FACULTY OF
BIOSCIENCE ENGINEERING

Expanding the metabolic engineering toolbox
to rationally and predictably modulate gene
expression at the transcriptional and
translational level

Gert Peters

GHENT
UNIVERSITY

UNIVERSITY
press

Expanding the metabolic engineering toolbox to rationally and predictably modulate gene expression at the transcriptional and translational level

Gert Peters

Dutch translation of the title

Uitbreiding van de beschikbare *metabolic engineering* technieken om rationeel en voorspelbaar genexpressie aan te passen op het transcriptionele en translationele niveau.

Cover illustration

Leigh Prather (shutterstock)

Woord vooraf

Het schrijven van je woord vooraf is het eindpunt van het schrijven van je doctoraat. Licht aan het einde van de tunnel. Gedurende de jaren maakte idealisme plaats voor realisme. Zo is een doctoraat een proces van vallen en opstaan waar je, in hoge mate, alleen voor staat. Toch zou het eindresultaat onmogelijk zijn zonder de bijdrage van andere personen, waarvoor ik oprecht dankbaar ben. Ik probeer het kort te houden.

Zo wil ik mijn promotoren, Marjan en Jeroen, bedanken om mij de kans te geven om eigen ideeën te onderzoeken in een ongedwongen sfeer. Ik heb de voorbije jaren met veel plezier onderzoek gedaan en een goede begeleiding is daarbij essentieel. Waarvoor dank. Daarnaast is er de laatste jaren veel veranderd. De metabolic engineering groep is onophoudelijk gegroeid tot een groep van gepassioneerde jonge onderzoekers. Een continue zoektocht naar innovatie binnen de synthetische biologie (met bijhorende hoge standaarden) gaat er hand in hand met een goede sfeer. Voor deze setting ben ik dankbaar.

Eén iemand heeft daarbij een grote invloed gehad: Jo. Bedankt voor de jarenlange begeleiding en ondersteuning, je vakkennis en perfectionisme hebben dit doctoraat naar een hoger niveau gebracht. Verder hebben verschillende mensen binnen de MEMO groep een invloed gehad op mijn doctoraat. In den beginne was er Hendrik, die al snel zijn ervaringen als senior PhD student deelde. Tijdens mijn masterthesis startten Frederik en Pieter aan hun persoonlijk avontuur, wat de basis zou vormen van de verdere groei van de MEMO groep. Frederik, mijn officieuze thesisbegeleider, bezorgde me de nodige wijsheden (*mijn metje zegt altijd...*) die me toelieten om de PhD tijd te overleven, zowel binnen het labo als daarbuiten. Pieter zorgde met zijn constante zoektocht naar nieuwigheden (*alles voor vijf jaar geleden is prehistorie...*) voor de nodige impulsen om innovatief, relevant en scherp te blijven. Tot zover de 'oude' garde, diegenen die voor mij zijn gestart aan hun doctoraat. Het drietal Bob, Brecht en Thomas. Bob (de speciale) verwonderde me telkens weer. De mifare saga, de NFC chip of de '*ik heb de*

3D printer in mijn slaapkamer gezet, anders worden de buren wakker van het geluid vrees ik...': het zijn maar enkele onvergetelijke voorbeelden. Brecht, altijd te vinden om voor een gesprek, over wetenschap of *thunder hooves rage*. Ook bedankt voor de inhoudelijke input rond biosensoren. Thomas die de synthetische biologie in gist op de kaart probeert te zetten, immens respect voor het doorzettingsvermogen. En bedankt om te gidsen in Heuvelland en Frans-Vlaanderen, doen we nog eens. Het volgende drietal, David, Van Brempt en Tom. David, de Chuck Berry van de groep (nee, niet op alle vlakken), even gepassioneerd door metabolic engineering als door rockabilly. Van Brempt, van comic relief tot volwaardig wetenschapper, het ging snel. Goed voor een aantal uitspraken die (zelfs) ik hier niet durf te herhalen. Tom, weinig woorden nodig, gewoon doordoen. Ook de vreemde eend Mol mag niet ontbreken, bedankt. Na jaren dan toch die (door Marjan verhoopte) vrouwelijke inbreng, Lien en Chiara. Veel succes in de toekomst, genoeg goede mensen om alles in goede banen te leiden. Gedurende al die jaren kon ik ook rekenen op Wouter en Dries D., de rotsen in de branding binnen de MEMO groep. Wouter, bedankt voor alle hulp over al die jaren. Nooit was er iets teveel. En altijd met diezelfde karakteristieke glimlach en met een voorbeeldige werkethiek. Daarnaast bedankt voor de vele fijne momenten naast het labo. Als er iemand in staat is om een feestje te bouwen ben jij het, en dat is nog heel voorzichtig omschreven. Dries D., jouw hulp maakte het laatste jaar iets draaglijker. Merci voor alles. Bewonderenswaardig hoe je moleculaire biologie verheft tot een echte competitie, die je altijd winnend afsluit.

De herinneringen aan de mindere momenten tijdens mijn doctoraat worden gelukkig overvleugeld door de memorabele gebeurtenissen. Iedereen die ze heeft meegemaakt weet waarover ik het heb. Talrijke muzieknummers zullen tot het einde der tijden herinneringen oproepen aan lang vervlogen tijden. Het motto '*work hard, play hard*' kon je nergens letterlijker waarnemen dan op het labo. Hiervoor ben ik bijzonder dankbaar. De bureau van het klein labo zoals ik het me altijd zal herinneren: Gilles, Dries VH, Eric, Maarten D en Catherine. Met veel te veel in een te kleine ruimte maar altijd bereid voor een babbeltje of om de laatste laboperikelen op te lossen. Ook de glycodirect en biosurf collegas mogen daarbij niet ontbreken. Met een bijzondere vermelding voor Verhaeghe en Margo, bedankt voor de minder traditionele muzikale uitstapjes en interdisciplinaire discussies. Ook Robin mag ik niet vergeten bedanken als onderdeel van de IWT lichting van 2013. Steevast aanwezig op vrijdagavond en op de verkenningen in de Vlaamse Ardennen. En bedankt om mee een afscheidsfeestje te organiseren, het werd unaniem top bevonden. Sophie Roelants, bedankt voor het eeuwige enthousiasme. Sofie De Maeseneire, bedankt om ondanks alles het leven in het labo te proberen verbeteren. Verder

wil ik Joeri bedanken voor de begeleiding en alle hulp gedurende de jaren bij alle mogelijke wetenschappelijke problemen. Dat ik naast synthetische biologie ook vaardigheden heb kunnen verwerven op andere gebieden heb ik ook te danken aan Gaspard. Die altijd klaar stond om te helpen bij één of ander informatica probleem, ook zelden iemand zo geamuseerd gezien bij het declareren van de variabele PATH in zowat mijn eerste bash script. Classic. Wim wens ik te bedanken om tijdens zijn les dat 1,3-propaandiol verhaal (met die triose fosfaat isomerase mutant) te vertellen, wat mijn interesse voor metabolic engineering heeft gewekt.

Daarnaast wil ik mijn familie en vrienden bedanken. De laatste jaren heb ik veel activiteiten links laten liggen, hopend op begrip. Bedankt voor alle steun en de kansen die ik heb gekregen de voorbije, soms moeilijke, jaren.

Als laatste wil ik Magali bedanken. Of waar een summer school over de bio-economie niet allemaal goed voor is. Bedankt voor alle steun en hulp. Als koppel ieder bezig zijn met je eigen doctoraat en dat nog over exact dezelfde periode leek voor velen een bijzondere uitdaging. We hebben het toch maar mooi gedaan. Bedankt om een luisterend oor te bieden op moeilijke momenten, al dan niet gerelateerd met mijn doctoraat. Ik kan me een leven zonder jou niet meer voorstellen. Bedankt voor alles.

Zo, dat was het dan. Letterlijk de laatste woorden van mijn boekje. Einde van een zwaar maar ook plezant hoofdstuk. Het was me een genoegen.

Gert

23 april 2017

Contents

Acronyms

AR   activation ratio

CDS   coding DNA sequence

CRISPR   clustered regularly interspaced short palindromic repeats

CRISPRi   clustered regularly interspaced short palindromic repeats interference

CRP   cAMP receptor protein

crRNA   CRISPR RNA

DOE   design of experiments

DWP   deep well plates

EA   free energy of the tiRNA monomer

EAA   free energy of the tiRNA-tiRNA dimer

EAB   free energy of the tiRNA-UTR dimer

EIS   intermolecular binding seed energy

ETS   total seed energy

FAA   formation energy of the tiRNA-tiRNA dimer

FAB   formation energy of the tiRNA-UTR dimer

FACS   fluorescence activated cell sorting

FRET   Förster resonance energy transfer

Fru6P   fructose 6-phosphate

GlcNAc   *N*-acetylglucosamine

GlcNAc6P *N*-acetylglucosamine 6-phosphate

HPLC	high-performance liquid chromatography
HTH	helix-turn-helix
IRES	internal ribosome entry site
LB	lysogeny broth
LBA	LB agar
ManNAc	*N*-acetylmannosamine
MFE	minimum free energy
Neu5Ac	*N*-acetylneuraminic acid
OD	optical density
OLS	ordinary least squares
ORF	open reading frame
ori	origin of replication
PAM	protospacer adjacent motif
PAU	probability availability of UTR
PCC	Pearson correlation coefficient
PLS	partial least squares
PT	paired termini
R^2	coefficient of determination
RBS	ribosome binding site
RBS11	RBS coverage of length 11
RBS5	RBS coverage of length 5
RSS	residual sum of squares
SELEX	systematic evolution of ligands by exponential enrichment
sgRNA	small guide RNA
sRNA	small RNA
STAR	small transcription activating RNA

TALEN	transcription activator-like effector nuclease
TF	transcription factor
tiRNA	translation inhibiting RNA
UDP-GlcNAc	UDP-N-acetylglucosamine
UTR	untranslated region
ZF	zinc finger

Chapter 1

Outline

Spurred by environmental and socio-economical concerns, there is a continuing strive for a more sustainable society with environment-friendly technologies. An important technology in this global challenge is industrial biotechnology, which uses microbial cells or enzymes for the fabrication of industrially relevant products from renewable resources. The possibilities of industrial biotechnology were greatly expanded at the end of the 20th century with the emergence of metabolic engineering, allowing rewiring the metabolism of the microbial cell factory to maximize the product yield. Several new production processes were developed using metabolic engineering, allowing sustainable production of an ever increasing number of both bulk and fine chemicals, previously considered impossible to be produced in an economically feasible way[1–5].

In order to achieve a maximal productivity the biosynthetic pathway needs to be optimized by balancing the required enzymes to obtain a maximal flux towards the product of interest. Typically, the production pathway is embedded in the native, heavily regulated, metabolism of the host organism, causing a decrease in productivity due to undesired biochemical conversions and native counterproductive regulation. Thus, the creation of custom microbial cell factories is a daunting task, which heavily depends on techniques to alter both the production pathway and the native metabolism of the host organism. To this end, metabolic engineering traditionally relies on techniques based on proteins, which are recently complemented with novel RNA-based parts to alter microbial metabolism.

Recently, synthetic biology, a research field that aims to systematically engineering biology in a standardized way, has resulted in numerous emerged technologies that allow effectively, rationally and predictably engineer biological systems[6,7]. Specifically, cutting edge technologies allow fast and affordable DNA synthesis, improved methods for DNA assembly, and high-throughput sequencing but applying these technologies to reliably engineer microbial cells in a custom way remains challenging[8–10]. As such, the development of custom microbial cell factories still requires reliable and efficient tools to rationally engineer gene expression in a predictable way, which improves the forward engineering capacities in the field of metabolic engineering.

Hence, the main goal of this PhD dissertation is to develop and optimize metabolic engineering tools, enabling forward engineering of microbial cell factories. To this end, engineering principles for both RNA-based or protein-based tools and devices were developed, optimized and evaluated in view to ultimately rewire the metabolism of microbial cell factories in a custom way. The recently emerged RNA-based tools are still in its infancy compared to the protein-based tools. To validate molecular detection using

RNA devices, theophylline was used as a proof of concept molecule. This methylxanthine drug, which was previously used to treat respiratory diseases, was one of the earliest to be detectable using RNA.

The protein-based technology was validated by another proof of concept molecule: *N*-acetylneuraminic acid (Neu5Ac), a vital sugar moiety with various roles in important biological processes. As a result of its ubiquity in nature, Neu5Ac has numerous applications in pharmaceutical and food industries, which are currently hindered by the limited production possibilities. The alternative production method via microbial cell factories holds the promise to overcome the limitations for market exploration and entry of Neu5Ac in the lucrative life science market, i.e., cost-effective production and sufficient amounts of Neu5Ac for bio-testing. An overview of this doctoral research is depicted in Figure 1.1, showing the different approaches to enlarge the metabolic engineering toolbox with methods allowing modulation of gene expression in a rational and predictable way.
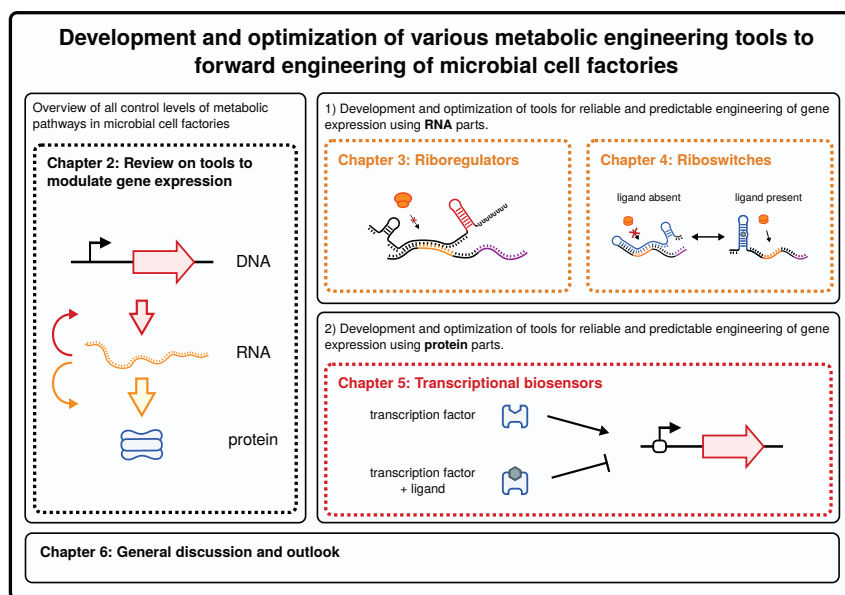
Figure 1.1: An overview of the different chapters described in this PhD dissertation. The different approaches to obtain the main objectives of this doctoral research: development and optimization of metabolic engineering tools to forward engineer at (1) the transcriptional (red) and (2) the translational level (orange).

Since RNA is not longer viewed merely as data carrier, a paradigm shift is gradually ongoing as the exploitation of RNA devices for metabolic engineering is steadily rising. Its enormous potential as standardized, designable, composable and orthogonal device is garnering interest due to its programmable nature. **Chapter 2** gives an extensive overview of the recently developed synthetic RNA devices, their advantages, current limitation in designing and applying them and their emerging applications in the field of metabolic engineering. Furthermore, these RNA based parts and tools are compared to generally used DNA and protein based parts and tools to optimize microbial cell factories. These RNA devices are becoming the necessary armamentarium for contemporary metabolic engineering. Despite several recent success stories the huge potential of RNA devices in terms of speed, scalability, flexibility, orthogonality, portability, modularity, etc. is not yet fully being exploited as several issues of experimental and theoretical nature still persist.

One of these limitations is that no clear design principles exist to build pure RNA devices to modulate translation due to the lack of knowledge on the functionality of these RNA devices impeding their general applicability. To this end, in **Chapter 3** repressing RNA riboregulators are developed from scratch to overcome this lack of knowledge on their functionality enabling forward engineering. More specific, a workflow is established for the *de novo* design of pure repressing riboregulators, referred to as translation inhibiting RNAs (tiRNAs), which repress gene expression by blocking translation initiation. In a first approach, several structural and thermodynamic tiRNA features were evaluated for their influence on the tiRNA functionality using an ordinary least squares regression model. Subsequently, to improve the reliability of *de novo* forward engineering of repressing riboregulators, a sequence-function model was constructed to link tiRNA functionality to the defined tiRNA features. Therefore, both structural and thermodynamic tiRNA features were used in a partial least squares (PLS) regression model, which was evaluated using an independent test set.

In **Chapter 4**, the RNA regulator toolbox is further broadened by an *in silico* approach to create *in vivo* functioning riboswitches using *in vitro* selected aptamers. Traditionally, ligand-responsive RNA devices heavily depend on screening combinatorial libraries and expert knowledge, hindering predictable forward engineering. Moreover, these trial-and-error approaches become impractical as the complexity of programmed biological systems increases, emphasizing the need for automatically designable RNA devices. Here, a computational workflow was created and optimized applying RNA bioinformatics. More specific, the riboswitch characteristics of 5' untranslated regions (UTRs)

were evaluated using both structural and thermodynamical features. Subsequently, this workflow was evaluated by testing several selected riboswitch candidates *in vivo*.

Monitoring cellular behavior and eventually properly adapting cellular processes is key to handle the enormous complexity of todays metabolic engineering questions. In this context, transcriptional biosensors bear the potential to augment and accelerate current metabolic engineering strategies, catalyzing vital advances in industrial biotechnology. The development of such transcriptional biosensors typically starts with exploring Natures richness for ligand responsive transcription factors (TFs) which are subsequently reengineered to control a specific reporter gene or actuator. However, as most biosensors are largely composed of native sequences, techniques to engineer biosensors are missing, which makes tuning of the main characteristics of these (natural) transcriptional biosensors, i.e. the response curve and ligand specificity, in view of specific industrial biotechnology applications, challenging. In **Chapter 5**, various biosensor engineering approaches to create biosensors responsive to Neu5Ac were explored based on a naturally occurring TF. To evaluate the effectiveness of the developed biosensors, a Neu5Ac producing strain was constructed.

In a final chapter (**Chapter 6**), the approaches followed in this doctoral research are assessed and the different tools to optimize microbial cell factories are evaluated. Guidelines to further apply these tools in order to optimize the Neu5Ac production platform and improve their industrial applicability are formulated.

Chapter 2

Literature Review

Contents

This chapter has been published as:

Abstract

Synthetic biology, in close concert with systems biology, is revolutionizing the field of metabolic engineering by providing novel tools and technologies to rationally, in a standardized way, reroute metabolism with a view to optimally converting renewable resources into a broad range of bio-products, bio-materials and bio-energy. Increasingly, these novel synthetic biology tools are exploiting the extensive programmable nature of RNA, vis-à-vis DNA- and protein-based devices, to rationally design standardized, composable, and orthogonal parts, which can be scaled and tuned promptly and at will. This review gives an extensive overview of the recently developed parts and tools for i) modulating gene expression ii) building genetic circuits iii) detecting molecules, iv) reporting cellular processes and v) building RNA nanostructures. These parts and tools are becoming necessary armamentarium for contemporary metabolic engineering. Furthermore, the design criteria, technological challenges, and recent metabolic engineering success stories of the use of RNA devices are highlighted. Finally, the future trends in transforming metabolism through RNA engineering are critically evaluated and summarized.

2.1   Introduction

Long after Crick[11] stated the central dogma of molecular biology, RNA has been viewed merely as data carrier, required to translate genetic information encoded in DNA into proteins. However, the complex role of RNA in the regulation of cellular metabolism has gradually begun to unravel, as over the last decades numerous regulatory RNAs were discovered, modulating cellular responses in various ways. For instance, small RNA (sRNA) regulators modulate protein expression through base pairing, riboswitches react to the availability of certain metabolites and clustered regularly interspaced short palindromic repeats (CRISPR) serves as an immune system[12–16].

Since these revelations, the large potential of RNA is increasingly being exploited spurred by recent technological advances. These cutting edge technologies such as fast and affordable DNA synthesis, improved methods for metabolic pathway assembly and high-throughput sequencing enable harnessing the full potential of RNA in synthetic biology[8–10]. The research field of synthetic biology aims at enabling the systematic engineering of biology through combining standardized basic components[6,7]. This standardization process allows automation and is considered critical to effectively, rationally and predictably engineer these biological systems[17–21]. As a result, considerable efforts have been invested in creating composable, tunable, scalable and reliable parts, which are either stored in repositories (e.g., BioBricks registry[22] and BIOFAB collection[23,24]) or *ad hoc* designed (e.g., ribosome binding site (RBS) calculator[25,26]). The combination of these concepts from synthetic biology with metabolic engineering and systems biology has recently driven major advances in industrial biotechnology[6,27–30].

Several new processes have been developed through metabolic engineering for the sustainable production of an ever increasing number of both bulk and fine chemicals, previously considered impossible to be produced economically from renewable resources[1–5]. The development of production hosts through metabolic engineering typically requires massive tuning of the cellular metabolism in order to optimize the flux towards the desired product. To this end, over the past decade numerous parts and tools have been developed for both rationally and combinatorially *i)* altering the host's native genome, *ii)* optimizing the flux through *de novo* pathways, *iii)* controlling expression using genetic circuits and *iv)* selecting optimal producers from vast libraries. However, whereas most approaches in synthetic biology and metabolic engineering primarily utilized DNA and protein parts to create complex biological systems for various applications[31–37], a paradigm shift is gradually ongoing as the use of RNA components for these purposes is

steadily rising.

Exploiting the enormous potential of RNA is garnering interest due to its programmable nature[38,39]. As such, the underlying structure function relationship makes RNA highly designable, enabling reliable construction of standardized, composable, and orthogonal parts, which can be scaled and tuned at will. Consequently, RNA regulators are effective tools to reprogram existing biological systems or to build completely new ones[40–46]. For example, these RNA devices outperform protein regulators in rewiring cells due to their designability and scalability[47–50]. After years of focusing on techniques to engineer RNA devices with a desired function, this research field is mature enough to start seeing applications becoming reality[51–56]. In this contribution, an overview is given of the current advances in synthetic RNA devices, the current limitations in designing and applying them, and their emerging applications in the field of biotechnological engineering. These RNA parts are becoming necessary weaponry for contemporary metabolic engineering. Furthermore, design principles, technological challenges, and recent metabolic engineering successes of the use of RNA devices are highlighted. Finally, the future trends in transforming microbial metabolism through RNA engineering are critically evaluated and summarized with a focus on metabolic engineering. In this field, techniques are mainly developed in model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* but modern metabolic engineering efforts are increasingly exploiting these technologies in more uncommon organisms with interesting characteristics.

2.2 Modulating gene expression using RNA

Over the years, most approaches in synthetic biology and metabolic engineering primarily utilized DNA and protein parts to create complex biological systems for various applications[31–37]. These earliest efforts mainly consist of modulating protein expression through transcriptional regulators on the DNA level to achieve the desired phenotypic behaviour. To properly construct these and more complex systems, such as genetic circuits and metabolic pathways, predictable and reliable techniques controlling protein synthesis are needed. The primordial importance of these tools is clear as the ever increasing complexity of genetic circuits and metabolic pathways makes a trial-and-error approach infeasible[25]. In this context, forward engineering of protein-based transcriptional regulators is challenging due to the complex DNA-protein interactions involved, which are hard to predict because of the numerous degrees of freedom involved[57]. In comparison, RNA serves as a great alternative building block with superior programma-

bility and scalability. For instance, translation initiation is strongly linked to the RNA structure in the translation initiation region, containing the RBS[58]. Additionally, intrinsic transcription termination is the result of the formation of a stem loop, which results in the disruption of the binding between RNA and template DNA[24,26].

Consequently, the ability to engineer gene expression largely depends on the knowledge of the sequence-structure relationship. The RNA structure consists of secondary and tertiary interactions, of which the former are thermodynamically more stable and form much faster, leading to a sequential folding process[59]. This allows decoupling of secondary and tertiary structure prediction, facilitating modelling efforts. These approaches are further supported by advanced RNA structure determination techniques, which currently allow high-throughput native structure determination *in vivo*[60,61]. Therefore, biophysical models are being developed with ever increasing detail, up until the recent incorporation of the biologically relevant cotranscriptional RNA folding process[62,63]. Consequently, RNA folding is one of the best modelled processes in biophysics, enabling reliable prediction of RNA interactions[64]. As such, RNA secondary structure prediction tools are used in forward engineering of various RNA parts[25,26,49,65–67]. This kind of microbial engineering typically revolves around modulating specific gene expression to attain precise goals.

In metabolic engineering, gene activity is typically removed or reduced to minimize certain undesired biochemical conversions, hereby maximizing the production of value-added compounds. To achieve maximal product formation, several genes related to the biosynthetic pathway require optimization, which results in an efficient conversion of substrates, intermediates and cofactors into the desired product[28,29]. Traditional strategies rely on DNA modifications to modulate gene expression but RNA can serve as a great alternative due to its modularity and programmable nature[68–70]. Over the years, several RNA-based tools were developed to decrease or increase gene expression either prior or posterior to transcription. Figure 2.1 compares protein-based gene expression modulation with RNA-based technology, controlling transcription and translation. In the sections below these various RNA devices which modulate expression on the different controlling levels are discussed and compared in depth.
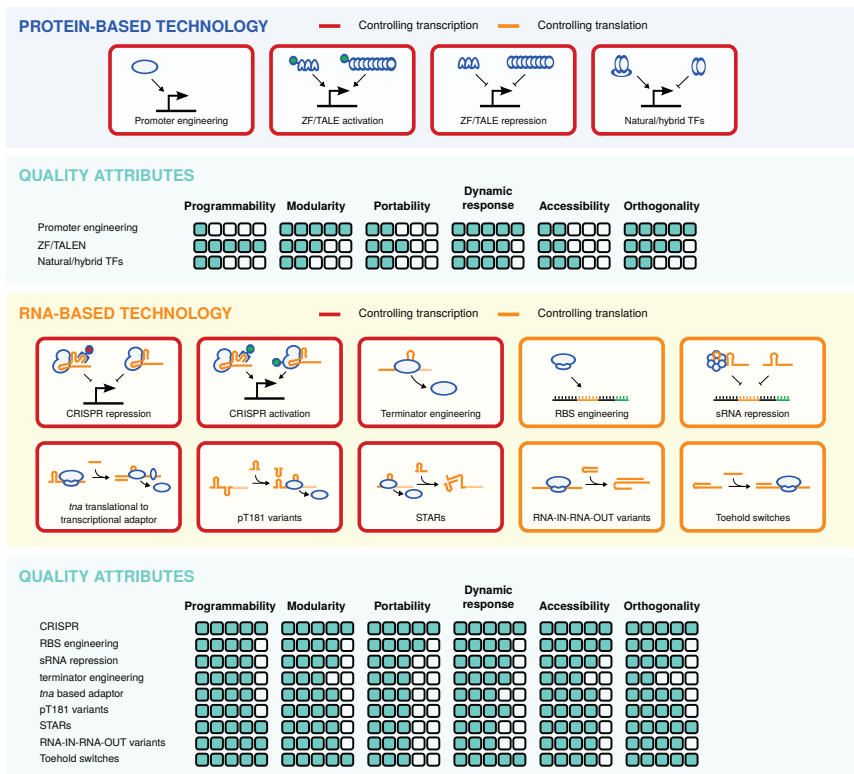
Figure 2.1: Comparison between protein- and RNA-based gene expression modulation technologies to control translation and transcription. Protein-based technologies mainly comprise promoter engineering[71], zinc finger (ZF) & transcription activator-like effector activation[72,73], ZF & transcription activator-like effector nuclease (TALEN) repression[72,74], and natural/hybrid transcription factors (TFs)[75,76]. RNA-based technologies comprise clustered regularly interspaced short palindromic repeats (CRISPR) repression[53,65,77], CRISPR activation[53,77], terminator engineering[24,26], ribosome binding site (RBS) engineering[23,25], small RNA (sRNA) repression[70,78], tna translational to transcriptional adaptor[79], pT181 variants[47], small transcription activating RNAs (STARs)[80], RNA-IN-RNA-OUT variants[49], and toehold switches[81]. The quality attributes evaluated are programmability (degree to which the function of the part is controllably adjustable), modularity (degree to which components can be separated and recombined), portability (degree to which technology is interchangeable between different host organisms), dynamic response (obtainable expression range convered by the technology), accessibility (degree to which the technology is easily applicable by non-experts), orthogonality (degree to which parts of the same family are tuned to noninterference while maintaining the same function).

2.2.1 Translational repression using synthetic sRNAs

The ability to specifically reduce gene expression is an indispensable tool for microbial engineering. A large group of naturally repressing sRNAs are *trans*-encoded, which are expressed on another location than where they exert their function. This property, along with the programmable nature of RNA, makes these an interesting target for synthetic gene regulation.

Early gene silencing efforts focused on controlling post-transcriptional processes, more specifically on designing RNA molecules which hybridize the translation initiation region. It is long known that the secondary structure in this region around the RBS heavily influences translation efficiency[58]. The ease in designing hybridizing RNA molecules drove several gene silencing efforts mainly comprising long antisense RNAs fully complementary to the target site[82–84]. Antisense RNA allows precise downregulation, hereby redirecting metabolite fluxes towards the metabolic phenotype of interest[78,85,86]. Although these approaches allow gene silencing, they generally lack applicable design principles and have variable repression efficiencies. To improve the inhibitory efficiency of synthetic *trans*-expressed sRNAs, several attempts employed the binding site of the naturally occurring *E. coli* Hfq chaperone protein, which naturally facilitates base pairing between sRNA and the target mRNA resulting in translational blockage or mRNA decay[87].

Man *et al.*[88] first described a semi-rational approach comprising an *in silico* screening of sRNA candidates containing a target site, a Hfq binding site, and an intrinsic terminator. After performing this computational analysis, *in vivo* assays showed that 4 out of 16 sRNAs candidates repressed gene expression with more than 60 %. Further developments focused more on screening efforts by using various naturally occurring Hfq-binding motifs to create combinatorial libraries. These libraries were subsequently used to select sRNAs which effectively repress endogeneous genes employing a fluorescent reporter gene translationally fused to the targeted mRNA. In contrast to the previously mentioned efforts, Na *et al.*[70] reported a general methodology to rationally design synthetic sRNAs which translationally modulate gene expression. To allow effective gene repression, these synthetic sRNAs contain a naturally existing Hfq-binding scaffold and a target-binding sequence complementary to the target site. Here, this target site comprises the translation initiation region with the aim of preventing ribosome binding. Optimal design rules were developed hereby enabling forward engineering of efficient *trans*-acting and Hfq-dependent synthetic sRNAs[89]. These artificial gene regulators form an alternative to traditional DNA-based gene knockout experiments by allowing simple

and fast gene knockdowns. In addition, this technique also allows downregulation of genes essential for proper cell function, which is often impossible by gene deletion. Another advantage originates from the plasmid-based expression of these artificial sRNAs, which allows high-throughput studies of knockdowns in various strains. As such, synthetic sRNAs were first used to test multiple gene knockdowns in various *E. coli* strains for improved tyrosine production (up to 2 g/L). Second, a sRNA-based strategy was used to perform large-scale target identification (up to 130 genes) followed by gene expression optimization to improve cadaverine production resulting in a 55 % increase in production[70]. These proofs of concept show the advantage of using translational RNA regulators over traditional, laborious and time-consuming genetic engineering strategies due to the ease in design and implementation, which allows high-throughput experiments on a shorter time scale. Moreover, in contrast to DNA-based gene knockouts, knockdowns using synthetic sRNAs are inducible and reversible, showing the advantages of these RNA regulators.

2.2.2 Programmable RNA-guided transcriptional modulation

Until recent, most RNA-based gene modulation efforts targeted the translational level, however, this changed due to recent advances in genome engineering techniques[90–92]. Various genome engineering methods are based on the naturally occurring CRISPR, which spurred the development of more programmable gene silencing techniques on the transcriptional level. CRISPR systems are a family of DNA repeats, which provide acquired immunity against viruses and plasmids in most archaeal ($\sim$ 90 %) and bacterial ($\sim$ 40 %) genomes[16]. Such specific immune systems are often adjacent to *cas* genes (CRISPR-associated) typically encoding endonucleases, helicases, polymerases and polynucleotide binding proteins. Depending on the type of CRISPR/Cas system, these Cas proteins can form a complex, in various ways, with CRISPR RNA (crRNA) allowing specific target recognition[16,90]. These crRNAs need maturation which occurs differently in the three existing types of CRISPR/Cas systems. Type I and III systems require specialized Cas endonucleases to process pre-crRNAs to mature crRNAs[90]. In contrast, type II systems rely on *trans*-activating crRNA (tracrRNA) complementary to repeat sequences in pre-crRNA to trigger RNase III processing to form mature crRNA[90,93]. Because this mechanism mainly relies on RNA interactions, which can be easily predicted, reprogramming these type II systems for genome engineering or gene expression modulation is straightforward. Therefore, type II CRISPR-Cas systems were used for site-specific genome editing using various Cas9 endonucleases, hereby introducing permanent mod-

ifications in various genomes[90–92]. This type of genome engineering provides a more efficient and easier way to construct chromosomal alterations compared to traditional protein-based techniques, i.e., ZFs and TALENs[94].

To use type II CRISPR systems for custom gene silencing, an inactivated Cas9 endonuclease from *Streptococcus pyogenes* allows specific RNA-guided transcriptional blockage by co-expression with a small guide RNA (sgRNA)[65]. This so called clustered regularly interspaced short palindromic repeats interference (CRISPRi) system allows tunable and efficient repression of specific genes without detectable off-target effects by interfering with transcriptional elongation and RNA polymerase binding. In addition, CRISPRi allows specific interference with TF binding sites[65,95].

Similarly to CRISPRi, Bikard *et al.*[77] used an inactivated Cas9 protein (dCas9) to block transcription of specific genes. In addition to transcriptional repression, genes can be activated by translationally fusing the ω-unit of RNA polymerase to dCas9. After optimizing the distance between binding site and promoter, a dCas9-ω C-terminal fusion protein showed up to 23-fold activation in *E. coli*[77]. Similar strategies using protein fusion of dCas9 and effector domains also enabled gene activation or repression in eukaryotic cells[96,97]. This approach using fusion proteins restricts modulation to one direction, either increasing or decreasing gene expression. To allow distinct types of regulation on multiple gene targets at once, Zalatan *et al.*[53] extended sgRNAs to include RNA scaffolds. These engineered RNA modules have distinct protein-docking sites which allow the construction of protein architectures *in vivo*[98]. By combining these RNA motifs with CRISPR sgRNA, modular RNA scaffolds can be created enabling custom targeting and tunable regulation. These so called CRISPR RNA scaffolds were used to program the metabolic flux in the branched violacein pathway[53]. The synthesis of the four possible end-products of this pathway is controlled by introducing two different control points using desired regulation at specific loci. The CRISPR RNA scaffold regulation platform exemplifies the modularity and programmability of RNA to engineer microbial gene expression: chromosomal targets are recognized by sgRNA-DNA base pairing and modular RNA scaffolds recruit specific proteins. In comparison to other targeted gene regulation techniques, RNA-guided CRISPR gene modulation excels because of the easy design and simple construction of sgRNAs allowing simultaneous modulation of several targeted genes within a short timespan. Additionally, in comparison to traditional DNA-based gene modification techniques, CRISPR-based regulators can be inducible and are reversible, enabling dynamic control over gene expression. Recently, a combinatorial assembly method containing CRISPR arrays was used to simultaneously downregulate

multiple genes, resulting in improved production of malonyl-CoA-derived chemicals[99]. Similarly, CRISPRi was used to increase poly(3-hydroxybutyrate-co-4-hydroxybutyrate) production by repressing multiple genes[100].

