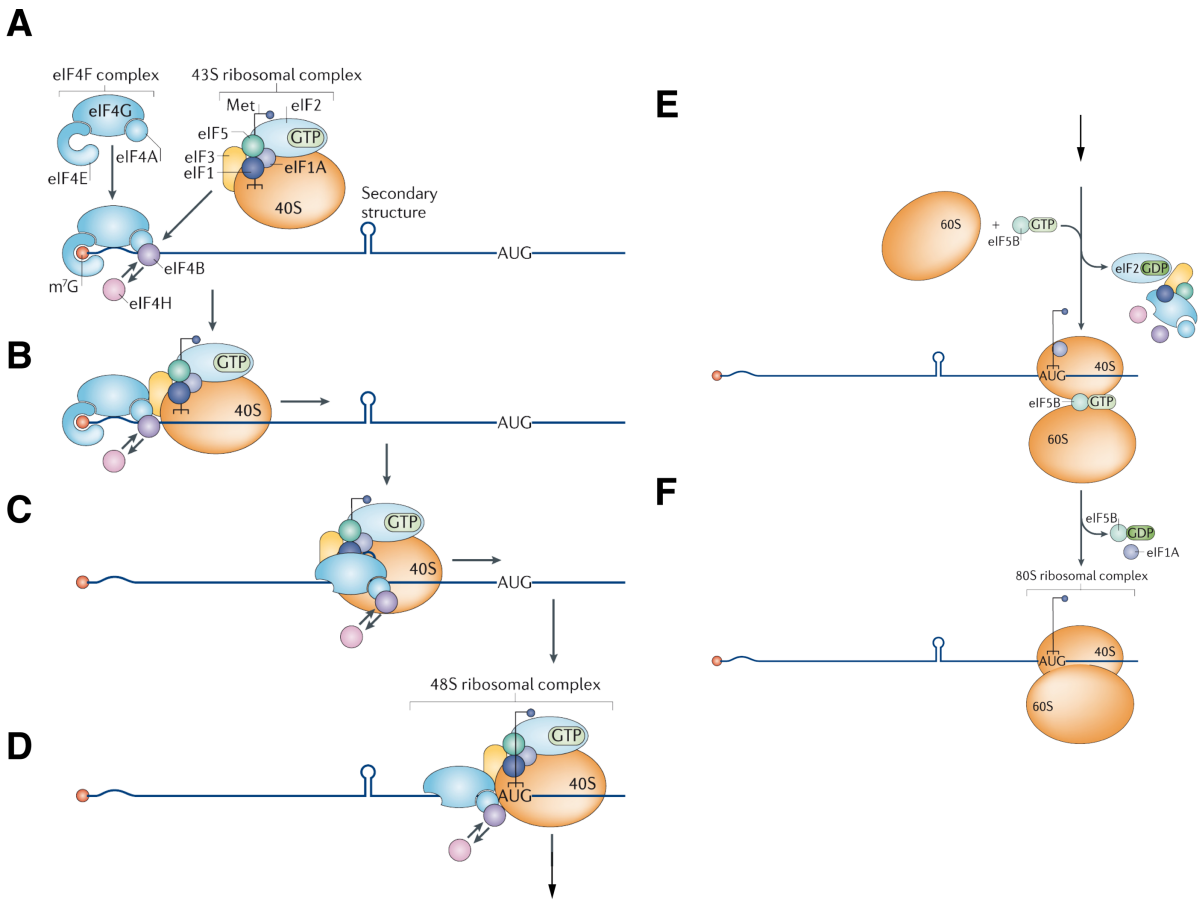


Figure 25. Initiation of translation in eukaryotes (Major steps). **A**. Formation of initiation factor complex and 43S ribosomal complex. **B**. Binding of 43S ribosomal complex to 5' mRNA. **C** and **D**. Scanning of 43S complex in 5'-3' direction towards AUG start codon and subsequent formation of 48S ribosome complex. **E**. Recruitment of 60S ribosome complex to the 48S ribosomal complex resulting in formation of 80S ribosomal complex. **F**. Fully assembled 80S elongation complex ready for entering the elongation phase. Modified from [117].

## Excursion: Relevance of Kozak sequence for translation initiation

A similar sequence pattern as the Shine-Dalgarno sequence is found in eukaryotes, which is termed - after the name of the first investigator - Kozak sequence. The Kozak sequence has been shown to be highly conserved in most vertebrates with subtle variations within the different taxa [118]. Its common sequence is as follows:

| +1 |
| --- |
| Kozak consensus sequence: 5'- GCC(A/U)CC**AUG**G -3' |

In an optimal context it contains a purine at -3 and a guanine at the +4 position, relatively to the adenine of the AUG triplet which is designated at +1 position [119].

Here the highest conserved position in this motif is found within -3 positions upstream of the AUG start codon. In eukaryotes, AUG represents the start codon without exception.

Otherwise, in this context Kozak *et al.* have been shown that in higher eukaryotes the first AUG only serves as start codon when the 5' proximal codon is in a favorable context [120,121]. The Kozak sequence plays a crucial role in the initiation of translation but is not considered as an RBS. This also applies to initial recruitment and binding of the 40S subunit to the mRNA which itself either occurs at the 5' cap or via an internal ribosome entry site (IRES) located within the mRNA [122,123]. Rather, the Kozak sequence is suggested to participate in the recognition of the start codon by the ribosome. However, there are reports suggesting that in various eukaryotic systems, the pre-initiation complex can lack the start codon of the open reading frame, possibly resulting in translation start from an alternative start site [124].
The initial scanning process of ribosome complexes close to the 5' cap and its associated need of unwinding secondary structures as well as the Kozak sequence are relevant for the design for the mammalian toehold switch project presented in chapter 4.2.

**Excursion: The influence of secondary structures on translation initiation**

At this point, the ability of ribosome complexes to unwind hairpins close to the 5' cap should be elaborated in more detail. Once the 43S ribosome complex has formed, it starts scanning the 5' UTR of mRNA. During this process, secondary structures, being abundantly present within mRNA need to be passed and unwound (Figure 26 below). Unwinding weak secondary structures is strongly ATP and RNA helicase activity dependent [125]. By focusing on *in vitro* studies Kozak analyzed two different hairpin structures on their influence on formation of 43S ribosome initiation complex and overall translation efficiency [126]. It turned out that a weak duplex structure of $\Delta G$ = -30 kcal/mol, if placed close (12 nt) to the 5' cap, did strongly prevent mRNA from recruiting 40S subunits as they might not be able to bind. Once inserted further downstream from the 5'cap (52 nt) the same hairpin could be overcome by 40S ribosomes. This suggests that the position of secondary structures strongly influences the unwinding capacity of ribosomes. However, a stronger hairpin ($\Delta G$ = -61 kcal/mol) placed further downstream from 5' cap (72 nt) turned out to be too stable to be unwound by 40S ribosomes. The author suggested that such a hairpin does not prevent a 40S subunit from binding, but results in its stalling 5' of the duplex structure. Nevertheless, to some extent, such strong secondary structures can be disrupted by 80S ribosomes during the elongation step, proposing the hypothesis that the fully assembled ribosome complex possesses a greater capability to unwind secondary structures [126]. Similar studies were carried out more recently in live mammalian cell lines [127]. Again, translation efficiency was significantly decreased when hairpin strengths were increased by $\Delta G$ = 10 kcal/mol (shifting $\Delta G$ = -25 kcal/mol to

ΔG = -35 kcal/mol), especially when hairpins were inserted close to 5' cap. Furthermore, it became apparent that the GC content of secondary structures influences translation efficiency. Hairpins with thermal stability and hairpin-to 5'cap distance kept constant, but in which the GC content was increased, resulted in decreased translation efficiency. The field came up with the hypothesis that this is due to impeded local unzipping processes by ribosomes as GC duplexes possess stronger bonds. Experiments carried out in yeast systems indicate that stronger hairpins (ΔG ≥ -67 kcal/mol) even inhibit 80S ribosomes at the elongation phase of translation [128]. In here, the Kozak consensus sequence comes into play again. As already mentioned above, the Kozak sequence is crucial for translation initiation as it comprises the AUG start codon. Thus, insertion of secondary structures close to the 5' cap with the Kozak sequence sequestered within the secondary structure are supposed to have an impact on translation, whereas secondary structures placed downstream of Kozak sequence should still enable translation.

Those effects were studied in chapter 4.2 by examining the complex mechanisms of translation initiation in eukaryotes with focus on the relevance of Kozak sequence and secondary structures.

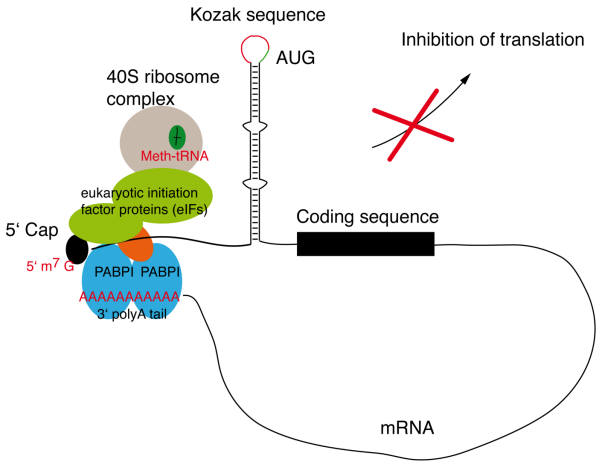


Figure 26. Hairpin positioned close to 5' cap hindering translation of downstream mRNA.

### 2.2.3 The relevance of microRNA

MicroRNAs (miRNAs) belong to a class of small non-coding RNA (ncRNA) species known to negatively regulate gene expression. This is mediated by binding to complementary mRNA sequences within 3' UTRs which target them for degradation and therefore prohibit mRNA translation [129-131]. These ncRNAs were discovered in 1993 in the model system *C. Elegans,* and till date have been verified in over 140 different species, including plants, thereby presenting highly phylogenetically conserved molecules [132,133].

**Biogenesis of miRNA**

In the following the biogenesis of miRNA will be described [129,130,134]. In animals, transcription of miRNA is mediated by RNAP II in the nucleus, followed by 5' capping and 3' polyadenylation. This results in an endogenous hairpin transcript possessing a stem loop of 33-35 bp, termed primary miRNA, see Figure 27 [129,135]. Still in the nucleus, the primary miRNA is processed and cleaved by microprocessor, a complex consisting of the RNase type III enzyme Drosha and its essential cofactor protein DGCR8 (including accessory factors). This results in a 60-70 nt long hairpin precursor pre-miRNA [135-137]. This precursor transcript is subsequently exported into the cytoplasm by exportin V protein, together with the GTP-binding protein Ran-GRP. Further processing is mediated by the RNase III enzyme DICER I [138]. DICER I subsequently cleaves off the stem loop structure leaving a ~22 nt long miRNA duplex [139]. DICER I activity has been shown to be associated with the transactivation-responsive RNA binding protein (TRBP). TRBP improves cleavage accuracy, and moreover it is involved in subsequent association with Argonaute (AGO) proteins [140]. The precursor miRNA duplex is loaded onto the AGO1-4 proteins, resulting in formation of an effector complex known as RNA-induced silencing complex (RISC). The RISC is an important mediator of various mRNA silencing pathways [141]. Following unwinding of the double-stranded miRNA, the mature miRNA known as the guide strand stays within the RISC, joint by the GW182 protein family. The complement of the guide strand is rapidly degraded. The choice of the catalytically active guide strand has been reported to depend on the thermodynamic stability of the two ends of the RNA duplex with preference to the unstable terminus of the 5' end as well as to the first nucleotide, which is preferred to be uracil [142]. The guide strand residing within the RISC can target complementary mRNA for post-transcriptional gene silencing which happens in the processing bodies (P-bodies). P-bodies

represent distinct foci within the cytoplasm of any eukaryotic cell. They are thought to contain crucial enzymes involved in mRNA turnover and are assumed to be spots for translational suppression, mRNA decay and silencing of RNA [129,143]. Apart from the above described canonical biogenesis pathway of miRNA, also non-canonical pathways have been reported [144].

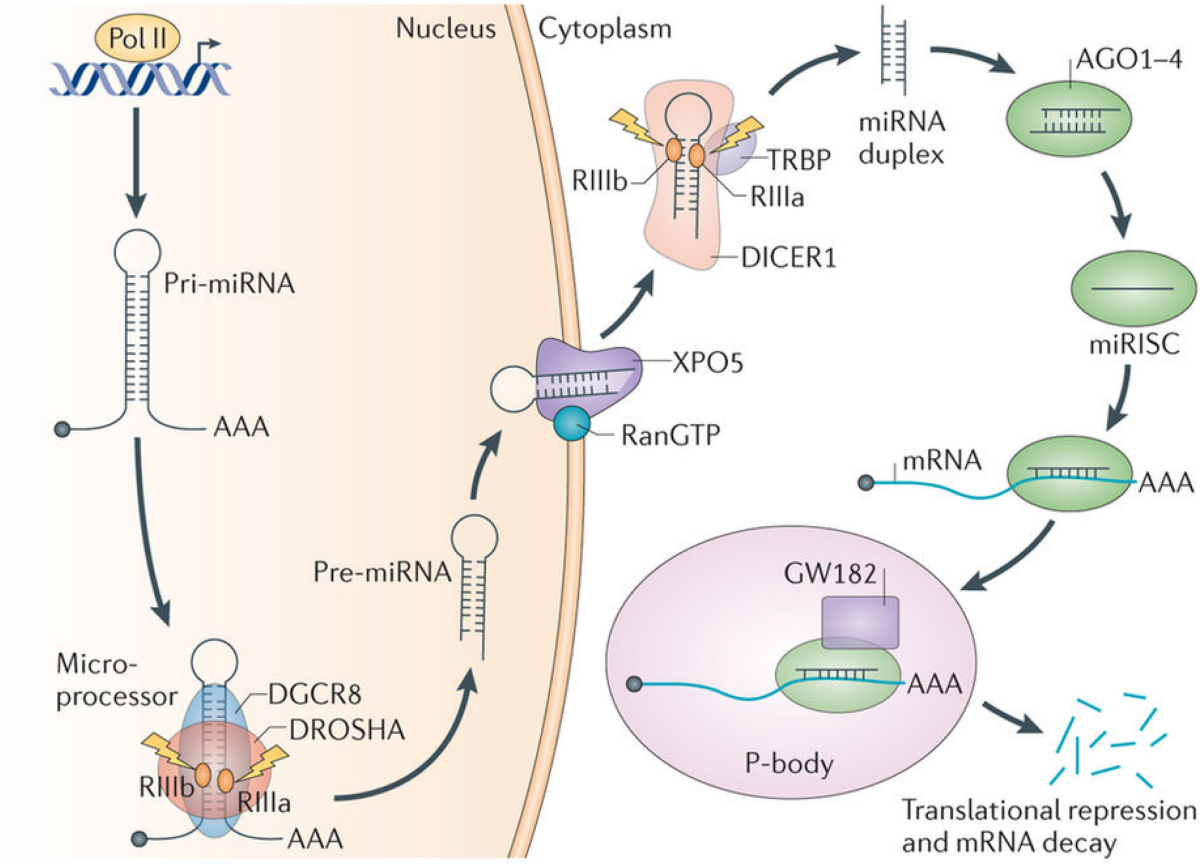


Figure 27. Biogenesis pathway of microRNA. In the canonical miRNA biogenesis, RNAP II transcribes a primary hairpin loop containing a miRNA construct which subsequently is processed, exported and due to processing steps loaded into the RISC with a catalytically active single-stranded guide strand. This mature guide strand now can target complementary mRNAs and mediate its translational repression and mRNA decay, presumably in foci as referred to P-bodies [130].

**The role of microRNA in disease development, especially cancer**

Mutations of proteins involved in the maturation process can cause substantial cellular malfunctions. Malfunctions and dysregulation in the microRNA pathway are connected to developmental, cardiac and neuronal diseases, and various types of solid cancer [129,130]. These can occur throughout the different steps during maturation. Especially in cancer, microRNAs may function as oncogenes or better known as onco-miRs in case of their uncontrolled overexpression and amplification. On the other hand, there is also evidence that miRNAs act

as tumor suppressors and can be downregulated in malignant cells. Also (epi)-genetic mutations and alterations either positively or negatively influence miRNA transcription at various levels within the biogenesis.

**miRNA-21 as suitable model input feeding the immobilized DSD cascade**

In my thesis, we decided to focus on the microRNA-21 which is one of the early found miRNAs in mammals. Overexpression of mir-21 affects crucial cellular signaling pathways such as AKT and MAPK as it inhibits phosphatases, resulting in uncontrolled amplification of kinases, which in turn phosphorylate downstream effectors [145,146]. We chose miRNA-21 as "realistic" target sequence for triggering the immobilized DSD cascade which therefore could be referred as a biosensing device, similar to DNA logic circuits already mentioned above that also used microRNAs as input signals [79]. It should be noted that the sequence length of miRNA-21 is 22 nt, which, in order to feed the system with a natural input, was not optimized. Its sequence is as follows: **UAGCUUAUCAGACUGAUGUUGA**

## 2.3 Synthetic biology

### 2.3.1 Molecular biology as an engineering discipline

Various definitions describe the field of synthetic biology. This fairly new branch of biological research applies classic engineering approaches to study (artificial) biological systems and manipulate them so that they become controllable. They are constructed to either carry out *de-novo* designed tasks or possess novel functionalities [147]. Hereby researches are intending make use of similar strategies and workflows known to engineering by taking advantage of mindsets and terms like standardization, characterization and abstraction:

- Standardization – Use of a toolbox containing widely used biological parts in which entities like genes are dissected into functional modules with well-defined properties like promoters, coding sequences or terminators. These go along with modes of connectivity, comparable to electrical parts with the goal to enable redesign or *de-novo* design of biological systems.
- Characterization – For understanding and prediction of function, these parts have to be thoroughly characterized and their operation within biological networks needs to integrated.
- Abstraction – Standardization as well as characterization of these building blocks aims at and allows an in-depths understanding and simplification of biological networks by dissecting complex biological systems into its single components and looking at it from an engineer's point of view.

The origins of synthetic biology can be backtracked to 1961 when Francois Jacob and Jacques Monod investigated the role of the lac operon in *E. coli* in gene expression. They realized that mutations occurring in LacZ and lacY structural genes did alter the amino acid sequence of proteins thus affecting lactose metabolism. This is considered as the first time proof of the existence of regulatory circuit mechanisms that enable the cell to react upon environmental changes of state [148]. In the following years, transcriptional regulation in bacteria was intensively studied and awarded with the Nobel prize in medicine and physiology in 1965 as a result "*for their* (Jacob, Lwoff, Monod) *discoveries concerning genetic control of enzyme and virus synthesis*" [149]. As understanding of molecular biology grew, new molecular biology tools as well as techniques arose, especially in molecular cloning. Particularly useful was the possibility to amplify and replicate DNA sequences of choice by PCR, which was awarded

with the Nobel prize in 1993 [150]. This knowledge and new technologies of e.g. (whole) DNA sequencing as well as the evolution of bioinformatic tools enabled the field to rapidly expand and study regulation of cellular systems and networks, mainly in *E. coli* and yeast strains of which the genomes were sequenced in 1997 and 1998, respectively [151,152]. Over time the library of genetic parts gradually increased and enabled the creation of artificial genetic circuits with increasing complexity.

The field of modern synthetic biology emerged in the late 1990's when the techniques and knowledge made it possible to transfer classic engineering approaches to cell and molecular biology. This opened a new research area which adopted terms and language from electric engineering – see defined terms above [153].

But instead of circuits for electronic devices, simple genetic circuits were built. Herein those operations could be described by elementary mathematical models. Due to its genetic simplicity and limited genomic size as well as the broad knowledge about its general biology and regulatory mechanisms, *E. coli* initially was the perfect organism to e.g. study simple genetic switches.

**Early stage genetic circuits**

One major breakthrough was achieved by Collins and coworkers in the year 2000 who designed a synthetic bistable gene-regulatory network in *E. coli* that is based on the predictions of a simple mathematical model, known as toggle switch [154]. This system is composed of two repressors (LacI and CI) and two constitutive promoters in which each promoter is inhibited by the repressors that is transcribed by the opposing promoter (Figure 28). Switching occurs by using an inducer (IPTG (Isopropyl β-D-1-thiogalactopyranoside)) which disengages lacI, and heat which disengages cI of the current active repressor, to displace one of the repressors from its operator. Induction by either IPTG or heat allows the opposing repressor to be maximally transcribed until it solidly represses the originally active promoter resulting in a transition from bistability to monostability. This work can be considered as a shift from the classic protein engineering approach as it focuses on the manipulation of a genetic network architecture.

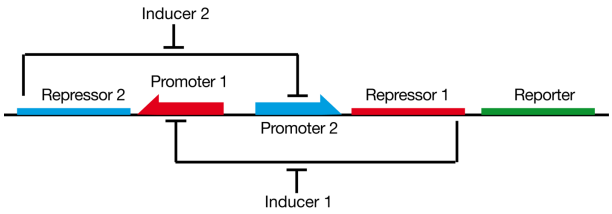


Figure 28. Design of the toggle switch. Repressor 1 inhibits transcription from Promoter 1 and is induced by Inducer 1. Repressor 2 inhibits transcription from Promoter 2 and is induced by Inducer 2. Modified from [154].

In the same time falls another pioneering work by Elowitz and Leibler who designed a synthetic oscillating circuit composed of three transcriptional repressor-promoter systems in *E. coli* which they called repressilator [155]. In there, lacI repressor protein inhibits transcription of second repressor gene tetR which in turn represses transcription of cI repressor gene, resulting in a negative feedback loop and periodic induction of GFP synthesis as readout. This work represents an example of taking advantage of naturally occurring components and its implementation into an artificial genetic network with new functional properties. Looking at those studies described above it becomes obvious that major parts used in there were similar (like promoters or GFP expression as fluorescence readout). This demonstrates that the use of parts from a common toolbox, as it is done in engineering, has been a fundamental part of synthetic biology from its early beginnings. Subsequently, a lot of effort was dedicated in the creation of more complex circuits and logic gates recapitulating the idea of using well characterized and standardized parts.

### 2.3.2 Post-transcriptional control of gene expression

One interesting direction in synthetic biology is the rational control of gene expression at the post-transcriptional and translational level. As translation initiation can strongly be influenced by secondary structures within the RBS (see chapter 2.2.2), a new class of post-transcriptional RNA-based regulators was presented by Collins and coworkers in 2004 underlining the important contribution of noncoding RNA species to control of gene expression (Figure 29) [156]. Inspired by endogenous riboregulators a modular artificial system was designed that could repress and activate translation *in vivo,* enabling precise control of gene expression through rationally designed RNA-RNA interactions [157]. Hereby repression is achieved by the formation of an RNA stem loop structure (crRNA) within the 5' UTR that sequesters the RBS of mRNA. This repressed mRNA inhibits 30S ribosomal subunit recognition of the RBS, thereby preventing subsequent steps in translation initiation. Expression of the downstream gene (in this case GFP for flow cytometry readout) is controlled by transcription of a trans-acting noncoding RNA (taRNA) produced from a second promoter. The sequence of taRNA was designed to target and hybridize to the stem loop of crRNA, forming a RNA duplex leading to a conformational change in the crRNA. Hence, the stem loop is unwound and the RBS exposed and accessible for ribosomes which can then initiate translation. Circuits able to control gene regulation at the post-transcriptional level became a model for a series of subsequent works.

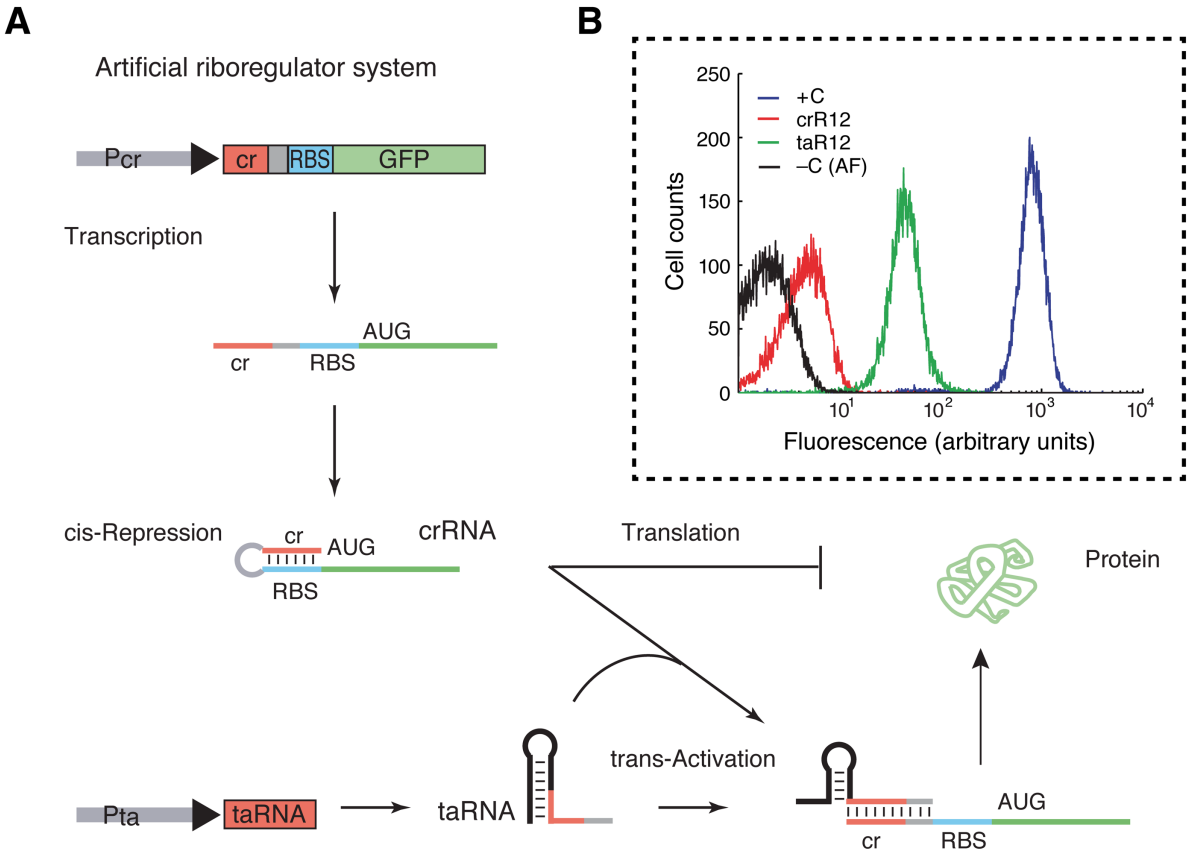


Figure 29. Design of an artificial modular riboregulator system. **A**. Principle of post-transcriptional control of gene expression. A cis repressing sequence (cr) is inserted upstream of RBS and downstream of a promoter (Pcr). Upon transcription, a stem-loop is formed which blocks ribosome binding and translation (crRNA). A second promoter (Pta) transcribes a short trans-acting RNA which is complementary to the cr domain of crRNA. By a linear-loop interaction, the previously sequestered RBS is exposed and as consequence translation can initiate. **B**. Flow cytometry measurements of trans-activation. Autofluorescence of cells in black (-C), Positive control of cells without cr sequence in blue (+C), Cis-repressed cells in red (no induction of Pta), trans-activated cells after induction of Pta in green. Modified from [158].

The design of the artificial riboregulator system was further exploited and fine-tuned allowing e.g. a more pronounced control of gene expression as well as the design of genetic switch boards that independently control expression of multiple genes in parallel [159,160]. These works were also aiming towards an improved orthogonality of regulators resulting in less cross-talk of the parts of which they are composed of. An accompanying effect were the accelerated response times of switching between ON and OFF states.

A similar work aiming for an orthogonal platform which is regulating gene expression relying on the sense-antisense RNA system is shown below [161]. This system is derived from the antisense RNA-mediated translation and copy number control system (RNA-IN-RNA-OUT) of the insertion sequence 10 (IS10) in *E. coli* [162]. Here transposase expression is inhibited by an antisense RNA (RNA-OUT) which can hybridize with the translation initiation region of transposase mRNA thus leading to blockage of ribosome binding and translation repression [163]. RNA-OUT is characterized by a hairpin encompassing a loop domain which is

complementary to the 5’ end of the sense RNA-IN, thus acting as translational repressor of downstream gene once expressed and bound to RNA-IN mRNA. Mutalik and coworkers used this naturally occurring control of gene regulation to design a set of orthogonal constructs in which accordant cognate RNA-IN-RNA-OUT pairs specifically bind to each other, resulting in strong repression of downstream protein (Super-folder green fluorescent protein (SFGFP)). In contrast, non-cognate pairs show little crosstalk, visible by substantial protein expression levels. As a result of an optimization process, they selected 23 RNA-IN-RNA-OUT mutant constructs and proved strong repression of 80% of cognate pairs. Furthermore, the authors stated that specificity of sense-antisense RNA interaction relies to a great extent on the thermodynamic stability of the entire duplex of RNA-IN-RNA-OUT and a 5-bp duplexed GC rich seed region, see Figure 30B.

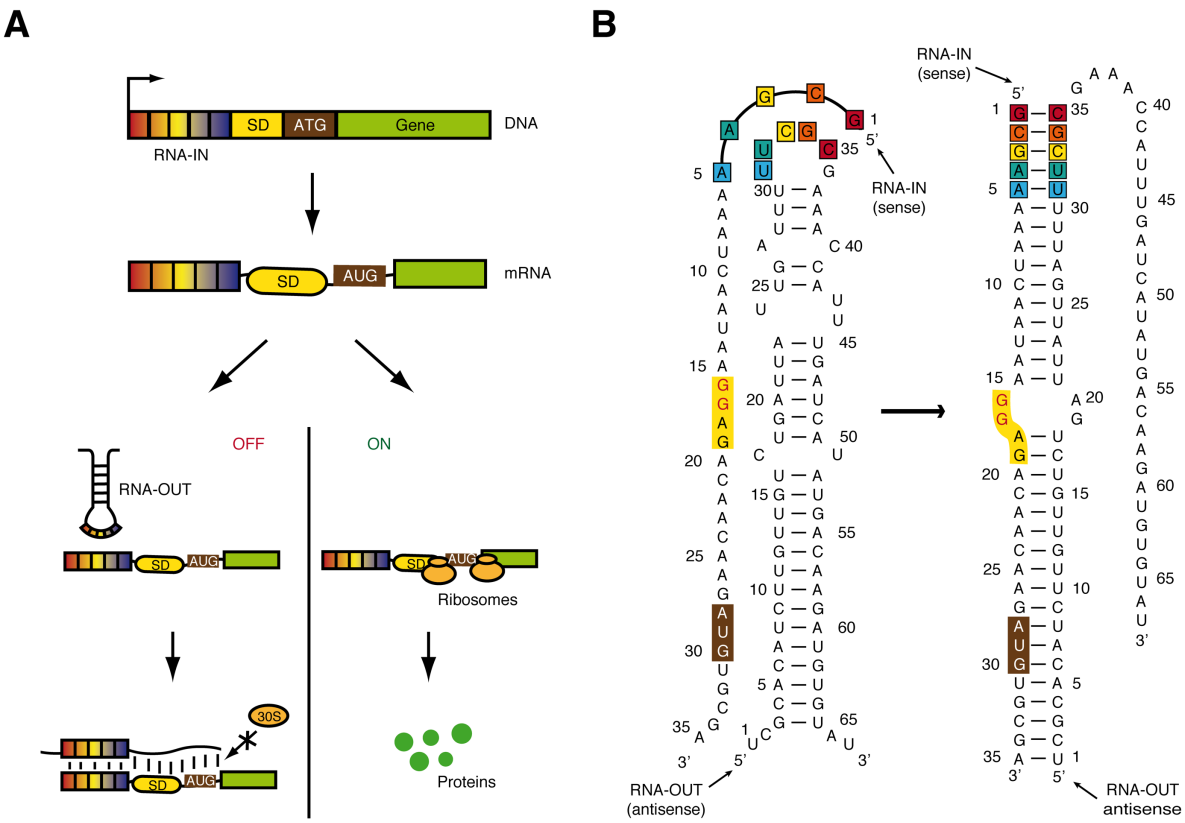


Figure 30. Illustration of the antisense RNA–mediated translation control system. **A**. RNA-IN-SFGFP encodes for SFGFP and transcription of mRNA mediates initiation of translation of downstream protein (right). Transcribed antisense RNA-OUT specifically binds to Shine-Dalgarno sequence and AUG start codon of RNA-IN-SFGFP mRNA thus preventing 30S ribosomal subunit binding and initiation of translation (left). **B**. Model of RNA-IN-RNA-OUT base pairing with seed recognition region between interacting sequences (in colored boxes). Modified from [161].

Translational repression on the post-transcriptional level will be taken up in chapter 4.1 by presenting bacterial toehold switches functioning as repressors of translation.

**Excursion: Constraints in sequence designs limiting design space**

Works presented above can be considered as a new type of programmable genetic regulators mediating control of gene expression by RNA-RNA interactions at the post-transcriptional level. Nevertheless, at that time, protein-based transcriptional regulators have been shown to possess dynamic ranges which were 10-fold higher than for previously shown RNA-based switches [3]. Furthermore, the number of possible independent orthogonal sequences was limited thus preventing the creation of large libraries of parts exhibiting minimal crosstalk [164]. Even though Mutalik *et al.* had presented 23 mutant pairs of their RNA-IN-RNA-OUT regulating system to show a high degree of orthogonality, significant crosstalk of non-cognate pairs could be detected within their heatmap [161]. One reason for this crosstalk can certainly be found in sequences that are involved in the actual sense-antisense interactions responsible for translational repression. Here repression was solely mediated by Watson-Crick base pairing to the RBS including AUG start codon. Any sequence supposed to displace the repressing sequence needed to contain a RBS, thus limiting design space. Considering a typical interaction domain of 30 nt, the sequence space contains over $10^{18}$ potential regulatory elements thus definitively providing room for improvement of sequence design [3].

Another reason hindering the development of an improved orthogonality of parts can be found in structural motifs used in that time. Works mentioned above mostly relied on U-turn loop structures, especially the YUNR motif (Pyrimidine-Uracil-Nucleotide-Purine) which are favoring loop-loop and loop-linear interactions, see also Figure 30B. These motifs represent a ubiquitous antisense RNA regulation system in prokaryotes and promote bi-molecular, tertiary interactions with a target RNA [165,166]. It has been found in the anticodon of tRNA, facilitating codon-anticodon base pairing during translation initiation as well as within the structure of hammerhead ribozyme. This motif is characterized by a sharp reversal of the RNA phosphate backbone due to two intraloop hydrogen bonds. Subsequent bases are presented in a solvent exposed stacked configuration which provide a scaffold for rapid interaction with complementary RNA. It seemed natural to use such motifs that favor duplex formation in RNA-RNA-based riboregulators, even though Mutalik and coworkers proved that the YUNR motif is not mandatory for solid function of their system [161]. These two examples reflecting intrinsic sequence constraints are illustrated in subsequent Figure 31.

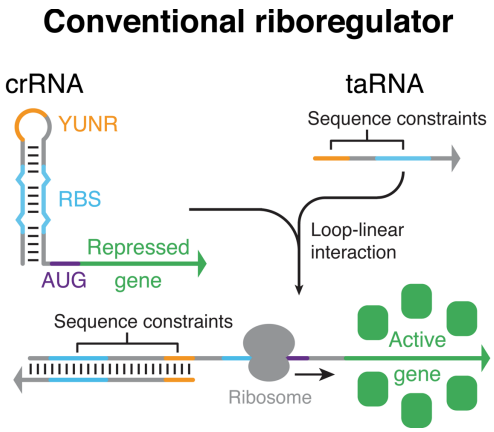


Figure 31. Schematic design principles of conventional riboregulators with sequence constraints indicated. Sequestration of RBS in the stem (in blue) and the YUNR motif in the loop (orange). Taken from [3].

### Toehold switches

The discrepancy of the need for an increased, orthogonal library of biological parts and the actual lack of readily available biological components inspired Green and coworkers to rethink conventional riboregulators leading to a design of what they called toehold switches. The concept will be elaborated in detail as it takes in an important part in this thesis (chapter 4).

### Basic design and mechanism of toehold switch and trigger RNA

Compared to previous examples which rather relied on biological control structures, the toehold switch is mainly founded on artificial mechanisms, taking advantage of linear-linear interactions instead of so far employed loop-linear or loop-loop interactions [58]. The linear-linear interaction applies the previously described toehold-mediated strand displacement reactions on the RNA level (chapter 2.1.3). This mechanism is characterized by the exposure of single-stranded toehold domains, making U-turn motifs for reaction initiation obsolete. The basic design relies on a secondary structure which forms a duplex stem loop, termed switch RNA, resulting in a configuration that prevents translation of a downstream gene. A cognate trans-acting RNA, the trigger RNA, binds to the switch RNA, opens up the hairpin and in turn by presenting RBS promotes translation of the downstream gene. Schematics of switch and trigger RNA are shown in Figure 32A. The original switch RNA is characterized by a hairpin structure inserted close to the 5' end of mRNA which harbors RBS and start codon followed by a 21 nt linker sequence coding for low molecular weight amino acids and the actual coding sequence for the gene of interest – in this case GFP-ASV. Upstream of the hairpin, a 12 nt

single-stranded sequence, domain a serves as a toehold for reaction initiation with cognate trigger RNA resulting in RBS and start codon exposure, allowing subsequent translation initiation. The hairpin structure itself represents the translation repressing module including sequestration of the unpaired RBS within a 11 nt loop (domain b). Domain b is designed so that the AUG start codon remains unpaired in a 3 nt bulge. Sequences immediately before and after AUG (6 and 9 base pairs, respectively) are sequestered within the RNA duplex thus raising the potential of possible trigger sequences. Domain a and b build up a sequence length of 30 nt which is complementary to the 30 nt sequence of the trigger RNA. The trigger RNA contains upstream of domains b* and a* a 10 nt strong hairpin loop structure, intended to stabilize and enhance switch-trigger interaction.

Figure 32B illustrates the mechanism of toehold switch activation upon its interaction with a cognate trigger RNA by a linear-linear interaction and subsequent translation of the repressed gene. Linear-linear interactions are initiated by hybridization of the toehold domain a of the switch RNA and its complement a* of the trigger RNA associated with previously described toehold-mediated strand displacement reactions. This is followed by a branch migration process, finally leading to exposure of the RBS.

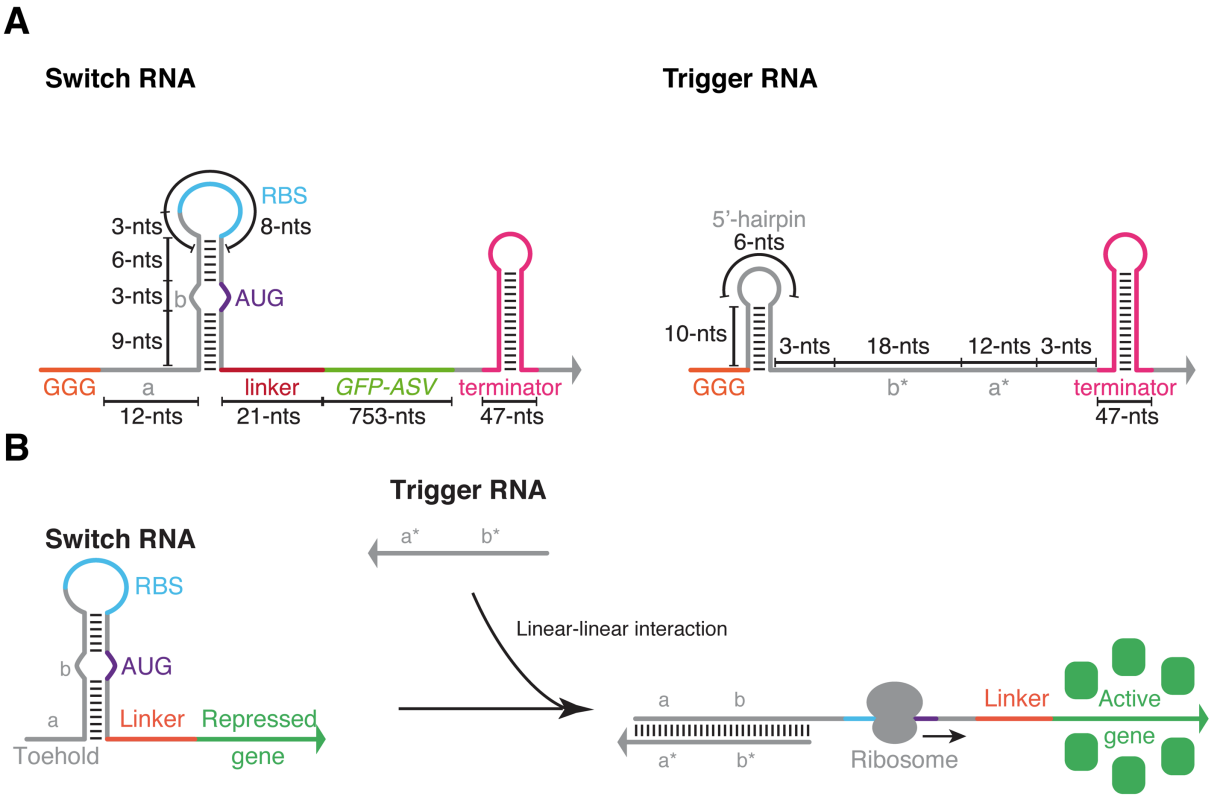


Figure 32. Design schematics and mechanism of the toehold switch with major functional parts highlighted. **A**. Design of switch and trigger RNA. **B**. Mechanism of linear-linear switch-trigger interaction and subsequent initiation of translation. Modified from [3].

The NUPACK nucleic acid sequence design tool was used for library design. Monte Carlo simulations helped to identify components showing the least crosstalk [21].

### Performance of toehold switches *in vivo*

Figure 33A shows the performance of toehold switches in *E. coli* Star DE3 (containing a truncated mutation of RNase E), which were co-transformed with switches on a medium copy colA origin plasmid and trigger on a high copy colE1 origin plasmid. Upon induction with IPTG, T7 RNAP promoted transcription of switch and trigger RNA resulting in their interaction and eventual expression of GFP. Activation of the switch into its ON state results in a similarly high GFP expression as the positive control in which GFP expression is not repressed, whereas the OFF state shows only a slight increase compared to autofluorescence. Similar results for switch-trigger activation are presented in Figure 33B displaying the mode values for three selected switches. In flow cytometry, the mode value allows to visualize and compare recorded events of cell populations varying in number by plotting the differences in relative percentage, and not in absolute numbers. In total 20 out of 168 toehold switches exceeded ON/OFF ratios of 100.

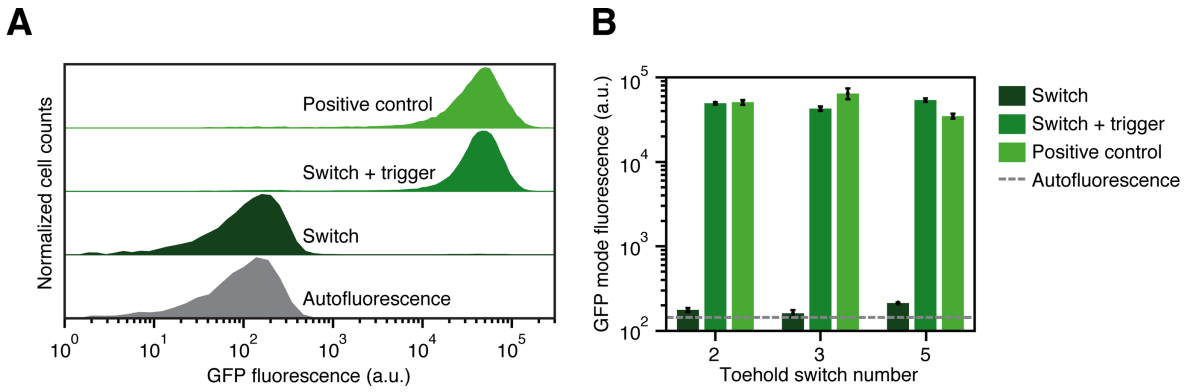


Figure 33. *In vivo* performance of toehold switches. **A**. Histogram representation of flow cytometry showing GFP expression of a representative toehold switch compared to *E. coli* autofluorescence and a positive control. **B**. GFP mode fluorescence levels measured for switches in their ON and OFF states in comparison to positive control and autofluorescence. Modified from [3].

In the same way, results of the repressor toehold switch project, coming up in chapter 4.1 were validated.

## Toehold switch orthogonality

Evaluation of the capability of toehold switches to function as a library of orthogonal sets is shown in Figure 34. 26 toehold switch pair combinations which, referring to previous *in silico* screening were supposed to show low level of crosstalk were tested pairwise in *E. coli*. Figure 34A represents images of GFP expression of bacterial colonies on induced LB plates whereas Figure 34B reflects flowcytometric analyzation of cognate (diagonal) and non-cognate switch-trigger combinations. Here crosstalk is calculated by taking GFP fluorescence from a given switch-trigger pair and dividing it by the GFP output determined for the switch bound to its cognate trigger. There is less than 12% crosstalk for the full set of tested 26 pairs which represents a significant improvement compared to previous works of e.g. Mutalik and coworkers [161].

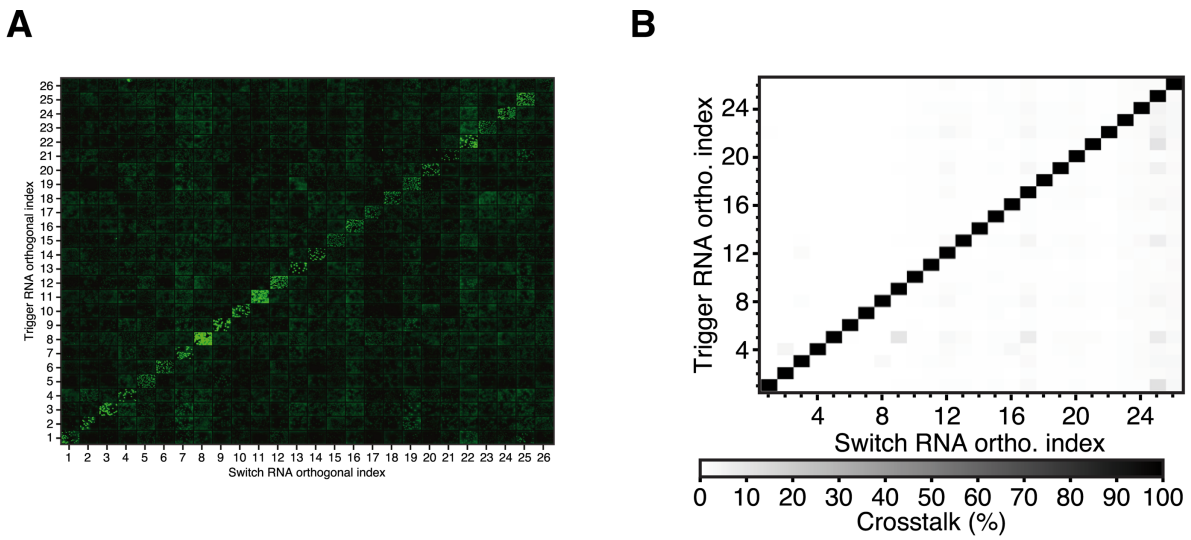


Figure 34. Orthogonality of toehold switches. **A**. GFP fluorescence of *E. coli* bacteria co-transformed with a set of 26 switch-trigger pairs. Cognate pairs represent the diagonal whereas the off-diagonal depicts non-cognate pairs, showing only low fluorescence. **B**. Quantification of crosstalk of these 26 switch-trigger pairs via flow cytometry. Determination of crosstalk was done by taking GFP output measured for a given trigger-switch combination and dividing it by the GFP output measured for the switch with its cognate trigger. Modified from [3].

## Excursion: Parameters determining toehold switch performance

In the course of tuning and forward engineering the first generation of toehold switches, Green *et al.* analyzed parameters that affected toehold switch performance. Key points will be addressed in the following, see Figure 35.

**Hairpin loop size**

Based on previous findings, the pre-RBS region between trigger binding site and RBS of the switch RNA was characterized by changing the loop size at its 5' end. It turned out that increasing the loop size improves the ON state GFP expression level. On the other hand, increasing the loop beyond some point also goes along with more leakiness of the toehold switch, which the authors attribute to the increase of entropy as a result of the longer loop. The longer the loop, the more likely proper folding of the hairpin is negatively affected thus causing reduction of repression. Extending the single-stranded region near the RBS might favor stable docking by the ribosome and thus enhance background translation. Experimental results showed that best ON/OFF ratios were achieved for loop sizes of 15-18 nt, reaching saturation at greater loop sizes.

**Sequence pattern at the top and the bottom of the stem**

Similarly, the stem top and bottom regions are affected. Analysis on this revealed that weak A-U base pairs enhance translation efficiency. This is supported by a work showing that sequences upstream of RBS possessing A-U bases can be beneficial for translation rate as they cooperate with the Shine Dalgarno sequence [167]. In addition, the authors refer to studies analyzing the ribosome-mRNA complexes that bases up to 35 nt upstream of AUG can be protected by the 30S ribosome subunit [168]. By considering these sequence-specific and steric factors, being relevant in a sequence region between the trigger binding site and the RBS of the switch RNA (pre-RBS region), the ON state expression might be further optimized.
Enhanced ON/OFF GFP fluorescence for switches with an A-U base pair at the top of the hairpin was shown to be correlated with low free energy ($\Delta G_{RBS\text{-}linker}$) in the region between RBS to the end of the linker sequence. This free energy reflects the amount of energy which is necessary to be overcome by the ribosome.

**Position and length of the trigger binding site**

Another finding which is worth considering when designing toehold switch-based devices relies in the length the unwound stem. Not surprisingly, increased expression levels are observed once repression of the start codon is decreased by trigger binding. Maximum activation is observed for trigger lengths that unwind 14-16 nt of the stem, leaving a 2-4 bp

long duplex configuration of the stem. This observation is attributed to spontaneous unwinding at 37°C leading to similar effects as those described with switches harboring increased loop sizes.

### Toehold length

When recapitulating the "theory" of strand displacement reactions (chapter 2.1.3) it becomes obvious that increasing toehold binding domain a by varying the 3' end of trigger domain a* is correlated with a moderate increase of expression in a regime between 6-8 nt and a strong increase at 12 nt domain length.

These parameters are important to consider when designing such type of post-transcriptional regulators of translation and were also included in the design of the toehold repressors (chapter 4.1).

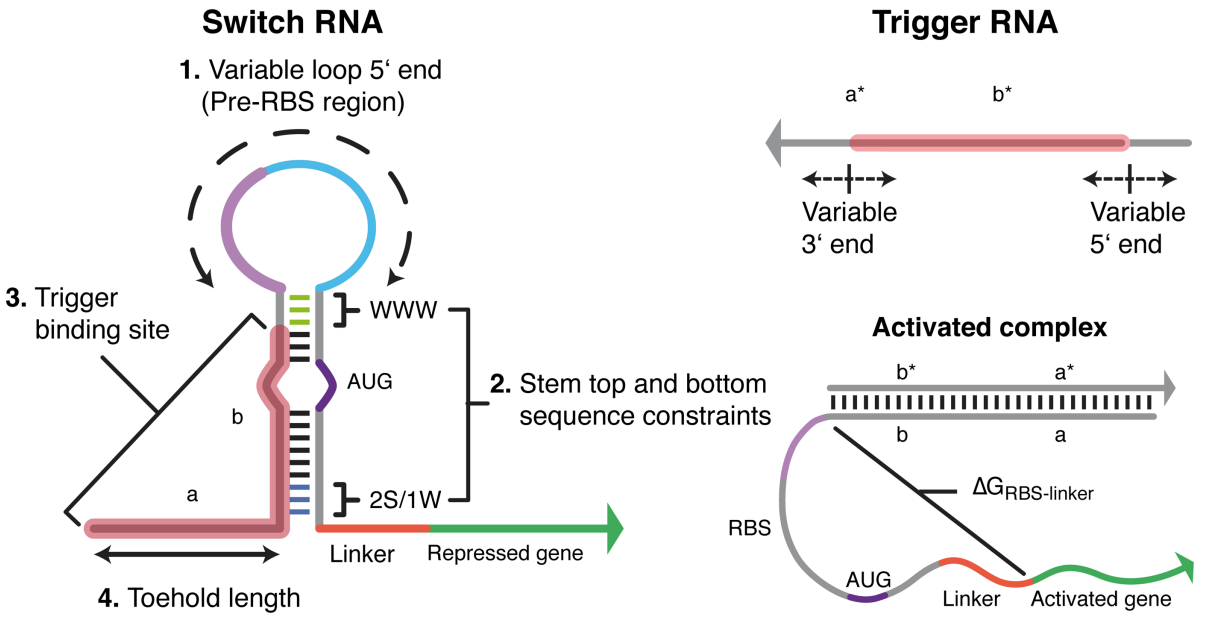


Figure 35. Important parameters determining performance of toehold switch. These include the loop size (**1**), sequence composition of the stem top and bottom (**2**), trigger binding site (**3**) and toehold length (**4**). "WWW" stands for three weak base pairs, "2S/1W" stands for two strong and one weak base pair. Modified from [3].

### Implementation of toehold switches as ribocomputing devices

Most recently the potential of toehold switches was expanded towards its implementation into *in vivo* logic circuits, see Figure 36 [169]. Toehold switches can be utilized as layered ribocomputing circuits that are able to perform AND, OR or NOT operations, thus allowing sophisticated regulation of translation. Using a language usually known from electronic

computers, the circuit receives input signals, which are the ensemble of trigger RNAs. In turn after activation of the gate RNA (known as the switch RNA representing the central signal-processing unit), an output signal in form of translated GFP is created.

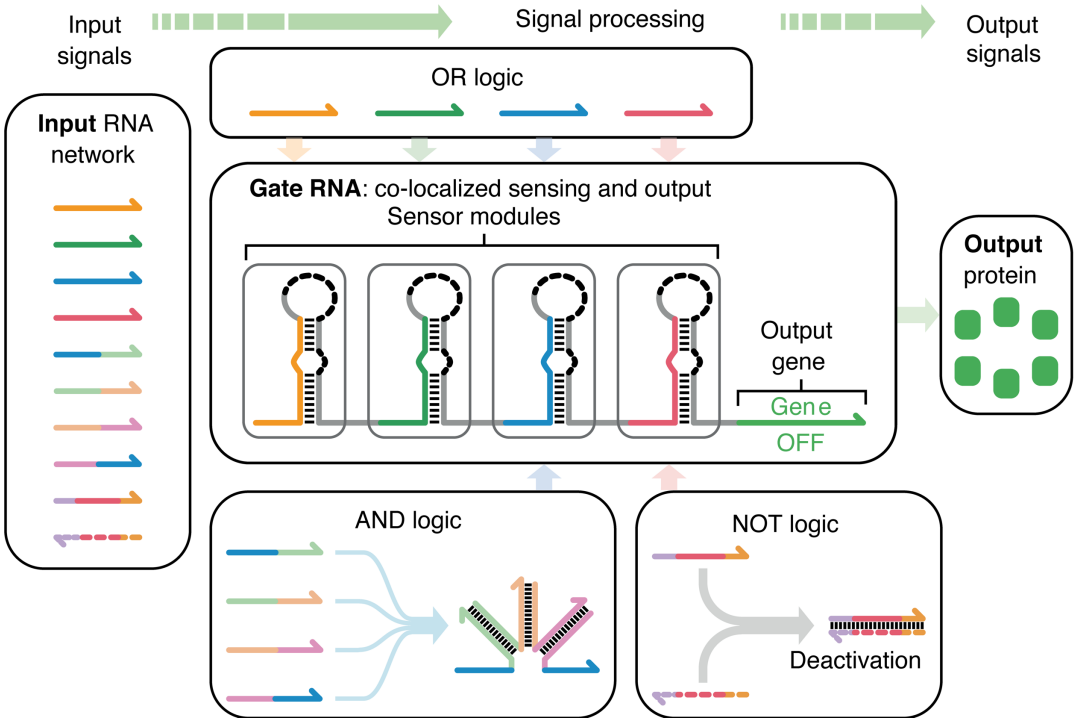


Figure 36. Implementation of toehold switches as ribocomputing devices able to perform OR, AND or NOT logic operations. A network of trigger RNA sequences serve as input which during signal processing create an output via GFP protein expression. The gate RNA, known as switch RNA serves as the central sensing unit being colocalized with the output sensor module. Modified from [169].

### Activation of toehold switches by functional mRNAs

The idea of controlling gene expression at the post-transcriptional level by RNA-RNA interactions can further be exploited towards real-world applications. Sequences originating from synthetic design approaches can be replaced by functional mRNA sequences. Design and reaction mechanism are shown in Figure 37. Obviously, mRNA sequences are fixed and switch-trigger duplex formation cannot be optimized thus making toehold binding more complicated. As a consequence, thorough beforehand *in silico* screening becomes mandatory. Existing secondary structures which are highly abundant in mRNAs negatively affect thermodynamics driving branch migration and can cause internal base pairing of toehold sequences or of sequences downstream of the hairpin module. To overcome such issues the thermodynamic equilibrium can be shifted towards an activated mRNA sensor by increasing the toehold length up to 30 nt (original toehold domain length is 12 nt). This strategy aims at

shifting the attention of single-stranded regions for initiation of binding from the trigger mRNA to the toehold switch itself. Another strategy can involve the design of a RNA refolding mechanism leading to a reduction of the energetic barrier for switch activation and an increased mRNA stability.

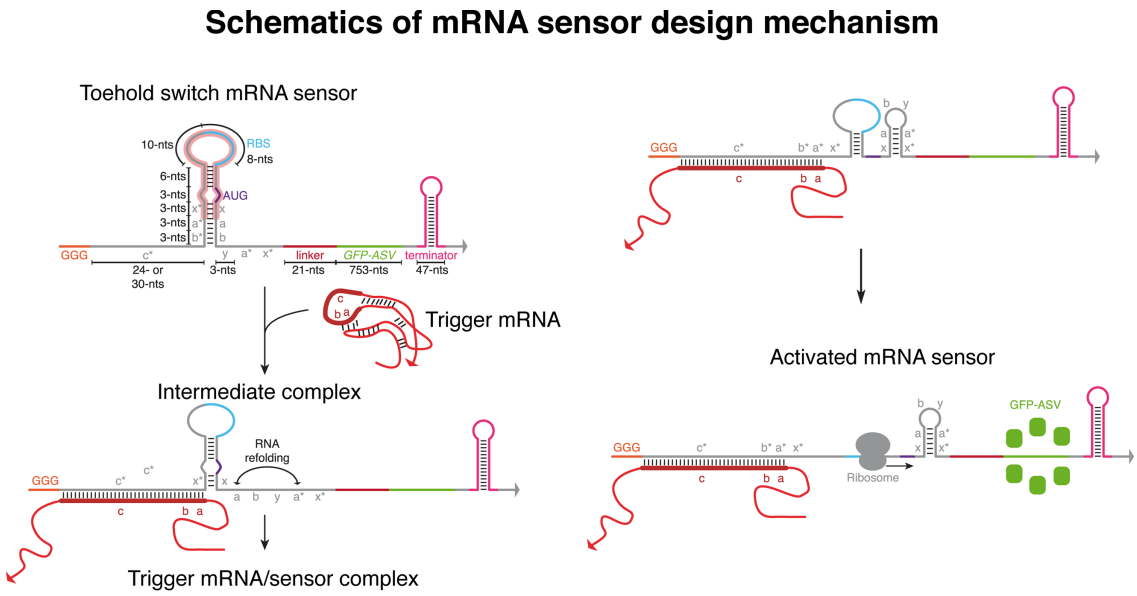


Figure 37. Schematics of mRNA sensor design mechanism. Upon binding of a trigger mRNA, an intermediate complex is formed with trigger mRNA bound to toehold binding domain c/c*, accompanied with an RNA refolding mechanism promoting the activated mRNA sensor state. Modified from [3].