However, CRISPR requires a specific protospacer adjacent motif (PAM) sequence for each Cas9 which limits the availability of target sites. In addition, the 14-nt recognition sequence is too short to obtain unique sequences in large genomes (for instance, the human genome)[95]. These drawbacks can be counteracted by employing Cas9 homologs with different PAMs, which expands the number of possible target sequences[91,95,101]. To further overcome this restriction, a set of fully orthogonal Cas9 proteins was characterized, which showed fewer PAM requirements for the *N. meningitidis* ortholog, making it a suitable starting point for protein engineering efforts. The way towards engineering enhanced CRISPR technology was further paved by functional and structural studies on the sgRNA, target DNA, and Cas9 complex[102,103]. Additionally, future efforts characterizing additional Cas9 orthologs are expected to further expand engineering capabilities[104].

2.2.3   Predictably tuning gene expression using RNA parts

Constructing custom made biological system calls for more engineering tools than technologies interfering with gene expression alone. For instance, creating microbial cell factories requires precise optimization of gene expression to balance the production pathway[5,28,29,51,105,106]. Here, traditional trial-and-error approaches become impractical as custom programmed cells execute increasingly complex functions[25]. To facilitate the development of these systems, tools and parts allowing reliable forward engineering are needed. A quantitative framework for the forward engineering of complex biological systems requires a thorough knowledge on how gene expression is regulated from DNA to protein. On the transcriptional level, various proteins, such as RNA polymerase and TFs, play an essential role. This makes linking the promoter sequence to protein production over a wide range of conditions very challenging due to the difficult physical modelling of DNA-protein interactions, which is a result of the large number of degrees of freedom involved[57]. Although several approaches successfully model this transcriptional process to some extent, the complex nature of transcription remains hard for forward engineering of promoter sequences[57,107–109]. As stated before, RNA can serve as a versatile alternative, which allows accurate control of protein expression. Moreover, RNA practically bridges DNA to protein, controlling intrinsic transcription termination and translation initiation. The latter enables precise control of translation, the former is required to reliably stop transcription.

Predictably controlling intrinsic transcriptional termination

To avoid undesired transcriptional readthrough, absolute transcription termination is essential for predictable microbial engineering, which requires strong terminators. To this end, effective termination requires efficient dissociation of the transcription elongation complex, which mostly occurs using intrinsic terminators. In bacteria, this type of termination comprises a short hairpin followed by a U-rich sequence, which in contrast to Rho-dependent terminators do not require any assistance of proteins. As a consequence, intrinsic terminators are much more designable, although details of the molecular mechanism of Rho-independent termination, despite decades of research, remain unknown[110]. The ever increasing size of custom biological systems demands however a large set of available terminators. To enlarge this terminator collection, numerous natural and synthetic terminators were systematically characterized to quantify their design constraints[24,26]. Using this data, biophysical models mainly comprising RNA thermodynamics simulation, including simulations of co-transcriptional RNA folding dynamics, were developed. This resulted in a substantial number of new strong terminators and a better understanding of the sequence-function relationship, which enabled the more reliable forward design of intrinsic terminators[24,26].

Programmable control of translation initiation rate

In addition to being essential in transcription termination, RNA architecture plays a crucial role in translation initiation. For years, research focused on elucidating important factors influencing the translation initiation rate, showing that the region surrounding the RBS is determinative for protein production. Translation initiation is mostly the rate limiting step, which is influenced by the start codon, the affinity for 16S rRNA, spacing between start codon and Shine-Dalgarno sequence and mRNA thermodynamics[58,111,112]. This thorough knowledge of the working mechanism, allowed the development of accurate biophysical models[25,113,114]. Pioneering work in this field showed the predictive capabilities of thermodynamic models, mainly based on secondary structure predictions, allowing precise quantification of the translation initiation rate[25]. As such, this mathematical framework enables forward design of RBSs, showing accurate prediction within a factor of 2.3 over a range of 100,000-fold in protein expression[25]. This approach was later extended using newly acquired insights in the working mechanism of ribosome binding upstream of the Shine-Dalgarno sequence[115]. Other modelling approaches showed similar predictive power, further extending the capabilities to rationally design synthetic RBSs[113,114]. However, using these RBSs in different ge-

netic contexts results in different protein levels, emphasizing the need for *ad hoc* RBS design[25,114]. To facilitate forward engineering, software implementations are available enabling forward and reverse engineering[25,114,116]. In conclusion, these approaches to model translation initiation allow forward engineering of RBS parts, although the predictability of these techniques is still limited[25,115].

Aside from rational designs with a specific protein expression level, these tools successfully generate reliable RBS libraries for combinatorial metabolic engineering[117]. These techniques were successfully used to combinatorially optimize pathways in various host organisms[118–120]. By using these highly informative RBS libraries, the whole search space in multi-protein optimization problems can be efficiently covered, avoiding a combinatorial explosion. Moreover, to effectively search through multi-protein expression spaces with custom search parameters, an algorithm was developed to design sequences based on RBS Library Calculator, which allows modelling the sequence-activity relationship of a biosynthetic pathway in various bacteria[51]. Subsequently, model predictions were used to design new pathway variants with the aim of finding the expression optimum. As proof of concept, this approach was successfully used to balance the neurosporene pathway, finding optimal protein expression levels hereby removing metabolic bottlenecks[51]. In a similar approach, Ng *et al.*[121] constructed an improved Entner-Doudoroff pathway by employing the RBS Library calculator to efficiently search the 5-dimensional protein expression space. This combinatorial approach resulted in a 25-fold increased NADH regeneration rate, which subsequently resulted in a 97 % increased terpenoid titer[121]. These kind of effective and reliable search methods for multi-protein systems, mainly based on RNA thermodynamics, play a primordial role in modern microbial engineering to precisely balance gene expression levels[5,28,29,105,106]. Mathematical modelling of translation initiation rates allows constructing reliable RBS libraries. In contrast, forward engineering of RBS parts using these methods show limited predictability, a result of imperfect knowledge on the biophysical rules regarding translation initiation[25,115]. Moreover, large scale analysis of common promoter and RBS parts showed RNA and protein levels within twofold of the expectations in 80 % and 64 % of the time, with larger deviations for the worst constructs[18]. The scale of this approach enables screening genetic parts for desired behaviour but also emphasizes the need for more predictive and composable parts to reduce the number of design-build-test cycles. Various methods, employing RNA devices, aim to resolve this bottleneck, hereby enabling a bottom-up approach to building complex genetic circuitry[23,122,123]. To improve the predictability of gene expression, genetic elements can be physically sep-

arated from their context by specifically cleaving mRNA. This physical separation was accomplished using CRISPR and ribozymes[122,123]. The former comprises RNA-guided cleavage of mRNA using CRISPR technology, while the latter autocatalytically cleaves the mRNA, leading to predictable gene expression. However, these approaches do not rule out the largest source of variability, namely interactions between the 5' untranslated region (UTR) and the gene of interest[124]. Consequently, Mutalik *et al.*[23] mainly aimed to further increase the reliability of translation initiation by using bicistronic designs. In this translationally coupled architecture, which overlaps the downstream gene by 1 base pair, the ribosome synthesizes a short leader peptide by binding an upstream RBS. Subsequently, translation is reinitiated by a second RBS, which is entirely encoded in the coding DNA sequence (CDS) of the leader peptide, resulting in reliable protein expression of the gene of interest. Among others, these bicistronic designs were used to create various libraries of genetic elements controlling transcription and translation initiation. Curation of these parts resulted in a set of genetic elements showing more reliable protein expression across a 1,000 fold dynamic range. This particular collection together with the previously mentioned technologies emphasize the indispensable role of RNA in precisely tuning translation.

Rational engineering of mRNA stability

Besides being crucial in transcription termination and translation initiation, RNA architecture plays a vital role in mRNA stability. Early metabolic engineering efforts showed that introducing hairpins in the 5' UTR prolongs the half-life of mRNA[125]. Using this finding, Carrier & Keasling[126] created 5' UTR libraries, resulting in variable mRNA halflives, showing the importance of RNA structure on mRNA degradation. As such, RNase cleavage sites were used in combination with other intergenic control mechanisms to control gene expression. These libraries of tunable intergenic regions (TIGRs) vary over a 100-fold range in expression and enabled the balancing of the mevalonate pathway, resulting in a sevenfold increase in production[106]. State-of-the-art RBS designs incorporate various design factors to minimize mRNA stability changes, including shortening unprotected mRNA regions and removal of potential RNase binding sites[51]. To conclude, RNA allows programmable control over protein expression, which is possible due to the in-depth knowledge on transcriptional termination, translation initiation, and mRNA stability.

2.3 Towards building increasingly complex genetic circuitry using RNA parts

In addition to gene expression modulation, RNA serves as an ideal building block to construct predictable biological circuits due to its programmable nature. These highly specific RNA-RNA interactions enable the facile bottom-up construction of genetic circuits. Isaacs *et al.* [127] pioneered in the field of RNA genetic circuitry by creating engineered riboregulators, comprising a *cis* repression module by inserting a complementary sequence upstream of the RBS. The resulting stem loop structure requires *trans* expression of small noncoding RNAs to relief repression of the downstream gene. This early approach shows the potential of RNA as a scalable component for building biological systems. In addition, these kind of RNA devices encapsulate various properties required to predictably construct biological systems [17]. Based on the pioneering work of Isaacs *et al.* [127], cells were reprogrammed using RNA to count biological inputs. Here, synthetic riboregulators were used to construct a genetic counter, which could be recorded using recombinases [48]. To further elaborate the riboregulator technology, a programmable platform was created, allowing orthogonal control of multiple genes in physiological conditions. These RNA devices were used to tightly control expression of various proteins, enabling the construction of various microbial behaviours [128]. The broad range of applications of these synthetic riboregulators was shown among others, by building a Boolean logic controlled kill-switch and a genetic switchboard, which regulates the carbon flux through the three *E. coli* glucose catabolic pathways [128,129]. In addition, riboregulators were used to develop a new biocontainment strategy with multilayered safeguards, which tightly control multiple essential genes in *E. coli*, enabling safe use of genetically modified organisms [130].

Building increasingly complex biological systems requires an expansion of the number of RNA regulators, which additionally need to be highly specific and unable to cross react. To increase the amount of available orthogonal RNA regulators, Mutalik *et al.* [49] rationally created variants of the naturally occurring RNA-IN-RNA-OUT antisense system of IS10 (see Figure 2.2D). This translational regulation comprises an antisense RNA (RNA-OUT), which blocks the RBS on the RNA-IN mRNA, hereby repressing transposase expression. In this study, 23 sense RNA-IN and 23 antisense RNA-OUT mutants were constructed and subsequently tested. Subsequently, model predictions were successfully used for the forward engineering of orthogonal riboregulators, hereby showing the ability to predictably build RNA devices using data-driven sequence-activity models [49]. De-

spite the advantages of using RNA, the aforementioned conventional riboregulators have a limited dynamic range in comparison to protein-based genetic switches[131]. This limitation is predominantly caused by imposed sequence constraints in the design of these riboregulators, which involves blocking the RBS by base pairing. These riboregulators are required to contain a RBS sequence, which reduces the number of sequences suited as orthogonal riboregulator. To overcome this limitation, Green *et al.*[81] reported a new type of riboregulator, the toehold switch, which forms a stem in the 5' UTR downstream of the RBS, without involving the RBS (see Figure 2.2E). This alternative design expands the sequence space to obtain functional riboregulators, allowing the construction of switches with improved orthogonality and activation. Moreover, a first-generation library was analyzed to derive design features, enabling data-driven forward engineering of highly orthogonal toehold switches with dynamic ranges comparable to protein-based transcriptional regulators. These RNA-based genetic switches were successfully used to regulate endogeneous genes, to sense sRNAs and, to construct complex genetic circuits[52,81]. In addition, toehold switches were used to create various programmable *in vitro* diagnostics, such as glucose sensors and strain-specific Ebola virus sensors[52,81]. Also, to facilitate field use of RNA regulators, a paper-based platform was developed using commercially available cell-free systems, enabling distribution of synthetic biology tools beyond laboratory environments[52]. In conclusion, the previously mentioned efforts show the versatility of RNA regulated genetic circuits, which have the advantage of programmability. RNA regulators can be simulated and designed *in silico* into functional genetic circuits. Even without data-driven models, functional riboregulators can be designed *de novo* using simple thermodynamic models. Moreover, using an objective function composed of factors derived from literature, mainly formation and activation energy involved in riboregulation, sufficed to computationally design functional RNA devices[132].

All previously mentioned riboregulators control translation through occlusion of the RBS, hereby modulating the translation initiation process. A different type of riboregulators are hammerhead ribozymes, which control translation initiation through conditional self-cleavage. These RNA regulators can be redesigned to control gene expression depending on small *trans* expressed RNAs, which is important for the construction of higher-order genetic circuits[133]. Shen *et al.*[134] developed a scalable automated methodology to design RNA-mediated signal transduction systems using so called regazymes, comprising ribozymes and aptamers, capable of sensing both sRNAs and small molecules. These *de novo* designed RNA devices are used in various complex gene circuits

built for a custom purpose, once again showing the programmable capabilities of RNA over protein parts.

Other types of RNA devices regulate transcriptional elongation, which, in contrast to translational regulators, can control synthesis of multiple RNAs, coding or noncoding. In addition, transcriptional regulators are inherently composable but difficult to engineer due to the poorly understood mechanism of RNA polymerase interference[79]. To eliminate this bottleneck, Liu *et al.*[79] developed an adaptor, which allows controlling transcription using translational riboregulators. This conversion is accomplished using the naturally occurring *tna* regulation, mainly comprising a leader peptide (*tnaC*). Full translation of this peptide blocks a Rho factor binding site, allowing transcription of the downstream genes in the operon. This mechanism was combined with translational riboregulators, forming the adaptor from translational to transcriptional control. Subsequently, this adaptor successfully converted the previously described RNA-IN-RNA-OUT translational riboregulators into transcriptional controllers[49]. Additionally, coupling this system to an aptamer, an oligonucleotide sequence which binds a specific small molecule, resulted in ligand dependent transcriptional regulation[79]. Another approach designed orthogonal variants of the natural occurring antisense-mediated transcription attenuation system from the pT181 plasmid (see Figure 2.2F), independently controlling transcription. These transcriptional riboregulators were used to create logic gates, allowing the propagation of antisense RNA signals in complex genetic circuits[47]. In order to increase the number of available orthogonal RNA transcriptional attenuators, Takahashi & Lucks[135] fused RNA binding regions from naturally occurring antisense regulated translational regulators to transcriptional attenuator from plasmid pT181. As such, 11 transcription attenuators were created and tested, resulting in a maximum of 7×7 mutually orthogonal riboregulators and design principles for these transcriptional regulators[135]. Similarly, Chappell *et al.*[80] used naturally occurring transcriptional attenuation systems to create STARs (see Figure 2.2G). Although there are no known naturally occurring sRNAs that activate transcription, four highly orthogonal regulators were engineered that disrupt intrinsic terminator formation. In addition, a sequence-function model was derived from systematic sequence modification, enabling forward engineered STARs to target terminators found on the *E. coli* genome for activation and the construction of transcriptional logic gates[80].

Various types of orthogonal riboregulators, controlling transcription or translation, were used to construct larger genetic circuits, which include multiple types of Boolean logic gates[47,80,81,128,132]. Besides these genetic circuits built from purely RNA regulators, the

previously described CRISPR/Cas technology can also be programmed to perform the same function. The potential of CRISPR/Cas transcription regulation was first shown in human cells, enabling the construction of complex transcriptional cascades[136]. In a similar strategy, tools using CRISPR/Cas were integrated with multiple RNA regulatory devices to construct programmable complex genetic circuits[137]. As a first real life application of CRISPR/Cas genetic circuitry, Liu *et al.*[138] constructed a modular AND gate to interpret cellular information from two promoters, hereby detecting bladder cancer cells. Subsequently, a gene network was constructed to inhibit bladder cancer cell growth, induce apoptosis and decrease cell motility[138]. After these early efforts in human cells, Nielsen & Voigt[139] constructed the first complex genetic circuits in microbial cells by using five synthetic promoters, which are orthogonally repressed by specifically designed sgRNAs. This enables up to 440-fold on-target repression, with minor off-target effects. Subsequently, these RNA-guided devices were used to construct various logic gates, including the Boolean-complete NOR gate, which allows building any computational device. As proof of concept, these basic components were used to program more complex behaviour containing up to 3 Boolean gates[139]. To conclude, the CRISPR/Cas and previously mentioned riboregulator technologies are indispensable tools for the construction of *ad hoc* tunable genetic circuits by fully harnessing the programmable nature of RNA. These recent advances show RNA parts as a great alternative to traditional protein-based genetic circuitry. Figure 2.2 shows a functionally complete NOR gate constructed using RNA or protein parts. In addition, several previously discussed RNA components controlling transcription and translation are shown.

As outlined above, enormous progress has been booked in recent years on the development of both orthogonal and complex/nonlinear regulators, tunable at will to obtain the desired dose-response curves. These features are considered essential for the construction of complex genetic circuits[141]. In this context, the development of RNA-based nonlinear logic Boolean OR, NOT, AND, AND NOT and NOR gates has been of utmost importance[80,132]. Such types of cooperative and/or complex RNA-based control have also been achieved using, for example, tandem riboswitches[142–144], despite the fact that response curves of sRNAs tend to be linear[145]. Hence, increasingly complex regulatory mechanisms are and will be designed by assembling RNA in a modular fashion to attain nonlinearly responding circuit components. This will render the use of transcription factors for the construction of sophisticated genetic circuits superfluous[142]. It should however be mentioned that sRNAs and transcription factors can work seamlessly together in hybrid synthetic circuits or that the differences in sRNA and transcription factor charac-
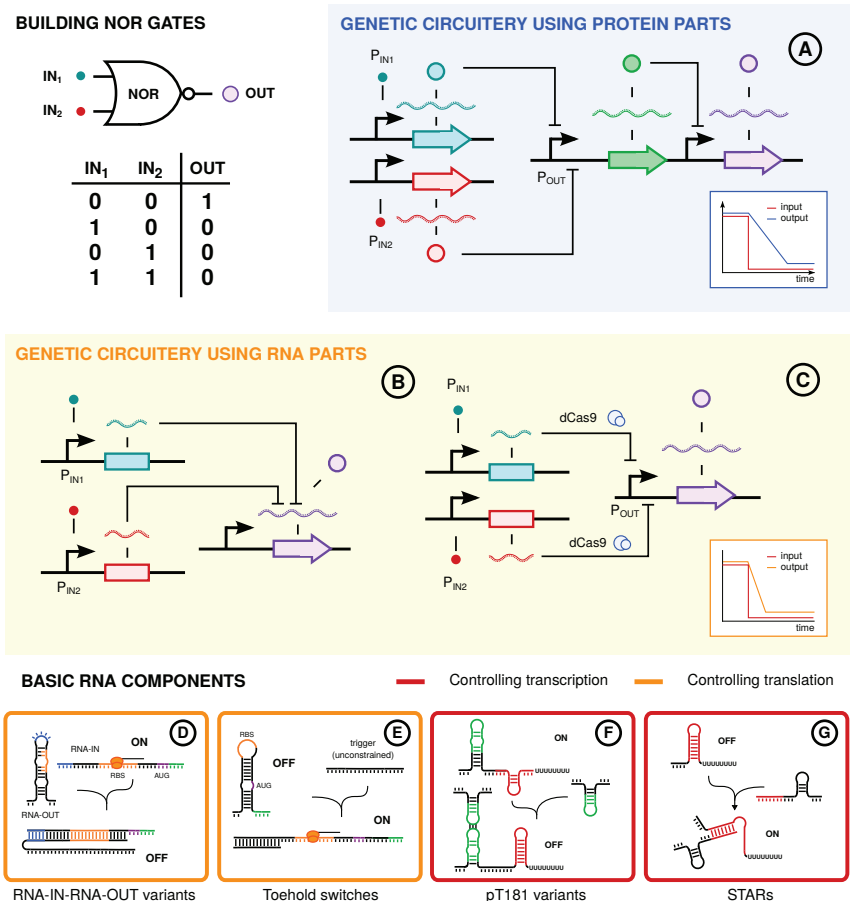
Figure 2.2: NOR gates built using protein or RNA parts. This functionally complete Boolean operator can be constructed using A) naturally occuring protein parts [140], B) solely riboregulators [79], and C) RNA-guided clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 [139]. Below, recently developed basic RNA components are depicted allowing repression and activation of both transcription [47,80] and translation [49,81]: D) RNA-IN-RNA-OUT variants [49], E) toehold switches [81], F) pT181 variants [47], and G) STARs [80].

teristics can be exploited to obtain the desired regulatory behaviour [145]. In addition, it has been demonstrated that multiple RNA-based signals can be used simultaneously and independently [104,132,146–148], i.e., orthogonally, which is mandatory for large circuits as they typically require multiple-input and multiple-output signals. Furthermore, meth-

ods to fine-tune these circuit components with a view to obtaining biologically relevant dose-response curves have been elaborated[149,150]. Hence, it is considered only a matter of time before more sophisticated (hybrid) RNA-based circuits will be constructed. In this respect, such RNA-based feedback[151,152] and feedforward loops[153] constitute an integral part of nature's toolbox to regulate cellular processes.

However, as the complexity of these circuits increases, so does the need to rationally design and fine-tune them with a view to obtaining the desired response. This will not only require extensive model-based optimization to evaluate the dynamics of the genetic circuit *in vivo*[154], but also the collection of *in vivo* and *in vitro* omics measurements to characterize the circuit and its various components in detail[155] and to properly identify the aforementioned models. For example, to this end, software tools such as AutoBio-CAD[20] have been developed, enabling the automated design of such genetic circuits.

2.4   Ligand-dependent gene expression using RNA

When control involves interaction with small molecules protein-based technologies are currently typically used to engineer microbial systems. These technologies mostly comprise transcription-factor-based and Förster resonance energy transfer (FRET) biosensors, which are constructed from naturally occurring binding domains[156,157]. Other prime examples of molecule sensing RNA regulators are riboswitches. These genetic switches are completely folded RNA molecules which control gene expression in response to changes in metabolite concentrations[15,158–160]. In nature, riboswitches mostly reside in the 5' UTR of mRNAs and typically exert regulation in a *cis* configuration[159,161]. These regulatory elements, controlling transcription or translation, comprise two structurally linked domains: an aptamer and a gene expression platform. The highly conserved aptamer domain selectively binds the target metabolite with high affinity. Generally, the aptamer part conformationally rearranges upon ligand binding leading to RNA conformational changes in the expression platform which in turn modulate downstream events. This way of controlling gene expression is widespread in nature and a broad range of riboswitch classes have been discovered over the years[15,158,162]. However, synthetic conditional gene expression systems were created several years before the first reports on natural riboswitches[163]. These artificial riboswitches are typically based on aptamers selected by systematic evolution of ligands by exponential enrichment (SELEX)[164,165]. This generally applicable procedure shows the major advantage of using RNA over protein-based technologies, which are limited by the availability of

protein sequences that specifically bind the ligand[156,157,166]. In contrast, over the years, SELEX has been employed to develop a wide range of oligonucleotides that bind to several classes of molecules[167–169]. A number of these selected aptamers have been used to control gene expression of which Werstuck & Green[170] first described a synthetic riboswitch using an RNA aptamer that specifically binds a cell-permeable Hoechst dye. This earliest description of synthetic RNA regulator marks the beginning of a new practice in synthetic biology where RNA is engineered to sense a molecule and subsequently modulate gene expression[41,42,163,171]. The first described riboswitches were built by trial and error, as such, various aptamers were inserted downstream the gene of interest and subsequently tested[170,172–175].

Until now, the lack of knowledge on the working mechanisms of riboswitches and the complex nature of cellular processes makes fully rational designs challenging. Consequently, the development of gene switches heavily depends on high-throughput screenings of combinatorial libraries[176–183]. Buskirk *et al.*[184] pioneered with the *in vivo* evolution of a RNA regulator to increase gene expression in *S. cerevisiae*. This approach was later extended to find riboswitches through selection by fusing a RNA regulator to a known tetramethylrosamine (TMR) aptamer[176]. The huge potential of genetic selection systems for synthetic riboswitch development was again shown by Desai & Gallivan[177], who used the theophylline aptamer[185] to construct the first RNA gene switch in *E. coli* using β-galactosidase, hereby enabling blue white screening. Subsequently, this selection was used to select an improved riboswitch from larger combinatorial libraries[178]. The Gallivan lab later proposed several other selection procedures with higher throughput based on cell motility (approximately 600,000 cells) and fluorescence activated cell sorting (up to 10^9 cells per day)[179,180]. Nomura & Yokobayashi[181] developed and optimized an alternative method relying on dual selection, which was used to create complex riboswitches capable of sensing and responding to two small molecules according to Boolean logics[182,186]. An entirely different screening approach reengineered a natural riboswitch into several orthogonal selective riboswitches[187]. This technique was used to tunably co-express two proteins in an orthogonal way and proved to function in several organisms[188,189]. Most selection methods for *in vivo* riboswitches use aptamers either found in nature or selected using SELEX. Although SELEX experiments are well established, few methods exist that select aptamers with *in vivo* regulatory activities[183,190]. To fill this void, Weigand *et al.*[183] combined *in vitro* selection and *in vivo* screening, resulting in a riboswitch for neomycin.

Although most riboswitches were developed using a selection procedure, an increasing amount of effort is invested in rational approaches [149,150,191,192]. The first rationally constructed gene switch, which was based on a ligand-dependent one-nucleotide slipping mechanism model, showed the potential of rationally designing riboswitches due to the inherent programmable nature of RNA molecules [191]. Another rational approach describes a specific type of riboswitches, the antiswitches. These small *trans*-acting RNAs are ligand-dependent riboregulators which can be designed to turn gene expression 'on' or 'off' when a ligand is available. The mode of regulation of an antiswitch is based on conformational dynamics of the RNA structures. This enabled specific tuning of the antiswitch by changing its thermodynamic properties, in particular the stability of two stems (an aptamer and an antisense stem). Several forward engineering constructs followed, showing that altered thermodynamical properties of riboswitches influences the dynamic behaviour in a predictable way. Besides this predictability, the modularity of the platform was shown using different aptamers and reporter genes [149]. In addition to the antiswitch platform, Smolke and collaborators developed several other ligand dependent RNA-based gene regulatory platforms. For instance, more ligand dependent riboregulators were rationally created based on hammerhead ribozymes, RNA interference and RNase III cleavage, showing the large variety of available ligand responsive RNA devices [150,193,194].

These early examples of rationally constructed riboswitches show the potential of programming RNA to link metabolite concentrations to gene regulation. However, tuning these specific RNA devices remains challenging. To overcome this limitation, Beisel *et al.*[150] developed a mathematical framework by quantitatively modelling the kinetics of riboswitch function. It was shown that riboswitch performance is dictated by the competition between reversible and irreversible rate constants. The former reflect conformational switching and ligand association while the latter comprise irreversible processes such as mRNA degradation and destabilization [192]. These design principles for riboswitches, in combination with recently unveiled mechanistic insights into the structure-function relationship, spurs the *de novo* design of riboswitches [178,180,195]. Consequently, recent approaches show the feasibility of building custom ligand binding RNA devices from scratch [66,196–198]. As such, secondary structure predictions are used to create functional riboswitches. For instance, an artificial translational riboswitch was designed to modulate the internal ribosome entry site (IRES) in wheat germ extract. The choice to modulate the IRES is due to its similarity with prokaryotic RBSs, which are much better understood and thus more programmable. Translation is blocked by dis-

rupting the functional conformation of the IRES when the ligand is unavailable which disables ribosome binding and subsequent translation initiation. When the ligand becomes available, a stable loop is formed forcing the IRES in its functional structure. Various designs using several aptamers were built showing the modularity and versatility of this approach[196]. Another approach rationally fuses naturally occurring *trans*-acting non-coding RNAs (ncRNAs) to aptamers, hereby creating translational and transcriptional riboswitches[197]. Therefore, design principles were described to fuse naturally occurring translation (IS10) and transcriptional (pT181) ncRNAs with aptamers to form a ligand sensing RNA device. To enable ligand dependent riboregulation, structural interactions were designed between the ncRNA motif and the aptamer. These so called pseudoknot interactions are disrupted when the specific ligand binds to the aptamer, hereby altering transcription or translation[197]. Another, more advanced, method uses an *in silico* pipeline to design *de novo* transcriptional riboswitches. Here, a sequence library was constructed comprising an aptamer connected to a randomized spacer and a sequence partially complementary to the aptamer. This library of riboswitch candidates was computationally screened using folding simulations based on several criteria relating to thermodynamics and riboswitch structure, which ultimately resulted in multiple functional gene switches controlling transcription[66]. This last approach marks the first promising computer assisted design of a functional riboswitch, which can be improved by implementing more design parameters derived from characterization studies. A similar approach used RNAiFold, an inverse folding algorithm able to determine RNA sequences satisfying certain structural constraints, to reliably construct functional hammerhead ribozymes[198]. These recent examples of computationally designed riboswitches once again show the versatility of using the RNA aptamer technology to build ligand responsive riboregulators. Figure 2.3 shows the different types of riboswitches and how these RNA devices are developed.

As the technology to build custom riboswitches matures, an ever increasing number of applications emerge. For instance, synthetic riboswitches can be used as high-throughput screening systems, which typically are bottlenecks in both enzyme and metabolic engineering[105,156,199]. Various *ad hoc* engineered riboregulators were used for noninvasive *in vivo* detection of metabolite accumulation[193]. Analogously, Michener & Smolke[56] used a comparable RNA switch to screen a large enzyme library, which resulted in a 33 fold increase in caffeine demethylase activity, hereby showing the potential of ligand dependent RNA switches in facilitating enzyme engineering. To resolve the same bottleneck related to high throughput screening, L-lysine and L-tryptophan riboswitches were
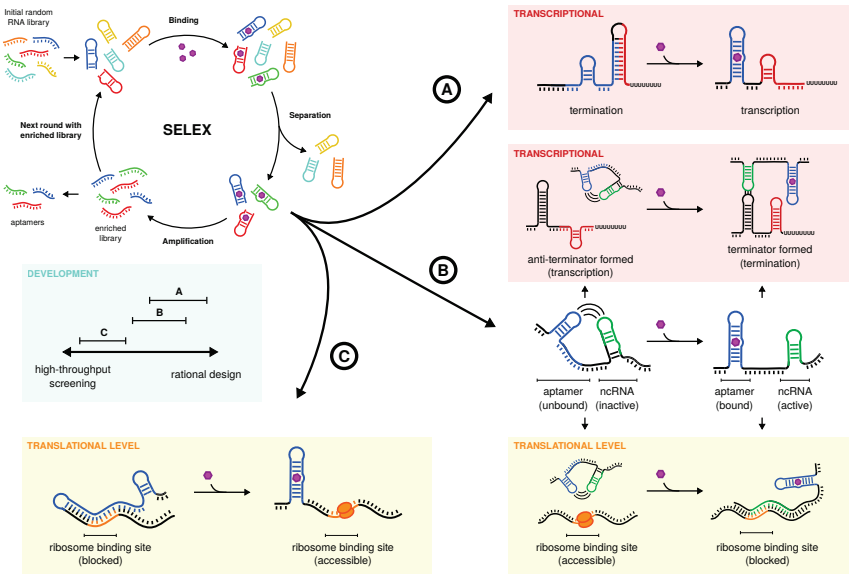
Figure 2.3: Overview of the various approaches to create ligand-dependent gene expression using RNA. Generally, these RNA devices are developed based on *in vitro* selected aptamers. As such, A) transcriptional riboswitches can be computationally designed[66], B) naturally occuring ncRNAs can be engineered to control transcription and translation[197] and C) various translational riboswitches can be selected through high-througput screening efforts[178–180,182,183].

selected from a combinatorial library using the previously mentioned *tetA* dual selection system. Subsequently, the L-lysine riboswitch was used to enrich a population with pathway optimized strains, showing the potential of riboswitches in metabolic engineering[54]. In a similar effort, Eckdahl *et al.*[200] used a theophylline riboswitch to select the best producing strains from a population of designed possible producers. The versatility of using riboswitches as screening devices was shown again by using the natural *glmS* ribozyme in yeast to select *N*-acetylglucosamine producing strains[201].

Another potential application lies in dynamic pathway control in metabolic engineering. Most cell factories are constructed using static control of gene expression, which leads to productivity losses due to its susceptibility to environmental changes. In addition, the biosynthetic pathways of most targets in metabolic engineering contain toxic intermediates. Accumulation of these products imposes stress on the cell, compromising cell growth and pathway productivity[202]. To resolve these issues, tailor made RNA devices can be used to dynamically control biosynthetic pathways. As such, various aptazyme-

regulated expression devices (aREDs) were programmed to quantitatively control gene expression[203]. The potential of these aREDs to control the *p*-aminostyrene pathway was later investigated using simulation analysis, which indicates that this kind of dynamic control can yield over 10 fold improvements over static control[55]. Besides this simulation study, a L-lysine riboswitch was successfully used as dynamic control point, resulting in an up to 63% increased L-lysine production[204].

2.5 Reporting physiology using RNA

The discovery and development of fluorescent proteins as reporter genes revolutionized the field of molecular biology[205]. Numerous fluorescent protein labels are used for molecular imaging of live cells to unravel complex biological processes. However, using protein tags reports gene expression on the protein level, but yields no information on the increasingly important RNA level[206,207]. In addition, using RNA does not require translation, resulting in faster response times, which is advantageous in time-dependent studies (see figure 2.4)[208,209].

To unravel the complex cellular functions of various RNAs, Paige *et al.*[209] developed RNA mimics of green fluorescent protein (GFP). In GFP, a fluorophore, 4-hydroxybenzylidene 2,3-imidazolinone (HBI), is formed autocatalytical from three residues, enabling fluorescence[208,209]. To mimic this fluorescence, various analogs of HBI were prepared, exhibiting no toxicity nor fluorescence upon incubation with cells. By performing SELEX on these HBI derivatives, aptamers were identified that bind and activate fluorescence of these fluorophores. The resulting collection of RNA-fluorophore complexes, termed Spinach, span a large fraction of the visible spectrum, which enables the use of RNA as a fast fluorescent reporter with considerable resistance to photobleaching[209]. However, the first developed generation of Spinach were dim when used to tag RNA in cells due to thermal instability and misfolding. To overcome this limitation, Spinach2 was developed, which was used to unveil the dynamics of toxic CGG repeat-containing RNAs[213]. Subsequently, the spectral properties of both Spinach and Spinach2 were further improved by testing various novel fluorophores, which allowed the characterization of the important structural features. As such, the spectral features of the RNA-fluorophore complex can be tuned at will by using specific plug-and-play fluorophores[214]. The previously mentioned RNA-fluorophore complexes were merely selected on their *in vitro* binding ability, which results in poor functionality in living cells[190]. To select better *in vivo* functioning RNA-fluorophore complexes, Filonov *et al.*[190] developed a procedure
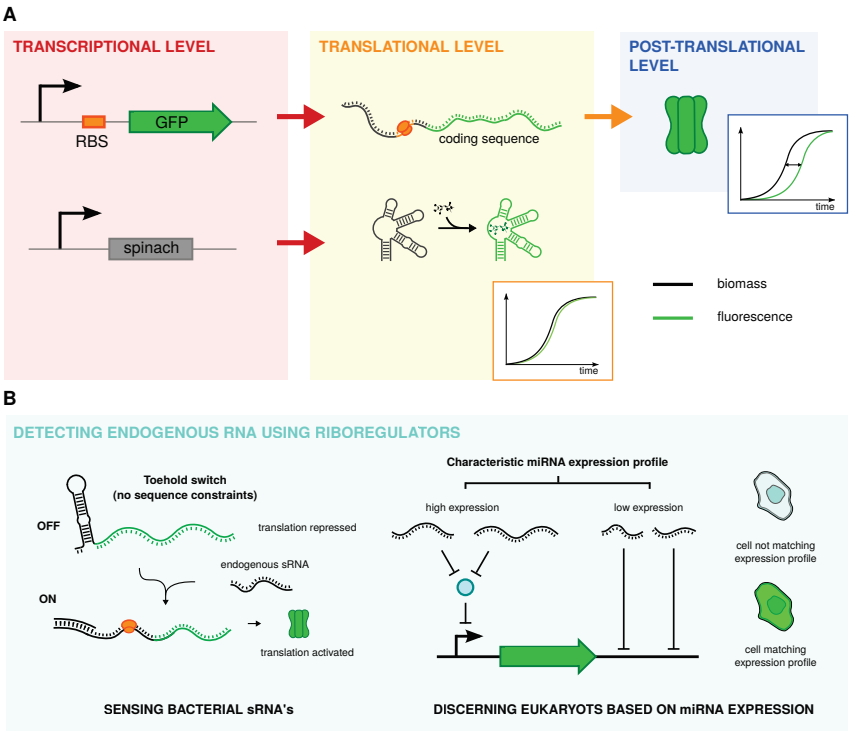
Figure 2.4: Schematic overview of RNA reporters and riboregulators reporting RNA signatures. A) A comparison between protein and RNA-based reporters on the transcriptional, translational and post-translational level. As example, reporter systems using GFP[205] and spinach[209] are shown with hypothetical fluorescence data. B) An overview of riboregulators capable of detecting endogenous RNA. Toehold switches are shown, allowing *in vivo* monitoring of specific bacterial small RNAs (sRNAs) due to their lack of sequence constraints[81]. In addition, a RNA-based circuit is shown capable of detecting a specific miRNA expression profile characteristic for a specific cell type. Moreover, based on the miRNA signature this circuit can discern between eukaryotic cells based on the miRNA expression pattern[210–212].

which involves the regular selection based on thermodynamic binding, followed by an *in vivo* selection process using FACS. This selection approach resulted in several new RNA reporters, termed Broccoli, showing robust folding in the presence of low magnesium concentrations and increased fluorescence in comparison to Spinach2[190]. Elaborating on these RNA reporters, Rogers *et al.*[215] developed a method to monitor RNA assembly and processing by splitting the Spinach aptamer. The resulting two nonfunctional halves can restore fluorescence once assembled, showing the potential for monitoring

assembly of RNA.

Besides serving as pure reporters, RNA-fluorophore complexes can be used to sense small molecules. Paige *et al.* [216] pioneered in this matter by coupling various aptamers to Spinach, resulting in RNA-based sensors, which sense small molecules and subsequently exhibit fluorescence. In this approach, aptamers binding adenosine, adenosine 5'-diphosphate, S-adenosylmethionine, guanine, guanosine 5'-triphosphate were fused to Spinach, using one of the three stem loops of this RNA reporter as an entry point. This stem loop was replaced by various transducers, which link Spinach to the aptamers. Using this method, metabolite sensors were obtained exhibiting up to 32-fold activation in fluorescence upon binding their respective ligand [216]. Similarly, other labs developed Spinach based sensors for cyclic di-GMP and cyclic AMP-GMP using naturally occurring riboswitch domains [217,218]. This latest advance in fluorescent imaging inside the cell shows the designability of pure RNA sensors, which enables measuring metabolites. Conventional metabolite sensing uses protein-based technologies, for instance, FRET, for which there is no direct approach to design devices allowing real time sensing of metabolites at high spatial resolution [166,216]. Overall, using FRET technology for sensing metabolites is not universally applicable as it requires a protein that specifically binds the molecule of interest [219]. Song *et al.* [206] expanded this technology by developing Spinach-based sensors for the detection of intracellular proteins, hereby enabling the development of generalizable RNA-based sensors for the detection of metabolites and proteins [220]. Recently, RNA reporters were combined with riboswitch technology, replacing the expression platform of these gene switches with Spinach, which, upon ligand binding, directly causes fluorescence [221,222]. Aside from the fluorophore-RNA complex technology, several types of RNA switches enable reporting specific physiological states [81,210,211,223]. For instance, the recently developed toehold switches technology lacks sequence constraints, allowing detection of endogenous RNA in bacteria. Monitoring intracellular levels of the RyhB sRNA shows the potential of riboregulators for *in vivo* detection of RNA signatures [81]. Furthermore, years of fundamental research on using RNA to process molecular information [48,223–225] has resulted in several gene circuits sensing endogeneous miRNAs [210]. Xie *et al.* [210] pioneered by constructing a synthetic regulatory circuit which senses RNA expression levels, indicative of HeLa cells, allowing specific identification of cancer cells. If the specific RNA signature matches, a cellular response is triggered, resulting in specific apoptosis of cells, leaving non-cancerous cells unharmed. This pioneering work shows the potential of using riboregulators reporting RNA signatures as intelligent therapeutics. To facilitate the usage of RNA as therapeu-

tic, various enabling technologies were developed to reduce undesired effects associated with traditional gene delivery methods[211,212]. For example, Wroblewska *et al.* [211] used RNA-binding proteins to create a simplified post-transcriptional only version of the circuit constructed by Xie *et al.* [210], avoiding potentially harmful genomic integration. These recent advances show the huge potential of RNA devices as next generation gene therapy by detecting specific RNA signatures and subsequently act upon these signals.

2.6 RNA scaffolds

Molecular crowding in the cytoplasm slows down diffusion processes considerably, hereby limiting translation and subsequently, growth physiology[226]. Hence, in order to optimize the flux through a metabolic pathway co-localization of biochemical reactions can be advantageous[227]. To this end, metabolons have evolved in nature, which allow efficient channeling of pathway intermediates between consecutive enzymatic reactions. Besides these metabolons, cellular localization can also be achieved by eukaryotic organelles and various bacterial protein-based compartments, e.g., bacterial microcompartments (BMCs)[228–230]. These metabolons are temporary complexes of sequential enzymes in a metabolic pathway, which perform multiple conversion steps without complete equilibration with the bulk cellular fluids[231], hereby locally increasing metabolite concentrations. The latter is particularly beneficial when it concerns cytotoxic, unstable intermediates, and metabolites exerting feedback regulation. To mimic these metabolons, protein fusions were initially developed. For example, Zhai *et al.* [234] developed a *nahK*-tr*glmU* fusion protein for the efficient production of UDP-*N*-acetylglucosamine *in vitro* with a yield of 88% from *N*-acetylglucosamine, ATP and UTP. However, despite some successful examples these protein fusions are utterly inadequate in terms of flexibility, as degrees of freedom like stoichiometry and distance and orientation between proteins catalyzing consecutive reactions and the potential of larger multidimensional scaffolds cannot fully be exploited. In this respect, RNA and to a lesser extent DNA scaffolds offer clear advantages over protein scaffolds. Not only, since over time a vast library of aptamers has been created[235], but more importantly since RNA scaffolds exhibit the required structural flexibility and can be (more) rationally designed[236]. These features are considered essential to further advance these emerging applications. Additionally, in eukaryotes, DNA scaffolds can only be used in the nucleus, a limitation that RNA scaffolds do no have[232]. Moreover, figure 2.5 compares the various types of techniques to create metabolons. These RNA scaffolds can be either
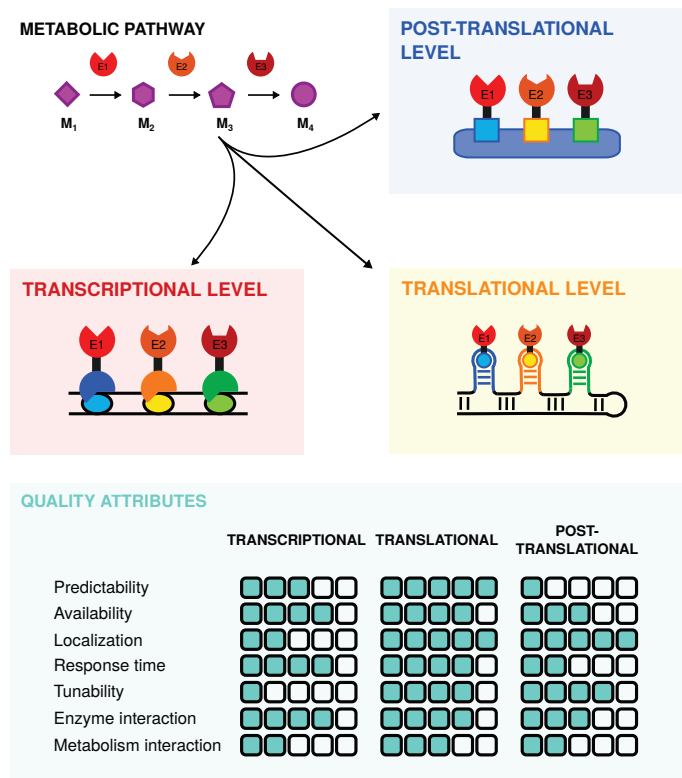
Figure 2.5: Overview of the available techniques to create synthetic metabolons, hereby facilitating biochemical conversion. As such, various scaffold techniques, on the DNA[232], RNA[233] and protein[228] level, are compared. The quality attributes evaluated are predictability (degree to which the part functionality can be estimated), availability (estimation of the technology available to build a metabolon), localization (degree to which the location of the metabolon is adjustable), response time (maturation time needed to obtain functional metabolon), tunability (degree to which the function can be controllably adjusted), enzyme interaction (degree to which the enzyme interact with other parts of the metabolon system), metabolism interaction (degree to which the metabolon interacts with the metabolism of the host organism).

discrete or more complex[236]. The former consist of a non-coding RNA which contains multiple copies of the chosen aptamers. The number of, the distance between, and orientation of the folded aptamers are tunable. While, the latter are composed of typically short (cross)-polymerizable RNA molecules. Extended multidimensional structures can be formed merely by using a limited set of different palindromic sequences[236].

Using these RNA scaffolds considerable production increases have yet been obtained, e.g., Sachdeva *et al.* [237] obtained a 80% production increase of a two-enzyme pentadecane pathway by reorienting the acyl-ACP reductase fusion, clearly indicating that success strongly depends on the scaffold's geometry as certain aptamer conformations rendered the transfer of certain metabolic intermediates less favourable. Hence, it should be clear that RNA and protein modeling software is indispensable to more rationally and successfully design these scaffolds [236] and extensive characterization to evaluate the scaffold's outcome is mandatory [98]. To this end, techniques like fluorescence complementation, which allows to visualize interactions among multiple proteins *in vivo*, and atomic force microscopy and transmission electron microscopy [98]. Other applications include, among others, enhanced solubilization of heterologously hard to express proteins, by fusing chaperonne proteins, such as DnaJ, to a RNA binding domain that specifically targets a RNA sequence in the 3' UTR region of the mRNA coding for the heterologous protein [238] and CRISPR-associated RNA scaffolds [53], as was previously discussed.

2.7 Conclusion

Empowered by the recent advances in synthetic biology, systems biology and protein engineering, metabolic engineering is tackling ever-increasing complex optimization problems, often considered not resolvable hitherto. To properly address these multivariate and non-linear optimization questions second generation RNA-based parts and tools are being designed to create tailor made cells applicable in biotechnology and biomedicine. More specifically, these RNA devices are used to optimally reroute metabolism by modulating gene expression, building genetic circuits, building scaffolds and to detect molecules and report cellular processes, which outperform contemporary alternatives. The translation of the fundamental knowledge on RNA, e.g. folding hybridization results in the creation of a set of second generation RNA tools with improved reliability, programmability, *etc.* outperforming the pioneering work on riboregulators.

These second generation tools are considered essential to deal with the enormous complexity encountered, as they typically contribute to diminishing the vast search space or speeding-up its exploration. For example, increasingly complex self-regulating cybernetic components, which control metabolism in response to various cellular signals, are applied for the creation of so called smart (combinatorial) systems and libraries, as flux distributions leading to a priori non-optimally performing phenotypes are ruled out.

However, with this increase in complexity so harshen the requirements for these novel parts and tools in terms of of speed, scalability, flexibility, designability, cost-efficiency, orthogonality, portability, *etc.* In view of the almost perfect match with the features of RNA, its engineering is developing from an academic fiddling into an essential metabolic engineering tool. At an amazing pace, the fundamental knowledge on RNA folding, stability, *etc.* and its pivotal role in controlling cellular processes is being translated into real-life applications. Where it took decades to establish and optimize gene modulation tools based on DNA devices such as promoter libraries and subsequently transfer them to other organisms, its remarkable how fast these RNA-based tools are being developed and transferred form one host to another, e.g. CRISPR based tools are yet available for all domains of life (archaea, bacteria, plant, human cells).

However, despite several recent success stories the potential of RNA use is not yet fully being exploited as several issues of experimental and theoretical nature still persist. In this respect, RNA engineering is expected to profit greatly from the dedicated collection of data. Compared to the other engineering devices based on DNA or protein, relatively few examples exist of detailed characterization studies. Such mustered data will

contribute to a further standardization of RNA components enabling more predictable engineering. Merely one example, the impact of translation and codon usage on RNA stability should be mapped in detail.

These efforts will also result in a better understanding of the predominant processes affecting RNA device performance and consequently to more reliable software tools. Though helpful *in silico* tools have already been developed, e.g., RBS calculator[25], most of them still suffer from imperfect knowledge on key processes affecting RNA device performance. In this respect evolutions in the field of computer sciences, e.g., big data analysis and parallel computing and in the field of RNA quantification, e.g., RNA-seq, will help to elucidate the functioning of RNA.

Acknowledgements

Chapter 3

Exploration of the feature space of *de novo* developed post-transcriptional riboregulators

Contents

Abstract

Metabolic engineering increasingly depends upon RNA technology to customely
rewire the metabolism to maximize production. To this end, pure riboregulators allow
dynamic gene repression without the need of a potentially burdensome coexpressed
protein like typical Hfq binding small RNAs and clustered regularly interspaced short
palindromic repeats technology. Despite this clear advantage, no clear general de-
sign principles are available to *de novo* develop repressing riboregulators, limiting
the availability and the reliable development of these type of riboregulators. Here,
to overcome this lack of knowledge on the functionality of repressing riboregula-
tors, translation inhibiting RNAs are developed from scratch. These *de novo* de-
veloped riboregulators explore features related to thermodynamical and structural
factors previously attributed to translation initiation modulation. In total, 12 struc-
tural and thermodynamic features were defined of which six features were retained
after removing correlations from an *in silico* generated riboregulator library. From
this translation inhibiting RNA library, 18 riboregulators were selected using a ex-
perimental design and subsequently constructed and co-expressed with two target
untranslated regions to link the translation inhibiting RNA features to functionality.
The pure riboregulators in the design of experiments showed repression up to 6 %
of the original protein expression levels, which could only be partially explained by
a ordinary least squares regression model. To allow reliable forward engineering,
a partial least squares regression model was constructed and validated to link the
properties of translation inhibiting RNA riboregulators to gene repression. In this
model both structural and thermodynamic features were important for efficient gene
repression by pure riboregulators. This approach enables a more reliable *de novo*
forward engineering of effective pure riboregulators, which further expands the RNA
toolbox for gene expression modulation.

3.1 Introduction

Over the last decade, synthetic biology and systems biology spurred major advances
in metabolic engineering, resulting in several economically competitive production pro-
cesses for both bulk and fine chemicals from renewable resources, revolutionizing indus-
trial biotechnology[1–5]. In this context, interfering with the native metabolism of the pro-
duction host is a necessity to redirect the metabolic flux towards the product of interest
with a view to maximizing productivity[28,29]. Traditionally, tuning the cellular metabol-
ism has been done through gene deletions, which is impossible for numerous essential
genes often related to various biosynthetic pathways[70,239]. As such, maximizing various
production pathways requires tools able to specifically reduce gene expression. To this
end, zinc fingers and transcription activator-like effectors were engineered to dynami-
cally control transcription of a specific gene through DNA-binding proteins[72,74]. These
custom gene expression regulators are outperformed by recently emerged clustered reg-
ularly interspaced short palindromic repeats interference (CRISPRi), an adaptation of
the type II clustered regularly interspaced short palindromic repeats (CRISPR) system
controlling transcription through reversible binding of a RNA-guided deactivated Cas9
nuclease to DNA[65]. Various metabolic engineering efforts in multiple organisms used
this CRISPRi technology to successfully repress a series of specific genes in a dynamic
way, hereby ameliorating the desired product formation[99,100,240].

Alternative approaches to control gene expression on the translational level employ
small RNAs (sRNAs) to repress protein production by blocking translational initiation,
enabling metabolic flux redirection at will[70,88,241,242]. Similar to the CRISPRi technol-
ogy, which requires a small guide RNA (sgRNA) able to bind to the dCas9 protein,
these types of sRNAs also require a protein binding RNA motif as they rely on the sta-
bilizing Hfq protein. This dependence on coexpressed proteins might cause increased
metabolic burden, which can lead to long term genetic instability and unexpected be-
haviour[100,243,244]. To reduce these undesired effects, gene expression modulation sys-
tems are preferred that solely rely on RNA[245]. These pure riboregulators require less
cellular resources by avoiding the extra translation step, hereby lowering metabolic bur-
den[245]. For example, riboregulator technology is successfully used to precisely down-
regulate specific genes, hereby redirecting metabolite fluxes towards the phenotype of
interest[78,85,86]. Also, pure riboregulators, which do not require proteins, harness large
potential for the construction of fast responding RNA circuitery[49,128,241,246].

Early pure riboregulators were designed to hybridize the translation initiation region.

The RNA architecture in this region plays a pivotal role in the translation initiation process, enabling gene expression through RNA-RNA interactions[58]. This apparent link between RNA structure and biological function, in combination with the ease and reliability of RNA secondary structure prediction, drove several gene silencing efforts solely using RNA. However, successful attempts to modulate gene expression using solely *trans* expressed RNA employed a variety of features[82–84]. As such, interfering with translation initiation using solely RNA-RNA interactions has been attributed to various features of the *trans* expressed RNA molecule[49,82,83,127,132,247,248].

These features are classified as either structural or thermodynamic features. Several structural features of riboregulators modulating translation initiation through RNA-RNA interactions include post-transcriptional ribosome binding site (RBS) occlusion[127,132], formation of paired termini structures[247], and manipulation of the structural accessibility of the target site[82,248]. Besides structural constraints, various thermodynamic features were previously used to design and optimize translation interfering riboregulators, mainly comprising formation and activation energies[49,132]. Formation energies are typically obtained by estimating the minimum free energy (MFE)[132]. Despite the importance of activation energy, various estimation methods for the activation energy were previously used to create functional riboregulators[49,132]. These methods rely on the initial monomeric structures and are based on the assumption that the unbound nucleotides in this state initiates the RNA-RNA complex formation[49,132].

This broad range of employed features indicates the lack of consensus in literature, which limits the general applicability of the current design rules for pure riboregulators (without using coexpressed proteins). For instance, simply expressing the antisense strand does not fully repress gene expression on the post-transcriptional level[78,85]. As such, various types of riboregulators suitable for metabolic engineering purposes were created using a number of different riboregulator design features, once again indicating the lack of consensus in literature on the development of riboregulators[49,81,132]. Overall, these riboregulators are either developed from a natural existing RNA regulator chassis or created *de novo*, the latter being the most interesting as this enables forward engineering in a broader context[81,132]. Moreover, only activating riboregulators were created *de novo*, which limits the construction of genetic circuitry using solely RNA.

Here, we propose a framework for the *de novo* design of pure riboregulators, referred to as translation inhibiting RNAs (tiRNAs), which repress gene expression by blocking translation initiation. To develop this predictive framework, the influence of all features previously attributed to post-transcriptional gene modulation were analyzed in a de-

sign of experiments (DOE). This experimental design allows exploration of the feature landscape and evaluation of their influence on gene repression. First, using a library of tiRNA, all features were analyzed *in silico* to create a collection of features with maximal information content. Next, the performance of *de novo* designed tiRNAs was evaluated *in vivo*, and used to construct an ordinary least squares (OLS) and a partial least squares (PLS) model which links riboregulator features to gene repression.

3.2 Methods

3.2.1 Strains and growth conditions

E. coli strain DH10B (Invitrogen) was used for both plasmid construction and fluorescence measurement purposes. Unless otherwise stated, all products were purchased from Sigma-Aldrich (Diegem, Belgium). For plasmid construction and fluorescence measurements strains were grown in lysogeny broth (LB) and MOPS EZ rich medium (Teknova, Bioquote, York, United Kingdom) at pH 7.4, respectively at 37°C with shaking. LB was composed of 1 % tryptone-peptone (Difco, Erembodegem, Belgium), 0.5 % yeast extract (Difco) and 1 % sodium chloride (VWR, Leuven, Belgium). LB agar (LBA) plates contain the same components as LB with the addition of 1 % agar. If required, medium was supplemented with 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} kanamycin.

3.2.2 Plasmids

pTarget plasmids were medium-copy vectors (pBR322 origin of replication (ori) and ampicillin resistance marker, originating from pSB6A1[249]) with proD[250] as promoter and BBa_B1006[24] as terminator for tiRNA expression (see Supplementary Figure A.1 for more details), and pSilence plasmids were low-copy vectors (pSC101 ori and kanamycin resistance marker, originating from pCL1920[251]) with proB[250] as promoter, *mKate2*[252] as reporter gene, rnpB T1[24] as terminator, and the target 5' untranslated region (UTR) (see Supplementary Figure A.2 for more details). The reporter mKate2 was used due to its low background and good fluorescent protein properties (brightness and maturation time)[252]. A schematic overview of the two plasmid types used in this study (pSilence and pTarget) is shown in Supplementary Figure A.3.

The control plasmids used in this study were pBlank$_1$ and pBlank$_2$, which are the same vectors as the pSilence and pTarget plasmids, respectively. The pBlank$_1$ plasmid comprises only the vector backbone and pBlank$_2$ contains the *mKate2* open reading frame (ORF) and rnpB T1[24] as terminator, thus without promoter and UTRs. All plasmids used in this study were constructed using Golden Gate[253] and CPEC[254] assembly. DNA oligonucleotides were commercially ordered from IDT (Leuven, Belgium) and DNA sequences of every constructed plasmid were verified using sequencing services (Macrogen Inc., Amsterdam, The Netherlands). All tiRNA sequences used in this study are listed in Supplementary Table A.1. Details of the plasmids and DNA sequences used in this study are listed in Supplementary Table A.2 and A.3, respectively.

3.2.3 *In vivo* fluorescence and optical density (OD) measurements

For *in vivo* assessment of translational inhibition, strains were plated on LBA plates containing $100\,\mu g$ ml^{-1} ampicillin and $50\,\mu g$ ml^{-1} kanamycin. After overnight incubation, three colonies were inoculated in $150\,\mu l$ MOPS EZ rich medium, covered by a Breathe-Easy sealing membrane (Sigma-Aldrich), and grown overnight on a Compact Digital Microplate Shaker (Thermo Scientific) at 800 rpm and 37°C. Subsequently, these cultures were 1:100 diluted in $150\,\mu l$ of fresh MOPS EZ rich medium and grown on a Compact Digital Microplate Shaker until late log phase (6 h) at 800 rpm and 37°C. Subsequently, fluorescence and OD were measured using a Tecan M200 pro microplate reader. Precultures were grown in Greiner bio-one (Vilvoorde, Belgium) polystyrene F-bottom 96 well plates. Fluorescence and OD measurements were performed after growth in Greiner bio-one (Vilvoorde, Belgium) black μclear 96 well plates. For measuring mKate2 expression an excitation wavelength and an emission wavelength of 588 nm and 633 nm were used, respectively. OD was measured at a wavelength of 700 nm to reduce bias in estimates of cell abundance[255].

3.2.4 Fluorescence data analysis

For fluorescence measurements, two types of controls were used on every 96-well microtiter plate, i.e., a MOPS EZ rich medium blank and *E. coli* DH10B cells without fluorescent protein expression (contains pBlank$_1$ and pBlank$_2$ plasmids). The medium blank was used to correct the background OD (OD$_{bg}$) of the medium. The fluorescence of the strain without fluorescent protein expression (FP$_{bg}$) was used to correct for the background fluorescence of *E. coli*. For all strains fluorescence per OD was calculated as follows:

$$\left(\frac{\text{FP}}{\text{OD}}\right)_{\text{corrected}} = \frac{\text{FP} - \text{FP}_{bg}}{\text{OD} - \text{OD}_{bg}} \tag{3.1}$$

The relative protein expression was defined as follows:

$$\text{Relative protein expression (\%)} = \frac{\left(\frac{\text{FP}}{\text{OD}}\right)_{\text{corrected}} \text{ with riboregulator}}{\left(\frac{\text{FP}}{\text{OD}}\right)_{\text{corrected}} \text{ without riboregulator}} \times 100 \tag{3.2}$$

3.2.5 Feature quantification using RNA bioinformatics

For each tiRNA candidate 12 features were calculated (see Table 3.1 for detailed definitions and Supplementary Figure A.4), which were determined based on various previous riboregulator efforts described in literature[49,82,127,132,132,248]. The tiRNA features

are classified in two main groups: thermodynamic properties and structural constraints. All intra- and intermolecular interactions between RNA molecules were predicted using RNAfold[256] and RNAcofold[257], respectively. Both RNA secondary structure prediction algorithms were available through the Vienna RNA package[258] and were used with only the options –noLP -d2 and an accuracy of 10^{-100}, all other settings are set to the default setting. Suboptimal structures of RNA molecules are drawn with probabilities equal to their Boltzmann weights using RNAsubopt[259]. The intermolecular binding between the unbound part of the tiRNA and the UTR is estimated by the RNAup algorithm[260]. All calculations were done using Python scripting on an Intel Xeon E5-2670 (2.60GHz) Linux (Debian) server. Details on the quantification of thermodynamic and structural properties of tiRNA molecules are available in Supplementary Methods A.1.1.

3.2.6 Statistical calculations and experimental design

All statistical calculations and analyses were performed in R. Unless otherwise stated, error bars represent the standard deviation (n=3). All coefficient of determinations (R^2s) were calculated using the hydroGOF package in R.

Experimental design

The 2^{6-2} fractional factorial design, which comprises solely UTR_1, was generated using the R package FrF2[261]. In the DOE, the -1 and 1 state of the factors were defined as the 0.1 and 0.9 p-quantiles of the original feature distribution, respectively. The center points are designed to be [0,0,0,0,0,0], where 0 represents the average of the absolute values of the tiRNA features in the initial library of 1,500,000 possible tiRNA candidates. For each feature in the experimental design all data points (X_i) were centered and scaled based on the 0.1 and 0.9 p-quantiles ($q_{X,0.1}$ and $q_{X,0.9}$) of the original distribution of feature X (Equation 3.3).

$$\tilde{X}_i = \frac{X_i - (q_{X,0.9} + q_{X,0.1})/2}{(q_{X,0.9} - q_{X,0.1})/2}$$

(3.3)

The centered features $\tilde{X}$ were only used in the analysis of the DOE. All data points of the 2^{6-2} fractional factorial design are shown in Supplementary Table A.4. The features FAB and EIS were multiplied by -1 to obtain positive regression coefficients as these were expected to be negatively correlated with tiRNA performance.

Table 3.1: Detailed definitions of all features (free energy of the tiRNA monomer (EA), free energy of the tiRNA-tiRNA dimer (EAA), free energy of the tiRNA-UTR dimer (EAB), formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), total seed energy (ETS), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 5 (RBS5), RBS coverage of length 11 (RBS11), paired termini (PT), and the length of the translation inhibiting RNA (tiRNA) (L)) used in the initial *in silico* screening of the tiRNA library for repression of a target untranslated region (UTR). This UTR contains a ribosome binding site (RBS) and controls the coding DNA sequence (CDS) of the reporter protein. All binding probabilities of monomers and dimers are respectively derived from base pairing probability matrices estimated by RNAfold[256] and RNAcofold[257].

Name	Definition of the tiRNA feature
EA	ΔG_{tiRNA}; free energy of the tiRNA monomer, calculated using RNAfold[256].
EAA	$\Delta G_{\text{tiRNA-tiRNA}}$; free energy of the tiRNA-tiRNA dimer, calculated using RNAcofold[257].
EAB	$\Delta G_{\text{tiRNA-UTR}}$; free energy of the tiRNA-UTR dimer, calculated using RNAcofold[257].
FAA	Formation energy of the tiRNA-tiRNA dimer; $\Delta G_{\text{tiRNA-tiRNA}}$ - 2 ΔG_{tiRNA}.
FAB	Formation energy of the tiRNA-UTR dimer; $\Delta G_{\text{tiRNA-UTR}}$ - ΔG_{tiRNA} - ΔG_{UTR}. With ΔG_{UTR} defined as free energy of the UTR (including first 50 nucleotides of the CDS RNA), calculated using RNAfold[256].
ETS	Average minimal total energy (gains from intermolecular binding and needs to open the binding site) for the binding of unbound parts of the tiRNA monomer to the target UTR, calculated using RNAup[260] for 100 suboptimal structures randomly drawn from Boltzmann ensemble[259].
EIS	Average minimal energy gained from intermolecular binding of unbound parts of the tiRNA monomer to the target UTR, calculated using RNAup[260] for 100 suboptimal structures randomly drawn from Boltzmann ensemble[259]
PAU	Weighted average of the relative number of unbound nucleotides in the UTR monomer with the relative number of nucleotides bound by the tiRNA molecule in the tiRNA-UTR dimer complex as weights.
RBS5	The RBS coverage (relative number of bound nucleotides) in the region $C_{\text{RBS}}-2$ to $C_{\text{RBS}}+2$, where C_{RBS} is defined as the weighted average of the nucleotides in the UTR bound by the 16S rRNA.
RBS11	The RBS coverage (relative number of bound nucleotides) in the region $C_{\text{RBS}}-5$ to $C_{\text{RBS}}+5$, where C_{RBS} is defined as the weighted average of the nucleotides in the UTR bound by the 16S rRNA.
PT	Average number of bound nucleotides between the first and the second half of the a tiRNA sequence calculated for 100 suboptimal structures randomly drawn from Boltzmann ensemble[259].
L	The length of the tiRNA (nt)

Regression models

Ordinary least squares (OLS) regression   The OLS regression was done in R. The
OLS regression model was calibrated using the absolute (unprocessed) values of FAB for
all data points, including data from target UTR_2. Eq. 3.4 depicts the linear relationship
obtained from the OLS regression, where Y_j is the relative protein expression when
tiRNA j is present, $X_{j,1}$ is feature FAB of tiRNA j, β_0 and β_1 are regression coefficients
and ϵ_j is an error term.

$$\log_{10}(Y_j) = \beta_0 + \beta_1 X_{j,1} + \epsilon_j \tag{3.4}$$

Partial least squares (PLS) regression   The PLS regression was done in R with the
package pls[262]. The PLS model was validated by splitting the data set from UTR_1 and
UTR_2 in a training set and validation set (5:1 ratio). Subsequently, the training set
was used to create the model by leave-one-out cross validation where predictors were
scaled prior to regression (by dividing each variable the sample standard deviation). In
PLS regression the matrix of predictors **X** is decomposed into orthogonal score matrix
T (projection of **X**) and loadings matrix **P**, circumventing potential collinearities in the
data set:

$$\mathbf{X} = \mathbf{T}\,\mathbf{P} \tag{3.5}$$

Subsequently, **Y** is regressed on the scores **T** (and not **X**). The specific PLS algorithm
used is kernel PLS, which was described by Dayal *et al.* [263].

3.3   Results and Discussion

The *trans* expressed tiRNAs are *de novo* designed to inhibit translation initiation of a
gene of interest, the rate-limiting step in translation[264], as depicted in Figure 3.1A.
Contrary to previous efforts to construct repressing riboregulators, these RNA devices
are constructed from scratch without a functional chassis, which is often based on a
natural occurring RNA regulation device[47,49,79]. To enable reliable forward engineering
of tiRNAs, a workflow to improve the *de novo* development of repressing riboregulators
through DOE guided exploration of the sequence space was developed and optimized
(see Figure 3.1B-D). First, possibly important features for translational inhibiting ri-
boregulators are derived from literature. Secondly, the number of features are reduced
by removing correlations. Subsequently, this reduced set of tiRNA properties is used in
an experiment designed to unravel design principles to build effective tiRNA molecules.
In the DOE, tiRNAs are constructed that explore the feature space in an intelligent way.
Ultimately, thoroughly analyzing the performance of the constructed tiRNAs with vary-
ing features can improve the knowledge on the structure-function relationship, which
correlates to better predictability of *de novo* created riboregulators[49,81,132].

3.3.1   Identification of determinative features of repressing riboregu-
lators

In total, 12 potentially determinative features of efficient tiRNA were identified and
derived from literature (see Figure A.4 and Table 3.1 for more details). These 12 features
represent all design principles previously used in riboregulator construction.   Five out
of the 12 indentified features are based on structural properties. Namely, two features
are defined to quantify RBS coverage, i.e. RBS5 and RBS11, which is the average base
pairing probability in the region of the RBS with length 5 and 11, respectively. The third
feature quantifies the amount of paired termini (PT) and the availability of the UTR is
evaluated by the PAU property. The last structural feature is determined by the length of
the tiRNA (L). The remaining seven defined features are based on properties relating to
thermodynamics. The energy required for the formation of the tiRNA-tiRNA dimer and
the tiRNA-UTR dimer are defined as FAA and FAB, respectively. These formation energies
are calculated based on the estimated Gibbs free energy of the final dimer and both
initial monomer states, which are described by the EA, the EAA, and the EAB features.
In addition, two features describe the activation energy: the intermolecular binding seed
energy (EIS) and the total seed energy (ETS). Noteworthy, most previous approaches
to specify design rules for riboregulators only take the MFE structures into account,
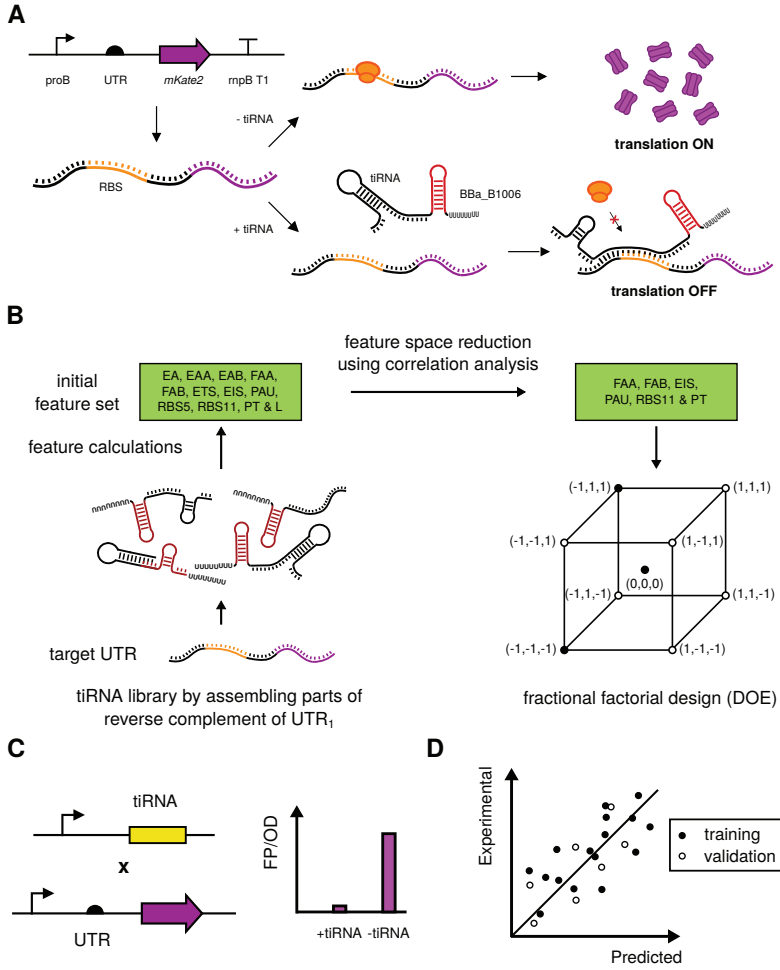
Figure 3.1: A) Schematic overview of the translation inhibiting RNA (tiRNA) working mechanism B) Workflow for the *in silico* selection of the tiRNAs comprising the design of experiments (DOE) to unravel design principles. The defined tiRNA features (free energy of the tiRNA monomer (EA), free energy of the tiRNA-tiRNA dimer (EAA), free energy of the tiRNA-UTR dimer (EAB), formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), total seed energy (ETS), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 5 (RBS5), RBS coverage of length 11 (RBS11), paired termini (PT), and the tiRNA length (L)) are calculated for a tiRNA library created based on a specific target 5' untranslated region (UTR). C) *In vivo* assessment of the tiRNAs in the designed experiment D) Linking features to tiRNA performance through modeling.

simplifying the Boltzmann ensemble of RNA secondary structures and the corresponding
complex dynamic energy landscape of regulatory RNAs[265]. The workflow followed here
to improve the *de novo* development of repressing riboregulator through DOE guided
exploration of the sequence space is depicted in Figure 3.1B. Here, simplifications were
minimized by taking the Boltzmann ensemble into account as much as possible.