The potential of toehold switches acting as sensors for functional mRNA sequences was also investigated in chapter 4.1. Here, binding of trigger mRNAs to forward-engineered toehold switches led to subsequent repression of GFP expression. Overall, the concept of toehold switches can be applied to more realistic settings when using naturally occurring mRNA sequences together with synthetic genetic circuits, thus opening this type of devices for diagnostic purposes.

### Applicability of toehold switches

The possibility to use naturally occurring mRNA sequences to trigger and activate toehold switches was used as an opportunity to implement this system into an *in vitro* based cell free platform [170]. There, the cell-free synthetic gene networks together with the toehold switch system were embedded onto the cellulose matrix of paper, opening up possibilities for *in vitro* diagnostics distributed outside a laboratory environment. In particular, gene parts of toehold switches as well as bacterial transcription and translation machinery including accessory

proteins can be freeze-dried onto paper. Pardee, Green and coworkers have shown that full length active mRNA sequences can be detected with paper freeze-dried toehold switches. Their potential use outside fundamental research was proven by designing toehold switches able to work as sensor for the viral pathogen Ebola. The designed toehold switch was able to distinguish between different strains of the pathogen. This technique offers a diagnostic tool which is inexpensive, distributable and easy to manufacture, having been proven to remain functional up to one year at room temperature. This approach was further expanded for detecting clinically relevant concentrations ([c] ≈3 fM) of Zika virus mRNA [171]. Again, the principle of toehold switch activation by cognate viral mRNA was applied onto cell-free protein expression platform. Detection was made visible for the naked eye by designing the sensors to regulate translation of lacZ enzyme which induces a change in color by converting a yellow substrate (chlorophenol red-b-D-galactopyranoside) into a purple product (chlorophenol red). This example of successful activation of the toehold switch sensor by taking a live Zika virus as initial input can be considered as a substantial innovation in the field of *in vitro* diagnostic tools and offers potential for further applications.

### 2.3.3 Recent advances in mammalian synthetic biology

So far, the focus of synthetic biology studies presented above was on implementing genetic circuits, and especially mechanisms for post-transcriptional control of gene expression in prokaryotic systems. Due to the significantly higher complexity of the transcription and translation machinery in eukaryotes (refer to chapter 2.2.2), similar advances have not yet been reached in mammalian cells. Mechanisms that have been proven to work in prokaryotic systems cannot simply be carried over into eukaryotic settings. Nevertheless, there have been recent advances in mammalian synthetic biology, such as control of transcription based on transcription factors, either naturally occurring ones (e.g. LacI, TetR) or programmable, synthetic ones (Zinc-finger protein containing factors, Transcription Activator Like Effectors). This was accompanied by the engineering of synthetic promoters [172].

The recently discovered CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) mechanism offers unforeseen possibilities for gene editing and control of gene expression. This adaptable immune mechanism is found in many bacteria, serving them as a protection against invading foreign nucleic acids, originating from viruses or plasmids [173]. Here, a Cas9 protein can specifically target DNA by binding a short guide RNA that is complementary to the desired DNA target sequence. Cas9 possesses endonuclease activity and thus can induce double strand breaks within the targeted genomic DNA. This new gene editing platform has been shown to efficiently target endogenous genes in bacteria, cell lines and even led to genetic alterations in whole organisms [174-176]. Its therapeutic potential is currently explored. Still these investigations involved control of gene expression on the DNA level.

Tools for RNA regulation on a post-transcriptional level mainly involve aptamers which are found in nature in the form of riboswitches. These motifs can, upon binding to small molecules and proteins, regulate activity of RNA effectors in *cis* or in *trans* by cleaving or splicing mechanisms [177]. Interestingly, insertion of RNA motifs which can be recognized by RNA-binding proteins close to the 5' cap affects translation efficiency [178]. Mammalian gene expression can also be controlled by small RNA motifs called ribozymes (e.g. Hammerhead ribozyme) which exhibit self-cleaving activity. When inserted at different positions of a reporter mRNA, these motifs result in mRNA cleavage, decay and reduced gene expression, see Figure 38 [179]. Another recent example of mammalian synthetic biology is a programmable single-cell mammalian biocomputer which integrates transcription and translation control components, transcription factors and RNA binding proteins, respectively. This system is

wired up into a combinatorial circuit with NOT, AND and NAND expression logic. Here, binding of input molecules to transcription factors and their subsequent inhibition leads to translation inhibition of output reporter gene [180].

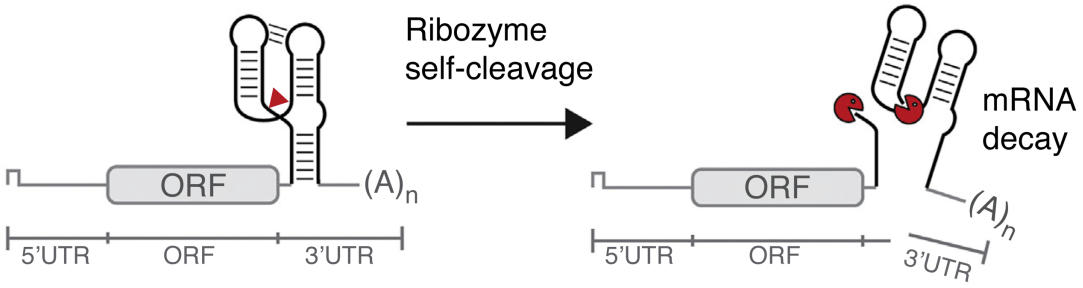


Figure 38. Ribozyme-based control of gene expression by insertion of Hammerhead ribozyme into 3' UTR of mammalian mRNA leads to its cleavage and consequently to decrease of gene expression. Schematics modified from [179].

However, post-transcriptionally controlling translation in mammalian systems solely relying on rationally designed RNA-RNA interactions as it has been successfully implemented in bacteria, (known as toehold switches) still remains unrivaled. To address this challenge was part of my project at the Wyss Institute For Biologically Inspired Engineering. Connected preliminary results on regulation of translation by insertion of RNA hairpin motifs close to 5' cap structure will be presented in chapter 4.2.

# 3 Robustness of localized DNA strand displacement cascades

The work of subsequent chapter is part of the "Sonderforschungsbereich" 1032 (SFB1032) and has been published in ACS Nano. It can be accessed online under following link: http://pubs.acs.org/doi/abs/10.1021/nn503073p

## 3.1 Motivation

Cellular environments in bacteria and especially in eukaryotic cells are crowded and possess complex properties. Consequently, in order to efficiently execute a reaction flux, it is known that colocalization of reactants can strongly improve kinetics and efficiency of chemical processes. Well-defined arrangement of reaction partners as it is the case for scaffolded signaling cascades or compartmentalization of organelles poses an essential advantage for effective signal propagation and dynamics of intra- and extracellular reaction processes.

Hence, the project is inspired to experimentally study the influence of spatial organization and local molecular interaction on the dynamics of reaction and signaling networks effects on a nanometer scale. By investigating the effect of spatial organization of reaction components, the potential benefit of colocalization and increased local concentration in a defined reaction vessel was evaluated. In order to study proximity effects of colocalized reaction components *in vitro*, the advantages of DNA nanotechnology, especially the DNA origami technique as a tool originating from structural DNA nanotechnology [2] and DNA strand displacement (DSD) reactions as a well-studied mechanism in dynamic DNA nanotechnology [54,58] were applied. These results will be presented in the following section.

Here, potential benefits of colocalization of reaction partners are analyzed by immobilizing a two-stage DSD reaction cascade comprised of a "sender" and a "receiver" gate onto a DNA origami platform (Figure 39). In this system, the addition of a DNA or RNA input strand triggers the release of an intermediate signal strand from the first gate of the cascade. The intermediate strand can activate a second gate on the DNA origami platform and release an output strand in a subsequent strand displacement process (Figure 42). Potentially, the cascade could be extended by more steps thus allowing more sophisticated circuits mimicking pathways with multiple loops like glycolysis and reaction processes such as amplifications or logic gates. In this context one has to consider that complexity of the design increases

disproportionally by the number of steps the cascade is designed to carry out. Consequently, undesired cross reactions have to be minimized and orthogonality of non-interacting domain regions needs to be optimized.

Results have shown that overall kinetics are considerably sped up for an intermediate spatial separation between the gates (21.5 nm), whereas colocalization does not have a significant effect on kinetics for larger distances (32 nm and 42.5 nm). At even smaller distances (10.5 nm), direct physical overlap of the gates results in strong intraorigami leak reactions, which precludes a stable preparation of the DSD cascades. Experiments revealed that for the smallest stable gate distance (21.5 nm) the DSD cascade is robust with respect to excessive amounts of interfering DNA strands that are partially sequence-complementary to the intermediate signal strand S. Despite the design of a single sender and receiver gate, a redundant architecture with four receiver gates immobilized on a single DNA origami platform was tested which turned out to considerably speed up the system accompanied with an improved robustness of the system. The findings gave insight into how proximity of immobilized neighboring DNA strands influences desired as well as undesired crosstalk. Potentially, these findings will contribute to a deeper understanding of cellular signaling pathways in *in vivo* conditions.

## 3.2 Outline and experimental setting

### 3.2.1 Assembly of the immobilized DSD cascade

**Geometric arrangement on the DNA origami platform**

To immobilize the components of the DSD cascade a scaffolded rectangular DNA origami substrate served as platform of choice. Representative schemes of the DNA origami structure used for this project are shown in Figure 39. The origami is a planar, single-layered structure which has been twist corrected in order to minimize curvature within the structure due to helical constraints, modified from [2]. As in the original DNA origami design the dimensions of the platform are ≈ 90 nm in length (264 bp) along the longitudinal axis and ≈ 65 nm in width resulting from 24 double helices in parallel assuming an effective helix diameter of around 3 nm and an averaged rise per base pair of 0.34 nm. In Figure 39A, the exact geometric arrangement of the cascade elements on the DNA origami platform is illustrated with distances between the gates indicated in base pairs using the DNA origami design software caDNAno [33]. Colored marks (A0 – H23) show the positions of protruding extended staples to which the cascade elements can be hybridized. C12 (in red) represents the central staple and starting point of the cascade (gate 1). Possible arrangements and distances of the end points of the cascade (gate 2) are designed to be 31 bp, 63 bp, 94 bp and 125 bp from the starting point which corresponds to 3, 6, 9, 12 full helical turns or to a distance 10.5 nm, 21.5 nm, 32 nm and 42.5 nm, respectively. Moreover, the cascade and its connecting extended staples were designed such that not only the distance was variable, but also the stoichiometry between sender and receivers for smaller gate distances with the hypothesis of potentially speeding up reaction kinetics. In Figure 39B, a top view of the immobilized DNA strand displacement cascade is shown with the DNA sender gate 1 being possibly hybridized to an extended staple at the position indicated in red. All investigated positions for receiver gates 2 are depicted in blue. In the three-dimensional illustration, the protruding adaptor sequences of the gate complexes to which sender and receiver gates can be hybridized are depicted as loose staple extensions (Figure 39C).

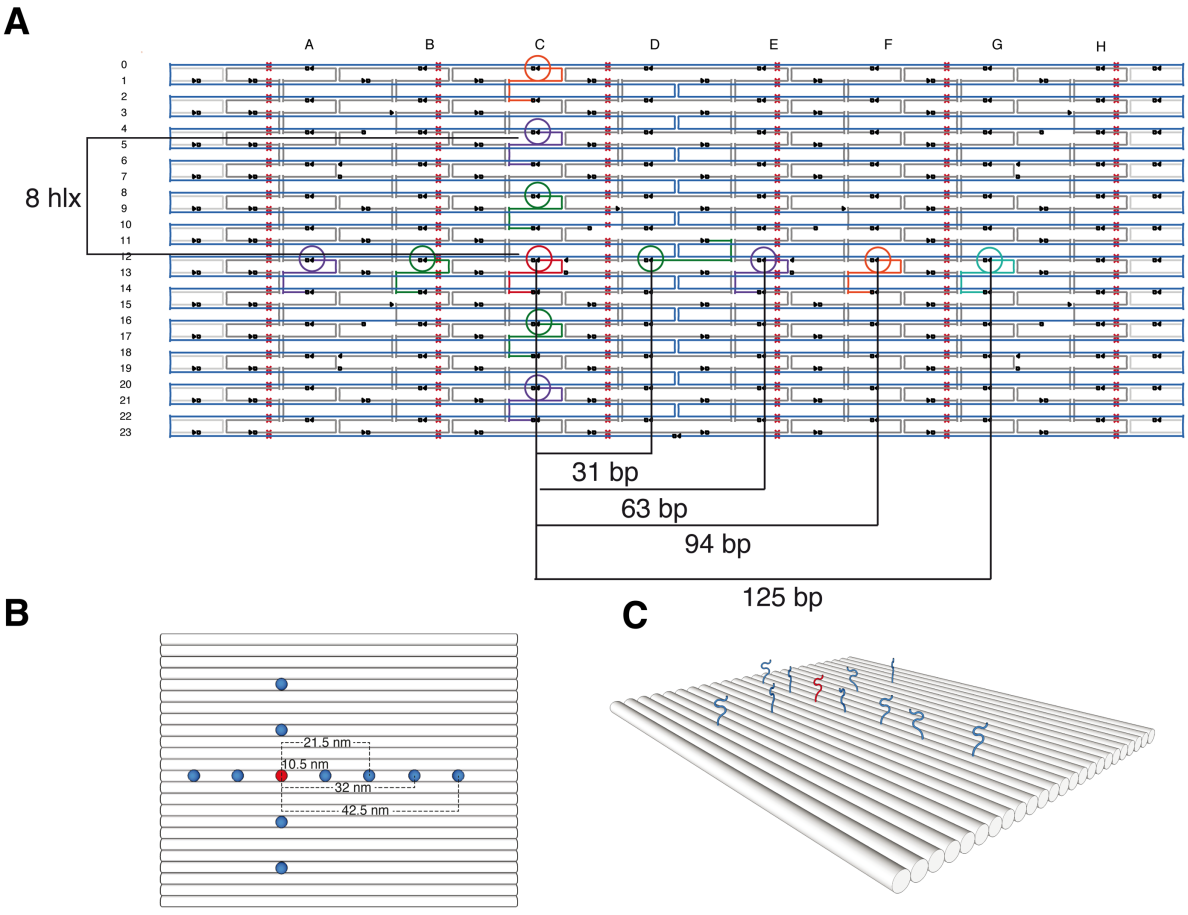


Figure 39. Design and illustration of the DNA origami structure used as scaffold for immobilization of sender and receiver gates. **A**: CaDNAno design of the DNA origami structure with the positions of extended staple strands, which are indicated as colored rings [33]. Positions are given in helix rows from A0 – H23. These serve as anchor sites for hybridization of cascade elements. The position at C12 (marked in red) indicates the starting position of the cascade to which gate 1 can be hybridized to. Extended staple strands to which gate 2 complexes can bind are positioned around gate 1 and were designated to exhibit distances of 10.5 nm (green), 21.5 nm (violet), 32 nm (orange) and 42.5 nm (turquoise). **B**: Topview of the rectangular structure –with anchor points for gate 1 (red) and gate 2 complexes (blue). Here calculated distances and geometric arrangements between gate 1 and gate 2 that were examined in this project are shown to be 10.5 nm, 21.5 nm, 32 nm and 42.5 nm. **C**: Three-dimensional illustration of the platform. Protruding staple extensions display the positions to which gate 1 (red position in the center) and gate 2 complexes (blue positions surrounding the center) can be hybridized to accordant individual settings.

### Attachment of the sender and receiver gate onto the DNA origami platform

In this part, the overall assembly process of the cascade and its attachment onto the DNA origami substrate is elaborated. In the initial state of the cascade after the completion of assembly, the sender gate is loaded with a signal strand S, and the receiver gate is hybridized to an output strand O (Figure 40**)**. Both S and O are modified with fluorescence quencher molecules (Q1 and Q2, respectively), which quench the fluorescence of the fluorophores F1 and F2 on the gate strands. These gate strands are hybridized to the staple extensions protruding from the DNA origami platform. Upon addition of an input strand I, signal strand S is released which in turn displaces strand O from gate 2 via toehold-mediated branch

migration. In principle, the "dummy"-output O could be extended by an additional output sequence in order to trigger downstream processes and thus serve as a "real" output making the design much more challenging though.

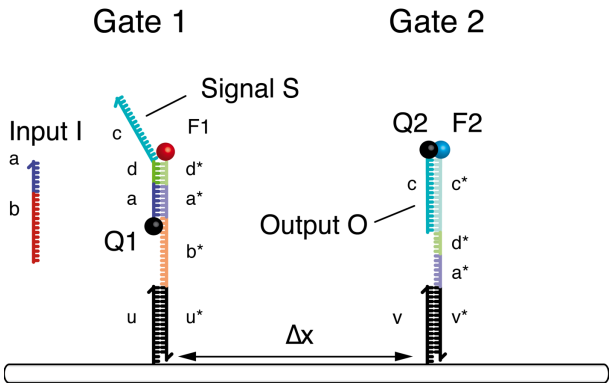


Figure 40. DNA strands and domains involved in the DSD cascade shown in its fully assembled initial state. Sender gate 1 is hybridized to a staple extension with sequence domain u, receiver gate 2 is hybridized to a staple extension v at a distance Δx from the sender. The sequence domain lengths are a: 7nt, b: 15 nt, c: 26 nt, d: 5 nt, u, v: 16 nt. For fluorescent readout, gate 1 and gate 2 are labeled with fluorophores F1 and F2, while signal and output strands carry fluorescence quenchers Q1 and Q2.

The architecture of the cascade is similar to that of previously developed toehold-exchange reactions [54,181]. The release of output may either occur on the same origami platform via local transfer of the signal strand S or on a different platform via interorigami diffusion of the signal strand. In order to feed the cascade with a "realistic" input potentially suitable for biosensing applications, the sequence of the input strand I was chosen to be that of the regulatory microRNA-21 (chapter 2.2.3). From a design perspective, this naturally occurring sequence is not optimized for the performance of the cascade and thus constrains the sequences for the remaining part of design.

**Assembly of the immobilized DSD cascade**

The different domains interacting with each other are already highlighted in Figure 40 and each play a crucial role for the assembly process. Intrinsic to the nature of the strand displacement process, the repeated use of DNA sequences becomes unavoidable. In the initial state of the cascade a*, c and d* occur twice. Experimental data show that simple colocalization of the gates on the DNA origami substrate already results in considerable strand displacement leakage, in which signal strand S is transferred from gate 1 to gate 2 even in the absence of input I (Figure 60). Follow up theoretical considerations have shown that due to these sequence constraints the assembly process has to be well-wrought and carried out

in a specific manner (Figure 41). The folded and purified DNA origami platform was initially incubated with a construct consisting of three previously assembled DNA sequences, termed F1, signal strand S (modified with Q1) and an auxiliary protection strand P (containing sequence domain c* and a toehold for deprotection) by hybridization with extended staple domain u (Figure 41A). Undesired gate interactions were strongly reduced by this passivation for inter-gate distances above 21.5 nm, but being still significant for the 10.5 nm distance (Figure 60, Figure 61).

The next step of the assembly process involved hybridization of the F2-Q2 construct representing the output strand O with extended staple domain v (Figure 41B). Even though the signal strand S was passivated by protection strand P, a subtle leakage throughout all tested configurations could not completely be excluded (see Figure 61). For that reason, incubation time was limited to 20 min at room temperature. The final step required the displacement of protection strand P by adding the deprotection strand P* in excess (containing domains z* and c) (Figure 41C and D). The deprotection process was monitored experimentally (Figure 62). Displacement of Q2 by I was explicitly excluded by the design of the sequences and was shown to have no impact on the fluorescence of F2 (Figure 65).

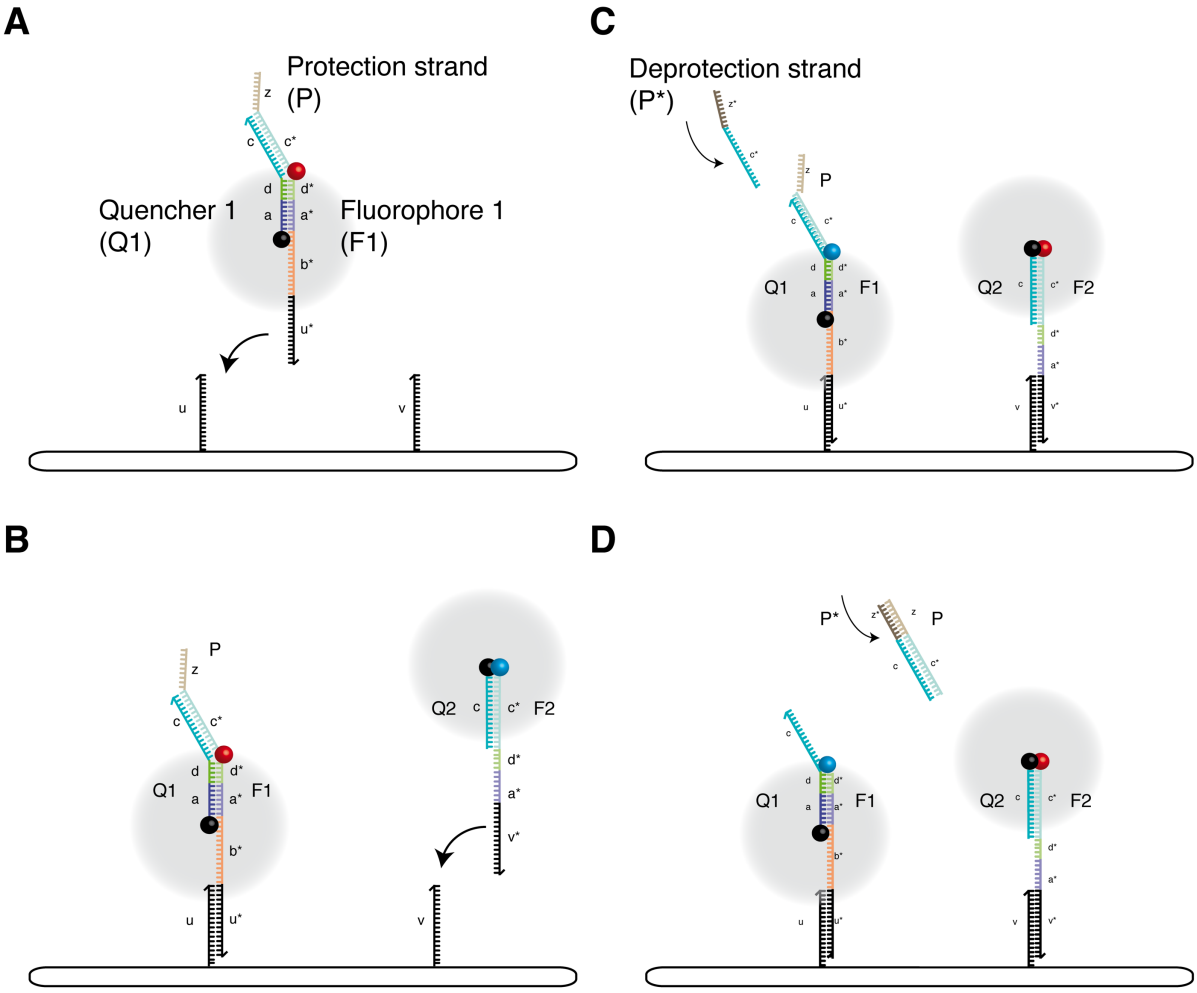


Figure 41. Assembly process of the immobilized DSD cascade. Schematic illustration of individual steps. **A**. Hybridization of the sender gate 1 with the protection strand P passivating domain c of signal strand S to the DNA origami solution, modified by extended staples with sequences u and v. **B**. After purification of the DNA origami substrate with gate 1 attached, receiver gate 2 is added, and after another incubation the solution is purified again. **C** and **D**. Finally, a deprotection strand P* is added of which its domains z* and c are complementary to the protection strand. This step turns the cascade into an activated state, to be triggered by input I.

### 3.2.2 Designated reaction process

In this section, the detailed reaction progress of the cascade is explained which can be divided into three main steps. Activation of the cascade is triggered by adding the input strand I to the solution (Figure 42). Input I will bind to toehold domain b*, at its 5' end resulting branch migration and subsequent displacement of signal strand S via toehold-mediated strand displacement reactions.

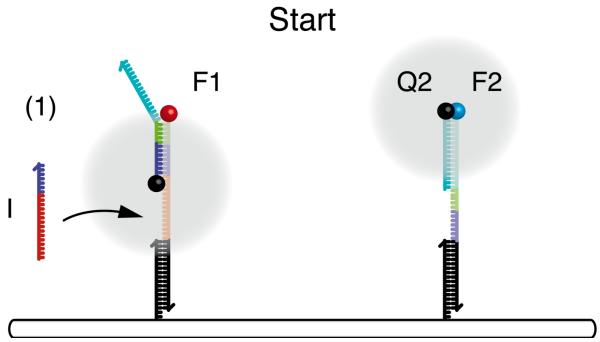


Figure 42. Progress of the reaction cascade. Activation of the cascade by addition of input I.

In this configuration, due to displacement of signal strand S, which is modified by the quencher Q1, the fluorescence signal of F1 increases. As a consequence, signal strand S freely diffuses within the reaction volume getting lost in bulk solution, or locally transfers to a receiver gate, either on the same DNA origami or on another DNA origami platform via inter-origami transfer (Figure 44). Here, thermodynamic aspects and sequence complementarity play an important role, posing some limitations to the system with respect to its scalability towards more complex cascades or circuits. Configuration of sender gate 1 before and after adding input strand I are shown in Figure 43. Also included is the minimum free energy (MFE) structure, calculated by NUPACK at 20°C [21]. For sender gate 1 with signal strand S bound, its minimum free energy is -46 kcal/mol, whereas a state in which input strand I is bound to sender gate 1 a minimum free energy of -60.4 kcal/mol is calculated thus representing a thermodynamically favored configuration.

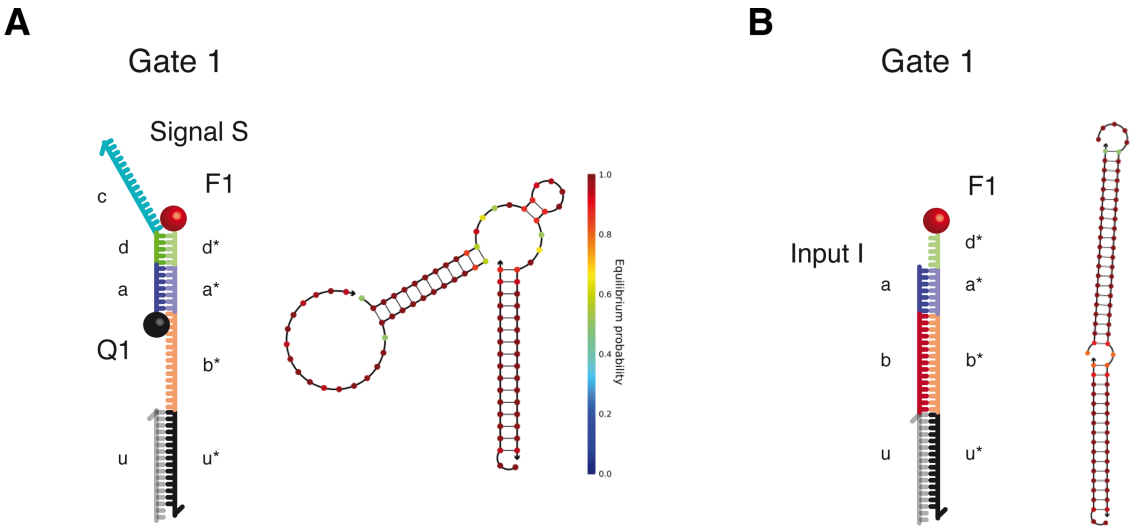


Figure 43. Composition of sender gate 1 before (**A**) and after (**B**) adding input strand I. Besides schematic illustrations, NUPACK calculations with MFE structures are included.

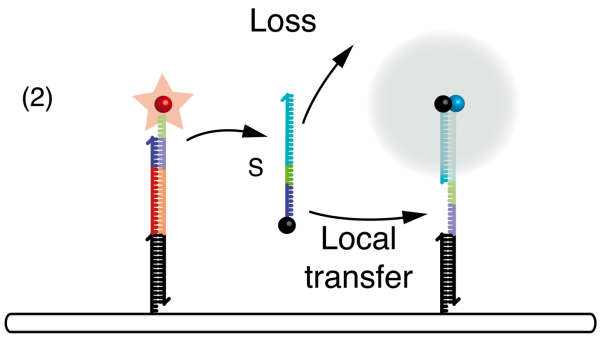


Figure 44. Progress of the reaction cascade. Release of signal strand S by Input strand I either causes its loss into bulk solution or its "local diffusion" and subsequent displacement of output strand O. Release of S is accompanied by a fluorescence increase of F1.

When signal strand S encounters receiver gate 2 – in the case of intraorigami interaction, signal strand S displaces output strand O via toehold-mediated strand displacement reaction, similar to initiation of the cascade. As for initial triggering of the cascade, this process is detectable in the fluorescence spectrometer as output strand O (modified with Q2) is displaced by signal strand S thus resulting in an increase of F2 (Figure 45). A closer look at receiver gate 2 including thermodynamic considerations is shown in Figure 46. Here the free energy of gate 2 in its initial configuration has been calculated to be -54.3 kcal/mol and in its final state with signal strand S bound -72.6 kcal/mol thus shifting equilibrium towards displacement of output strand.

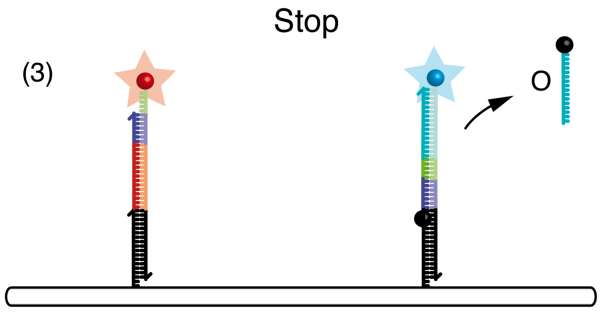


Figure 45. Progress of the reaction cascade. Release of output strand O leads to an increase in fluorescence of F2.

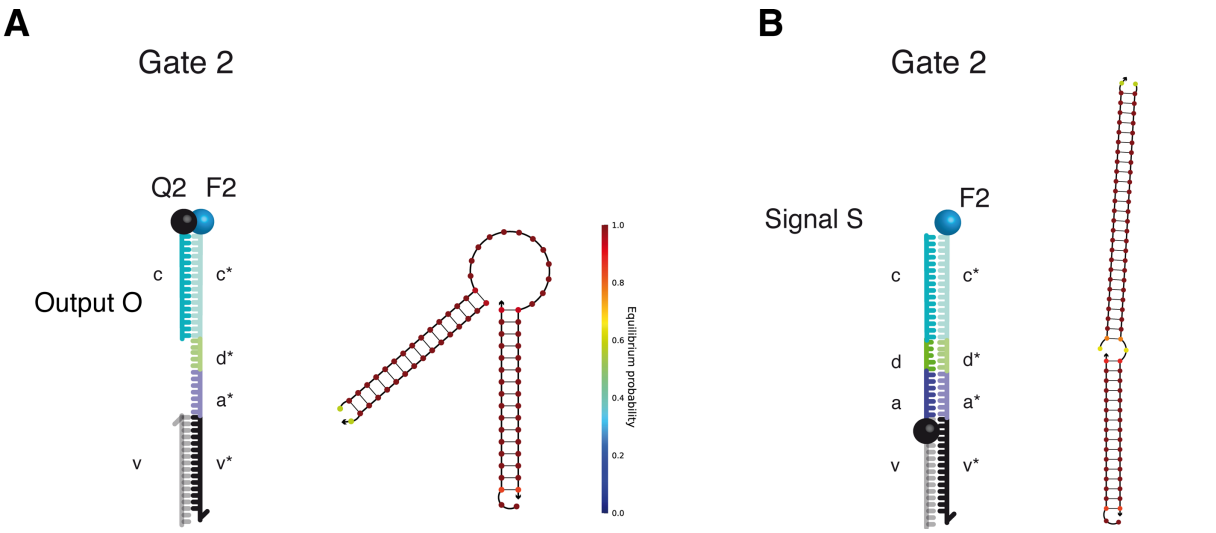


Figure 46. Composition of receiver gate 2 before (**A**) and after (**B**) signal strand S hybridized to F2. Besides schematic illustrations, NUPACK simulations with MFE structures are included.

## 3.3 Results

### 3.3.1 Distance dependence of cascade kinetics

One of the central goals of this project was the investigation of localized and cascaded DNA strand displacement reactions and its dependence on their geometrical and spatial arrangement on the DNA origami substrate. Distances were tested according to the helically possible attachment sites for gate 1 and gate 2. Studied distances were ranging from 10.5 nm to 42.5 nm for the smallest and largest distance, respectively. The distances were limited on the size and geometry of the DNA origami substrate. Measurements were mostly carried out by fluorescence spectroscopy.

**Normalization procedure for fluorescence spectroscopic experiments**

The preparation process and related experiments investigating its assembly revealed that intra-origami crosstalk strongly depends on the designated distance of sender and receiver gate (Figure 61). Especially, for the 10.5 nm gate distance, proximity-induced intraorigami strand transfer prevented the preparation of a defined initial state that was stable on the time-scale of the experiments. This also affected starting and end read out levels for the different gate distances. As a consequence, cascade kinetics was analyzed in detail only for the larger distances $\Delta x \approx 21.5$, 32 and 42.5 nm for which normalization was done.

All fluorescence traces were normalized to vary between a starting level $\xi\,(t = 0) = 0$ and an end level $\xi\,(t \rightarrow \infty) = 1$, where the fluorescence end levels were determined for each experiment individually. $\xi\,(t)$ represents the fractional progress of the displacement reactions. Respective end levels for F2 were determined for distances $\Delta x \approx 21.5$ nm, 32 nm, and 42.5 nm in a separate experiment by adding signal strand S in excess at t = 5700 s and resulting in toehold-mediated strand displacement of output strand O (Figure 47A). Besides normalization for F2, also the quenching efficiency of Q2 was analyzed (Figure 47B). A sample was prepared to which only F2 was hybridized to. At t = 0 s, Q2 was added in excess which – as expected resulted in a significant decrease in fluorescence by over 95%. This result was also taken into consideration thus further improving reliability and accuracy of the experiments. The gate distance of 10.5 nm was not normalized as strong leak reactions were observed. For the normalized comparison of gate distances, final values of $\xi$ were taken at t = 4500 s (Figure 49). Due to observed intraorigami leakage reactions, the starting fluorescence levels

varied slightly between the experiments with higher starting levels for the F2 signal for shorter gate distances, especially for the distance of 10.5 nm which revealed significant crosstalk – in particular during the assembly process.

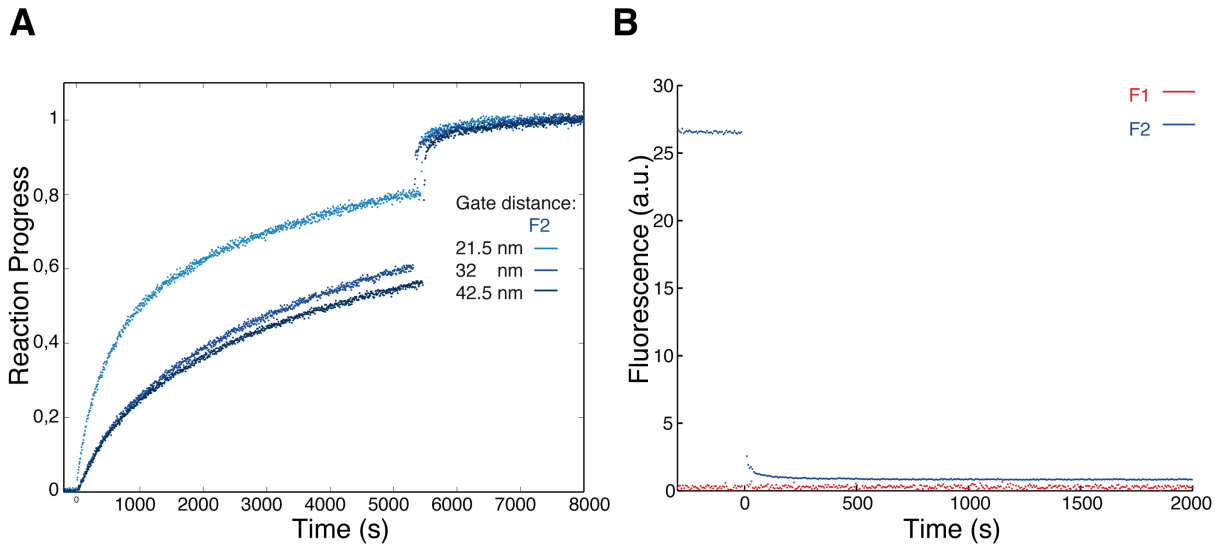


Figure 47. Normalization process. **A**. Normalization procedure for F2. At t = 5700 s, signal strand S was added in excess to the solution resulting in toehold-mediated strand displacement of output strand O. **B**. Investigation of the quenching efficiency of Q2 (output O). Addition of Q2 in excess at t = 0 s to a previously prepared DNA origami sample to which only F2 was hybridized to results in significant decrease in fluorescence. This was also taken into consideration for the normalization process necessary for Figure 49.

### A) Looking at kinetics of F1

Distance dependence of cascade kinetics will be discussed in the following with one receiver gate immobilized. In each case, the cascade was first activated by removal of the protection strands P from gate 1, and then triggered by the addition of input strands in large excess ([I] ≈ 470 nM) over the origami platforms present at [O] ≈ 4.8 nM (Figure 62). In addition to the normalized data shown in Figure 49, same data with 10.5 nm distance included are shown without having been normalized (Figure 50). In another experiment, cascade kinetics are analyzed in which protection strand P is still bound to signal strand S (Figure 51), mimicking a toehold-free strand displacement reaction. As expected, the kinetics of the first release step are independent of the distance between sender and receiver. Indeed, the corresponding fluorescence traces monitored by F1 were almost identical throughout all experiments. In this case, the kinetic curves could not be well fit with a simple rate law, but appeared to result from a superposition of several processes. It potentially also reflects contributions from several fluorescent species in different quenching states. As a rough measure of the reaction speed, the half-times t ½ of the reactions ($\xi\ (t_{1/2}) = 0.5$) were determined, which were found

to vary between 30 s and 40 s for DNA inputs, and were close to 100 s for RNA inputs. If the apparently more complicated release kinetics are ignored and the half-times are related to an effective second order rate constant $k_{on} = (c \cdot t_{1/2})^{-1}$ (with c = [I]), a $k_{on} \approx 6 \times 10^4\ M^{-1}s^{-1}$ and $\approx\ 2 \times 10^4\ M^{-1}s^{-1}$ for DNA and RNA are obtained, respectively. Known from literature, these values are in the lower range of typical hybridization reaction rates. One possible reason for these relatively slow kinetics may be found in the presence of secondary structure in the fixed sequence of input strand that is derived from the miR-21. To further analyze the different kinetics caused by DNA and RNA input, the MFE of the input I was calculated by using the NUPACK tool. It is known that RNA is more likely to form secondary structures, besides the fact that RNA-RNA interactions are more stable than DNA-DNA interactions. Indeed, the MFE of the DNA and RNA input, possessing the same sequence differ from each other. The free energy of the secondary structure of the DNA input is -2.71 kcal/mol which is almost half of the MFE of RNA input which is calculated to possess a free energy of -5.31 kcal/mol. The presence of such secondary structure in miR-21 might be one of the reasons for the relatively slow kinetics.

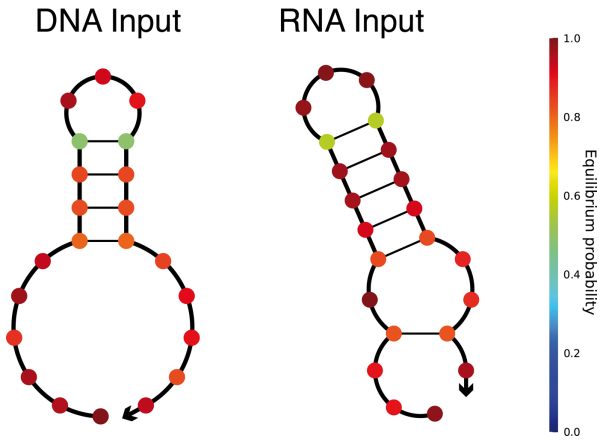


Figure 48. Minimum free energy structure of DNA Input (ΔG = -2.71 kcal/mol) and RNA input (ΔG = -5.31 kcal/mol), calculated by NUPACK [21].

**B) Looking at kinetics of F2**

The second step of the cascade turned out to be more complex as it involves the intermediate signal strand S, which triggers the release of the output strand O. In contrast to the first step, the signaling step appears to be sensitive to the distance between the gates. Especially for the DNA input, the 21.5 nm gate distance resulted in faster overall kinetics ($t_{1/2}$ = 1000 s) than for the larger distances (32 and 42.5 nm), for which the kinetics were relatively similar ($t_{1/2}$ =

3410 s and 4010 s, respectively). Interestingly, for all the RNA inputs, all three distances resulted in similarly slow reactions with half-times $t_{1/2}$ = 3770, 4000, 5000 s for 21.5, 32, 42.5 nm, respectively. The ratio of ≈ 100 between half-times for the slow signaling reactions (i.e., all but the DNA input Δx=21.5 nm case) and the release step corresponds very well to the concentration ratio between input and signal strands ([I]/[O] = 470 nM/4.8 nM ≈ 100). This observation strongly suggests that in these cases the intermediate signal strands are not transferred "locally" on the origami platforms, but are first released to the bulk and then return to the output gates in a conventional hybridization reaction. While the "slow" F2 curves in Figure 49 roughly follow second order kinetics, the signaling reaction obtained for the DNA input at a 21.5 nm gate distance appears to contain a fast initial phase followed by a slow phase with similar kinetics as for the larger gate distances. This indicates that in this case at least a fraction of the signaling strands is transferred locally, "on-platform", from input gate 1 to output gate 2. As will be discussed below, the most likely cause for this effect is a direct physical interaction of the gates at this distance, which allows local strand transfer through some intermediary complexes (Figure 69). Interestingly, such a local transfer does not appear to be possible for the RNA input at the same gate distance. In contrast to the DNA case, upon hybridization of RNA input to gate 1, a slightly shorter A-form helical section (with ≈ 0.24-0.28 nm/bp as opposed to 0.34 nm/bp in B-DNA) with sequence b/b* will be formed ([14,15]. Conceivably, this prevents a direct physical interaction of gate 1 and gate 2. Alternatively, the more strongly binding RNA input may displace signal strand from gate 1 more effectively, and thus preclude the formation of an intermediate complex.

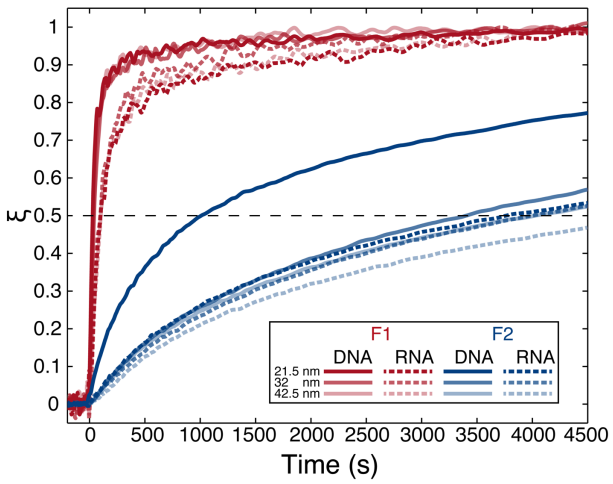


Figure 49. Progress of the cascade reaction with DNA and RNA inputs for distances Δx ≈ 21.5 nm, 32 nm, and 42.5 nm. Fluorescence of sender gate fluorophore F1 and the receiver fluorophore F2 (see figures above) are shown for the various distances as indicated by the colors. Release of the signal strand from gate 1 (red traces) is ≈ 100 times faster than capture of the signal strands (blue traces), as input strands are added in 100-fold excess. The trace recorded for the Δx ≈ 21.5 nm gate

distance for DNA inputs displays a fast initial phase, which may be interpreted by a “local”, intra-origami signal strand transfer.

### C) Kinetics of F2 (not normalized)

All major conclusions regarding cascade kinetics in this project are based on measurements shown in Figure 49 for which the described normalization procedure was applied. Traces represented in Figure 50 show same data as in Figure 49 including the shortest gate distance of 10.5 nm being not normalized. It becomes visible that starting levels of F2 increase with decreasing gate distance. Especially the 10.5 nm case suffered from significant crosstalk accompanied by unwanted side reactions via intra-origami leak interaction during assembly. This gives justification why this distance was not further considered for kinetic analysis.

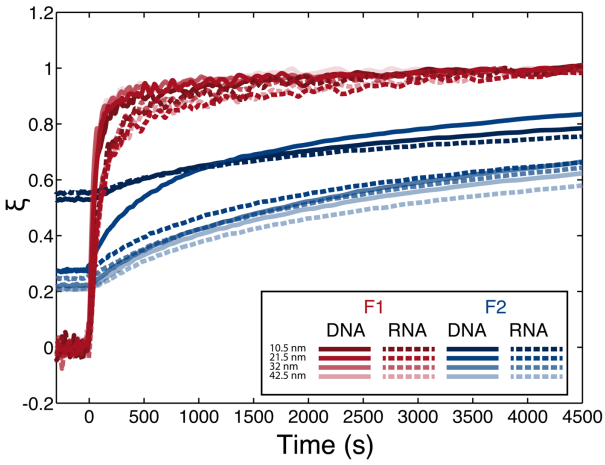


Figure 50. Cascade reaction for distances Δx ≈ 10.5 nm, 21.5 nm, 32 nm, and 42.5 nm. In contrast to Figure 49 above, the fluorescence traces for F2 were not normalized to 0, which shows their differing starting levels for the different gate distances. The 10.5 nm distance obviously is strongly affected by the intra-origami leak interaction during the assembly and is not further analyzed in terms of kinetics.

### D) Special case: Protection strand still bound to signal strand S

A situation was investigated in which the protection strand has not been displaced from gate 1, which usually represents the last step in the assembly protocol. This should mimic a toehold-free strand displacement process. As before, kinetics of the signal release step is independent of the distance between the gates whereas kinetics of F2 differed quite a bit from above data. With protection strand P still hybridized to signal strand S domain c remains passivated. Nevertheless, toehold domain b/b* is still accessible for invading input strand I. Strand displacement of the output strand O is considerably slowed down, possibly due to the

partially double-stranded character of S-P resulting in almost linear kinetics (Figure 51). Moreover, no increased speed for the 21.5 nm construct could be observed as it has been shown for the "unprotected" samples. Nonetheless, consistent with previous observations, for all distances, starting levels for F1 and F2 increased with decreasing distance of sender and receiver gate.

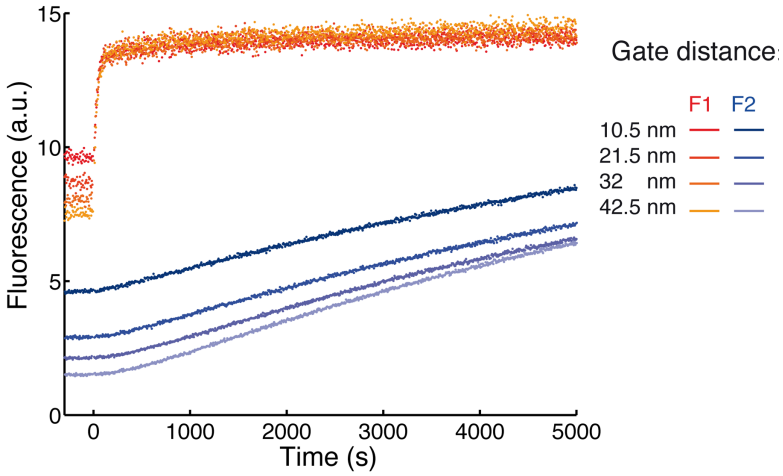


Figure 51. Investigation of the kinetics of the DSD cascade on the DNA origami platform with the protection strand P still bound to gate 1. Addition of the input (t = 0 s) results in strongly reduced strand displacement of the output by construct S-P due to its partially double-stranded character. The fluorescence starting levels increase with decreasing distance between gate 1 and gate 2.

## Gel electrophoretic characterization of the immobilized DSD cascade

The immobilized DSD cascade was also analyzed via gel electrophoresis (Figure 52). First it was confirmed that the DSD cascades were successfully immobilized onto the correctly folded DNA origami structures (yellow). Second, it could be shown that reactions take place on the origami structures. Addition of input strand I resulted in an increase of fluorescence of F2 on the gel bands representing the immobilized DSD cascade reactions (blue), clearly visible for the gate distance of 1 x 21.5 nm (lanes 2 and 3).

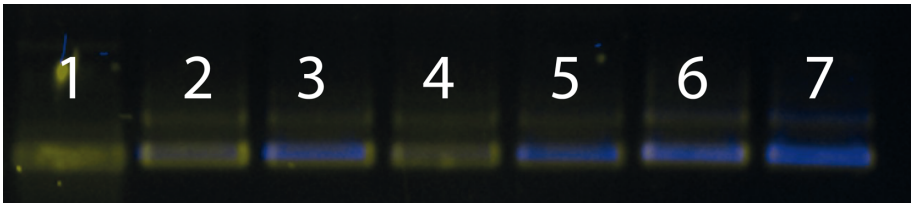


Figure 52. Gel electrophoresis of the DNA origami structures before and after addition of input strand I and incubation for 45 min. The image is an overlay of a laser scanning image at 532 nm for tracking F2 (blue) and an ordinary gel photograph imaged after SybrGold staining (yellow). **1**. Scaffold (M13mp18), **2**. 1 x 21.5 nm without input, **3**. 1 x 21.5 nm with input, **4**. 1 x 42.5 nm without input, **5**. 1 x 42.5 nm with input, **6**. 4 x 21.5 nm without input, **7**. 4 x 21.5 nm with input. The F2

intensity in lane 7 is higher than in lane 6, but there is considerable background caused by leakage and the presence of 4 labels in this case.

### 3.3.2 Robustness of the cascade

In order to further test the hypothesis of a "local transfer" of signal strands and to analyze its robustness against disturbing DNA strands, a series of competition experiments with origami-supported cascades for the different distances was performed. The undisturbed reaction pathway of the cascade has already been described above. Again, one has to discriminate between inter- and intraorigami diffusion which is always prone to diffusional loss. The schematic drawing in Figure 53 illustrates the idea of introduction of competing DNA strands. In these experiments, an increasing amount of competitor strands is added to the solution (together with input I), containing sequence domains a*-d* thus being partially complementary to the signal strands. A potential release of signal strands by the competitor strands *via* strand invasion at the input gate was found to be negligible (Figure 67). In theory, this interaction has been designed to be strong enough for disturbing the performance of the DSD cascade, consequently inhibiting "successful", direct transfer of signal strand S to gate 2.

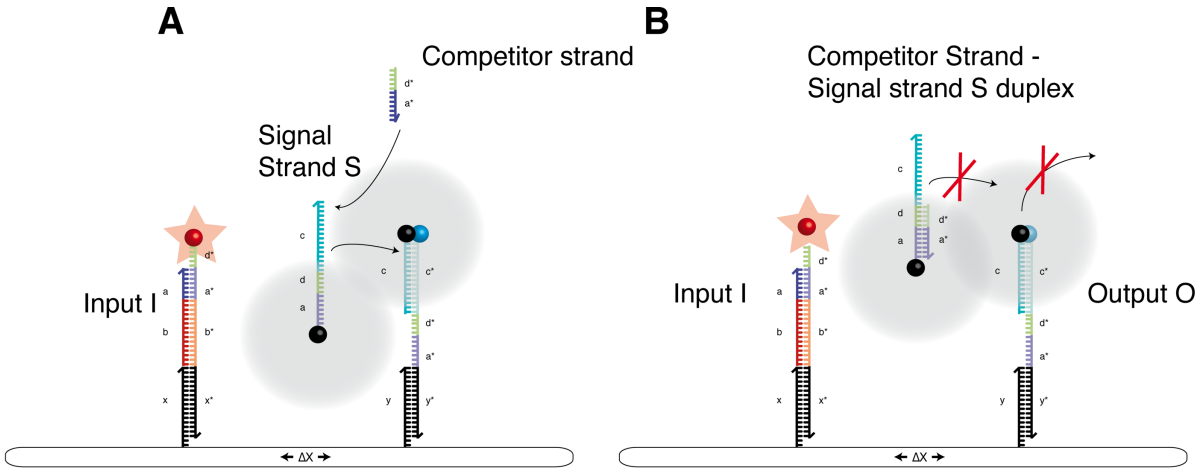


Figure 53. Expected reaction process after adding competitor strands together with input strand I to the system. **A**. Competitor strand is partially complementary to signal strand S and thus can hybridize to it within domains d and a. **B**. Duplex formation of signal strand S with the competitor strand interferes with displacement of output O.

The traces were expected to look like the following simple schematic drawings. Figure 54A reflects the performance of the cascade without competition strand whereas Figure 54B shows how the cascade might perform after introduction of competing strands with increasing concentration. It was expected that a globally diffusing fraction of the signal strands would be

affected by the presence of competitors more strongly than a potential "local" fraction. Here it is supposed that competitor strands possibly hybridize to freely diffusing DNA strands which are not directly interacting with the cascade gates at distances greater than 20 nm thus interrupting the system.

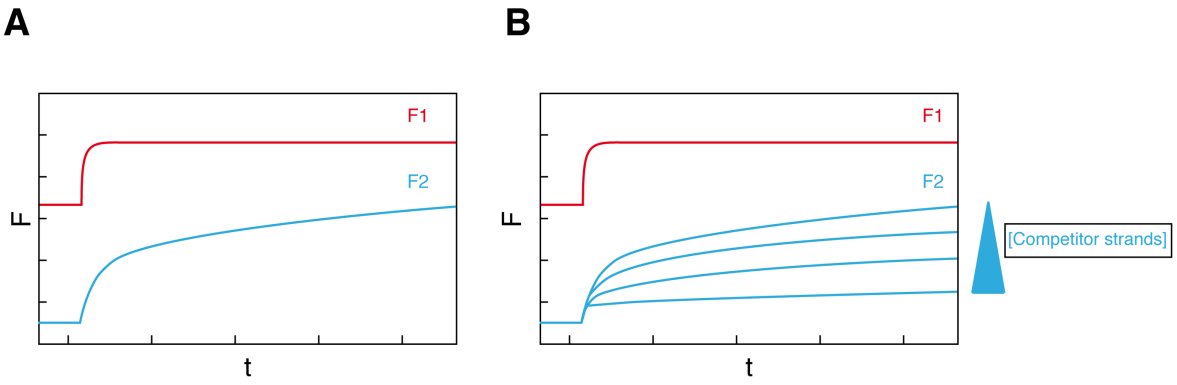


Figure 54. Basic illustration showing the expected impact of the introduction of the competitor strands. **A**. Performance of the cascade without competing strands in an undisturbed setting. **B**. Expected performance of the cascade with increasing concentrations of competitor strands.

Indeed, as it is shown in Figure 55 the release of output strands is almost completely suppressed for large amounts of competitor strands (≈ 1 μM) for the 32 nm and 42.5 nm gate distances, whereas still a significant fraction of output is generated for the 21.5 nm distance even at excessive competitor concentrations (3.7 μM). This is consistent with the assumption that the local transfer of signal strands for these distances occurs faster than hybridization with a competitor strand. Thus, colocalization at 21.5 nm not only speeds up the reaction and allows a fast local transfer of signal strands, but also significantly increases the robustness of the cascade with respect to a competing reaction pathway. This may be seen as a "dynamic" robustness as opposed to the "static" robustness of the larger gate distances with respect to intra-origami leakage. We may interpret the fraction of signal $f$ in Figure 55A that cannot be suppressed by the addition of competitor strands (marked in red) as caused by the locally transferred signaling strands. Hence, a strand transfer efficiency η can be defined as a function of competitor strands $\eta([C]) = \frac{f([C])-1}{f(0)-1}$. The locally transferred fraction then corresponds to the value $\eta_\infty$ for long times and at excessive competitor concentrations. In fact, the kinetics of this fraction alone is identical to that of the release step, which indicates that for the local signaling process input binding is rate-limiting.

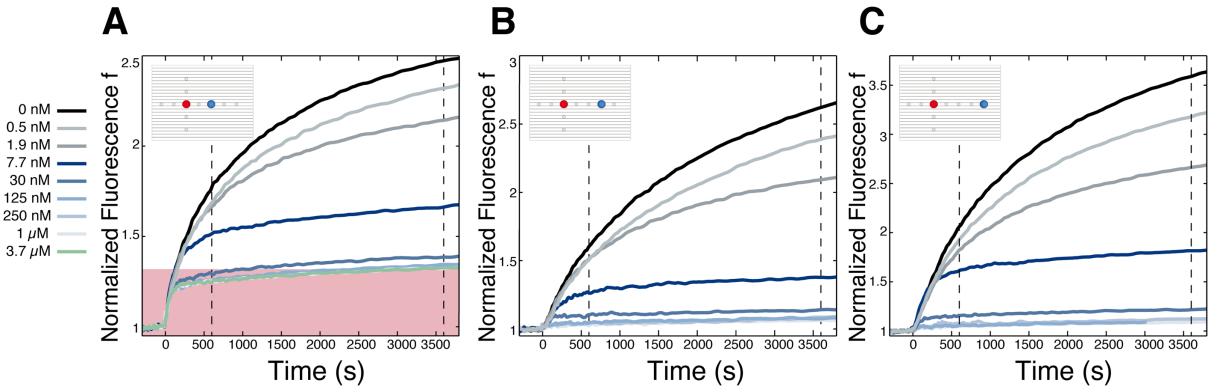


Figure 55. Probing strand transfer robustness with competitor strands partially complementary to the signal strands for distances **A**. 21.5 nm, **B**. 32 nm, **C**. 42.5 nm. Fluorescence signals f are normalized to their respective starting values. The signal of the output fluorophore F2 is strongly affected by the interference with competitor strands. However, while the signal is completely suppressed for excessive competitor concentrations in **B** and **C**, it remains finite for the shorter gate distance $\Delta x = 21.5$ nm in **A**. The kinetics of the robust "local" signaling fraction (marked in red) is identical to that of the release process shown in Figure 49.

These findings are also supported by experiments investigating an inter-origami diffusion for which two DNA origami platforms were prepared, each modified with either the sender gate 1 or the receiver gate 2 (see Figure 56). These two specimens were mixed in 1:1 stoichiometry. Subsequent measurements with different concentrations of competition strands revealed that at high concentrations (≈ 1 μM) the cascade is heavily disturbed and its robustness comparable to the larger distances of 32 nm and 42.5 nm, clearly differing from the gate distance of 21.5 nm.

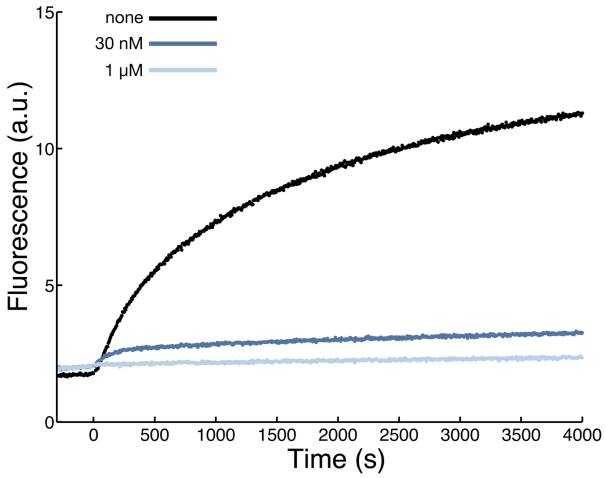


Figure 56. Preparing two origami platforms each either modified with sender or receiver gate. Study of inter-origami diffusion. In the presence of high competitor strand concentrations, the cascade is highly disturbed, comparable with distances $\Delta x \approx 32$ nm and $\Delta x \approx 42.5$ nm.