3.3.2   Feature space reduction using correlation analysis

A library of 1,500,000 unique possible tiRNAs with length 20, 30 or 40 nucleotides
(nt) was created *in silico* based on UTR$_1$ (see Supplementary Table A.5). To this end,
sequences were generated by successively combining different randomly chosen parts
(length >= 2 nt) of the reverse complement of UTR$_1$, as effective riboregulators typically
contain parts of the reverse complement of the target UTR.

The amount of correlations between the various features was reduced by analyzing ex-
isting correlations between all features, and subsequently removing correlations above
a set threshold of 0.75. This was done by calculating the Pearson correlation coeffi-
cients (PCCs) (see Figure 3.2). The correlations between FAA, EA, and EAA, between
FAB and EAB, and between EIS and ETS are caused by one or more features being used
in the calculation of another feature. Also, RBS5 and RBS11 are correlated, which can
be explained by the fact that the RBS region covered by RBS5 is also covered by RBS11.
Finally, the length of tiRNA (L) is correlated with EAB as the stability of the tiRNA-UTR
complex increases (lower Gibbs free energy) with the tiRNA length. The feature space
was reduced, while minimizing information loss, by removing correlations between var-
ious features. To this end, one feature of a set of correlated features was selected ($|$PCC$|$
> 0.75). The reduced set of the tiRNA features X_i with limited correlations, i.e. FAA,
FAB, EIS, PAU, RBS11, and PT, was used in a DOE to unravel the features with the most
influence on the repression efficiency of these pure riboregulators (see Figure 3.1B).
More specifically, a fractional factorial 2-level design was used with a resolution of IV
(2^{6-2} design), comprising two center points and 16 factorial points. After rescaling all
features (see Section 3.2.6 and Supplementary Table A.4 for details), the 18 best suiting
data points were selected from the library of 1,500,000 tiRNA candidates. The density of
all tiRNA features of the complete constructed library with the 0.1 and 0.9 p-quantiles is
depicted in Supplementary Figure A.5. Because the features are calculated based on the
sequence of a generated tiRNA candidate, the factors cannot be set to a specific value.
Instead, suitable sequences were selected from the tiRNA candidate library based on the
residual sum of squares (RSS) between the real data point of the experimental design
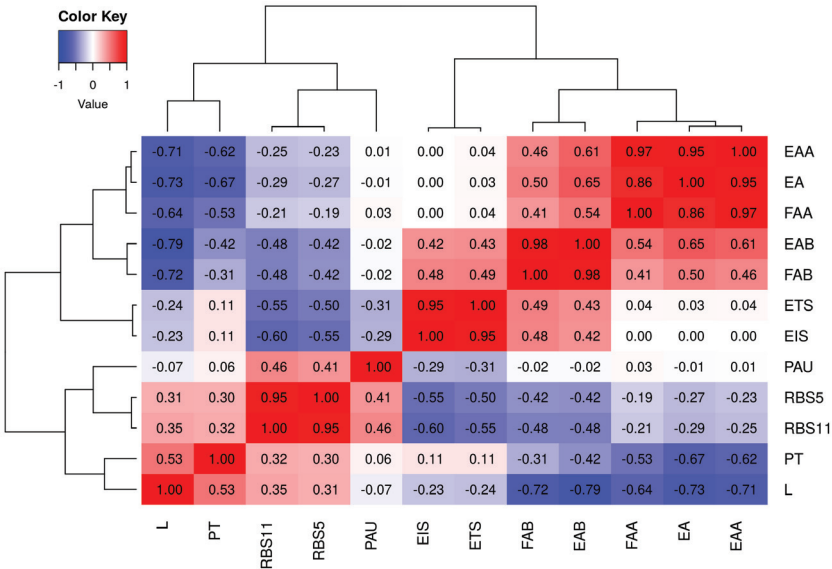
Figure 3.2: Heatmap of the Pearson correlation coefficients (PCCs) between all features of the translation inhibiting RNA (tiRNA) library, with L = length, EA = free energy of the tiRNA monomer, EAA = free energy of the tiRNA-tiRNA dimer, EAB = free energy of the tiRNA-UTR dimer, FAA = formation energy of the tiRNA-tiRNA dimer, FAB = formation energy of the tiRNA-UTR dimer, ETS = total seed energy, EIS = intermolecular binding seed energy, PAU = probability availability of UTR, RBS5 = RBS coverage of length 5, RBS11 = RBS coverage of length 11, and PT = paired termini (see Table 3.1 and Supplementary Figure A.4).

and the actual features a this specific candidate (overall, average RSS is 0.59). The selected tiRNA sequences (one feature was selected from features with a PCC| above 0.75) with their corresponding theoretical data point are depicted in Supplementary Table A.1 and Figure 3.3A, respectively.

3.3.3   *In vivo* analysis of tiRNA performance

Subsequently, the performance of these 18 tiRNAs in the DOE was evaluated *in vivo* as depicted in Figure 3.3A. In this experimental setup, the tiRNA molecules are expressed from pSilence plasmids carrying the pBR322 ori, which have an approximately fourfold higher copy number compared to the pSC101 ori of the pTarget plasmids utilized for UTR expression[249], and are under the control of the proD promoter, which showed 8.4 fold higher transcriptional activity compared to the proB promoter used for UTR

expression[250]. The overall higher relative tiRNA expression (compared to its target
UTR) was chosen based on the fact that *trans* acting sRNAs typically require relatively
higher expression of the sRNA compared to its target, in both natural and synthetic sRNA
regulation systems[266,267].

To enlarge the data set, the pSilence plasmids were co-transformed with the pTarget
plasmid containing UTR_1 or UTR_2 (a truncated version of UTR_1, see Supplementary
Table A.5), respectively, evaluating the repression efficiency of the tiRNAs in the DOE.
Compared to UTR_1, UTR_2 results in 3.3 times less production of fluorescent protein in ab-
sence of any riboregulator (see Supplementary Figure A.6) although the thermodynamic
stability of UTR_1 is much higher than UTR_2 (-27.6 and -17.3 kcal/mol, respectively),
which is in contrast to previous studies inversely relating translation to the Gibbs free
energy of the UTR[25,268]. Moreover, the UTR_2 forms much less base pairs in the region
around the Shine-Dalgarno (SD) sequence (see Supplementary Figure A.6), making the
RBS possibly more accessible.

The outcome of the designed experiment is depicted in Figure 3.3C. The activity of the
de novo designed riboregulators shows that almost all tiRNAs were active. Specifically,
eight of the 18 tiRNAs targetting UTR_1 inhibit the translation initiation of UTR_1 with
more than 75 %. The most repressing tiRNAs reduce protein expression of UTR_1 to
about 6 % of the original expression level. The highest dynamic range of translation
repression among all data points is 16, which is higher than previously described re-
pressing riboregulators[49]. Moreover, these tiRNA are created *de novo*, without using a
naturally occurring functional chassis.

Overall, the repression levels on UTR_2 are comparable to those of UTR_1, indicating that
the truncated part distal to the RBS BBa_B0032 is less important for riboregulator activ-
ity. There is a clear difference in repression efficiency between $tiRNA_1$ ([-1,-1,-1,-1,-1,-
1]) and $tiRNA_{16}$ ([1,1,1,1,1,1]), showing the importance of at least one of the selected
features for translational inhibition. Another interesting fact is the high repression rate
of $tiRNA_{17}$ and $tiRNA_{18}$, which are the center points of the DOE. The good performance
of these center points indicate that choosing less extreme values can also result in effec-
tive translational repression.

A

B

	FAB	FAA	EIS	PAU	RBS11	PT
1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	1
4	1	1	-1	-1	-1	-1
5	-1	-1	1	-1	1	-1
6	1	-1	1	-1	-1	1
7	-1	1	1	-1	-1	1
8	1	1	1	-1	1	-1
9	-1	-1	-1	1	-1	1
10	1	-1	-1	1	1	-1
11	-1	1	-1	1	1	-1
12	1	1	-1	1	-1	1
13	-1	-1	1	1	1	1
14	1	-1	1	1	-1	-1
15	-1	1	1	1	-1	-1
16	1	1	1	1	1	1
17	0	0	0	0	0	0
18	0	0	0	0	0	0

C
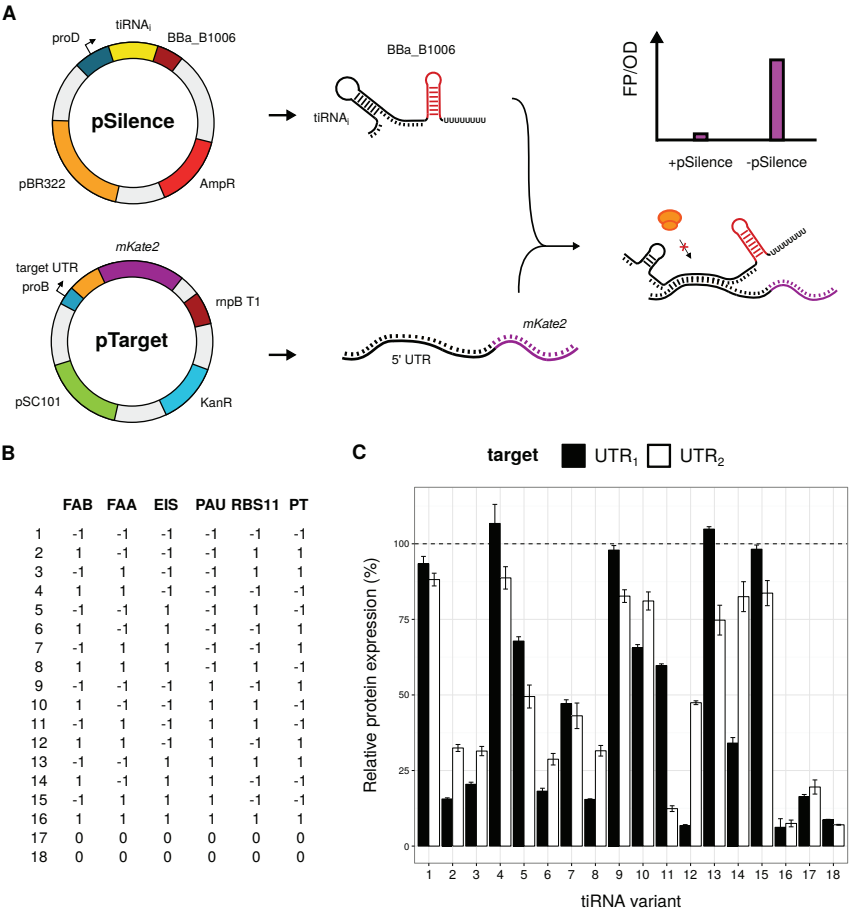
Figure 3.3: The results from the designed experiment to unravel the principles for translation inhibiting RNA (tiRNA) design. (A) The practical execution of the design of experiments (DOE). All tiRNAs, representing a data point in the DOE, are coexpressed with the target untranslated region (UTR) and the riboregulator efficiency is determined. (B) All tiRNAs with corresponding feature data point in the DOE, where the features are calculated using UTR_1. (C) The tiRNA performance, expressed in relative protein expression, where lower expression represents more effective tiRNAs. The performance was evaluated using both UTR_1 and UTR_2. The 100 % relative protein expression represents the protein expression in absence of the tiRNA. Error bars represent standard deviation (n=3).

3.3.4 Linking features to tiRNA activity

To unravel underlying design principles of repressing tiRNAs, a OLS linear regression analysis was performed in a first approach. To this end, a linear regression model was applied using all data points in the experimental design. All relative expression percentages from all data points of the experimental design are plotted against all normalized features (with only UTR_1 as target) in Supplementary Figure A.7. Relative expression percentages plotted against all absolute features for all data (including the repression percentages of UTR_2) are depicted in Supplementary Figure A.8. Only two factors in the linear model had a significant influence, namely FAB ($p < 0.05$) and PT ($p < 0.1$). Factors FAB and PT also had significant influence on several other reported riboregulator systems with or without the aid of Hfq[49,70,81,132,242,247]. When using only these two features in a linear regression model, only the factor FAB was significant ($p < 0.05$), while factor PT turned out to be not significant ($p > 0.1$). As the factor FAB is based on thermodynamic properties, it was hypothesized that the relation between FAB and the relative protein expression is exponential. Therefore, a linear model was used to relate the logarithmic of relative protein expression percentage to the tiRNA feature FAB (see Equation 3.4). The outcome of this OLS regression is depicted in Figure 3.4. Despite the significant influence of FAB in the DOE, this basic model is still unable to explain all tiRNA functionality which is reflected by the fact that the majority of the data points are not within the 95 % confidence interval of the OLS model.

Data driven approaches using regression methods have previously been successful in biological engineering[107,269] and, more specifically, forward design of various RNA devices[49,81,270]. Therefore, in a second approach, PLS regression was performed. To maximize the information possibly linked to tiRNA activity, the 12 defined features were included.

To perform the PLS regression, all data points from UTR_1 and UTR_2 were split into two subsets: one set used for model calibration, i.e. training set, and one independent set used for model validation, i.e. test set. The latter set was selected by randomly picking tiRNAs from three groups which are ordered based on the averaged gene expression of both UTR_1 and UTR_2. Before regression, the absolute values of the tiRNA features were scaled through division by the sample standard deviation. Model calibration was done using all 30 data points from the training set and uses 12 features ($k = 12$) describing tiRNA performance. The final model contained 4 latent variables and was selected based on the root mean squared error of prediction and the explained Y variance. By using the training set, a final PLS model contains 63.9 % of the X variance, which explained 50.4 %
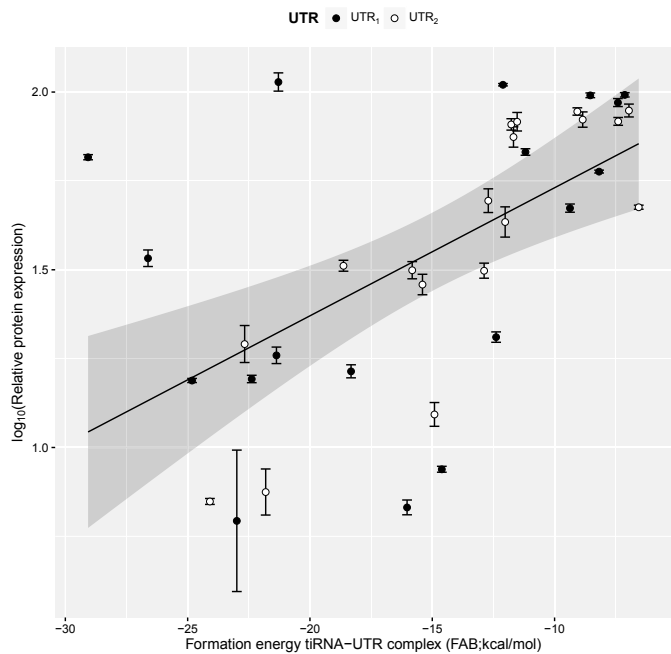
Figure 3.4: Plot of the ordinary least squares (OLS) regression of the linear model, linking $\log_{10}$ of
the relative protein expression to the translation inhibiting RNA (tiRNA) feature formation energy of
the tiRNA-UTR dimer (FAB). All data points were used, including the effect of the tiRNAs on both
untranslated region 1 (UTR_1) and UTR_2. The gray area depicts the 95 % confidence interval of the
OLS linear regression. Error bars represent the standard deviation (n=3).

variance of the response variable and a R^2 (describing the model efficiency) of 0.50 (see
Figure 3.5). To validate this PLS model, the independent validation set was used to
assess the quality of the PLS model. The R^2 of this validation set was 0.69, indicating that
the model successfully explains tiRNA activity. To identify the most important factors
in the PLS regression model, all estimated regression coefficients are calculated. The
regression coefficients of the 12 tiRNA features are shown in Supplementary Figure A.9.
The cumulative loadings of the 4 components and the biplot of the first two components
are depicted in Supplementary Figure A.10 and A.11, respectively. From these estimates
the formation energy of the UTR-tiRNA complex is again inversely correlated to the final
protein expression as both regression coefficients of EAB and FAB are positive. This link
between dimer stability and riboregulator performance was also previously observed

in other RNA devices[49,132]. Other observations are the negative relation between FAA
and protein expression, indicating that a stable tiRNA-tiRNA dimer complex decreases
tiRNA efficiency. Besides these thermodynamic factors, structural features PAU and PT
are inversely correlated to protein expression. Thus, as described in literature, target
nucleotide availability and the number of paired termini in the riboregulator monomer
is important for repression efficiency[82,247,248]. Contrary to previous studies, activation
energy and total RBS occlusion has a rather limited influence on gene repression.

Overall, the PLS modelling approach employed here successfully predicts tiRNA activ-
ity based on the described 12 features, which were defined based on literature. How-
ever, various features used in previously described efforts were quantified using differ-
ent methods[49,132]. This lack of standardized methods to determine thermodynamic and
structural features of riboregulators complicates forward engineering of riboregulators.
Also, the diverse range of features required to explain tiRNA functionality is an indi-
cation of the complex nature of the regulatory mechanism of riboregulation. As such,
RNA regulation might require properties unknown today, which might be discovered
using recently developed technologies allows detailed structural analysis of riboregula-
tors with a high-throughput. For instance, SHAPE-Seq allows *in vivo* characterization of
RNA structure by coupling chemical probing techniques to next-generation sequencing
technology[271,272].
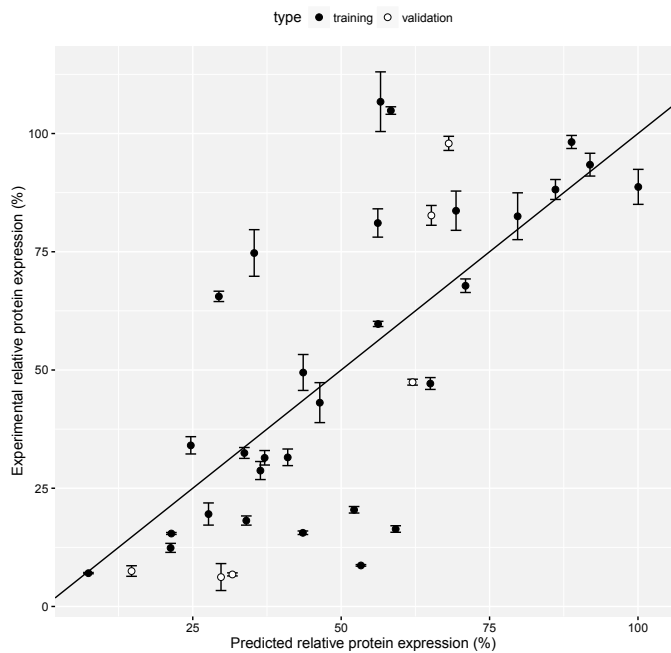
Figure 3.5: Validation of partial least squares (PLS) regression model predicting relative protein expression from 12 predictors (features of translation inhibiting RNA (tiRNA)). Plot of experimental versus predicted relative protein expression via PLS model for the training set (used for model calibration) and the validation set (used to test the model efficiency; coefficient of determination (R^2) equal to 0.69). Error bars represent the standard deviation (n=3).

3.4   Conclusions

The developed approach allows *de novo* design of translation inhibiting riboregulators,
which further broadens the RNA regulator toolbox. From the 18 constructed tiRNAs
molecules designed in the DOE eight tiRNAs repressed protein production with more
than 75 %. In contrast to all previously described pure riboregulator efforts for trans-
lation inhibition, these riboregulators are not derived from a naturally functional sRNA
chassis, which indicates the broad applicability of the findings of this study. To fur-
ther improve riboregulator design several basic modelling approaches were employed.
However, these basic efforts were unable to fully explain tiRNA performance, indicating
the complexity of riboregulator repression. Previous efforts often rely on several cri-
teria to engineer riboregulators of various types with varying success[49,80,81,132]. Based
on these efforts, 12 features were defined and used in a DOE to explore the tiRNA
feature space. Subsequently, to improve the reliability of *de novo* forward engineering
of repressing riboregulators, a sequence-function model was constructed to link tiRNA
functionality to the defined tiRNA features. To this end, both structural and thermody-
namic tiRNA features were used in a PLS regression model, which was evaluated using
an independent test set (R^2 equal to 0.69). The success of this data driven approach
indicates the importance of machine learning techniques in modern synthetic biology
to grasp the ever increasing complexity of biological design. Furthermore, the complex
nature of riboregulation and the limited knowledge of the underlying working mecha-
nisms makes engineering RNA devices challenging. To this end, novel technologies (for
instance SHAPE-Seq) enable high-throughput study of the structure-function relation-
ship of various types of riboregulators in detail by combining RNA structural probing
techniques and next-generation sequencing technology, allowing better prediction of
riboregulator performance[271,272].

Chapter 4

Computer-aided development of ligand-responsive RNA devices

Contents

Abstract

Technologies to precisely modulate gene expression in response to ligand concentrations form indispensable tools in biotechnology. In this context, recently emerged ligand responsive riboregulators, more specifically riboswitches, are an attractive alternative to the traditional protein regulators. However, the development of these regulators heavily relies on high-throughput screening, which makes the development process of these riboswitches very laborious. To overcome this limitation, an *in silico* approach was developed to help create *in vivo* functioning riboswitches from *in vitro* selected aptamers. First, design principles were derived from exisiting riboswitches, which forms the basis of the used objective function. This objective function quantifies potential riboswitch activity using RNA structural and thermodynamical predictions, which enables *in silico* screening for riboswitches with these properties. To estimate the effectiveness of this method, 29 potential riboswitches were computationally selected with various characteristics, which were subsequently constructed and evaluated *in vivo*. As a proof of concept, a theophylline aptamer was used to create *in vivo* functional riboswitches. The vast majority of the created riboswitches were functional out of the box (12 out of the 29 created riboswitches have an activation ratio (AR) above five) with ARs comparable to riboswitches developed trough the laborious screening procedures. Despite the fact that the algorithm developed here allows the design of functional riboswitches with a reasonable chance, globally linking the AR of the created riboswitch to structural and thermodynamic properties remains challenging.

4.1 Introduction

Over the past few years, tremendous progress in the field of synthetic biology spurred numerous breakthroughs in metabolic engineering, cell therapy, disease control and diagnostics[30,52,273,274]. These efforts are mainly supported by technologies to precisely engineer gene expression in response to various types of input signals. In this context, ligand responsive devices, capable of specifically sensing small molecules and subsequently altering gene expression, form indispensable tools in various research fields. For instance, biosensors capable of measuring intracellular metabolites facilitate efficient strain development by enabling selection of good producers from mutant libraries[156]. Also, ligand responsive devices allow the construction of complex genetic circuits dynamically controlling cellular metabolism for biomedical or metabolic engineering purposes[202,275]. The majority of available biosensors are based on naturally occurring transcription factors, either used with their natural ligand or expanded to non-native ligands through protein engineering[157,276,277].

Recently, synthetic RNA devices have emerged as a versatile alternative to these traditional protein based biosensors for various biotechnological purposes[241]. Regulating gene expression by RNA is particularly attractive due to the advanced state of RNA folding algorithms and the relatively clear relationship between structure and biological function. Additionally, RNA regulators control expression on the transcriptional and translational level, in various configurations, and by various regulatory mechanisms which enables construction of a wide range of synthetic devices for numerous purposes. One of these types of regulation are riboswitches, which are naturally occurring RNA molecules controlling gene expression in response to specific small molecules (ligands)[15,158]. These types of ligand responsive RNA devices rely on an aptamer part to specifically interact with small molecules. These aptamers, either naturally occurring or developed synthetically through the SELEX procedure, can be used to develop gene switches on one transcript, in *cis*, or in *trans* when expressed elsewhere[79,149,180,182,197]. However, the development of these ligand sensing RNA devices heavily depends on screening combinatorial libraries and expert knowledge, hindering predictable forward engineering of genetic circuitry. Moreover, these trial-and-error approaches become impractical as the complexity of programmed biological systems increases, emphasising the need for automatically designable RNA devices[25].

One interesting group of riboswitches are translational riboswitches, regulatory segments of the 5' untranslated region (UTR) containing a sensing domain (aptamer) for

ligand interaction. The supposed working mechanism of translational (activating) ribo-switches, which is derived from literature, is depicted in Figure 4.1[178,180,195,278]. Upon transcription, the ligand binds to the aptamer and the mRNA gets trapped in an open (ON) configuration, which has an accessible ribosome binding site (RBS) hereby allow-ing translation. When the ligand is absent, the closed (OFF) configuration forms and the RBS gets occluded, blocking ribosome binding and translation.
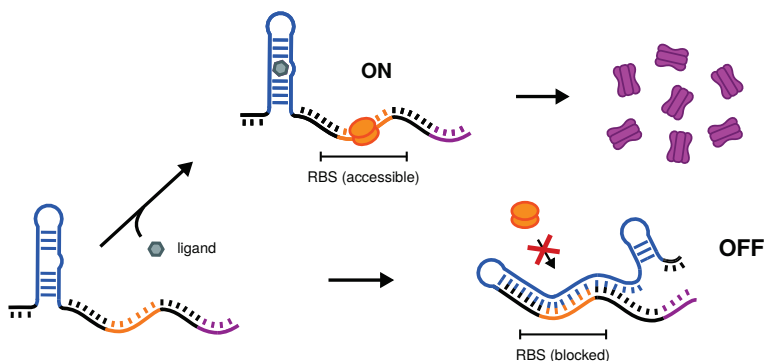
Figure 4.1: Schematic overview of the working mechanism of translational riboswitches used in the design algorithm. The translational riboswitch functions as followed: the mRNA is transcribed and the riboswitch ends up in the ON or OFF state depending on the availability of the specific ligand. When the specific ligand is available, its binding to the aptamer makes the ribosome binding site (RBS) accessible for ribosome binding. Contrary, when the ligand is unavailable, the riboswitch ends up in the minimum free energy (MFE) structure, which blocks the RBS from ribosome binding.

Various efforts have shown the potential of automated *de novo* design of functional ri-boregulators to build complex genetic circuits[49,81,132]. Related to riboswitches, both translational and transcriptional riboswitches have been developed using computer aided design. Both approaches use thermodynamic models to select functional riboswitches *in silico* from *in vitro* selected aptamers[66,279]. However, reliable computational design of effective RNA devices capable of detecting small molecule remains challenging[66,198,279]. Details on the exact working mechanisms of the diverse group of translational riboswit-ches are rarely clear, making forward design of riboswitches challenging[178,180,195,278]. As such, synthetic riboswitch mechanisms have been attributed to both kinetic and ther-modynamic models[178,180,195,279]. Apart from the riboswitch mechanism, ligand induced RBS occlusion also plays an important role in the gene expression modulation of various riboswitches[178,180,195].

In this work, we developed a computational method to aid in the design of *in vivo* func-

tional translational riboswitches from *in vitro* selected aptamers. To this end, a computational workflow was created using RNA bioinformatics to evaluate the riboswitch characteristics of UTRs, which comprises both structural and thermodynamical features. Subsequently, this workflow was evaluated by testing several selected riboswitch candidates *in vivo*. As proof of concept, various riboswitches were selected using the previously *in vitro* selected theophylline aptamer[185]. To improve the predictability of the developed method, various riboswitch properties used in the computational design phase were evaluated to explain riboswitch performance.

4.2   Methods

4.2.1   *In silico* design of translational riboswitches

For the automated design of translational riboswitches a search algorithm was used to find the global minimum of a defined objective function (see 4.3.1), which comprises both thermodynamic and structural terms derived from RNA secondary structure predictions. The initial riboswitch candidate contains completely randomized nucleotides for all regions except the fixed sequence constraints. Every new candidate is created by mutating 1 to all (randomly chosen) possible mutable (defined in the sequence constraints) nucleotides. The riboswitch performance of mutated candidates are quantified using the objective function and subsequently, evaluated using a Metropolis Monte Carlo simulated annealing algorithm[280].

Details on the search algorithm are available in Supplementary Methods (see section B.1.1). All RNA thermodynamics were predicted using RNAfold[256] from the Vienna RNA package[258] with the options –noLP -d2 and the accuracy was set to 10^{-100}, all other settings are set to the default setting. Gibbs free energies mentioned in this study are ensemble free energies. All calculations were done using Python scripting on a Intel Xeon E5-2670 (2.60GHz) Linux (Debian) server.

4.2.2   Strains, plasmids and growth conditions

Unless otherwise stated, all products were purchased from Sigma-Aldrich (Diegem, Belgium). *E. coli* strain DH10B (Invitrogen) was used for both plasmid construction and fluorescence measurement purposes. For plasmid construction and fluorescence measurements strains were grown in lysogeny broth (LB) and MOPS EZ rich medium (Teknova, Bioquote, York, United Kingdom) at pH 7.4, respectively at 37°C with shaking. LB was composed of 1 % tryptone-peptone (Difco, Erembodegem, Belgium), 0.5 % yeast extract (Difco, Erembodegem, Belgium) and 1 % sodium chloride (VWR, Leuven, Belgium). LB agar (LBA) plates contain the same components as LB with the addition of 1 % agar. As required, medium was supplemented with 100 μg ml^{-1} ampicillin.

Details of the constructed plasmids are presented in Supplementary Figure B.1 and Supplementary Figure B.2. All plasmids used in this study were constructed using CPEC[254] assembly. All translational riboswitches were constitutively expressed from a medium-copy vector (pBR322 origin of replication and ampicillin resistance marker, originating from pSB6A1[249]) using pFAB39[23] as promoter, *mKate2*[252] as reporter gene, and rnpB1

87

T1 as terminator[24]. The reporter mKate2 was used due to its low background and good fluorescent protein properties (brightness and maturation time)[252]. Between this promoter and the reporter gene riboswitch sequences were inserted seamlessly using single stranded DNA assembly[117]. Details of the plasmids and important DNA sequences used in this study are listed in Supplementary Table B.1 and B.2, respectively. DNA oligonucleotides were commercially ordered from IDT (Leuven, Belgium) and DNA sequences of every constructed plasmid were verified using sequencing services (Macrogen Inc., Amsterdam, The Netherlands).

4.2.3   *In vivo* fluorescence and optical density (OD) measurements of riboswitches

For *in vivo* assessment of translational riboswitches, strains were plated on LBA plates containing $100\,\mu$g ml^{-1} ampicillin and incubated overnight at 37°C. From each strain three colonies were inoculated in $150\,\mu$l MOPS EZ rich medium containing ampicillin and grown overnight on a Compact Digital Microplate Shaker (Thermo Scientific) at 800 rpm. Subsequently, these cultures were 1:300 diluted in $150\,\mu$l of fresh MOPS EZ rich medium in absence and presence (2 mM) of theophylline and grown until stationary phase (12 h) on a Compact Digital Microplate Shaker (Thermo Scientific) at 800 rpm. Finally, fluorescence and OD measurements were performed using a Tecan m200 pro microplate reader (for mKate2: excitation 588 nm and emission 633 nm). Precultures were grown in Greiner bio-one (Vilvoorde, Belgium) polystyrene F-bottom 96 well plates. For all growth experiments sterile 96-well flat-bottomed microtiter plates (Greiner, Leuven, Belgium) were used, which were covered using a breathe-easy sealing membrane (Sigma-Aldrich). Fluorescence and OD measurements were performed after growth in Greiner bio-one (Vilvoorde, Belgium) black μclear 96 well plates. All microtiter plates contained a control without fluorescent protein expression for background correction.

4.2.4   Fluorescence data analysis

First, fluorescence measurements were normalized through division by OD at 600 nm (FP/OD). On every 96-well microtiter plate *E. coli* DH10B cells without fluorescent protein expression (contains pBlank) were grown in MOPS EZ rich medium. The fluorescence of this strain without fluorescent protein expression (FP_{bg}/OD_{bg}) was used to correct for the background fluorescence of *E. coli*. For all strains fluorescence per OD

(arbitrary units) was calculated relative to the blank as shown in Equation 4.1.

$$\left(\frac{FP}{OD}\right)_{relative} = \frac{FP/OD}{FP_{bg}/OD_{bg}}$$

(4.1)

From the $FP/OD_{relative}$ in the presence of $0\,mM$ and $2\,mM$ theophylline the activation ratio (AR) is calculated:

$$AR = \frac{(FP/OD)_{relative} \text{ with } 2\,mM \text{ theophylline}}{(FP/OD)_{relative} \text{ with } 0\,mM \text{ theophylline}}$$

(4.2)

4.3 Results and Discussion

Translational riboswitches are post-transcriptional RNA regulators, which control gene expression on the RNA level based on specific binding of the molecule of interest to the aptamer part of the riboswitch. The algorithm developed creates an *in vivo* functional riboswitch by designing the nucleotide regions up- and downstream of the aptamer sequence, enabling translation when the ligand is available and RBS occlusion when the ligand is absent. The search for a translational riboswitch modifies all nucleotides in the nucleotide regions that are designed by the algorithm until a suitable solution is found (see Figure 4.2). These designed regions are both located in the 5' UTR and linked by the aptamer sequence, which is fixed and based on a previously developed aptamer. As a proof of concept, the theophylline aptamer, an aptamer previously employed to create *in vivo* riboswitches, was used to create *in vivo* functional riboswitches[178,180,185,195].
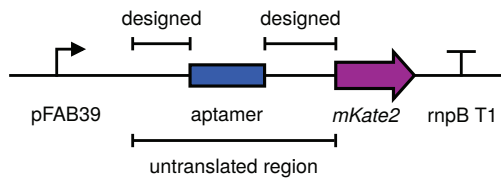
Figure 4.2: The designed part of the translational riboswitch, depicted in DNA. The UTR is designed during the search for a functional riboswitch, comprising both up- and downstream regions of a fixed aptamer region. This designed riboswitch is subsequently expressed using the pFAB39 promoter[23], *mKate2* gene[252], and the rnpB T1 terminator[24]

The computational method to design translational riboswitches comprises an objective function and a search algorithm. The objective function is used to estimate the performance of a certain riboswitch candidate, which was used to evaluate possible riboswitch activity of a specific UTR *in silico*. The objective function used here to evaluate potential riboswitch activity was defined using prior knowledge on the working mechanisms of synthetic riboswitches[178,180,182,195,279], comprising both thermodynamical factors and structural properties. Next, a search algorithm based on simulated annealing is used to efficiently search the large sequence space to find a minimum of the objective function[280].