### 3.3.3 Redundant receiver gates

The overall performance of the cascade can be further improved by introducing redundant receiver gates on the origami platforms. A larger number of colocalized targets is expected to increase the probability of locally transferring signal strands and thus decrease the diffusive loss of signals to the bulk. For the competition experiment shown in Figure 57, four copies of gate 2 were immobilized on the DNA origami platform, each ≈ 21.5 nm apart from the sender gate 1. Compared to the case of a single receiver gate at the same distance, the kinetics appears to be faster and the displacement of the output strand O shows a fast convergence to the maximum fluorescence level of F2. Moreover, the larger number of receiver gates seems to further increase the robustness of the cascade with respect to high competitor strand concentrations. As before, a similar experiment for the 10.5 nm gate distance (Figure 58) was corrupted by excessive leakage already during the preparation of the cascade, while additional receivers at 32 nm distance (data not shown) did not reveal any effect. In Figure 55 and Figure 57 relative fluorescence values normalized to the initial fluorescence are displayed rather than the reaction progress $\xi$ defined above. It can be seen that in the absence of competitor strands the overall fluorescence increase is highest for the larger distances, whereas it is lowest for the multiple receiver case. This is caused by the correspondingly higher starting fluorescence values for the shorter distances, which in turn results from intra-origami leakage during preparation and purification (Figure 62). Furthermore, due to the presence of four (partially quenched) fluorophores F2 in Figure 57 the relative change with respect to the initial value is smaller.

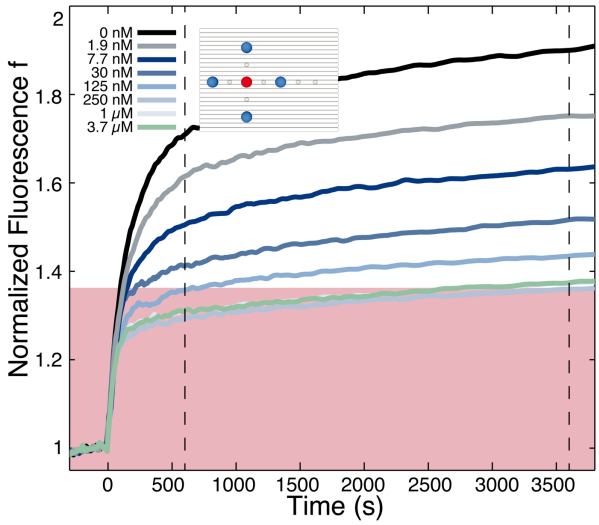


Figure 57. A competitor strand assay performed for a cascade with four receiver gates at a distance of Δx ≈ 21.5 nm. In terms of "surviving" signal strand fraction transferred even in the presence of excessive competitor strands, the redundant cascade is more robust.

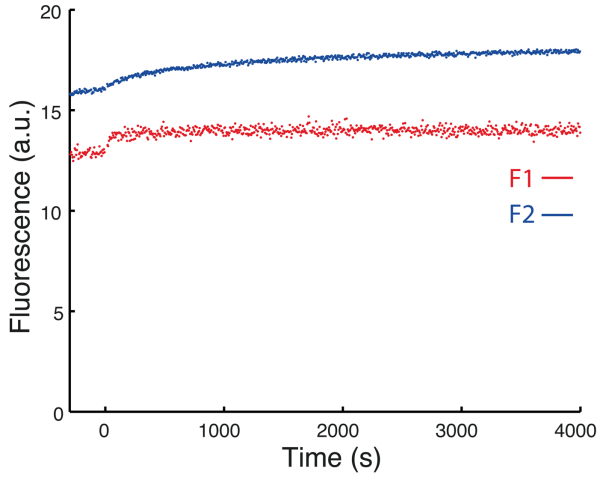


Figure 58. Competitor experiment assay for a cascade with four receiver gates at a distance of Δx ≈ 10.5 nm. Here, data are not normalized in order to underline the very high starting levels of F1 and F2 as a result of the significant crosstalk during assembly which made quantitative analysis impossible.

A summary of the competition experiments is shown in Figure 59, where transfer efficiencies $\eta([C])$ defined above are compared for the various cascade geometries at two different time points. The higher dynamic robustness of the shorter gate distance is clearly reflected. Specifically, for the redundant cascade geometry with four output gates at $\Delta x \approx 21.5$ nm, more than $\eta = 40\%$ of the available signal strands are transferred even in the presence of excessive concentrations of interfering strands. In terms of η, there is essentially no difference between the 32 nm and the 42.5 nm gate distance. The efficiency of signal transfer in the presence of competitor strands is expected to scale with their concentration [C] as:

$\eta\,([C]) = \eta_\infty + \frac{1-\eta_\infty}{1+\gamma[C]}$, where $\gamma = \frac{k_c}{k_{on}[O]}$ is determined by the ratio of the reaction rates between sender and competitor strands $k_c$, the rate $k_{on}$ for hybridization on the origami platforms and their concentration [O]. As can be seen in Figure 59, the measured values for $\eta$ at $t = 60$ min (which approximate the long-term equilibrium values) follow this behavior very well.

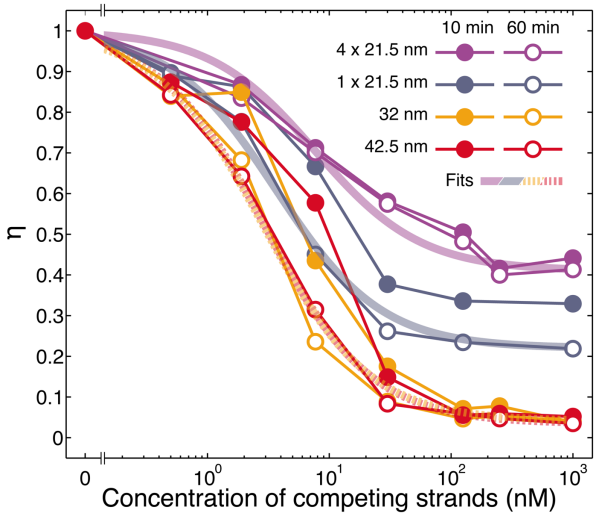


Figure 59. Effect of competing strands on the efficiency of signal transfer for the different distances and spatial arrangements on the DNA origami platform. η is defined as the relative fluorescence increase f-1 in the presence of competitor strands normalized to the undisturbed signal at the same time point (η is shown here for t = 10 min and t = 60 min to also capture kinetic effects). Signal transfer is most efficient for the 21.5 nm gate distance. In the redundant case (with four receiver gates), η is constant after 10 min. The thick, continuous lines display the theoretically expected reduction of η at long times with increasing competitor concentration. The curves are fits of the simple competition model described in the text to the "long term" values η (t = 60 min). The corresponding fit parameters were $\eta_\infty$ = 0.4, 0.2, 0, 0 and γ = 0.13 $nM^{-1}$, 0.27 $nM^{-1}$, 0.34 $nM^{-1}$, 0.33 $nM^{-1}$ for the 4 x 21.5 nm, 1 x 21.5 nm, 32 nm and 42.5 nm case, respectively.

### 3.3.4 Experimental investigation of the assembly

Experiments disclosed that especially during the assembly of the cascade at distances smaller than 22 nm, considerable crosstalk could not be avoided. The introduction of the protection strand P significantly reduced those unwanted leak reactions. Nevertheless, it was of particular interest to monitor the assembly process of the cascade in the fluorescence spectrometer thereby supporting assumptions stated before. Gained insights significantly helped to improve the overall performance of the cascade.

## The importance of the protection strand

### Cascade leakage during assembly without the protection strand P

During the early stage of this project, sender and receiver gates were added simultaneously to the DNA origami solution. At this early stage of the project one was not aware of the fact that mixing together both gates at the same time would cause extreme leak reactions before considering any actual experiment. Preliminary fluorescence-spectrometric experiments carried out in that way did result in a minimal absolute change in fluorescence. Consequently, an improved protocol was established, involving a stepwise addition of the gates to the DNA origami substrates. First the sender gate 1 was hybridized to the purified DNA origami solution to which after another purification the receiver gates 2 were added, followed by an incubation of the solution for 2 h at room temperature. The result is shown for a gate distance of 21.5 nm (Figure 60). Even in a sequential addition of the gates and a short incubation for 20 min, the undesired crosstalk during the actual experiment turned out to remain significant.

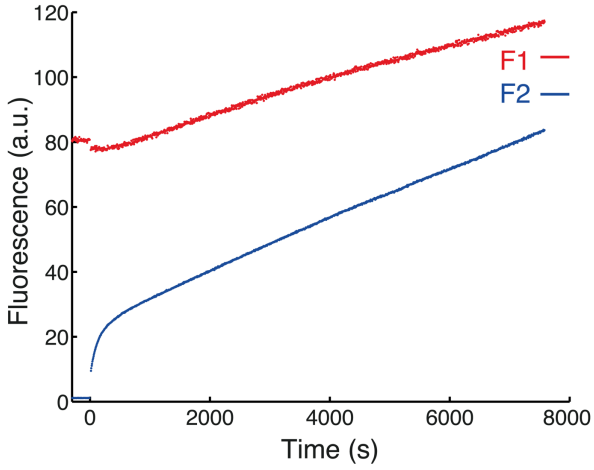


Figure 60. Cascade leakage without the protection strand P attached to gate 1 for a gate distance of 21.5 nm. F2-Q2 (t = 0 s) was added in 2x excess to the DNA origami structure (40 nM) and the change of F1 and F2 was monitored. In the absence of P, a very strong interaction of gate 1 and gate 2 was observed. This underlines the importance of the protection strand.

### Cascade leakage during assembly with protection strand P

By interpreting the results, a protection strand was introduced which was designed to passivate domain c. Subsequently, leakage was recorded for all four distances within a time frame of 3.5 h. Consistent to related experiments, leakage increased with decreasing gate distance and was highest for the 10.5 nm distance, again proving its unavoidable physical interaction. Interestingly, starting values for fluorophore 2 (gate 2) were at the same level.

This is an indication of comparable and reliable sample preparation. Among others, this experiment proved that a gate distance smaller than 20 nm is not suitable for this particular project outline. Overall, the introduction of the protection strand significantly improved the performance of the cascade, even though, the assembly protocol now involved more steps.

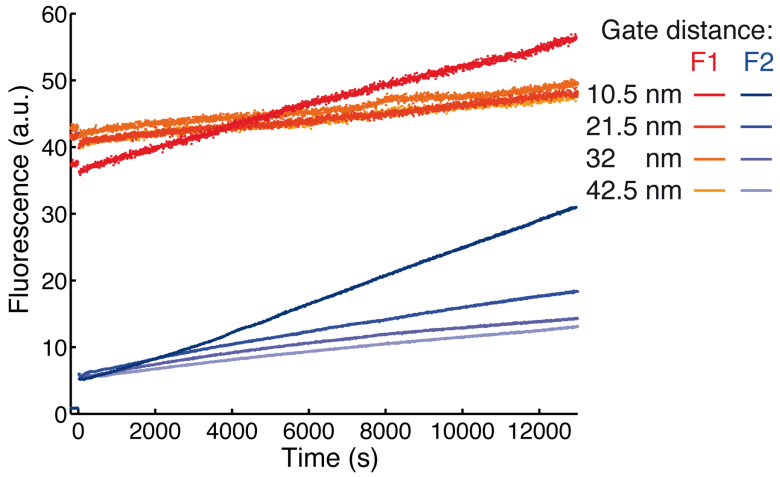


Figure 61. Cascade leakage during assembly. F2-Q2 (t = 0 s) was added in 2x excess to the DNA origami structure (30 nM) and the change of F1 and F2 was monitored over 3.5 hours. Reducing the gate distances results in a strong physical interaction of the gate complexes. The 10.5 nm distance showed considerable leakage and thus was not taken into further consideration for kinetic analysis.

**Deprotection of gate 1 and delayed triggering of the cascade**

Apart from observing the assembly process "in real-time" it was also interesting to resolve the subsequent deprotection process and triggering of the cascade. These two important steps were carried out in one single experiment. The complete process – for all four cascade distances is depicted in Figure 62. Again, starting levels of fluorescence increase with decreasing gate distance with the 10.5 nm distance standing out. Interestingly, the deprotection process only had a significant effect on the fluorescence of F2 which rapidly increased by addition of the input strand I.

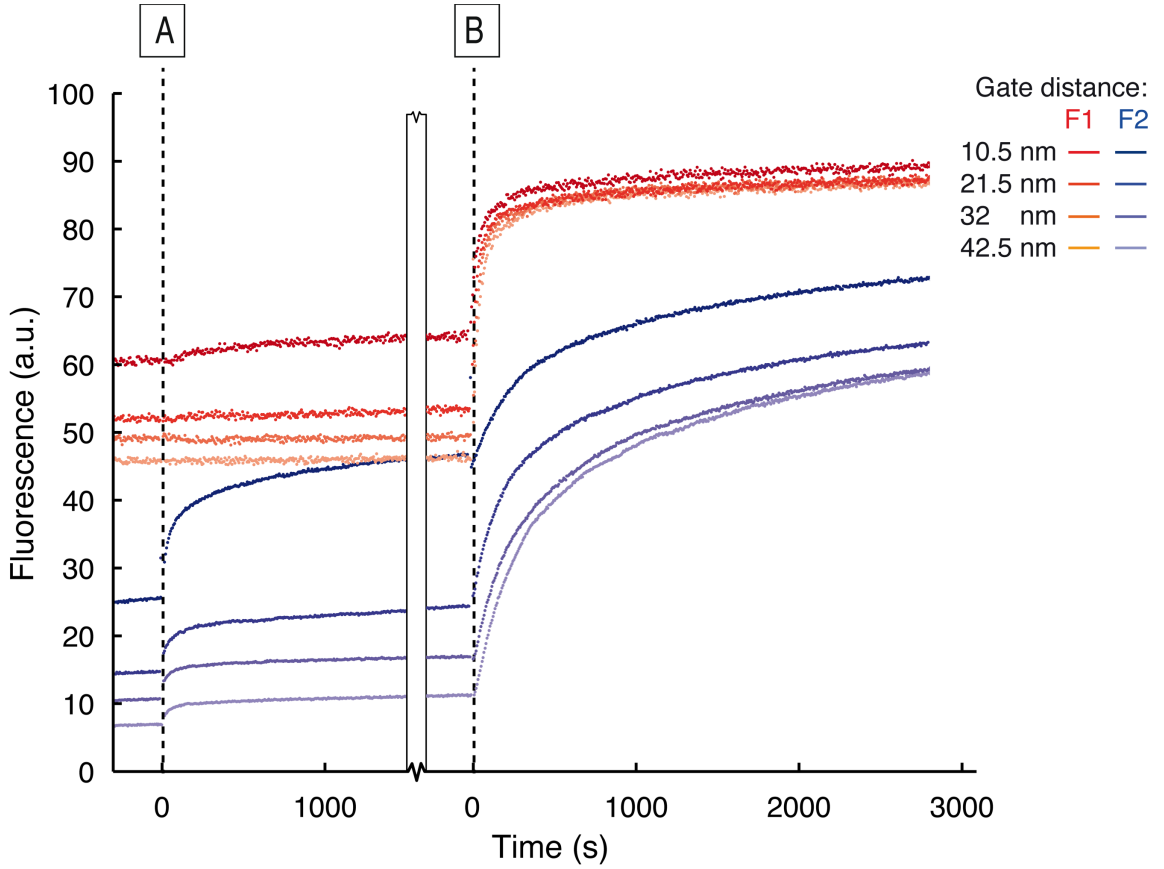


Figure 62. Deprotection of gate 1 and delayed triggering of the cascade. At (**A**), 90 nM of the deprotection strand was added to 30 nM of assembled DNA origami structures. After 30 min, (**B**), 470 nM of input was added to initiate the cascade (the curves were aligned to the same starting point). These two reactions were performed for all four different distances in one experiment. The starting levels of fluorescence increases with decreasing the distance of gate 1 and gate 2 thus proving again their cross talk during the assembly process, especially the 10.5 nm distance gate.

### 3.3.5 Further characterization of cascade performance

#### AFM imaging of the immobilized DSD cascade

Besides characterization of the cascade by fluorescence spectroscopy, it was also intended to visualize immobilized DSD gates via atomic force microscopy (AFM). Initially, AFM imaging was performed with the DSD gates immobilized on the DNA origami substrate without any changes within the fabrication process. Unfortunately, the gate complexes could not be resolved properly. This observation becomes obvious as the flexibility of gate complexes due to their partial double-stranded character hampers exact high resolution imaging. As a consequence, a different strategy was applied that turned out to be more promising. In this case, gate constructs were "mimicked" and replaced by DNA oligonucleotides modified with biotin at their 3' end. These 16 nt long oligonucleotides were hybridized to the extended staple strands of the DNA origami structures to which streptavidin

protein solution was added (Figure 63). As the interaction of streptavidin and biotin is one of the strongest non-covalent interactions in nature ($K_D \approx 10^{-14}$ M), we expected to image streptavidin molecules bound to biotinylated extended DNA strands on the DNA origami platform [182].

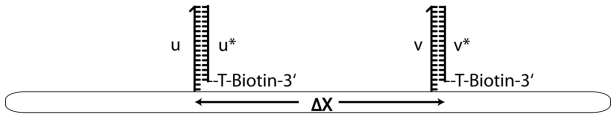


Figure 63. Schematic illustration of DNA origami constructs to which biotinylated DNA constructs were hybridized, and which subsequently after incubation with streptavidin were imaged via AFM.

It turned out that it remains challenging to resolve the positions of the gate constructs in good quality. One example is shown in Figure 64. Here a sample with one sender and four receiver gates at a distance of 21.5 nm has been imaged. Bound streptavidin proteins mimicking positions of sender and receiver gate on the DNA origami rectangle are resolved as white "spots" as high as ≈ 8 nm.

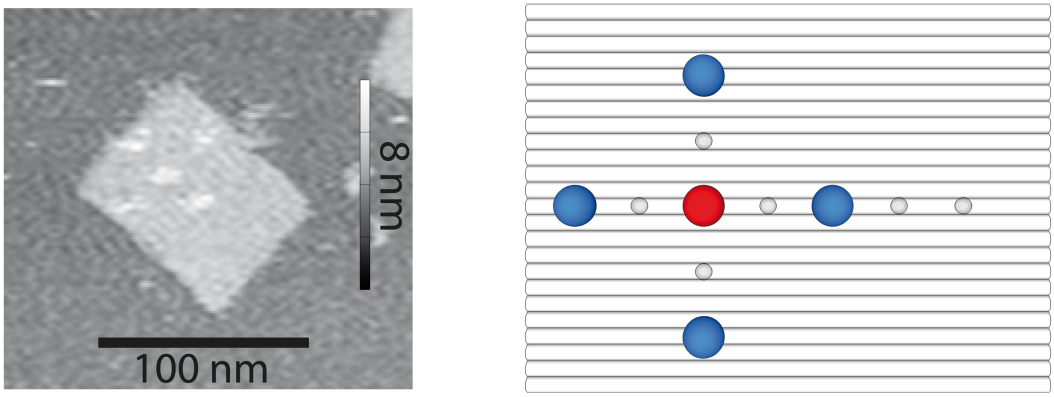


Figure 64. AFM image of a DNA origami platform modified with DNA gates carrying biotinylated strands (4 x 21.5 nm gate distance), which were visualized by labeling with streptavidin. Corresponding schematic illustration of the actual sample is included.

**Further fluorescence spectrometric experiments**

In order to ensure that the main results are interpreted correctly, several experiments were carried out helping to exclude false negative results and allowed further characterization of the cascade. Some crucial experimental data are shown in the following. For those supporting measurements, the samples were not normalized.

**Performance of the cascade with only receiver gate 2 immobilized**

First it was ensured that addition of the input strand I exclusively displaces signal strand S and does not cross-react with the receiver gate 2 by displacing output strand O. For this a DNA origami sample was prepared to which only receiver gate 2 was hybridized. Subsequent addition of input I did not result in signal change of F2 (t = 0 s) whereas signal strand S in excess (t = 1750 s) resulted in a significant increase of fluorescence (Figure 65).

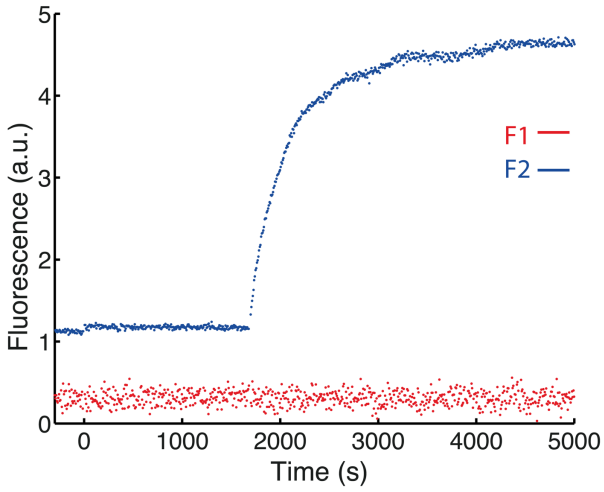


Figure 65. Performance of the cascade with only receiver gate 2 immobilized. A sample was prepared to which only F2-Q2 was hybridized. Adding the input strand (t = 0 s) does not result in a change in fluorescence. Adding the signal strand S (t = 1750 s) displaces output strand O by toehold-mediated strand displacement resulting in an increase in fluorescence of F2.

**Origami platform without extended staples for sender and receiver gates hybridization**

For verifying the purity of the immobilized DSD cascade, a DNA origami platform was prepared in which extended staples necessary for the attachment of gate complexes were replaced by standard staples lacking hybridization domains for the gate complexes. The assembly protocol (purification steps and stepwise addition of sender and receiver gates) remained the same. Fluorescence levels appeared to be close to background and addition of input strand I (t = 0 s) did not lead to a change of fluorescence. This proved that the purification protocol worked well as neither sender and receiver gates were remaining in solution nor they could bind to the DNA origami structure (Figure 66).

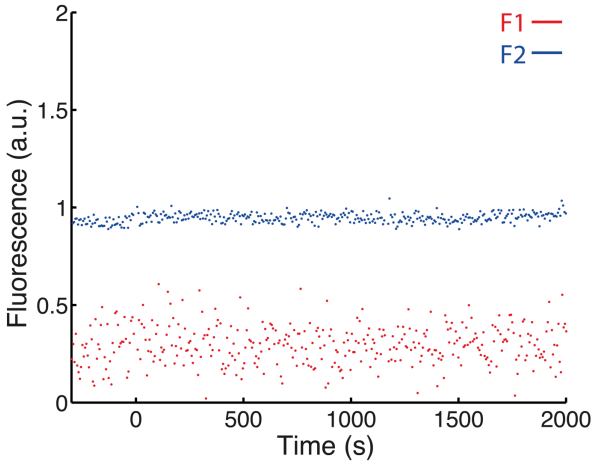


Figure 66. Origami platform without extended staples necessary for sender and receiver gate immobilization. Addition of input strand I (t = 0 s) did not result in any change of fluorescence.

### Influence of the competitor strand on the cascade

An additional experiment included the verification that the presence of the competitor does not have a detrimental effect on the performance of the cascade. Competitor strands at high concentration (≈ 1μM) were added to the solution (t = 0 s) (Figure 67). As anticipated, no effect of the competitor strand without simultaneous presence of input strand I could be detected.

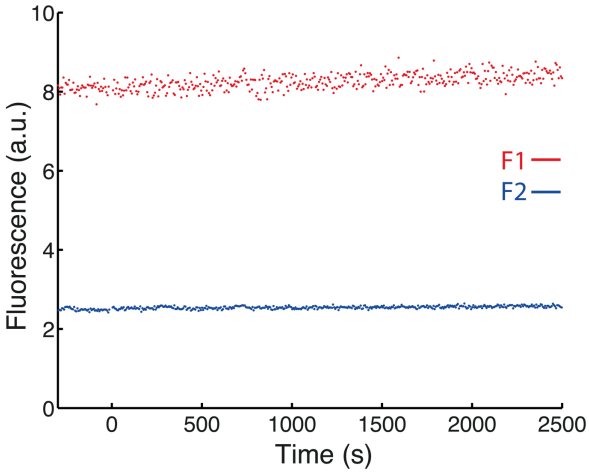


Figure 67. Influence of the competitor strand on the cascade. Addition of competitor strands (t = 0 s) at high concentration (≈ 1 μM) did not result in any undesired crosstalk.

## 3.4 Discussion

The experiments demonstrate that colocalization of components of a DNA strand displacement cascade onto a DNA origami platform has a strong effect on its performance for small gate distances and also affects the assembly process. In order to rationalize this observation, one has to consider the geometric dimensions of the various components of the cascade (Figure 68). In the initial state of gate 1, sequence domains u/u*, a/a* and d/d* are in a duplex conformation, while b* and c are single-stranded. In a relaxed conformation, a single-stranded DNA section of $n$ nucleotides is assumed to have a typical "length" of $\approx (b_k\, l)^{1/2}$, where $b_k \approx 1.6\, nm$ is the Kuhn length of single-stranded DNA, [16,183] and $l = n \times 0.6\, nm$ is its contour length [16]. Considering also the dithymidine spacers between the origami platform and the gates, in a relaxed linear conformation, the length of gate 1 is estimated to be ≈ 18.6 nm (in the protected state ≈ 20.1 nm), while gate 2 has a of length ≈ 15.6 nm. Fully stretched to their contour lengths, gate 1 and gate 2 have lengths ≈ 29.3 nm and ≈ 19.3 nm (Figure 68A). Therefore, already in a relaxed conformation, the two gates can physically interact at their "tips". At distances above ≈ 33 nm, however, the strand orientation of the signal strand S does not allow for a direct transfer to gate 2 *via*, *e.g.*, duplex fraying followed by strand displacement (Figure 68B). For the shortest distance investigated in this study – 10.5 nm – the direct physical overlap of the DSD gates (with comparatively large dimensions around ≈ 20 nm) leads to excessive strand displacement interactions already in the absence of input strands. This makes a stable preparation of a defined initial state of the cascade virtually impossible. Leak reactions presumably proceed fraying of duplex ends followed by strand invasion of complementary sequence domains of neighboring DNA gates **(**Figure 69**).** In solution such processes only occur very slowly (with a displacement rate of order 1-10 $M^{-1}s^{-1}$ [63]), but their rate is significantly enhanced due to the huge effective concentrations of the reactants on the origami platform. The mean distance between reactants with a molar concentration $c$ is $d = (c \cdot N_A)^{-1/3}$ which is ≈ 0.7 µm at c = 5 nM ($N_A$ is Avogadro's constant). Thus, the distance between sender and receiver gates on the origami platforms is more than an order of magnitude smaller, corresponding to an increase in local concentration by a factor of $2 \times 10^5$ for Δx = 10.5 nm (and still $3 \times 10^4$ for Δx = 21.5 nm). This situation is quite analogous to the speed-up of unfavorable reactions in DNA-templated synthesis [90,184].

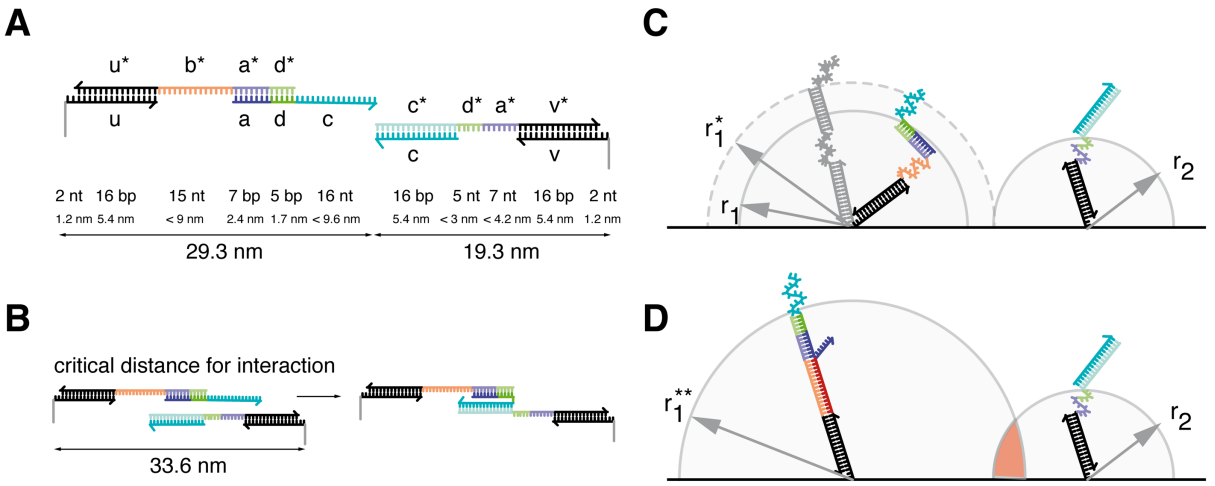


Figure 68. Geometric considerations. **A**. The two gate complexes and their respective sequence domains. Fully stretched gate 1 would extend by 29.3 nm, while gate 2 would have a length of 19.3 nm (using 0.34 nm/bp for dsDNA and 0.6 nm/nt for ssDNA). **B**. Direct interaction between the gates is only possible below a gate distance of ≈ 33.6 nm (assuming the unlikely case where both gate complexes are fully extended and lie horizontally on the origami platform) – at this distance the sequence domains d/d* of both gates can "touch each other" and therefore signal strand S can assume an orientation in which it can displace the output strand on gate 2. **C**. In a typical situation, the single-stranded sections will be relaxed, and thus gate extensions will be smaller than in the extreme cases of A and B. In a relaxed conformation, in which the dsDNA sections of gate 1 are co-linear (shown in gray), the extension up to sequence domain d would be $r_1^* = 14.5$ nm. In general, however, the two duplexes will be randomly oriented, and the root mean square extension will be only $r_1^* \approx 8.7$ nm. The extension of gate 2 (again up to domain d*) will be $r_2 \approx 7.5$ nm. **D**. Hybridization of DNA input to gate 1 stretches the gate complex ($r_1^{**} = 15.8$ nm) and thus strongly enhances the interaction between the gates for the 21.5 nm distance. For an RNA input, $r_1^{**} \approx$ 14.3 nm-14.9 nm.

While for large distances (32 nm, 42.5 nm in this study), a stable preparation of an initial state is possible, colocalization in this case does not provide any kinetic advantage, and the DSD cascade proceeds with a similar speed as in a diffusion-controlled bulk reaction. In these experiments, an overall beneficial effect of colocalization is only observed for an intermediate gate distance of ≈ 21.5 nm, for which enhanced kinetics of the cascade reaction and an increased robustness with respect to interference by a competitive reaction is found. The experiments suggest that for the 21.5 nm gate distance signal strands are transferred locally from gate 1 to gate 2 of the same platform. A variety of scenarios for this local transfer are possible. As shown in Figure 68, already in the initial state gate 1 and gate 2 – stretched to their contour lengths – can interact slightly. However, the single-stranded sections (b* on gate 1 and a*d* on gate 2) will be collapsed and an excursion to large extensions is improbable. Hybridization of input I to gate 1 renders the gate complex into a full duplex with larger (mean) extension, and thus the probability of interaction between the gates increases. At this point gate 1 and gate 2 may form an intermediate complex, in which signal S is "spooled over" without ever completely detaching from the origami platform. As indicated in Figure 69, such a transfer can happen in a variety of different ways, which cannot be distinguished in

the experiments. Alternatively, strand S could be completely displaced by the input, and then "locally diffuse" to the adjacent target gate. The efficiency of such a process depends on the ratio of the diffusion and the reaction time-scale. Assuming a diffusion coefficient for the 28 nt long signal strand S of order $D \approx 100\,\mu m^2/s$, a typical time-scale associated with diffusive processes over a distance $\Delta x$ is given by $t_D \sim \Delta x^2/D$, which is only 1 µs for $\Delta x = 100\,nm$, and 40 ns for $\Delta x = 20\,nm$. Hence, a released signal strand will quickly explore its neighborhood on the origami platform by diffusion and also quite likely bump into a receiver gate [185]. DNA hybridization, however, is not diffusion-limited and occurs at a considerably slower time-scale, [186,187] and thus most of the encounters between signal and gate will not result in the successful nucleation of a helix.

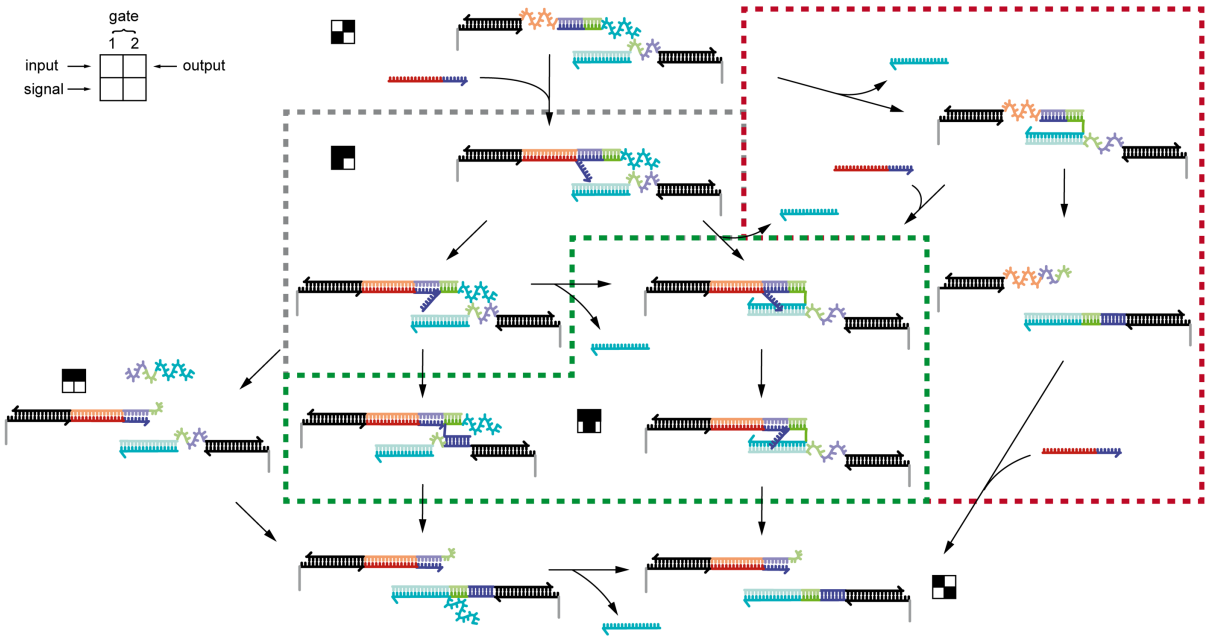


Figure 69. DSD reaction pathways. Various processes may be involved in the progress of the DSD cascade. At the smallest gate distances, a leak pathway is possible (red box), in which output is displaced from gate 2 directly by strand invasion from the neighboring gate 1. At small and intermediate distances, hybridization of input to gate 1 promotes the formation of an intermediate complex between the gates, which leads to a local transfer of signal S from gate 1 to gate 2 (green box). This may occur via toehold-mediated strand displacement, or simply via fraying and strand invasion. At large distance, only a diffusive pathway (displayed on the left) is possible. This figure includes the state symbols used for the simplified kinetic model employed in Figure 70, a legend for these symbols is shown in the upper left corner.

Also the fact that an RNA input – resulting in only a slightly shorter input-gate 1 complex than the DNA input – does not speed up the cascade reaction disagrees with a "local diffusion" scenario (which should be more or less the same for DNA and RNA inputs), and therefore supports a local strand transfer, which requires a direct physical linkage between gate 1 and gate 2. Regardless of the microscopic details of the "on-platform" transfer, one can

phenomenologically summarize the reactions involved in the DSD cascade in the simplified kinetic scheme displayed in Figure 70 – for modeling and parameter estimation please see the next section. In this scheme, binding of an input strand to an origami platform (with rate $k_{on}$) sets gate 1 into an "activated state". From this, signal strand S can either detach by spontaneous dissociation of the 5 bp d/d* duplex (rate $k_{off}$), or create an intermediate complex with gate 2 (rate $k_{loc}$). From this complex, signal strand is transferred completely to gate 2 with rate $k_{trans}$, accompanied by the release of output strand (which is not included explicitly in the model as it is not rate-limiting here). Alternatively, strands S released to the bulk re-bind to available origami platforms in a conventional hybridization reaction (again with rate $k_{on}$). As can be seen from the simulated kinetic curves in Figure 70, this model captures the experimental observations from Figure 49 very well using realistic parameter settings. Specifically, the fitted $k_{on} = 5.7 \times 10^4\, M^{-1}s^{-1}$ agrees with the rough estimate based on $t_{1/2}$ above, and $k_{off} = 5.3 \times 10^3 s^{-1}$ corresponds with an off-rate expected for a 5 bp duplex. $k_{trans} = 0.0017\ s^{-1}$ indicates that the transfer event itself is relatively slow. The fraction of locally transferred strands $\frac{k_{loc}}{k_{loc}+k_{off}} \approx 59\%$ in the model is larger than that estimated from the competition experiments (*i.e.*, $\eta_\infty$), which indicates that competitor strands might interfere also with local transfer processes.

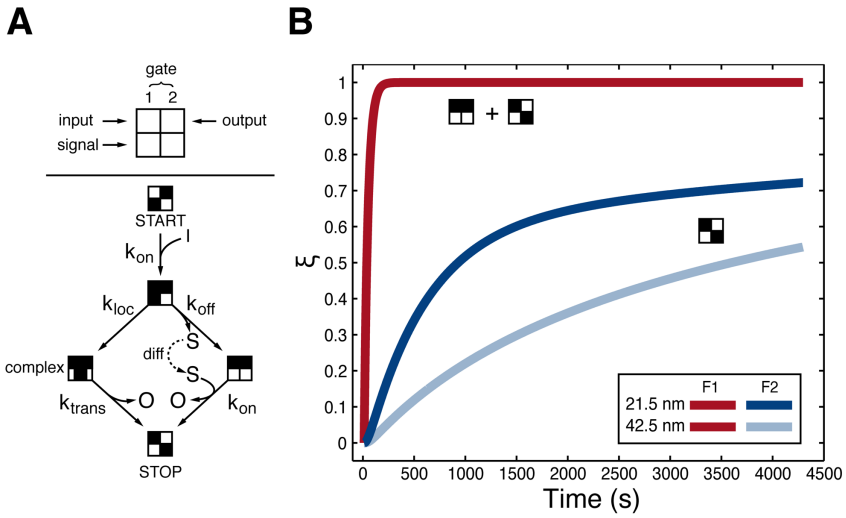


Figure 70. Simplified kinetic model of the cascade. **A**. Schematic illustration of the gate composition and the dominant pathways that occur after binding of the input strand. **B**. Simulated kinetic curves for gate distances of 21.5 nm and 42.5 nm. Rate constants are $k_{on}$ = 5.7 × $10^4$ $M^{-1}$ $s^{-1}$, $k_{off}$ = 5.3 × $10^3$ $s^{-1}$, $k_{trans}$ = 0.0017 $s^{-1}$ for all curves and $k_{loc}$ = 7.6 × $10^3$ $s^{-1}$ for the shorter and $k_{loc}$ = 0 $s^{-1}$ for the larger distance.

**Modeling and Parameter Estimation**

The kinetics of the reaction scheme shown in Figure 70 above was modeled using MATLAB with the following set of ordinary differential equations (ODEs):

$$\dot{I} = -k_{on} \cdot I \cdot O_1$$

$$\dot{O}_1 = -k_{on} \cdot I \cdot O_1$$

$$\dot{O}_1^* = +k_{on} \cdot I \cdot O_1 - k_{off} \cdot O_1^* - k_{loc} \cdot O_1^*$$

$$\dot{O}_{complex} = k_{loc} \cdot O_1^* - k_{trans} \cdot O_{complex}$$

$$\dot{O}_2 = +k_{off} \cdot O_1^* - k_{on} \cdot O_2 \cdot S$$

$$\dot{S} = +k_{off} \cdot O_1^* - k_{on} \cdot O_2 \cdot S$$

$$\dot{O}_3 = +k_{trans} \cdot O_{complex} + k_{on} \cdot O_2 \cdot S$$

All symbols are to be interpreted as reaction rates or concentrations of the represented species, respectively. I stands for input strands, $k_{on}$ is the rate at which input strands bind to gate 1 and at which freely diffusing signal stands S can bind to gate 2. $O_1$ represents DNA origami structures with signal strands bound to gate 1, before any input strand has bound. When I binds, $O_1$ changes into the "activated state" $O_1^*$. This state can change into the on-platform complex $O_{complex}$ with rate $k_{loc}$. Alternatively, $O_1^*$ can release the signal strand with rate $k_{off}$ and thus change into $O_2$ while producing freely diffusing signal strands S. From $O_{complex}$ the signal strands can transfer and hence are only bound at gate 2, producing $O_3$ at rate $k_{trans}$. $O_2$ can bind S with rate $k_{on}$ and also turn into $O_3$. Depending on the geometric arrangement, the fraction of directly transferred signal strands varies. This is implemented with the parameter $f_r$ which relates $k_{loc}$ to $k_{off}$ *via* $k_{loc} = k_{off} \cdot \frac{f_r}{1-f_r}$. The ODE set is solved with *MATLAB*'s integrated *ode23s* solver. *MATLAB*'s least-squares optimization function *lsqcurvefit* is used to optimize the four parameters $k_{on}$, $k_{off}$, $k_{trans}$ and $f_r$ simultaneously with respect to the measured cascade reaction progress of the 21.5 nm and the 42.5 nm distance.

## 3.5 Closing remarks

This project has shown that localization on an origami platform influences the performance of a DNA strand displacement cascade in various ways. For very small distances, at which the stages of the cascade are in direct physical contact, preparation of the cascade constructs is corrupted by strong intra-origami leak reactions. For distances at which the DSD gates are well separated there is no leak, but also no benefit of colocalization. The cascade reaction is sped up for intermediate distances, at which the two gates just barely interact with each other in the absence of an input strand. Hybridization of input and sender gate presumably improves the overlap between sender and receiver and facilitates the formation of an intermediate complex, in which the signal strand is spooled over from its start to its end position. Most importantly, this type of localized signal transfer makes the cascade reaction robust with respect to excessive amounts of complementary interfering strands in solution, which may be important for utilization of similar mechanisms in complex environments, or when running several alternative processes in parallel.

These findings also suggest that the performance of the cascade could be optimized by fine-tuning the distances between the gates, and also by deliberately utilizing length and conformational changes that reduce or increase the interaction between neighboring gates. Since a complete release of intermediate strands is not desirable, all strands comprising a circuit would have to be kept immobilized on the platform all the time – and in this sense an immobilized signaling cascade would not differ considerably from DNA walker or motor systems [188-194]. A different gate architecture with improved strand design and using protective secondary structures could reduce leak reactions and thus facilitate smaller gate distances. Some of the leak occurring after deprotection of the DSD cascade could potentially be avoided by "clamping" the c domain in a similar manner as demonstrated by Qian and Winfree for seesaw gates, which strongly reduced undesired strand invasion reactions [195]. Conceivably, however, introduction of more stable secondary structures would diminish the kinetic advantage conveyed by colocalization in the first place. An alternative would be to simply take leakage into account in the design of the cascades and dynamically assemble or disassemble the gates on the platforms "on the fly", *i.e.*, during operation of the circuits. The challenge is to create kinetic pathways, in which desirable processes occur at a much faster rate than undesired ones.

## 3.6 Materials

### 3.6.1 *In silico* design considerations

To study the performance of an immobilized DSD cascade it is necessary to first rationally and carefully design each of the elements. Among several freely accessible online analysis tools, the nucleic acid sequence design software (NUPACK) has been proven to be a solid program for *in silico* design and analysis of strand displacement reactions [21]. The following principles of sequence design and considerations specific to the DSD cascade were taken into account:

**Sequence dependency of input strand**

One of the main ideas of this cascade relies on its potential application as biosensing device for clinically relevant small RNAs. Among those the class of microRNA plays an important role (chapter 2.2.3). Potentially, an elevated level of – in this case – miR-21 will activate the cascade and trigger subsequent strand displacement reactions, which in the end can be monitored by a fluorescence readout. Apparently, the sequence of the miR-21 is fixed and thus predefines the consecutive design of the cascade. The RNA sequence of 22 nt was divided into segments - so-called domains - which defined subsequent design steps. Domains were designated with small alphabetical letters. The corresponding complementary sequence was annotated with a *star* (*).

**Restrictions when using dye-modified sequences**

For monitoring the reaction progress, 3' and 5' ends of designed oligonucleotides were modified with fluorescent dye molecules. It has been shown that the nucleotide adjacent to a dye molecule influences the quantum yield of the fluorophore. This has been observed in particular for guanosine at the 3' end. Electron transfer occurs from 2' aminopurine to guanosine in proximity to purines, [196]. Due to this fact, the sequence pattern of Atto dye-modified oligonucleotides was optimized and guanine bases explicitly excluded as terminal bases.

**Sequence design and orthogonality of designated reaction partners**

Apart from the constrained sequence of the input strand I and the considerations concerning fluorophores, it is crucial to rationally design the reaction components of the cascade. It was ensured that sequences did not contain undesired secondary structures or reveal crosstalk and self-complementarity. It was also ensured that sequences only reacted with their designated complementary strands and not with any other oligonucleotides that exhibited a similar sequence pattern.

All these considerations were made when designing the cascade *in silico* by using the NUPACK sequence design tool which was also applied for subsequent sequence analysis. Moreover, this software was used to simulate reactions in order to minimize unwanted side reactions and to optimize reaction conditions at a chosen temperature of 20°C and a salt concentration of 1M NaCl.

## 3.6.2 Sample preparation

Synthetic DNA and RNA sequences built up the basis for this project and are specified as follows. For folding the DNA origami structures, a 7249 nt long single-stranded DNA sequence, serving as scaffold strand, originating from the genome of the lambda bacteriophage M13 (M13mp18) was used at a stock concentration of 100 nM, dissolved in $_{dd}H_20$. The scaffold DNA was produced by F. Praetorius at TUM, stored at -20°C and thawed prior to use. Staple strands, serving as connecting strands in order to form the desired shape of the DNA origami structure were purchased from MWG Eurofins Genomics, Ebersberg, Germany at a stock concentration of 100 µM, dissolved in $_{dd}H_20$ and stored at -20°C. Extended staples necessary for hybridization of the gate complexes to the DNA origami structures as well as dye-modified (quencher and fluorophore) oligonucleotides were ordered at Biomers Ulm at a concentration of 100 µM and stored at -20°C. These vendor specifications also apply to DNA and RNA sequences which were used as input signal for the cascade.

**Preparation of the DNA origami structures**

For folding the DNA origami structures, scaffold strand and staple strands were mixed as indicated in Table 3 together with folding buffer, containing 12.5 mM $MgCl_2$ and 1 x TAE (40 mM Tris, 20 mM acetic acid, 1mM EDTA, pH 8.0) which was diluted from a 10 x stock solution (125 mM $MgCl_2$, 10 x TAE). This buffer was used throughout the whole preparation protocol.

| | Concentration [nM] | Molar excess over scaffold |
|---|---|---|
| Scaffold (M13mp18) | 48 | - |
| Staple strands | 200 | 4x |
| Extended staple strands | 400 | 8x |

Table 3. Concentration of scaffold and staple strands (including its molar excess) used for folding the DNA origami structures.

The folding reaction was performed by one pot thermal annealing in a thermal cycler with a denaturing step to 70°C for 5 min and subsequent cooling steps of 0.5 °C/min until reaching 40°C and a final temperature of 20°C. Structures were purified from remaining excess staples by a two-step PEG precipitation procedure [197] and subsequently resolved in folding buffer. Concentration of purified DNA origami structures was measured by using a nanophotometer.

**Preparation of the DNA strand displacement gates**

Gate complexes of which the cascade is composed of were prepared separately prior to hybridization to the DNA origami structures. For this, original stock concentrations of sequences composing gate 1 and gate 2 complexes were diluted 1:5 in folding buffer, resulting in an approximate concentration of 20 µM. Exact concentration of these sequences was determined by a spectrophotometer. Gate 1 and gate 2 complexes were mixed in separate tubes and prepared according following protocol. At the stage of assembly, gate 1 is composed of three oligonucleotides comprising a fluorophore labeled, a quencher labeled and an unlabeled protective strand. The fluorophore labeled strand, F1 (modified with Atto647N at the 5' end) was designed to be partially complementary to a quencher labeled DNA strand, given the name signal strand S (labeled with BBQ-650 at the 5' end). In order to minimize unwanted strand displacement reactions between the two gates during the assembly process, a

protective strand P hybridizing to a single-stranded toehold domain of S was added to the mixture. The ratio for this construct consisting of F1, S, P was adjusted to be 3 µM, 4.5 µM and 7,5 µM, respectively. Gate 2 is composed of a fluorophore labeled DNA strand, F2 (Atto 532 at the 5' end) to which a quencher modified DNA strand, output O (BHQ-1 at the 3' end) is hybridized. Here concentrations are adjusted to be 3 µM for F2 and 6 µM for Q2. Once prepared, gate 1 and gate 2 mixtures were incubated in folding buffer for at least 1 h at room temperature.

**Attachment of the gate complexes on to the DNA origami platform**

Attachment involves several purification steps of which the order has to be kept. First, the complex F1-S-P was added to the solution containing DNA origami platforms in a 2x molar excess in respect to the DNA origami platforms, and incubated for at least 1 h at room temperature. To remove unbound gate 1 complexes, the samples were purified by PEG precipitation twice and resuspended in folding buffer. F2-Q2 constructs were added to the solution in 1.5x molar excess over the DNA origami structures which subsequently were incubated for 20 min sharp at room temperature, followed by another two-time PEG precipitation step. Finally, a deprotection strand, fully complementary to strand P was added in 3x excess, resulting in "activation" of the cascade. Samples were incubated for 1 min at room temperature. For further measurements, the samples were diluted in folding buffer to a final concentration of 5 nM, rapidly frozen in liquid nitrogen and stored at -80°C.

**Preparation of input strands I and serial dilutions of competition strands**

Input strands, synthetic analogs of the miRNA-21 (DNA and RNA) for triggering the cascade were diluted in folding buffer to a concentration of 10.7 µM. For experiments investigating the robustness of the cascade input strands were mixed with competitor strands in a specific manner. 50 µl of competitor strands (stock concentration 100 µM) were mixed with 50 µl of folding buffer. For obtaining a serial dilution this step was repeated for 12 times. To each dilution, 6 µl of input strand I was added, thus keeping the concentration for I same as without competitor strands. All input solutions were stored at 4°C.

## 3.7 Methods

### 3.7.1 Fluorescence spectroscopy

One widely applied analytical and very sensitive method to study (bio)-chemical processes is fluorescence spectroscopy [198]. As the name already indicates, the technique is used to study molecules possessing fluorescence properties. Fluorescence represents one specific case of luminescence. Organic fluorescent molecules are typically based on an aromatic hydrocarbon structure which can be excited by an incoming photon from a light source like a laser or a lamp with a specific wavelength and energy. This causes the molecule to undergo an electronic transition from a low energetic ground state S0 into an excited state of higher energy known as S1. The process of absorption takes place within a very short time frame of $10^{-15}$ s, whereas the lifetime of the excited state is within the regime of $10^{-10}$-$10^{-7}$ s. In the excited state, the electron first relaxes vibrationally to the lowest energy level in the S1 state. Fluorescence describes the emission of a photon by relaxation from the excited S1 state to the S0 state. The wavelengths of a fluorescence emission spectrum are higher than those of its absorption due to the loss of energy. This gap between the maximum of the first absorption and the maximum of fluorescence is known as the *Stokes shift*. Other important processes occurring after excitation of a molecule are internal conversion, phosphorescence or intersystem crossing.

All these processes are illustrated in a simplified Jablonski diagram (Figure 71). The emitted energy in terms of fluorescence can be detected by using spectroscopic techniques that excite molecules of choice at a specific wavelength and detect their emission spectra, enabling a thorough characterization of biomolecules in bulk.

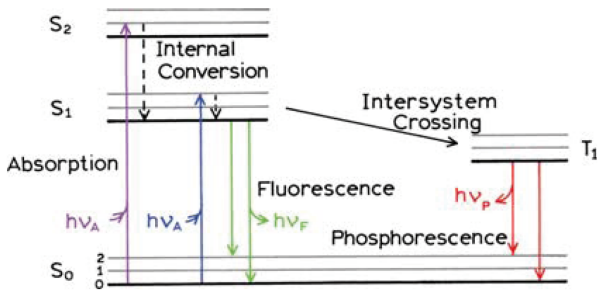


Figure 71. Jablonski diagram depicting different energy levels of a molecule or atom after excitation by photons of certain wavelength and energy. Taken from [199].

For the DSD cascade, characterization and data analysis predominantly relied on fluorescence spectroscopy that was carried out using a Cary Eclipse spectrometer (Agilent Technologies Deutschland GmbH, Böblingen, Germany). For each measurement, the excitation wavelength for fluorophore 1 (F1 – Atto532) which corresponds to gate 1, was set to 532 nm with a band pass filter of +/- 10 nm. Emission was recorded at 554 nm +/- 10 nm. For gate 2, whose output was monitored with an Atto647N dye linked to the "F2" strand, excitation was at 647 nm +/- 10 nm and emission at 669 nm +/- 10 nm. Excitation and emission spectra for F1 (Atto532) and F2 (Atto647N) are shown in Figure 72.

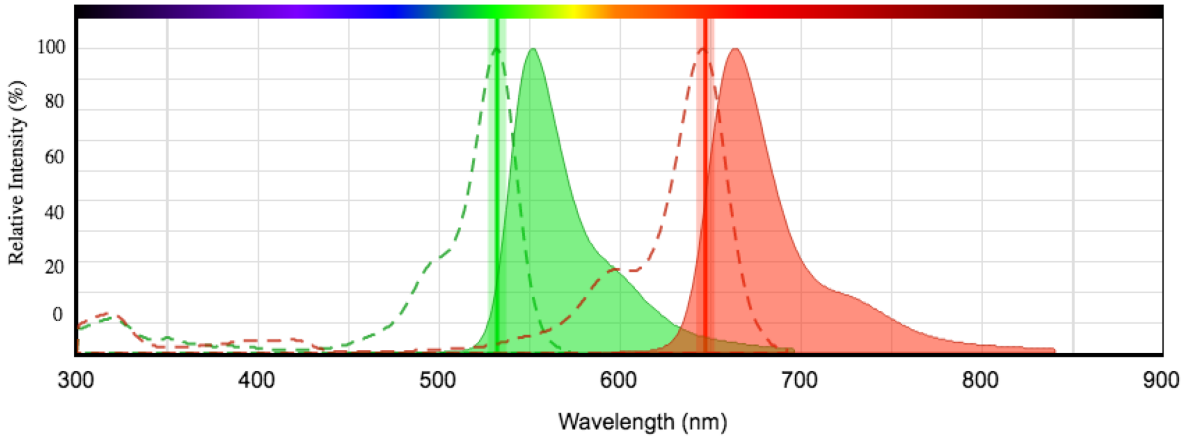


Figure 72. Excitation and emission spectrum of Atto532 and Atto647N dyes, created using the online spectra viewer by AAT Bioquest.

The sample volume for all kinetic measurements was 65 µl to which 3 µl of input strands (total volume 68 µl) were added into a fluorescence cuvette (105.254-QS, Hellma GmbH & Co. KG, Müllheim, Germany). Before and after each measurement, the cuvette was thoroughly washed with Hellmanex and Isopropanol.

### 3.7.2 Atomic force microscopy (AFM)

For proper confirmation of structure assembly, DNA origami structures and gate complexes were analyzed by atomic force microscopy (AFM). AFM represents an opto-mechanical imaging technology which has proven to be a versatile tool to study surface interactions in material sciences, both in air and liquid conditions [200]. Moreover, also biological materials can be investigated via AFM measurements even though nucleic acids based constructs display a challenging material as these soft materials can easily be destructed during the actual measurement. The basic principle of AFM is described in the following [201]. A cantilever to

which a tip is attached at the bottom part, scans an area of a sample surface row by row (Figure 73A). Usually, the sample is moved along the XY axes below the tip, which is achieved by a nanometer-precise scanning mechanism, controlled by a piezo crystal. The cantilever itself is fixed at a chip (Figure 73B). In AFM tapping (AC) mode, the tip oscillates and taps the surface. Changes in sample topography and material properties influence the deflection of the cantilever, which is recorded by a detector. This information is processed and used by an integrated feedback mechanism to keep a constant distance between tip and sample by moving the sample holder in z direction. The topography of the actual sample is then reconstructed from the movement in z direction. The deflection of the cantilever depends on the attraction or repulsion between the sample material and the tip. The forces between atoms or molecules at close distance are well described by the Lennard-Jones potential, which includes attractive Van-der-Waals forces and short-range Pauli repulsion. The movement and deflection of the cantilever, which results from the interaction of the tip with the surface, is measured via a laser beam that is reflected at the top side of the cantilever and monitored by a photo diode. The obtained information in turn is processed and finally translated into an image of the sample topography.

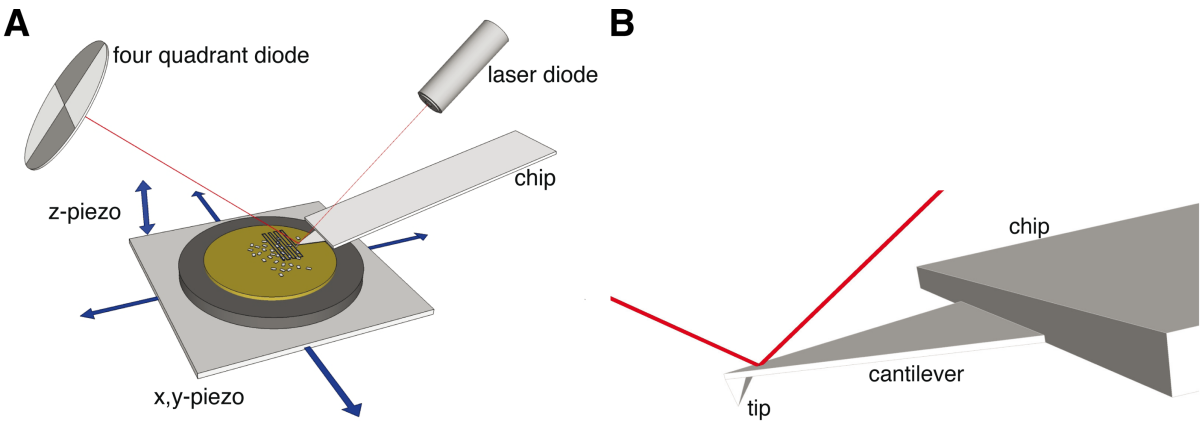


Figure 73. **A**. Schematic illustration of the measurement principle of an AFM. **B**. Reflection of the laser beam that has been positioned on top of the cantilever. Taken from [201].

For the experiments presented, an Asylum Research Cypher scanning probe microscope was used. Initially the DNA origami structures were imaged including the immobilized gate complexes without any further adaptions. Due to the flexibility of the partially single strand gate complexes, those structures could not be imaged as intended, thus requiring the application of biotinylated DNA strands to better visualize the strands in AFM via streptavidin. For biotinylation, the DNA origami structures were folded as described before. Subsequently, samples were incubated with biotinylated oligonucleotides for 1 h at room

temperature, followed by two PEG precipitation steps and final dilution in folding buffer to a concentration of 2.5 nM. For imaging, 5 µl of origami solution were added to a freshly cleaved mica surface to which 60 µl of folding buffer was added. Prior to imaging, 2 µl of 500 nM streptavidin solution was added in order to initiate binding of streptavidin to the biotin modified oligonucleotides. The measurement was performed in liquid AC mode with an Olympus microcantiliver BL-AC40TS-C2.

### 3.7.3 Gel electrophoresis

Agarose gel electrophoresis represents a common tool in molecular biology for the analysis of nucleic acids, restriction digestion or purification steps. As the phosphate residues of the DNA backbone are negatively charged at neutral pH, nucleic acid fragments can be separated by applying an electric field resulting in its migration towards the anode. Originally, agarose is found in red algae. This material is composed of polysaccharides and forms, after heating in TAE or TE buffer, a matrix with pores that enable the separation of DNA strands. When electric current is applied, smaller DNA strands move more rapidly through the gel than larger DNA strands and migrate further through the gel towards the anode. Consequently, DNA fragments can be separated accordant to their size. DNA strands of the same length form distinct bands on the gel, which are visualized by staining with intercalating dyes.
Here, this technique was applied to confirm correct folding of the DNA origami structures and successful hybridization of the gates. 1% agarose was dissolved in 100 ml folding buffer. After hardening of the gel, 24 µl of 5 nM, PEG purified DNA origami solution was mixed with 5 µl of 40% sucrose and subsequently loaded into the gel chambers. To avoid disassembly of gate complexes, the gel was run in an ice-cooled water bath for 1.5 h. For analyzing the gate change of fluorescence intensity before and after binding of input I, the gel was imaged with a laser scanner (Typhoon FLA 9500) with excitation at a wavelength of 532 nm. Afterwards, the gel was stained with SybrGold, an intercalating dye binding with high affinity to DNA and RNA, and imaged again, demonstrating the correct folding of the DNA origami structures.

# 4 Toehold switches

Several projects involving toehold switches in bacterial cells as well as in mammalian cells have been performed during a research visit at Wyss institute For Biologically Inspired Engineering. The contents of chapter 4.1 are extracted from a joint paper draft with following preliminary title: "*Toehold Repressors: De-Novo-Designed Programmable Controller of Gene Expression*". Data and results which I have partially contributed to, or which are relevant to the reader in order to get the big picture are presented in the following.
Data shown in chapter 4.2 are preliminary results on the investigation of artificial riboregulators in mammalian cells. This project represents a sub aim of the Unified Nucleic Acid Computation System (UNACS) project. It was intended to further study post-transcriptional control of gene expression by implementing previously described toehold switches into eukaryotic systems. Excerpts of preliminary experimental work was carried out in a team with major contributions by myself.

## 4.1 Bacterial toehold switches as repressors of translation

### 4.1.1 Motivation

Gene expression is a complex process describing the flow of genetic information, also known under the term "central dogma of biology", of which the details in its entirety have been barely understood (chapter 2.2.1). A discipline desiring to control and manipulate each step of gene expression by redefining the classic molecular biology workflow, represents synthetic biology. Synthetic biology as it has been elaborated in chapter 2.3 aims to take advantage of current understanding of molecular biology in order to create programmable biological systems that are capable to carry out specific (*de-novo* designed) tasks or which possess novel and complex functionalities, following an engineering-driven approach. Since its beginnings in the early 2000's modern synthetic biology has created a large repertoire of regulatory systems with increasing focus on regulation of gene expression. Even though these early stage genetic networks were mostly composed of widely used regulatory elements, their compatibility and orthogonality as well as their scalability was limited. The recurring lack of programmability and predictability, hindering engineering of complex genetic systems has been addressed by employing RNA molecules for the construction of orthogonal regulatory elements. Due to its predictable Watson-Crick base pairing rules and the well-described

thermodynamic parameters, RNA can form the basis of *de-novo*-designed synthetic regulators of gene expression.

Pioneering works used sense-antisense RNA interactions around the translation initiation site of mRNA to either activate or repress translation [158,161]. However, these early stage mechanisms still revealed limitations for the creation of large libraries due to sequence constraints and employed reaction mechanisms. Recently, these issues have been approached and solved by the implementation of toehold switches [3]. These *de-novo* designed regulators, relying on arbitrary sequences, fulfilled the requirements of being highly programmable, and therefore were instrumental for creation of more complex circuits as well as for a potential *in vitro*-based diagnostic tool [169,171].

Inspired by previous works, a set of *de-novo* designed translational repressors will be presented in the following, which builds on the design strategy of the original toehold switches. Similar to the activating toehold switches, a predictive data-driven *in silico* modeling from quantitative assays in *E. coli* was employed. Key determinants of functionality were identified by model integration of *in vivo* reporter assays. Interacting sequences were optimized and designed to be entirely synthetic, and by this circumvented sequence constraints of naturally occurring structural motifs. Moreover, regulation of translation solely relied on toehold-mediated strand displacement reactions of two rationally designed RNA species, based on a linear-linear nucleic acid based mechanism that is originating from dynamic DNA nanotechnology [58]. Those design principles were applied to create a set of repressors which efficiently represses translation in *E. coli*, and which provided the basis for the creation of a large library of orthogonal constructs. Moreover, toehold repressors successfully proved its potential as functional sensors for native mRNA. Overall, a novel family of forward-engineered riboregulators displaying a high degree of orthogonality was constructed, based on current knowledge of thermodynamics and nucleic-acid interactions. These findings could contribute to further extend the library of highly programmable and standardized parts to control post-transcriptional gene expression *in vivo*, and help extend such designed biological elements for real-world synthetic biology applications, possibly in diagnostics or even therapy. In the following, design and results of this project are shown as excerpts from a paper draft with an emphasis on my contributions to this project.

### 4.1.2 Design and reaction mechanism

Figure 74 demonstrates the design and reaction mechanism of the original toehold switch on which the design of the toehold repressor is based on. The toehold switch RNA consists of a hairpin loop structure and is inserted upstream of gene of interest. In its initial configuration, translation is inhibited as the RBS is sequestered within a single-stranded loop. Additionally, the AUG start codon is located unpaired within the hairpin stem which itself forms a RNA duplex by hybridization of domains b and b*. At the 5' end of the switch RNA a single-stranded toehold domain a, serves as initial binding site of the trigger RNA. The trigger RNA includes the toehold binding domain a* and an extended single-stranded domain, b*. Upon hybridization of a cognate trigger RNA to the toehold switch via domains a/a* and b/b*, toehold-mediated branch migration reaction results in unwinding of the hairpin. Subsequent exposure of RBS and start codon allows ribosome binding and initiation of translation of downstream gene of interest.

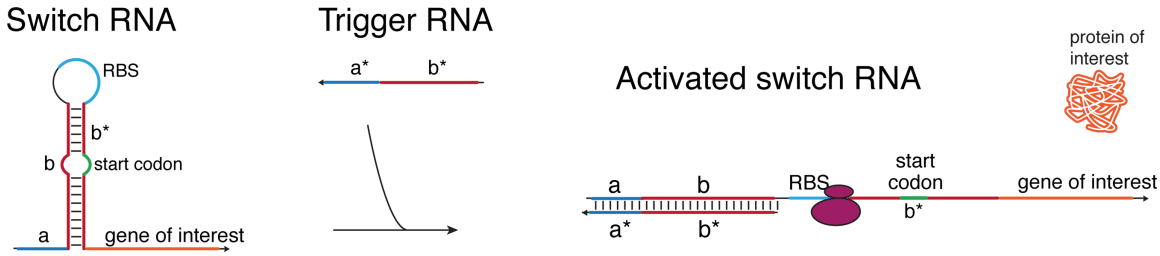


Figure 74. Design and reaction schematics of the original toehold switch. Translation is repressed by formation of a stem loop structure which is characterized by RNA base pairing before and after the AUG start codon, and sequestration of the RBS within the single-stranded loop. Addition of a cognate trigger RNA results in hybridization of the toehold domain a of switch RNA with its complementary domain a* of the trigger RNA, followed by subsequent initiation of toehold-mediated strand displacement and branch migration mechanisms. This makes the RBS and start codon accessible for ribosome binding, enabling translation of downstream gene of interest.

#### Design of the toehold repressor switch

The toehold repressor was designed based on understanding of RNA thermodynamics and optimized analogous to the previously described activating toehold switch library, see Figure 75A. Here it was ensured that the switch-trigger pair forms a stem loop structure which is similar to the toehold switch. In its initial state, this secondary structure is located upstream of the RBS and the start codon. Consequently, in the absence of the trigger RNA, this configuration allows downstream ribosome binding, followed by translation of a gene of interest. Figure 75B shows MFE structures of a toehold repressor switch in its untriggered (left), and triggered state (right) at 37°C, simulated by NUPACK [202]. Interaction domains are

highlighted. Folding free energies for switch RNA and repressed switch RNA are calculated with -42.8 kcal/mol and -103. 88 kcal/mol, respectively.

The detailed toehold repressor design includes a stable hairpin structure based on the hybridization of the 18 nt repressing domain b bound to its complementary sequence b* in the absence of the trigger RNA. Domain b is available to bind around the start codon to maintain the typical toehold switch conformation upon trigger hybridization. In addition to domain b, the stem is extended by a 12 nt long domain c, base-paired with c* hence further shifting the thermodynamic equilibrium towards the active state. This modification results in a slightly longer duplex stem, compared to the activating toehold switch. At its 5' end, upstream of the duplex hairpin, a 15 nt single-stranded toehold domain a* serves as initial interaction site with the cognate trigger RNA. Respectively, a 45 nt long trigger RNA is composed of domains a, b, c which are fully complementary to the duplex stem domains b* and c* as well as to toehold a* of the switch RNA. Upon binding of the trigger RNA to the switch RNA toehold-mediated branch migration reactions are initiated resulting in the formation of a second hairpin structure which is shifted further downstream. In the trigger-bound repressed state, the RBS sequence and the start codon are sequestered in a 12 nt loop and the stem region, respectively. In this configuration ribosome binding is inhibited, leading to repression of translation. Notably no sequences complementary to RBS and start codon are required here which allows an arbitrary choice of potential trigger sequences.

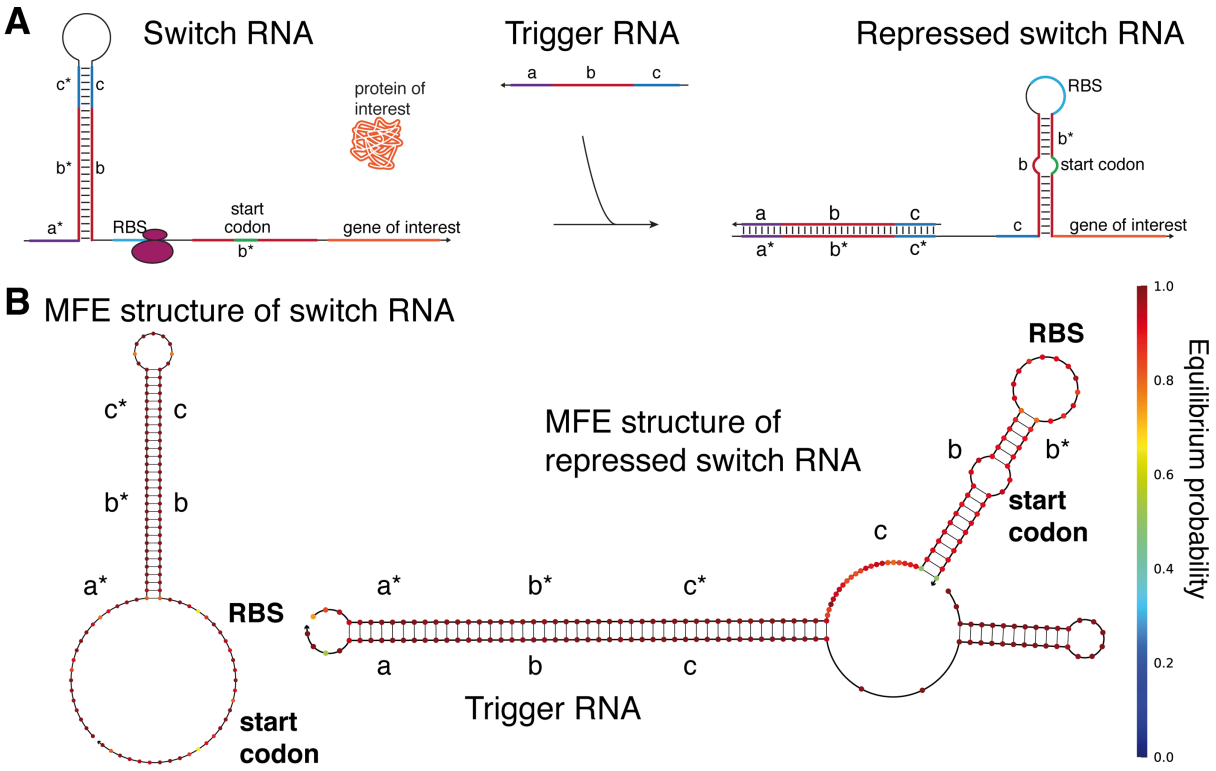


Figure 75. Design and reaction schematics of the toehold repressor. **A**. Schematic illustration. Initially the RBS is accessible for ribosome binding allowing translation of gene of interest with a stem loop located upstream of RBS and the AUG start codon. Addition of a cognate trigger RNA results in toehold-mediated strand displacement reactions beginning at the 5' end of the switch with its domains a/a* leading to a conformational change of the stem loop. This is followed by formation of a secondary structure downstream the initial hairpin stem, now encompassing the RBS in the single-stranded loop and the start codon within the stem loop. Translation of gene of interest is hindered. **B**. MFE structures of switch RNA (left) and repressed switch RNA (right) upon binding of the cognate trigger RNA using the NUPACK [21].

Following specification of the overall secondary structure and underlying thermodynamic parameters, a set of toehold repressors was generated using the *in silico* nucleic acid sequence design tool NUPACK [21]. Subsequently, the performance of toehold repressor constructs was tested under *in vivo* conditions by predominantly transforming plasmids into *E. coli* BL21 Star DE3. This strain possesses an inducible genomic T7 RNAP and is deficient in RNase E. Repressor switches were cloned into a medium copy plasmid, whereas a high copy plasmid served as backbone for the trigger RNA, in order to maximize expression of small RNA species. Upon induction with IPTG, switch and trigger RNA were produced by T7 RNAP. Output GFP expression was monitored by flow cytometry and the performance metric was calculated as fold reduction of GFP output.

### 4.1.3 Results

#### 4.1.3.1 Repression of translation of toehold repressors

In initial tests, 45 toehold repressor constructs have been tested out and characterized *in vivo*. The performance of each is illustrated in Figure 76A. The mode fluorescence value was taken from histograms generated by flow cytometry experiments and served for calculation of the ON/OFF ratios of each toehold switch design. For exact determination, cellular autofluorescence was not subtracted from either the ON or OFF state fluorescence. Error bars represent standard deviation resulting from three biological replicates. Among the first set of toehold repressors two constructs had an ON/OFF ratio greater than 100, five greater than 50 and 21 exceeded ON/OFF levels of 10. It could be shown that thermodynamic parameters correlated well with switch performance.

To further improve switch performance which makes such devices more suitable for real-world applications, key parameters responsible for high ON/OFF ratios were identified. Candidate thermodynamic parameters were analyzed via linear regression coefficients against the performance metric of first generation repressors. Regressions using at least three different parameters were required for obtaining $R^2$ values above ~0.3. By using up to three different parameters, over 1 million different linear regressions were screened. The three-parameter linear fit revealing the strongest correlation ($R^2 = 0.39$) with the experimental data is shown in Figure 76B. Key parameters determining high GFP fluorescence were subsequently identified, but are not further discussed in the following (as I have not been involved in these optimization process and as consequence details remain undisclosed at the time of thesis submission). However, for further reading, it is crucial to include these steps into the thesis.

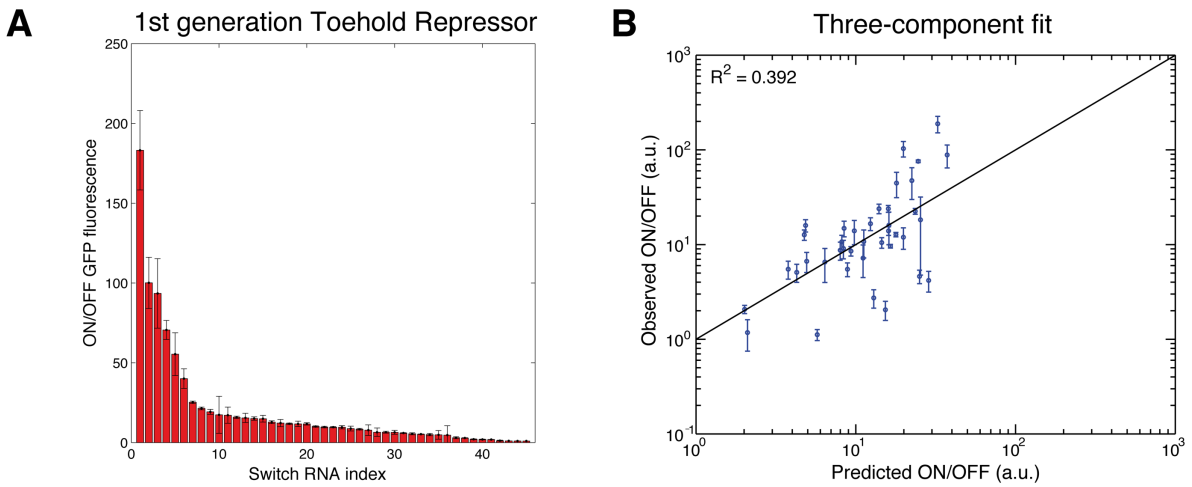


Figure 76. *In vivo* characterization and thermodynamic analysis. **A**. ON/OFF GFP fluorescence levels obtained 3 h after induction for 45 first-generation toehold repressors. Relative errors for ON and OFF states are the result from three biological replicates. **B**. Correlation between experimentally observed performance of repressors and predicted performance of repressors using a three-parameter regression model.

Findings of the optimization process were integrated into the selection process of the second-generation toehold repressors. A set of 96 forward-engineered toehold repressors was characterized by flow cytometry. ON/OFF GFP fluorescence levels for the second-generation repressors 3 hours after induction with IPTG are depicted in Figure 77A. Out of these designs, eight exceeded a GFP fluorescence output greater than 100, 16 greater than 50 and 82 greater than 10 thus proving a significant increase in overall behavior. Compared to the first-generation set, the average ON/OFF ratio of second-generation library increased from 20 to 40, respectively. This improvement gave evidence that no *in vitro* evolution is necessary as well as no large-scale screening experiments for realization of a highly programmable *in vivo* regulatory system in bacteria. Furthermore, the effectiveness of the selection criteria was quantified by calculation of the percentage of forward-engineered toehold repressors with ON/OFF ratios exceeding a given minimal level and subsequent comparison to the library of first-generation toehold repressors. As expected, the overall yield for high-performance repressors is higher for the second-generation library (Figure 77B). In this chart, the gray dashed horizontal line shows the percentage of the ON/OFF ratio of repressor switches exceeding a minimal level of output fluorescence which was set to be at least 10-fold or equaling 90% repression. Out of these, three particularly well-performing switches are presented by displaying its mode GFP fluorescence in the untriggered (ON state) as well as in the triggered (OFF state) (Figure 77C). These top repressor constructs underscore the high repression and scalability of such systems. Standard deviations are indicated as error bars and are the result of a set of three biological replicates.

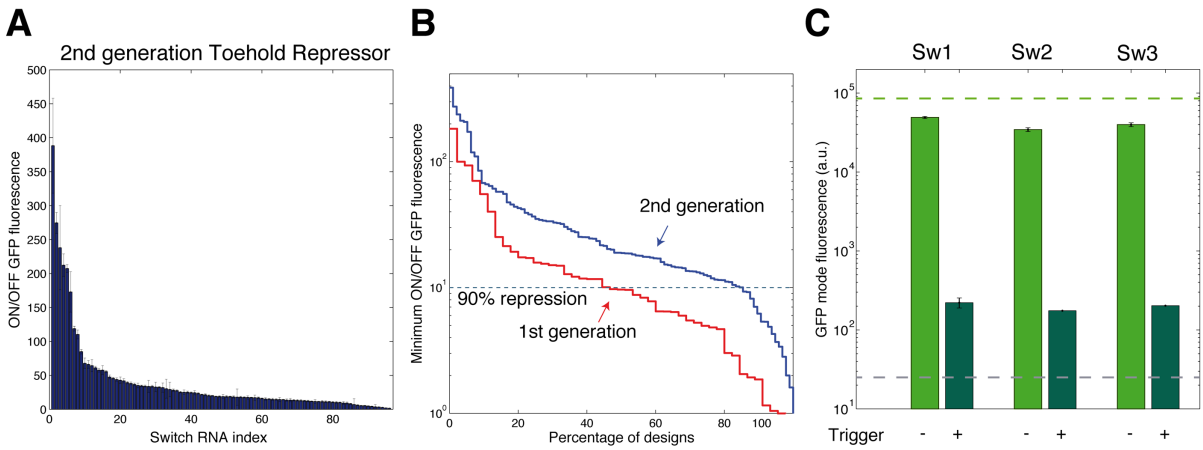


Figure 77. Toehold repressor performance in *E. coli*. **A**. ON/OFF fluorescence levels of forward-engineered second-generation library of a set of 96 constructs 3h after induction which have been selected via *in silico* optimization. **B**. Comparison of first and second generation toehold repressors with ON/OFF ratios exceeding 10-fold expression, indicated by the gray dashed line indicating a 90% repression. **C**. Display of GFP mode fluorescence of three high performance toehold repressors in the untriggered (dark green) and triggered (light green) state. Standard deviations are indicated as error bars and are the result of three biological replicates.

### 4.1.3.2 Toehold repressors as mRNA sensors

In 2014, Green *et al.* refined toehold switches as sensors for detection of native mRNA [3]. Based on these insights, the capability of toehold repressors to sense native mRNA sequences was exploited (Figure 78A). In contrast to rationally designed trigger sequences in the original design, mRNA trigger sequences are predetermined. In theory, toehold repressors can be designed in such a way that parts of native mRNAs, upon binding to the switch result in strong repression. However, native mRNAs are characterized by strong secondary structures which hinder the binding of the trigger and subsequent branch migration. Consequently, kinetics and thermodynamics of toehold repressors and mRNA interactions had to be carefully incorporated in the design procedure, and several modifications to the original design had to be done. To address these issues, the following main steps were considered for the mRNA-sensing toehold repressors:

1. To facilitate switch-trigger binding and initiation of the subsequent branch migration process, the length of the toehold domain of the repressor was increased from 15 to 27 nt.
2. The branch migration domain length was increased from 30 to 32 nt. This results in thermodynamic energy gain of toehold-binding interaction between switch and trigger and consequently favors cognate switch-trigger interactions.

3. *In silico* screening served for identification of native mRNA sequence domains which contain minimum 60 nt with minimum secondary structures to increase the probability of proper binding to the toehold domain.

Results of these adjustments are presented in Figure 78B. Two different mRNA sequences turned out to give reasonable results: The antibiotic β-lactamase (*bla*) mRNA which confers ampicillin resistance and the kanamycin resistance protein (*kanR*) mRNA which confers resistance to antibiotic kanamycin. Translation of GFP output is repressed upon binding of the *bla* sensors to cognate *bla* trigger mRNA transcripts (shown in green bars), whereas translation of mCherry output is repressed upon binding of the *kanR* sensors to cognate *kanR* trigger mRNA (shown in violet bars). Among the tested *bla* and *kanR* sensors two specific trigger constructs resulted in a fold-change repression of at least 10 which can be considered as a significant change, Figure 78C highlights the dynamic range of eight different kanR trigger mRNAs sequences capable to bind to cognate mRNA sensors and induce repression of translation. Accordant mCherry fluorescence ON/OFF ratios are shown for cells induced with IPTG after three and five hours in blue and green, respectively. In most cases, overall performance increases with induction time. Figure 78D represents a flow cytometry histogram generated by FlowJo software. Here median mCherry expression is shown for cells having been transformed with kanR mRNA sensors. Cognate kanR trigger mRNA results in strong (10-fold) repression (red), compared to non-cognate trigger mRNA (green). Other mRNA constructs as potential triggers were tested, including transcripts encoding *gfp* as well as the gene conferring resistance to the antibiotic spectinomycin. Repression was not significant, unfortunately.

The results suggest that native mRNA sequences, once they have been thoroughly screened *in silico* can serve as trigger to switch the toehold repressor constructs and thus provide a reasonable fluorescence output signal. For real-world applications, it might be desirable to replace these mRNA triggers by sequences relevant in diagnostics or therapy.

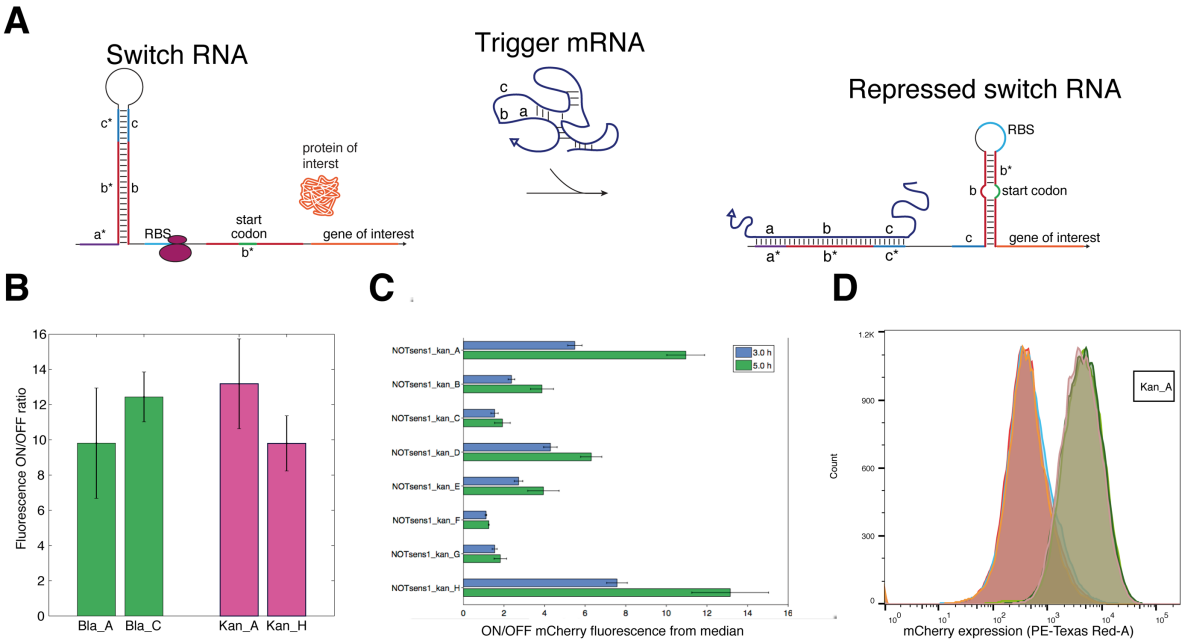


Figure 78. Toehold repressors as mRNA sensors. **A**. Design and reaction pathway of toehold repressors used for mRNA sensing. Trigger binding domains a, b, and c within the trigger mRNA are indicated. **B**. Repression fold-change ON/OFF ratio of fluorescence output for toehold repressor sets used for mRNA sensing. Fold change repression is shown for detection of ampicillin resistance (shown in green bars) resulting in GFP output and kanamycin resistance (shown in red bars) resulting in mCherry fluorescence. Relative errors for ON/OFF ratios are obtained by adding the relative errors of the sensor ON and OFF state fluorescence measurements in quadrature. Relative errors for ON/OFF states are the result from standard deviations of three biological replicates. **C**. ON/OFF ratios for mRNA sensors detecting eight different kanR mRNA transcripts. Median ON/OFF mCherry fluorescence is shown three (blue) and five hours (green) after induction with IPTG. Relative errors for ON/OFF states are the result from standard deviations of three biological replicates. **D**. Histogram of flow-cytometric measurements of cells transformed with KanR sensors and kanR trigger mRNA showing median mCherry expression in its repressed state (red) and un-repressed state (green).

### 4.1.4 Discussion and conclusion

Results presented above show the successful demonstration of a *de-novo* designed riboregulator family capable of repressing translation at the post-transcriptional level. Analogous to the activating toehold switches, an arbitrary, optimized sequence can serve as trigger RNA which eliminates sequence constraints of previous works that have relied on naturally occurring RNA motifs or base pairing of the RBS and start codon. However, several adjustments had to be made for the repressing toehold switches, such as increasing the trigger RNA length from 30 nt to 45 nt. However, increasing interaction domains of toehold repressors with trigger RNAs might also lead to increased cross-talk of non-cognate trigger RNAs. Toehold repressors in their untriggered state expose the RBS and the start codon, being accessible for ribosome binding. As secondary structures around the RBS and the start codon highly affect translation efficiency, shorter homology domains of about 10 nt might be sufficient to significantly repress translation from the untriggered state. Hence, variation of trigger RNA length might significantly affect specificity as well as cross-talk. Following *in silico* predictions of structures and aspects of RNA thermodynamics several toehold repressor switches were designed that displayed a high dynamic range of repression. Based on initial *in vivo* screening, key parameters responsible for high ON/OFF ratios were extracted. These designs were subject to further optimization. This resulted in an improved set of toehold repressors which subsequently formed the basis for the design of mRNA sensors. In addition to results shown above, the construction of an orthogonal library as well the potential of toehold repressors to function within logic circuits (2-input NAND gate) was investigated. Orthogonality and logic circuit experiments are not shown in this thesis. Overall, toehold repressor switches are composed of programmable modules, allowing its seamless integration with existing biological parts and thus match to the inherent goal of synthetic biology to create a common toolbox of standardized biological components.

### 4.1.5 Materials

#### Strains

Following *E. coli* strains were used in this project: BL21 Star DE3 (F-ompT hsdSB (rB-mB-) gal dcm rne131 (DE3); Invitrogen), BL21 DE3 (F-ompT hsdSB (rB-mB-) gal dcm (DE3); Invitrogen), and DH5α (endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17 (rK-mK+) λ-).

#### Culture conditions

Bacteria were cultured in LB medium at 37°C with appropriate antibiotics as follows: ampicillin (50 μg/ml), spectinomycin (25 μg/ml), and kanamycin (30 μg/ml).

#### Toehold repressor library construction

Gibson assembly and PCR were used for construction of plasmids. Single-stranded DNA was purchased from Integrated DNA Technologies (IDT) and subsequently used for assembly of DNA templates for toehold repressor switch and trigger RNA expression. Amplification of synthetic DNA sequences was performed via PCR. Insertion into plasmid backbones was done via Gibson assembly using 30 bp long homology domains [203]. Validation of all plasmids which were cloned into the *E. coli* DH5α strain was done via sequencing. Plasmid backbones were used from following commercial vectors originating from EMD Millipore: pET15b with ampicillin resistance, ColE1 origin, pCOLADuet with kanamycin resistance, ColA origin and pCDFDuet with spectinomycin resistance, CDF origin. The reporter for the toehold repressor switch was a GFP variant: GFPmut3b-ASV, which is a GFPmut3b with an ASV degradation tag. [204] The mRNA sensor detecting kanamycin resistance used mCherry as reporter output. Relevant sequences used in plasmids are given in the appendix below.

#### Toehold repressor expression

Expression of toehold repressor switch and trigger RNAs was performed via T7 RNAP in BL21 Star DE3 strain, which is an RNase-deficient strain. Induction of T7 RNAP was done by the addition of isopropyl β–D-1-thiogalactopyranoside (IPTG). For some experiments the toehold repressor switch and trigger RNAs were also investigated in the BL21 DE3 strain,

with T7 RNAP induced by addition of IPTG. Cells were cultivated overnight in 96 well plates with a shaking incubator at 900 rpm and 37°C. The following day, cultures were subsequently diluted by 100-fold into fresh LB medium with antibiotics and regrown for 80 min at 900 rpm, 37°C. Cultures were induced with 0.1 mM IPTG and subsequently put back into the shaker at 900 rpm and 37°C until the actual flow cytometry measurement at specific times post induction was performed.

**Flow cytometry measurements and analysis**

For flow-cytometric measurements, a BD LSRFortessa cell analyzer with a high-throughput sampler was used. For actual measurements, the culture was diluted by a factor of 65 into phosphate-buffered saline (PBS). Detection of cells was done using a forward scatter (FSC) trigger with recording of a minimum of at least 30.000 cells for each measurement. Gating of cell population was performed according to their FSC and side scatter (SSC) distributions as previously described [3]. Circuit performance was measured by GFP or mCherry fluorescence levels of accordant gated cell populations. The obtained GFP or mCherry fluorescence histograms showed unimodal population distributions of which the geometric mean was used from to extract the average fluorescence across the approximately log-normal fluorescence distribution. For each sample, at least three biological replicates were conducted. Evaluation of ON/OFF ratio of GFP or mCherry fluorescence levels was achieved by taking the averaged fluorescence output of a toehold repressor switch with a non-cognate trigger and its division by its fluorescence output with a cognate trigger construct. Repression of output fluorescence in percent was calculated by averaging the fluorescence output of toehold repressor switch with a cognate trigger and dividing it by the fluorescence output with a non-cognate trigger. Prior to determination of the ON/OFF ratio and percent repression, cellular autofluorescence was not subtracted.

## 4.2 Riboregulators in mammalian cells

### 4.2.1 Motivation

Chapter 4.1 focused on post-transcriptional control of gene expression in bacteria by redesigning the original activating toehold switch into a repressing toehold switch. This mechanism was transferred into mammalian cells. To date, successful attempts to regulate gene expression in eukaryotic cells solely relying on RNA-RNA interactions are limited, and mostly involve riboswitches, or ribozymes possessing self-cleaving activity (chapter 2.3.3). These systems are based on naturally occurring structural motifs and mechanisms, and have limited potential to be rationally engineered. However, translation initiation is a complex process in eukaryotic systems and its control represents a desirable target in synthetic biology. Already 30 years ago it has been shown that translation efficiency is strongly influenced by secondary structures within the 5' UTR of mRNA [118,120,126]. In particular, this applies to hairpins close to the 5' cap of mRNA. This indicates that secondary structures strongly influence ribosomal recruitment and assembly as well as the positioning at a favorable start codon. Translation initiation involves mRNA scanning of the 5' UTR by the 43S complex until a start codon is encountered. This process is accompanied by unwinding existing secondary structures and subsequent assembly of the active 80S elongation complex. The capability of ribosome complexes to unwind secondary structures strongly correlates with the stage of assembly as well as on the actual secondary structure strength and its position within the mRNA transcript [127]. In particular, 80S elongation complexes have been shown to break up secondary structures that could not be disrupted by the 43S ribosome complexes [126]. In a similar context, the Kozak consensus sequence has to be mentioned. This ~10 nt highly conserved sequence includes the AUG start codon and plays an important role in translation initiation, sequence-specifically impacting translation efficiency [120].

A simple transfer of the bacterial toehold switch into mammalian cells is not possible as translation machinery in eukaryotes substantially differs from its counterpart in prokaryotes (chapter 2.2.2). In this regard, the ability of hairpin structures to hinder translation as well as the importance of the Kozak consensus sequence for initiation of translation represent fundamental challenges for this project. Thus, the overall design and experimental approach had to be adjusted accordantly. To facilitate the understanding of the context, terms such as "switch" or "trigger" as well as sequence domain nomenclature are based on the repressing toehold switch.

The first part of this chapter investigates the impact of secondary structures on translation efficiency. Nine different hairpins varying in lengths and respective strength were designed. Each design included sequestration of the Kozak consensus sequence within the single-stranded loop. Successfully assembled plasmids were transiently transfected into HeLa and HEK293 cell lines. Fluorescence output was measured by flow cytometry and fluorescence microscopy. The strong dependence of translation efficiency on hairpin strength could be confirmed. Obtained results were fundamental for the next step, attempting translational control. Analogous to the bacterial toehold switch, translational activation was proposed by the design of trigger RNA strands able to interact with the switch hairpin RNA via linear-linear mechanism. Unwinding of the stem via toehold-mediated strand displacement reactions and branch migration were aimed to enable ribosome recruitment and assembly, resulting in initiation of translation and subsequent gene expression. In initial flow cytometry experiments targeting for an activation of gene expression were inconclusive. Further investigations on the facilitation of switch-trigger RNA interaction was performed. For visualization and localization within the cell, accordant mRNA transcripts were imaged using the RNA FISH technique. These experiments indicated the successful transcription of both, the switch and the trigger RNA. A series of experiments in which switch and trigger RNA were first transcribed *in vitro* and subsequently tested in various *in vitro* translation kits did not give additional evidence of switch-trigger interaction.

Consequently, an in-depth investigation of the mammalian translation machinery is necessary to understand, predict and control interaction of rationally designed RNA species under such conditions. Nevertheless, transferring the toehold switch from bacteria into mammalian systems would open up an innovative strategy to control gene expression. Building on a *de-novo* gene regulatory system using rationally designed sequences, a successful implementation of mammalian toehold switches would offer new possibilities for diagnostic or even therapeutic applications, e.g. detection of disease related mRNAs or microRNAs.

### 4.2.2 Design and reaction mechanism

Based on bacterial toehold switches, the design had to be adjusted to fit the requirements in mammalian synthetic biology of which the concept will be elaborated in the following.

**Design at the DNA level**

Plasmids are based on a high copy plasmid (Figure 79). Transcription of mRNA is mediated by RNAP II binding to a constitutive CMV promoter. This type of promoter is known for its strong and constitutive expression in mammalian cells and originates from the *Cytomegalovirus* genus. Respective sequence encoding for the switch DNA is inserted 54 nt downstream of the transcription start site. For analysis of the impact of switch hairpins on translation, the sequence of a fluorescent mCherry reporter protein was inserted downstream of each hairpin module.

For this project a bicistronic vector was used and enabled simultaneous expression of two proteins originating from one single mRNA transcript. An Internal Ribosomal Entry Site (IRES), allowing ribosome recruitment independently of the 5' cap was placed downstream of the mCherry sequence. IRES motifs originate from polioviruses and enable initiation of translation of viral RNA in the host organism [205-207]. Downstream of the IRES, the sequence for GFP protein was inserted. Termination of transcription was ensured by a SV40 terminator. Kanamycin/neomycin resistance gene was used as selection marker. Subsequent transcription via RNAP II results in an mRNA transcript as follows: *Switch-(Kozak consensus sequence)-mCherry-IRES-GFP.*

This plasmid design served as basis for any further plasmid modifications.

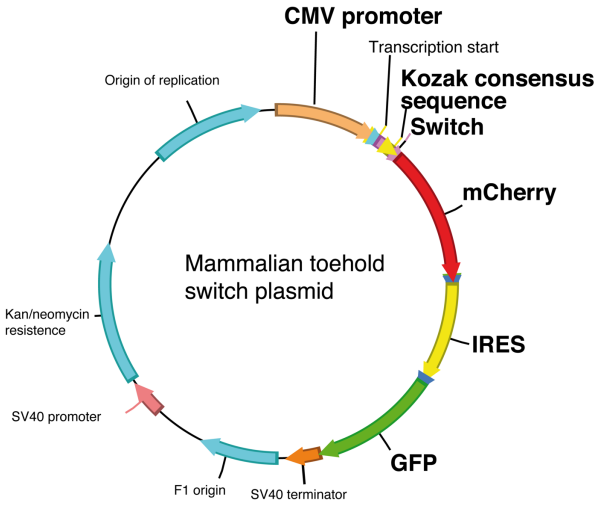


Figure 79. Plasmid map encoding the mammalian toehold switch. Relevant sequences are marked in bold.

## Design at the RNA level

In the subsequent paragraph, the design characteristics of the mammalian toehold switch will be described. Figure 80A depicts a representative mRNA transcript with a cis-repressing hairpin inserted 57 nt downstream of the transcription start site and 32 nt upstream of mCherry mRNA. Each hairpin switch contains a single-stranded loop. Within this loop, the Kozak consensus sequence including the AUG start codon is sequestered, which makes it unfavorable for ribosome binding and subsequent initiation of translation. The exact Kozak sequence used in this project is as follows: CGCCACC**AUG**G. It was expected that translation of mCherry output correlates with hairpin strength as ribosome complexes become less efficient to unwind respective secondary structures. Figure 80B depicts MEF structures of the weakest (left) and strongest (right) hairpin, calculated by the NUPACK analysis tool at 37°C and 1M NaCl [21]. In contrast, expression of downstream GFP reporter protein was independent of hairpin strengths as its translation is initiated by ribosome binding to an IRES, known to be decoupled from secondary structures close to the 5' cap of mRNA.

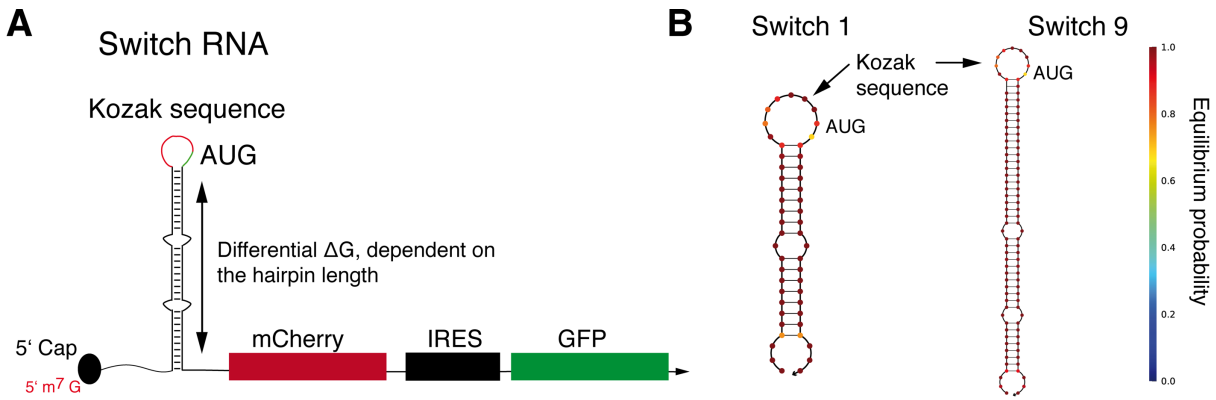


Figure 80. **A**. Schematic illustration of the switch mRNA transcript including the position of the hairpin **B**. MFE structures of the weakest and strongest hairpin calculated by NUPACK [21].

Nine different constructs were designed of which the length of the duplex stem varied from 17 bp up to 41 bp, disrupted by bulges due to single base mismatches, resulting in minimum free energies ranging from -27.1 kcal/mol up to -68.9 kcal/mol, respectively. Table 4 summarizes all tested hairpins with respective hairpin length and minimum free energies, calculated by NUPACK [21].

| | Sw 1 | Sw 2 | Sw 3 | Sw 4 | Sw 5 | Sw 6 | Sw 7 | Sw 8 | Sw 9 |
|---|---|---|---|---|---|---|---|---|---|
| Length (bp) | 17 | 20 | 24 | 26 | 29 | 32 | 35 | 38 | 41 |
| ΔG (kcal/mol) | -27.1 | -34 | -39.6 | -44.1 | -49.1 | 53.7 | -60 | -64 | -68.9 |

Table 4. List of all nine tested switches (Sw) with hairpin lengths and respective free energies at 37°C and 1M NaCl, calculated by NUPACK [21].

**Differentiation between “pre-Kozak” and “post-Kozak” switches**

**“pre-Kozak” switch**

At this point two different variants of switch designs are introduced. Figure 81A illustrates the design in which the Kozak consensus sequence is sequestered in the single-stranded loop and initiation of translation is supposed to strongly rely on the accessibility of this unique sequence motif. This type is identical to the design described in Figure 80 above, and from now on will be referred to as “pre-Kozak” switch.

**“post-Kozak” switch**

To investigate the influence of the Kozak sequence on translation initiation, a variant of the pre-Kozak switch was constructed, termed “post-Kozak” switch (Figure 81B). Individual plasmid design and hairpin lengths as well as the sequestered Kozak sequence are identical to the pre-Kozak switch. The difference lies in the insertion of an additional Kozak consensus sequence 29 nt upstream of each hairpin which is intended to function as additional recognition site for ribosome binding. Hence ribosome recruitment and assembly of the 80S complex is supposed to be favored. The 80S elongation complex has been shown to open up stronger hairpins than the 43S complex [126-128]. As the hairpin is located downstream of this additional Kozak sequence, this variant will be referred to as “post”.

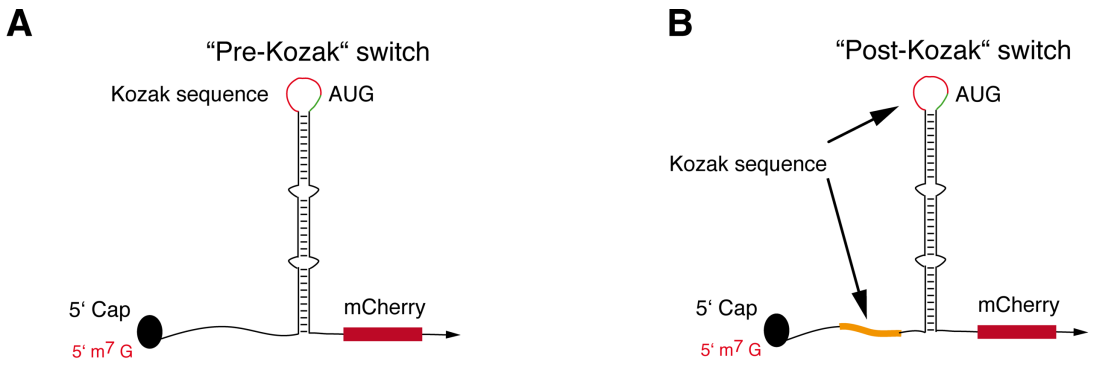


Figure 81. Comparison of the two mammalian toehold switch variants. **A**. “Pre-Kozak” switch with a unique Kozak consensus sequence sequestered within the single-stranded loop of the hairpin. **B**. “Post-Kozak” switch possessing an additional Kozak consensus sequence upstream of the hairpin.

Successfully assembled plasmids with encoded pre- and post-Kozak switches were subsequently transfected into mammalian HeLa and HEK cell lines. Quantitative and qualitative analysis was performed via flow cytometry as well as fluorescence microscopy.

### 4.2.3 Results

#### 4.2.3.1 Analysis of pre- and post-Kozak switches

**Flow cytometry analysis of transfected HeLa cells**

The following flow cytometry results demonstrate the influence of the different mammalian toehold switch constructs in HeLa and HEK293 cell lines. Histograms shown in Figure 82 represent gated living singlet HeLa cells, transfected with pre- and post-Kozak switch plasmids 1-9. Gating steps are shown in Figure 108. FlowJo analysis software was used for subsequent data editing. Overall protocol and data processing was kept consistent for the entirety of flow cytometry measurements.

**Pre-Kozak switches**

Translation of mCherry is highest for pre-Kozak switch 1. This supports the hypothesis that weak secondary structures close to 5' cap do not negatively influence translation activity of ribosomes. Interestingly, this tendency can be observed up to switch 4 which is calculated with a MFE of ΔG -44.1 kcal/mol accordant to NUPACK, and thus represents a relatively stable secondary structure (pre-Kozak switch 2 was corrupted and should be neglected at this point). Interestingly switch 5 displayed a drastic decrease of fluorescence. An increased translation efficiency of switches 6 and 7 in comparison to switch 5 is noticeable. This again points out switch 5 as an outlier of which the exact reason raises an interesting scientific question. Potentially the individual sequence of the secondary structure imposes an energetic barrier for the ribosome. Switches 8 (ΔG = -64 kcal/mol) and 9 (ΔG = -68.9 kcal/mol) inhibit translation most efficiently and consequently represent interesting targets for further investigation.

**Post-Kozak switches**

Post-Kozak switches displayed a slightly different behavior compared to pre-Kozak constructs. Switches 1-4 show similar high translation efficiencies for mCherry output as it is the case for pre-Kozak switches. This further proves the observation that weak hairpins can be broken up by fully assembled 80S elongation complexes, as those could be formed after recognition of the Kozak consensus sequence by 43S ribosome subunits upstream of the

hairpin loop. Again, switch 5 construct caused a reduced mCherry expression. Stronger hairpins (switches 6-9) reveal various levels of weak expression levels combined with a tendency of formation of a second peak at lower expression levels. This "pseudo-bimodal" distribution is different compared to the unimodal distribution of respective pre-Kozak switches, especially for post-Kozak switch 9. Potentially, ribosome subunits compete for assembly at the preceding Kozak consensus sequence. Possibly translation initiation is aborted as the secondary structure inserted downstream is too strong to be completely unwound thus hindering efficient translation of mCherry protein.

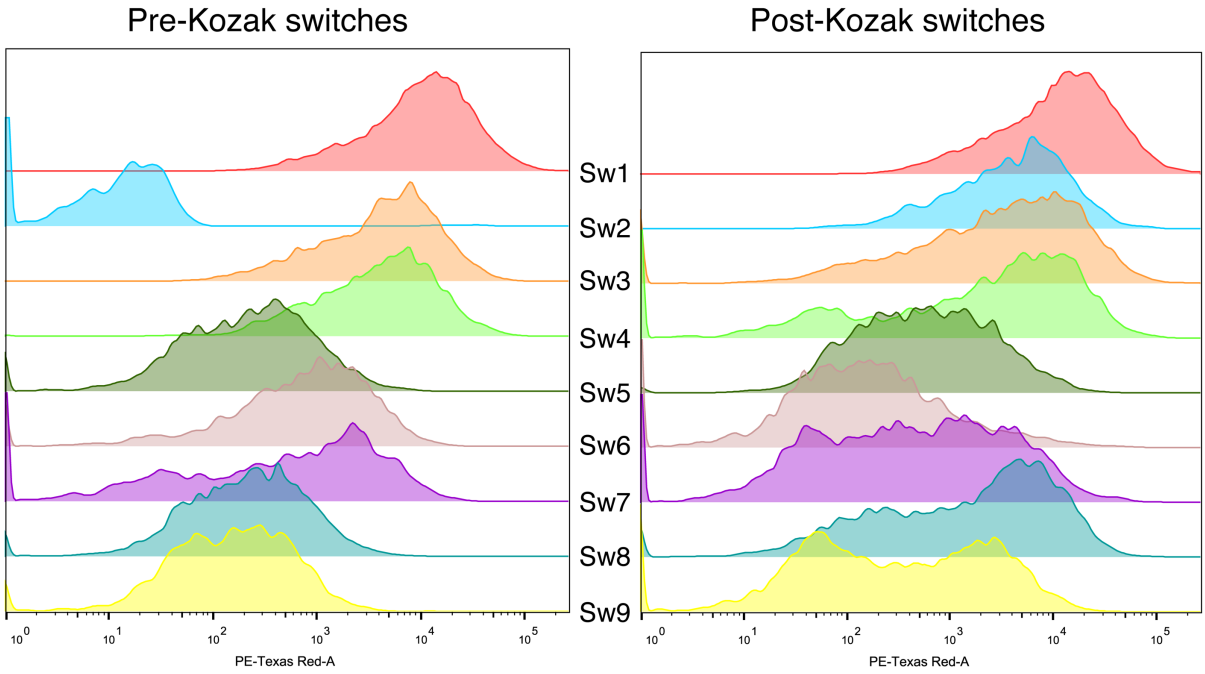


Figure 82. Flow cytometry analysis of HeLa cells, transfected with pre- and post-Kozak switches (Sw). Flow cytometry was performed 48 h after transfection. Fluorescence measurements were normalized reflecting the relative intensity of mCherry protein expression, represented by the mode intensity of the PE-Texas Red channel. Data processing and gating has been performed by using FlowJo software.

Comparing pre- and post-Kozak constructs of switch 8, the drastically reduced fluorescence of the pre-Kozak construct compared to the post-Kozak construct becomes obvious (Figure 83). One might hypothesize that the scanning 43S ribosome complex once it encounters the strong hairpin structure of pre-Kozak switch stops, and "falls off" the mRNA as it cannot unwind the secondary structure. On the contrary, the additional Kozak consensus sequence in the post-Kozak construct allows the 43S complex to recognize the AUG start codon, formation of the 48S initiation complex, which in turn recruits the 60S subunit. The resulting elongating 80S elongation complex may more efficiently unwind the hairpin and initiate translation.

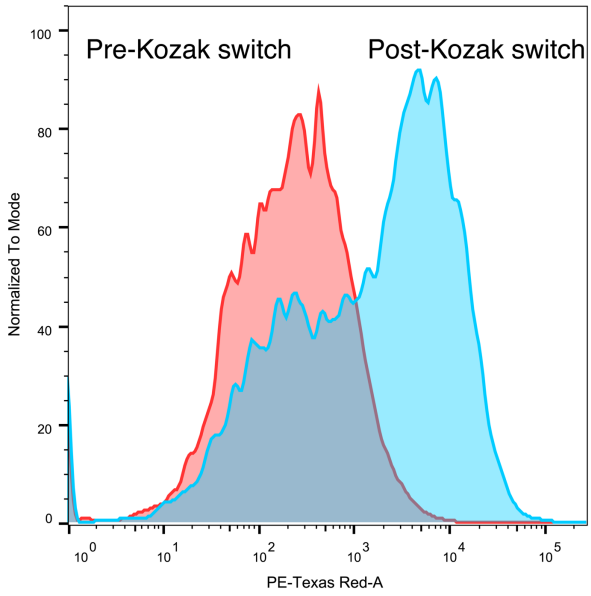


Figure 83. Comparison of switch 8 pre-Kozak (in red) and post-Kozak (in blue). Data are normalized to mode, reflecting the relative fluorescence intensity. Accordant data extracted from Figure 82 above.

## Fluorescence microscopy

In addition to flow cytometry characterization of pre- and post-Kozak constructs, transfected HeLa cells were analyzed by fluorescence microscopy. Respective images shown in Figure 84 support previous data. Pre- and post-Kozak switches 1-4 show a strong expression of mCherry protein, visible by the overlay of mCherry (red) and GFP (green). GFP is constantly expressed, independent of the hairpin strength. Cells transfected with switches 7-9 mostly express GFP, whereas mCherry expression is significantly reduced, indicating that the hairpins impose substantial hindrance for the ribosome subunit assembly. Again, cells transfected with strong post-Kozak switches, especially constructs 8 and 9 reveal a significantly higher expression level of mCherry, compared to corresponding pre-Kozak switches.

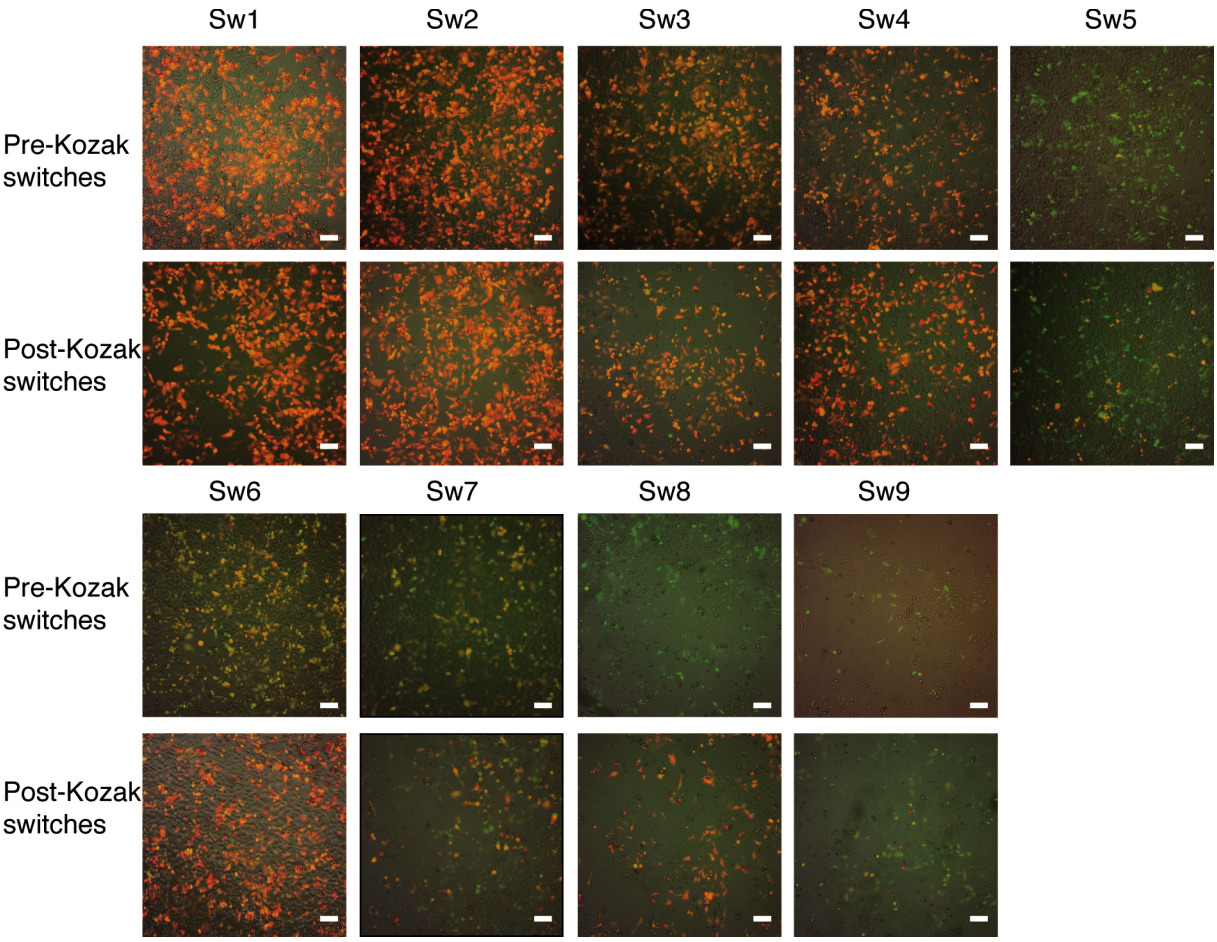


Figure 84. Fluorescence microscopy images of HeLa cells, transfected with pre-and post-Kozak switches (Sw). Images are the result of an overlay of the green (GFP) and red (mCherry) channel. Imaging was performed 48 h after transfection. Scale bars = 100 µm.

**Flow cytometry analysis of transfected HEK 293 cells**

Flow cytometry experiments with the complete set of pre- and post-Kozak switches, were additionally performed in HEK 293 cells. As this cell line is easy to maintain and amendable to various transfection protocols it was used for all subsequent experiments. Results are shown in Figure 85A. High mCherry expression levels were observed for pre-Kozak constructs 1-4, followed by repressed mCherry translation of cells transfected with pre-Kozak switch 5 plasmid. As pre-Kozak switch 5 caused a comparably high repression as in HeLa cells, possibly due to its specific sequence, a closer look into the underlying reason might be interesting. Pre-Kozak hairpins 8 and 9 exhibit drastically decreased expression levels. For cells transfected with post-Kozak constructs, similar results were obtained as for transfected HeLa cells with highest mCherry expression for switch 1, followed by a surprisingly lowered expression for switch 2. Interestingly, post-Kozak constructs with stronger hairpins cause a second “peak” at lower fluorescence intensities, comparable to observations made in HeLa

cells before. Moreover, post-Kozak switches 8 and 9 reveal a significantly increased mCherry expression thus strengthening the hypothesis that ribosomes once they are able to form 80S elongation complexes can unwind strong secondary structures. Cells transfected with plasmid "-mCherry" represent a negative control as this plasmid lacks the sequence encoding for mCherry protein. This construct was used as gating reference and to exclude possible interference of GFP and mCherry fluorescence channels. Overall, these results are consistent within HeLa and HEK293 cells.

In addition to mCherry expression, GFP expression was plotted (Figure 85B). GFP expression levels should be independent of the hairpin strength as the IRES allows independent and constitutive expression GFP. GFP expression was roughly equal throughout pre- and post-Kozak switches. In addition, these measurements confirm that GFP expression is decoupled from mCherry expression without interference, which would result in an undesired crosstalk requiring compensation in the gating steps. Hence, data obtained for mCherry expression are not influenced by GFP expression. However, for this specific experiment, absolute GFP expression intensity of "-mCherry" control is lower than for the switch constructs.