To this end, nucleotide sequences were *de novo* designed up- and downstream of the *in vitro* selected aptamer sequence (see Figure 4.3A). Within these sequence constraints, the simulated annealing algorithm searches for the best possible riboswitch candidate

(comprising the whole UTR) according to the objective function (see Figure 4.3B). These designed sequences perform best as riboswitch, according to the objective function. Finally, these potential riboswitches are expressed *in vivo* and evaluated by determining the activation ratio (AR), which is defined as the ratio of the OD normalized fluorescence in the presence (2 mM) and absence of theophylline (see Figure 4.3C).

4.3.1 Development of the objective function

Similar strategies to automatically design RNA devices *de novo* based on an objective function containing thermodynamic and structural properties were previously successfully applied to activating riboregulators and transcriptional riboswitches[66,132]. However, contrary to the numerous applications of translational riboswitches in biotechnology, the computer-aided design of translational riboswitches to speed up the development of these riboswitches is still in an infant state.

To determine the relevance of thermodynamic and structural properties used in the objective function, design principles were derived from literature and data on various previously described riboswitched efforts was used to evaluate these design criteria. The relation between riboswitch performance and thermodynamic features of UTRs was verified by analyzing previously described riboswitches *in silico*[195]. First, the MFE structure (theoretically the most abundant structure in the Boltzmann ensemble) was predicted along with the corresponding Gibbs free energy (ΔG_{MFE}). Second, the Gibbs free energy is determined with the aptamer part of the riboswitch constrained to the secondary structure which allows ligand binding ($\Delta G_{CONSTRAINED}$). During these calculations the first nucleotides of the CDS is taken into account, which is important since these nucleotides of the CDS have a significant influence on translation initiation[25]. Next, the difference in Gibbs free energy was calculated and plotted against the AR observed by Mishler & Gallivan[195], which were determined using a different reporter gene than that used in this study. The estimated difference in Gibbs free energy ($\Delta\Delta G$) is related to the AR, as shown in Supplementary Figure B.3.

Because of this important role of thermodynamics in the riboswitch functionality, thermodynamic properties ΔG_{MFE} and $\Delta G_{CONSTRAINED}$ were included in the objective function for translational riboswitch design (see Supplementary Figure B.4A). For the calculation of $\Delta G_{CONSTRAINED}$ the constrained aptamer part of the riboswitch is defined as the secondary structure of this aptamer part that allows ligand binding. The thermodynamic
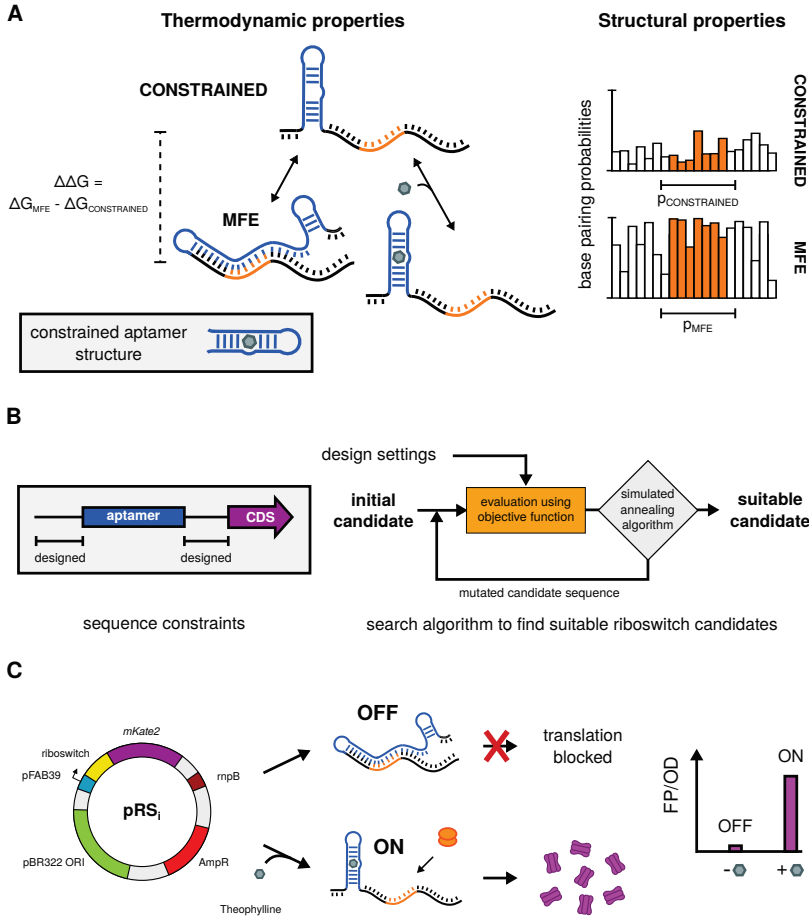
Figure 4.3: A) Schematic overview of the criteria derived from previously described riboswitches [178,180,182,195,279], which form the core of the objective function for automated riboswitch design. These criteria comprise thermodynamic properties like $\triangle\triangle G$, the difference between Gibbs free energy in the minimum free energy (MFE) configuration and the Gibbs free energy in the aptamer constrained configuration and structural properties like untranslated region (UTR) base pairing probabilities in the constrained ($p_{CONSTRAINED}$) and the MFE (p_{MFE}) configuration. B) Representation of the workflow to find riboswitch candidate controlling a coding DNA sequence (CDS) of interest by efficiently searching the sequence space, which best suite the desired objectives. Candidates generated based on the defined sequence constraints are evaluated by an objective function, which is based on user defined design settings, and subsequently evaluated by the simulated annealing algorithm, ultimately resulting in a suitable candidate. C) Overview of the plasmids constructed and the *in vivo* evaluation of the riboswitch performance of the *in silico* selected candidates.

property for the design of translational riboswitches is defined as followed:

$$\Delta\Delta G = \Delta G_{MFE} - \Delta G_{CONSTRAINED} \qquad (4.3)$$

During the design phase, the thermodynamic score ($score_{\Delta\Delta G}$) is calculated as followed:

$$score_{\Delta\Delta G} = |\Delta\Delta G_{target} - \Delta\Delta G_{estimated}| \qquad (4.4)$$

Where $\Delta\Delta G_{target}$ is the target $\Delta\Delta G$ set for the design of riboswitches and $\Delta\Delta G_{estimated}$ is the estimated $\Delta\Delta G$ for a specific UTR.

Besides thermodynamical properties, riboregulation typically relies on structural changes, which subsequently alter gene expression post-transcriptionally and translation initiation [127,132]. Structural fluctuations involve changes in base pairing probabilities, which are estimated using RNA bioinformatics tools [256,257]. Here, the switching ability of previously described riboswitches is analyzed by calculating the base pairing probabilities in the MFE (p_{MFE}) and in the aptamer constrained secondary structure ($p_{CONSTRAINED}$). The constrained structure contains the structure of the aptamer part, which is constrained to the structure that allows binding to the corresponding ligand.

To verify the influence of structural changes, all base pairing probabilities were calculated for both the MFE and the constrained configuration of previously described riboswitches (see Supplementary Figure B.5) [195]. There is a clear difference in base pairing probabilities between the MFE and constrained configuration in the nucleotide region downstream of the aptamer region with less base pairing in the constrained configuration. The upstream region of the aptamer region locally interacts with the 5' part of the aptamer region in the MFE configuration. In the constrained configuration there are interactions between the fixed up- and variable downstream regions of the aptamer region (see Supplementary Figure B.5). Besides the defined switching region in the riboswitch, the objective function also comprises a structural term to prevent undesired interactions between the CDS and the riboswitch, which would possibly limit the modularity of the riboswitch. Namely, the $n_{unpaired}$ region is defined as the first nucleotides of the CDS of the downstream gene.

Overall, the structural properties in the objective function are base pairing probabilities of the total ensemble in both the constrained ($p_{CONSTRAINED}$) and the MFE state (p_{MFE}) (see Supplementary Figure B.4B). For the design of translational riboswitches two regions of target nucleotides are defined ($n_{switching}$ and $n_{unpaired}$), which are the nucleotide regions that are involved in the switching motion and are unpaired in the MFE configu-

ration, respectively:

$n_{switching}$: nucleotides of UTR with p == 0 in the constrained state

and p == 1 in the MFE state.

$n_{unpaired}$: nucleotides of UTR with p == 0 in the MFE state.

The $score_{switching}$ is defined as:

$$score_{switching} = \frac{\sum_{i=1}^{n_{switching}} 1 - p_{MFE} + p_{CONSTRAINED}}{n_{switching}} \tag{4.5}$$

The $score_{unpaired}$ is defined as:

$$score_{unpaired} = \frac{\sum_{i=1}^{n_{unpaired}} p_{MFE}}{n_{unpaired}} \tag{4.6}$$

Overall the objective function is defined as followed:

$$score_{total} = score_{\Delta\Delta G} w_1 + score_{switching} w_2 + score_{unpaired} w_3 \tag{4.7}$$

Where w_1, w_2, and w_3 are weights equal to 1, 1 and 0.5 respectively. These weights were assigned because preventing the CDS from interacting with the riboswitch was considered a secondary goal, besides the primary goal of designing a functional riboswitch. All scores were normalized to 5 (arbitrary chosen) before calculating $score_{total}$

4.3.2 Design of riboswitch candidates

To verify the performance of this computer-aided design strategy, various design targets were chosen, based on these design settings riboswitches were selected and subsequently tested *in vivo* (see Figure 4.3C). As a proof of concept, riboswitches are designed based on the *in vitro* selected theophylline aptamer (GGUGAUACCAGCAUCGU-CUUGAUGCCCUUGGCAGCACC) [185], which is an aptamer that has previously been used to create *in vivo* functional riboswitches [178,180,195]. To evaluate riboswitch activation, 2 mM theophylline is added to the medium, which represents a much lower intracellular concentration. Previous research on theophylline uptake showed that a external concentration of 10 μM results in an intracellular concentration of 7 nM, which increases

linearly until an extracellular concentration of 5 mM [177,281]. The required input settings for riboswitch design are $\Delta\Delta G_{target}$ and the nucleotide regions where the riboswitch switches between the bound and the unbound state ($n_{switching}$). Another input setting, the nucleotide region where the MFE configuration (OFF state) should be unbound ($n_{unpaired}$), is fixed to the first 12 nucleotides of the CDS. Details of all input settings used for the designed riboswitches are shown in Supplementary Table B.5.

First, the effectiveness of the search algorithm was evaluated. The evolution of the accepted score during the search for a minimum in the objective function are shown in Supplementary Figure B.6 for four randomly chosen designs, showing the convergence of the algorithm and the ability of this algorithm to effectively find a minimum of the objective function. Riboswitches RS_1 to RS_{29} are designed using these different input settings, these riboswitches are labeled "designed". To enlarge the explored riboswitch sequence space, various nucleotide sequences up- and downstream of the aptamer region were shuffled in new riboswitch designs (RS_{30} to RS_{38}), these riboswitches are labeled "reshuffled" (see Supplementary Figure B.6).

For all the designed riboswitches the $score_{total}$, $score_{\Delta\Delta G}$, $score_{switching}$, and $score_{unpaired}$ are pairwise plotted in Supplementary Figure B.7. This analysis shows that the energy criterium ($score_{\Delta\Delta G}$) is generally easier to satisfy than the base pairing criterium, meaning that obtaining a base pairing switch in 100 % of the structures in the Boltzmann criteria is harder than obtaining an exact difference in ΔG between the ON (constrained) and OFF (MFE) state ($\Delta\Delta G_{estimated}$). Consequently, the degree of base pairing switch determines the absolute value of the minimal $score_{total}$. A 3D scatterplot of all three scores of the designed riboswitches does not show any correlation between these scores and the ARs (see Supplementary Figure B.8).

4.3.3 *In vivo* riboswitch expression system

To evaluate the performance of the algorithm for translational riboswitch design, first an expression system was established to effectively assess riboswitch performance. This system requires both high and stable expression of riboswitches to sensitively detect the quality of the designed riboswitch. To this end, ARs are maximized using a strong constitutive transcriptional promoter to allow sensitive detection of functional riboswitches. More specifically, the promoter used for riboswitch expression is pFAB39 from the Bio-FAB promoter library, which previously showed high transcriptional activity [23]. Another important factor is the reporter gene, which needs to be easily detectable and sensitive. The reporter gene of choice is *mKate2*, encoding a far-red fluorescent protein, which has

a fluorescent spectrum with little interference of the host background[252]. Also, the riboswitch expression system comprises the medium copy pSB6A1 vector with the pBR322 origin of replication (ori), which is heavily regulated to guarantee stable protein expression[282]. This system was validated by expression of the previously described riboswitch D for theophylline[195], which has a AR of 157 ± 3 in this expression system, which is higher than the AR reported using the luciferase reporter system, showing the required sensitivity of this novel riboswitch expression system comprising several synthetic biological parts[195]. This riboswitch expression system was used for the *in vivo* evaluation of all riboswitches described in this study.

4.3.4 Evaluation of the developed *in silico* workflow

All designed and reshuffled riboswitches were constructed and tested *in vivo* along with various previously described riboswitches[195,283]; the ARs are shown in Figure 4.4. The observed fluorescence in the presence of 0 mM and 2 mM theophylline is depicted in Supplementary Figure B.9.

The designed and reshuffled riboswitches almost all activate translation, as only three constructed riboswitches have ARs not significantly higher than one (p = 0.05). All other riboswitches have ARs significantly higher than one. Moreover, 12 out of 29 designed riboswitches (41 %) have ratios above five, which is approximately the minimum AR of the riboswitches selected from literature and thus serves as a criterium for riboswitch performance. . Riboswitches design RS_1 and RS_2 have the same upstream sequence as the previously described riboswitches, which resulted in good ARs, comparable to riboswitches from literature (both above 50). These riboswitches (RS_1 and RS_2) are the only designs that contain a spacer upstream of the fixed theophylline part, indicating the importance of a good designed spacer between transcription start site and the aptamer part of a riboswitch. The best *de novo* developed riboswitch is RS_{32}, a reshuffled riboswitch, which has a AR of 86, higher than most of the previously described manually optimized riboswitches. The reshuffled RS_{32} contains the up- and downstream region of the aptamer part of RS_{20} and RS_{23}, respectively. Remarkably, both RS_{20} and RS_{23} have an AR below five, once again indicating the complex sequence-structure-function relationship of translational riboswitches.

4.3.5 Comparison with published riboswitches

Most *de novo* developed riboswitches have ARs that are much lower than those observed from riboswitches derived in literature. Despite the lower performance of the automati-
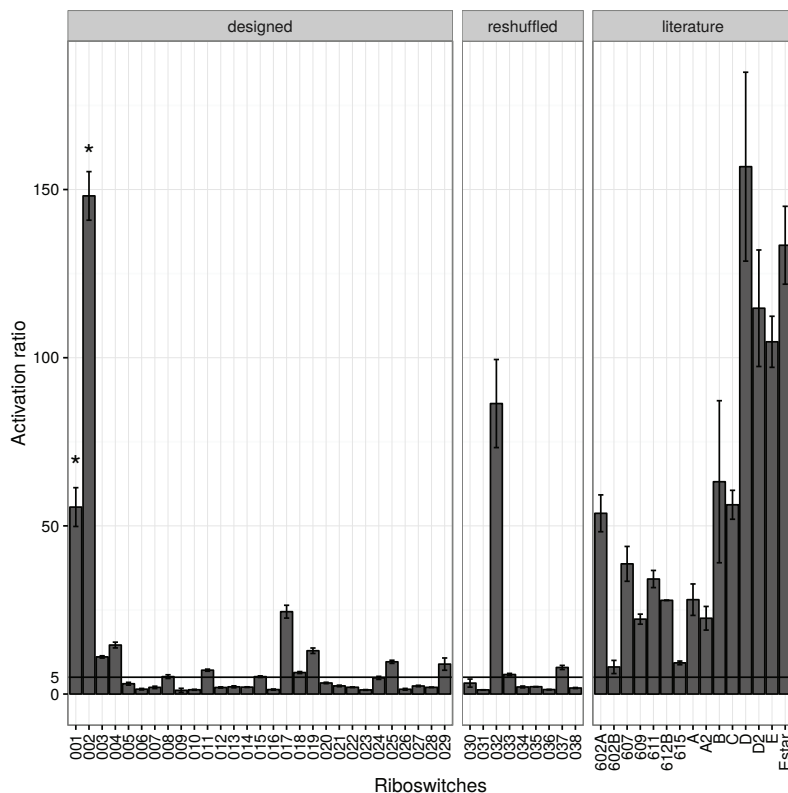
Figure 4.4: Barplot of the activation ratios of the riboswitches that were designed, reshuffled or defined in literature. All automated designed riboswitches are displayed on the left and previously described riboswitches are shown on the right. The riboswitches in the middle are constructed by reshuffling the regions up- and downstream of the aptamer of designed riboswitches. Both the designed, reshuffled, and the previously described riboswitches are expressed as depicted in Figure 4.3C. The horizontal line indicates a AR of 5. RS_1 and RS_2 have the same fixed spacer region upstream of the theophylline aptamer (*).

cally designed riboswitches here, it is important to note that the riboswitches from literature are developed through several rounds of high-throughput screening and manual optimization (largely dependent on trial and error approaches) [178,180,195,283]. In contrast to these laborious techniques, the riboswitches *de novo* designed in this study were created in a few days of computation and cloning, without any further optimization. To explain the lower ARs, the AR were plotted against the fluorescence emitted in the

ON and OFF state of the designed riboswitches, see Supplementary Figure B.10 and Supplementary Figure B.11, respectively. When comparing the fluorescence emitted in the ON state, the *de novo* designed riboswitches span a broader range of fluorescence compared to the riboswitches from literature. Thus, several designed riboswitches reach much higher levels of translation initiation while showing much lower ARs. This contradiction is explained by the fluorescence emitted in the OFF state, which is much lower for the riboswitches from literature than the designed riboswitches (see Supplementary Figure B.11). These designed riboswitches show high conditional protein translation, which is useful in several biotechnology fields. This aspect of riboswitch design is often neglected as the typically used evaluation criterium is the AR. The lower basal expression of the riboswitches from literature indicates a strict regulation when there is no ligand present. However, the fluorescence emitted in the OFF state is not correlated to the Gibbs free energy of the MFE configuration nor to the hybridization energy between 3' part of the 16S rRNA (ACCUCCUUA), $\Delta\Delta G$ and $\Delta\Delta G_{16S\ rRNA}$.

To unravel design principles, several thermodynamic features were calculated for the expressed riboswitches, both automated designed and from literature. Among these features the previously mentioned ensemble free energies in the MFE and the constrained conformation, ΔG_{MFE} and $\Delta G_{CONSTRAINED}$, respectively. But also the hybridization energy (estimated using RNAcofold [257]) of the MFE and the constrained configuration with the 3' end of the 16S rRNA ($\Delta G_{16S,MFE}$ and $\Delta G_{16S\ rRNA,CONSTRAINED}$), which has been linked to translation initiation rate [25,279]. Also the differences between the free energies of the two configuration states is considered for both the ensemble free energies and the hybridization energy between 3' part of the 16S rRNA (ACCUCCUUA), $\Delta\Delta G$ and $\Delta\Delta G_{16S\ rRNA}$, respectively. These six features are plotted against the AR of both designed riboswitches and riboswitches from literature (see Supplementary Figure B.12). Although there is no clear correlation between the observed AR and any of the features, it is clear that functional riboswitches have certain characteristics. For instance, all riboswitches have a $\Delta\Delta G$ below -6 kcal/mol. However, this threshold does not guarantee good riboswitch performance as several other determinative factors remain unclear. Remarkably, the previously reported link between $\Delta\Delta G$ and riboswitch performance, which is also shown in Supplementary Figure B.3, does not always apply. More specifically, using the expression system described here, it does not apply for all previously reported theophylline riboswitches, which is in contrast with the results depicted in Supplementary Figure B.3 [195,283].

Moreover, another developed thermodynamic model attempts to quantify the AR of ri-

boswitches using only thermodynamic factors based on the RBS calculator[25]. In comparison to the algorithm described here, which serves as an *in silico* screening method for translational riboswitches, the model described by Borujeni *et al.*[279] attempts to accurately quantify the AR of a riboswitch. However, this model for the prediction of translational riboswitch function is unable to accurately predict the ARs observed in this study (see Supplementary Figure B.13).   This once again illustrates the complexity of translational riboswitches and the infant state of the knowledge on these regulators, hindering the forward engineering capacity. For instance, cotranscriptional folding plays an important role in *in vivo* RNA structure formation, which is always ignored in the forward engineering of RNA riboregulators. In this context, recently emerged cotranscriptional SHAPE-seq technology enables elucidating cotranscriptional folding pathways of riboswitches, which was shown for the *Bacillus cereus crcB* fluoride riboswitch[284]. This technology is a complementary extension of the regular SHAPE-seq technique, which is an advanced next-generation sequencing technology that allows high-throughput RNA structural probing, enabling detailed characterization of the structure-function relationship *in vivo*[271]. At the moment, the sequence space of previously described riboswitches is small due to the small differences between all riboswitches, which leads to a rather narrowly applicable design rules. Using these techniques allows characterization of larger data sets, which enables the construction of a generally applicable sequence-function model. A larger data set can improve the forward engineering capabilities by expanding the sequence space explored in developed riboswitches.

4.4   Conclusions

The created riboswitches were all functional out of the box with several designed ri-
boregulators with ARs comparable to previously described riboswitches. Compared to
these riboswitches, this method had the advantage of speed. It only takes several days
to create these *de novo* designed riboswitches, whereas the riboswitches from literature
were developed trough numerous cycles of laborious optimization procedures. More-
over, the designed riboswitches show lower ARs mainly because of their higher basal
expression. Despite these limitations, these type of automated design methods show
great potential to aid the development of riboswitches. Overall, taking the results from
this study into account it remains challenging to explain riboswitch performance. Even
the most advanced physicochemical models are unable to accurately predict the AR of all
translational riboswitches[279]. Here, the developed automated design method allows *in
silico* selection of riboswitch candidates with a high probability to have a good riboswitch
performance. To increase the reliability of this algorithm, recently developed RNA struc-
tural probing techniques allow high-throughput analysis of translational riboswitches,
enabling unraveling more details on their working mechanism.

Chapter 5

Development of Neu5Ac-responsive biosensors based on the transcriptional regulator NanR

Contents

Abstract

Biosensors, controlling transcription in response to specific compounds, have various applications in metabolic engineering, including the construction of dynamic pathway regulation and high-throughput screening of combinatorial strain libraries. Various biosensors were previously created from naturally occurring transcription factors (TFs), which are largely composed of native sequences without the possibility to modularely optimize the response curve. This lack of design and engineering techniques hinders the development of custom biosensors. Here, novel biosensors were created that respond to *N*-acetylneuraminic acid (Neu5Ac), an important sugar moiety with various biological functions, by employing native and engineered promoters that interact with the TF NanR. To evaluate the effectiveness of the developed biosensors, a Neu5Ac producing strain was constructed. This strain was used to examine the response of the created biosensors to Neu5Ac, showing that seven out of eight created biosensors were functional and exhibited a different response. Moreover, three of these were successfully developed using a novel approach that inserts the NanR binding site in a constitutive promoter, resulting in functional biosensors comprising various modular defined parts without the interference of other TFs. This novel technique to obtain functional biosensors paves the way to modular biosensor optimization, enabling more reliable engineering of response curve characteristics. Overall, the repertoire of biosensors is expanded with seven functional biosensors capable of detecting Neu5Ac and it was shown that these transcriptional biosensors were obtained by combining different defined parts. Moreover, the modular composition of the engineered biosensors enabled modulating the response curve by changing the ribosome binding site (RBS) controlling NanR translation, which resulted in increased repression and activation in the absence and presence of intracellular Neu5Ac, respectively. Finally, the linear response of three created biosensors to an increasing amount of Neu5Ac production was shown, enabling the use of these biosensors for various metabolic engineering purposes.

5.1 Introduction

Over the last decades, biobased chemicals were introduced, driven by environmental concerns, energy security, and competitiveness of the chemical industry[285]. This recent paradigm shift regarding chemical synthesis is largely supported by metabolic engineering, which enables *ad hoc* rewiring of cellular metabolism to produce molecules of interest[286]. Recent advances in synthetic biology drove the development of several new microbial cell factories for molecules which were previously impossible to produce economically efficient[1–5]. These success stories largely depend on novel technologies relating to DNA synthesis, DNA assembly and combinatorial pathway optimization. In this context, the optimization of biosynthetic pathways by reducing flux imbalances, i.e. eliminating excessive intermediate production, is crucial to achieve maximal productivities[29]. Moreover, an optimal pathway flux requires optimal usage of cellular components to minimize metabolic burden and the related productivity losses[245]. Today, numerous techniques are available to combinatorially optimize production pathways by varying the expression of specific pathway genes, which results in different production efficiencies[28]. However, this combinatorial approach to strain optimization heavily relies on the availability of high-throughput screening techniques[29]. To overcome this limitation, various biosensors were created to detect a wide range of molecular specificities, allowing high-throughput screening of large strain libraries in search of mutants that produce the molecule of interest. Besides applications as high-throughput screening, biosensors also enable dynamic pathway regulation and adaptive laboratory evolution to improve product yields[202,287].

Over the last few years, biosensors for a range of molecules were created from naturally available TFs and subsequently applied in metabolic engineering strategies[156,157,288]. However, as most transcriptional biosensors are largely composed of native sequences, techniques to engineer these biosensors are missing, which makes obtaining optimal biosensor behaviour challenging. For instance, naturally occurring TF regulation is typically controlled by both local and global regulators, which causes unwanted interference in specific physiological conditions. This infant state of the transcriptional biosensor engineering field substantially hinders its applicability in metabolic engineering[289]. Also, despite the recent efforts to create transcriptional biosensors with various specificities, the availability of custom biosensors to sensitively detect a particular molecule of interest remains limited[157,288].

For instance, the number of transcriptional biosensors capable of detecting sugar moi-

eties are limited[290,291]. One such interesting group of sugar molecules with various applications are sialic acids, which are a group of neuraminic acid derivatives, a family of nine-carbon carboxylated monosaccharides. In this group of sugar derivatives, Neu5Ac is the predominant sialic acid found in mammalian cells, where it plays a vital role in various biological processes as the terminal ends of numerous glycoproteins and glycolipids[292]. These processes include binding of infectious microorganisms, tumor progression, metastasis formation, immune response regulation, and infant brain development[292–295]. This broad range of biological functions makes Neu5Ac an interesting target molecule for numerous applications in the pharmaceutical and food industry[296–299]. For instance, a Neu5Ac analogue was developed as a potent inhibitor of influenza, which is marketed as Zanamivir[296,297]. Also, Neu5Ac and its derivatives have potential for drug delivery and drug targetting applications[298]. Next to these applications, Neu5Ac and the related sialylated oligosaccharides are important nutrients[294,299,300].

Numerous TFs are known regulators of sugar degradation pathways, which might serve as a basis for biosensor development[301]. As such, there are several TF families known to regulate Neu5Ac degradation pathways in various phyla[302]. Among these regulators the best known TF NanR originates from Enterobacteriaceae, which regulates Neu5Ac related operons. In *E. coli*, the *nanATEK* operon, which is the primary degradation route by converting Neu5Ac in *N*-acetylglucosamine 6-phosphate (GlcNAc6P), is regulated by NanR[303,304]. This helix-turn-helix (HTH) based TF from the FadR/GntR family represses the *nanATEK-yhcH*, *nanCMS*, and *yjhBC* operons when Neu5Ac is not present[304].

Here, various biosensor engineering approaches to create transcriptional biosensors responsive to Neu5Ac were explored based on a naturally occurring TF. Several native promoters were used to construct novel transcriptional biosensors, which contain a fluorescent reporter gene. Subsequently, the native regulatory circuits were deconstructed in defined parts, enabling a more modular biosensor optimization. To obtain transcriptional biosensors with changed response curve characteristics and to remove unwanted interference of other regulators, several engineered biosensors were constructed by a top-down approach (through truncation) and by a bottom-up approach (modular biosensor construction). The latter approach paves the way for modular biosensor optimization of these biosensors by changing the position, multiplicity, and/or sequence of the TF binding site[289]. Furthermore, the optimization possibilities of these modular biosensors and the general applicability of the created biosensors are further explored.

5.2 Methods

5.2.1 Strains

E. coli strain K-12 MG1655 (ATCC) was used as the parent strain for all strain engineering experiments and plasmid construction. All DNA manipulations were performed following standard molecular cloning protocols. Site-directed chromosomal gene deletion was accomplished by homologous recombination using λ-Red recombinase (induced from pKD46)[305]. Linear DNA for recombination containing FRT-flanked antibiotic resistance casettes was generated by PCR amplification from pKD3 or pKD4 using primers to build in site-specific homologies towards the targeted loci. DNA oligonucleotides were purchased from IDT (Leuven, Belgium). Positive transformants were cured from the antibiotic resistance cassette through FLP recombinase (expressed on pCP20) induced recombination of FRT sites. All chromosomal deletions were verified by colony PCR and DNA sequencing (Macrogen Inc., Amsterdam, The Netherlands). The strains constructed were *E. coli* K-12 MG1655 $\Delta nanRATEK$ and *E. coli* K-12 MG1655 $\Delta nanR$. The primers used for gene deletions are described in Supplementary Table C.1.

5.2.2 Growth conditions

Unless otherwise stated, all products were purchased from Sigma-Aldrich (Diegem, Belgium). For strain engineering and plasmid construction strains were grown in lysogeny broth (LB) at 30°C with shaking. LB was composed of 1 % tryptone-peptone (Difco, Erembodegem, Belgium), 0.5 % yeast extract (Difco) and 1 % sodium chloride (VWR, Leuven, Belgium). LB agar (LBA) plates contain the same components as LB with the addition of 1 % agar.

For growth experiments a defined medium was used. This defined medium contained 2 g/L NH_4Cl, 5 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 7.3 g/L K_2HPO_4, 8.4 g/L MOPS, 0.5 g/L NaCl, 0.5 g/L $MgSO_4 \cdot 7\,H_2O$, and 16.5 g/L glucose$\cdot H_2O$, 1 ml/L trace element solution and 100 μL/L molybdate solution. Trace element solution used contained 3.6 g/L $FeCl_2 \cdot 4\,H_2O$, 5 g/L $CaCl_2 \cdot 2\,H_2O$, 1.3 g/L $MnCl_2 \cdot 2\,H_2O$, 0.38 g/L $CuCl_2 \cdot 2\,H_2O$, 0.5 g/L $CoCl_2 \cdot {}_6H_2O$, 0.94 g/L $ZnCl_2$, 0.0311 g/L H_3BO_4, 0.4 g/L $Na_2EDTA \cdot 2\,H_2O$, 1.01 g/L thiamine$\cdot$HCl and was sterilized with a bottle top filter (Corning PTFE filter, 0.22 μm). Molybdate solution (0.967 g/L $Na_2MoO_4 \cdot 2\,H_2O$) was sterilized with a bottle top filter (Corning PTFE filter, 0.22 μm). Carbon source and $MgSO_4$ were dissolved in 200 ml H_2O and autoclaved separately to avoid Maillard reaction. All other components (except trace element and molybdate solution) were dissolved in 800 ml H_2O and set to

pH 7 with KOH. After autoclaving, the sterile trace element and molybdate solution were added. If required, medium was supplemented with 100 μg ml^{-1} ampicillin and 25 μg ml^{-1} chloramphenicol. For all growth experiments longer than 24 h medium was supplemented with 100 μg ml^{-1} carbenicillin instead of 100 μg ml^{-1} ampicillin.

For flask experiments, precultures were grown in 50 ml centrifuge tubes containing 10 ml LB medium with the necessary antibiotic for selection pressure. Pre-cultures were grown overnight (16h) at 30 °C and 200 rpm (LS-X AppliTek orbital shaker, Nazareth, Belgium) and subsequently, used for 1 % inoculation of 100 ml glucose or glycerol defined medium in 500 ml shake flasks and grown at 30 °C and 200 rpm (LS-X AppliTek orbital shaker, Nazareth, Belgium). At regular intervals, samples for extracellular metabolites analysis were collected and optical density (OD) at 600 nm is determined. For 24-well deep well plates (DWP) experiments, precultures were grown in 50 ml centrifuge tubes containing 10 ml LB medium with the necessary antibiotic for selection pressure. Pre-cultures were grown overnight (16h) 30 °C and 200 rpm (LS-X AppliTek orbital shaker, Nazareth, Belgium) and subsequently, used for 1 % inoculation of 3 ml glucose or glycerol defined medium in 24-well DWP plates with sandwich covers (EnzyScreen, Heemstede, The Netherlands) and grown at 30 °C and 200 rpm (LS-X AppliTek orbital shaker, Nazareth, Belgium). In 24-well DWPs, cultures were sampled at regular intervals for extracellular metabolite analysis and OD measurement. OD was measured at 600 nm using a Jasco V-630Bio spectrophotometer (Easton, UK).

5.2.3   High-performance liquid chromatography (HPLC) analysis

Extracellular metabolite concentrations were determined through HPLC analysis (Prostar 230, Varian, Sint-Katelijne-Waver, Belgium) equipped with a refractive index detector. All metabolites, i.e. glucose, glycerol, acetate, Neu5Ac, *N*-acetylmannosamine (ManNAc), and, *N*-acetylglucosamine (GlcNAc), were separated on a Aminex HPX-87H column (Bio-rad, Temse, Belgium) with 25 mM H_2SO_4 as mobile phase at 80 °C and a flow rate of 0.5 ml/min.

5.2.4   Neu5Ac biosensor and pathway constructs

A schematic overview of all plasmid types used is depicted in Supplementary Figure C.1. All plasmids used were constructed using CPEC[254] assembly. Details of the plasmids constructed are listed in Supplementary Table C.2. The DNA sequences of every constructed plasmid was verified using sequencing services (Macrogen Inc., Amsterdam, The Nether-

lands). Details of all regulatory DNA sequences and all coding DNA sequences (CDSs) are listed in Supplementary Table C.3 and C.4, respectively.