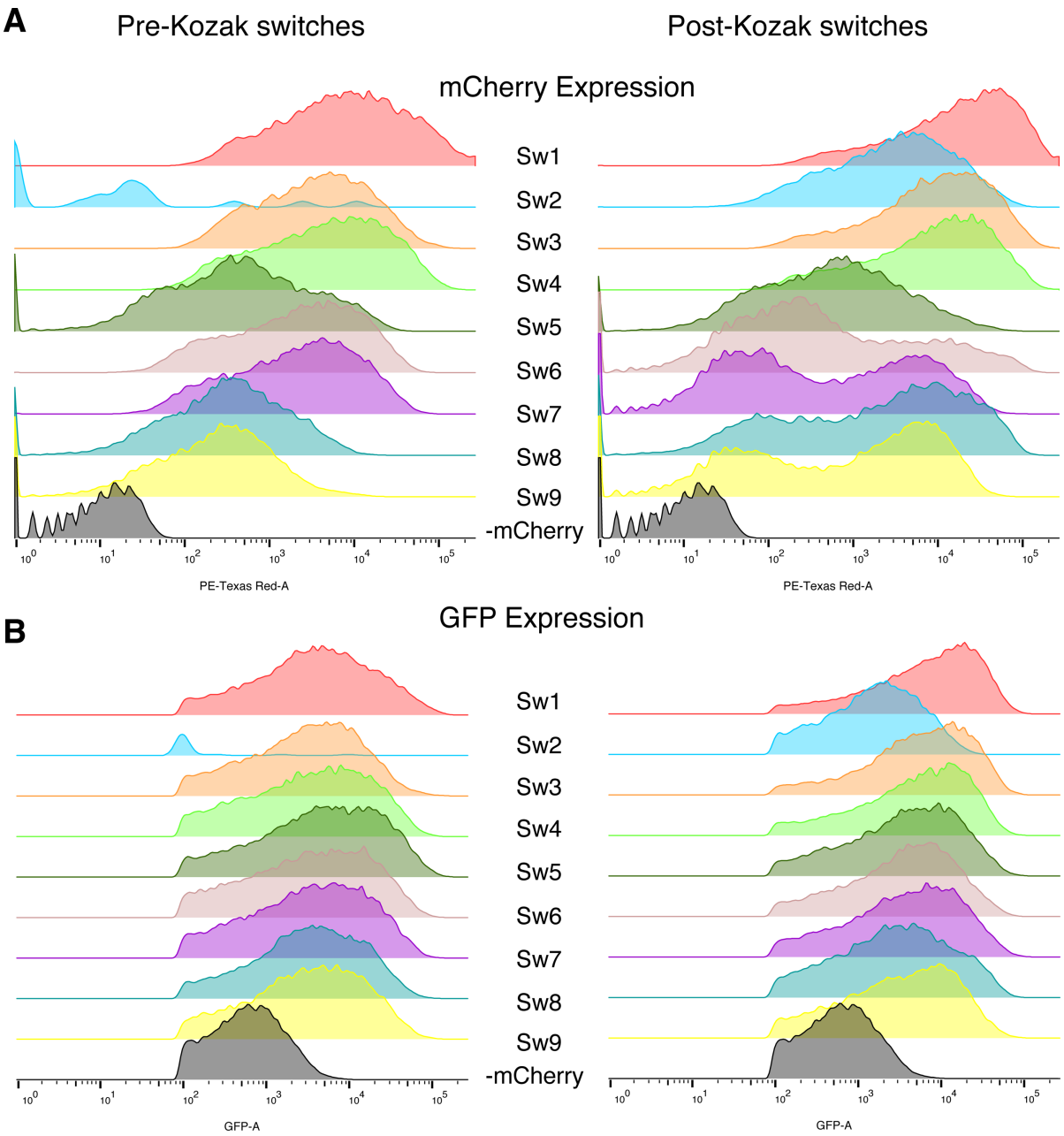


Figure 85. Flow cytometry analysis of HEK293 cells, transfected with pre- and post-Kozak switches (Sw). Flow cytometry was performed 48 h after transfection. **A.** Data set shows expression of mCherry protein, represented by the intensity of the PE-Texas Red channel. **B**. GFP expression levels for the same cell population. "-mCherry" control is included in which the gene sequence for mCherry protein has been deleted. Data processing and gating was performed by using FlowJo software.

### 4.2.3.2 Interaction of switch and trigger RNA

In this paragraph, the goal was to activate initiation of translation by interaction of the switch RNA with complementary trigger RNA strands. Switch 8 was chosen as its pre- and post-Kozak variants resulted in the most distinct differences in fluorescence ON/OFF ratios at reasonably high transfection rates (Figure 83 above).

The underlying mechanism is similar to the one already described in bacteria, as it is based on toehold-mediated strand displacement reactions and subsequent exposure of crucial sequences for translation initiation, see Figure 32 above. Figure 86 illustrates the putative mechanism for the mammalian toehold switch upon trigger binding which results in presentation of the Kozak consensus sequence upstream of the trigger-switch duplex comprising of domains a/a* and b/b*.

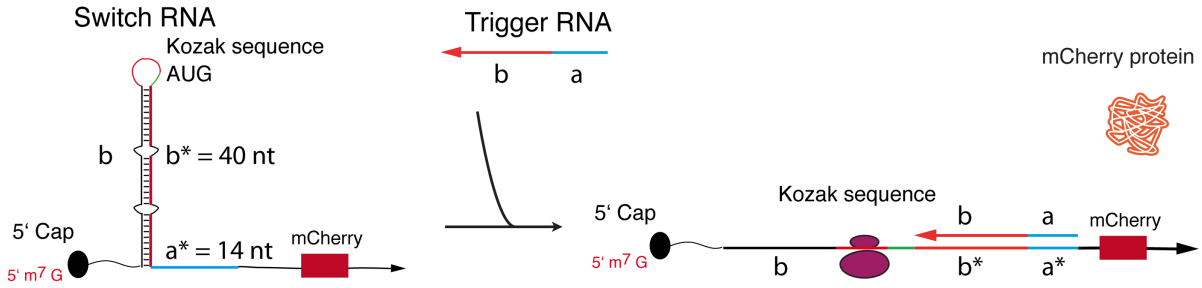


Figure 86. Schematic illustration of translation initiation by interaction of the pre-Kozak RNA switch with a complementary trigger RNA. Hybridization of the trigger RNA results in toehold-mediated strand displacement, unwinding of the hairpin and presentation of the Kozak consensus sequence to the ribosome subunits.

## Design at the DNA level

The plasmid backbone encoding for the trigger RNA including essential elements remains same as for the switch plasmid (Figure 87A). A variant of this trigger RNA construct additionally contained the sequence for GFP under control of an IRES. It was used for preliminary visualization experiments of the trigger mRNA in flow cytometry experiments as well as for RNA FISH experiments (Figure 87B).

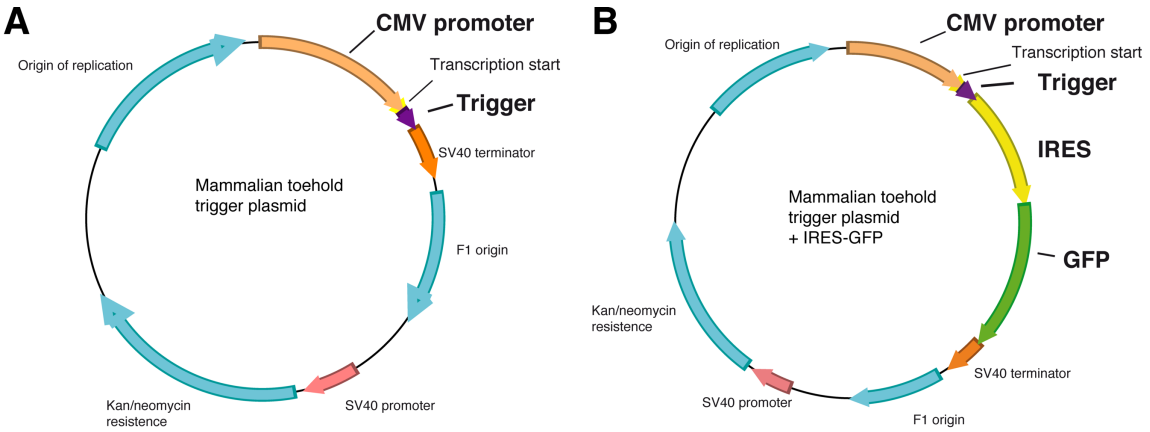


Figure 87. Plasmid maps coding for trigger RNA without (**A**) and with IRES-GFP (**B**).

**Design on the RNA level**

Four different trigger variants were designed and tested which differ in their interaction with the switch hairpin (Figure 88). They are termed “Trigger Full”, “Trigger Half”, “Trigger Junction” and “Trigger Partial Junction”. Specific functionality and design differences are explained in the following with MFE structures, calculated by NUPACK at 37°C and 1M NaCl. All trigger variants possess a conserved hairpin at their 5’ end.

**A. Trigger Full**

The most obvious way of switch-trigger interaction represents a complete unwinding of the stem loop by a complementary trigger sequence via toehold-mediated strand displacement and subsequent branch migration. This includes domains b* and a*. (Figure 88A). Initially existing bulges in the middle of the stem loop are used for additional base pairing. This causes the Kozak consensus sequence to become accessible for ribosome subunit assembly which is initiated by recognition of the start codon via the 43S ribosome complex, resulting in formation of the 48S complex. Subsequent joining of the 60S subunit forms the 80S elongation complex which in turn initiates translation. Eventually existing secondary structures that would have been too stable to be disrupted by the 43S complex have been proven to be penetrable for the 80S complexes [126]. The NUPACK simulation shows the binding of trigger to the pre-Kozak switch, with a MFE calculated with $\Delta G$ = -126.08 kcal/mol.

**B. Trigger Half**

Another variant of trigger RNA is shown in Figure 88B, termed Trigger Half. This trigger type partially opens up the stem loop resulting in a weaker hairpin of 22 bp which reflects hairpin strengths that have been shown to be passable by the scanning 43S ribosome complexes, allowing subsequent steps in ribosome recruitment. The interaction is characterized by partial binding of the trigger RNA to the hairpin stem domain b (19 nt) and toehold domain a (14 nt). Here, MFE is calculated with $\Delta G$ = -117.88 kcal/mol.

**C. Trigger Junction**

Similar to previous designs, the trigger Junction partially opens up the hairpin, resulting in a 21 nt strong hairpin stem with the Kozak consensus sequence still sequestered in the single-stranded loop (Figure 88C). The trigger Junction forms a three-armed junction, partially encompassing domains b*, b, and a with domain lengths of 6 nt, 18 nt and 14 nt, respectively. This structure is intended to result in reduced reversibility (ΔG = -120.88 kcal/mol).

**D. Trigger Partial Junction**

Trigger design Partial Junction represents a slight modification of Trigger Junction (Figure 88D). The trigger length is same as for the junction type trigger. The hairpin is partially unwound. The result is a weakened stem loop, harboring the Kozak consensus sequence which is accessible for the 43S ribosome complex. Trigger binding occurs via domains b = 23 nt and a = 14 nt. ΔG is calculated with -117.88 kcal/mol.

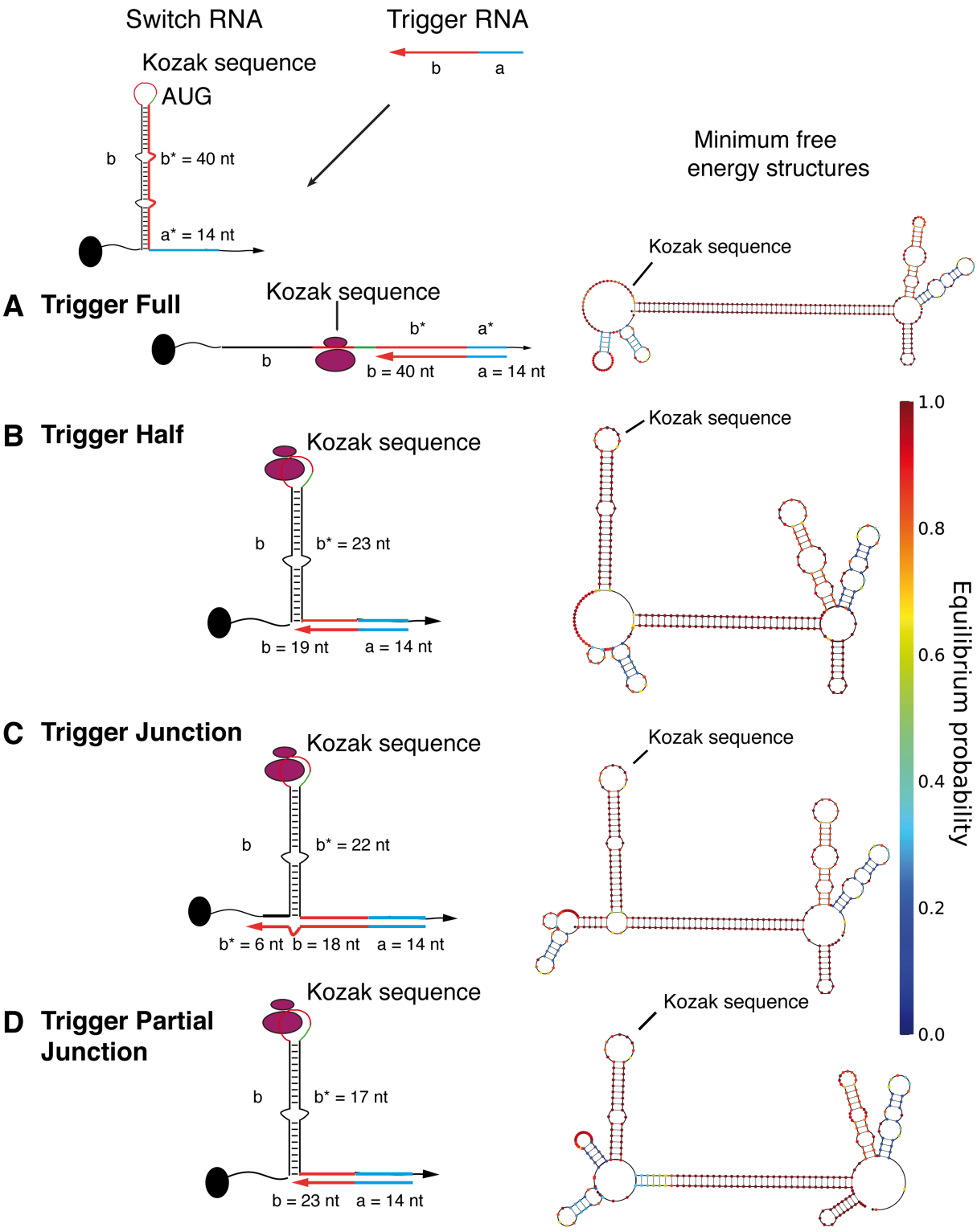


Figure 88. Schematic illustrations the four trigger RNA variants and their interaction with the switch RNA, including corresponding MFE structures, calculated by NUPACK. **A**. Trigger Full unwinds the hairpin completely. **B**. Trigger Half unwinds the hairpin partially. **C**. Trigger Junction unwinds the hairpin partially and forms 3-armed junction to reduce reversibility **D**. Trigger Partial Junction unwinds the hairpin partially and is the same length as the junction type trigger. Respective domain lengths and binding domains are highlighted.

For toehold switch-trigger interactions, a number of preliminary cotransfection experiments were performed of which excerpts will be presented in the following. Experimental readout

was recorded via flow cytometry, using the same protocol as for initial hairpin strengths characterization experiments. For this part of the project, solely HEK293 cells were cotransfected with switch and trigger plasmids as transfection efficiency for this cell line is usually high, and changes in the transfection process can easily be made. Initial characterization and optimization experiments revealed an influence of the ratio of transfected switch and trigger plasmid. The higher the ratio of trigger to switch DNA was set, the lower the overall expression levels for GFP. However, it was assumed that an excess of trigger plasmid DNA over switch plasmid DNA would favor an interaction of those two species, thus requiring a compromise. In subsequent cotransfection experiments the ratio of switch to trigger DNA was 1:4, resulting in transfected plasmid ratios of 300:1200 ng (using 24 well plate), respectively.

**Cotransfection of switch and trigger constructs**

For flow cytometry experiments, at least 50.000 cells were recorded 48 h after transfection. Cells were gated by selecting live singlet cells of which the populations were plotted and analyzed using the FlowJo analysis software. Data are depicted in histograms as well as in two-dimensional dot and contour plots (Figure 89). In a representative dot plot, (left) each dot represents one event, corresponding to a single cell. Contour plots (right) illustrate the density of a cell population in specific regime using equal probability contouring and give indication for the expectation regime of cellular distribution. Here, the Y-axis refers to GFP expression, whereas the X-axis depicts mCherry expression, represented by fluorescence intensities in the PE-Texas Red channel. This additional data representation facilitated data illustration, and delineates a common plot to depict data in flow cytometry.

Figure 89 depicts the distribution of cells transfected with pre-Kozak 8 (red) and pre-Kozak switch 1 (light purple). Data reveal the same trend as in section above. Pre-Kozak switch 8 strongly inhibits mCherry translation whereas pre-Kozak switch 1 does not negatively influence translation. These two constructs will be now used in a different context. Here, cells transfected with these constructs serve as benchmark reference. Pre-Kozak switch 1 is used as “positive control”, displaying high mCherry expression levels, whereas pre-Kozak switch 8 applies as “negative control”, representing a state of strongly inhibited translation.
Ideally, upon cotransfection of switch and trigger plasmid, an activation of the switch RNA by the trigger RNA enhances translation of mCherry which is detectable by a shift to the right

within the PE-Texas Red channel, towards those cell populations transfected with pre-Kozak switch 1.

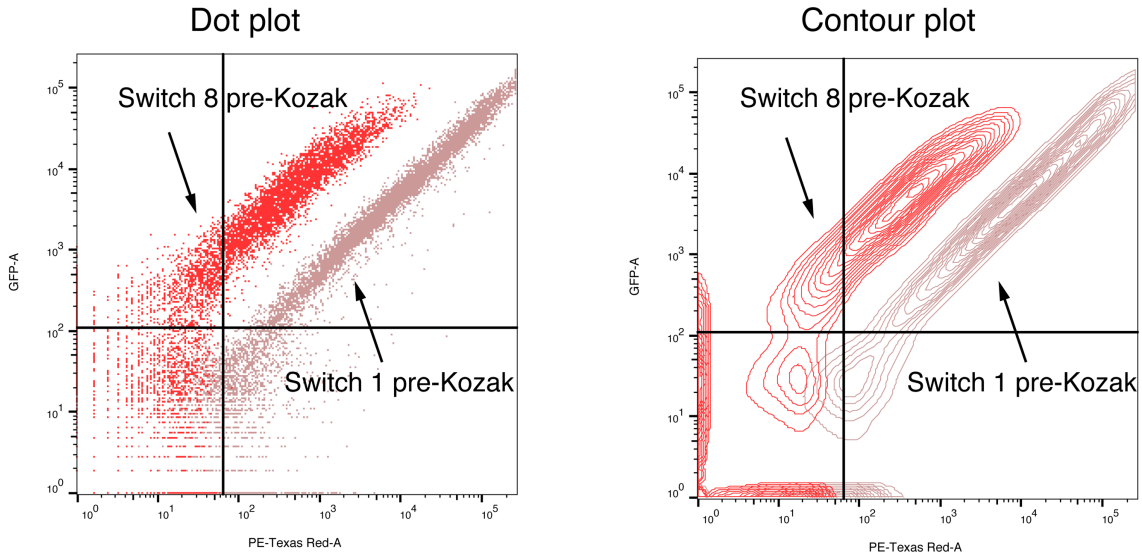


Figure 89. Two-dimensional data visualization of cells transfected with switch 1 pre-Kozak and switch 8 pre-Kozak, displaying GFP expression on the Y-axis and mCherry expression (PE-Texas Red) on the X-axis, using the FlowJo software.

**Cotransfection of switch and trigger under control of the CMV promoter**

Figure 90 and Figure 91 show HEK293 cells cotransfected with switch and trigger plasmids – both under control of a CMV promoter. All four variants of trigger constructs have been tested. An interaction of switch and trigger variants could not be detected as mCherry expression levels of cotransfected cells are even less compared to pre-Kozak switch 8. Trigger variants itself could not be differentiated within isolated cell populations. Moreover, it is clearly visible that GFP expression of cotransfected cells is decreased in comparison to single-transfected cells indicating a lowered switch mRNA transcript concentration. One possible explanation for this observation might be found in the promoters. As both, switch and trigger plasmids are under control of the CMV promoter which is RNAP II-driven, competition for the transcriptional machinery with the promoter on the other plasmid is conceivable.

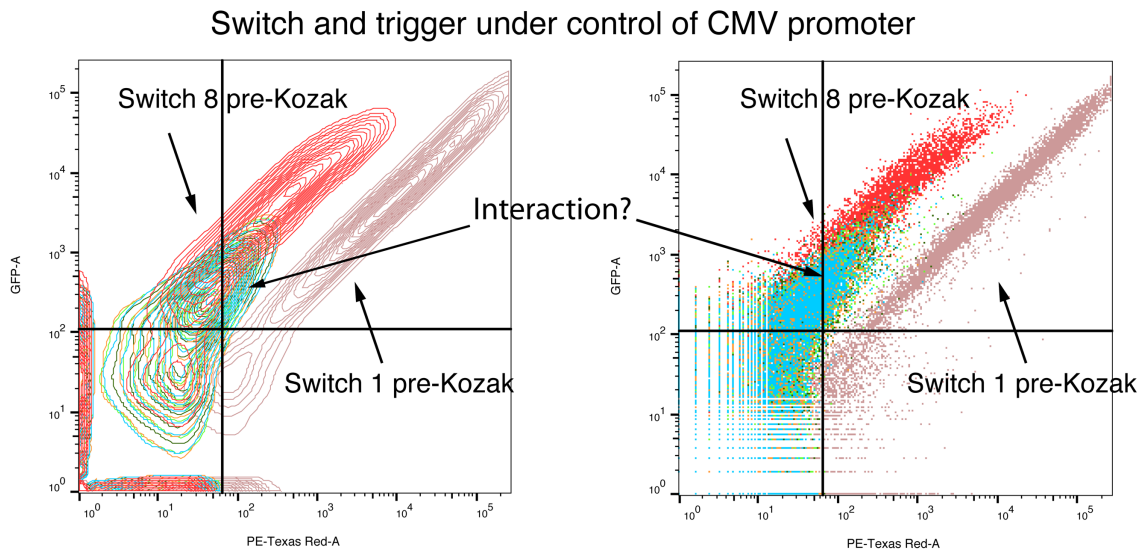


Figure 90. Cotransfection of HEK293 cells with switch and trigger plasmids. Expression levels of GFP and mCherry are represented in a contour plot (left) and dot plot (right). Potential activation of the switch is indicated by arrows.

Histograms manifest observations from Figure 90. Figure 91A as well as Figure 91B indicates that cotransfection of cells clearly affects expression of mCherry as well as GFP, respectively. This effect is possibly induced by the use of two identical promoters for different mRNA transcripts which increases the transcriptional load. Preliminary experiments in which the ratio of trigger to switch plasmid was increased did result in steadily decreased GFP expression levels, which would support the hypothesis of the transcription machinery competing for the same promoters.

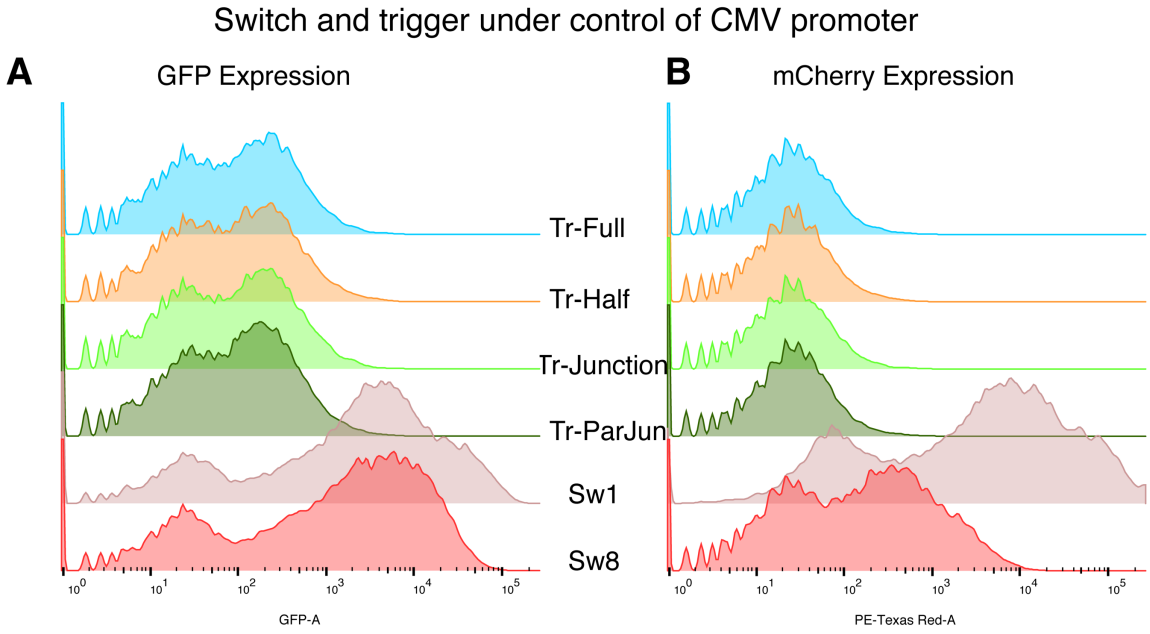


Figure 91. Representation GFP (**A**) and mCherry expression (**B**) of same data set in a half offset overlay histogram with the four different cotransfected trigger variants (Tr-F = Trigger Full, Tr-Half = Trigger Half, Tr-Junction = Trigger Junction, Tr-ParJun = Trigger Partial Junction) and the two controls (switch 1 and switch 8 below).

**Cotransfection with trigger plasmid modified with IRES-GFP**

Alongside with previous measurements, cotransfection experiments were performed with a modified trigger plasmid which contained the sequence for GFP under control of an IRES. The respective plasmid map has been introduced in Figure 87. Consequently, in the case of two mRNA transcripts encoding for GFP, the overall GFP expression level was expected to be enhanced which is confirmed by histogram Figure 92A. The overall signal intensity of GFP expression for cotransfected cells could be improved, which is most likely the result of GFP translation from the switch as well as from the trigger transcript, thus giving indication of trigger RNA production. However, fluorescence intensities were significantly reduced, compared to single-transfected cells, referring to the two controls. mCherry expression for double-transfected cells was comparably low as for previous measurements, compare Figure 91B with Figure 92B. This approach did not help facilitate an interaction. It rather served as indication of trigger RNA transcription.

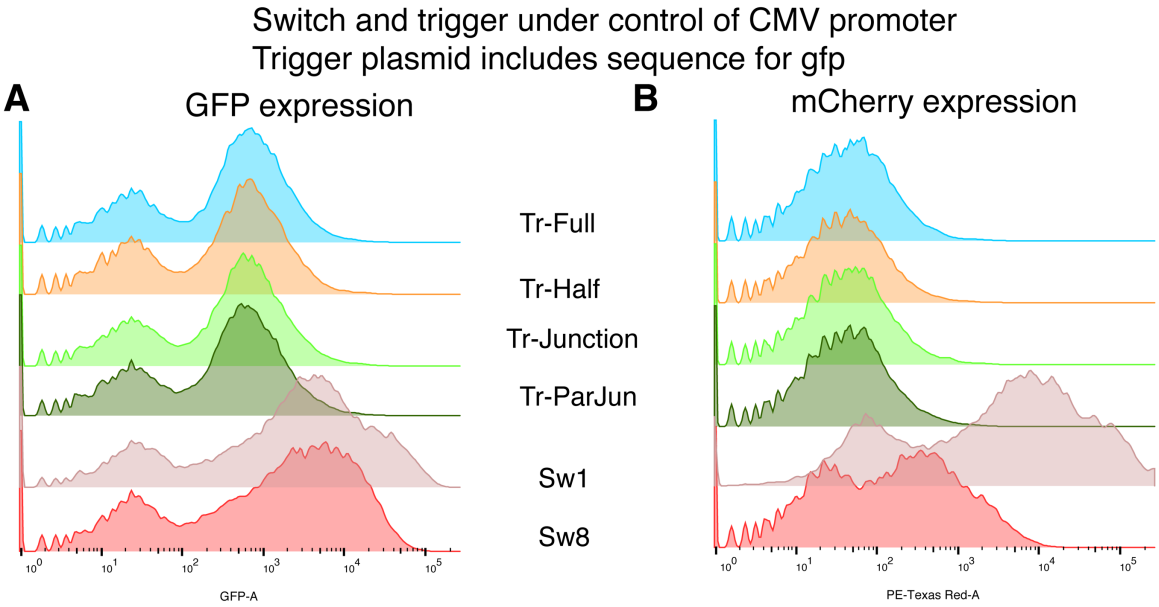


Figure 92. Cotransfection of HEK293 cells with switch and trigger plasmids in which the trigger plasmid contained the sequence encoding GFP under the control of an IRES. **A**. GFP expression of transfected cells in a half offset overlay histogram. **B**. mCherry expression of same data set. (Tr-F = Trigger Full, Tr-Half = Trigger Half, Tr-Junction = Trigger Junction, Tr-ParJun = Trigger Partial Junction).

To address observations made in experiments above and to facilitate an activation of toehold switch several ideas were tested which will be presented in the following.

## Modification of the trigger plasmid

### Replacement of CMV promoter by U6 promoter

Based on observations and conclusions made above one strategy involved the modification of the promoter regulating transcription of the trigger RNA. Consequently, the CMV promoter has been replaced by the human U6 promoter. This promoter type is known for its preference of transcribing small RNA sequences, recognized by RNAP III, instead of RNAP II as it is the case for the CMV promoter.

RNAP III is known to transcribe small RNA species such as tRNA, 5S rRNA and other small RNAs mainly required for cell growth, whereas RNAP II synthesizes mRNA [208]. These small RNA species share the characteristics of not being translated [209]. As the trigger RNA lengths targeting the pre-Kozak switch 8 varied between 62 nucleotides (Trigger Junction) and 78 nucleotides (Trigger Full), and the 5S rRNA is denoted with a length of approximately 120 nucleotides, it was assumed that the trigger RNA would be transcribed by the RNAP III driven U6 promoter as well [210]. Thus, the U6 promoter was cloned upstream of the trigger RNA sequences. The plasmid design is shown in Figure 93. Over the course of experiments the SV40 terminator sequence has been replaced by a series of five thymine spacers as this short sequence is known to sufficiently stop transcription of small RNAs under this type of promoter. Respective trigger sequences as well as the plasmid map remained same. Moreover, these modifications resulted in a smaller plasmid size.

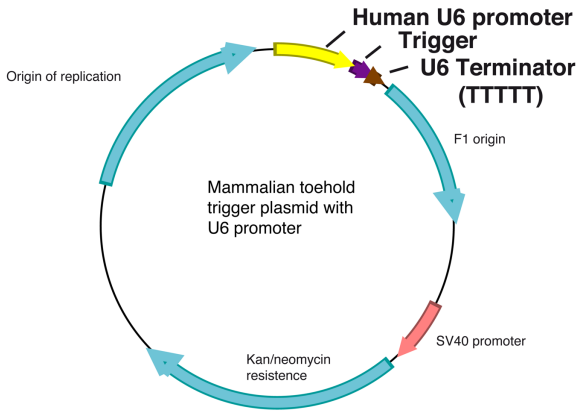


Figure 93. Plasmid map of the modified trigger construct with the trigger sequence under control of U6 promoter and five thymine bases (TTTTT) as termination sequence.

The principal idea of this attempt was to separate transcriptional machineries of switch and trigger plasmid under control of CMV promoter and U6 promoter, recognized by RNAP II

and RNAP III, respectively. This separation should reduce the transcriptional load for accordant RNAP and increase transcription efficiency of both, switch and trigger RNA. Ideally this will result in higher concentrations of RNA transcripts favoring an interaction of switch mRNA and short trigger RNA. Figure 94 compares GFP expression levels of transfected HEK293 cells in which both plasmids were either under control of a CMV promoter (A) or the switch was set under the CMV promoter and the trigger under control of U6 promoter (B). It becomes evident that GFP expression of cotransfected constructs with the trigger RNA under control of the U6 promoter, being subject to RNAP III was significantly higher as in constructs in which both, switch and trigger were under control of the CMV promoter, recognized by RNAP II.

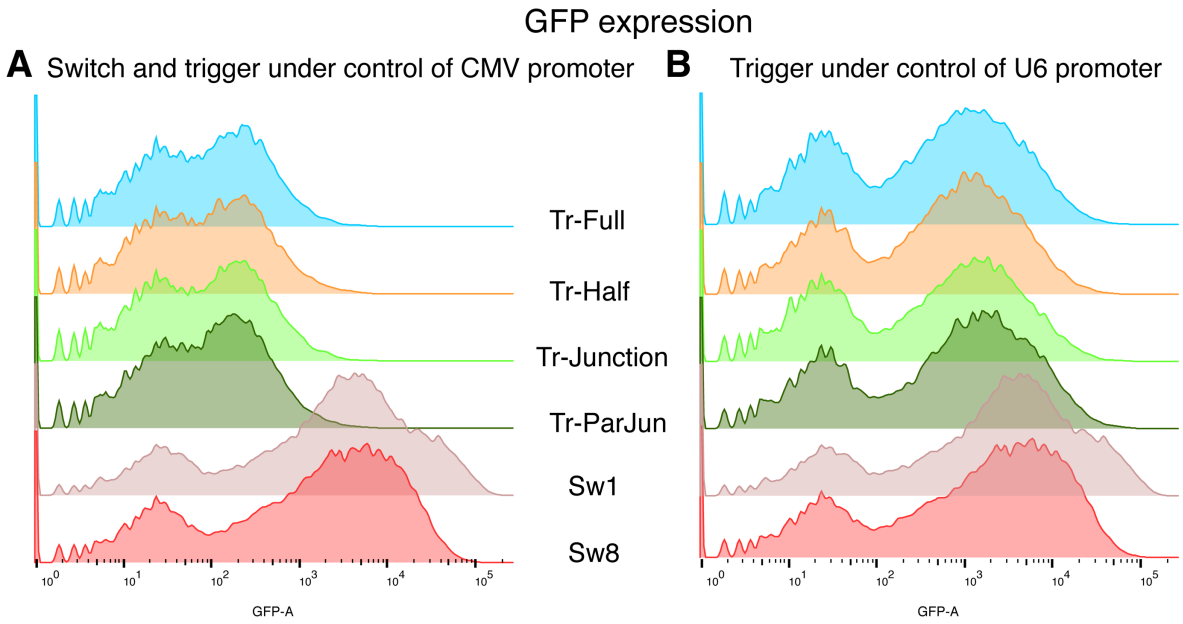


Figure 94. **A**. GFP expression of cotransfected HEK293 cells with both constructs, switch and trigger being under control of CMV promoter. **B**. GFP expression of transfected HEK293 cells with the trigger RNA controlled by the U6 promoter. (Tr-F = Trigger Full, Tr-Half = Trigger Half, Tr-Junction = Trigger Junction, Tr-ParJun = Trigger Partial Junction).

Increased overall GFP intensity also affected potential switch-trigger interaction, indicated by increased mCherry expression even though clear evidence of an activation of the switch by the trigger RNA remains to be proved, see Figure 95. Apart from improved GFP expression levels, a slight shift of cell populations towards increased mCherry expression can be detected along the diagonal axis even though differences of the individual trigger variants could not be identified, similar to initial experiments. Hence, it remains unclear and point of discussion whether in these cell populations the pre-Kozak switches have been activated by trigger binding. Moreover, mCherry expression levels for cells transfected with pre-Kozak switch 1 are an order of magnitude higher than in those cell populations in which an interaction is

speculated (Figure 96). Similar results were obtained in a series of cotransfection experiments with the trigger plasmid containing the U6 promoter as well as still the original SV40 terminator sequence.

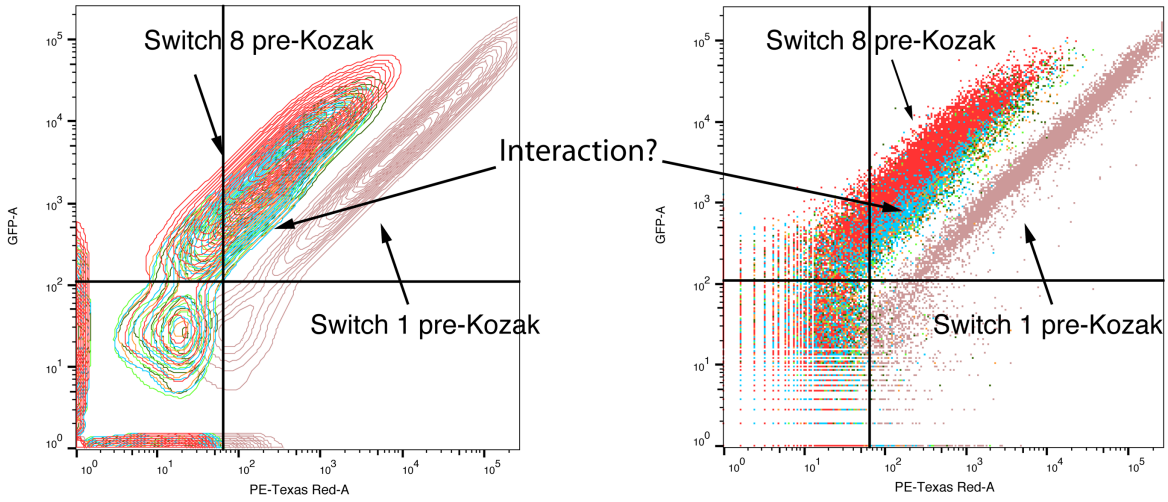


Figure 95. Contour and dot plot of cotransfected HEK293 cells with switch and trigger plasmids. The switch is under control of RNAP II machinery, whereas the trigger is transcribed via RNAP III.

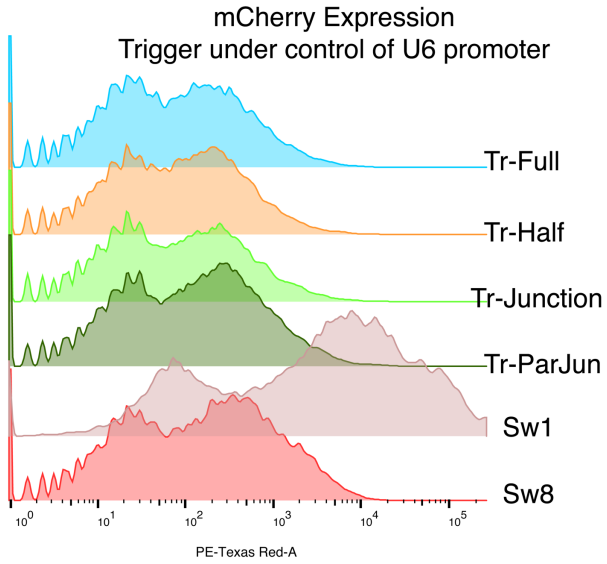


Figure 96. Representation of mCherry expression of same data set as in Figure 95 above in a half offset overlay histogram with the four different cotransfected trigger variants and the two controls (pre-Kozak switch 1 and 8). (Tr-F = Trigger Full, Tr-Half = Trigger Half, Tr-Junction = Trigger Junction, Tr-ParJun = Trigger Partial Junction).

### Transfection of *in vitro* transcribed trigger RNA together with switch plasmids

Another approach to potentially drive switch-trigger interaction towards the activated state was to use *in vitro* transcribed and processed trigger RNA and transfect it alongside with switch plasmid DNA. The general approach goes into the same direction as changing the

promoter of the trigger from CMV to U6. The idea of this attempt was to maximize the likelihood of switch and trigger interaction by decoupling their individual transcription and translation machinery. By possibly incrementing the intracellular concentration of switch and trigger RNA, an interaction of these two species might become more probable, especially as the dimensions and volumes of mammalian cells are orders of magnitude larger than those of bacteria. *In vitro* transcription is a well-established method to produce high concentrations of RNA. Modified (5' capped and 3' tailed) trigger RNA was transfected alongside with plasmids encoding for the pre-Kozak switch 8 (Figure 97). Amount and ratio of transfected switch and trigger RNA was kept constant as in previous experiments. Similar to previous figures, GFP expression of those constructs was investigated first (Figure 97A). Mixing *in vitro* transcribed RNA and plasmid DNA in the transfection process did not negatively affect GFP expression of switch plasmids in general. GFP expression intensities were reasonably high for a similar cell distribution as it was observed for the populations of the controls.
However, the absolute number of cotransfected cells expressing high GFP levels was slightly less, especially for cells transfected with Trigger Partial Junction.

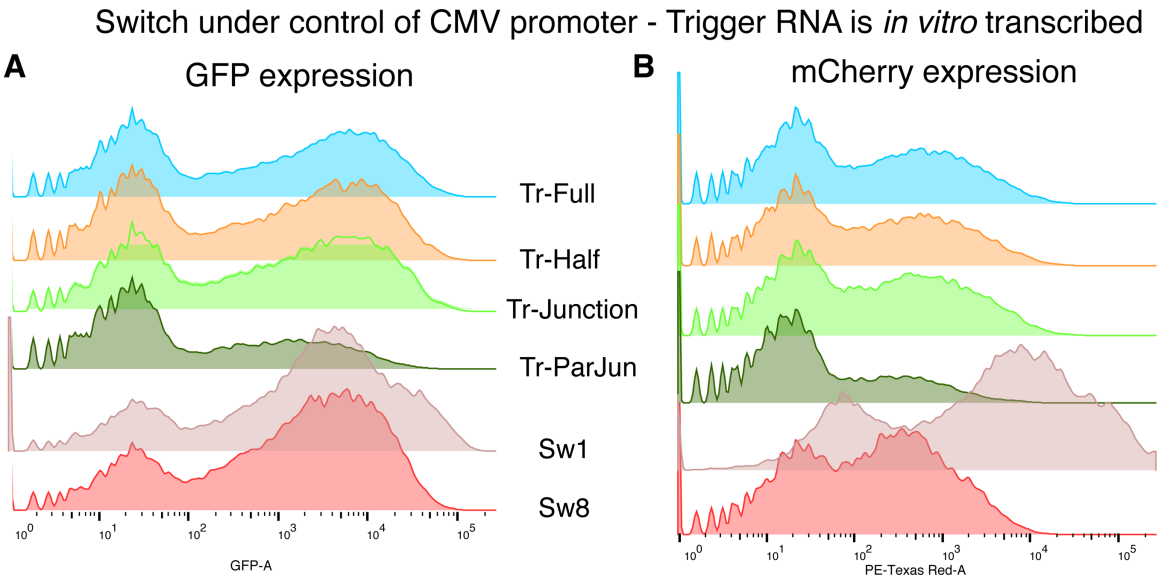


Figure 97. Cotransfection of HEK293 cells with switch transfected as plasmid DNA and trigger transfected as *in vitro* transcribed RNA. **A**. GFP expression of transfected cells in a half offset overlay histogram. **B**. mCherry expression of same data set. (Tr-F = Trigger Full, Tr-Half = Trigger Half, Tr-Junction = Trigger Junction, Tr-ParJun = Trigger Partial Junction).

mCherry expression of cotransfected cells is shown in Figure 97B. The data suggest a slight shift of cell populations to higher mCherry expression intensities, which, in addition are increased in comparison to the pre-Kozak switch 8 control. This trend of "activation" as similarly detected for U6 controlled trigger constructs can be observed in the two-dimensional

dot and contour plots in Figure 98. However, it remains uncertain whether in these cell populations, the *in vitro* transcribed trigger RNA actually causes an interaction with the switch RNA, resulting in the observed shift related to translational activation. Especially, the trigger RNA designs could not be distinguished in their performance even though the putative reactions mechanisms are different.

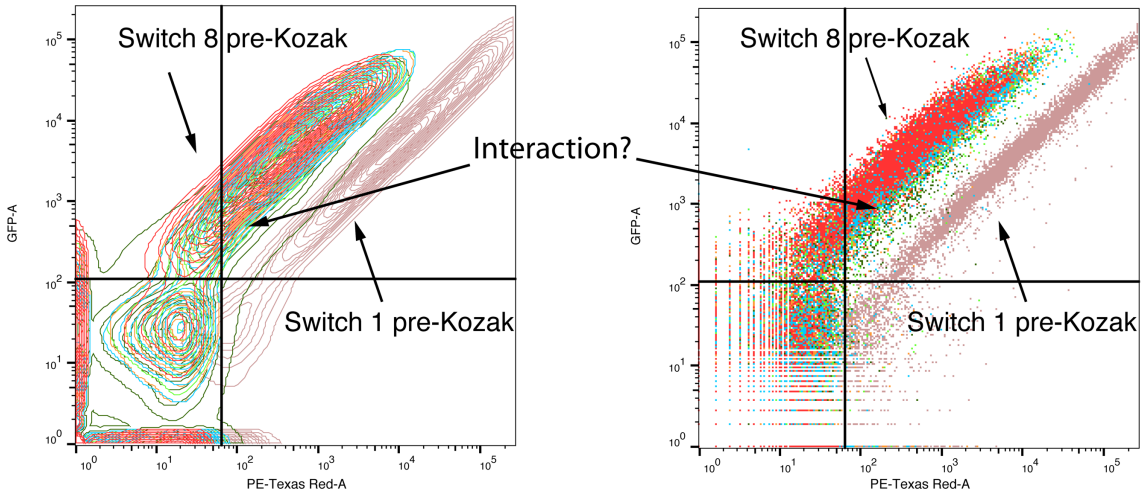


Figure 98. Cotransfection of HEK293 cells with switch DNA delivered as plasmid DNA and trigger transfected as *in vitro* transcribed RNA.

**Preliminary conclusions on switch-trigger interaction**

At this point a preliminary conclusion of flow cytometry investigation of switch-trigger interaction may be drawn. Translation efficiency of mCherry protein strongly correlates with hairpin strengths, especially for pre-Kozak switches 8 and 9, revealing highest translational repression. Based on these insights, pre-Kozak switch 8 was chosen for subsequent experiments aiming for an activation of translation by interaction of the switch RNA with trigger RNA. However, as pre-Kozak switch 5 caused substantial repression even though its hairpin strength is in the intermediate range, this construct might serve for another target switch for translational activation. Four different variants of trigger RNA were designed, each unwinding the repressing hairpin to a different extent. Described preliminary experiments could not show a notable activation of translation in cells that have been cotransfected with switch and trigger constructs. Especially, for constructs being under control of the same promoter type, cells did not reveal a measureable sign of activation. Here GFP expression of cotransfected cells was significantly reduced, compared to cells transfected only with the switch plasmid. Related to this observation, initial experiments reveal a dependence of GFP

expression on the ratio of switch to trigger. Increased trigger to switch ratios result in decreased GFP expression, which is indicating reduced switch mRNA transcription levels. Based on these observations, the assumption was made that switch and trigger RNA are competing for the same transcription machinery which then decreases the switch mRNA concentration. On the other side, it was speculated that an excess of trigger over switch RNA is fundamental for increased chances of interaction.

Subsequent strategies were addressing these observations. Modification of the trigger plasmid and replacement of the CMV promoter by the U6 promoter improved GFP expression levels. This might be explained by the separation of the transcriptional machineries as the U6 promoter is recognized by RNAP III, whereas the CMV promoter serves for recognition for RNAP II enzymes. The trend was observed in few cell populations displaying an increased mCherry expression. Another trial directed at the decoupling of transcription of switch and trigger RNA involved *in vitro* transcription of trigger RNA. Cotransfection of switch DNA and *in vitro* transcribed RNA resulted in some cell populations revealing increased mCherry expression, compared to the switch 8 pre-Kozak control. GFP expression levels were less affected as for example in the initial attempt, in which switch and trigger RNA were transcribed *in vivo* using the same promoter type. However, it remains difficult to draw clear conclusions from the results presented above. Even though the four different trigger designs differed in the binding mechanism with the switch RNA, no differences in their individual performance could be observed in the experiments. This is raises the question of their general interaction probability. Varying the position of the hairpin in respect to the 5' cap of mRNA might also influence the impact the translation efficiency of ribosomes before and after potential switch-trigger duplex formation. Shifting the toehold domain upstream of the hairpin might, upon trigger binding eliminate any secondary structure downstream of the exposed Kozak consensus sequence and thus favor ribosome assembly. Potentially, also newly formed switch-trigger duplexes impede ribosome scanning and assembly activity. Parts of this collaborative work involved preliminary off target experiments with non-cognate rigger RNA molecules which were mainly performed at a time, when my research visit at the Wyss Institute was already over. This approach aimed to investigate a sequence-specific interaction of switch and trigger RNA. However, those results were inconclusive and did not give further insights on how translation in mammalian cells can be controlled by the toehold switch mechanism, unfortunately. For the future, several strategies have to be executed. This includes the rethinking of the overall design approach. Even though post-transcriptional activation of gene expression via of toehold-mediated strand displacement reactions has

successfully been implemented in bacteria, it may place different requirements in mammalian cells of which the reasons are not known yet.

#### 4.2.3.3 Visualization of switch and trigger RNA by RNA FISH

The RNA FISH technique has its origins in DNA FISH (DNA Fluorescence In Situ Hybridization) and allows quantification and localization of mRNA transcripts at the single molecule level. Recently it became possible to even visualize small RNAs such as microRNAs [211]. The coupling of spatial localization and its quantification capability substantially underlines its advantage over quantification methods such as qPCR or Northern Blot.

RNA FISH relies on the hybridization of short DNA probe strands to a target RNA sequence. For visualization in the fluorescence microscope, these probes are labeled with fluorophores. In order to properly resolve the target sequence, it is necessary to label multiple probe strands with fluorophores either at the 5' or 3' end, and design those sequences such a way to allow optimal binding to the target RNA via Watson Crick base pairing. Only one probe strand bound to a RNA sequence would hardly be sufficient for confocal microscopic detection. RNA FISH was used to study the spatial arrangement and quantity of switch and trigger RNA. This might provide further information on the likelihood of potential interaction of switch and trigger RNA as it delivers a different type of data quality and quantity, compared to flow cytometry. In this project, it has been of particular interest to identify whether the RNA species have been exported out of the nucleus into the cytosol or remain in the nucleus, therefore preventing translation of the switch mRNA.

HeLa and HEK293 cells were transfected with either switch or trigger plasmid DNA following the same protocol as for flow cytometry experiments. After incubation for 48 h cells were fixed in methanol and incubated for 24 h together with a pool of 80 5' Cy3b-labeled probe strands. Each of the probe strand was 20 nt complementary to the corresponding mRNA. In addition, the probe strands were extended by two thymines, followed by a nine-nucleotide long DNA-PAINT docking site that eventually allows for RNA FISH PAINT experiments – which were not performed in this project. The final concentration of each individual probe strand was set to 100 nM, which after some preliminary experiments turned out to be a suitable concentration. Too high concentrations would hinder identification of single mRNA transcripts, resulting in aggregates and non-specific binding, using too low concentrations would result in intensities below detection limit. For analysis of switch mRNA, following domains were used for probe strand design, resulting in a set of 80 probe strands capable to hybridized with the mRNA. However, strong structures in the mRNA may prevent strong binding of individual probe strands. The switch mRNA domains accessible for

probe strand binding are as follows: switch length: ~95 nt, mCherry: 780 nt, IRES: 576 nt and GFP: 720 nt (Figure 99).


Figure 99. Sequence domains of switch mRNA theoretically accessible for target probe strand binding.

Trigger RNA offered substantially less domains that can serve as binding regions for probe strands. For this part of the project, the trigger construct containing GFP and IRES was used of which the transcribed mRNA could serve as target for the probe strands (Figure 100).


Figure 100. Sequence domains of trigger mRNA accessible for target probe strand binding.

Representative results are shown in Figure 101 for HeLa cells. Each sample is represented by a bright field image (DIC), by the Cy3b channel (excitation: 545 nm, emission: 572 nm), visualizing the probe strands having bound to the mRNA, and by a DAPI (4',6-diamidino-2-phenylindole) channel, (excitation: 353 nm, emission: 465 nm) which visualizes the nucleus. Constructs 1 pre-Kozak and 8 pre-Kozak as well as Trigger Full were targeted.

Bright field images depict cells in a healthy condition as they are single layered, having created a network of intercellular connections. In each of the samples which have been transfected with the switch constructs, it is clearly visible that probe strands did bind to the switch mRNA. In some cases, one can detect single spots which are supposed to represent individual mRNA molecules. The overlay gives evidence of the actual localization of mRNA. It seems that mRNA has been exported out of the nucleus as fluorescence spots can be found in the cytoplasm with a tendency to aggregate around the nucleus. Similar results have been gained for the Trigger 8 Full with mRNA molecules ubiquitously distributed within the cytoplasm, giving indication that nuclear export was successful. The overall fluorescence intensity in the Cy3b channel for probe strands targeting the trigger mRNA is less compared to probes targeting the switch RNA. This observation can be explained by the reduced sequence length of the trigger mRNA which in turn decreases potential probe strand binding sites and thus lowers the fluorescence signal.

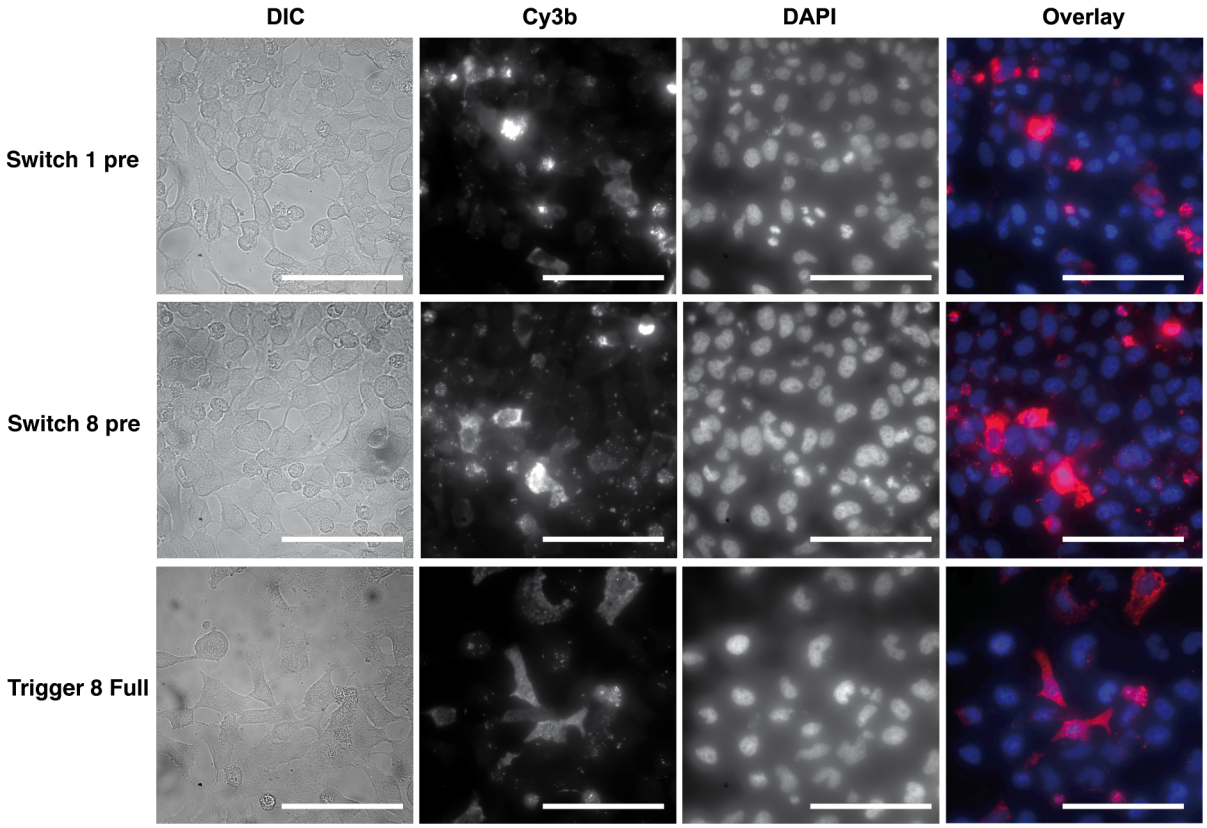


Figure 101. Fluorescence microscopy images of RNA FISH experiments targeting switch and trigger mRNA of transfected HeLa cells. DIC represents bright field, Cy3b represents probe strands, DAPI is used for staining the nucleus. Additionally, Cy3b and DAPI channels are merged and shown as overlay. Scale bars = 100 µm.

Figure 102 represents control experiments which were performed alongside with the samples to exclude false positive results. Figure 102A shows untransfected cells to which 100 nM of each individual probe strand were added. No fluorescence in the Cy3b channel could be detected. Thus, the Cy3b channel was excluded from the overlay in all controls. Overlays of bright field and nuclear staining reveal the healthy state of the cells. In Figure 102B, cells were transfected with pre-Kozak switch 8 plasmids but not treated with the probe strands. Here no fluorescence in the Cy3b channel could be detected, hence implying that fluorescence of actual samples results from binding of the probe strands to the specific target switch or trigger mRNA. The last control represents a double negative control without transfection of plasmid DNA and further treatment with probe strands (Figure 102C). No fluorescence is visible in the Cy3b channel. Overall viability of those cells is comparable to the actual samples thus ensuring that transfection and the subsequent fixation process did not harm the condition of the cells.

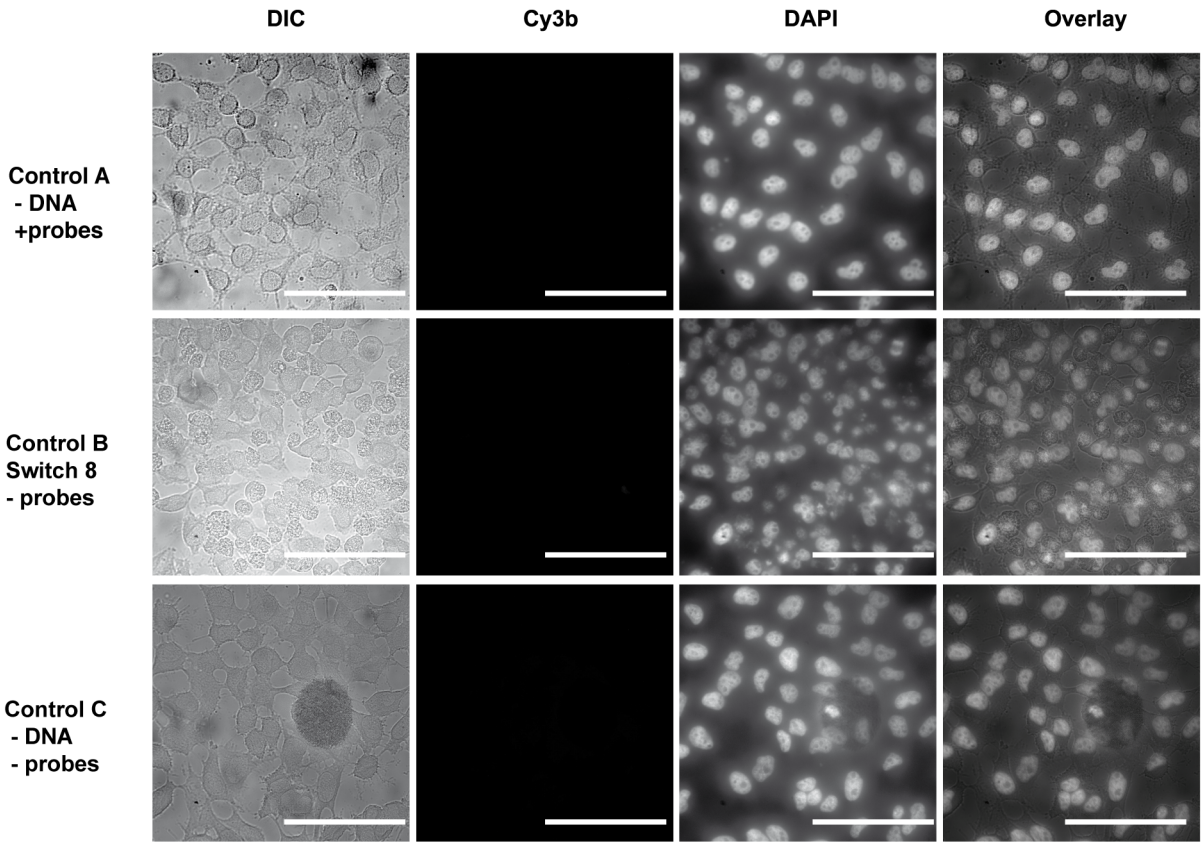


Figure 102. Control experiments. **A**. Control with untransfected cells to which probe strands were added. **B**. Control with cells transfected with pre Kozak switch 8 plasmid DNA, but which were not incubated with probe strands. **C**. Control of untransfected cells without incubation with probe strands. Scale bars = 100 μm.

The same experiments have been performed with HEK293 cells for pre-Kozak switch 8 and trigger 8 Full. Similar results were obtained with tendency for less prominent visualization for cells transfected with trigger constructs (Figure 103).

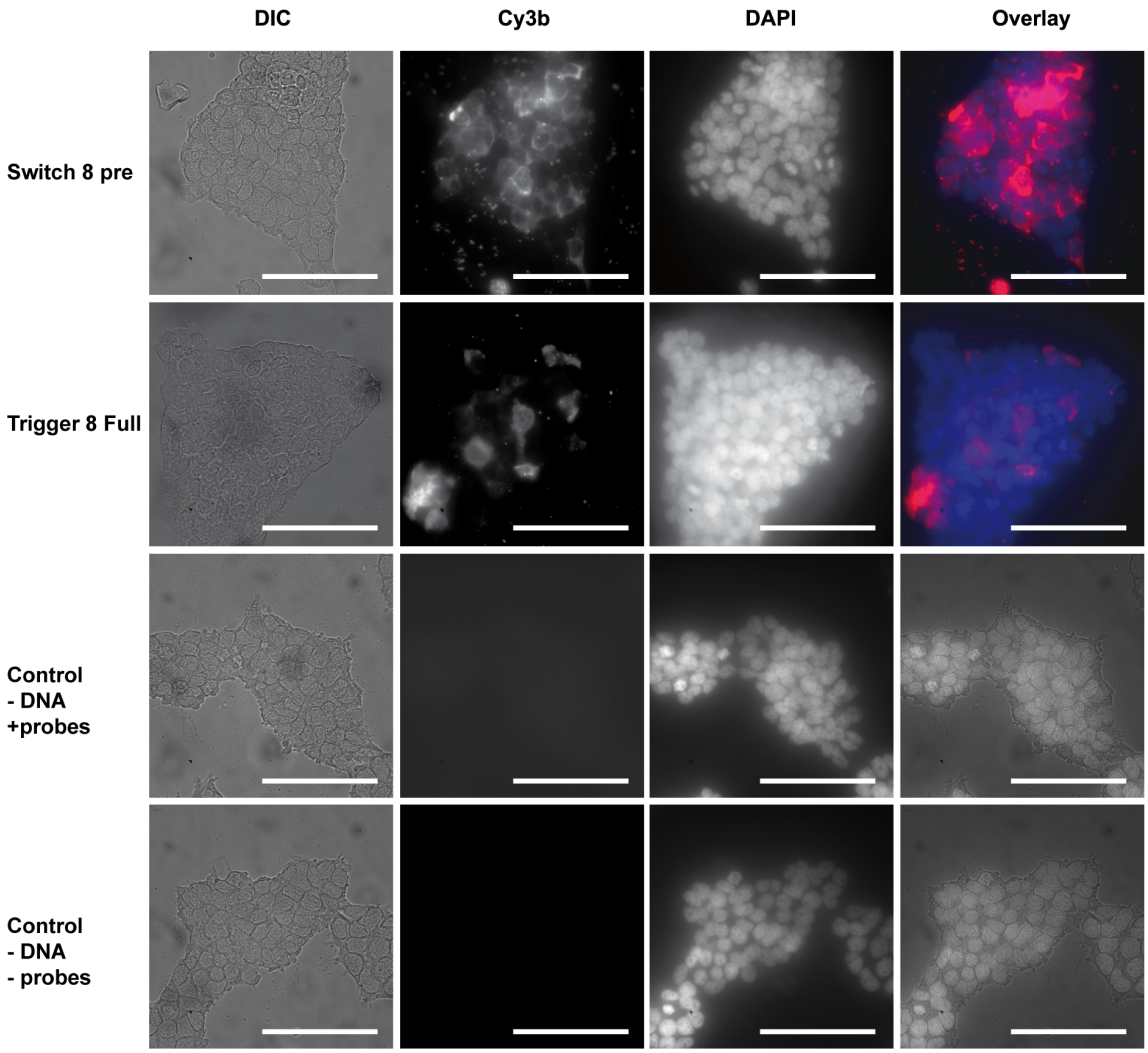


Figure 103. Fluorescence microscopy images of RNA FISH experiments targeting switch and trigger mRNA of transfected HEK293 cells, including controls. DIC represents bright field, Cy3b represents probe strands, DAPI is used for staining the nucleus. Additionally, Cy3b and DAPI channels are merged and shown as overlay. Scale bars = 100 µm.

### Preliminary conclusions on RNA FISH experiments

In parallel to flow cytometry measurements, RNA FISH was applied as additional quantitative method to visualize switch and trigger RNA molecules inside the cell. This was supposed to give some indication of the expression level of the RNA species as well as of their localization within the cell. Therefore, HEK293 and HeLa cell lines were transfected with pre-Kozak switches 1 and 8 as well as Trigger Full, followed by hybridization with the probe strands and subsequent imaging. Results indicate successful binding of dye-labeled DNA probe strands to the switch and trigger target RNA, visible by the loci within the cells. However, in those preliminary experiments the data do not allow an absolute and accurate quantification of mRNA molecules inside the cell as fluorescence spots were not clearly distinct and defined. Further optimization of the probe strand concentration and adjustments

in the protocol might result in a better resolution of distinguishable spots. Repeating the experiment by using a higher magnification might allow quantification of individual RNA molecules. Even so, the cellular distribution of target RNA was mostly homogenous with a tendency to aggregate around the nuclei. Controls reveal that the observed fluorescence signal has to originate from probe strand binding to the target, as for the untransfected control to which probe strands were added no fluorescence signal was detectable in the Cy3-b channel. These control measurements helped to clarify false positives. “Blurry regions” of high fluorescence might be the results of probe strands entering individual dead cells even though the cells appeared to be in good condition prior to fixation. For future attempts the design of different sets of dye-conjugated probe strands specifically targeting either switch RNA or trigger RNA sequence would enable to localize each of those species and prove eventual colocalization. As the probe strands were extended by a 9 nt long DNA-PAINT sequence, super resolution experiments would allow a more accurate positioning of switch and trigger RNA. In summary, the results indicate successful transcription of switch and trigger RNA which are uniformly distributed in the cells. However, this technique captures the cells in a fixed state hence not reflecting their actual *in vivo* state under physiological conditions, which might be more desirable.

#### *4.2.3.4 In vitro* studies of switch and trigger interaction

*In vivo* experiments remained challenging and provided unclear results. To further address these observations, several experiments under *in vitro* conditions were performed. *In vitro* studies allow investigation of transcription and translation under more defined conditions compared to the living system, especially in eukaryotes. Plasmids were modified by deleting IRES-GFP, which typically contains complex secondary structures thus being unfavorable for RNAP activity, followed by linearization. For efficient mRNA production, the promoter was replaced by a T7 promoter. Modified and linearized sequences encoding for the four trigger RNA variants were transcribed *in vitro* using common kits and protocols for *in vitro* RNA transcription. Purified trigger RNA was capped at its 5' end and tailed at its 3' end by adding ~ 200 adenine bases by using commercially available kits. 5' capping and 3' tailing represent crucial post-transcriptional modifications for RNA structures as those steps enhance stability of RNA and label them for nuclear export, respectively. Produced RNA transcripts were used for flow cytometry experiments (shown in Figure 97), *in vitro* translation assays in various cell lysates (HeLa cell lysate, rabbit reticulocyte lysate and wheat germ lysate) and simple binding assays using agarose gel electrophoresis. Unfortunately, results for *in vitro* translation assays were inconclusive and inconsistent within repetitive experiments, and thus are not further discussed in this thesis. This also applies to the binding assays which could not prove a measureable interaction of *in vitro* transcribed RNAs.

Nevertheless, Figure 104 shows an interesting result in which HeLa cells were cotransfected with plasmid DNA encoding for pre-Kozak switch 9 and 300 ng of *in vitro* transcribed trigger RNA. Figure 104A depicts cells in which the trigger RNA was modified by capping and tailing, whereas in Figure 104B cells were transfected with unmodified trigger RNA. Treatment with modified RNA does not decrease viability of the cells as they appear in a healthy state having formed intercellular connections, whereas unmodified RNA at same amounts causes significant cell death, visible by the mostly spherical cell clumps. Possibly, unmodified RNA is toxic to cells and activates signaling pathways related to apoptosis.

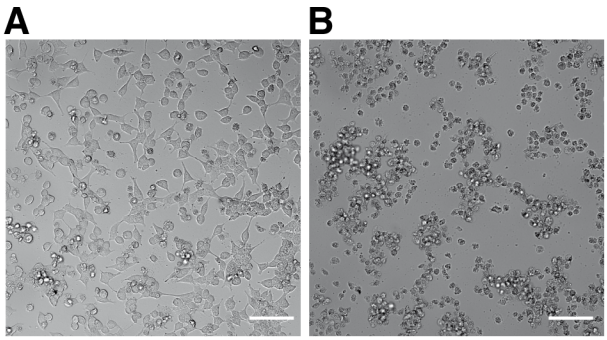


Figure 104. Comparison of cells either transfected with 5' capped and 3' tailed *in vitro* transcribed RNA (**A**) or with unmodified *in vitro* transcribed RNA (**B**), alongside with plasmid DNA encoding for pre-Kozak switch 9. For each, sample, 300 ng of RNA were used for transfection. Scale bars = 100 μm.

### 4.2.4 Discussion and conclusion

The project aimed at the implementation of riboregulators into eukaryotes. In order to control gene expression on the post-transcriptional level, a mechanism inspired by the bacterial toehold switch was introduced and studied in HeLa and HEK293 cell lines. Hairpin loop structures differing in strengths were designed in which the Kozak consensus sequence including the AUG start codon was sequestered in a single-stranded loop. Initial characterization experiments with pre-Kozak switches using flow cytometry and fluorescence microscopy revealed a significant correlation of translational repression with strong secondary structures. Post-Kozak switches which contained an additional Kozak consensus sequence upstream of each hairpin demonstrated a similar expression pattern for weak hairpins (switches 1-4). Though, in contrast to pre-Kozak switches, moderate expression levels for the strongest hairpins (post-Kozak switches 8 and 9) were still measurable. Results described above are consistent with previous works that systematically investigated the influence of secondary structures on translation efficiency close to the 5' cap [118-120,127].

However, the actual goal of this project was to control gene expression by the interaction of switch RNA with trigger RNA. Trigger RNA sequences were designed to be complementary to the hairpin stem of the switch RNA. Upon binding, the stem structure is supposed to be unwound which results in exposure of the Kozak consensus sequence to the scanning 43S ribosome complexes, following initiation of translation. As a consequence, four different trigger RNA variants were designed *in silico* to target the pre-Kozak switch 8. In the course of experimental investigations, various conditions and settings were investigated. Preliminary flow cytometry experiments could not prove a significant and reliable activation. Adjustments in the protocol and modification steps included replacing the CMV promoter of the trigger plasmid by the U6 promoter, as well as cotransfection of *in vitro* transcribed trigger RNA. In some cases, a small cell population seemed to be shifted to the activated state. However, a significant and unambiguous activation in fold changes as shown for the bacterial toehold switch were not achieved. It became clear that a simple transfer of the reaction mechanism from bacteria into in mammalian cells is not straight-forward.

There are a number of possible reasons of which some considerations will be discussed. First, dimensions of bacteria and mammalian cells might play a role. Bacteria (~1 × 3 µm for *E. coli*) are an order of magnitude smaller than mammalian cells of which the diameter for HEK293 cells is reported to be in the regime of 10 – 15 µm [212,213]. Consequently, respective volumes differ by a factor of ~3000 (1 $\mu m^3$ for *E. coli*, 3000 $\mu m^3$ for a HeLa cell, [214]). A

verifiable interaction between switch and trigger RNA might have failed due to the improbability of their encounter within the cell, which in addition is highly compartmentalized and characterized by distinct chemical milieus. At this point it should be mentioned that NUPACK simulations proving switch-trigger interactions cannot incorporate the conditions in a cell and thus are only suitable for a rough approximation. Proving the underlying strand displacement mechanism represents another critical point to investigate. Simple binding assays using synthetic DNA and RNA sequences, labeled with fluorophores might give further insights on interaction of switch and trigger strands. However, RNA FISH experiments successfully resolved switch and trigger RNA molecules in the cytoplasm and around the nucleus. Hence, it might be reasonably assumed that switch and trigger RNA are transcribed. However, it remains to be studied at what stage of RNA maturation switch and trigger RNA are likely to interact with each other, and how this process is controlled. This goes along with mRNA stability and half-life which is based on various *cis* and *trans* determinants such as structural motifs, or RNA binding proteins which influence interactions of RNA structures [215]. To better understand the challenges this mechanism is imposing, it might be interesting to provide information on the actual RNA structure and integrity of switch and trigger RNA by using structural characterization techniques such as SHAPE-Seq [216].

Apart from *in vivo* investigation, a series of *in vitro* transcription and translation experiments were carried out. Given the possibility to precisely control reaction conditions and concentration, switch and trigger RNA were transcribed and processed *in vitro*. Binding assays and subsequent analysis via agarose gel electrophoresis did not give anticipated indications of interaction. Similarly, *in vitro* translation assays performed in HeLa cell lysate, rabbit reticulocyte lysate and wheat germ extract resulted in inconclusive data which revealed a high variance within replicate measurements. A redesign of the mammalian toehold switch structure, which possibly shares more structural similarities to naturally occurring motifs, such as secondary structures found in Internal Ribosomal Entry Sites might be desirable. Possibly, yet unidentified factors are responsible for the so far encountered limitations. Initial experiments in yeast might be another possible approach as this organism represents a eukaryote of reduced complexity.

To summarize, it remains a challenging task to control translation in mammalian cells solely relying on this type of RNA-RNA interaction. Nevertheless, gene regulation in mammalian cells by the aids of *de-novo*-designed RNA switches would open up a new field in synthetic biology as well as for applications in diagnostics or therapy.

## 4.2.5 Materials

### 4.2.5.1 Bacterial culture

**Bacterial strain used for plasmid production**

Switch and trigger plasmids were transformed and produced in following *E. coli* strain: DH5α (endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17 (rK-mK+) λ-).

**Culture conditions for bacteria**

Bacteria were cultured in LB medium containing kanamycin (30 μg/ml) at 37°C.

**Mammalian toehold switch construction**

Gibson assembly and PCR were used for construction of plasmids. Single-stranded DNA was purchased from Integrated DNA Technologies (IDT) and subsequently used for assembly of DNA templates for toehold switch and trigger RNA expression. Amplification of synthetic DNA sequences was performed via PCR. Insertion into plasmid backbones was done via Gibson assembly using 30 bp long homology domains [203]. Relevant sequences used in plasmids are given in the appendix below.

### 4.2.5.2 Mammalian cell culture

Cell lines were cultivated at 37°C and 5% $CO_2$ in an incubator. Cells were grown in 10 ml DMEM medium containing 1x diluted PenStrep antibiotic solution (5000 U/ml Penicillin, 5000 μg/ml Streptomycin, ThermoFisher) in T75 flasks. As HEK293 and HeLa cells are adherent cells, culture flasks and dishes were used which are coated by collagen or a poly-D/L-Lysine matrix. As this modified surface creates a positive charge, typically negatively charged cells will more favorably attach to the surface. Reaching ~90% confluency cells were trypsinized by a 0.25% trypsin/EDTA solution. After removal of the old medium, cells were washed once with 3 ml PBS solution. Subsequently 3 ml of pre-warmed 0.25% trypsin/EDTA was added, followed by an incubation step for 5 min at 37°C. As trypsin is a serine protease, it cleaves peptide chains at the carboxyl side of the amino acids lysine or arginine. By adding it to the cells, the adhesive proteins of the matrix and surface adhesion proteins get cleaved,

causing the cells to lose their contact which results in their singularization into solution. For neutralizing the trypsin, 5 ml of DMEM medium was added, followed by centrifugation for 5 min at 350 g. Subsequently, cells were resuspended in 10 ml of fresh medium and transferred into new culture dishes. HEK293 and HeLa cells were split every 3-5 days in ratios between 1:2 and 1:10 dependent on the concentration and required number of cells.

**Counting cells**

The concentration of cell suspension as well determination of viability was determined by using an automated cell counter (Countess Automated Cell Counter, ThermoFisher Scientific). The measurement principle is derived from the Neubauer counting chamber which consists of a microscope slide with defined depth and grid. It allows cell counting in a defined volume. For the Neubauer chamber, live and dead cells are differentiated by using trypan blue. Dead cells or cells with damaged cell membrane will be filled with this reagent and thus appear as dark spots, making an exclusion from live cells easily visible. Cells were counted for a series of transfection experiments as the cell number influences overall transfection efficiency.

**Transfection of HEK293 and HeLa cells**

Transfection describes the transfer of gene vectors or RNA into mammalian cells by applying chemical or physical methods in order to study gene function or protein expression. Two different types of transfection exist: Transient and stable transfection. In a transient transfection, multiple copies of DNA or RNA of interest are delivered into the cell. However, nucleic acids do not integrate into the host chromosome. Transient transfection is usually accompanied by high expression levels of the introduced gene, which are only maintained for a few days after transfection. Thus, this type is not suitable for long term studies. Here, typically gene expression is analyzed within 12-96 h after transfection, depending on the vector and experiment.
In contrast, stable transfection results in the integration of the transfected DNA into the host chromosome. To obtain a stably transfected homogenous cell population, a selective marker along with the gene of interest is integrated into the plasmid. As only a few copies of the transfected DNA integrate into the hosts genome, expression levels of the transfected gene are

generally lower, compared to transiently transfected cells. However, the successfully integrated gene into the chromosome will be expressed over the time of cultivation.
For most transfection experiments a nonliposomal and low toxic based method was used, which is composed of a mixture of lipids and accessory components, dissolved in 80% ethanol (FuGene HD, Promega). Additionally, lipofectamine 3000 (ThermoFisher) was used in preliminary experiments which generally resulted in lower transfection efficiencies compared to FuGene HD.

**Principle of transfection process**

Among various non-physical methods such as calcium phosphate precipitation or DAE (Diethylaminoethyl)-dextran, lipid-based strategies using cationic lipid mixtures represent the most widely used technique.
The general principle of lipid-based transfection strategies relies on an interaction of lipid structures with the cell membrane forming an encapsulating complex (Figure 105). Cell membranes are amphiphilic lipid bilayer structures, composed of a polar, hydrophilic head group and a nonpolar, hydrophobic tail, resulting in an overall negative net charge at physiological pH. Hence, DNA or RNA molecules are unlikely to pass the cell membrane due to their negatively charged phosphate backbones. This effect is circumvented by mixing nucleic acids with cationic lipid formulations possessing positively charged head groups, enabling to overcome electrostatic repulsion [217,218]. This causes subsequent fusion of the lipid-DNA clustered complex with the cell membrane in a process known as endocytosis [218,219]. By this, nucleic acids are incorporated into the cell membrane and delivered into the nucleus by endosomal uptake and packaging. Subsequently, nucleic acids are released into the cytoplasm by endosomal escape and transferred towards the nucleus via the nuclear pore complex, finally leading to transcription and subsequent translation of the delivered gene.

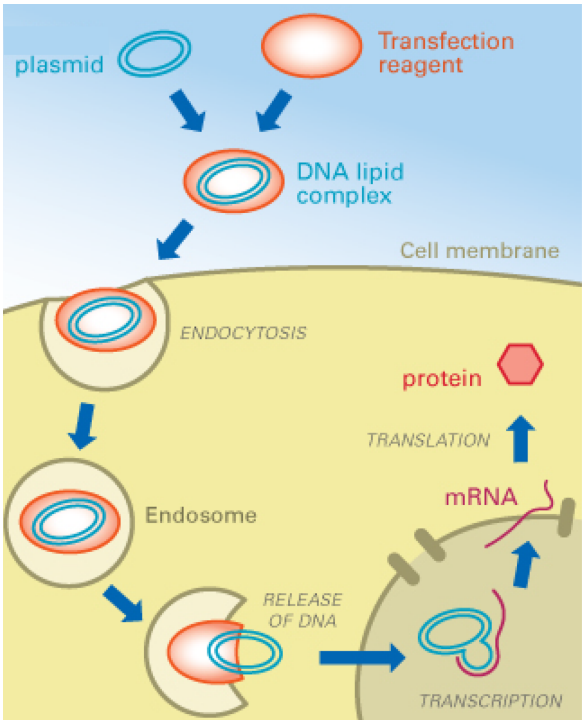


Figure 105. Schematic principle of lipid-based transfection. Modified from [220].

**Cell treatment prior to transfection**

Prior to transfection, cells were seeded into a 24 well plate. Cell number was either determined by a cell counter or its number was kept in a constant ratio throughout the splitting and seeding procedure. This constant ratio results in equal, reliable and comparable confluency throughout a definable period of time. Prior to transfection, equal numbers of cells were seeded in 1 ml of DMEM medium without antibiotics, followed by incubation overnight. The overall viability of cells was analyzed and the transfection agent was equilibrated at room temperature. The ratio of FuGene transfection reagent to DNA concentration was determined in preliminary experiments. For data shown in this thesis 4 µl of transfection agent per 1 µg of plasmid DNA turned out to give best transfection efficiencies.

The mix composed of the plasmid DNA and transfection reagent was prepared separately at a final volume of 26 µl. For transfection experiments, Opti-MEM cell culture medium without serum was used to which plasmid DNA and FuGene reagent was added. This mix was thoroughly mixed by pipetting up and down and subsequently incubated at room temperature for 10 min. Finally, the solution was added to the cells which were incubated overnight. After 24 h, cells were supplied with fresh medium. Flow cytometry experiments were typically performed 48 h after transfection.

### 4.2.6 Methods

#### 4.2.6.1 Flow cytometry

Flow cytometry represents an optical method, based on the light scattering properties of particles. It allows multi-parameter characterization on the single cell level, but on the same time delivers quantitative information. In general, particles down to a size of approximately 100-200 nm can be analyzed. In most cases, this method is used for bacteria and mammalian cells which are about 1 µm and 10 µm in size, respectively. A typical flow cytometer is composed of the following key components: A light source, a flow cell through which the cell suspension is directed, a photodetector which collects incoming photons, an analog-digital converter and computer system to collect the data and analyze the signal [221,222].

In most cases, the primary light source represents a laser beam emitting light at a wavelength of 488 nm. The scattered light provides elementary information of the cell or particle, as well as fluorescence information of dyes that are excited within this wavelength. The number of lasers varies between 7-10 for the most advanced flow cytometers, allowing analysis of up to 20 parameters in parallel for a single particle. The sample is singularized and confined to the center of the flow cell by the sheath fluid which creates by a constant stream at laminar flow, known as hydrodynamic focusing (Figure 106). Sheath fluid and sample are not mixed. Singularized particles or cells pass the flow cell and are hit by a laser beam, resulting in scattered light and emission of fluorescence.

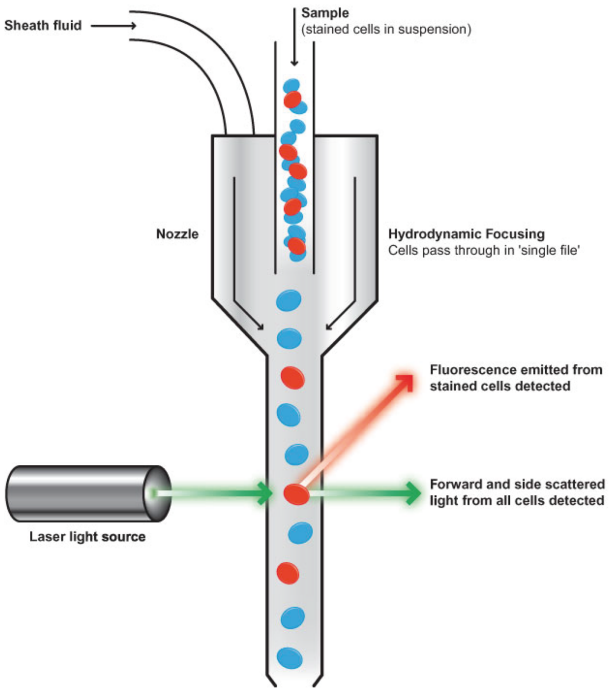


Figure 106. Schematic scheme of a flow cell. Hydrodynamic focusing results in singularization of cell the suspension. Samples pass the laser beam which results in forward and side scattered light and emission of fluorescence. Taken from [223].

The scattered light of the laser beam is divided into forward and side scattering and provides basic information on the individual cell as well as on different cell types, e.g. in tissue samples. Forward scattering determines the size of the cell, whereas side scattering gives indication of the granularity and internal texture of the cell by right-angled refraction. The fluorescence signal is measured by appropriate detectors, which are usually positioned at a right angle to the laser beam and sample stream (Figure 107). Emitted photons are recorded by a photodetector, mostly an avalanche photodiode or a photomultiplier. Hereby the photons hit a light sensitive photocathode and release electrons off its surface, described by the photoelectric effect. These photoelectrons are accelerated in an electric field and release multiple electrons by hitting additional electrodes (dynodes), resulting in multiplication of the input signal. A set of optical filters, beam splitters and dichroic mirrors ensures channeling of distinct fluorescence signals and thus a specified direction of accordant photons to the designated photodetector units. This allows analysis of multiple fluorescence parameters in parallel.

Effectively, light as analogous signal will be converted into an electrical signal, which now can be further processed. Finally, the signal can be evaluated as digital readout by a computer system.

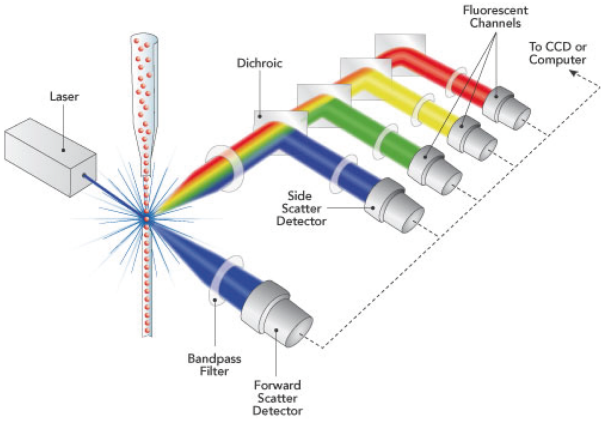


Figure 107. The optical system of a flow cytometer. While cells pass the laser beam, scattered light and fluorescence emission are channeled by optical filters and dichroic mirrors. Electrons are amplified by photomultipliers and collected by photodetectors which convert the analogous signal into an electrical output. Taken from [223].

## Applications

A very common method represents the application of fluorophore-conjugated antibodies specifically targeting a protein of interest, e.g. a surface receptor or an antigen. This allows analysis of cell cycle states or disease related protein expression levels. Moreover, flow cytometry allows monitoring of the changing DNA content within the cell cycle by using dye molecules that specifically bind to DNA. Despites indirect detection of target molecules by fluorophore-labeled antibodies it is also common to probe overexpression of proteins of which the plasmid DNA has been delivered into cells via transformation or transfection. This represents a widely-used method to study intracellular signaling pathways or synthetic gene circuits. In this thesis, flow cytometry was extensively applied to study gene regulation on the post-transcriptional level in bacteria and mammalian cells.

## Sample preparation for flow cytometry measurements

For flow cytometry experiments, transfected cells were trypsinized similar to procedure described before. Old medium was removed, cells washed with 1 ml of PBS solution to which subsequently 1 ml of 0.25% trypsin was added. After detachment of cells, 1 ml of fresh medium was added and the solution centrifuged for 5 min at 350 g. The supernatant was carefully discarded without disturbing the pellet. Finally, cells were resuspended in 300-1000 µl DMEM medium. To exclude potential interference with the laser settings of the flow cytometer, cells were resuspended in DMEM medium without phenol red. Prior to flow

cytometry measurements, 200 µl of cells were transferred into a 96 well plate. A BD LSRFortessa cell analyzer with a high-throughput sampler was used.

**Data analysis**

In this thesis, recorded events were predominantly processed using the FlowJo analysis software. The process is called gating – not to be confused with the term "gating" that for instance is used in electrophysiology. In following example (Figure 108) gating is shown for a set of recorded HeLa cells which were transfected with the bicistronic plasmid coding for mCherry and GFP. Here GFP served as internal control and reference. Figure 108A represents a typical distribution of HeLa cells. All events, representing individual cells are shown and processed in a dot plot with density distribution. First, cells are sorted according to their size (Forward Scatter Area, FSC-A) and granularity (Side Scatter Area, SSC-A). Living cells cluster in the upper center of the dot plot, with characteristic size and granularity. In this example, almost 60% of cells are alive at the point of measurement. This population is selected and depicted in another plot using different scatter settings (Forward Scatter Area, FSC-A and Forward Scatter Height, FSC-H), see Figure 108B. It might happen that two cells pass the flow chamber too quickly and thus cannot be distinguished by the laser as two separate cells. By selecting FSC-H versus FSC-A those doublet cells are excluded from the live singlet cell population. In this example as many as 95% of all cells appear as singlets which are selected for subsequent specification. This step further improves quality of data analysis. These two described steps are common procedures in live cell gating whereas all subsequent gating steps are unique to the individual experiment. In this particular case a negative control of untransfected cells was gated for mode GFP expression for which a threshold was set (Figure 108). Finally, these settings were applied to all samples ensuring normalization.

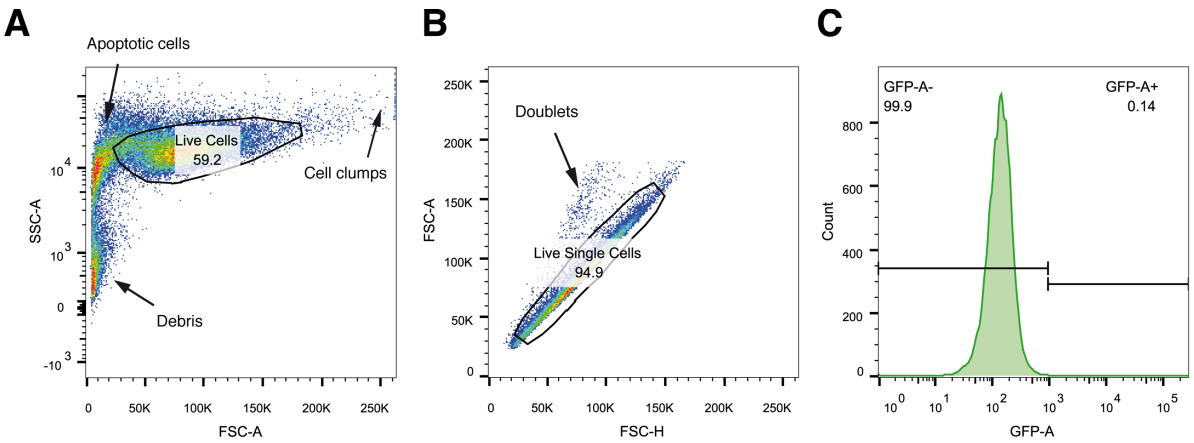


Figure 108. Processing of flow cytometry data via gating. **A**. Selection of live cells via SSC-A and FSC-A. **B**. Selection of live singlet cells via FSC-A and FSC-H. **C**. Selection for GFP positive and GFP negative cells in a histogram.

As an addition, some flow cytometers are equipped with a device allowing to sort cells, known as fluorescence activated cell sorting (FACS).

### 4.2.6.2 Fluorescence microscopy

Fluorescence microscopy represents a widely used optical method to probe fluorescent molecules within biological samples [224]. Based on the principle of fluorescence (see Figure 71 above), this technique allows for the analysis of individual properties of biological specimens with high sensitivity and specificity down to the single-molecule level. By, for instance specifically targeting organelles or proteins via fluorophore-conjugated molecules such as antibodies, cellular substructures as well as protein expression levels can be resolved and visualized. In this thesis, fluorescence microscopy was applied to study expression levels of GFP and mCherry protein in transfected HeLa and HEK293 cells. This study was further exploited by the RNA FISH method which also takes advantage of fluorescence microscopy to visualize mRNA using dye-labeled probe strands specifically hybridizing to the mRNA.

**Sample preparation for fluorescence microscopy experiments**

For fluorescence microscopy experiments, cells were carefully washed once with PBS. DMEM medium without phenol red was added to the cells prior to experiment.

**Sample preparation for RNA FISH experiments**

The protocol used in this project is based on the TurboFISH protocol [225]. For cell culture an 8 well Labtek chamber was used to which 75 µl DMEM medium was added. Subsequently, HeLa or HEK293 cells (100.000 cells/ml) were plated dropwise to the center of each chamber and incubated overnight at 37°C in order to reach a confluency of ~60%. Subsequent transfection was performed accordant to protocol above with 100 ng of plasmid DNA per well, using FuGene HD transfection reagent. Medium was replaced after 24 h, followed by the actual fixation process. Probe strands were labeled with Cy3b dye molecules at the 5' end and subsequently purified by high-performance liquid chromatography. For the labeling process 5 mg of Cy3b at a 25 nMol scale and a 1:10 ratio of DNA:dye were used. Fractions containing dye labeled probe strands were pooled. Concentration of the probe strand stock was determined by using a UV/VIS spectrometer and calculated with 70 µM. Prior to the measurements, probe strands were diluted to 100 nM in hybridization buffer. Hybridization steps were as follows. Cells were washed twice with 200 µl 1x PBS. 200 µl of pre-cooled methanol (-20°C) was added to the center of each chamber. The chamber was incubated in the -20°C freezer for 30 min. After fixation of the cells, the methanol was aspirated from the samples. Subsequently, 100 µl of hybridization buffer (containing probe strands) was added to the center of each well, followed by an incubation step over night at -20°C. Next day, the hybridization buffer was aspirated and 200 µl of pre-warmed wash buffer added to the cells. Samples were washed twice with 200 µl of wash buffer and subsequently incubated at 37°C for 5 min after each washing step. The wash buffer was removed and the samples washed once with 2x SSC buffer. For nuclear staining, 200 µl of DAPI stain at a concentration of 300 nM (diluted in 1x PBS) was added to the cells, followed by an incubation for 5 min and subsequent rinsing twice with 200 µl PBS. Subsequently, 200 µl of GLOX buffer without enzymes was added for equilibration, and the cells were incubated for 2 min at 37°C. The GLOX buffer was removed and 100 µl of GLOX buffer containing glucose oxidase and catalase enzymes (ratio 1:100 – 1 µl of 100 x GLOX (100x stock concentration with enzymes)) were added to each well. Finally, samples were imaged via fluorescence microscopy.

Buffer solutions required for RNA FISH:

**Wash buffer:** 2x SSC buffer (Saline-sodium citrate) with 10% formamide.

**Hybridization buffer:** 2x SSC buffer with 10% dextran sulfate and 10% formamide (stored in 4°C). For 10 ml volume: 1 ml SSC, 2 ml dextran, 1 ml deionized formamide, 6 ml $H_2O$.

**Antibleach buffer**: 10% glucose RNase free, 1M Tris HCl, pH 8.0 20x SSC, RNase free $H_2O$, glucose oxidase, 3.7 mg/ml, (Sigma), catalase, (Sigma) Trolox, 40 mM.

**GLOX buffer without enzymes**: For 10 ml of GLOX buffer stock solution: 8.5 ml nuclease $H_2O$, 1 ml 20x SSC, 400 µl of 10% glucose, 100 µl of 1M Tris and vortex.

**GLOX buffer with enzymes**: For 100 µl: 93 µl of GLOX buffer, 1 µl glucose oxidase, 1 µl catalase and 5 µl Trolox.

# Appendix

## Oligonucleotide Sequences for the immobilized DSD cascade

**Unmodified DNA Origami Staples** (All sequences are given in 5'-3' direction)

| Staple position | Sequence |
|---|---|
| 0[47]1[31] | AGAAAGGAACAACTAAAGGAATTCAAAAAAA |
| 0[79]1[63] | ACAACTTTCAACAGTTTCAGCGGATGTATCGG |
| 0[111]1[95] | TAAATGAATTTTCTGTATGGGATTAATTTCTT |
| 0[143]1[127] | TCTAAAGTTTTGTCGTCTTTCCAGCCGACAA |
| 0[175]0[144] | TCCACAGACAGCCCTCATAGTTAGCGTAACGA |
| 0[207]1[191] | TCACCAGTACAAACTACAACGCCTAGTACCAG |
| 0[239]1[223] | AGGAACCCATGTACCGTAACACTTGATATAA |
| 0[271]1[255] | CCACCCTCATTTTCAGGGATAGCAACCGTACT |
| 1[32]3[31] | AGGCTCCAGAGGCTTTGAGGACACGGGTAA |
| 1[64]3[71] | TTTATCAGGACAGCATCGGAACGACACCAACCTAAAACGA |
| 1[96]3[95] | AAACAGCTTTTTGCGGGATCGTCAACACTAAA |
| 1[128]3[127] | TGACAACTCGCTGAGGCTTGCATTATACCA |
| 1[160]2[144] | TTAGGATTGGCTGAGACTCCTCAATAACCGAT |
| 1[192]3[191] | GCGGATAACCTATTATTCTGAAACAGACGATT |
| 1[224]3[223] | GTATAGCAAACAGTTAATGCCCAATCCTCA |
| 1[256]3[263] | CAGGAGGTGGGGTCAGTGCCTTGAGTCTCTGAATTTACCG |
| 2[47]0[48] | ACGGCTACAAAAGGAGCCTTTAATGTGAGAAT |
| 2[79]0[80] | CAGCGAAACTTGCTTTCGAGGTGTTGCTAA |
| 2[143]1[159] | ATATTCGGAACCATCGCCCACGCAGAGAAGGA |
| 2[175]0[176] | TATTAAGAAGCGGGGTTTTGCTCGTAGCAT |
| 2[207]0[208] | TTTCGGAAGTGCCGTCGAGAGGGTGAGTTTCG |
| 2[239]0[240] | GCCCGTATCCGGAATAGGTGTATCAGCCCAAT |
| 2[271]0[272] | GTTTTAACTTAGTACCGCCACCCAGAGCCA |
| 3[32]5[31] | AATACGTTTGAAAGAGGACAGACTGACCTT |
| 3[96]5[95] | ACACTCATCCATGTTACTTAGCCGAAAGCTGC |
| 3[128]5[127] | AGCGCGATGATAAATTGTGTCGTGACGAGA |
| 3[160]4[144] | TTGACAGGCCACCACCAGAGCCGCGATTTGTA |
| 3[192]5[191] | GGCCTTGAAGAGCCACCACCCTCAGAAACCAT |
| 3[224]5[223] | TTAAAGCCAGAGCCGCCACCCTCGACAGAA |
| 4[47]2[48] | GACCAACTAATGCCACTACGAAGGGGGTAGCA |
| 4[63]6[56] | ATAAGGGAACCGGATATTCATTACGTCAGGACGTTGGGAA |
| 4[79]2[80] | GCGCAGACAAGAGGCAAAAGAATCCCTCAG |
| 4[111]2[112] | GACCTGCTCTTTGACCCCCAGCGAGGGAGTTA |
| 4[143]3[159] | TCATCGCCAACAAAGTACAACGGACGCCAGCA |
| 4[175]2[176] | CACCAGAAAGGTTGAGGCAGGTCATGAAAG |
| 4[207]2[208] | CCACCCTCTATTCACAAACAAATACCTGCCTA |
| 4[239]2[240] | GCCTCCCTCAGAATGGAAAGCGCAGTAACAGT |
| 4[255]6[248] | AGCCACCACTGTAGCGCGTTTTCAAGGGAGGGAAGGTAAA |
| 4[271]2[272] | AAATCACCTTCCAGTAAGCGTCAGTAATAA |
| 5[32]7[31] | CATCAAGTAAAACGAACTAACGAGTTGAGA |
| 5[96]7[95] | TCATTCAGATGCGATTTTAAGAACAGGCATAG |
| 5[128]7[127] | AACACCAAATTTCAACTTTAATCGTTTACC |

| | |
|---|---|
| 5[160]6[144] | GCAAGGCCTCACCAGTAGCACCATGGGCTTGA |
| 5[192]7[191] | CGATAGCATTGAGCCATTTGGGAACGTAGAAA |
| 5[224]7[223] | TCAAGTTTCATTAAAGGTGAATATAAAAGA |
| 6[47]4[48] | TACGTTAAAGTAATCTTGACAAGAACCGAACT |
| 6[79]4[80] | TTATACCACCAAATCAACGTAACGAACGAG |
| 6[143]5[159] | GATGGTTTGAACGAGTAGTAAATTTACCATTA |
| 6[175]4[176] | CAGCAAAAGGAAACGTCACCAATGAGCCGC |
| 6[207]4[208] | TCACCGACGCACCGTAATCAGTAGCAGAACCG |
| 6[239]4[240] | GAAATTATTGCCTTTAGCGTCAGACCGGAACC |
| 6[271]4[272] | ACCGATTGTCGGCATTTTCGGTCATAATCA |
| 7[32]9[31] | TTTAGGACAAATGCTTTAAACAATCAGGTC |
| 7[56]9[63] | ATGCAGATACATAACGGGAATCGTCATAAATAAAGCAAAG |
| 7[96]9[95] | TAAGAGCAAATGTTTAGACTGGATAGGAAGCC |
| 7[128]9[135] | AGACGACAAAGAAGTTTTGCCATAATTCGAGCTTCAA |
| 7[192]9[199] | ATACATACCGAGGAAACGCAATAAGAAGCGCATTAGACGG |
| 7[224]9[223] | AACGCAAAGATAGCCGAACAAACCCTGAAC |
| 7[248]9[255] | GTTTATTTTGTCACAATCTTACCGAAGCCCTTTAATATCA |
| 8[47]6[48] | ATCCCCCTATACCACATTCAACTAGAAAAATC |
| 8[79]6[80] | AATACTGCCCAAAAGGAATTACGTGGCTCA |
| 8[111]6[112] | AATAGTAAACACTATCATAACCCTCATTGTGA |
| 8[143]7[159] | CTTTTGCAGATAAAAACCAAAATAAAGACTCC |
| 8[175]6[176] | ATACCCAACAGTATGTTAGCAAATTAGAGC |
| 8[207]6[208] | AAGGAAACATAAAGGTGGCAACATTATCACCG |
| 8[239]6[240] | AAGTAAGCAGACACCACGGAATAATATTGACG |
| 8[271]6[272] | AATAGCTATCAATAGAAAATTCAACATTCA |
| 9[32]11[31] | TTTACCCCAACATGTTTTAAATTTCCATAT |
| 9[64]11[63] | CGGATTGCAGAGCTTAATTGCTGAAACGAGTA |
| 9[96]11[95] | CGAAAGACTTTGATAAGAGGTCATATTTCGCA |
| 9[160]10[144] | AGAGAGAAAAAAATGAAAATAGCAAGCAAACT |
| 9[224]11[223] | AAAGTCACAAAATAAACAGCCAGCGTTTTA |
| 9[256]11[255] | GAGAGATAGAGCGTCTTTCCAGAGGTTTTGAA |
| 10[47]8[48] | CTGTAGCTTGACTATTATAGTCAGTTCATTGA |
| 10[79]8[80] | GATGGCTTATCAAAAAGATTAAGAGCGTCC |
| 10[127]12[120] | TAGAGAGTTATTTTCATTTGGGGATAGTAGTAGCATTA |
| 10[143]9[159] | CCAACAGGAGCGAACCAGACCGGAGCCTTTAC |
| 10[175]8[176] | TTAACGTCTAACATAAAAACAGGTAACGGA |
| 10[191]12[184] | GAAACGATAGAAGGCTTATCCGGTCTCATCGAGAACAAGC |
| 10[207]8[208] | ATCCCAATGAGAATTAACTGAACAGTTACCAG |
| 10[239]8[240] | GCCAGTTAGAGGGTAATTGAGCGCTTTAAGAA |
| 10[271]8[272] | ACGCTAACACCCACAAGAATTGAAAATAGC |
| 11[32]13[31] | AACAGTTTTGTACCAAAAACATTTTATTTC |
| 11[64]13[63] | GATTTAGTCAATAAAGCCTCAGAGAACCCTCA |
| 11[96]13[95] | AATGGTCAACAGGCAAGGCAAAGAGTAATGTG |
| 11[224]13[223] | GCGAACCTCCAAGAACGGGTATGACAATAA |
| 11[256]13[255] | GCCTTAAACCAATCAATAATCGGCACGCGCCT |
| 12[47]10[48] | TAAATCGGGATTCCCAATTCTGCGATATAATG |
| 12[79]10[80] | AAATTAAGTTGACCATTAGATACTTTTGCG |
| 12[111]10[112] | TAAATCATATAACCTGTTTAGCTAACCTTTAA |
| 12[143]11[159] | TTCTACTACGCGAGCTGAAAAGGTTACCGCGC |

| | |
|---|---|
| 12[175]10[176] | TTTTATTTAAGCAAATCAGATATTTTTTGT |
| 12[207]10[208] | GTACCGCAATTCTAAGAACGCGAGTATTATTT |
| 12[239]10[240] | CTTATCATTCCCGACTTGCGGGAGCCTAATTT |
| 12[271]10[272] | TGTAGAAATCAAGATTAGTTGCTCTTACCA |
| 13[32]15[31] | AACGCAAAATCGATGAACGGTACCGGTTGA |
| 13[64]15[71] | TATATTTTGTCATTGCCTGAGAGTGGAAGATTGTATAAGC |
| 13[96]15[95] | TAGGTAAACTATTTTTGAGAGATCAAACGTTA |
| 13[120]15[127] | AAAGGCCGGAGACAGCTAGCTGATAAATTAATTTTTGT |
| 13[160]14[144] | GTAATAAGTTAGGCAGAGGCATTTATGATATT |
| 13[184]15[191] | GACAAAAGGTAAAGTAATCGCCATATTTAACAAAACTTTT |
| 13[224]15[223] | ACAACATGCCAACGCTCAACAGTCTTCTGA |
| 13[256]15[263] | GTTTATCAATATGCGTTATACAAACCGACCGTGTGATAAA |
| 14[143]13[159] | CAACCGTTTCAAATCACCATCAATTCGAGCCA |
| 14[271]12[272] | TTAGTATCACAATAGATAAGTCCACGAGCA |
| 15[32]17[31] | TAATCAGCGGATTGACCGTAATCGTAACCG |
| 15[96]17[95] | ATATTTTGGCTTTCATCAACATTATCCAGCCA |
| 15[128]17[127] | TAAATCAAAATAATTCGCGTCTCGGAAACC |
| 15[160]16[144] | ATCGCAAGTATGTAAATGCTGATGATAGGAAC |
| 15[192]17[191] | TCAAATATAACCTCCGGCTTAGGTAACAATTT |
| 15[224]17[223] | CCTAAATCAAAATCATAGGTCTAAACAGTA |
| 16[47]14[48] | ACAAACGGAAAAGCCCCAAAAACACTGGAGCA |
| 16[63]18[56] | CGGATTCTGACGACAGTATCGGCCGCAAGGCGATTAAGTT |
| 16[79]14[80] | GCGAGTAAAAATATTTAAATTGTTACAAAG |
| 16[111]14[112] | TGTAGCCATTAAAATTCGCATTAAATGCCGGA |
| 16[143]15[159] | GCCATCAAGCTCATTTTTTAACCACAAATCCA |
| 16[175]14[176] | TATAACTAACAAAGAACGCGAGAACGCCAA |
| 16[207]14[208] | ACCTTTTTATTTTAGTTAATTTCATAGGGCTT |
| 16[239]14[240] | GAATTTATTTAATGGTTTGAAATATTCTTACC |
| 16[255]18[248] | GAGAAGAGATAACCTTGCTTCTGTTCGGGAGAAACAATAA |
| 16[271]14[272] | CTTAGATTTAAGGCGTTAAATAAAGCCTGT |
| 17[32]19[31] | TGCATCTTTCCCAGTCACGACGGCCTGCAG |
| 17[96]19[95] | GCTTTCCGATTACGCCAGCTGGCGGCTGTTTC |
| 17[128]19[127] | AGGCAAAGGGAAGGGCGATCGGCAATTCCA |
| 17[160]18[144] | AGAAAACAAAGAAGATGATGAAACAGGCTGCG |
| 17[192]19[191] | CATTTGAAGGCGAATTATTCATTTTTGTTTGG |
| 17[224]19[223] | CATAAATCTTTGAATACCAAGTGTTAGAAC |
| 18[47]16[48] | CCAGGGTTGCCAGTTTGAGGGGACCCGTGGGA |
| 18[79]16[80] | GATGTGCTTCAGGAAGATCGCACAATGTGA |
| 18[143]17[159] | CAACTGTTGCGCCATTCGCCATTCAAACATCA |
| 18[175]16[176] | CTGAGCAAAAATTAATTACATTTTGGGTTA |
| 18[207]16[208] | CGCGCAGATTACCTTTTTTAATGGGAGAGACT |
| 18[239]16[240] | CCTGATTGCAATATATGTGAGTGATCAATAGT |
| 18[271]16[272] | CTTTTACAAAATCGTCGCTATTAGCGATAG |
| 19[32]21[31] | GTCGACTTCGGCCAACGCGCGGGGTTTTTC |
| 19[56]21[63] | TACCGAGCTCGAATTCGGGAAACCTGTCGTGCAGCTGATT |
| 19[96]21[95] | CTGTGTGATTGCGTTGCGCTCACTAGAGTTGC |
| 19[128]21[127] | CACAACAGGTGCCTAATGAGTGCCCAGCAG |
| 19[160]20[144] | GCAATTCACATATTCCTGATTATCAAAGTGTA |
| 19[192]21[191] | ATTATACTAAGAAACCACCAGAAGTCAACAGT |

| | |
|---|---|
| 19[224]21[223] | CTACCATAGTTTGAGTAACATTTAAAATAT |
| 19[248]21[255] | CGTAAAACAGAAATAAAAATCCTTTGCCCGAAAGATTAGA |
| 20[47]18[48] | TTAATGAACTAGAGGATCCCCGGGGGGTAACG |
| 20[79]18[80] | TTCCAGTCGTAATCATGGTCATAAAAGGGG |
| 20[111]18[112] | CACATTAAAATTGTTATCCGCTCATGCGGGCC |
| 20[143]19[159] | AAGCCTGGTACGAGCCGGAAGCATAGATGATG |
| 20[175]18[176] | ATTATCATTCAATATAATCCTGACAATTAC |
| 20[207]18[208] | GCGGAACATCTGAATAATGGAAGGTACAAAAT |
| 20[239]18[240] | ATTTTAAAATCAAAATTATTTGCACGGATTCG |
| 20[271]18[272] | CTCGTATTAGAAATTGCGTAGATACAGTAC |
| 21[32]23[31] | TTTTCACTCAAAGGGCGAAAAACCATCACC |
| 21[64]23[63] | GCCCTTCAGAGTCCACTATTAAAGGGTGCCGT |
| 21[96]23[95] | AGCAAGCGTAGGGTTGAGTGTTGTAGGGAGCC |
| 21[128]23[127] | GCGAAAAATCCCTTATAAATCAAGCCGGCG |
| 21[160]22[144] | TCAATATCGAACCTCAAATATCAATTCCGAAA |
| 21[192]23[191] | TGAAAGGAGCAAATGAAAAATCTAGAGATAGA |
| 21[224]23[223] | CTTTAGGGCCTGCAACAGTGCCAATACGTG |
| 21[256]23[255] | GCCGTCAAAAAACAGAGGTGAGGCCTATTAGT |
| 22[47]20[48] | CTCCAACGCAGTGAGACGGGCAACCAGCTGCA |
| 22[79]20[80] | TGGAACAACCGCCTGGCCCTGAGGCCCGCT |
| 22[143]21[159] | TCGGCAAATCCTGTTTGATGGTGGACCCTCAA |
| 22[175]20[176] | ACCTTGCTTGGTCAGTTGGCAAAGAGCGGA |
| 22[207]20[208] | AGCCAGCAATTGAGGAAGGTTATCATCATTTT |
| 22[239]20[240] | TTAACACCAGCACTAACAACTAATCGTTATTA |
| 22[271]20[272] | CAGAAGATTAGATAATACATTTGTCGACAA |
| 23[32]22[48] | CAAATCAAGTTTTTTGGGGTCGAAACGTGGA |
| 23[64]22[80] | AAAGCACTAAATCGGAACCCTAATCCAGTT |
| 23[96]22[112] | CCCGATTTAGAGCTTGACGGGGAAAAAGAATA |
| 23[128]23[159] | AACGTGGCGAGAAAGGAAGGGAAACCAGTAA |
| 23[160]22[176] | TAAAAGGGACATTCTGGCCAACAAAGCATC |
| 23[192]22[208] | ACCCTTCTGACCTGAAAGCGTAAGACGCTGAG |
| 23[224]22[240] | GCACAGACAATATTTTTGAATGGGGTCAGTA |
| 23[256]22[272] | CTTTAATGCGCGAACTGATAGCCCCACCAG |

## List of modified staples

The following list of staple represents the extended staples to which fluorophore modified strands were hybridized.

| Staple position | Sequence, Sequence domains are indicated above | Position on the origami map (Figure 39) |
|---|---|---|
| 14[111]12[112] | GAGGGTAGGATTCAAAAGGGTGAGACATCCAA<br>\| u \|<br>TTAATGGAATGGAATGAA | C12 (central staple) |
| 14[79]12[80] | GCTATCAGAAATGCAATGCCTGAATTAGCA<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | B12 (10.5 nm left) |
| 11[160]12[144] | CCAATAGCTCATCGTAGGAATCATGGCATCAA<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | D12 (10.5 nm right) |

| | | |
|---|---|---|
| 10[111]8[112] | TTGCTCCTTTCAAATATCGCGTTTGAGGGGGT<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | C8 (10.5 nm up) |
| 18[111]16[112] | TCTTCGCTGCACCGCTTCTGGTGCGGCCTTCC<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | C16 (10.5 nm down) |
| 7[160]8[144] | TTATTACGAAGAACTGGCATGATTGCGAGAGG<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | D8 (15 nm diagonal) |
| 14[47]12[48] | AACAAGAGGGATAAAAATTTTTAGCATAAAGC<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | A12 (21.5 nm left) |
| 14[175]12[176] | CATGTAATAGAATATAAAGTACCAAGCCGTT<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | E12 (21.5 nm right) |
| 6[111]4[112] | ATTACCTTTGAATAAGGCTTGCCCAAATCCGC<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | C4 (21.5 nm up) |
| 22[111]20[112] | GCCCGAGAGTCCACGCTGGTTTGCAGCTAACT<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | C20 (21.5 nm down) |
| 14[207]12[208] | AATTGAGAATTCTGTCCAGACGACTAAACCAA<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | F12 (32 nm right) |
| 2[111]0[112] | AAGGCCGCTGATACCGATAGTTGCGACGTTAG<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | C0 (32 nm up) |
| 14[239]12[240] | AGTATAAAGTTCAGCTAATGCAGATGTCTTTC<br>\| v \|<br>TTAGGGTAGGTAGTAGGG | G12 (42.5 nm right) |

## Dye modified oligonucleotides

| Name | Sequence, Sequence domains are indicated above |
|---|---|
| Quencher1, Q1 (BBQ-650) | \| a \| d \| c \|<br>BBQ-650-ATGTTGAGTTGAGGTGTAGGTAGGTAGT |
| Fluorophore 1, F1 (Atto 647N) | \| d* \| a* \| b* \| u* \|<br>Atto 647N-TCAACTCAACATCAGTCTGATAAGCTATTCATTCCATTCCATT |
| Quencher2, Q2 (BHQ-1) | \| c \|<br>GGTGTAGGTAGGTAGT-BHQ-1 |
| Fluorophore 2, F2 (Atto532) | \| c* \| d \| a* \| v* \|<br>Atto 532-ACTACCTACCTACACCTCAACTCAACATCCCTACTACCTACCCT |

## Additional unmodified oligonucleotides

| Name | Sequence, Sequence domains are indicated above |
|---|---|
| DNA-21, Input I | \| b \| a \|<br>TAGCTTATCAGACTGATGTTGA |
| RNA-21, Input I | \| b \| a \|<br>UAGCUUAUCAGACUGAUGUUGA |
| Competitor strand | \| d* \| a* \|<br>TCAACTCAACAT |

| | |
|---|---|
| Protection strand | \| c* \| z \|<br>CCCTACCTACTACCTACCTACACC |
| De-Protection strand | \| z* \| c \|<br>GGTGTAGGTAGGTAGTAGGTAGGG |