Biosensor plasmids

Native biosensor (pNB) plasmids and engineered biosensor (pEB) plasmids (see Supplementary Figure C.2) are medium-copy vectors (p15A origin of replication and chloramphenicol resistance marker, originating from pACYCDuet[306]), which are used to express the TF NanR[307] and the reporter gene *mKate2*[252]. The reporter mKate2 was used due to its low background and good fluorescent protein properties (brightness and maturation time)[252]. pNB_1 contains the native *nanR* operon (the native intergenic region upstream of *nanR* and the *nanR* CDS) and its downstream region, i.e., the intergenic region upstream of *nanA* (see Supplementary Figure C.3 for more details). On pNB_2 through pNB_4 and pEB_1 through pEB_4, NanR was expressed using BBa_J23114[308] as promoter, BBa_B0031[308] as RBS, and rnpB T1[24] as terminator and mKate was expressed using various promoters and the rnpB T1[24] terminator. Plasmids pNB_2 through pNB_4 contain native intergenic regions, i.e. native promoters and untranslated regions (UTRs) reported to contain NanR operators (*nanA*, *yjhB*, *nanC*, respectively) for mKate2 expression (see Supplementary Figure C.2A and C.4 for more details). Plasmids pEB_1 through pEB_4 contain engineered promoters (containing NanR operators) and UTRs for mKate2 expression (see Supplementary Figure C.2B and C.5 for more details). On all pEB plasmids and all pNB plasmids, except pNB_1, NanR and mKate2 were expressed in the opposite direction. All pO plasmids have the same content as their pNB or pEB counterpart, without the expression of NanR.

The three pEB_2 variants were constructed by replacing the BB_B0031 RBS with three different RBSs. As such, the $pEB_{2,1}$, $pEB_{2,2}$, and $pEB_{2,3}$ plasmids contain the BB_B0032[308], BB_B0030[308], and BB_B0034[308], respectively.

Pathway plasmids

The $pPathway_{medium}$ plasmid is a medium-copy vector (pBR322 origin of replication and ampicillin resistance marker, originating from pET22B (Merck Millipore, Overijse, Belgium)) used for the expression of the biosynthetic pathway of Neu5Ac. This pathway contains two enzymes (UDP-*N*-acetylglucosamine 2-epimerase and *N*-acetylneuraminic acid synthase), encoded by the *neuC* and *neuB1* gene, respectively[309]. The expression of these two genes was done polycistronicly, controlled by the (medium) promoter p14[107] and the rnpB T1 terminator[24] . The translation of the NeuB1 and the NeuC protein

is controlled by the T7 RBS[310] and the in-house designed intergenic region[306], respectively. The CDSs were designed for expression in *E. coli* K-12 MG1655 using the COOL algorithm[311]. All details of the pPathway$_{medium}$ plasmid are depicted in Supplementary Figure C.6 and C.7. Two variants of the pPathway$_{medium}$ plasmid were created by replacing the p14 promoter with the pFAB46* (mutated form of the pFAB46 promoter[23]) and pJ23105[308] promoters, resulting in the pPathway$_{high}$ and pPathway$_{low}$, respectively. The pBlank plasmid is the same medium-copy vector as pPathway plasmids (pBR322 origin of replication and ampicillin resistance marker, originating from pET22B (Merck Millipore, Overijse, Belgium)), without any DNA insert.

5.2.5 *In vivo* fluorescence and OD measurements

For *in vivo* assessment of the various transcriptional biosensors, strains were plated on LBA plates containing $100\,\mu g$ ml^{-1} ampicillin and 25 μg ml^{-1} chloramphenicol if required. After overnight incubation, strains were inoculated in $150\,\mu L$ LB and grown overnight on a Compact Digital Microplate Shaker (Thermo Scientific) at 800 rpm and 30 °C. Subsequently, these cultures were 1:120 diluted in $150\,\mu l$ of fresh defined medium and grown on a Compact Digital Microplate Shaker until stationary phase (24 h) at 800 rpm and 30 °C. Subsequently, fluorescence and OD were measured using a Tecan M200 pro microplate reader. Precultures were grown in Greiner bio-one (Vilvoorde, Belgium) polystyrene F-bottom 96 well plates. For all growth experiments sterile 96-well flat-bottomed microtiter plates (Greiner, Leuven, Belgium) were used, which were covered using a breathe-easy sealing membrane (Sigma-Aldrich). Fluorescence and OD measurements were performed after growth in Greiner bio-one (Vilvoorde, Belgium) black μclear 96 well plates. For measuring mKate2 expression an excitation wavelength and emission wavelength of 588 nm and 633 nm were used, respectively. OD was measured at a wavelength of 700 nm to reduce bias in estimates of cell abundance[255].

5.2.6 Fluorescence data analysis

For fluorescence measurements, two types of controls were used on every 96-well microtiter plate: a defined medium blank and *E. coli* K-12 MG1655 cells blank. The medium blank was used to correct the background OD (OD$_{bg}$) of the medium. The fluorescence of the strain without fluorescent protein expression (FP$_{bg}$) was used to correct for the background fluorescence of *E. coli*. For all strains fluorescence per OD was calculated as

follows:

$$\left(\frac{\text{FP}}{\text{OD}}\right)_{\text{corrected}} = \frac{\text{FP} - \text{FP}_{\text{bg}}}{\text{OD} - \text{OD}_{\text{bg}}} \tag{5.1}$$

5.3   Results and Discussion

The biosensors for the detection of Neu5Ac developed here are based on the regulatory action of the TF NanR, which controls the *nanATEK-yhcH*, *nanCMS*, and *yjhBC* operons in *E. coli* (see Figure 5.1)[304]. These gene clusters relate to the degradation of various sialic acids, including the most abundant mammalian sialic acid, Neu5Ac. The most important operon for Neu5Ac degradation is the *nanATEK-yhcH* cluster, which contains genes encoding for an aldolase (*nanA*), a permease (*nanT*), an epimerase (*nanE*), and a kinase (*nanK*)[303,304]. These genes are responsible for the uptake of Neu5Ac into the cytosol and the subsequent conversion into GlcNAc6P, which is further processed into the central metabolite fructose 6-phosphate (Fru6P) by NagA and NagB. The biological function of the last gene in the *nanATEK-yhcH* operon, (*yhcH*) is still unknown. The other clusters controlled by NanR are the *nanCMS* and *yjhBC* operons, which contain several genes with an unknown function. These genes are possibly involved in the conversion of less common sialic acids to Neu5Ac[304,312–314]. For instance, *nanS* encodes for an esterase required for growth of *E. coli* on *O*-acetylated sialic acids[314].
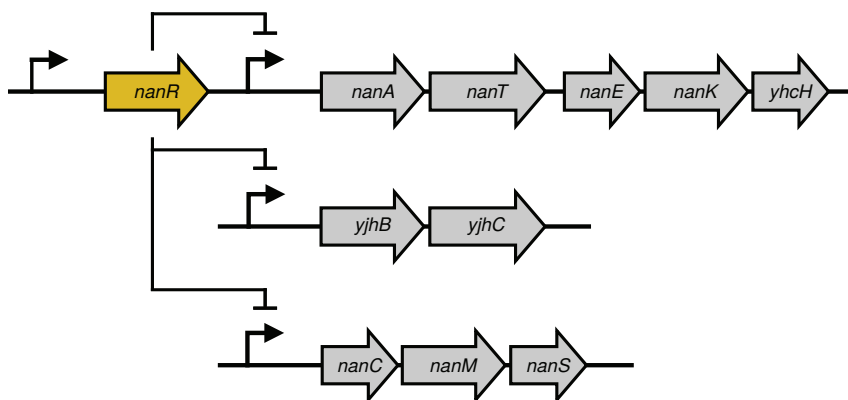
Figure 5.1: Schematic overview of the genetic organization of the sialoregulon, which is negatively regulated by the transcription factor (TF) NanR[304]. The sialoregulon comprises the *nanATEK-yhcH*, *nanCMS*, and *yjhBC* operons, which all relate, to a certain degree, to the degradation of sialic acids[304,312–314].

All these gene clusters are involved in *N*-acetylneuraminic acid (Neu5Ac) degradation and are repressed in *trans* by NanR, which is a TF from the FadR/GntR family of HTH regulators. The underlying binding mechanism of NanR was unraveled by Kalivoda *et al.*[307], who identified the TF operator region. This DNA binding region of NanR (the

nan box) contains three tandem repeats of a hexameric sequence GGTATA[307]. Further details were revealed through transcriptome, genetic, and biochemical analyses, showing the induction of the *nanATEK-yhcH*, *nanCMS*, and *yjhBC* operons by Neu5Ac. Critical nucleotides in the operator sequences were identified, which are required for binding to the HTH domain of NanR[304]. Also, translational fusions of *lacZ* to *nanA*, *yjhC* and, *nanS* showed increased β-galactosidase activity when Neu5Ac was present[304].

5.3.1   Biosensor design using native promoters

Biosensors were constructed to link the intracellular concentration of Neu5Ac to a fast and easily readable output signal without interference of the host organisms background. To this end, the fast maturing fluorescent protein mKate2 was used, which has a fluorescent spectrum with little interference of the host (*E. coli* MG1655) background[252]. Fluorescent proteins allow linking fluorescence to protein production and are typically used to characterize transcriptional biosensors in a sensitive way[156].   To link the production of the fluorescent protein to the intracellular concentration of Neu5Ac, various promoters and UTRs that contain NanR binding sites were tested to control mKate2 expression. The general mechanism of the constructed biosensor is depicted in Figure 5.2A. A schematic overview of mKate2 expression using the NanR-responsive promoters (without NanR expression) is depicted in Figure 5.2B.

The first biosensor (pNB$_1$) was constructed using the *nanA* intergenic region (including the first 36 nucleotides of the *nanA* CDS), containing the NanR operator sequence, CDS, and the native NanR promoter. In a second set of biosensor plasmids (pNB$_2$ through pNB$_4$), three native intergenic regions upstream of the *nanA*, *yjhB*, and *nanC* genes were used as transcriptional promoters and UTRs. To ensure all native NanR regulation is maintained in the biosensor, a translational fusion protein is used. This is achieved by coupling the first 36 nucleotides of the gene downstream of the used native intergenic region to *mKate2*. Contrary to pNB$_1$, NanR is expressed on pNB$_2$, pNB$_3$ and, pNB$_4$ using the BBa_J23114 promoter and the BBa_B0031 RBS, which both have a relatively low activity[308]. In contrast to the pNB$_1$ plasmid, where NanR and *mKate2* are expressed in same direction, NanR and mKate2 (with the first 13 residues of NanA) are expressed in the opposite direction on plasmids pNB$_2$ through pNB$_4$ (see Supplementary Figure C.2A for details on pNB$_1$ through pNB$_4$).
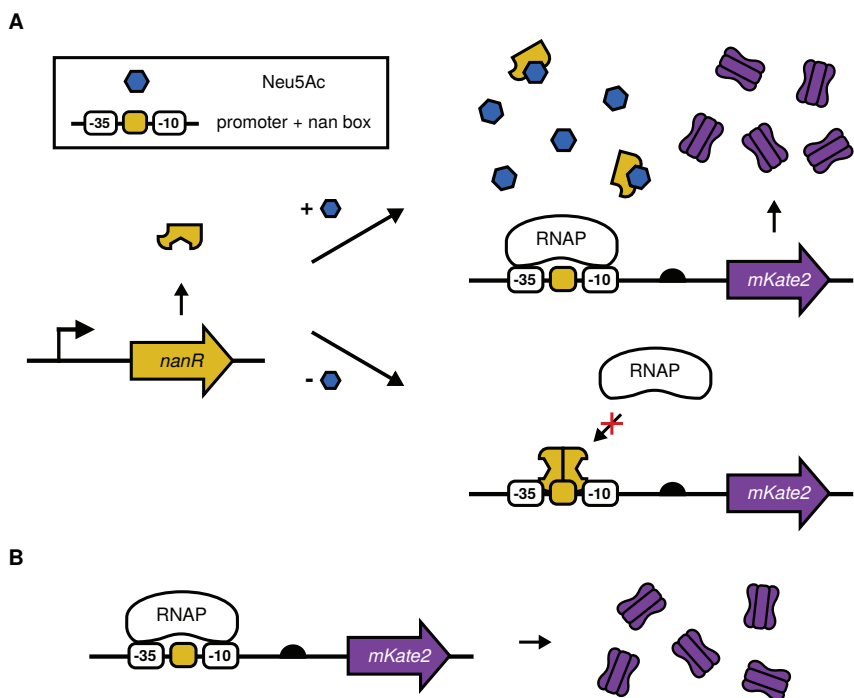
Figure 5.2: A) Schematic overview of the working mechanism of a biosensor based on the re-
pressing transcription factor (TF) NanR for the detection of *N*-acetylneuraminic acid (Neu5Ac). In
absence of Neu5Ac, the expression of NanR allows the repressor to bind to the *nan* box contain-
ing promoter[304], which hinders transcription and therefore mKate2[252] expression. When Neu5Ac
is present, it binds to NanR, allowing transcription of *mKate2* and subsequent fluorescence detec-
tion. B) Schematic overview of working mechanism of a NanR-responsive without the expression
of NanR.

5.3.2   Engineered promoters for improved Neu5Ac detection

Besides using three native promoter variants to generate three Neu5Ac biosensors with
different properties, various engineered promoters were developed to modulate the per-
formance of these biosensors. In addition, by deconstructing such a native biosensor
circuit into defined parts, more modular biosensor optimization is enabled by varying
the core promoter and/or RBS strength and position, multiplicity and/or sequence of
the *nan* box[289]. Furthermore, as such, potentially unwanted regulatory interference is
removed, which could broaden the applicability of the biosensor to various physiological
conditions. For instance, the best studied promoter with the largest observed Neu5Ac

induction contains a cAMP receptor protein (CRP) binding site[304,307], which is a regulatory protein involved in catabolite repression[315]. Two approaches were used to create NanR-responsive promoters for improved Neu5Ac detection, a bottom-up and top-down design approach.

The top-down approach was applied to the *nanA* promoter by truncating the promoter based on the available information on the TF operator sites present in this promoter. The NanR binding site (the *nan* box) was shown to coincide with the transcription start site (spanning from -15 to +9), which indicates that NanR binds on the first part of the UTR of *nanA*[307]. To remove transferring undesired native regulation, only the core element of the *nanA* promoter pNanA, which contains the NanR binding site without the CRP binding site, was used in plasmid pEB_1 (see Figure 5.3A).

The three other NanR-responsive promoters were created using a bottom-up approach, which was based on a previously described effort where TetR repressed promoters were created by integrating operator sites of their cognate repressor[75]. This approach used an *in vitro* method to identify the operator sites of a set of TetR repressors, which were subsequently incorporated in the strong constitutive promoter BBa_J23119[308] to generate repressor-promoter sets, which were then screened for cross-reactions[75]. Similarly, three NanR-responsive synthetic promoters were created by aligning the NanR binding site with the -10 and -35 element of the constitutive promoter BBa_J23119. To this end, the *E. coli nan* box[307], which was obtained from the *nanATEK-yhcH* operon of *E. coli*, and the consensus *nan* box from several NanR orthologs of different origin[304] were used. The *E. coli* and the consensus *nan* boxes were aligned with both the -10 and the -35 box of BBa_J23119, creating the three engineered promoters with integrated *nan* boxes (see Figure 5.3B).

For the *nan* boxes aligned to the -35 box of the pJ23119 promoter, two promoters were created ($pJ23119_{H35,1}$ and $pJ23119_{H35,2}$), which differ in only one nucleotide in the -35 box. Moreover, $pJ23119_{H35,1}$ fully contains the *nan* box and $pJ23119_{H35,2}$ contains the original -35 box to ensure effective transcriptional initiation. The engineered promoters $pJ23119_{H10}$, $pJ23119_{H35,1}$ and, $pJ23119_{H35,2}$ form the basis of the biosensor plasmids pEB_2, pEB_3, and pEB_4, respectively (see Supplementary Figure C.2B).

5.3.3   Evaluation of the constructed biosensors

First, the strength of the NanR-responsive promoters was evaluated without NanR expression using the corresponding pOperator plasmids. To this end, an *E. coli* MG1655

A

native intergenic region between *nanR* and *nanA*

CAP binding site   pNanA promoter elements   nan box (NanR binding site)   *nanA*

	-35 box		-10 box	
pNanA	AGCTTTCTGT	ATGGGG	TGTTGCTTAATTGATC**TGG**	**TATA**AC	A**GGTATA**AA**GGTATA**TCGTTTA
E. coli nan box			TCT**GG**	**TATA**AC	A**GGTATA**AA**GGTATA**TCGTTT

B

original BBa_J23119 constitutive promoter   →   hybrid BBa_J23119 with integrated nan box at the -35 box

		-35 box		-10 box	
BBa_J23119	gattcgttaccaa	ttgaca	gctagctcagtcctagg	tataat	gctagctactaga
E. coli nan box	TCT**GGT**	**ATA**ACA	**GGTATA**AA**GGTATA**TCG	TTT	
consensus nan box	nn**TGGT**	**ATA**ACA	**GGTATA**nA**G**n**TAnn**nnn	Tnn	
pJ23119$_{H35,1}$	gattcgtTCT**GGT**	**ATA**ACA	**GGTATA**AA**GGTAct**agg	tataat	gctagctactaga
pJ23119$_{H35,2}$	gattcgtTCT**GGT**	**AT**gACA	**GGTATA**AA**GGTAct**agg	tataat	gctagctactaga

C

original BBa_J23119 constitutive promoter   →   hybrid BBa_J23119 with integrated nan box at the -10 box

		-35 box		-10 box	
BBa_J23119	gattcgttaccaa	ttgaca	gctagctcagtcctagg	tataat	gctagctactaga
E. coli nan box			CT**GGTATA**ACAGG	**TATA**AA	**GGTATA**TCGTTT
consensus nan box		T	Tnann**TGGTATA**ACAGG	**TATA**nA	**G**n**TAnn**nnnTnn
pJ23119$_{H10}$	gattcgttaccaa	ttgaca	gctaCT**GGTATA**ACAGG	**TATA**AA	**GGTAg**ctactaga

Figure 5.3: Schematic overview of the creation of the engineered promoters. A) Detailed overview of the development of pNanA, which is a truncated version of the native intergenic region between *nanR* and *nanA*. pNanA (basis of the biosensor of plasmid pEB$_1$) comprises the core elements (the -10 and the -35 box) of the *nanA* promoter, including the *nan* box required for NanR binding but without the CRP binding site. B) The development of the three engineered promoters by integrating the *nan* box for NanR binding at the position of the -10 and the -35 elements.

without endogenous NanR expression (*E. coli* MG1655 Δ*nanR*) was created. The fluorescence emitted by Δ*nanR* strains containing the various promoters with a *nan* box are depicted in Figure 5.3. From all tested promoters containing a NanR operator the native *nanA* promoter (pO$_1$ and pO$_2$) has the strongest mKate2 expression. The trunca-

tion of the *nanA* promoter and substitution of the native *nanA* RBS resulted in a lowered protein expression (pO_5). The lower expression of the truncated *nanA* promoter (pO_5) compared to the native *nanA* promoter (pO_1) could be related to the removal of various regulation mechanisms, for instance a CRP binding site, which activates the *nanATEK* operon in low glucose conditions[304,315]. Overall, all tested promoters with a *nan* box, except for *nanC* ($p0_3$), result in high fluorescent protein expression in the absence of the TF NanR.
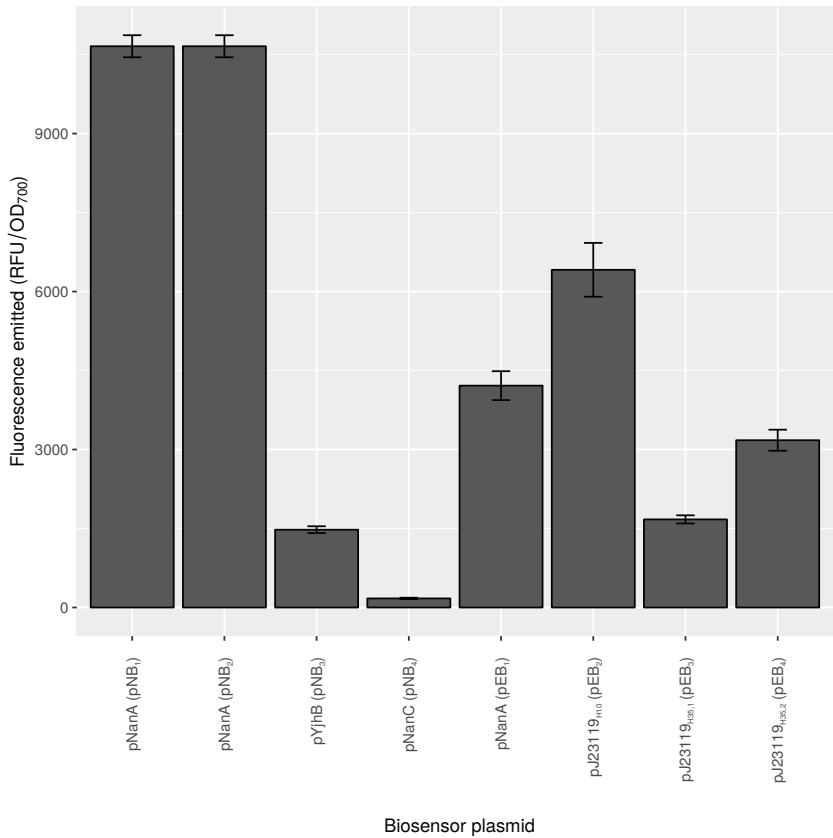
Figure 5.4: Barplot of fluorescence emitted by the eight created promoters, which all contain NanR binding sites. These plasmids contain the previously described promoters upstream of mKate2 without NanR expression (see Supplementary Figure C.1).

Next, the repression of the native and engineered promoters by the TF NanR was evaluated. This was done by comparing the strains containing pOperator to the strains containing the pBiosensor and pBlank plasmids. All constructed promoters were successfully repressed by the level of NanR expression obtained from the BBa_J23114 promoter and the BBa_B0031 RBS on the pBiosensor plasmids. This observation indicates that the repressor NanR is able to bind to all constructed pBiosensor plasmids, hereby blocking transcription of the gene downstream of the native and engineered promoters.

In functional biosensors, this blockage should be relieved when the ligand, in this case Neu5Ac, is present. To evaluate the response of the constructed biosensors on Neu5Ac, the fluorescence emitted by a Neu5Ac producing strain was compared to a wild-type strain, which does not produce Neu5Ac. To this end, the operon responsible for the degradation of Neu5Ac was deleted (*nanATEK*). This operon comprises the *nanA* gene (Neu5Ac aldolase), the *nanT* gene (sialic acid transporter), the *nanE* gene (GlcNAc6P epimerase) and, the *nanK* gene (ManNAc kinase)[303,316]. Subsequently, since the constructed strain does not accumulate Neu5Ac naturally under normal physiological conditions[303,316], a heterologous pathway was introduced to enable Neu5Ac production. This pathway comprises the *neuB1* and *neuC* genes originating from *Campylobacter jejuni*, which encode a Neu5Ac synthase and a GlcNAc6P 2-epimerase, respectively. The engineered pathway of Neu5Ac and the operon comprising the heterologous genes introduced are depicted in Supplementary Figure C.7A and C.7B, respectively. This Neu5Ac production strain is capable to produce Neu5Ac with titers of $1.39 \pm 0.06\,\text{g/l}$ in a 24 well DWP, which is comparable to titers obtained by other production strains[316]. Details on Neu5Ac production on DWP scale are depicted in Supplementary Figure C.8. The fluorescence emitted by *E. coli* $\Delta nanRATEK$ strains containing the eight constructed biosensor (both pNB and pEB) plasmids in the presence and absence of intracellular Neu5Ac (with the plasmids pPathway$_{\text{medium}}$ and pBlank, respectively) is shown in Figure 5.5.
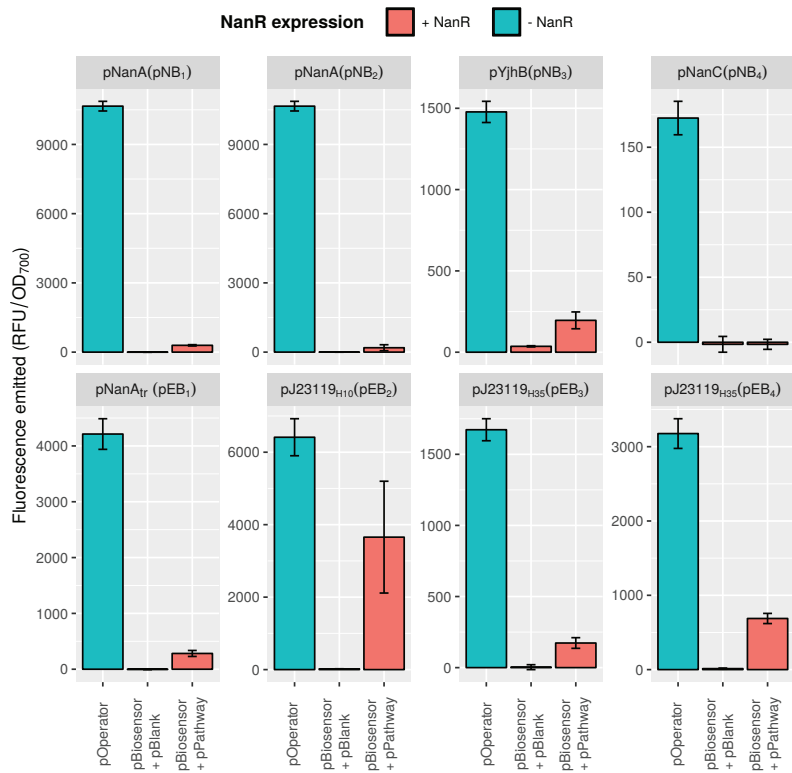
Figure 5.5: Barplot of fluorescence emitted by the eight created promoters, which all contain NanR binding sites. The effectiveness of the pBiosensor plasmids was evaluated by comparing the fluorescence emitted by a Neu5Ac producing (pPathway$_{medium}$ containing strain) and a wild-type *E. coli* strain (containing pBlank), which does not produce Neu5Ac. To evaluate the maximal fluorescence emitted by the constructed promoters, the fluorescence emitted by the pOperator plasmids was evaluated. Error bars represent standard deviations (n=3).

All biosensor plasmids except pNB$_2$ and pNB$_4$, are derepressed when Neu5Ac is present. One of the non-functional biosensors (p=0.99) contains the native *nanC* promoter. This promoter, compared to all other biosensor constructs, also emits lower fluorescence levels without NanR expression. Also the biosensor with non-native expression of NanR and the native *nanA* promoter is not derepressed significantly (p>0.05), which is remarkable since pNB$_1$ contains the same operator sequence. This could be explained by a higher expression of the TF, which makes derepression require a higher Neu5Ac concentration.

All other strains containing a functional pBiosensor ($p < 0.05$) and the pPathway$_{medium}$ plasmid emit a fraction of the fluorescence emitted when there is no NanR expression (strains with the pOperator plasmids, see Figure 5.4). Compared to the native biosensors, the engineered promoters exhibit higher protein expression in the presence of intracellular Neu5Ac. However, the strongest engineered promoter shows a large variance on the protein expression level (see pEB$_2$). Overall, the engineered biosensors show the feasibility of complex engineering of TF-responsive promoters. More specifically, engineering NanR-responsive promoters, further broadening the engineering capabilities for biosensor development[289]. Contrary to the native biosensors, the engineered biosensors are composed of defined parts, enabling modular optimization of the response curve by adjusting the position, multiplicity, and sequence of the binding site in the core promoter[289].

One biosensor (pEB$_2$) exhibits a higher level of activation than all other biosensors, indicating the high sensitivity of this promoter. All other biosensors show similar absolute levels of fluorescence activation in the presence of Neu5Ac. However, the applicability of all created Neu5Ac responsive biosensors depends on the response curve, of which only one point is available here. One important factor is sensitivity with one particular Neu5Ac biosensor showing high sensitivity: pEB$_2$. Another important property of a biosensor response curve is the saturation level, which is impossible to determine here. For instance, it is impossible to say whether these created biosensors already reached their saturation level in the Neu5Ac producing strain.

5.3.4   Engineering biosensor response

Based on the Neu5Ac biosensor with the highest sensitivity (pEB$_2$), three variants were constructed with a different RBS to control the translation initiation of NanR. The biosensors variants pEB$_{2,1}$, pEB$_{2,2}$, and pEB$_{2,3}$, which contain BBa_B0032, BBa_B0030, and BBa_B0034 as RBS, respectively. These biosensor variants were evaluated using the same experimental workflow as all other constructed biosensors, using a Neu5Ac producing (pPathway$_{medium}$) and non-producing strain (pBlank). The fluorescence emitted by the pEB$_2$ biosensors and its three variants (pEB$_{2,1}$, pEB$_{2,2}$, and pEB$_{2,3}$) in the presence and absence of intracellular Neu5Ac is depicted in Figure 5.6.

The RBS used to control NanR has a clear influence on the response of the biosensors in the presence of Neu5Ac, which increases as the strength of the RBS increases (see Supplementary Figure C.9). However, as the signal increases so does the variance on the biosensor response, indicative of a possible decrease in TF stability as a result of
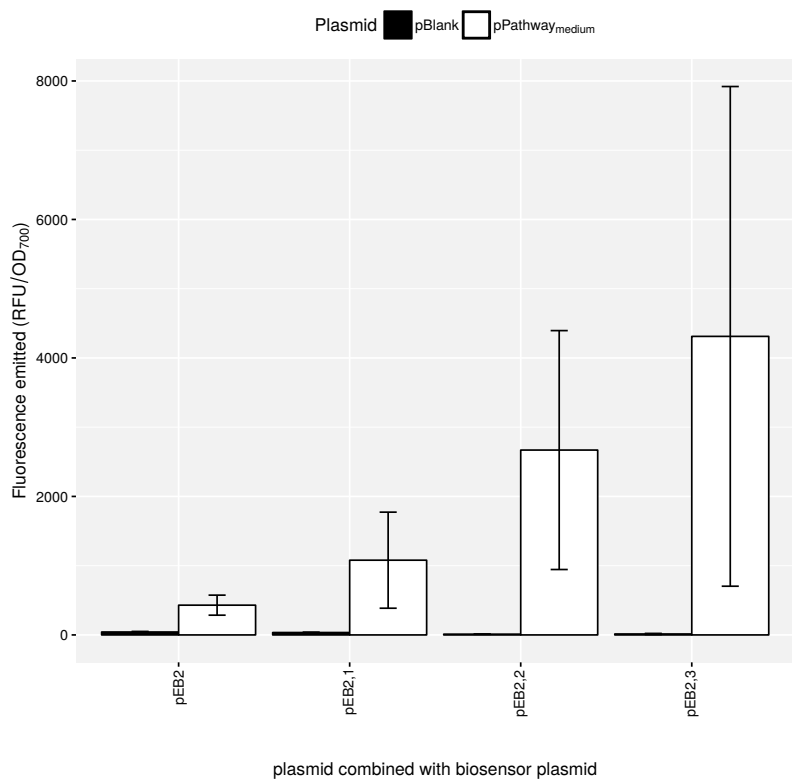
Figure 5.6: Barplot of fluorescence emitted by the three pEB_2 variants $pEB_{2,1}$, $pEB_{2,2}$, and $pEB_{2,3}$, which all contain NanR binding sites. The four biosensors are ordered based on the previously described strength of the RBSs used for NanR expression[308]. The effectiveness of the pBiosensor plasmids was evaluated by comparing the fluorescence emitted by a Neu5Ac producing ($pPathway_{medium}$ containing strain) and a wild-type *E. coli* strain (containing pBlank), which does not produce Neu5Ac. Error bars represent standard deviations (n=3).

metabolic burden. Another observation is the increased level of repression as the RBS strength increases, which is shown in Supplementary Figure C.9. This is an indication of an increased amount of available NanR molecules in the dimeric form, the form that allows binding to the NanR binding site and subsequently allows repression[304]. In contrast to the increased repression when the RBS strength is increased, the related rise in fluorescence in the presence of Neu5Ac is hard to explain. It is possible that the imposed burden of the NanR expression and the heterologous pathway has an effect but the logic

behind the response curve modulation remains unclear, quantifying metabolic burden would be required to assess this influence[245].

5.3.5   Neu5Ac production evaluation using biosensors

A typical application of biosensors in metabolic engineering is linking intracellular concentration of a specific metabolite to a measurable signal, which is fluorescence in this case. To assess the ability of the created Neu5Ac biosensors to link metabolite concentration to fluorescence, two mutants of the production strain were created, containing variants of the Neu5Ac production pathway plasmid. The pathway plasmids pPathway$_{high}$ and pPathway$_{low}$ were created by changing the p14 promoter with the pFAB46* (mutated form of the pFAB46 promoter[23]) and the pJ23105[308], respectively. First, the Neu5Ac production capacities of these pPathway plasmids were evaluated in $\Delta nanATEKR$ strains using a shake flask experiment (measured after 48 h). The amounts of Neu5Ac and acetic acid, the only detectable metabolites, produced by strains containing the three constructed pPathway plasmids are depicted in Supplementary Figure C.11.

To determine the relation between the genotype and the observed phenotype, mainly the Neu5Ac production, the strength of the promoters used is compared to the Neu5Ac production. The relative strength of these promoters was previously determined in-house, which was 1.91, 8.77 and 0.34 for the p14, the pFAB46*, and the pJ23105 promoters, respectively. The amount of extracellular Neu5Ac produced is related to the relative strength of the promoter used (see Supplementary Figure C.10). Also, the amount of acetic acid produced decreases as Neu5Ac production increases (see Supplementary Figure C.11), but there is still acetic acid produced, which shows the remaining potential to increase Neu5Ac production.

The applicability of the created biosensors to detect high Neu5Ac producers from a combinatorial library requires a proportional relationship between the amount of Neu5Ac production and the fluorescence produced by the biosensor. The plasmids pPathway$_{low}$, pPathway$_{medium}$, and pPathway$_{high}$, enabling 0.06 ± 0.03 g/L, 0.42 ± 0.07 g/L, and 1.4 ± 0.4 g/L extracellular production of Neu5Ac in a shake flask, respectively, were combined with the previously described pNB$_1$, pEB$_2$, and pEB$_4$ in a $E. coli$ $\Delta nanATEKR$ strain. The fluorescence emitted by these strains (ordered based on the Neu5Ac production) is depicted in Figure 5.7, which increases as the amount of Neu5Ac increases.
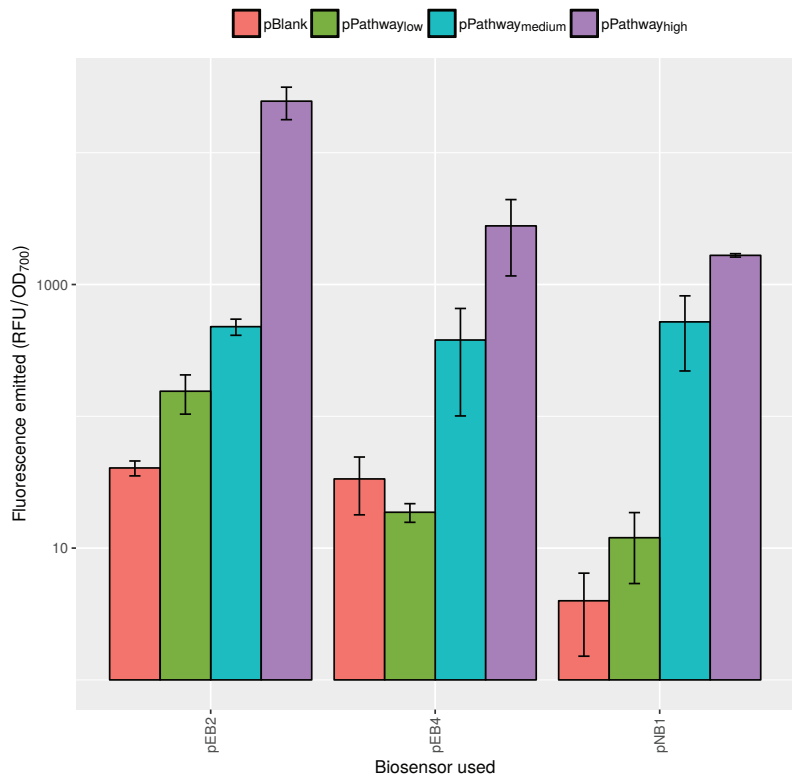
Figure 5.7: Bar plot (log scale used) of the fluorescence emitted by three created biosensor plasmids (pNB$_1$, pEB$_2$, and pEB$_4$) combined with three pathway variants (pPathway$_{low}$, pPathway$_{medium}$, and pPathway$_{high}$). The bar plots are ordered based on the amount of extracellular Neu5Ac produced. As a control a non-producing strain (containing pBlank) was used. Error bars represent standard deviation (n=3).