# Bacterial toehold switches as repressors of translation

## Switch sequences

| Name | Sequence |
|---|---|
| NOT_h073 | GCGCTAATACGACTCACTATAGGGAAGATAAAGAAAATGATAAGAGTGATGAAATGGTGTGATGT<br>AAGTACGGACGAGACTTACATCACACCATTTCATCCCTCTTATAAACAGAGGAGAATAAGAATGA<br>TGAAATGGAACCTGGCGGCAGCGCAA |
| NOT_h074 | GCGCTAATACGACTCACTATAGGGAAAGTGAAAGTAAGAATAAAGTGGAAAGAGACGAGAATAA<br>GAAGAGACATCATATCTTCTTATTCTCGTCTCTTTCTACTTTATAAACAGAGGAGAATAAAGATGA<br>AAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h075 | GCGCTAATACGACTCACTATAGGGAGAAATATGAAATGGATTGAAGGTGTGAAAGGGAGTAGTA<br>GGTTGATATTCACGCAACCTACTACTCCATTTCACAACTTCAATAAACAGAGGAGAATTGAAATG<br>GTGAAATGGAACCTGGCGGCAGCGCAA |
| NOT_h076 | GCGCTAATACGACTCACTATAGGGAATGAAATAATGGAGATAAGTAATAATGAGTCGGTGTGTGC<br>TGTTAAACAGATCAACAGCACACACCGTCTCATTATTACTTATAAACAGAGGAGAATAAGTATGA<br>ATGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h077 | GCGCTAATACGACTCACTATAGGGAAAGATAAAGAGAGTATAAAGAGTGATGGAGGGTATGATA<br>GTTGTACAATAAGAACAACTATCATACCTTCCATCAATCTTTATAAACAGAGGAGAATAAAGATG<br>GATGGAAGGAACCTGGCGGCAGCGCAA |
| NOT_h078 | GCGCTAATACGACTCACTATAGGGAAAGTGAAAGTGAAGATAAATGAAGTGAAGAGGTGTGTTG<br>CTGTAACAACACAGTACAGCGACACACCTCTTCACTTCATTTATAAACAGAGGAGAATAAATATG<br>GTGAAGAGGAACCTGGCGGCAGCGCAA |
| NOT_h079 | GCGCTAATACGACTCACTATAGGGAAAGTAAATGAAGTGATAGAAGATAGAAGTGCGATGTGTGT<br>TAGATTCAAACATTCTAACACACATCGTACTTCTATCTTCTATAAACAGAGGAGAATAGAAATGA<br>GAAGTACGAACCTGGCGGCAGCGCAA |
| NOT_h080 | GCGCTAATACGACTCACTATAGGGAGAGAAAATAGAGAAATAGATAGGAGGAAGTGGAGAGTGA<br>GAGAACTCTCAAATTTCTCTTACTCTCCACTTCCTCCTATCTATAAACAGAGGAGAATAGATATGA<br>GGAAGTGGAACCTGGCGGCAGCGCAA |
| NOT_h081 | GCGCTAATACGACTCACTATAGGGAAAGTAAGAAAGAGTATAAGAGAGATAGAAAGGTGTGCGC<br>GTATACTGACGAGATATACGAGCACACCTTTCTATCTCTCTTATAAACAGAGGAGAATAAGAATG<br>ATAGAAAGGAACCTGGCGGCAGCGCAA |
| NOT_h082 | GCGCTAATACGACTCACTATAGGGAGAAGTAAGAAGATAATAAAGGTTGATGAAGGGTGTAAGT<br>GTAGTATCTACACAACTACACTTACACCTTTCATCAACCTTTATAAACAGAGGAGAATAAAGATG<br>GATGAAAGGAACCTGGCGGCAGCGCAA |
| NOT_h083 | GCGCTAATACGACTCACTATAGGGAAAATAATGAGAAGAATAAAGTTGAAGATAAGGAGTTGTGT<br>AAAGATCATACTCCTTTACACAACTCCTTATCTTCTACTTTATAAACAGAGGAGAATAAAGATGAA<br>GATAAGGAACCTGGCGGCAGCGCAA |
| NOT_h084 | GCGCTAATACGACTCACTATAGGGAAAGTGAAAGTAAGTATAAAGTGTGATGAGTGGTGTGAATG<br>TTGTAACAAACATACAACAATCACACCTCTCATCATACTTTATAAACAGAGGAGAATAAAGATGG<br>ATGAGAGGAACCTGGCGGCAGCGCAA |
| NOT_h085 | GCGCTAATACGACTCACTATAGGGATGAGAAATAAGAAGATAATAGAAAGAAATGGCGGTGAAA<br>GAGTATCAGCCGTATACTCTATCACCGCTATTTCTTTCTATTATAAACAGAGGAGAATAATAATGA<br>GAAATAGCAACCTGGCGGCAGCGCAA |
| NOT_h086 | GCGCTAATACGACTCACTATAGGGAAGAAAGTAAGTGAAATAAAGTGTGAAGAGGCGTGTGATT<br>GTAGTCAAGAAACAACTACAATCACACGTCTCTTCAGACTTTATAAACAGAGGAGAATAAAGATG<br>GAAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h087 | GCGCTAATACGACTCACTATAGGGAAGAAAGTAAATGAGATAGAAGGAGATAAGAGGATGAGAG<br>TAAGAGCTTATTATTCTTACACTCATCCTCTTATCTTCTTCTATAAACAGAGGAGAATAGAAATGG<br>ATAAGAGGAACCTGGCGGCAGCGCAA |
| NOT_h088 | GCGCTAATACGACTCACTATAGGGAAAGTAAATGAAATGATAAAGTGTGTGAATGCGATGTGTGT<br>AGTTACTATTCACAACTACACACATCGTATTCACATACTTTATAAACAGAGGAGAATAAAGATGG<br>TGAATACGAACCTGGCGGCAGCGCAA |
| NOT_h089 | GCGCTAATACGACTCACTATAGGGAAAGATGAAATGAAGATAGAAGGTGGAAGTTGGAGTGTGT<br>GTAAGATTACGAAACTTACAAACACTCCTACTTCCACCTTCTATAAACAGAGGAGAATAGAAATG<br>GGAAGTAGGAACCTGGCGGCAGCGCAA |
| NOT_h090 | GCGCTAATACGACTCACTATAGGGAAGAAAGTAAATGAGATAAAGAATGAGAAGACGGATAGAG<br>AAGTTACAATAAATAACTTCACTATCCGTCTTCTCATTCTTTATAAACAGAGGAGAATAAAGATGG<br>AGAAGACGAACCTGGCGGCAGCGCAA |
| NOT_h091 | GCGCTAATACGACTCACTATAGGGATAAAGAATGAAGTGAATAGAGTGATAAAGGCGGTGATGA<br>TAAAGACTCAAAGGCTTTATTATCACCGTCTTTATCCCTCTATTAAACAGAGGAGAAATAGAATGA<br>TAAAGACGAACCTGGCGGCAGCGCAA |
| NOT_h092 | GCGCTAATACGACTCACTATAGGGAAGATGAAGAGAGATATAAGAGATGAAGAGAGCGTGTGTT<br>GTAGAGTTAATTTATCTACAACACACGCTCTCTTCAGCTCTTATAAACAGAGGAGAATAAGAATG<br>GAAGAGAGCAACCTGGCGGCAGCGCAA |
| NOT_h093 | GCGCTAATACGACTCACTATAGGGATGAAAGATGAAGTGATAGAAGATGAAGAGGCGATGATGT |

| | |
|---|---|
| | GTAGTTCAATCTTTACTACAAATCATCGTCTCTTCAACTTCTATAAACAGAGGAGAATAGAAATGG<br>AAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h094 | GCGCTAATACGACTCACTATAGGGAAAGAAGTAAGAAGAATAAAGTGTGAATGGGCGTGTGTAA<br>GTAAGAATACACTACTTACTAACACACGTCCATTCACACTTTATAAACAGAGGAGAATAAAGATG<br>GAATGGACGAACCTGGCGGCAGCGCAA |
| NOT_h095 | GCGCTAATACGACTCACTATAGGGAATGAAATAGAAAGAATAAGAAGTGAAAGTGCGATAGATA<br>AAGTTCACTTAATTAACTTTATCTATCGTACTTTCACTTCTTATAAACAGAGGAGAATAAGAATGG<br>AAAGTACGAACCTGGCGGCAGCGCAA |
| NOT_h096 | GCGCTAATACGACTCACTATAGGGAGAAGAATGAGAAGAATAAAGGATGAGAGAGCGGAGAAAT<br>GTGTAGACAAATTATACACACTTCTCCGTTCTCTCAGCCTTTATAAACAGAGGAGAATAAAGATGG<br>AGAGAACGAACCTGGCGGCAGCGCAA |
| NOT_h097 | GCGCTAATACGACTCACTATAGGGATGGATAAGAAGTGGATAAGATGGAAGATGACGAGAATAG<br>GTATGGACAATAAACATACCTATTCTCGTCATCTTCGATCTTATAAACAGAGGAGAATAAGAATG<br>AAGATGACGAACCTGGCGGCAGCGCAA |
| NOT_h098 | GCGCTAATACGACTCACTATAGGGAAGGTAAGAGAAGTGATAAATGTGAAGATAAGGTGTAGAT<br>GTGTAGGATATAAATACACAACTACACCTTATCTTCTCATTTATAAACAGAGGAGAATAAATATG<br>AAGATAAGGAACCTGGCGGCAGCGCAA |
| NOT_h099 | GCGCTAATACGACTCACTATAGGGAACTTGATTGGTGAAATTGGTTGTCTTTATTCGTATATTCGTT<br>TCGAAACAAATGAAACGAATATACGTATAAAGATAACCAATAAACAGAGGAGAATTGGTATGCTT<br>TATACGAACCTGGCGGCAGCGCAA |
| NOT_h100 | GCGCTAATACGACTCACTATAGGGAAAATAATGGGAGAGATAAGAGGGATAAAGAGGGTGAGAA<br>GTAGAAACTTAACCTCTACTACTCACCCTCTTTATCCCTCTTATAAACAGAGGAGAATAAGAATGA<br>TAAAGAGGAACCTGGCGGCAGCGCAA |
| NOT_h101 | GCGCTAATACGACTCACTATAGGGAAAGTGAATGAGATTATTATTGGAGGTATTTGGAGAGAATG<br>TTATACGGCACATATAACAATCTCTCCTAATACCTACAATAATAAACAGAGGAGAATTATTATGG<br>GTATTAGGAACCTGGCGGCAGCGCAA |
| NOT_h102 | GCGCTAATACGACTCACTATAGGGAGAGTAAGAAGATATATAGAAGTGCGAATGAGGTGTGTAA<br>GTAGTACAAATATCACTACTTACACACCTCATTCGCTCTTCTATAAACAGAGGAGAATAGAAATGC<br>GAATGAGGAACCTGGCGGCAGCGCAA |
| NOT_h103 | GCGCTAATACGACTCACTATAGGGAGAAGTGAGAAAGATATAGAGGGTGAATGGAGGAAGGTGT<br>GTAAGGAACAAGAACTTACATACCTTCCTCCATTCACCCTCTATAAACAGAGGAGAATAGAGATG<br>GAATGGAGGAACCTGGCGGCAGCGCAA |
| NOT_h104 | GCGCTAATACGACTCACTATAGGGATAAGAAAGAGTGGAATAGAGGTTGAGGAAGGCTTGAGTT<br>GTAGTACACACAGAACTACAACTCAAGCTTTCCTCAACCTCTATAAACAGAGGAGAATAGAGATG<br>GAGGAAAGCAACCTGGCGGCAGCGCAA |
| NOT_h105 | GCGCTAATACGACTCACTATAGGGAAGATGATGAAGAAGATTGGATTGAAGATAGGGTTGGTGAG<br>AGTATTTCAAATCTACTCTAACCAACCATATCTTCGATCCAATAAACAGAGGAGAATTGGAATGA<br>AGATATGGAACCTGGCGGCAGCGCAA |
| NOT_h106 | GCGCTAATACGACTCACTATAGGGATAAGAAGAGTAAGAATAGATGTGAAGAAGTGGTGTGATG<br>ATTTGGAACATAAACAAATCATCACACCACTTCTTCTCATCTATAAACAGAGGAGAATAGATATG<br>AAGAAGTGGAACCTGGCGGCAGCGCAA |
| NOT_h107 | GCGCTAATACGACTCACTATAGGGAAAGATGAAGAGGTAATAATGTGTGAAGAGTGCTTGTGTTG<br>TAGAACATTTAATTCTACAACACAAGCTCTCTTCATACATTATAAACAGAGGAGAATAATGATGG<br>AAGAGAGCAACCTGGCGGCAGCGCAA |
| NOT_h108 | GCGCTAATACGACTCACTATAGGGAAAATAATGAGTGGAATAAAGGGTAAAGAGTCGAGGTGTG<br>TAAGTGATACAAAGACTTACACACCTCGTCTCTTTAACCTTTATAAACAGAGGAGAATAAAGATG<br>AAAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h109 | GCGCTAATACGACTCACTATAGGGAGAATGAATAAGAAGATAAAGTGGAAAGAGGCGGTGAAAG<br>AAGTACCTTATAACTACTTCATTCACCGTCTCTTTCCACTTTATAAACAGAGGAGAATAAAGATGA<br>AAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h110 | GCGCTAATACGACTCACTATAGGGTTAAGACGCCTGAGAATAGAAAGAAATAAGTCGGAAGAAA<br>GTAGTGATCAATATACTACTTTCTTCCGTCTTATTTCTTTCTATAAACAGAGGAGAATAGAAATGA<br>ATAAGACGAACCTGGCGGCAGCGCAA |
| NOT_h111 | GCGCTAATACGACTCACTATAGGGAAAGATAAGAAAGTGATAAAGTGTGAAGAAGGCGTGTGTA<br>GTATAAGAACACAATATACTACACACGCTTTCTTCAAACTTTATAAACAGAGGAGAATAAAGATG<br>GAAGAAAGCAACCTGGCGGCAGCGCAA |
| NOT_h112 | GCGCTAATACGACTCACTATAGGGAGAATAAGAGAGATAATAAGAAAGAGTGAATGGGTATGTT<br>GTGTAAACGGCAACTACACAACATACCCTTTCACTCCTTCTTATAAACAGAGGAGAATAAGAATG<br>AGTGAAAGGAACCTGGCGGCAGCGCAA |
| NOT_h113 | GCGCTAATACGACTCACTATAGGGGATTTGAAAGTGAATATTGGCCCTGGACGGTCGATTGTTGTA<br>GATACTTCTTCAATCTACTACAATCGTCCGTCCAAGGCCAATAAACAGAGGAGAATTGGCATGGG<br>ACGGACGAACCTGGCGGCAGCGCAA |
| NOT_h114 | GCGCTAATACGACTCACTATAGGGGTAAAGATGAGTGTGATAAGAGGTAGAGAGGCGTGTGTGA<br>GTGTAGACATAAGATACACTAACACACGTCTCTCTAACTCTTATAAACAGAGGAGAATAAGAATG<br>AGAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h115 | GCGCTAATACGACTCACTATAGGGAGAAAGATAAGAAAGATAAAGAGTGGAATGAGGTGTGTGA<br>TAGTACAAACTCTTTACTATAACACACCTCATTCCATTCTTTATAAACAGAGGAGAATAAAGATGG<br>GAATGAGGAACCTGGCGGCAGCGCAA |
| NOT_h116 | GCGCTAATACGACTCACTATAGGGAAAGTAAGAAGTGAAATAAAGAATGAAGAAGGCGAAGTGT<br>GTAATCACAAATCTATTACATACTTCGCTTTCTTCACTCTTTATAAACAGAGGAGAATAAAGATGG<br>AAGAAAGCAACCTGGCGGCAGCGCAA |
| NOT_h117 | GCGCTAATACGACTCACTATAGGGATTAATGGGATGTGGATTGTAGTTAAGAAGAGGTGCGTAGC<br>TAGTTCTCAACATACTAGCCACGCACCACTTCTTATCTACAATAAACAGAGGAGAATTGTAATGAA<br>GAAGTGGAACCTGGCGGCAGCGCAA |
| NOT_h118 | GCGCTAATACGACTCACTATAGGGAGAGAAGATAAGAAGATAGAACGGAATAAGTGCGAAAGTA<br>GAAATAACATCAATATTTCTACTTTCGCTCTTATTCTGTTCTATAAACAGAGGAGAATAGAAATGA |

| | |
|---|---|
| | ATAAGAGCAACCTGGCGGCAGCGCAA |
| NOT_h119 | GCGCTAATACGACTCACTATAGGGAAGATGAAAGATAGAATAGAATGTAGAAGATGCGAAGTTG<br>ATGTTCAACAATTCAACATCAACTTCGCTTCTTCTAGATTCTATAAACAGAGGAGAATAGAAATGA<br>GAAGAAGCAACCTGGCGGCAGCGCAA |
| NOT_h120 | GCGCTAATACGACTCACTATAGGGATAAGAAAGATGGAGATTAGATGGAGAAAGGCCGGTTTATG<br>TAGATTCATCACATCTACAGAAACCGGTCTTTCTCGATCTAATAAACAGAGGAGAATTAGAATGA<br>GAAAGACCAACCTGGCGGCAGCGCAA |
| NOT_h121 | GCGCTAATACGACTCACTATAGGGAAGATGATAAGTGAAATAGAAGGTAGAGAGGCGTGTGTTG<br>AAAGTTTAATATCTACTTTCTACACACGTCTCTCTAACTTCTATAAACAGAGGAGAATAGAAATGA<br>GAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h122 | GCGCTAATACGACTCACTATAGGGAGAATAAGAGGAAATATAAAGAGTGAGAGTGCGGGATTGG<br>TTGTAATCACAATATACAACAAATCCCGTACTCTCAATCTTTATAAACAGAGGAGAATAAAGATG<br>GAGAGTACGAACCTGGCGGCAGCGCAA |
| NOT_h123 | GCGCTAATACGACTCACTATAGGGAAAGATAATGGGAAGATAATAGGTGGTGAGTGGAGTAAAT<br>GTGAACCTAAATTCTTCACACTTACTCCTCTCACCAGCTATTATAAACAGAGGAGAATAATAATGG<br>GTGAGAGGAACCTGGCGGCAGCGCAA |
| NOT_h124 | GCGCTAATACGACTCACTATAGGGAAAGTGGAGAAGAATATAGAGGGTGAGAGTGGGTTGAGAT<br>GTTGTATTAACTTAACAACAACTCAACCTACTCTCAGCCTCTATAAACAGAGGAGAATAGAGATG<br>GAGAGTAGGAACCTGGCGGCAGCGCAA |
| NOT_h125 | GCGCTAATACGACTCACTATAGGGATAAAGATGGAGAAGATAAGATTGGATGAAGGGTTGAGGG<br>TATATATCGATCTTATATACACTCAACCTTTCATCCAATCTTATAAACAGAGGAGAATAAGAATGG<br>ATGAAAGGAACCTGGCGGCAGCGCAA |
| NOT_h126 | GCGCTAATACGACTCACTATAGGGGTAAGAGAAGTAGTGATAAGAGGAATGAGAAGGATGTTTGT<br>GTATACTCGAAATATACACAAACATCCATCTCATTGCTCTTATAAACAGAGGAGAATAAGAATGA<br>TGAGATGGAACCTGGCGGCAGCGCAA |
| NOT_h127 | GCGCTAATACGACTCACTATAGGGGTAAAGAAGATGGAGATAAAGAGTGAATGGAGGGTGAAGT<br>GTGTAATAAACTACTACACATTTCACCCTCCATTCATTCTTTATAAACAGAGGAGAATAAAGATGG<br>AATGGAGGAACCTGGCGGCAGCGCAA |
| NOT_h128 | GCGCTAATACGACTCACTATAGGGATGGAGAAAGATGATATTGAAGGAGAATGGAGGGTTTGGA<br>GTGTTACAAACACTAACACTTCAAACCCTCCATTCTTCTTCAATAAACAGAGGAGAATTGAAATGG<br>AATGGAGGAACCTGGCGGCAGCGCAA |
| NOT_h129 | GCGCTAATACGACTCACTATAGGGAGAGATGAAAGAGTAATAAAGTGGGAAAGTAGGGTAAGGT<br>GTATGACTAGAACACATACACCTTACCCTACTTTCCTACTTTATAAACAGAGGAGAATAAAGATG<br>GAAAGTAGGAACCTGGCGGCAGCGCAA |
| NOT_h130 | GCGCTAATACGACTCACTATAGGGAGAAATAAGAAGAGTAATGAAAGTGAGAAGTGGTGTGTGA<br>ATATAATACAAACTTATATTTACACACCACTTCTCACTTTCATTAAACAGAGGAGAAATGAAATGG<br>AGAAGTGGAACCTGGCGGCAGCGCAA |
| NOT_h131 | GCGCTAATACGACTCACTATAGGGAAAGATAAGAAGGTAATAAGAATGAATGATGGCTGTGTGTA<br>TATAACAGACAACTATATATACACAGCTATCATTCTTTCTTATAAACAGAGGAGAATAAGAATGA<br>ATGATAGCAACCTGGCGGCAGCGCAA |
| NOT_h132 | GCGCTAATACGACTCACTATAGGGAGAAATATGGATGAAATAAAGTGGGAAATAGGGTGTGTGAT<br>GAATAAACATTTGATTCATAACACACCTTATTTCCCACTTTATAAACAGAGGAGAATAAAGATGG<br>AAATAAGGAACCTGGCGGCAGCGCAA |
| NOT_h133 | GCGCTAATACGACTCACTATAGGGAGAAGAATAAGAGTGATAAAGTGTAGATGTGCGGTTGTTGA<br>TTTCTATGAAATTGAAATCTACAACCGTACATCTAAACTTTATAAACAGAGGAGAATAAAGATGA<br>GATGTACGAACCTGGCGGCAGCGCAA |
| NOT_h134 | GCGCTAATACGACTCACTATAGGGAGGAAGAGTAAGTGGATAGTGGAAGAAGTGTGGTATGTAA<br>GTTGTAACAAAGCGACAACTCACATACCTCACTTCTTCCACTATAAACAGAGGAGAATAGTGATG<br>GAAGTGAGGAACCTGGCGGCAGCGCAA |
| NOT_h135 | GCGCTAATACGACTCACTATAGGGAAAATAAGATGGAGAATAAAGGGAGATGGAAGGGTGAGAG<br>AGAAACTAACACAATTTCTCACTCACCCTTCCATCTTCCTTTATAAACAGAGGAGAATAAAGATGG<br>ATGGAAGGAACCTGGCGGCAGCGCAA |
| NOT_h136 | GCGCTAATACGACTCACTATAGGGATGAAAGATAAGAGAATAGATGAGAGAAGTGGGTTTGAAT<br>ATGATCTATCTAACATCATAATCAAACCTACTTCTCCCATCTATAAACAGAGGAGAATAGATATGA<br>GAAGTAGGAACCTGGCGGCAGCGCAA |
| NOT_h137 | GCGCTAATACGACTCACTATAGGGAGAAAGTGGAAAGATATAGAGTGGATGGAGAGGTGTGAGT<br>GTGTTAATACATACAACACATTCACACCTCTCCATCTACTCTATAAACAGAGGAGAATAGAGATG<br>ATGGAGAGGAACCTGGCGGCAGCGCAA |
| NOT_h138 | GCGCTAATACGACTCACTATAGGGAAAGAGTGGAGAAGTAATGAAGGTGAGATGTGGAGTGTGA<br>GTTGTGACATAAATACAACTAACACTCCTCATCTCATCTTCATTAAACAGAGGAGAAATGAAATG<br>GAGATGAGGAACCTGGCGGCAGCGCAA |
| NOT_h139 | GCGCTAATACGACTCACTATAGGGAGAAGTAAGAAAGTGATAGATGTTGAAGAGTGCTGTGTGTA<br>TTGTACAGAAAGAACAATAAACACAGCTCTCTTCATCATCTATAAACAGAGGAGAATAGATATGG<br>AAGAGAGCAACCTGGCGGCAGCGCAA |
| NOT_h140 | GCGCTAATACGACTCACTATAGGGAGAAGAGTAAGAAGAATAGAATGGAATAGAAGGTTGATGG<br>TAGTAAATTACAACTACTACAATCAACCTTCTATTCTATTCTATAAACAGAGGAGAATAGAAATGA<br>ATAGAAGGAACCTGGCGGCAGCGCAA |
| NOT_h141 | GCGCTAATACGACTCACTATAGGGAGAATGGAGAGAGTAATAAGATGGAAGATAAGGACAGAAT<br>GCAAAGTGTTTAAATTTGCATTCTGTCCTTATCTTCCATCTTATAAACAGAGGAGAATAAGAATGA<br>AGATAAGGAACCTGGCGGCAGCGCAA |
| NOT_h142 | GCGCTAATACGACTCACTATAGGGAGAGAATTGAGATATATTAGATGGAGGAGATCGTTTGTTAT<br>TAGTTCTCTACTTACTAATAACAAACGTTCTCCTCCATCTAATAAACAGAGGAGAATTAGAATGAG<br>GAGAACGAACCTGGCGGCAGCGCAA |
| NOT_h143 | GCGCTAATACGACTCACTATAGGGGTAACTAAGAATACTATAAAGGGCACGGAAAGGATTAGGA<br>AATTGAATTTAATACAATTTACTAATCCTTTCCGTGGCCTTTATAAACAGAGGAGAATAAAGATGA |

| | |
|---|---|
| | CGGAAAGGAACCTGGCGGCAGCGCAA |
| NOT_h144 | GCGCTAATACGACTCACTATAGGGAGAGAAAATAGAAAGAATGGAGGTGAATGGAGGTGTGTGA<br>TGTGTAATATAAAGACACATAACACACCTCCATTCAGCTCCATTAAACAGAGGAGAAATGGAATG<br>GAATGGAGGAACCTGGCGGCAGCGCAA |
| NOT_h145 | GCGCTAATACGACTCACTATAGGGAAGAATGGAGAGGTAATAAAGAATGAGAAGACGTGTGAGA<br>GGATTGAAATAAAGAATCCTATCACACGTCTTCTCATTCTTTATAAACAGAGGAGAATAAAGATG<br>GAGAAGACGAACCTGGCGGCAGCGCAA |
| NOT_h146 | GCGCTAATACGACTCACTATAGGGATGAAGAATGAAGTGATAAGAAGTAGAAGAACGTGTAAAT<br>GTAATTCTACAAACATTACATTTACACGTTCTTCTACTTCTTATAAACAGAGGAGAATAAGAATGA<br>GAAGAACGAACCTGGCGGCAGCGCAA |
| NOT_h147 | GCGCTAATACGACTCACTATAGGGATGGATAAAGATGAAATAAGAGGTGAAGATGGCTGTTGGTA<br>TGTTAATAAATCGAACATAACAACAGCTATCTTCATCTCTTATAAACAGAGGAGAATAAGAATGG<br>AAGATAGCAACCTGGCGGCAGCGCAA |
| NOT_h148 | GCGCTAATACGACTCACTATAGGGAGTAGAAGAAATGGAATAAAGTTTGAGATGAGGTTGTTGCT<br>TAGTGAAACGATGACTAAGAAACAACCTCATCTCACACTTTATAAACAGAGGAGAATAAAGATGG<br>AGATGAGGAACCTGGCGGCAGCGCAA |
| NOT_h149 | GCGCTAATACGACTCACTATAGGGAGAAGAGTAGTGAAGATTGAAATTGAAAGAGGCGAAGTTT<br>GTGATCAAATAACAATCACAAACTTCGCTTCTTTCAGTTTCAATAAACAGAGGAGAATTGAAATG<br>GAAAGAAGCAACCTGGCGGCAGCGCAA |
| NOT_h150 | GCGCTAATACGACTCACTATAGGGAGAATGAAAGAAGTGATAAGTGTGATGAAGTGCGTTGTTAT<br>GTATCGACTAAACATACATAACAACGCTCTTCATCTCACTTATAAACAGAGGAGAATAAGTATGA<br>TGAAGAGCAACCTGGCGGCAGCGCAA |
| NOT_h151 | GCGCTAATACGACTCACTATAGGGAGAAATAGTGGAAGAATAAGATTGGAGATGTGGAGGGTTG<br>GTATAGACGAACAATATACCAACCCTCCTCATCTCCCATCTTATAAACAGAGGAGAATAAGAATG<br>GAGATGAGGAACCTGGCGGCAGCGCAA |
| NOT_h152 | GCGCTAATACGACTCACTATAGGGAGAGATGATGGAAGAATAAAGAGTGAGATGAGGTTGTGTA<br>GTAAGGAGAAATTACTTACTACACAACCTCATCTCACTCTTTATAAACAGAGGAGAATAAAGATG<br>GAGATGAGGAACCTGGCGGCAGCGCAA |
| NOT_h153 | GCGCTAATACGACTCACTATAGGGAGAGATGAAGAGAATATTGGAGAGGAATGGAGGTTGATGA<br>GTTGTACTTCAGACACAACTAATCAACCTCCATTCCTCTCCAATAAACAGAGGAGAATTGGAATG<br>GAATGGAGGAACCTGGCGGCAGCGCAA |
| NOT_h154 | GCGCTAATACGACTCACTATAGGGAGAAGAATAAGGAAGATAAAGAATGAAGGAGGCGAAGGTA<br>TAGTTGACAACTTAAACTATACCTTCGCTTCCTTCAGTCTTTATAAACAGAGGAGAATAAAGATGG<br>AAGGAAGCAACCTGGCGGCAGCGCAA |
| NOT_h155 | GCGCTAATACGACTCACTATAGGGAAGAATGAAGAAGATATAGAAGAAGTAAGAAGGTGTGTAA<br>GTAATTACAAACACATTACTTACACACCTTCTTACTTCTTCTATAAACAGAGGAGAATAGAAATGG<br>TAAGAAGGAACCTGGCGGCAGCGCAA |
| NOT_h156 | GCGCTAATACGACTCACTATAGGGATGGAAAGTGATAAGATAAAGTGTAAAGAGTCGTGTGTTGT<br>AAGTCAATACAATACTTACAACACACGTCTCTTTATACTTTATAAACAGAGGAGAATAAAGATGA<br>AAGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h157 | GCGCTAATACGACTCACTATAGGGAGAATAAATGTGGAGATAATGTGTGAAAGATCGTGTGTTTA<br>TGTTACAAACACTAACATAAACACACGTTCTTTCAGACATTATAAACAGAGGAGAATAATGATGG<br>AAAGAACGAACCTGGCGGCAGCGCAA |
| NOT_h158 | GCGCTAATACGACTCACTATAGGGGTAAAGATAAGAGTGATAAGAGAGGATGAGTCGTTGTGTGT<br>AAGTTACTTAAACACTTACACACAACGTCTCATCCTCTCTTATAAACAGAGGAGAATAAGAATGG<br>ATGAGACGAACCTGGCGGCAGCGCAA |
| NOT_h159 | GCGCTAATACGACTCACTATAGGGAAGAGAGTAAAGAAGATAAGAGTGATAAAGTGCTGTGTTGT<br>AGTGCATATAAATCACTACAACACAGCTCTTTATCCCTCTTATAAACAGAGGAGAATAAGAATGA<br>TAAAGAGCAACCTGGCGGCAGCGCAA |
| NOT_h160 | GCGCTAATACGACTCACTATAGGGAATGAAAGAAAGGATATAAGGTGGGAAATGTGGGAGTGTA<br>AGAGACACTTCAATTCTCTTACACTCCCTCATTTCCGACCTTATAAACAGAGGAGAATAAGGATGG<br>AAATGAGGAACCTGGCGGCAGCGCAA |
| NOT_h161 | GCGCTAATACGACTCACTATAGGGAGAGAATGGAGAAAGATAGAAAGGATGGAAGGGATAGTGA<br>GAGATCTCATTCACATCTCTAACTATCCTTTCCATCCTTTCTATAAACAGAGGAGAATAGAAATGA<br>TGGAAAGGAACCTGGCGGCAGCGCAA |
| NOT_h162 | GCGCTAATACGACTCACTATAGGGAGAAAGGTAAAGAATATAGAAAGTGTAGAGTGGTGTGTAAT<br>TGTGCATAACTTTCACAATAACACACCTCTCTACACTTTCTATAAACAGAGGAGAATAGAAATGGT<br>AGAGAGGAACCTGGCGGCAGCGCAA |
| NOT_h163 | GCGCTAATACGACTCACTATAGGGAGCGCTTTAATTCATATATTTCAGACTAGTACCGTGTTTCTTT<br>GTGAATATGAGACAAAGAAACACGGTACTAGTCCGAAATATAAACAGAGGAGAATATTTATGACT<br>AGTACCAACCTGGCGGCAGCGCAA |
| NOT_h164 | GCGCTAATACGACTCACTATAGGGATAAGAATGAAGTGGAATGAAGATGAAAGAGGCAAGTGTG<br>TAGTTCTTCATATCAACTACACACTTGCTTCTTTCAACTTCATTAAACAGAGGAGAAATGAAATGG<br>AAAGAAGCAACCTGGCGGCAGCGCAA |
| NOT_h165 | GCGCTAATACGACTCACTATAGGGAGAAGTAAGAAAGTGATAAGATGTAGAAATGCCGTGATGTG<br>TTATCAGCAAGAAATAACAAATCACGGTATTTCTATATCTTATAAACAGAGGAGAATAAGAATGA<br>GAAATACCAACCTGGCGGCAGCGCAA |
| NOT_h166 | GCGCTAATACGACTCACTATAGGGAGAGAAATGAGAAGTAATGAAGGTGAGAAGGCGAGAAAGA<br>TAAAGGACTTTAGACTTTATATTTCTCGTCTTCTCACCTTCATTAAACAGAGGAGAAATGAAATGG<br>AGAAGACGAACCTGGCGGCAGCGCAA |
| NOT_h167 | GCGCTAATACGACTCACTATAGGGAAAGAATGGAAGAAGATAGAACTAAAGGATGGGATGTGGG<br>AATGTTATCTTTATACATTCTCACATCCTATCCTTTCGTTCTATAAACAGAGGAGAATAGAAATGA<br>AGGATAGGAACCTGGCGGCAGCGCAA |
| NOT_h168 | GCGCTAATACGACTCACTATAGGGAATAAAGATGGAGAAATAAAGAATAGAGAGACGGAAAGAA<br>TAGTAAATCCATACTACTATACTTTCCGTCTCTCTAGTCTTTATAAACAGAGGAGAATAAAGATGA |

| | |
|---|---|
| | GAGAGACGAACCTGGCGGCAGCGCAA |

## Trigger Sequences

| Name | Sequence |
|---|---|
| NOT_t073 | GCGCTAATACGACTCACTATAGGGAGTAGCCAAACAAGCGTTTGGCTACTATAACTTACATCAC ACCATTTCATCACTCTTATCATTTTCTTTATCTTCTTTAGCATAACCCCTTGGGGC |
| NOT_t074 | GCGCTAATACGACTCACTATAGGGCGATAGCTATACCTGGATAGCTATCGCATTCTTCTTATTCT CGTCTCTTTCCACTTTATTCTTACTTTCACTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t075 | GCGCTAATACGACTCACTATAGGGCGACCGACGGCTCGTCCCGTCGGTCGCGGCAACCTACTAC TCCCTTTCACACCTTCAATCCATTTCATATTTCTAATTAGCATAACCCCTTGGGGC |
| NOT_t076 | GCGCTAATACGACTCACTATAGGGCCAATGGACAACTATTTGTCCATTGGAAGAACAGCACAC ACCGACTCATTATTACTTATCTCCATTATTTCATTTATTAGCATAACCCCTTGGGGC |
| NOT_t077 | GCGCTAATACGACTCACTATAGGGCGTGTGGTCGGACCTACGACCACACGACAACAACTATCA TACCCTCCATCACTCTTTATACTCTCTTTATCTTTATTTAGCATAACCCCTTGGGGC |
| NOT_t078 | GCGCTAATACGACTCACTATAGGGTCCCTCAAGAAACAGATCTTGAGGGAATATACAGCAACA CACCTCTTCACTTCATTTATCTTCACTTTCACTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t079 | GCGCTAATACGACTCACTATAGGGACCCTATGCGACTAACCGCATAGGGTAAATCTAACACAC ATCGCACTTCTATCTTCTATCACTTCATTTACTTTATTTAGCATAACCCCTTGGGGC |
| NOT_t080 | GCGCTAATACGACTCACTATAGGGCCTAACCTTTGGAATCAAAGGTTAGGCAATTCTCTCACTC TCCACTTCCTCCTATCTATTTCTCTATTTTCTCTATTTAGCATAACCCCTTGGGGC |
| NOT_t081 | GCGCTAATACGACTCACTATAGGGTCCCGGAAACCGAATCGTTTCCGGGAATTTATACGCGCAC ACCTTTCTATCTCTCTTATACTCTTTCTTACTTTATTTAGCATAACCCCTTGGGGC |
| NOT_t082 | GCGCTAATACGACTCACTATAGGGCGAATTCAGTTATGACACTGAATTCGCGAACTACACTTAC ACCCTTCATCAACCTTTATTATCTTCTTACTTCTCTTTAGCATAACCCCTTGGGGC |
| NOT_t083 | GCGCTAATACGACTCACTATAGGGTGGCGCGCTAGAGGTGTAGCGCGCCAATACTTTACACAA CTCCTTATCTTCAACTTTATTCTTCTCATTATTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t084 | GCGCTAATACGACTCACTATAGGGCGATCATAACCAATAGGTTATGATCGATAACAACATTCAC ACCACTCATCACACTTTATACTTACTTTCACTTTCTTTAGCATAACCCCTTGGGGC |
| NOT_t085 | GCGCTAATACGACTCACTATAGGGAGACAAGCCAACTCAATGGCTTGTCTCATTACTCTTTCAC CGCCATTTCTTTCTATTATCTTCTTATTTCTCATCACTAGCATAACCCCTTGGGGC |
| NOT_t086 | GCGCTAATACGACTCACTATAGGGTCATCTTATCGCTCAGGATAAGATGAATAACTACAATCAC ACGCCTCTTCACACTTTATTTCACTTACTTTCTTTATTAGCATAACCCCTTGGGGC |
| NOT_t087 | GCGCTAATACGACTCACTATAGGGCGCGAGCTAGGATGGACTAGCTCGCGCATTCTTACTCTCA TCCTCTTATCTCCTTCTATCTCATTTACTTTCTTACTTAGCATAACCCCTTGGGGC |
| NOT_t088 | GCGCTAATACGACTCACTATAGGGAGGCGCTGAATACTCTTTCAGCGCCTATTAACTACACACA TCGCATTCACACACTTTATCATTTCATTTACTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t089 | GCGCTAATACGACTCACTATAGGGTCATTCACCAACACGATGGTGAATGACGACTTACACACA CTCCAACTTCCACCTTCTATCTTCATTTCATCTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t090 | GCGCTAATACGACTCACTATAGGGCCGGATACCAAATATATGGTATCCGGCTAAACTTCTCTAT CCGTCTTCTCATTCTTTATCTCATTTACTTTCTTCACTAGCATAACCCCTTGGGGC |
| NOT_t091 | GCGCTAATACGACTCACTATAGGGACGGTTAAGTGATGCGACTTAACCGTACACTTTATCATCA CCGCCTTTATCACTCTATTCACTTCATTCTTTATCACTAGCATAACCCCTTGGGGC |
| NOT_t092 | GCGCTAATACGACTCACTATAGGGTCAGTCGAGTTGTTGGACTCGACTGAATATCTACAACACA CGCTCTCTTCATCTCTTATATCTCTCTTCATCTTAATTAGCATAACCCCTTGGGGC |
| NOT_t093 | GCGCTAATACGACTCACTATAGGGTCACTAGTGGCTGACACCACTAGTGAATAACTACACATCA TCGCCTCTTCATCTTCTATCACTTCATCTTTCATATTTAGCATAACCCCTTGGGGC |
| NOT_t094 | GCGCTAATACGACTCACTATAGGGAGTGTGTAGCATAAATGCTACACACTCGACTTACTTACAC ACGCCCATTCACACTTTATTCTTCTTACTTCTTTATTTAGCATAACCCCTTGGGGC |
| NOT_t095 | GCGCTAATACGACTCACTATAGGGTCGCGTTTCCAGATTTGGAAACGCGATTTAACTTTATCTA TCGCACTTTCACTTCTTATTCTTTCTATTTCATTATTTAGCATAACCCCTTGGGGC |
| NOT_t096 | GCGCTAATACGACTCACTATAGGGCCTTAGCGAATTAAATTTCGCTAAGGCATTACACATTTCT CCGCTCTCTCATCCTTTATTCTTCTCATTCTTCTATTTAGCATAACCCCTTGGGGC |
| NOT_t097 | GCGCTAATACGACTCACTATAGGGCGGGAAGTCTCTAGGTAGACTTCCCGCAACATACCTATTC TCGTCATCTTCCATCTTATCCACTTCTTATCCATCTTTAGCATAACCCCTTGGGGC |
| NOT_t098 | GCGCTAATACGACTCACTATAGGGTCCAGGATAAGAATATTTATCCTGGAAATTACACATCTAC ACCTTATCTTCACATTTATCACTTCTCTTACCTTATTTAGCATAACCCCTTGGGGC |
| NOT_t099 | GCGCTAATACGACTCACTATAGGGCGTTAAACAGCGCATCCTGTTTAACGAAAGAAACGAATA TACGAATAAAGACAACCAATTTCACCAATCAAGTTATATAGCATAACCCCTTGGGGC |
| NOT_t100 | GCGCTAATACGACTCACTATAGGGCGAGAGATGTATACTAACATCTCTCGCAATCTACTTCTCA CCCTCTTTATCCCTCTTATCTCTCCCATTATTTTCACTAGCATAACCCCTTGGGGC |
| NOT_t101 | GCGCTAATACGACTCACTATAGGGTCCAGCTCCCATGAAAGGGAGCTGGACAAATAACATTCT CTCCAAATACCTCCAATAATAATCTCATTCACTTTAAATAGCATAACCCCTTGGGGC |
| NOT_t102 | GCGCTAATACGACTCACTATAGGGCCGTTTGTGCTCGAACGCACAAACGGAATACTACTTACAC ACCTCATTCGCACTTCTATATATCTTCTTACTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t103 | GCGCTAATACGACTCACTATAGGGCAGCCAGGTTTGTTAGAACCTGGCTGCGACTTACACACCT TCCTCCATTCACCCTCTATATCTTTCTCACTTCTAATTAGCATAACCCCTTGGGGC |
| NOT_t104 | GCGCTAATACGACTCACTATAGGGCCAAGCAACGTGTCGGCGTTGCTTGGCAAACTACAACTC AAGCCTTCCTCAACCTCTATTCCACTCTTTCTTATATTTAGCATAACCCCTTGGGGC |
| NOT_t105 | GCGCTAATACGACTCACTATAGGGCGACTCCCATTATGTCATGGGAGTCGCACTACTCTCACCA ACCCTATCTTCAATCCAATCTTCTTCATCATCTTAATTAGCATAACCCCTTGGGGC |
| NOT_t106 | GCGCTAATACGACTCACTATAGGGACAGACGGTAAGCGGTTACCGTCTGTAAACAAATCATCA CACCACTTCTTCACATCTATTCTTACTCTTCTTATATTTAGCATAACCCCTTGGGGC |

| | |
|---|---|
| NOT_t107 | GCGCTAATACGACTCACTATAGGGCCTCTCTGGCCAGGAGGCCAGAGAGGATATCTACAACAC<br>AAGCACTCTTCACACATTATTACCTCTTCATCTTTAACTAGCATAACCCCTTGGGGC |
| NOT_t108 | GCGCTAATACGACTCACTATAGGGCGGTTGCTTATGGGCGTAAGCAACCGCAAACTTACACAC<br>CTCGACTCTTTACCCTTTATTCCACTCATTATTTTCACTAGCATAACCCCTTGGGGC |
| NOT_t109 | GCGCTAATACGACTCACTATAGGGCCACAATTTCAGCATGGAAATTGTGGCATTACTTCTTTCA<br>CCGCCTCTTTCCACTTTATCTTCTTATTCATTCTCACTAGCATAACCCCTTGGGGC |
| NOT_t110 | GCGCTAATACGACTCACTATAGGGTGTGGCTATCTGGTGCGATAGCCACATATACTACTTTCTT<br>CCGACTTATTTCTTTCTATTCTCAGGCGTCTTAATTATAGCATAACCCCTTGGGGC |
| NOT_t111 | GCGCTAATACGACTCACTATAGGGTCTAGATCGGTCGCTCCCGATCTAGAAATTATACTACACA<br>CGCCTTCTTCACACTTTATCACTTTCTTATCTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t112 | GCGCTAATACGACTCACTATAGGGAGCCAAGCCATCTGAATGGCTTGGCTCAATACACAACAT<br>ACCCATTCACTCTTTCTTATTATCTCTCTTATTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t113 | GCGCTAATACGACTCACTATAGGGTGACCTGTCGAGGGCTCGACAGGTCATATATCTACAACA<br>ATCGACCGTCCAGGGCCAATATTCACTTTCAAATCACATAGCATAACCCCTTGGGGC |
| NOT_t114 | GCGCTAATACGACTCACTATAGGGCTTAGTGTGCTATGCAGCACACTAAGCATTACACTCACAC<br>ACGCCTCTCTACCTCTTATCACACTCATCTTTACATTTAGCATAACCCCTTGGGGC |
| NOT_t115 | GCGCTAATACGACTCACTATAGGGCGACTCAACCTGCTGAGGTTGAGTCGAACTACTATCACAC<br>ACCTCATTCCACTCTTTATCTTTCTTATCTTTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t116 | GCGCTAATACGACTCACTATAGGGTCAGTTCCTAGCATCTTAGGAACTGATATATTACACACTT<br>CGCCTTCTTCATTCTTTATTTCACTTCTTACTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t117 | GCGCTAATACGACTCACTATAGGGTGGCACTAGGCGTTACCCTAGTGCCATATACTAGCTACGC<br>ACCTCTTCTTAACTACAATCCACATCCCATTAATAAGTAGCATAACCCCTTGGGGC |
| NOT_t118 | GCGCTAATACGACTCACTATAGGGTCAGATGTGATGCCTCTCACATCTGAAACATTTCTACTTT<br>CGCACTTATTCCGTTCTATCTTCTTATCTTCTCTTATTAGCATAACCCCTTGGGGC |
| NOT_t119 | GCGCTAATACGACTCACTATAGGGTCTAAGAATAGGTTTATATTCTTAGAAATAACATCAACTT<br>CGCATCTTCTACATTCTATTCTATCTTTCATCTTACTTAGCATAACCCCTTGGGGC |
| NOT_t120 | GCGCTAATACGACTCACTATAGGGTGCGGTTTGGTGCCCTCCAAACCGCAATTTCTACATAAAC<br>CGGCCTTTCTCCATCTAATCTCCATCTTTCTTATAACTAGCATAACCCCTTGGGGC |
| NOT_t121 | GCGCTAATACGACTCACTATAGGGACTGGATGCAGGACGTTGCATCCAGTCATACTTTCAACAC<br>ACGCCTCTCTACCTTCTATTTCACTTATCATCTTTATTAGCATAACCCCTTGGGGC |
| NOT_t122 | GCGCTAATACGACTCACTATAGGGAGACCGGGAATCCACATTCCCGGTCTCCGTACAACCAATC<br>CCGCACTCTCACTCTTTATATTTCCTCTTATTCTTTATAGCATAACCCCTTGGGGC |
| NOT_t123 | GCGCTAATACGACTCACTATAGGGCGTACCCTTGCGTATCCAAGGGTACGCGATTCACATTTAC<br>TCCACTCACCACCTATTATCTTCCCATTATCTTTATTTAGCATAACCCCTTGGGGC |
| NOT_t124 | GCGCTAATACGACTCACTATAGGGCGTAGAGGTAACTATATACCTCTACGCAAACAACATCTCA<br>ACCCACTCTCACCCTCTATATTCTTCTCCACTTTAATTAGCATAACCCCTTGGGGC |
| NOT_t125 | GCGCTAATACGACTCACTATAGGGCGTTCTTGAGAAATAGCTCAAGAACGCAAATATACCCTC<br>AACCCTTCATCCAATCTTATCTTCTCCATCTTTATACTTAGCATAACCCCTTGGGGC |
| NOT_t126 | GCGCTAATACGACTCACTATAGGGTCTGGTCATCAATCATGATGACCAGACTAATACACAAAC<br>ATCCTTCTCATTCCTCTTATCACTACTTCTCTTACATTTAGCATAACCCCTTGGGGC |
| NOT_t127 | GCGCTAATACGACTCACTATAGGGCTAAGCTTATGGGTTAATAAGCTTAGCAATACACACTTCA<br>CCCTCCATTCACTCTTTATCTCCATCTTCTTTACTTATAGCATAACCCCTTGGGGC |
| NOT_t128 | GCGCTAATACGACTCACTATAGGGCGTAACTTGCACCTGGGCAAGTTACGCATAACACTCCAA<br>ACCCTCCATTCTCCTTCAATATCATCTTTCTCCATCACTAGCATAACCCCTTGGGGC |
| NOT_t129 | GCGCTAATACGACTCACTATAGGGCCGCCAGCAACGGGAATTGCTGGCGGCGACATACACCTT<br>ACCCTACTTTCCCACTTTATTACTCTTTCATCTCTTTATAGCATAACCCCTTGGGGC |
| NOT_t130 | GCGCTAATACGACTCACTATAGGGCGAACTAGTGTAGTTCCACTAGTTCGATATATATTCACAC<br>ACCACTTCTCACTTTCATTACTCTTCTTATTTCTTATTAGCATAACCCCTTGGGGC |
| NOT_t131 | GCGCTAATACGACTCACTATAGGGTCAGATCAATAACAAGATTGATCTGACAATATATACACA<br>CAGCCATCATTCATTCTTATTACCTTCTTATCTTTCACTAGCATAACCCCTTGGGGC |
| NOT_t132 | GCGCTAATACGACTCACTATAGGGCTGCACTGCACTCAATTGCAGTGCAGCAAATTCATCACAC<br>ACCCTATTTCCCACTTTATTTCATCCATATTTCTTATTAGCATAACCCCTTGGGGC |
| NOT_t133 | GCGCTAATACGACTCACTATAGGGAGGTATGTACTACCAAGTACATACCTAACGAAATCAACA<br>ACCGCACATCTACACTTTATCACTCTTATTCTTCTTTATAGCATAACCCCTTGGGGC |
| NOT_t134 | GCGCTAATACGACTCACTATAGGGCGGTACCCGTACGTGAACGGGTACCGCAAACAACTTACA<br>TACCACACTTCTTCCACTATCCACTTACTCTTCCTAATTAGCATAACCCCTTGGGGC |
| NOT_t135 | GCGCTAATACGACTCACTATAGGGCGATTCTCTGTTAGGTCAGAGAATCGCAATTTCTCTCTCA<br>CCCTTCCATCTCCCTTTATTCTCCATCTTATTTTACTTAGCATAACCCCTTGGGGC |
| NOT_t136 | GCGCTAATACGACTCACTATAGGGACGGGCCTGTAATTCAACAGGCCCGTCTTATCATATTCAA<br>ACCCACTTCTCTCATCTATTCTCTTATCTTTCATCGCTAGCATAACCCCTTGGGGC |
| NOT_t137 | GCGCTAATACGACTCACTATAGGGCGACAGTGCGATAGGACGCACTGTCGCATAACACACTCA<br>CACCTCTCCATCCACTCTATATCTTTCCACTTTCTAATTAGCATAACCCCTTGGGGC |
| NOT_t138 | GCGCTAATACGACTCACTATAGGGCCGCAAGCTCAAATCAGAGCTTGCGGCAAACAACTCACA<br>CTCCACATCTCACCTTCATTACTTCTCCACTCTTTATATAGCATAACCCCTTGGGGC |
| NOT_t139 | GCGCTAATACGACTCACTATAGGGCCTGGGTCTATGCGGCTAGACCCAGGATTACAATACACA<br>CAGCACTCTTCAACATCTATCACTTTCTTACTTCTAATTAGCATAACCCCTTGGGGC |
| NOT_t140 | GCGCTAATACGACTCACTATAGGGCGCACGAGGATAGCAATCCTCGTGCGAACTACTACCATC<br>AACCTTCTATTCCATTCTATTCTTCTTACTCTTCTTATTAGCATAACCCCTTGGGGC |
| NOT_t141 | GCGCTAATACGACTCACTATAGGGTTGAAGCTTTCCACGGAAAGCTTCAACTGTTTGCATTCTG<br>TCCTTATCTTCCATCTTATTACTCTCTCCATTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t142 | GCGCTAATACGACTCACTATAGGGTGCCCGCGCATAAGCGTGCGCGGGCAATAACTAATAACA<br>AACGATCTCCTCCATCTAATATATCTCAATTCTCTAACTAGCATAACCCCTTGGGGC |
| NOT_t143 | GCGCTAATACGACTCACTATAGGGAGTCGCCAATCCGCACATTGGCGACTTGGCAATTTCCTAA<br>TCCTTTCCGTGCCCTTTATAGTATTCTTAGTTACACATAGCATAACCCCTTGGGGC |
| NOT_t144 | GCGCTAATACGACTCACTATAGGGCCCGTGGTGTTGTCTTACACCACGGGCAAACACATCACAC |

| | |
|---|---|
| | ACCTCCATTCACCTCCATTCTTTCTATTTTCTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t145 | GCGCTAATACGACTCACTATAGGGCGCCATTAACACCACTGTTAATGGCGCTAAATCCTCTCAC<br>ACGTCTTCTCATTCTTTATTACCTCTCCATTCTTACTTAGCATAACCCCTTGGGGC |
| NOT_t146 | GCGCTAATACGACTCACTATAGGGTCCGTCATTATTTCTATAATGACGGAATTATTACATTTACA<br>CGTTCTTCTACTTCTTATCACTTCATTCTTCATTTATAGCATAACCCCTTGGGGC |
| NOT_t147 | GCGCTAATACGACTCACTATAGGGAGTCCCTGGTTTAATAACCAGGGACTCATAACATACCAAC<br>AGCCATCTTCACCTCTTATTTCATCTTTATCCATATTTAGCATAACCCCTTGGGGC |
| NOT_t148 | GCGCTAATACGACTCACTATAGGGCGGTGTGTTCGGTGCCGAACACACCGATAACTAAGCAAC<br>AACCTCATCTCAAACTTTATTCCATTTCTTCTACTAATTAGCATAACCCCTTGGGGC |
| NOT_t149 | GCGCTAATACGACTCACTATAGGGACGAGGCGCAGACTTGTGCGCCTCGTCATATCACAAACTT<br>CGCCTCTTTCAATTTCAATCTTCACTACTCTTCTTTATAGCATAACCCCTTGGGGC |
| NOT_t150 | GCGCTAATACGACTCACTATAGGGTCCCTAGGAAACCGTCTTCCTAGGGATAAATACATAACA<br>ACGCACTTCATCACACTTATCACTTCTTTCATTCTATATAGCATAACCCCTTGGGGC |
| NOT_t151 | GCGCTAATACGACTCACTATAGGGCCAAGTAGAGGGCGTGCTCTACTTGGCAATATACCAACC<br>CTCCACATCTCCAATCTTATTCTTCCACTATTTCTAACTAGCATAACCCCTTGGGGC |
| NOT_t152 | GCGCTAATACGACTCACTATAGGGCGTGGAGATCTTATAAGATCTCCACGAAACTTACTACACA<br>ACCTCATCTCACTCTTTATTCTTCCATCATCTCTTATTAGCATAACCCCTTGGGGC |
| NOT_t153 | GCGCTAATACGACTCACTATAGGGTCGCAACTCATCATGCTGAGTTGCGACAAACAACTCATCA<br>ACCTCCATTCCTCTCCAATATTCTCTTCATCTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t154 | GCGCTAATACGACTCACTATAGGGTCGGTGTAACACGAACGTTACACCGACATAACTATACCTT<br>CGCCTCCTTCATTCTTTATCTTCCTTATTCTTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t155 | GCGCTAATACGACTCACTATAGGGTCCAGTCCTTAAGAACAAGGACTGGAATTATTACTTACAC<br>ACCTTCTTACTTCTTCTATATCTTCTTCATTCTTACTTAGCATAACCCCTTGGGGC |
| NOT_t156 | GCGCTAATACGACTCACTATAGGGCCAGCGCCTAGACACATAGGCGCTGGACAACTTACAACA<br>CACGACTCTTTACACTTTATCTTATCACTTTCCATTATTAGCATAACCCCTTGGGGC |
| NOT_t157 | GCGCTAATACGACTCACTATAGGGCGCGGTCCGAATCCTTTCGGACCGCGGACAACATAAACA<br>CACGATCTTTCACACATTATCTCCACATTTATTCTAATTAGCATAACCCCTTGGGGC |
| NOT_t158 | GCGCTAATACGACTCACTATAGGGCGGAGCAAGGCATATTCCTTGCTCCGAATACTTACACACA<br>ACGACTCATCCTCTCTTATCACTCTTATCTTTACATTTAGCATAACCCCTTGGGGC |
| NOT_t159 | GCGCTAATACGACTCACTATAGGGCGGAACTCTTGCGACTAAGAGTTCCGAAACACTACAACA<br>CAGCACTTTATCACTCTTATCTTCTTTACTCTCTTACTTAGCATAACCCCTTGGGGC |
| NOT_t160 | GCGCTAATACGACTCACTATAGGGCCGAGCCGCCGTTCAGGGCGGCTCGGCAATCTCTTACACT<br>CCCACATTTCCCACCTTATATCCTTTCTTTCATTACTTAGCATAACCCCTTGGGGC |
| NOT_t161 | GCGCTAATACGACTCACTATAGGGCCTGATAGGCGCGCTGGCCTATCAGGCGTATCTCTCACTA<br>TCCCTTCCATCCTTTCTATCTTTCTCCATTCTCTTATTAGCATAACCCCTTGGGGC |
| NOT_t162 | GCGCTAATACGACTCACTATAGGGCCTAGCTATAGGATGATATAGCTAGGAAGCACAATTACA<br>CACCACTCTACACTTTCTATATTCTTTACCTTTCTACTTAGCATAACCCCTTGGGGC |
| NOT_t163 | GCGCTAATACGACTCACTATAGGGAGTACCACAGCGTTATCTGTGGTACTGAAACAAAGAAAC<br>ACGGTACTAGTCTGAAATATATGAATTAAAGCGCTAAATAGCATAACCCCTTGGGGC |
| NOT_t164 | GCGCTAATACGACTCACTATAGGGTCATTTAGATTAACAAATCTAAATGATTTAACTACACACT<br>TGCCTCTTTCATCTTCATTCCACTTCATTCTTATTATTAGCATAACCCCTTGGGGC |
| NOT_t165 | GCGCTAATACGACTCACTATAGGGTCTGCTATTTACACTCAAATAGCAGAATTATAACACATCA<br>CGGCATTTCTACATCTTATCACTTTCTTACTTCTTATTAGCATAACCCCTTGGGGC |
| NOT_t166 | GCGCTAATACGACTCACTATAGGGCTTCCAGGTAGCCGTTTACCTGGAAGCAACTTTATCTTTC<br>TCGCCTTCTCACCTTCATTACTTCTCATTTCTCTATATAGCATAACCCCTTGGGGC |
| NOT_t167 | GCGCTAATACGACTCACTATAGGGCGATTTCACTCCGACAAGTGAAATCGCATACATTCCCACA<br>TCCCATCCTTTAGTTCTATCTTCTTCCATTCTTTATTTAGCATAACCCCTTGGGGC |
| NOT_t168 | GCGCTAATACGACTCACTATAGGGCGTGTCAGTCCCAGCTGACTGACACGCTATACTATTCTTT<br>CCGTCTCTCTATTCTTTATTTCTCCATCTTTATTACTTAGCATAACCCCTTGGGGC |

## Sequences for PCR

| Name | Sequence |
|---|---|
| Trigger Insert PCR | |
| Fwd Primer | AACGTGTACGGGCTATCTGGCTTTCGTTGCGCTAATACGACTCACTATAGGG |
| Rev Primer | CCCGTTTAGAGGCCCCAA |

| Name | Sequence |
|---|---|
| Switch Insert PCR | |
| Fwd Primer | GAAGTCTAACGCTGCTCTGGGCTAACTGTCGCGCTAATACGACTCACTATAGGG |
| Rev Primer | TTTACGCATCTTTTGCGCTGC |

| Name | Sequence |
|---|---|
| Trigger Backbone PCR | |
| Fwd Primer | TAGCATAACCCCTTGGGGC |
| Rev Primer | GCAACGAAAGCCAGATAGCCCGTACACGTTCATAAGGGAGAGCGTCGAGATC |

| Name | Sequence |
|---|---|
| Switch Backbone PCR | |
| Fwd Primer | AACCTGGCGGCAGCGCAAAAGATGCGTAAAGGAGAAGAACTTTTCACT |
| Rev Primer | GACAGTTAGCCCAGAGCAGCGTTAGACTTCATAAGGGAGAGCGTCGAGATC |