The observed increase as the amount of Neu5Ac produced increases confirms the proportional operating range of these biosensors, indicating the applicability of these biosensors as a high-throughput screening tool to select the best producers from a combinatorial mutant library. However, it remains unclear whether the biosensors used here are saturated by the levels of Neu5Ac present in the strains containing pPathway$_{high}$. Moreover, when comparing the extracellular concentration of Neu5Ac in shake flask using the pathway variants to the fluorescence emitted, biosensor plasmids pNB$_1$ and pEB$_4$ are closer to saturation compared to biosensor pEB$_2$ (see Figure 5.8). This is confirmed by the increase in the fluorescence emitted between the pPathway$_{medium}$ and pPathway$_{high}$, which is much higher for pEB$_2$ than the two other biosensors tested.
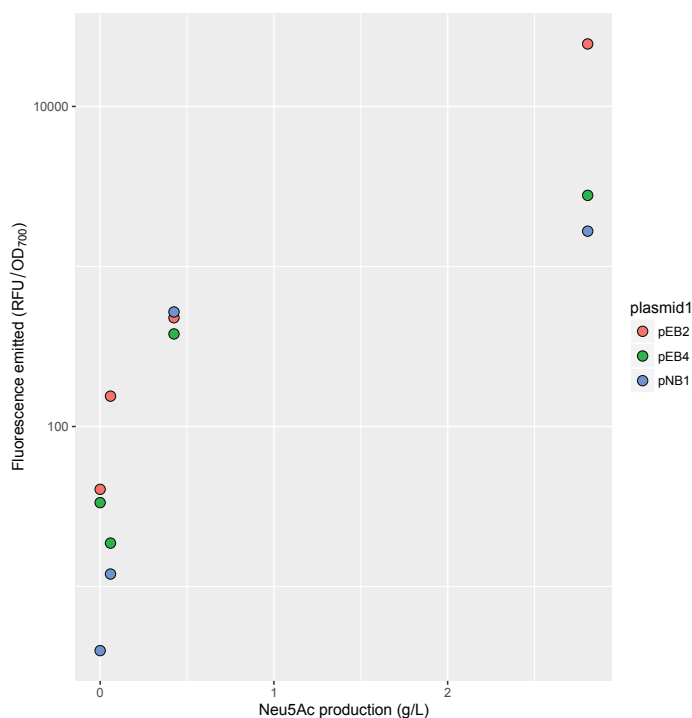
Figure 5.8: Plot of the fluorescence emitted against the amount of extracellular *N*-acetylneuraminic acid (Neu5Ac) produced by strains containing pathway variants (log scale used). Three created biosensor plasmids (pNB$_1$, pEB$_2$, and pEB$_4$) were combined with three pathway variants (pPathway$_{low}$, pPathway$_{medium}$, and pPathway$_{high}$).

5.4 Conclusions

Various biosensors were constructed using either native or engineered promoters, which contain binding sites of the TF NanR. This regulator naturally represses the *nanATEK-yhcH*, *nanCMS*, and *yjhBC* operons, which are all related to the degradation of various sialic acids. When Neu5Ac, the most abundant sialic acid, is present the repression by NanR is relieved. This mode of regulation was used to create several novel biosensors which link the expression of a fluorescent protein to the presence of Neu5Ac. All created (native and engineered) promoters were functional and all were successfully repressed by NanR expression. To examine the Neu5Ac responsiveness of the created biosensors a Neu5Ac producing strain was created and evaluated. Seven out of the eight created biosensors were able to discriminate between the Neu5Ac producing strain and the wild-type strain based on the emitted fluorescence, showing the desired biosensor functionality. More specifically, all four engineered biosensors are functional and exhibit a varying response to the Neu5Ac present in the production strain. Three out of the four engineered biosensors are created by inserting the NanR binding site in a known constitutive promoter, which, contrary to the native biosensors, results in modular biosensors composed of defined parts without undesired interference of other TFs. This approach deconstructs the native biosensors into well characterized modules, enabling modular biosensor optimization to change response curve characteristics by adjusting the promoter, the RBS strength, and the position, multiplicity, and sequence of the NanR binding site. To show the engineering capacities created by the biosensor deconstruction, the response of pEB_2 was modulated by changing the RBS controlling the NanR expression, which resulted in a decrease and increase in fluorescence in the absence and presence of Neu5Ac, respectively. Finally, the applicability of three created biosensors as a high-throughput screening was proven by combining these biosensors with various pathway plasmid variants, which allow producing up to 1.4 ± 0.4 g/L extracellular Neu5Ac. Moreover, all three biosensors showed a proportional relation to the amount of Neu5Ac produced extracellularly, a critical property of transcriptional biosensors for various applications in metabolic engineering.

Chapter 6

General discussion and outlook

RNA as a versatile alternative to traditional protein technology

The main goal of this PhD dissertation was to develop and optimize metabolic engineering tools, enabling forward engineering of microbial cell factories. To this end, engineering principles for both RNA-based or protein-based tools and devices were developed, optimized and evaluated in view of rewiring the metabolism of microbial cell factories in a custom way. This helps to engineer biology in a standardized way, allowing automated strain development, which is considered critical to shorten development time of production strains. To this end, previous efforts focused on creating composable, tunable, scalable and reliable biological parts, allowing the construction of complex biological systems with custom behaviour. Traditionally, these systems are built and reengineered using protein components, which function typically rely on protein-DNA or protein-protein interactions that are hard to predict. This limited programmability is a major disadvantage of protein-based parts, which is a major driving force for the recent development of RNA parts for synthetic biology. Despite the great interest in the last few years, RNA synthetic biology is still in its infancy compared to their protein counterparts. However, RNA has an advantage in programmability due to its function being related to their structure, which is easily predictable *de novo* compared to proteins. In contrast to RNA regulators, synthetic biology has traditionally used protein parts due to the large availability of protein parts and knowledge available on protein regulators. This is a result of the recent discovery of abundant RNA regulation, which is much harder to unravel than protein parts. For instance, protein regulators are encoded in an open reading frame (ORF), which are hard to ignore, whereas RNA is encoded in non-coding regions, which were largely overlooked in the past. Also, transcriptional regulators are well known compared to RNA regulation, making them perfect targets to link gene expression to the availability of a specific molecule. For instance, the programmable nature of RNA especially allows predictable control of transcription termination and particularly translation initiation rate, an indispensable tool for production pathway balancing in metabolic engineering.

In this context, pioneering efforts rely on thermodynamic models to design custom RBSs with the desired strength. To increase the accuracy of these tools, a machine learning approach was employed by combining fluorescence activated cell sorting (FACS) with next-generation sequencing, which drastically increases the throughput and the subsequent amount of data. Through machine learning, this resulted in forward engineering tools for RBS with increased accuracy and precision[270]. Recently, the democratization and throughput of enabling technologies such as flow cytometry, DNA synthesis, and,

next-generation sequencing allow the generation of large amounts of data, requiring big data analysis through machine learning, which increases the predictability of forward engineering tools. This is a recent trend observed throughout the whole field of metabolic engineering, which increasingly embraces the novel synthesis and sequencing technologies to decrease development times of microbial cell factories. Another factor in strain development is the continuing expansion of the metabolic engineering toolbox, which increasingly employs various types of recently discovered RNA devices.

In **Chapter 2** an extensive overview is given of the recently emerged RNA technology to modulate gene expression, build genetic circuitry, sense molecules, report physiological processes, and build nanostructures. In comparison to their protein counterparts, RNA tools are good alternatives due to their superior progammability and scalability. However, overall, the field of RNA synthetic biology is still in an infant state and still making a transition from development through trial and error to development increasingly based on forward engineering, resulting in devices with improved properties. Moreover, the previously mentioned novel sequencing and synthesis technologies are increasingly being exploited to increase designability. One specific development with huge impact on biotechnology in general the last few years is the birth of RNA-guided gene expression modulation on both the transcriptional and the translational level, which are facilitated by dCas9 and Hfq proteins. These techniques have clear advantages over traditional DNA-based gene knockouts, as they allow inducible and reversible knockdowns, which expands the metabolic engineering toolbox, allowing strain engineering strategies previously impossible. However, the dependence on protein parts for proper gene expression modulation might cause an undesirable stress effect (metabolic burden)[245]. Moreover, the use of RNA-guided protein parts also requires specific complex design constraints to allow the necessary RNA-protein interactions and proper functionality, which limits the inherent programmability of RNA[65,242].

Towards forward engineering of RNA genetic circuitry

To bypass the need for coexpressed protein parts and the burdensome and relatively slow translation step, various novel riboregulator parts were developed that allow the construction of genetic circuitry using solely RNA. Despite the fact that several types of regulators are available, activating or repressing on the translational or the transcriptional level (see **Chapter 2** for details), the dynamic range and availability of these riboregulators remains limited. Moreover, various riboregulators are reengineered using a natural existing chassis, limiting the applicability. To overcome these limitations,

repressing riboregulators were developed *de novo* in **Chapter 3** by efficiently exploring possibly important properties through a design of experiments (DOE). This approach yielded several functional riboregulators with dynamic ranges comparable or better than previous efforts, making these RNA regulators interesting parts to construct RNA only circuitry. These possibly important riboregulator features were defined based on previously described efforts, which used these different properties as design principles. This broad range of previously used design principles for riboregulators hinders the application of this technology for non-experts. To elucidate the importance of these design criteria, the *in vivo* performance of all riboregulators was used to build a partial least squares (PLS) model, linking various riboregulator features to their repression efficiency. This enables to identify the important factors from the less important factors but also shows the applicability of big data analysis to further improve riboregulator designs. Moreover, the fact that basic regression techniques do not allow linking features to riboregulator performs shows the complex nature of riboregulation, which could not be sufficiently explained using regular ordinary least squares (OLS) regression.

The developed model is suitable for *de novo* forward engineering of repressing riboregulators, which has never been described before (previous efforts either described activating riboregulators or use a naturally occurring chassis for repressing riboregulator development). However, this model was verified for two (related) UTRs and the general applicability of the PLS model for completely different UTRs should be assessed. An interesting future perspective of this riboregulator technology and its sequence-function model is the creation of orthogonal families of riboregulators. This could be done by creating orthogonal riboregulator-UTR pairs, containing a riboregulator that specifically represses an UTR without interacting with any other UTR. However, **Chapter 3** shows the potential of advanced data analysis techniques in synthetic biology but the forward engineering capacities might not reach far beyond the data used in the model. Obtaining generally applicable models requires high-throughput data generation, which is enabled by recently developed techniques. For instance, the recently developed SHAPE-Seq technique combines *in vivo* RNA structural probing and next-generation sequencing technology, allowing high-throughput interrogation to further elucidate the sequence-structure-function relationship of riboregulators[271,272]. Combining this technology with FACS to analyze large strain libraries could yield sufficient data to build models with a broader applicability and better accuracy. In addition, the recently emerged sequencing and gene synthesis technologies with a huge throughput lift the possibilities of metabolic engineering to unprecedented heights. However, these technologies remain largely un-

used, indicating the need for further investments. Another aspect to this paradigm shift in metabolic engineering is coping with these large amounts of data, which could reach much larger proportions than ever seen before.

Improving riboswitch development through bioinformatics

Another type of RNA regulator with great potential are riboswitches, which link gene expression to the availability of specific ligands (see **Chapter 2** for more information). One type of riboswitches are translational riboswitches, which are typically located in the 5'UTR and include an aptamer region to allow ligand interaction. Curiously, before discovering naturally occurring translational riboswitches, various synthetic riboswitches were created from *in vitro* selected aptamers, which were selected from a combinatorial library through systematic evolution of ligands by exponential enrichment (SELEX). This transition from *in vitro* selected aptamer to *in vivo* functional riboswitches has been challenging ever since, mainly relying on *in vivo* selection such as FACS. To facilitate the development of translational riboswitches, **Chapter 4** describes an *in silico* workflow that was created and evaluated for the computer-aided design of translational riboswitches. This algorithm is able to design translational riboswitches with a high probability of activating translation initiation in the presence of the ligand of interest, which was theophylline as a proof of concept. Yet, despite the newly developed translational riboswitches, linking riboswitch activity to structural or thermodynamic features remains challenging. Moreover, the riboswitch with the highest activation ratio (AR) was created by reshuffling regions from two different riboswitches, indicating the complex nature of translational riboswitch that is still largely unknown. The observed ARs of the designed riboswitches are much lower than those of the riboswitches described in literature, indicating the limitations of the automated design algorithm. However, the obtained absolute expression levels in the ON state are much higher for the designed riboswitches, showing that the observed ARs is largely determined by the expression level in the OFF state. Moreover, the riboswitches described in literature are the result of a laborious optimization that requires numerous resources compared to the automated design algorithm that can be obtained in a few days. Yet, the *in silico* approach is still in an infant state and requires several improvements to increase predictability of the algorithm. For instance, a recently published study on the applicability of the computational design of transcriptional riboswitches showed that only aptamers adopting a minimum free energy (MFE) structure are suitable for riboswitch design approaches using RNA bioinformatics, showing the complexity of automated riboswitch design and the need

for improvements[317]. Expanding the model to include previously unknown physiological processes can also improve the reliability of the design algorithms. For instance, recently emerged cotranscriptional SHAPE-seq technology allows discovering RNA folding pathways, which were shown to be important for proper riboswitch performance[284]. Overall, the *in silico* selection approach developed in **Chapter 4** is much less laborious than the typical manual screening and optimization of riboswitches.

Another factor determining the precision of the riboswitch design algorithms are the simplification that are made, as only the secondary structures are considered during the design process due to technical feasibility. A democratization of advanced molecular dynamics techniques could be useful here, as riboswitches form flexible tertiary structures that interact in 3D with their specific ligand. As such, cutting edge RNA 3D structure predictions and docking algorithms could be used, similarly to the use of protein 3D structures and docking algorithms in protein engineering. For instance, one of the best algorithms (included in the ROSETTA software suite), fragment assembly of RNA with full atom refinement (FARFAR), allows accurate *de novo* prediction of the 3D RNA structure[318,319]. However, the length of the accurately predicted RNA sequences is short (<20 nucleotides) and the computational resources required are significant, making high-throughput infeasible. Besides modelling the 3D structure of riboswitches, the interaction with the ligand needs to be simulated, which is possible using docking algorithms. Specifically, the latest version of the DOCK, DOCK 6.0, include techniques to simulate RNA-small molecule complexes[320]. Generally, most molecular dynamics techniques are historically focused on more rigid proteins, although several algorithms are gradually including force fields specific for RNA. Despite the limited development of RNA molecular dynamics, the potential of molecular dynamics for *in silico* aptamer selection was previously shown as a proof of concept[321]. Overall, advances in RNA molecular dynamics and the democratization of computational resources are undoubtedly a major opportunity for RNA synthetic biology. Still, similarly to the contribution of bioinformatics tools to protein engineering, a very advanced research field with years of evolution in molecular dynamics and related technology, the use of algorithms will facilitate molecular design but will always require wet lab work to test the obtained molecular devices. Over the years algorithms to design RNA devices can improve but will, similarly to protein engineering, rarely be able to perfectly predict the performance of biological devices. Yet, bioinformatics tools are indispensible in RNA synthetic biology but are still in an infant state compared to protein bioinformatics techniques.

Biosensors as an enabling technology in metabolic engineering

Similarly to translational riboswitches, transcriptional biosensors employ proteins to link small molecule concentration to gene expression. Therefore, biosensors are useful tools as sensor or genetic switch in metabolic engineering. To this end, various biosensors were developed using naturally occurring TFs to allow detection of a range of molecules. However, the number of available biosensors remains limited and the biosensor engineering principles are small, hindering response curve engineering possibilities. One interesting group of molecules are sialic acids, a family of nine-carbon carboxylated monosaccharides with Neu5Ac as most important member with numerous applications in pharmaceutical and food industry. To detect Neu5Ac, various biosensors were developed in **Chapter 5** based on native NanR responsive promoters and by deconstructing these native promoter in defined parts. These biosensors created using a bottom-up approach allows modular biosensor optimization, further expanding the engineering tools to obtain desired response curves. More specifically, the feasibility of TF binding site insertions in a constitutive promoter was shown, which paves the way for changing the position, multiplicity, and/or sequence and the subsequent changes in response curve. Combining these findings with *in vitro* binding site identification techniques, further expands the possibilities to create new biosensors with custom response curves[75]. Biosensor engineering remains limited to what is available in nature. Luckily, a large number of regulatory systems are widely available for a huge number of molecules. The main requirement here is that molecules are available in nature. However, developing regulators for new-to-nature molecules would require advanced protein engineering techniques. To this end, compared to the field of RNA synthetic biology, bioinformatic tools play a more prominent role in the field of biosensors due to the mature state of protein molecular dynamics. This was shown by a recent effort that used computational protein design to reengineer the inducer specificity of the allosteric TF LacI to respond to new inducer molecules[277]. Moreover, the biosensor engineering possibilities can be further expanded by databases such as the recently created AlloRep, which collect information on the data on (mutated) regulators[322].

Regarding Neu5Ac, the applications of this sugar moiety are hindered by its limited availability due to insufficient production technologies, an effective microbial cell factory removes this hurdle. The newly created biosensor for Neu5Ac could be used as a high-throughput screening method or as a regulator in a dynamically regulated biosynthetic pathway. Both these metabolic engineering strategies could improve the productivity of the created microbial cell factories, which produced up to $2.8 \pm 0.8\,\text{g/L}$ extracellular

Neu5Ac without any further optimization. It was shown that the biosensors for Neu5Ac could serve as an initial screen for combinatorial libraries of the production pathway to obtain better Neu5Ac producers. Another important facet besides strain engineering is the optimization of the microbial production process, which could also improve the performance of a specific strain. Moreover, an interesting evolution in the field of metabolic engineering is the combination of genetic elements and growth conditions in one DOE to optimize microbial production, indicating the importance of optimizing both strain properties and growth conditions simultaneously[323,324].

In general, various strategies are possible in strain engineering, which typically range from a rational to a combinatorial approach. Since DNA assembly is no longer a bottleneck, the ideal strategy depends on the knowledge and resources available. For instance, when a protein regulator is available for the molecule of interest a combinatorial approach is more feasible than when no regulator was available. On the other hand, a rational approach requires a high degree of knowledge, which is typically unavailable (even for model organisms). In case there is a protein regulator described for a specific molecule, protein parts have a clear advantage over RNA to form a biosensor. Yet, in case no such regulator is available, RNA has the advantage to create a biosensor, for instance, based on translational riboswitches, which have a relatively simple working mechanism. However, the development of *in vivo* functional riboswitches for a molecule of choice remains challenging, with bottlenecks that are present all over the development process. For instance, the number of available aptamers that function *in vivo* is limited, which shows that a specific SELEX procedure for fast generation of aptamers that could function *in vivo* are needed. At the *in vivo* end of development, rapidly turning *in vitro* functional aptamers into *in vivo* functional riboswitches remains laborious despite the development of recent bioinformatics technology. Overall, RNA technology for the construction of microbial cell factories has the intrinsic advantage over their protein counterparts that RNA does not require the burdensome translation process. This is particularly important as the field of metabolic engineering tackles increasingly complex production pathways, an irreversible evolution that is further supported by state of the art synthesis and sequencing technologies. In general, metabolic engineering is undergoing a shift in speed as these novel technologies are increasingly embraced, which, along with advanced tools like those developed in this doctoral research, significantly reduces development times of production strains.

Appendix A

Exploration of the feature space of *de novo* developed post-transcriptional riboregulators

A.1 Supplementary Methods

A.1.1 Quantification of thermodynamic properties

In general, complex formation energy ΔG^{ij}_{form} can be calculated as follows:

$$\Delta G^{ij}_{form} = \Delta G^{ij}_{dimer} - (\Delta G^{i}_{monomer} + \Delta G^{j}_{monomer}) \qquad \text{(A.1)}$$

Where ΔG^{ij}_{dimer}, $\Delta G^{i}_{monomer}$ and $\Delta G^{j}_{monomer}$ are the estimated Gibbs free energy of the final dimer, and both initial monomer states, respectively. The features describing thermodynamic properties of translation inhibiting RNA (tiRNA) are free energy of the tiRNA monomer (EA), free energy of the tiRNA-tiRNA dimer (EAA), free energy of the tiRNA-UTR dimer (EAB), formation energy of the tiRNA-tiRNA dimer (FAA) and formation energy of the tiRNA-UTR dimer (FAB) (see Table 3.1).

A.1.2 Quantification of activation energy

The activation energy is estimated by the hybridization energy of consecutively unbound nucleotides of the tiRNA complex, the seed region, with the UTR. To get a good representation of the Boltzmann ensemble a random sample of 100 suboptimal structures of the tiRNA molecule are drawn with probabilities equal to their Boltzmann weights using the RNAsubopt algorithm[259]. This is done via stochastic backtracking in the partition function using RNAsubopt[259]. The intermolecular binding between the unbound part of the antisense and the UTR is estimated using the RNAup algorithm, which first calculates the energy requires to 'open' the binding site and subsequently calculates the minimal energy gained from intermolecular binding[260]. This is done for all possible unbound sequences with a length between two and six nucleotides of the tiRNA monomer. Two features are calculated for each of the 100 suboptimal structures: EIS and ETS. To get a general feature for a tiRNA molecule the average is calculated of the minimal intermolecular binding seed energy (EIS) and total seed energy (ETS) of the 100 structures.

A.1.3 Calculation of structural tiRNA features

Besides these previously mentioned thermodynamic features, some structural properties were previously described to improve translation initiation repression using solely RNA interactions.

UTR availability   One possible determinative feature relates to the availability of the target region in the mRNA, which was previously used as design principle for silencing small RNAs (sRNAs) as the mRNA is not in a constant state but rather in a state of constant structural fluctuation near MFE structures[248,325–327]. To account for a certain volatility of regions in the mRNA, Johnson & Srivastava[248] analyzed suboptimal mRNA structures to find regions which alter their structure without significantly changing the Gibbs free energy of the global structure. These regions are assumed to be more accessible and matched naturally occurring antisense target sites[248]. In the DOE the availability of the UTR nucleotides are accounted for by the $P_{\text{availability}}$ term, probability availability of UTR (PAU), which is calculated based on the partition function of the tiRNA-UTR dimer and the UTR monomer. For each nucleotide of the UTR the availability (relative number of unbound nucleotides) in the UTR monomer complex ($P_{\text{UTR,available}}$) and the coverage (relative number of nucleotides bound by the antisense molecule) in the tiRNA-UTR dimer complex ($P_{\text{UTR,tiRNA-coverage}}$) is determined. Based on these two terms the overall availability term (PAU; $P_{\text{availability}}$) is determined (weighted average of $P_{\text{UTR,available}}$ with $P_{\text{UTR,tiRNA-coverage}}$ as weights).

$$P_{\text{availability}} = \frac{\sum_0^{n_{\text{UTR}}} P_{\text{UTR,available}} \times P_{\text{UTR,tiRNA-coverage}}}{\sum_0^{n_{\text{UTR}}} P_{\text{UTR,tiRNA-coverage}}} \tag{A.2}$$

RBS coverage   To quantify the RBS coverage, two features were defined (RBS coverage of length 5 (RBS5) and RBS coverage of length 11 (RBS11)) as the base pairing probability in the region of the RBS. For the calculation of both features the weighted average of the nucleotides in the UTR bound by the 16S rRNA forms the center of the RBS region. Based on this center the RBS coverage in the regions $C_{\text{RBS}}-5$ to $C_{\text{RBS}}+5$ (RBS11) and from $C_{\text{RBS}}-2$ to $C_{\text{RBS}}+2$ (RBS5) is calculated based on the partition function estimated by RNAfold[256].

Paired termini (PT)   Another structural feature improving translational repression is the presence of paired termini in the silencing RNA molecule, which is strongly related to its thermodynamical stability, resulting in efficient gene silencing[84,247]. The feature PT is calculated by again drawing a random sample of 100 suboptimal structures of the tiRNA molecule with probabilities equal to their Boltzmann weights (using RNAsubopt[259]). From these 100 structures the average number of bound nucleotides between the first and the second half of the tiRNA sequence is calculated.

A.2 Supplementary Figures

CACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTTGCTGGA
TAACTTTACGGGCATGCATAAGGCTCGTATAATATATTCATAGTCTTTAGAAAGTTAA
AATTATTAAGGGAAACCTGCCTACTGAAAAAAAAAACCCCGCCCCTGACAGGGCGGGG
TTTTTTTT

Figure A.1: Overview of a DNA construct for translation inhibiting RNA (tiRNA) expression. As an example, the insert of pSilence₁ is displayed with the proD promoter[250], tiRNA₁ riboregulator, and BB_B1006 terminator[24].

CTAGAGCACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTT
GCTGGATAACTTTACGGGCATGCATAAGGCTCGTAATATATATTCAGGGAGACCACAA
CGGTTTCCCTCTACAAATAATTTTGTTTAACTTTTACTAGAGTCACACAGGAAAGTAC
TAGATGGTTAGCGAGCTGATCAAAGAAAACATGCACATGAAACTGTATATGGAAGGCA
CCGTGAATAACCACCACTTTAAATGTACCAGCGAAGGTGAAGGTAAACCGTATGAAGG
CACCCAGACCATGCGTATTAAAGCAGTTGAAGGTGGTCCGCTGCCGTTTGCATTTGAT
ATTCTGGCAACCAGCTTTATGTATGGCAGCAAAACCTTTATTAACCATACCCAGGGTA
TCCCGGATTTTTTTCAAACAGAGCTTTCCGGAAGGTTTTTACCTGGGAACGTGTTACCAC
CTATGAAGATGGTGGTGTTCTGACCGCAACCCAGGATACCAGTCTGCAGGATGGTTGT
CTGATTTATAATGTGAAAATTCGCGGTGTGAACTTTCCGAGCAATGGTCCGGTTATGC
AGAAAAAAACCCTGGGTTGGGAAGCAAGCACCGAAACCCTGTATCGGGCAGATGGTGG
TCTGGAAGGTCGTGCAGATATGGCACTGAAACTGGTTGGTGGTGGTCATCTGATTTGC
AATCTGAAAACCACCTATCGTAGCAAAAAACGGGCAAAAAATCTGAAAATGCCTGGCG
TGTATTATGTTGATCGTCGTCTGGAACGTATTAAAGAGGCAGATAAAGAAACCTATGT
GGAACAGCATGAAGTTGCAGTTGCACGTTATTGTGATCTGCCGAGCAAACTGGGTCAC
CGCTGATAACCATGGGCTAGCGGTTTGAAGGGTATTGGTCGGTCAGTTTCACCTGATT
TACGTAAAAACCCGCTTCGGCGGGTTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGAC
TGTC

Figure A.2: Overview of a DNA construct for untranslated region (UTR) expression. As an example, the insert of pTarget₁ is displayed with the proB promoter[250], UTR₁[250], *mKate2* reporter gene[252], and rnpB T1 terminator[24].
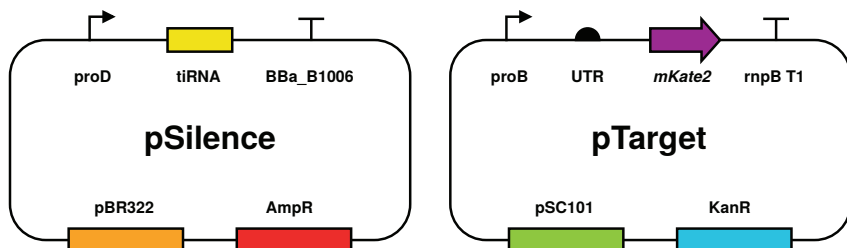
Figure A.3: Schematic overview of the two plasmid types used in this study. The plasmids pSilence and pTarget are used to respectively express translation inhibiting RNAs (tiRNAs) and the target untranslated regions (UTRs) upstream of the reporter gene *mKate2*[252]. pSilence comprises a medium-copy vector (pBR322 origin of replication and ampicillin resistance marker, originating from pSB6A1[308]) using proD[250] as promoter and BBa_B1006[24] as terminator. pTarget comprises a low-copy vector (pSC101 origin of replication and kanamycin resistance marker) with proB[250] as promoter, *mKate2*[252] as reporter gene, and rnpB T1[24] as terminator. Details of all tiRNAs, plasmids and important DNA sequences used in this study are listed in Supplementary Table A.1, A.2 and A.3, respectively.
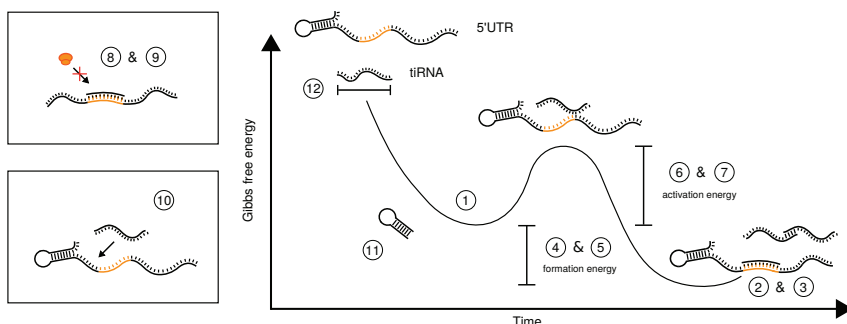
Figure A.4: Schematic overview of all potentially determinative features of translation inhibiting RNAs (tiRNAs): 1) free energy of the tiRNA monomer (EA), 2) free energy of the tiRNA-UTR dimer (EAB), 3) free energy of the tiRNA-tiRNA dimer (EAA), 4) formation energy of the tiRNA-UTR dimer (FAB), 5) formation energy of the tiRNA-tiRNA dimer (FAA), 6) intermolecular binding seed energy (EIS), 7) total seed energy (ETS), 8) RBS coverage of length 5 (RBS5), 9) RBS coverage of length 11 (RBS11), 10) probability availability of UTR (PAU), 11) paired termini (PT), and 12) tiRNA length (L).

Appendix A. Exploration of the feature space of *de novo* developed post-transcriptional riboregulators
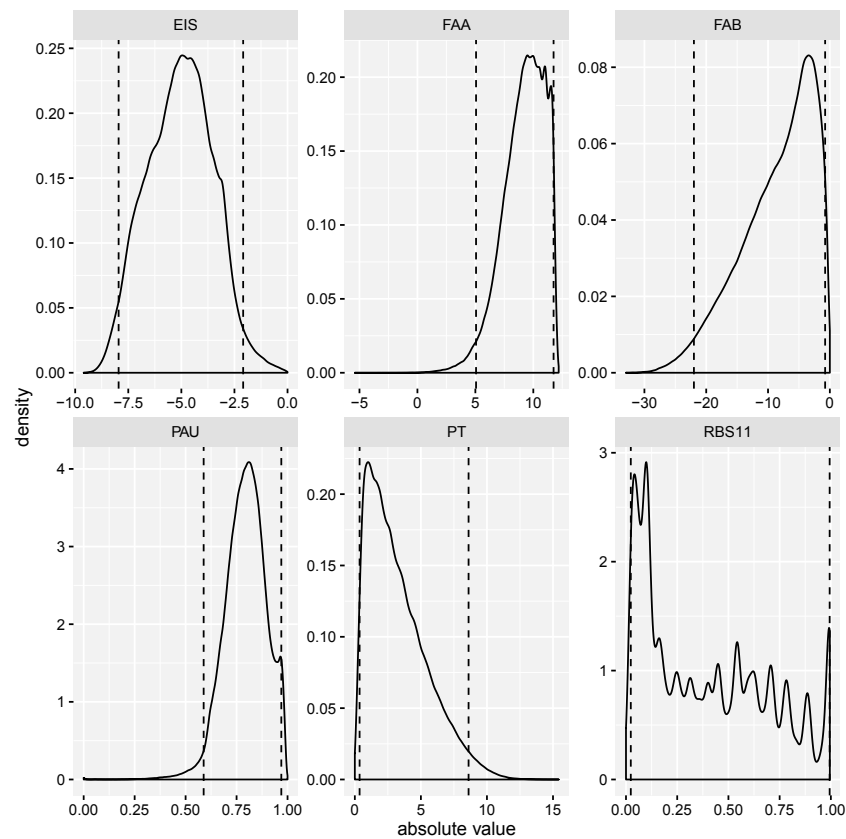
Figure A.5: Density of all features of tiRNA with the 0.1 (-1 level) and 0.9 p-quantiles (+1 level) indicated as vertical stripped lines.

	predicted stability (ΔG)	fluorescence (FP/OD)
UTR_1	-27.6 kcal/mol	27000 ± 1200
UTR_2	-17.3 kcal/mol	8100 ± 1200
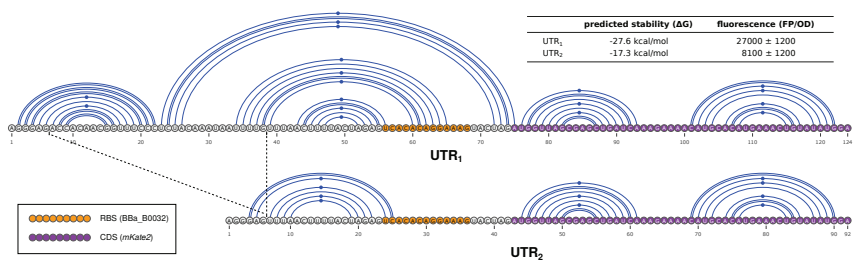
Figure A.6: Predicted minimum free energy (MFE) structures of the two mRNAs used in this study, comprising the untranslated region (UTR) and the coding DNA sequence (CDS) part. Both UTRs contain the ribosome binding site (RBS) BBa_B0032. The predicted Gibbs free energy ($\triangle$G) is depicted (top right) along with the experimentally determined fluorescence $(FP/OD_{700})_{corrected}$ for both UTR_1 and UTR_2.