## Riboregulators in mammalian cells

## Switch sequences

| Switch | Pre-Kozak hairpin sequence | Minimal hairpin sequence | Post-Kozak hairpin sequence |
|---|---|---|---|
| Switch 1 | TACGCTTACGACAAGTTTTCAGATTTCAAC CGCACTTCAACCTACTCCGCCACCATGGGA GTAGGTAGAAGTGCGCAATTAGTTAGTAAC CTGGCTGCAGCTCAGATGGTTTCTAAAGGA GAAGAAGATAACATGGCCATCATCAAGGA GTTCAT | AACCGCACTTC AACCTACTCCG CCACCATGGG AGTAGGTAGA AGTGCGCAA | TACCGGTCGCCACCATGGAATACGC TTACGACAAGTTTTCAGATTTCAAC CGCACTTCAACCTACTCCGCCACCA TGGGAGTAGGTAGAAGTGCGCAAT TAGTTAGTAACCTGGCTGCAGCTCA GATGGTTTCTAAAGGAGAAGAAGA TAACATGGCCATCATCAAGGAGTTC AT |
| Switch 2 | TACGCTTACGACAAGTTTTCAGATTTCAAC CGTACCCTCACCCGCTCCATCCGCCACCAT GGGATGGAGCTGGTTAGGGTACGATCAATT AGTTAGTAACCTGGCTGCAGCTCAGATGGT TTCTAAAGGAGAAGAAGATAACATGGCCAT CATCAAGGAGTTCAT | AACCGTACCCT CACCCGCTCCA TCCGCCACCAT GGGATGGAGC TGGTTAGGGTA CGATC | TACCGGTCGCCACCATGGATACGCT TACGACAAGTTTTCAGATTTCAACC GTACCCTCACCCGCTCCATCCGCCA CCATGGGATGGAGCTGGTTAGGGTA CGATCAATTAGTTAGTAACCTGGCT GCAGCTCAGATGGTTTCTAAAGGAG AAGAAGATAACATGGCCATCATCA AGGAGTTCAT |
| Switch 3 | TACGCTTACGACAAGTTTTCAGATTTCAAC GCCTATACATACTACCGACCTTACCGCCAC CATGGGTAAGGTCGGTAGTAAGTATAGGCA TCAATTAGTTAGTAACCTGGCTGCAGCTCA GATGGTTTCTAAAGGAGAAGAAGATAACAT GGCCATCATCAAGGAGTTCAT | AACGCCTATAC ATACTACCGAC CTTACCGCCAC CATGGGTAAG GTCGGTAGTA AGTATAGGCA TC | TACCGGTCGCCACCATGGATACGCT TACGACAAGTTTTCAGATTTCAACG CCTATACATACTACCGACCTTACCG CCACCATGGGTAAGGTCGGTAGTAA GTATAGGCATCAATTAGTTAGTAAC CTGGCTGCAGCTCAGATGGTTTCTA AAGGAGAAGAAGATAACATGGCCA TCATCAAGGAGTTCAT |
| Switch 4 | TACGCTTACGACAAGTTTTCAGATTTCAAC CGCCACATACTCCAGACAAATCCACCTCGC CACCATGGAGGTGGATGTGTCGGGAGTATG TGGCGATCAATTAGTTAGTAACCTGGCTGC AGCTCAGATGGTTTCTAAAGGAGAAGAAG ATAACATGGCCATCATCAAGGAGTTCAT | AACCGCCACA TACTCCAGACA AATCCACCTCG CCACCATGGA GGTGGATGTGT CGGGAGTATG TGGCGATC | TACCGGTCGCCACCATGGATACGCT TACGACAAGTTTTCAGATTTCAACC GCCACATACTCCAGACAAATCCACC TCGCCACCATGGAGGTGGATGTGTC GGGAGTATGTGGCGATCAATTAGTT AGTAACCTGGCTGCAGCTCAGATGG TTTCTAAAGGAGAAGAAGATAACA TGGCCATCATCAAGGAGTTCAT |
| Switch 5 | TACGCTTACGACAAGTTTTCAGATTTCAAC GCCAAATCACCAGCCACCTACGAGATCTAC CGCCACCATGGGTAGATCTGGTAGGTGGCT GGAGATTTGGCATCAATTAGTTAGTAACCT GGCTGCAGCTCAGATGGTTTCTAAAGGAGA AGAAGATAACATGGCCATCATCAAGGAGTT CAT | AACGCCAAAT CACCAGCCAC CTACGAGATCT ACCGCCACCAT GGGTAGATCT GGTAGGTGGC TGGAGATTTGG CATC | TACCGGTCGCCACCATGGATACGCT TACGACAAGTTTTCAGATTTCAACG CCAAATCACCAGCCACCTACGAGAT CTACCGCCACCATGGGTAGATCTGG TAGGTGGCTGGAGATTTGGCATCAA TTAGTTAGTAACCTGGCTGCAGCTC AGATGGTTTCTAAAGGAGAAGAAG ATAACATGGCCATCATCAAGGAGTT CAT |
| Switch 6 | TACGCTTACGACAAGTTTTCAGATTTCAAC CGCCACTCATACACGCTCATACATCCGCAA CCTCGCCACCATGGAGGTTGCGTATGTATG GGCGTGTAGGAGTGGCGATCAATTAGTTAG TAACCTGGCTGCAGCTCAGATGGTTTCTAA AGGAGAAGAAGATAACATGGCCATCATCA AGGAGTTCAT | AACCGCCACTC ATACACGCTCA TACATCCGCAA CCTCGCCACCA TGGAGGTTGC GTATGTATGGG CGTGTAGGAG TGGCGATC | TACCGGTCGCCACCATGGATACGCT TACGACAAGTTTTCAGATTTCAACC GCCACTCATACACGCTCATACATCC GCAACCTCGCCACCATGGAGGTTGC GTATGTATGGGCGTGTAGGAGTGGC GATCAATTAGTTAGTAACCTGGCTG CAGCTCAGATGGTTTCTAAAGGAGA AGAAGATAACATGGCCATCATCAA GGAGTTCAT |
| Switch 7 | TACGCTTACGACAAGTTTTCAGATTTCAAC GCACCCTTACACGACCTACGACATAGATCC CGATCACGCCACCATGGTGATCGGGTTCTA TGTCGTAGTTCGTGTAAGGGTGCATCAATT AGTTAGTAACCTGGCTGCAGCTCAGATGGT TTCTAAAGGAGAAGAAGATAACATGGCCAT CATCAAGGAGTTCAT | AACGCACCCTT ACACGACCTA CGACATAGAT CCCGATCACGC CACCATGGTG ATCGGGTTCTA TGTCGTAGTTC GTGTAAGGGT GCATC | TACCGGTCGCCACCATGGATACGCT TACGACAAGTTTTCAGATTTCAACG CACCCTTACACGACCTACGACATAG ATCCCGATCACGCCACCATGGTGAT CGGGTTCTATGTCGTAGTTCGTGTA AGGGTGCATCAATTAGTTAGTAACC TGGCTGCAGCTCAGATGGTTTCTAA AGGAGAAGAAGATAACATGGCCAT CATCAAGGAGTTCAT |

| | | | |
|---|---|---|---|
| Switch 8 | TACGCTTACGACAAGTTTTCAGATTTCAAC<br>GCCTTACCCTCATCCTAATCTTCACCTCCAA<br>TCTTACTCCGCCACCATGGGAGTAAGAGTG<br>GAGGTGAAGATTGGGATGAAGGTAAGGCA<br>TCAATTAGTTAGTAACCTGGCTGCAGCTCA<br>GATGGTTTCTAAAGGAGAAGAAGATAACAT<br>GGCCATCATCAAGGAGTTCAT | AACGCCTTACC<br>CTCATCCTAAT<br>CTTCACCTCCA<br>ATCTTACTCCG<br>CCACCATGGG<br>AGTAAGAGTG<br>GAGGTGAAGA<br>TTGGGATGAA<br>GGTAAGGCAT<br>C | TACCGGTCGCCACCATGGATACGCT<br>TACGACAAGTTTTCAGATTTCAACG<br>CCTTACCCTCATCCTAATCTTCACCT<br>CCAATCTTACTCCGCCACCATGGGA<br>GTAAGAGTGGAGGTGAAGATTGGG<br>ATGAAGGTAAGGCATCAATTAGTTA<br>GTAACCTGGCTGCAGCTCAGATGGT<br>TTCTAAAGGAGAAGAAGATAACAT<br>GGCCATCATCAAGGAGTTCAT |
| Switch 9 | TACGCTTACGACAAGTTTTCAGATTTCAAC<br>CGTCATCCATATCCATATCTCTAACCTCCAC<br>CTCCGCATATCCGCCACCATGGGATATGCG<br>GAGGTGGAGGTTATAGATGTGGATAAGGAT<br>GACGATCAATTAGTTAGTAACCTGGCTGCA<br>GCTCAGATGGTTTCTAAAGGAGAAGAAGAT<br>AACATGGCCATCATCAAGGAGTTCAT | AACCGTCATCC<br>ATATCCATATC<br>TCTAACCTCCA<br>CCTCCGCATAT<br>CCGCCACCATG<br>GGATATGCGG<br>AGGTGGAGGT<br>TATAGATGTGG<br>ATAAGGATGA<br>CGATC | TACCGGTCGCCACCATGGATACGCT<br>TACGACAAGTTTTCAGATTTCAACC<br>GTCATCCATATCCATATCTCTAACC<br>TCCACCTCCGCATATCCGCCACCAT<br>GGGATATGCGGAGGTGGAGGTTAT<br>AGATGTGGATAAGGATGACGATCA<br>ATTAGTTAGTAACCTGGCTGCAGCT<br>CAGATGGTTTCTAAAGGAGAAGAA<br>GATAACATGGCCATCATCAAGGAGT<br>TCAT |

## Relevant plasmid sequences

| Name | DNA Sequence |
|---|---|
| Kozak sequence | CGCCACCATGG |
| pCMV | TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAA<br>CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTAT<br>GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC<br>CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG<br>CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG<br>TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC<br>GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACT<br>TTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT<br>ATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATC |
| IRES | GCCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGT<br>CTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTC<br>TTGACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAG<br>GAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAAC<br>CCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGC<br>ACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATT<br>CAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTGCA<br>CATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTT<br>TCCTTTGAAAAACACGATAATACC |
| Terminator | CTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA |
| Eukaryotic mCherry: | ATGGTTTCTAAAGGAGAAGAAGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATG<br>GAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCAC<br>CCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCA<br>GTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTT<br>CCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGG<br>ACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACG<br>GCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGC<br>GCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAA<br>GACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACA<br>TCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACC<br>GGCGGCATGGACGAGCTGTACAAGAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCG<br>GGCCCGGGATCCACCGGATCTAGATAA |
| Human U6 promoter | GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGA<br>ATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG<br>TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG<br>ATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG |
| pre-Kozak switch 8 | TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAA<br>CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTAT<br>GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC<br>CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG<br>CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG<br>TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC<br>GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACT<br>TTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT<br>ATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCTAGCGCTACCGGTTACGCTTACGACAA<br>GTTTTCAGATTTCAACGCCTTACCCTCATCCTAATCTTCACCTCCAATCTTACTCCGCCACCATGGGAGT<br>AAGAGTGGAGGTGAAGATTGGGATGAAGGTAAGGCATCAATTAGTTAGTAACCTGGCTGCAGCTCAGA |

| | |
|---|---|
| | TGGTTTCTAAAGGAGAAGAAGATAACATGGCCATCATCAAGGAGTTCATgcgcttcaaggtgcacatggagggctcCGT<br>GAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCA<br>AGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACG<br>GCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCT<br>TCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGC<br>AGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGC<br>AGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGC<br>GAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAA<br>GGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAA<br>CGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACG<br>AGCTGTACAAGAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCA<br>CCGGATCTAGATAAAGTTATGGCGCGCCAATTCCGCCCCTCTCCCTCCCCCCCCCCCTAACGTTACTGGCC<br>GAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGG<br>CAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCC<br>AAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAAC<br>AACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAA<br>GCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGT<br>GGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCC<br>ATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAAC<br>GTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATAATACCATGTATCGATCC<br>ACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGA<br>CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC<br>TGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG<br>GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG<br>AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG<br>AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA<br>CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGA<br>AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC<br>CACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACC<br>CAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC<br>GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAACTGATCATAATCAGCCATACCACATTTGTA<br>GAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATT<br>GTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA<br>ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTAAA<br>TTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAG<br>GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTT<br>TGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGG<br>CGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAA<br>TCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGG<br>AAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAAC<br>CACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCG<br>CGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGA<br>TAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGT<br>CAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA<br>GTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCA<br>ATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCA<br>TTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTA<br>TTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATG<br>AGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCT<br>ATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCA<br>GGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGC<br>GCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGG<br>AAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGA<br>GAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGA<br>CCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATG<br>ATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCC<br>GACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGC<br>TTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACC<br>CGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC<br>CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAA<br>ATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGG<br>TTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAG<br>TTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGC<br>TATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTC<br>GGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTT<br>CCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAG<br>GCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAG<br>GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGA<br>GCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCT<br>TGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTC<br>CGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCC<br>ACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGC<br>CAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTC<br>GGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC<br>TACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGC<br>GGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCC<br>TGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGG |

| | |
|---|---|
| | AAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCC TGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT |

## Mammalian switch plasmid construction

### Backbone PCR

| Name | Sequence |
|---|---|
| eTk_BB_fwd | ATTAGTTAGTAACCTGGCTGCAGC |
| eTk_BB_rev | GTTGAAATCTGAAAACTTGTCGTAAGCG |

### Insert PCR

| Name | Sequence |
|---|---|
| eTk_hpin_fwd | TACGCTTACGACAAGTTTTCAGATTTCAAC |
| eTk_hpin_rev | ATCTGAGCTGCAGCCAGGTTACTAACTAAT |

| Name | Sequence |
|---|---|
| eTk_hp_1 | CGCTTACGACAAGTTTTCAGATTTCAACCGCACTTCAACCTACTCCGCCACCATGGGAGTAGGTAG AAGTGCGCAATTAGTTAGTAACCTGGCTGCAGC |
| eTk_hpin_2 | CGCTTACGACAAGTTTTCAGATTTCAACCGTACCCTCACCCGCTCCATCCGCCACCATGGGATGGA GCTGGTTAGGGTACGATCAATTAGTTAGTAACCTGGCTGCAGC |
| eTk_hp_3 | CGCTTACGACAAGTTTTCAGATTTCAACGCCTATACATACTACCGACCTTACCGCCACCATGGGTA AGGTCGGTAGTAAGTATAGGCATCAATTAGTTAGTAACCTGGCTGCAGC |
| eTk_hp_4_fwd | CGCTTACGACAAGTTTTCAGATTTCAACCGCCACATACTCCAGACAAATCCACCTCGCCACCATGG AGGTG |
| eTk_hp_5_fwd | CGCTTACGACAAGTTTTCAGATTTCAACGCCAAATCACCAGCCACCTACGAGATCTACCGCCACCA TGGGTAGA |
| eTk_hp_6_fwd | CGCTTACGACAAGTTTTCAGATTTCAACCGCCACTCATACACGCTCATACATCCGCAACCTCGCCA CCATGGAGGTT |
| eTk_hp_7_fwd | CGCTTACGACAAGTTTTCAGATTTCAACGCACCCTTACACGACCTACGACATAGATCCCGATCACG CCACCATGGTGATC |
| eTk_hp_8_fwd | CGCTTACGACAAGTTTTCAGATTTCAACGCCTTACCCTCATCCTAATCTTCACCTCCAATCTTACTC CGCCACCATGGGAGTA |
| eTk_hp_9_fwd | CGCTTACGACAAGTTTTCAGATTTCAACCGTCATCCATATCCATATCTCTAACCTCCACCTCCGCAT ATCCGCCACCATGGGATAT |
| eTk_hp_4_rev | GCTGCAGCCAGGTTACTAACTAATTGATCGCCACATACTCCCGACACATCCACCTCCATGGTGGCG AGG |
| eTk_hp_5_rev | GCTGCAGCCAGGTTACTAACTAATTGATGCCAAATCTCCAGCCACCTACCAGATCTACCCATGGTG GCGGTA |
| eTk_hp_6_rev | GCTGCAGCCAGGTTACTAACTAATTGATCGCCACTCCTACACGCCCATACATACGCAACCTCCATG GTGGCGAGG |
| eTk_hp_7_rev | GCTGCAGCCAGGTTACTAACTAATTGATGCACCCTTACACGAACTACGACATAGAACCCGATCACC ATGGTGGCGTGA |
| eTk_hp_8_rev | GCTGCAGCCAGGTTACTAACTAATTGATGCCTTACCTTCATCCCAATCTTCACCTCCACTCTTACTC CCATGGTGGCGGAG |
| eTk_hp_9_rev | GCTGCAGCCAGGTTACTAACTAATTGATCGTCATCCTTATCCACATCTATAACCTCCACCTCCGCAT ATCCCATGGTGGCGGAT |

## Trigger sequences

| Name | Sequence |
|---|---|
| Trigger 8 Full | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATGCCTTACCTTCATCCCAATCTTCACCTCCACT CTTACTCC |
| Trigger 8 Half | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATGCCTTACCTTCATCCCAAT |
| Trigger 8 Partial Junction | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATGCCTTACCTTCATCCCAATCTTCACC |
| Trigger 8 Junction | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATGCCTTACCTTCATCCCAATATTAGGA |
| Trigger 9 Full | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATCGTCATCCTTATCCACATCTATAACCTCCACC TCCGCATATCC |
| Trigger 9 Half | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATCGTCATCCTTATCCACATCTAT |
| euk9 Partial Junction | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATCGTCATCCTTATCCACATCTATAACCTCC |
| Trigger 9 Junction | CCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATCGTCATCCTTATCCACATCTATAGAGATA |

## Trigger plasmid construction

**Backbone Primers**

| Name | Sequence |
|---|---|
| **with GFP Backbone** | |
| BB_euk_trig_rev | GTTGCTCGGCGAGTGCCGAGCGGGGATCTGACGGTTCACTAAACCAGC |
| BB_euk_trig_with_GFP_fwd | ATCTAGATAAAGTTATGGCGCGCC |
| **without GFP Backbone** | |
| BB_euk_trig_rev | GTTGCTCGGCGAGTGCCGAGCGGGGATCTGACGGTTCACTAAACCAGC |
| BB_euk_trig_wo_GFP_fwd | CTGATCATAATCAGCCATACCACATTTGT |

**Trigger Primers**

| Name | Sequence |
|---|---|
| euk8_full_fwd_52C | CAGATCCCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATGCCTTACCTTCA |
| euk9_full_fwd_52C | CAGATCCCCGCTCGGCACTCGCCGAGCAACTAACTAATTGATCGTCATCCTTATCC |
| **With GFP revPrimers** | |
| euk8_full_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGGGAGTAAGAGTGGAGGTG |
| euk9_full_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGGGATATGCGGAGGTGGAG |
| euk8_partial_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGATTGGGATGAAGGTAAGGC |
| euk9_partial_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGATAGATGTGGATAAGGATGACG |
| euk8_partjxn_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGGGTGAAGATTGGGATGAAGG |
| euk9_partjxn_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGGGAGGTTATAGATGTGGATAAGG |
| euk8_junction_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGTCCTAATATTGGGATGAAGGTAAGGC |
| euk9_junction_wo_GFP_rev_52C | TACAAATGTGGTATGGCTGATTATGATCAGTATCTCTATAGATGTGGATAAGGATGACG |
| **Without GFP revPrimers** | |
| euk8_full_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATGGAGTAAGAGTGGAGGTG |
| euk9_full_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATGGATATGCGGAGGTGGAG |
| euk8_partial_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATATTGGGATGAAGGTAAGGC |
| euk9_partial_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATATAGATGTGGATAAGGATGACG |
| euk8_partjxn_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATGGTGAAGATTGGGATGAAGG |
| euk9_partjxn_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATGGAGGTTATAGATGTGGATAAGG |
| euk8_junction_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATTCCTAATATTGGGATGAAGGTAAGGC |
| euk9_junction_w_GFP_rev_52C | GGAATTGGCGCGCCATAACTTTATCTAGATTATCTCTATAGATGTGGATAAGGATGACG |

## Editing the trigger plasmid

**Exchanging CMV by the U6 promoter**

| Name | Sequence |
|---|---|
| U6_Trigger_Vector fwd | GGCTTTATATATCTTGTGGAAAGGACGAAACACCGCCCGCTCGGCACTCG |
| U6_Trigger_Vector rev | TGCAAATATGAAGGAATCATGGGAAATAGGCCCTCATGCATGGCGGTAATACGGT |
| U6_Trigger_U6 insert fwd | CTGATTCTGTGGATAACCGTATTACCGCCATGCATGAGGGCCTATTTCCCATGATTC |
| U6_Trigger_U6 insert rev | TCAATTAGTTAGTTGCTCGGCGAGTGCCGAGCGGGCGGTGTTTCGTCCTTTCCAC |

**Primers for *in vitro* constructs**

| Name | Sequence |
|---|---|
| Fwd Primer | TAATACGACTCACTATAGGGAGACTGGTTTAGTGAACCGTCAGATC |
| Rev Primer | GCAGTGAAAAAAATGCTTTATTTGTGAAATT |
| **Remove IRES+GFP** | |
| 032314_Remove_IRES+GFP_R | GGAATTGGCGCGCCATAA |
| 032314_Remove_IRES+GFP_F | CTGATCATAATCAGCCATACCACATT |

**Primers to remove mCherry**

| Name | Sequence |
|---|---|
| RemovemCh_F | AGTTATGGCGCGCCAATTC |

| | |
|---|---|
| RemovemCh_R | ACCGGTAGCGCTAGCTAG |

**Primers to remove SV40 Terminator and add TTTTT to the 5' end**

| Name | Sequence |
|---|---|
| RemovemSV40_F | TTTTTATTCTAGTTGTGGTTTGTCCAAACT |
| 8Full_Remove_SV40_R | GGAGTAAGAGTGGAGGTGAAGA |
| 8Half_Remove_SV40_R | ATTGGGATGAAGGTAAGGCATC |
| 8Junc_Remove_SV40_R | TCCTAATATTGGGATGAAGGTAAGGC |
| 8HJ_Remove_SV40_R | GGTGAAGATTGGGATGAAGGTAAG |

**Probe strands for RNA FISH**

| Sequence Name | Sequence | %GC |
|---|---|---|
| mcherry_gfp_mrna_1 | gaccccgtaattgattactaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_2 | actccatatatgggctatgaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_3 | gccatttaccgtaagttatgTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_4 | ttattgacgtcaatgggcggTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_5 | ttggcgttactatgggaacaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_6 | gtaaatactccacccattgaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_7 | atacacttgatgtactgccaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_8 | tcattgacgtcaatagggggTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_9 | tcatgtactgggcataatgcTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_10 | agatgtactgccaagtaggaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_11 | catcaccatggtaatagcgaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_12 | gaaatccccgtgagtcaaacTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_13 | caaactcccattgacgtcaaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_14 | gaaagtcccgttgattttggTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_15 | actaaaccagctctgcttatTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_16 | cgcatgaactccttgatgatTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_17 | catgaactgaggggacaggaTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_18 | gggaaggacagcttcaagtaTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_19 | gcttcaccttgtagatgaacTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_20 | gaggtgatgtccaacttgatTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_21 | attcgaagcttgagctcgagTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_22 | cgcgccataactttatctagTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_23 | cctcacattgccaaaagacgTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_24 | tgctcgtcaagaagacagggTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_25 | tttggcgagaggggaaagacTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_26 | cttccttcacgacattcaacTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_27 | aaagggtcgctacagacgttTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_28 | gtgtatcttatacacgtggcTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_29 | ctatccaactcacaacgtggTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_30 | ttgaggagagccatttgactTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_31 | atcagatcccatacaatgggTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_32 | atgtaaagcatgtgcaccgaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_33 | gcctagacgtttttttaaccTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_34 | tcaaaggaaaaccacgtcccTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_35 | atggtggatcgatacatggtTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_36 | tggcggacttgaagaagtcgTTATACATCTA | 55,00% |

| | | |
|---|---|---|
| mcherry_gfp_mrna_37 | cttgaagaagatggtgcgctTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_38 | ttgtcggccatgatatagacTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_39 | tcagttacttgtacagctcgTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_40 | aggtgtgggaggttttttaaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_41 | aagggattttgccgatttcgTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_42 | actggaacaacactcaacccTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_43 | ttggagtccacgttctttaaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_44 | gattagggtgatggttcacgTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_45 | gttccgatttagtgctttacTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_46 | cgtcaagctctaaatcggggTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_47 | tcagggttattgtctcatgaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_48 | accctaactgacacacattcTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_49 | gagatgcatgctttgcatacTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_50 | atgcatgctttgcatacttcTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_51 | ggactatggttgctgactaaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_52 | ccaaaaaagcctcctcactaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_53 | tctcttgatcgatctttgcaTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_54 | atcatgcgaaacgatcctcaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_55 | gacaggtcggtcttgacaaaTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_56 | agtgacaacgtcgagcacagTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_57 | caggagcaaggtgagatgacTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_58 | tcagccatgatggatactttTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_59 | atgaatccagaaaagcggccTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_60 | cttcagcaatatcacgggtaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_61 | taaagcacgaggaagcggtcTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_62 | aaggcgatagaaggcgatgcTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_63 | aaatctcgtgatggcaggttTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_64 | caacctttcatagaaggcggTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_65 | cgaagaactccagcatgagaTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_66 | cccgcgtttatgaacaaacgTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_67 | tgagtaacctgaggctatggTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_68 | tcctttgatcttttctacggTTATACATCTA | 40,00% |
| mcherry_gfp_mrna_69 | agagttggtagctcttgatcTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_70 | ggacagtatttggtatctgcTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_71 | ggtgctacagagttcttgaaTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_72 | ggtaacaggattagcagagcTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_73 | gtaagacacgacttatcgccTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_74 | gtaactatcgtcttgagtccTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_75 | taaagataccaggcgtttccTTATACATCTA | 45,00% |
| mcherry_gfp_mrna_76 | tcaagtcagaggtggcgaaaTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_77 | ccctgacgagcatcacaaaaTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_78 | ggccaggaaccgtaaaaaggTTATACATCTA | 55,00% |
| mcherry_gfp_mrna_79 | atcaggggataacgcaggaaTTATACATCTA | 50,00% |
| mcherry_gfp_mrna_80 | catggcggtaatacggttatTTATACATCTA | 45,00% |

# Acknowledgements

Zu allererst möchte ich mich bei Herrn Professor Friedrich Simmel für die Betreuung meiner Doktorarbeit bedanken. Du hast mir die Möglichkeit gegeben, meine Doktorarbeit in einem sehr innovativen und interdisziplinären Forschungsfeld anzufertigen. Die Projekte im Bereich der DNA Nanotechnologie, die ich bearbeitet habe, haben mein Wissen über das Leben sehr erweitert. Auch der Blick eines Physikers auf die Biologie hat mich stets inspiriert. Anfangs war mir noch nicht bewusst, wie interdisziplinär und international meine Promotionszeit an Deinem Lehrstuhl werden würde. Dadurch, dass Du mir ermöglicht hast, für fast 1 Jahr ans Wyss Institute zu gehen um dort an Projekten der Synthetischen Biologie zu forschen, konnte ich einen weiteren spannenden Bereich kennenlernen; von der interkulturellen Erfahrung mal ganz abgesehen. Diese Zeit hat mich nachhaltig geprägt. Herzlichen Dank für alles. Dank gebührt in diesem Zusammenhang auch dem SFB1032, dem Graduiertenprogramm und die damit verbundene finanzielle Unterstützung.

Bedanken möchte ich mich auch bei Susanne Kinzel, insbesondere als es um bürokratische und formale Hürden bezüglich meines USA Aufenthaltes (z.B. ein Fax nach MUC verschicken...) ging, als auch für so manche aufmunternden Gespräche.

Helene Budjarek danke ich für Rat und Tat, wenn es um Fragen innerhalb und außerhalb des Labors und des täglichen Lebens ging.

Allen ehemaligen und aktuellen Kollegen möchte ich für die Zusammenarbeit danken; darunter fällt natürlich auch das Büro 2.012. Wir haben zusammen viel gearbeitet, viel erreicht, viel erlebt und viel durchlebt. Ich danke Euch für die gemeinsame Zeit, Eure Hilfe während des Schreibens meiner Doktorarbeit, und für die entstandenen Freundschaften, welche mir viel bedeuten.

Bedanken möchte ich mich auch bei Herrn Professor Peng Yin für das Hosting in seiner Arbeitsgruppe am Wyss Institute für Biologically Inspired Engineering. In dieser sehr intensiven Zeit habe ich viel Neues gelernt und kennengelernt. Bedanken möchte ich mich in diesem Atemzug auch bei all meinen damaligen Kollegen am Wyss Institute und der „Coffee-Club Gang“.

Danke auch an meine Freunde außerhalb des Lehrstuhls, nah und fern.

Egal wo ich mich gerade befand, wie mein Gemütszustand war, wenn z.B. ein Projekt mal sehr zäh voranging; ich wusste immer wo der Fels in der Brandung zu finden war und ist. Ich danke mit allem was ich habe, meinen Eltern.

**References**

1 Watson, J. D. & Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964-967 (1953).
2 Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302, doi:10.1038/nature04586 (2006).
3 Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: de-novo-designed regulators of gene expression. *Cell* **159**, 925-939, doi:10.1016/j.cell.2014.10.002 (2014).
4 Bloomfield, V. A., Crothers, D. M. & Tinoco, I. *Nucleic Acids: Structure, Properties, and Functions*. (University Science Books, 2000, 2000).
5 Michael M. Cox, D. L. N. Lehninger Principles of Biochemistry: 6th Edition. *WH Freeman* (2012).
6 Martin, T. G. Functional Synthetic DNA Nanostructures. 221 (2014).
7 Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res* **34**, 564-574, doi:10.1093/nar/gkj454 (2006).
8 Kilchherr, F. *et al.* Single-molecule dissection of stacking forces in DNA. *Science* **353**, doi:10.1126/science.aaf5508 (2016).
9 LaBean, T. H. Handbook of Nanomaterials Properties doi:10.1007/978-3-642-31107-9 (2014).
10 Chen, H. *et al.* Understanding the mechanical properties of DNA origami tiles and controlling the kinetics of their folding and unfolding reconfiguration. *Journal of the American Chemical Society* **136**, 6995-7005, doi:10.1021/ja500612d (2014).
11 Mandelkern, M., Elias, J. G., Eden, D. & Crothers, D. M. The dimensions of DNA in solution. *J Mol Biol* **152**, 153-161 (1981).
12 Pabo, C. O. & Sauer, R. T. Protein-DNA recognition. *Annu Rev Biochem* **53**, 293-321, doi:10.1146/annurev.bi.53.070184.001453 (1984).
13 Chandrasekaran, R. & Arnott, S. The structure of B-DNA in oriented fibers. *J Biomol Struct Dyn* **13**, 1015-1027, doi:10.1080/07391102.1996.10508916 (1996).
14 Dickerson, R. E. *et al.* The anatomy of A-, B-, and Z-DNA. *Science (New York, N.Y.)* **216**, 475-485 (1982).
15 Milman, G., Langridge, R. & Chamberlin, M. J. The structure of a DNA-RNA hybrid. *Proceedings of the National Academy of Sciences of the United States of America* **57**, 1804-1810 (1967).
16 Murphy, M., Rasnik, I., Cheng, W., Lohman, T. & Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical journal* **86**, 2530-2537 (2004).
17 Krishnan, Y. & Simmel, F. C. Nucleic acid based molecular devices. *Angewandte Chemie (International ed. in English)* **50**, 3124-3156, doi:10.1002/anie.200907223 (2011).
18 SantaLucia, J. & Hicks, D. The thermodynamics of DNA structural motifs. *Annual review of biophysics and biomolecular structure* **33**, 415-440, doi:10.1146/annurev.biophys.32.110601.141800 (2004).
19 SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95**, 1460-1465 (1998).
20 Crothers, D. M. & Zimm, B. H. Theory of the Melting Transition of Synthetic Polynucleotides: Evaluation of the Stacking Free Energy. *J Mol Biol* **9**, 1-9 (1964).
21 Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *Journal of computational chemistry* **32**, 170-173, doi:10.1002/jcc.21596 (2011).
22 Voet, D. & Rich, A. The crystal structures of purines, pyrimidines and their intermolecular complexes. *Prog Nucleic Acid Res Mol Biol* **10**, 183-265 (1970).
23 Liu, Y. & West, S. C. Happy Hollidays: 40th anniversary of the Holliday junction. *Nat Rev Mol Cell Biol* **5**, 937-944, doi:10.1038/nrm1502 (2004).
24 Seeman, N. C. Nucleic acid junctions and lattices. *Journal of theoretical biology* **99**, 237-247, doi:10.1016/0022-5193(82)90002-9 (1982).
25 Neville R. Kallenbach, R.-I. M., Nadrian C. Seeman. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **305**, 305, doi:10.1038/305829a0 (1983).
26 Chen, J. H. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, doi:10.1038/350631a0 (1991).
27 Fu, T. J. & Seeman, N. C. DNA double-crossover molecules. *Biochemistry* **32**, 3211-3220, doi:10.1021/bi00064a003 (1993).
28 Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539-544, doi:10.1038/28998 (1998).
29 Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414-418, doi:10.1038/nature08016 (2009).
30 Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725-730, doi:10.1126/science.1174251 (2009).
31 Pan, K. *et al.* Lattice-free prediction of three-dimensional structure of programmed DNA assemblies. *Nat Commun* **5**, 5578, doi:10.1038/ncomms6578 (2014).
32 Sanderson, K. Bioengineering: What to make with DNA origami. *Nature* **464**, 158-159, doi:10.1038/464158a (2010).
33 Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Research* **37**, 5001-5006 (2009).

34 Guo, P. The emerging field of RNA nanotechnology. *Nat Nanotechnol* **5**, 833-842, doi:10.1038/nnano.2010.231 (2010).
35 Jasinski, D., Haque, F., Binzel, D. W. & Guo, P. Advancement of the Emerging Field of RNA Nanotechnology. *ACS Nano* **11**, 1142-1164, doi:10.1021/acsnano.6b05737 (2017).
36 Geary, C., Rothemund, P. W. & Andersen, E. S. RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345**, 799-804, doi:10.1126/science.1253920 (2014).
37 Hong, F., Zhang, F., Liu, Y. & Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem Rev*, doi:10.1021/acs.chemrev.6b00825 (2017).
38 Schlichthaerle, T., Strauss, M. T., Schueder, F., Woehrstein, J. B. & Jungmann, R. DNA nanotechnology and fluorescence applications. *Curr Opin Biotechnol* **39**, 41-47, doi:10.1016/j.copbio.2015.12.014 (2016).
39 Langecker, M. *et al.* Synthetic lipid membrane channels formed by designed DNA nanostructures. *Science (New York, N.Y.)* **338**, 932-936, doi:10.1126/science.1225624 (2012).
40 Hernández-Ainsa, S. *et al.* DNA Origami Nanopores for Controlling DNA Translocation. *ACS Nano* **7**, 6024-6030, doi:10.1021/nn401759r (2013).
41 Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nature methods* **11**, 313-318, doi:10.1038/nmeth.2835 (2014).
42 Jungmann, R. *et al.* Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano letters* **10**, 4756-4761, doi:10.1021/nl103427w (2010).
43 Dai, M., Jungmann, R. & Yin, P. Optical imaging of individual biomolecules in densely packed clusters. *Nature nanotechnology* **11**, 798-807, doi:10.1038/nnano.2016.95 (2016).
44 Steinhauer, C., Jungmann, R., Sobey, T. L., Simmel, F. C. & Tinnefeld, P. DNA origami as a nanoscopic ruler for super-resolution microscopy. *Angew Chem Int Ed Engl* **48**, 8870-8873, doi:10.1002/anie.200903308 (2009).
45 Schmied, J. J. *et al.* DNA origami nanopillars as standards for three-dimensional superresolution microscopy. *Nano Lett* **13**, 781-785, doi:10.1021/nl304492y (2013).
46 Schmied, J. J. *et al.* Fluorescence and super-resolution standards based on DNA origami. *Nat Methods* **9**, 1133-1134, doi:10.1038/nmeth.2254 (2012).
47 Voigt, N. V. *et al.* Single-molecule chemical reactions on DNA origami. *Nature nanotechnology* **5**, 200-203, doi:10.1038/nnano.2010.5 (2010).
48 Liu, H. & Liu, D. DNA nanomachines and their functional evolution. *Chem Commun (Camb)*, 2625-2636, doi:10.1039/b822719e (2009).
49 Bath, J. & Turberfield, A. J. DNA nanomachines. *Nat Nanotechnol* **2**, 275-284, doi:10.1038/nnano.2007.104 (2007).
50 Kay, E. R., Leigh, D. A. & Zerbetto, F. Synthetic molecular motors and mechanical machines. *Angew Chem Int Ed Engl* **46**, 72-191, doi:10.1002/anie.200504313 (2007).
51 Mao, C., Sun, W., Shen, Z. & Seeman, N. C. A nanomechanical device based on the B-Z transition of DNA. *Nature* **397**, 144-146, doi:10.1038/16437 (1999).
52 Srinivas, N. *et al.* On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research* **41**, 10641-10658, doi:10.1093/nar/gkt801 (2013).
53 Zhang, D. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society* **131**, 17303-17314, doi:10.1021/ja906987s (2009).
54 Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry* **3**, 103-113, doi:10.1038/nchem.957 (2011).
55 Krishnan, Y. & Simmel, F. C. Nucleic Acid Based Molecular Devices. *Angewandte Chemie-International Edition* **50**, 3124-3156, doi:10.1002/anie.200907223 (2011).
56 Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science (New York, N.Y.)* **266**, 1021-1024 (1994).
57 Simmel, F. C., Yurke, B. & Sanyal, R. J. Operation kinetics of a DNA-based molecular switch. *J Nanosci Nanotechnol* **2**, 383-390 (2002).
58 Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605-608, doi:10.1038/35020524 (2000).
59 Green, C. & Tibbetts, C. Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Res* **9**, 1905-1918 (1981).
60 Radding, C. M., Beattie, K. L., Holloman, W. K. & Wiegand, R. C. Uptake of homologous single-stranded fragments by superhelical DNA. IV. Branch migration. *J Mol Biol* **116**, 825-839 (1977).
61 Srinivas, N. *et al.* On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic acids research* **41**, 10641-10658, doi:10.1093/nar/gkt801 (2013).
62 Vologodskii, A. V. & Cozzarelli, N. R. Conformational and thermodynamic properties of supercoiled DNA. *Annu Rev Biophys Biomol Struct* **23**, 609-643, doi:10.1146/annurev.bb.23.060194.003141 (1994).
63 Reynaldo, L. P., Vologodskii, A. V., Neri, B. P. & Lyamichev, V. I. The kinetics of oligonucleotide replacements. *Journal of molecular biology* **297**, 511-520, doi:10.1006/jmbi.2000.3573 (2000).
64 Šulc, P., Ouldridge, T. E., Romano, F., Doye, J. P. & Louis, A. A. Modelling toehold-mediated RNA strand displacement. *Biophysical journal* **108**, 1238-1247, doi:10.1016/j.bpj.2015.01.023 (2015).
65 Lesnik, E. A. & Freier, S. M. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry* **34**, 10807-10815 (1995).
66 Yurke, B. & Jr, A. P. Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines* **4**, 111-122 (2003).

67 Zhang, D. Y., Turberfield, A. J., Yurke, B. & Winfree, E. Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science* **318**, 1121-1125 (2007).
68 Prasad, R. *et al.* DNA polymerase beta -mediated long patch base excision repair. Poly(ADP-ribose)polymerase-1 stimulates strand displacement DNA synthesis. *J Biol Chem* **276**, 32411-32414, doi:10.1074/jbc.C100292200 (2001).
69 Liu, Y. *et al.* Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta. *J Biol Chem* **282**, 13532-13541, doi:10.1074/jbc.M611295200 (2007).
70 Jiang, M., Ma, N., Vassylyev, D. G. & McAllister, W. T. RNA displacement and resolution of the transcription bubble during transcription by T7 RNA polymerase. *Mol Cell* **15**, 777-788, doi:10.1016/j.molcel.2004.07.019 (2004).
71 Schoen, I., Krammer, H. & Braun, D. Hybridization kinetics is different inside cells. *Proc Natl Acad Sci U S A* **106**, 21649-21654, doi:10.1073/pnas.0901313106 (2009).
72 Simmel, F. C. & Yurke, B. Using DNA to construct and power a nanoactuator. *Phys Rev E Stat Nonlin Soft Matter Phys* **63**, 041913, doi:10.1103/PhysRevE.63.041913 (2001).
73 Yurke, B. & Mills, A. Using DNA to Power Nanostructures. *Genetic Programming and Evolvable Machines* **4**, 111-122, doi:10.1023/a:1023928811651 (2003).
74 Yan, H., Zhang, X., Shen, Z. & Seeman, N. C. A robust DNA mechanical device controlled by hybridization topology. *Nature* **415**, 62-65, doi:10.1038/415062a (2002).
75 Chakraborty, B., Sha, R. & Seeman, N. C. A DNA-based nanomechanical device with three robust states. *Proc Natl Acad Sci U S A* **105**, 17245-17249, doi:10.1073/pnas.0707681105 (2008).
76 Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science (New York, N.Y.)* **335**, 831-834, doi:10.1126/science.1214081 (2012).
77 Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73-76, doi:10.1038/nature07971 (2009).
78 Grossi, G., Dalgaard Ebbesen Jepsen, M., Kjems, J. & Andersen, E. S. Control of enzyme reactions by a reconfigurable DNA nanovault. *Nat Commun* **8**, 992, doi:10.1038/s41467-017-01072-8 (2017).
79 Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585-1588 (2006).
80 Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368-372, doi:10.1038/nature10262 (2011).
81 Agapakis, C. M., Boyle, P. M. & Silver, P. A. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat Chem Biol* **8**, 527-535, doi:10.1038/nchembio.975 (2012).
82 Sweetlove, L. J. & Fernie, A. R. The spatial organization of metabolism within the plant cell. *Annu Rev Plant Biol* **64**, 723-746, doi:10.1146/annurev-arplant-050312-120233 (2013).
83 Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* **11**, 114-119 (2001).
84 Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* **332**, 680-686, doi:10.1126/science.1198701 (2011).
85 Engelman, J. A., Luo, J. & Cantley, L. C. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat Rev Genet* **7**, 606-619, doi:10.1038/nrg1879 (2006).
86 Conrado, R. J., Varner, J. D. & DeLisa, M. P. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Current opinion in biotechnology* **19**, 492-499, doi:10.1016/j.copbio.2008.07.006 (2008).
87 Negron, L., Patchett, M. L. & Parker, E. J. Expression, Purification, and Characterisation of Dehydroquinate Synthase from Pyrococcus furiosus. *Enzyme Res* **2011**, 134893, doi:10.4061/2011/134893 (2011).
88 Buchner, A., Tostevin, F., Hinzpeter, F. & Gerland, U. Optimization of collective enzyme activity via spatial localization. *The Journal of chemical physics* **139**, 135101, doi:10.1063/1.4823504 (2013).
89 Buchner, A., Tostevin, F. & Gerland, U. Clustering and Optimal Arrangement of Enzymes in Reaction-Diffusion Systems. *Physical Review Letters* **110**, art. no. 208104, doi:10.1103/PhysRevLett.110.208104 (2013).
90 Li, X. & Liu, D. R. DNA-templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angewandte Chemie-International Edition* **43**, 4848-4870 (2004).
91 Wilner, O. I. *et al.* Enzyme cascades activated on topologically programmed DNA scaffolds. *Nature nanotechnology* **4**, 249-254, doi:10.1038/nnano.2009.50 (2009).
92 Fu, J., Liu, M., Liu, Y., Woodbury, N. W. & Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *Journal Of The American Chemical Society* **134**, 5516-5519, doi:10.1021/ja300897h (2012).
93 Linko, V., Eerikäinen, M. & Kostiainen, M. A. A modular DNA origami-based enzyme cascade nanoreactor. *Chemical Communications* **51**, 5351-5354, doi:10.1039/C4CC08472A (2015).
94 Zhang, Y., Wang, Q. & Hess, H. Increasing Enzyme Cascade Throughput by pH-Engineering the Microenvironment of Individual Enzymes. *ACS Catalysis* **7**, 2047-2051, doi:10.1021/acscatal.6b03431 (2017).
95 Thubagere, A. J. *et al.* A cargo-sorting DNA robot. *Science* **357**, doi:10.1126/science.aan6558 (2017).
96 Chatterjee, G., Dalchau, N., Muscat, R. A., Phillips, A. & Seelig, G. A spatially localized architecture for fast and modular DNA computing. *Nat Nanotechnol* **12**, 920-927, doi:10.1038/nnano.2017.127 (2017).
97 Miescher, F. Ueber die chemische Zusammensetzung der Eiterzellen. *Hoppe-Seyler's medizinisch-chemischen Untersuchungen* **Heft 4**, S. 441–460 (1871).
98 Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* **122**, 565-581, doi:10.1007/s00439-007-0433-0 (2008).
99 Griffith, F. The Significance of Pneumococcal Types. *J Hyg (Lond)* **27**, 113-159 (1928).

100 Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J Exp Med* **79**, 137-158 (1944).
101 Chargaff, E., Vischer, E. & et al. The composition of the desoxypentose nucleic acids of thymus and spleen. *J Biol Chem* **177**, 405-416 (1949).
102 Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed Proc* **10**, 654-659 (1951).
103 Hershey, A. D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* **36**, 39-56 (1952).
104 Crick, F. Ideas on Protein Synthesis. doi:https://doi.org/10.1371/journal.pbio.2003243.g001 (1956).
105 Cobb, M. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol* **15**, e2003243, doi:10.1371/journal.pbio.2003243 (2017).
106 Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A* **47**, 1588-1602 (1961).
107 Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
108 Berg, J. M., Tymoczko, J. L. & Stryer, L. Biochemistry. (2012).
109 O'Donnell, M., Langston, L. & Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol* **5**, doi:10.1101/cshperspect.a010108 (2013).
110 Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491 (1988).
111 C, K. Cell and molecular biology: concepts and experiments. *Wiley, Hoboken* (2004).
112 Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes & Development* **25**, 1770-1782, doi:10.1101/gad.17268411 (2011).
113 Shine, J. & Dalgarno, L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* **71**, 1342-1346 (1974).
114 Sonenberg, N. & Dever, T. E. Eukaryotic translation initiation factors and regulators. *Curr Opin Struct Biol* **13**, 56-63 (2003).
115 Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5**, 827-835, doi:10.1038/nrm1488 (2004).
116 Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-127, doi:10.1038/nrm2838 (2010).
117 Parsyan, A. *et al.* mRNA helicases: the tacticians of translational control. *Nat Rev Mol Cell Biol* **12**, 235-245, doi:10.1038/nrm3083 (2011).
118 Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**, 8125-8148 (1987).
119 Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* **266**, 19867-19870 (1991).
120 Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292 (1986).
121 Kozak, M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J* **16**, 2482-2492, doi:10.1093/emboj/16.9.2482 (1997).
122 Hellen, C. U. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev* **15**, 1593-1612, doi:10.1101/gad.891101 (2001).
123 Scientific, T. Ribosomal Binding Site Sequence Requirements.
124 Zur, H. & Tuller, T. New universal rules of eukaryotic translation initiation fidelity. *PLoS Comput Biol* **9**, e1003136, doi:10.1371/journal.pcbi.1003136 (2013).
125 Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev* **16**, 2906-2922, doi:10.1101/gad.1020902 (2002).
126 Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* **9**, 5134-5142 (1989).
127 Babendure, J. R., Babendure, J. L., Ding, J.-H. H. & Tsien, R. Y. Control of mammalian translation by mRNA structure near caps. *RNA (New York, N.Y.)* **12**, 851-861, doi:10.1261/rna.2309906 (2006).
128 Hosoda, N. *et al.* Translation termination factor eRF3 mediates mRNA decay through the regulation of deadenylation. *J Biol Chem* **278**, 38287-38291, doi:10.1074/jbc.C300300200 (2003).
129 Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509-524, doi:10.1038/nrm3838 (2014).
130 Lin, S. & Gregory, R. I. MicroRNA biogenesis pathways in cancer. *Nat Rev Cancer* **15**, 321-333, doi:10.1038/nrc3932 (2015).
131 Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857-866, doi:10.1038/nrc1997 (2006).
132 Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).
133 Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**, D152-157, doi:10.1093/nar/gkq1027 (2011).
134 Macfarlane, L.-A. A. & Murphy, P. R. MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics* **11**, 537-561, doi:10.2174/138920210793175895 (2010).

135 Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-4060, doi:10.1038/sj.emboj.7600385 (2004).
136 Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-240, doi:10.1038/nature03120 (2004).
137 Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-3027, doi:10.1101/gad.1262504 (2004).
138 Bohnsack, M. T., Czaplinski, K. & Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185-191 (2004).
139 Park, J. E. *et al.* Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**, 201-205, doi:10.1038/nature10198 (2011).
140 Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740-744, doi:10.1038/nature03868 (2005).
141 Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343-349, doi:10.1038/nature02873 (2004).
142 Hu, H. Y. *et al.* Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* **10**, 413, doi:10.1186/1471-2164-10-413 (2009).
143 Eulalio, A., Behm-Ansmant, I., Schweizer, D. & Izaurralde, E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol* **27**, 3970-3981, doi:10.1128/MCB.00128-07 (2007).
144 Yang, J. S. & Lai, E. C. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol Cell* **43**, 892-903, doi:10.1016/j.molcel.2011.07.024 (2011).
145 Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2257-2261, doi:10.1073/pnas.0510565103 (2006).
146 Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838, doi:10.1038/nature03702 (2005).
147 Cameron, E. D., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nature Reviews Microbiology* **12**, 381-390, doi:10.1038/nrmicro3239 (2014).
148 Monod, J. & Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harbor symposia on quantitative biology* **26**, 389-401 (1961).
149 Medicine, N. P. i. P. o. The Nobel Prize in Physiology or Medicine 1965. *Nobel Media AB 2014* **Nobelprize.org** (1965).
150 Bartlett, J. M. & Stirling, D. A short history of the polymerase chain reaction. *Methods Mol Biol* **226**, 3-6, doi:10.1385/1-59259-384-4:3 (2003).
151 Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1462 (1997).
152 Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563-547 (1996).
153 McAdams, H. H. & Shapiro, L. Circuit simulation of genetic networks. *Science* **269**, 650-656 (1995).
154 Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342, doi:10.1038/35002131 (2000).
155 Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338, doi:10.1038/35002125 (2000).
156 Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22**, 841-847, doi:10.1038/nbt986 (2004).
157 Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952-956, doi:10.1038/nature01145 (2002).
158 Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nature biotechnology* **22**, 841-847, doi:10.1038/nbt986 (2004).
159 Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R. & Collins, J. J. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc Natl Acad Sci U S A* **107**, 15898-15903, doi:10.1073/pnas.1009747107 (2010).
160 Callura, J. M., Cantor, C. R. & Collins, J. J. Genetic switchboard for synthetic biology applications. *Proc Natl Acad Sci U S A* **109**, 5850-5855, doi:10.1073/pnas.1203808109 (2012).
161 Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B. & Arkin, A. P. Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol* **8**, 447-454, doi:10.1038/nchembio.919 (2012).
162 Kittle, J. D., Simons, R. W., Lee, J. & Kleckner, N. Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the anti-sense RNA. *J Mol Biol* **210**, 561-572 (1989).
163 Ma, C. & Simons, R. W. The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. *EMBO J* **9**, 1267-1274 (1990).
164 Chappell, J., Watters, K. E., Takahashi, M. K. & Lucks, J. B. A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Curr Opin Chem Biol* **28**, 47-56, doi:10.1016/j.cbpa.2015.05.018 (2015).
165 Gutell, R. R., Cannone, J. J., Konings, D. & Gautheret, D. Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J Mol Biol* **300**, 791-803, doi:10.1006/jmbi.2000.3900 (2000).
166 Franch, T., Petersen, M., Wagner, E. G., Jacobsen, J. P. & Gerdes, K. Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J Mol Biol* **294**, 1115-1125, doi:10.1006/jmbi.1999.3306 (1999).
167 Vimberg, V., Tats, A., Remm, M. & Tenson, T. Translation initiation region sequence preferences in Escherichia coli. *BMC Mol Biol* **8**, 100, doi:10.1186/1471-2199-8-100 (2007).

168 Huttenhofer, A. & Noller, H. F. Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J* **13**, 3892-3901 (1994).
169 Green, A. A. *et al.* Complex cellular logic computation using ribocomputing devices. *Nature* **548**, 117-121, doi:10.1038/nature23271 (2017).
170 Pardee, K. *et al.* Paper-based synthetic gene networks. *Cell* **159**, 940-954, doi:10.1016/j.cell.2014.10.004 (2014).
171 Pardee, K. *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165**, 1255-1266, doi:10.1016/j.cell.2016.04.059 (2016).
172 Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature Reviews Molecular Cell Biology* **15**, 95-107, doi:10.1038/nrm3738 (2014).
173 Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331-338, doi:10.1038/nature10886 (2012).
174 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).
175 Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular cell* **54**, 698-710, doi:10.1016/j.molcel.2014.04.022 (2014).
176 Sander, J. D. & Joung, K. J. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355, doi:10.1038/nbt.2842 (2014).
177 Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **4**, doi:10.1101/cshperspect.a003566 (2012).
178 Endo, K., Stapleton, J. A., Hayashi, K., Saito, H. & Inoue, T. Quantitative and simultaneous translational control of distinct mammalian mRNAs. *Nucleic Acids Res* **41**, e135, doi:10.1093/nar/gkt347 (2013).
179 Auslander, S. & Fussenegger, M. Synthetic RNA-based switches for mammalian gene expression control. *Curr Opin Biotechnol* **48**, 54-60, doi:10.1016/j.copbio.2017.03.011 (2017).
180 Auslander, S., Auslander, D., Muller, M., Wieland, M. & Fussenegger, M. Programmable single-cell mammalian biocomputers. *Nature* **487**, 123-127, doi:10.1038/nature11149 (2012).
181 Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *Journal Of The American Chemical Society* **131**, 17303-17314, doi:10.1021/ja906987s (2009).
182 Green, N. M. Avidin. *Adv Protein Chem* **29**, 85-133 (1975).
183 Dessinges, M.-N. *et al.* Stretching Single Stranded DNA, a Model Polyelectrolyte. *Physical Review Letters* **89**, art. no. 248102, doi:10.1103/PhysRevLett.89.248102 (2002).
184 Hansen, M. H. *et al.* A Yoctoliter-Scale DNA Reactor for Small-Molecule Evolution. *Journal of the American Chemical Society* **131**, 1322-1327, doi:10.1021/ja808558a (2009).
185 Berg, H. C. *Random Walks in Biology*. (Princeton University Press, 1993).
186 Pörschke, D. & Eigen, M. Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic--oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *Journal of Molecular Biology* **62**, 361-381 (1971).
187 Craig, M., Crothers, D. & Doty, P. Relaxation Kinetics of Dimer Formation by Self Complementary Oligonucleotides. *Journal Of Molecular Biology* **62**, 383-401 (1971).
188 Tian, Y., He, Y., Chen, Y., Yin, P. & Mao, C. A DNAzyme that walks processively and autonomously along a one-dimensional track. *Angewandte Chemie-International Edition* **44**, 4355-4358 (2005).
189 Green, S., Bath, J. & Turberfield, A. Coordinated Chemomechanical Cycles: A Mechanism for Autonomous Molecular Motion. *Physical Review Letters* **101**, art. no. 238101, doi:10.1103/PhysRevLett.101.238101 (2008).
190 Yin, P., Choi, H. M., Calvert, C. R. & Pierce, N. A. Programming biomolecular self-assembly pathways. *Nature* **451**, 318-322 (2008).
191 Omabegho, T., Sha, R. & Seeman, N. C. A Bipedal DNA Brownian Motor with Coordinated Legs. *Science* **324**, 67-71, doi:10.1126/science.1170336 (2009).
192 Lund, K. *et al.* Molecular robots guided by prescriptive landscapes. *Nature* **465**, 206-210, doi:10.1038/nature09012 (2010).
193 Wickham, S. F. J. *et al.* Direct observation of stepwise movement of a synthetic molecular transporter. *Nature Nanotechnology* **6**, 166-169 (2011).
194 Dannenberg, F., Kwiatkowska, M., Thachuk, C. & Turberfield, A. in *DNA Computing and Molecular Programming* Vol. 8141 *Lecture Notes in Computer Science* (eds David Soloveichik & Bernard Yurke) Ch. 3, 31-45 (Springer International Publishing, 2013).
195 Qian, L. & Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **332**, 1196-1201, doi:10.1126/science.1200520 (2011).
196 Nazarenko, I., Pires, R., Lowe, B., Obaidy, M. & Rashtchian, A. Effect of primary and secondary structure of oligodeoxyribonucleotides on the fluorescent properties of conjugated dyes. *Nucleic acids research* **30**, 2089-2195 (2002).
197 Stahl, E., Martin, T. G., Praetorius, F. & Dietz, H. Facile and scalable preparation of pure and dense DNA origami solutions. *Angewandte Chemie (International ed. in English)* **53**, 12735-12740, doi:10.1002/anie.201405991 (2014).
198 Valeur, B. Molecular Fluorescence: Principles and Applications. *Wiley-VCH Verlag GmbH* (2001).
199 Lakowicz, J. R. *Principles of fluorescence spectroscopy*. (Springer, 2006).
200 Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys Rev Lett* **56**, 930-933, doi:10.1103/PhysRevLett.56.930 (1986).
201 Kopperger, E. a. A. R., Ali AFM Studies and Surface Assisted Self-Assembly of DNA Nanostructures. (2015).

202 Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *Journal of computational chemistry* **32**, 170-173, doi:10.1002/jcc.21596 (2011).
203 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).
204 Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **64**, 2240-2246 (1998).
205 Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320-325, doi:10.1038/334320a0 (1988).
206 Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* **62**, 2636-2643 (1988).
207 Jackson, R. J., Howell, M. T. & Kaminski, A. The novel mechanism of initiation of picornavirus RNA translation. *Trends Biochem Sci* **15**, 477-483 (1990).
208 Paule, M. R. & White, R. J. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res* **28**, 1283-1298 (2000).
209 Vannini, A. *et al.* Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell* **143**, 59-70, doi:10.1016/j.cell.2010.09.002 (2010).
210 Szymanski, M., Barciszewska, M. Z., Erdmann, V. A. & Barciszewski, J. 5S Ribosomal RNA Database. *Nucleic Acids Res* **30**, 176-178 (2002).
211 Renwick, N. *et al.* Multicolor microRNA FISH effectively differentiates tumor types. *J Clin Invest* **123**, 2694-2702, doi:10.1172/JCI68760 (2013).
212 Mateus, A., Matsson, P. & Artursson, P. Rapid measurement of intracellular unbound drug concentrations. *Mol Pharm* **10**, 2467-2478, doi:10.1021/mp4000822 (2013).
213 Reshes, G., Vanounou, S., Fishov, I. & Feingold, M. Cell shape dynamics in Escherichia coli. *Biophys J* **94**, 251-264, doi:10.1529/biophysj.107.104398 (2008).
214 Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38**, D750-753, doi:10.1093/nar/gkp889 (2010).
215 Ross, J. mRNA stability in mammalian cells. *Microbiol Rev* **59**, 423-450 (1995).
216 Mortimer, S. A., Trapnell, C., Aviran, S., Pachter, L. & Lucks, J. B. SHAPE-Seq: High-Throughput RNA Structure Analysis. *Curr Protoc Chem Biol* **4**, 275-297, doi:10.1002/9780470559277.ch120019 (2012).
217 Kabanov, A. V. & Kabanov, V. A. DNA complexes with polycations for the delivery of genetic material into cells. *Bioconjug Chem* **6**, 7-20 (1995).
218 Labat-Moleur, F. *et al.* An electron microscopy study into the mechanism of gene transfer with lipopolyamines. *Gene Ther* **3**, 1010-1017 (1996).
219 Yeung, T. *et al.* Membrane phosphatidylserine regulates surface charge and protein localization. *Science* **319**, 210-213, doi:10.1126/science.1152066 (2008).
220 Ibidi. Transfection.
221 Ormerod, M. G. *Flow cytometry: a basic introduction*.
222 Macey, M. G. *Flow cytometry: Principles and applications.*, (Totowa, N.J: Humana Press, 2007).
223 Rieger, A., Baghirova, Sabina, Burshtyn, Debby. <https://flowcytometry.med.ualberta.ca/> (
224 Lichtman, J. W. & Conchello, J. A. Fluorescence microscopy. *Nat Methods* **2**, 910-919, doi:10.1038/nmeth817 (2005).
225 Shaffer, S. M., Wu, M. T., Levesque, M. J. & Raj, A. Turbo FISH: a method for rapid single molecule RNA FISH. *PLoS One* **8**, e75120, doi:10.1371/journal.pone.0075120 (2013).