Appendix A. Exploration of the feature space of *de novo* developed post-transcriptional riboregulators
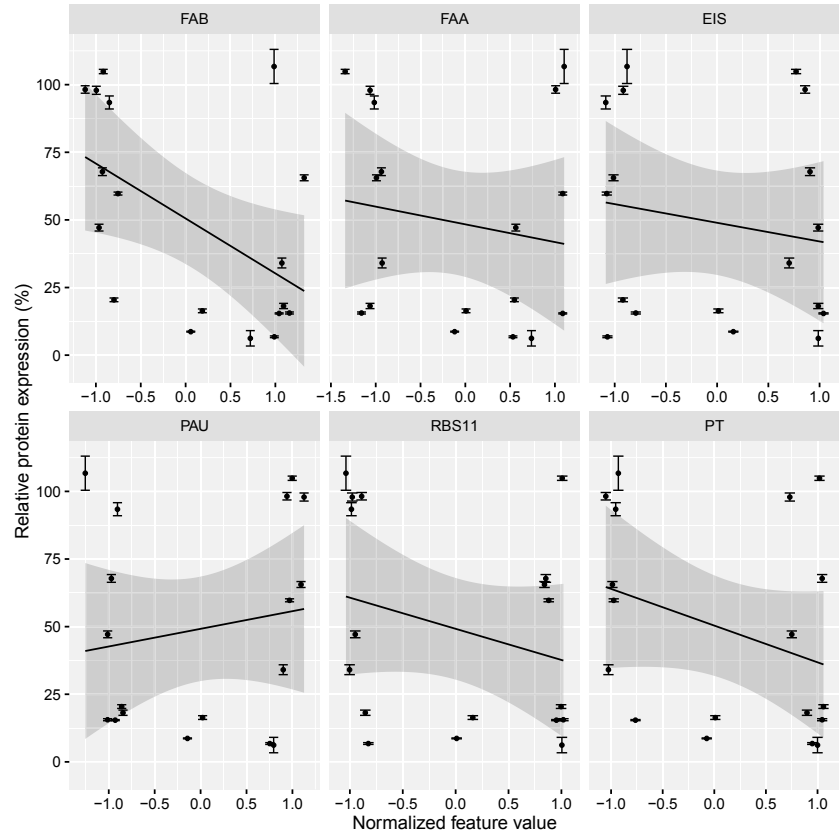
Figure A.7: Plot of the percentage relative expression of all data points in the experimental design (comprising solely UTR$_1$) against the normalized translation inhibiting RNA (tiRNA) features of the reduced feature set. The six factors used in the design of experiments (DOE) are the features in the reduced feature set (formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 11 (RBS11), and paired termini (PT)). The gray area depicts the 95 % confidence interval of the linear regression between the relative protein expression and each normalized feature value. Error bars represent standard deviation (n=3).
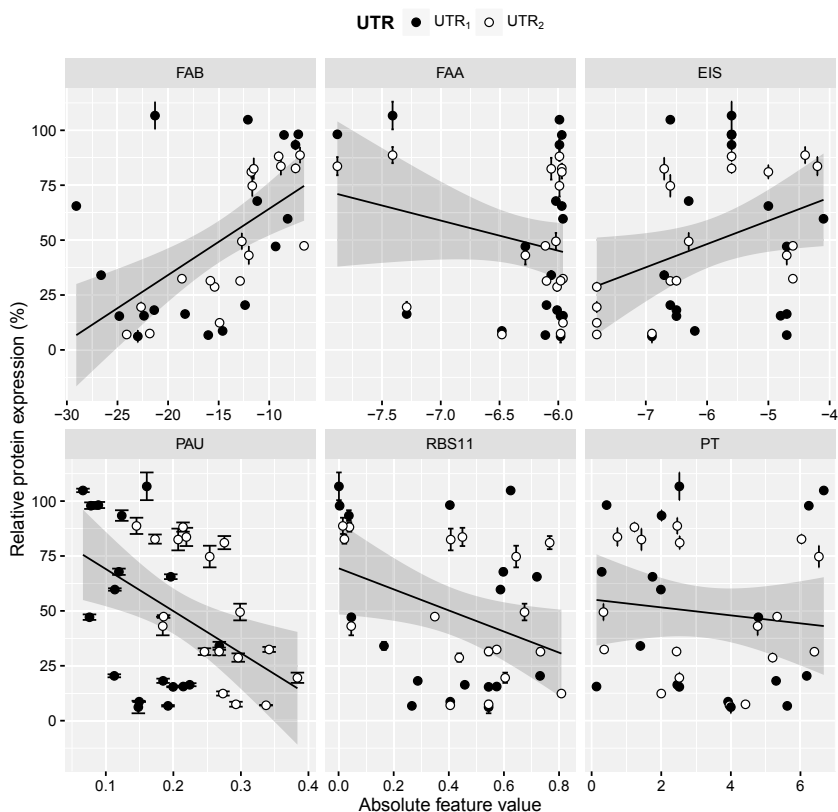
Figure A.8: Plot of the percentage relative expression of all data points against the absolute translation inhibiting RNA (tiRNA) features of the reduced feature set. The six factors used in the design of experiments (DOE) are the features in the reduced feature set (formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 11 (RBS11), and paired termini (PT)). The gray area depicts the 95 % confidence interval of the linear regression between the relative protein expression and each absolute feature value. Error bars represent standard deviation (n=3).
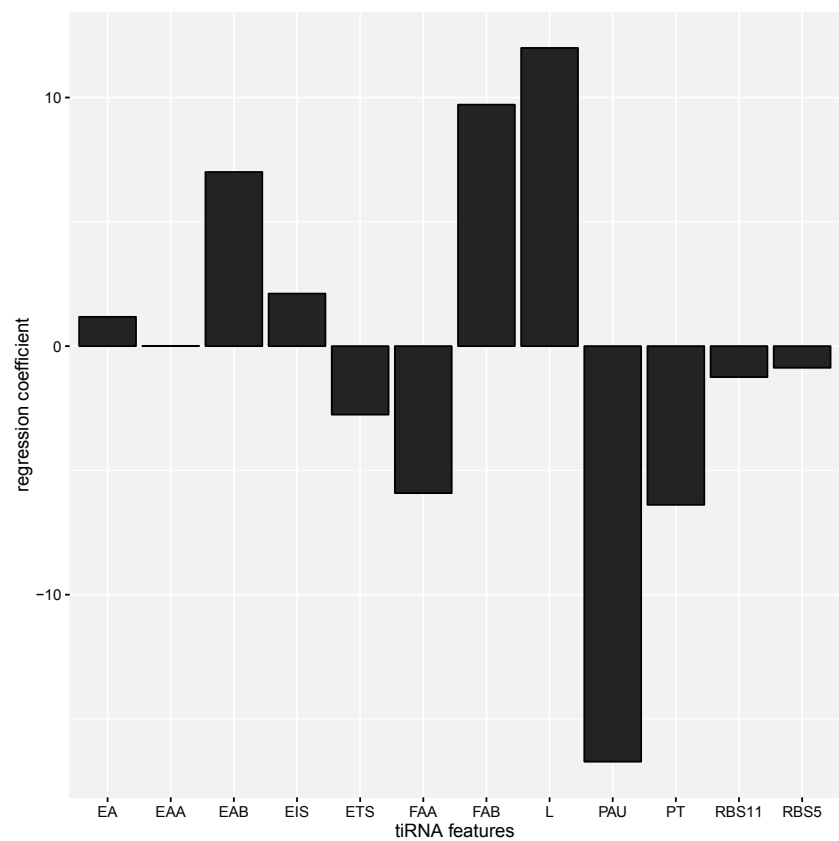
Figure A.9: The estimated PLS regression coefficients of the eleven identified translation inhibiting RNA (tiRNA) features. Detailed definitions of all features (free energy of the tiRNA monomer (EA), free energy of the tiRNA-tiRNA dimer (EAA), free energy of the tiRNA-UTR dimer (EAB), formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), total seed energy (ETS), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 5 (RBS5), RBS coverage of length 11 (RBS11), paired termini (PT), and tiRNA length (L)) are available in Table 3.1.
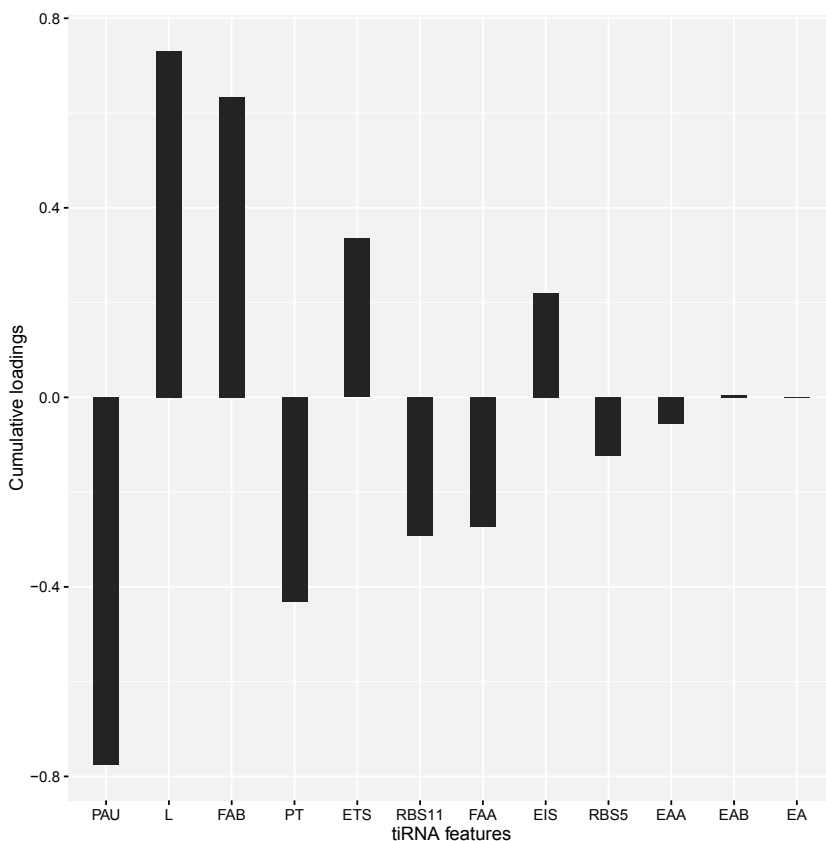
Figure A.10: Cumulative loadings of the four components used in the PLS model. Detailed definitions of all features (EA, EAA, EAB, FAA, FAB, ETS, EIS, PAU, RBS5, RBS11, PT, and tiRNA length (L)) are available in Table 3.1.
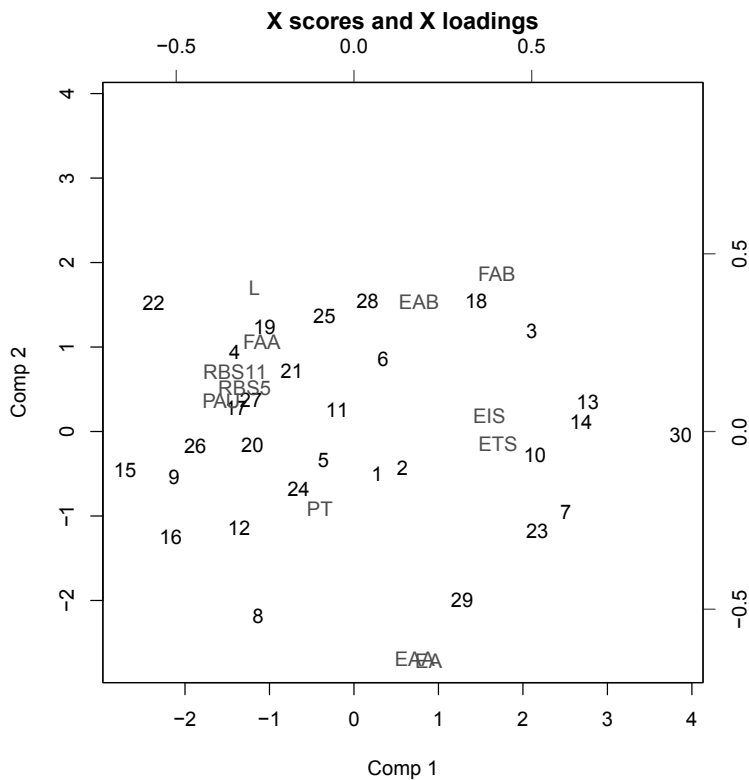
Figure A.11: Biplot of the first two components of the PLS regression model. Detailed definitions of all features (EA, EAA, EAB, FAA, FAB, ETS, EIS, PAU, RBS5, RBS11, PT, and tiRNA length (L)) are available in Table 3.1.

A.3 Supplementary Tables

Table A.1: Overview of all translation inhibiting RNA (tiRNA) molecules, comprising the experimental design to discover design rules for gene silencing using solely RNA.

Name	RNA sequence (5'→3')
tiRNA$_1$	UAGUCUUUAGAAAGUUAAAAUUAUUAAGGGAAACCUGCCU
tiRNA$_2$	GUACCUGUGUGGUAAAAGUUAAACAAUAUCCGUUGUGGUC
tiRNA$_3$	ACUUCCUGUAGUAAAAGUAUUGUGGUGUCU
tiRNA$_4$	AGUAAAGUUAAGAGGGAAACGUUGUGGUC
tiRNA$_5$	UUCUGUGUAGCAAAUUGUCCGUGUGGUCUC
tiRNA$_6$	UCUAGCCUGUUAAACAAAAUUAUUUGUGAGGGAAACUCCC
tiRNA$_7$	AUCUACUCUAGUAAAGUGGU
tiRNA$_8$	CCUGUGUUAAAAGUUAAACAAAAUUAUUGU
tiRNA$_9$	UUUACUCUAGUUAAAAUAGAGGGAAACCUG
tiRNA$_{10}$	UACUUUCGUGGUAAAAUAAUGUAGAGGGAAACCGUUGUGG
tiRNA$_{11}$	GUACUUUUGUGCGUCUCCCU
tiRNA$_{12}$	CAUCUAGUACUAAAAAAUUAUUUGUAGAGU
tiRNA$_{13}$	UCCUGUGUCUUUAUUGUGGGAAACCUGGUC
tiRNA$_{14}$	UAGUACCUAGUAAUUAAACAUGUAGAGGGAAACCGUUGUG
tiRNA$_{15}$	CUACUCUAGUACCGUUGGUC
tiRNA$_{16}$	CCUGUGUGACUCUUAUUUGUAGAUUGUGGU
tiRNA$_{17}$	UCUAGUUUUCCUAGUAAAAUUAAACAAAAUUAAACUCCCU
tiRNA$_{18}$	AGUACUGUGUCUCUAGUAUAAGAUCUCCCU

Table A.2: Overview of all plasmids used in this study.

Name	Content (5'→3')	Backbone
pSilence$_1$	proD - tiRNA$_1$ - BBa_B1006	pBR322
pSilence$_2$	proD - tiRNA$_2$ - BBa_B1006	pBR322
pSilence$_3$	proD - tiRNA$_3$ - BBa_B1006	pBR322
pSilence$_4$	proD - tiRNA$_4$ - BBa_B1006	pBR322
pSilence$_5$	proD - tiRNA$_5$ - BBa_B1006	pBR322
pSilence$_6$	proD - tiRNA$_6$ - BBa_B1006	pBR322
pSilence$_7$	proD - tiRNA$_7$ - BBa_B1006	pBR322
pSilence$_8$	proD - tiRNA$_8$ - BBa_B1006	pBR322
pSilence$_9$	proD - tiRNA$_9$ - BBa_B1006	pBR322
pSilence$_{10}$	proD - tiRNA$_{10}$ - BBa_B1006	pBR322
pSilence$_{11}$	proD - tiRNA$_{11}$ - BBa_B1006	pBR322
pSilence$_{12}$	proD - tiRNA$_{12}$ - BBa_B1006	pBR322
pSilence$_{13}$	proD - tiRNA$_{13}$ - BBa_B1006	pBR322
pSilence$_{14}$	proD - tiRNA$_{14}$ - BBa_B1006	pBR322
pSilence$_{15}$	proD - tiRNA$_{15}$ - BBa_B1006	pBR322
pSilence$_{16}$	proD - tiRNA$_{16}$ - BBa_B1006	pBR322
pSilence$_{17}$	proD - tiRNA$_{17}$ - BBa_B1006	pBR322
pSilence$_{18}$	proD - tiRNA$_{18}$ - BBa_B1006	pBR322
pTarget$_1$	proB - UTR$_1$ - *mKate2* - rnpB T1	pSC101
pTarget$_2$	proB - UTR$_2$ - *mKate2* - rnpB T1	pSC101
pBlank$_1$	/	pBR322
pBlank$_2$	*mKate2* - rnpB T1	pSC101

Table A.3: Overview of all important DNA sequences

Name	Sequence (5'→3')	Reference
mKate2	ATGGTTAGCGAGCTGATCAAAGAAAACATGCACATGAAAGCTGATCAAAGAAAACATG CACATGAAACTGTATATGGAAGGCACCGTGAATAACCACCACTTTAAATGTACCAGCG AAGGTGAAAGCTGATCAAAGAAAACATGCACATGAAACTGTATATGGAAGGCACCGTG AATAACCACCACTTTAAATGTACCAGCGAAGGTGAAGGTAAACCGTATGAAGGCACCC AGACCATGCGTATTAAAGCAGTTGAAGGTGGTCCGCTGCCGTTTGCATTTGATATTCT GGCAACCAGCTTTATGTATGGCAGCAAAACCTTTATTAACCATACCCAGGGTATCCCG GATTTTTTTCAAACAGAGCTTTCCGGAAGGTTTTACCTGGGAACGTGTTACCACCTATGA AGATGGTGGTGTTCTGACCGCAACCCAGGATACCAGTCTGCAGGATGGTTGTCTGATT TATAATGTGAAAATTCGCGGTGTGAACTTTCCGAGCAATGGTCCGGTTATGCAGAAAA AAACCCTGGGTTGGGAAGCAAGCACCGAAACCCTGTATCCGGCAGATGGTGGTCTGG AAGGTCGTGCAGATATGGCACTGAAACTGGTTGGTGGTGGTCATCTGATTTGCAATCT GAAAACCACCTATCGTAGCAAAAAACCGGGCAAAAAATCTGAAAATGCCTGGCGTGTAT TATGTTGATCGTCGTCTGGAACGTATTAAAGAGGCAGATAAAGAAACCTATGTGGAAC AGCATGAAGTTGCAGTTGCACGTTATTGTGATCTGCCGGAGCAAACTGGGTCACCGCTG ATAA	[252]
proB	CTAGAGCACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTT GCTGGATAACTTTACGGGCATGCATAAGGCTCGTAATATATATTC	[250]
proD	CTAGAGCACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTT GCTGGATAACTTTACGGGCATGCATAAGGCTCGTATAATATATTC	[250]
rnpB T1	TCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAG GGGCAGAAAGATGAATGACTGTC	[24]
BBa_B1006	AAAAAAAAACCCCGCCCCTGACAGGGCGGGGTTTTTTTT	[24]
UTR₁	AGGGAGACCACAACGGTTTCCCTCTACAAATAATTTTGTTTAACTTTTACTAGAGTCAC ACAGGAAAGTACTAG	[250]
UTR₂	AGGGAGATTGACTTTTACTAGAGTCACACAGGAAAGTACTAG	this work

Table A.4: Detailed description of the 2^{6-2} fractional factorial design to unravel translation inhibiting RNA (tiRNA) design principles. This experimental design for six 2-level factors comprises 16 regular runs (tiRNA$_{1-16}$) and two center points (tiRNA$_{17}$ and tiRNA$_{18}$). The six factors used in the design of experiments (DOE) are the features in the reduced feature set (formation energy of the tiRNA-tiRNA dimer (FAA), formation energy of the tiRNA-UTR dimer (FAB), intermolecular binding seed energy (EIS), probability availability of UTR (PAU), RBS coverage of length 11 (RBS11), and paired termini (PT)).

	FAB	FAA	EIS	PAU	RBS11	PT
tiRNA$_1$	-1	-1	-1	-1	-1	-1
tiRNA$_2$	1	-1	-1	-1	1	1
tiRNA$_3$	-1	1	-1	-1	1	1
tiRNA$_4$	1	1	-1	-1	-1	-1
tiRNA$_5$	-1	-1	1	-1	1	-1
tiRNA$_6$	1	-1	1	-1	-1	1
tiRNA$_7$	-1	1	1	-1	-1	1
tiRNA$_8$	1	1	1	-1	1	-1
tiRNA$_9$	-1	-1	-1	1	-1	1
tiRNA$_{10}$	1	-1	-1	1	1	-1
tiRNA$_{11}$	-1	1	-1	1	1	-1
tiRNA$_{12}$	1	1	-1	1	-1	1
tiRNA$_{13}$	-1	-1	1	1	1	1
tiRNA$_{14}$	1	-1	1	1	-1	-1
tiRNA$_{15}$	-1	1	1	1	-1	-1
tiRNA$_{16}$	1	1	1	1	1	1
tiRNA$_{17}$	0	0	0	0	0	0
tiRNA$_{18}$	0	0	0	0	0	0

Table A.5: Untranslated regions (UTRs) used in this study

Name	Sequence (5'→3')
UTR$_1$	AGGGAGACCACAACGGUUUCCCUCUACAAAUAAUUUUGUUUAACUUUUACUAGA GUCACACAGGAAAGUACUAG
UTR$_2$	AGGGAGAUUGACUUUUACTAGAGTCACACAGGAAAGUACUAG

Appendix B

Computer-aided development of ligand-responsive RNA devices

B.1 Supplementary Methods

B.1.1 Search algorithm

The search algorithm used to efficiently find a global minimum of the objective function is a Metropolis Monte Carlo simulated annealing algorithm (see Algorithm 1 for pseudocode).

Define $E(x)$; mutate(x); random(x) with x = riboswitch sequence;
Define $score_{max}$; $p_{accept,initial}$; iterations; $T_{initial}$; T_{final}; slope;
begin
    $T = T_{initial}$;
    $sequence_{accepted}$ = random(sequence constraints);
    $score_{accepted}$ = $E(sequence_{accepted})$;
    for *iterations - 1* **do**
        $sequence_{mutated}$ = mutate($sequence_{accepted}$);
        $score_{mutated}$ = $E(sequence_{mutated})$;
        if *$score_{mutated} < score_{accepted}$* **then**
            $sequence_{accepted}$ = $sequence_{mutated}$;
            $score_{accepted}$ = $score_{mutated}$;
        else
            ΔE = $score_{mutated}$ - $score_{accepted}$;
            p = min(1, exp(-ΔE/T));
            random_float = random.uniform(0,1);
            if *random_float $<= p$* **then**
                $sequence_{accepted}$ = $sequence_{mutated}$;
                $score_{accepted}$ = $score_{mutated}$;
            end
        end
        T = T*slope
    end
end

Algorithm 1: Pseudocode describing the Metropolis Monte Carlo simulated annealing algorithm used to find sequences with minimum in the defined objective function. Definitions functions (with x a riboswitch sequence): $E(x)$ = riboswitch objective function; mutate(x) = sequence mutation function; random(x) = random nucleotide selection. Definitions parameters: $score_{max}$ = maximal score; $p_{accept,initial}$ = initial acceptance probability if ΔE == 10% of $score_{max}$; iterations = number of iterations (50,000); $T_{initial}$ = -($score_{max}$*0.1)/log($p_{accept,initial}$); T_{final} = 0.01; slope = $(T_{final}/T_{initial})$**(1/iterations).

This algorithm simulates the metallurgic annealing process where controlled cooling maximized crystal size and minimizes defect formation. Overall, solutions are accepted if their score (score$_{\text{mutated}}$) is lower (better) than the previously accepted solution (score$_{\text{accepted}}$). If the score$_{\text{mutated}}$ is higher (worse) than score$_{\text{accepted}}$, the new (mutated) solution is accepted with following probability:

$$p_{\text{accept}} = e^{\frac{-\Delta E}{T}} \tag{B.1}$$

Where ΔE is equal to score$_{\text{mutated}}$ - score$_{\text{accepted}}$ and T is equal to the temperature. The initial acceptance probability of a ΔE equal to 1 was set to 0.2. Thus T$_{\text{initial}}$ is caluclated as follows:

$$T_{\text{initial}} = -\frac{1}{\log(0.2)} \tag{B.2}$$

The cooling scheme is logarithmic from the initial temperature to 0.01 (T$_{\text{final}}$). Temperature profile is calculated as follows:

$$T_{i+1} = T_i \times \text{slope} \tag{B.3}$$

Where slope is calculated as follows:

$$\text{slope} = \left(\frac{T_{\text{final}}}{T_{\text{initial}}}\right)^{\frac{1}{\text{iterations}}} \tag{B.4}$$

Where iterations are the number of iterations of simulated annealing. By default, iterations were set to 50.000.

B.1.2 Other settings

Riboswitch candidates where transcription of the random part downstream of the aptamer part interferes with the formation of the constrained aptamer secondary structure are not accepted (simulated by folding in steps of four nucleotides). Also, nucleotide repeats longer than four are not allowed in candidate sequences. Overall, the input settings were sequence constraints (fixed parts comprise the aptamer part and the first 12 nucleotides of the CDS), the thermodynamic property $\Delta\Delta G_{\text{target}}$ (Eq. 4.4), and the nucleotide regions n$_{\text{switching}}$ and n$_{\text{unpaired}}$, which are involved in the switching motion and are unpaired in the MFE configuration, respectively.

B.2   Supplementary Figures
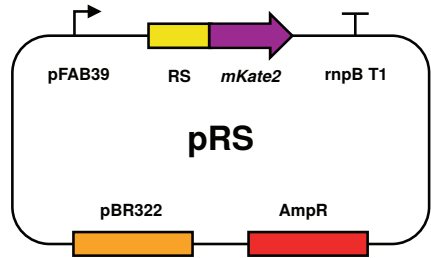
Figure B.1: Schematic overview of the plasmid type pRS used for the expression of riboswitches in this study. The plasmid (pBR322 vector[282]) uses pFAB39[23] as promoter, *mKate2*[252] as reporter gene and rnpB T1[24] as terminator. Riboswitches were cloned seamlessly between pFAB39 and *mKate2*.

Figure B.2: Overview of a DNA construct for riboswitch expression. As an example, the insert of pRS_1 is displayed with the pFAB39 promoter, RS_1 riboswitch, *mKate2* reporter gene, and rnpB T1 terminator.
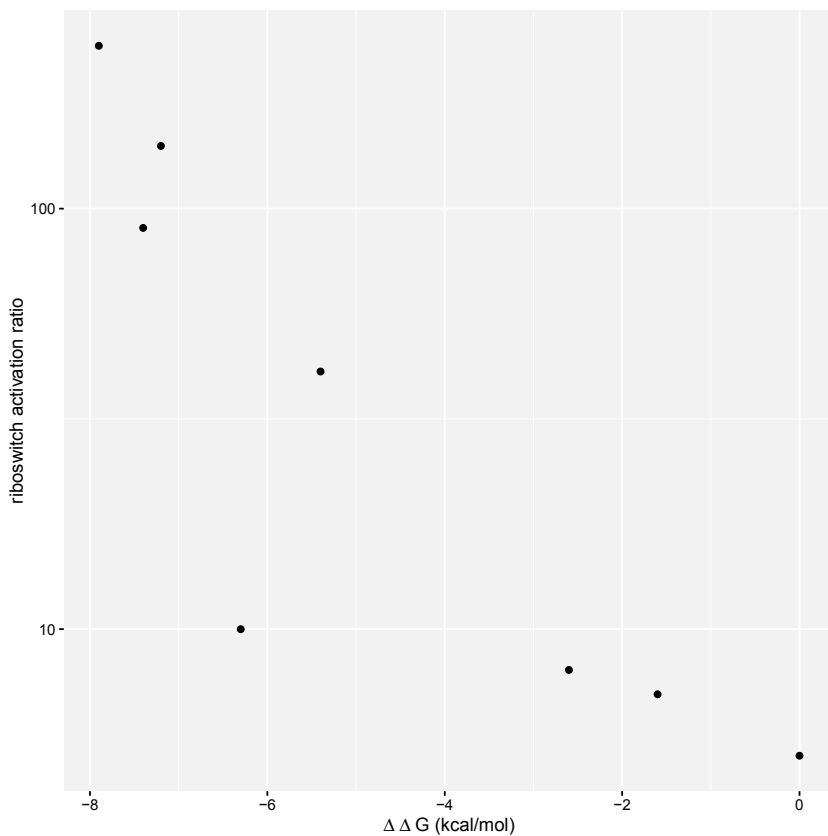
Figure B.3: Plot of $\Delta\Delta G$ (difference between predicted ΔG_{MFE} and predicted $\Delta G_{CONSTRAINED}$) against the activation ratio of previously described riboswitches [195]. Activation ratios displayed were defined and determined in Mishler & Gallivan [195].
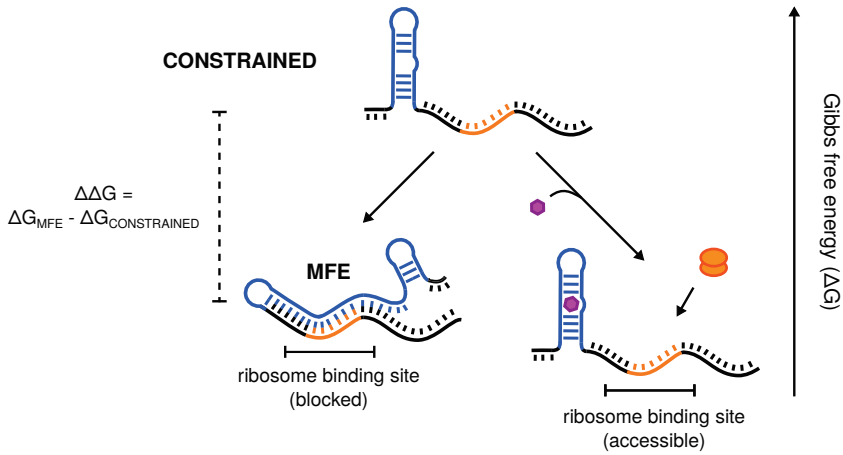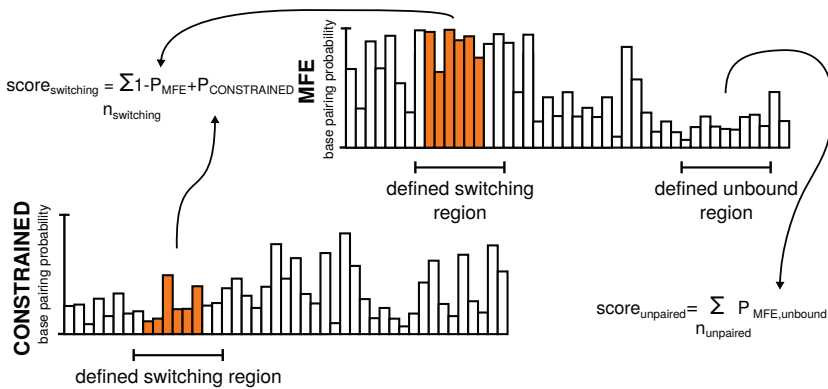
Figure B.4: Schematic overview of the objective function used for the automated design of *cis* encoded translational riboswitches. This objective function comprises both thermodynamical factors and structural properties.
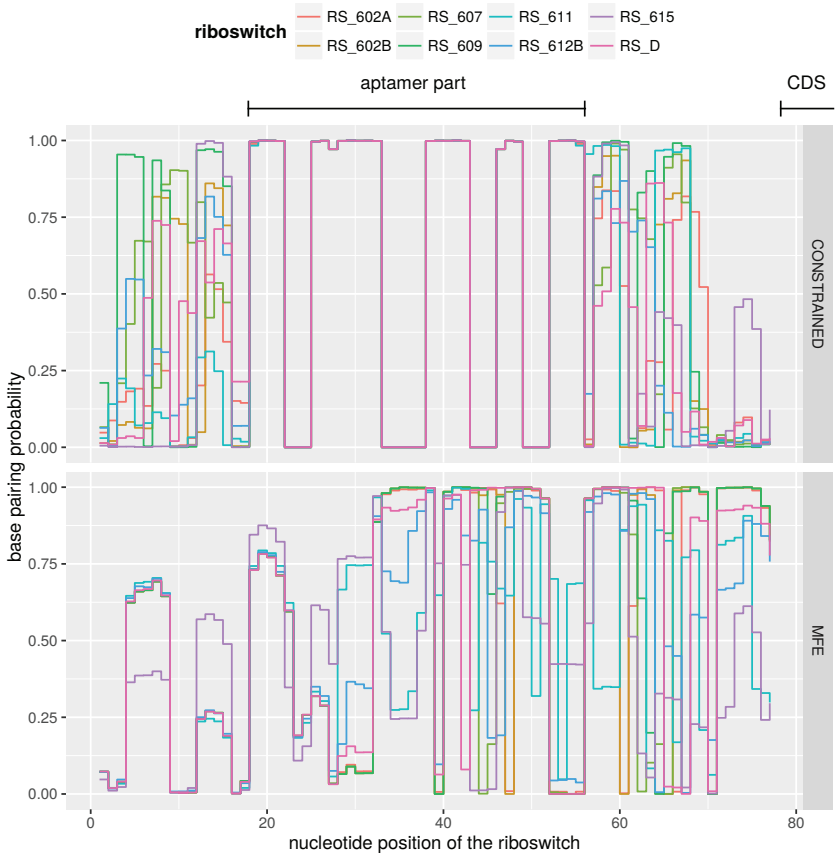
Figure B.5: Plot of the predicted base pairing probability in the minimum free energy (MFE) and the constrained form of previously described riboswitches[195]. The fixed aptamer region and the coding DNA sequence (CDS) is indicated above the plot.
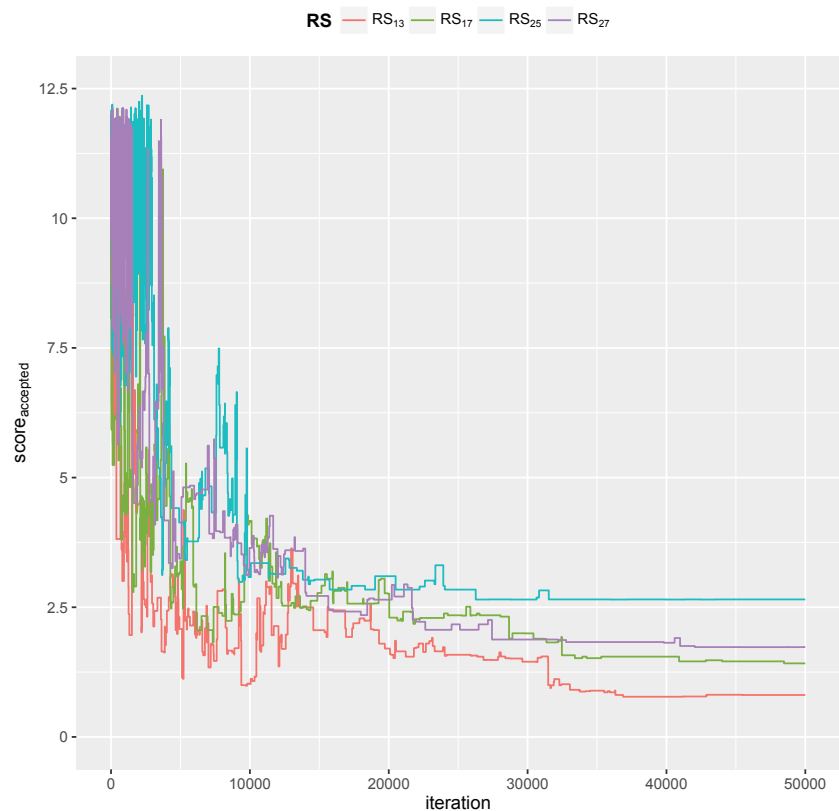
Figure B.6: Plot of the score of the accepted solution versus the iteration of the simulated annealing algorithm for 4 randomly chosen riboswitch designs (as an example).
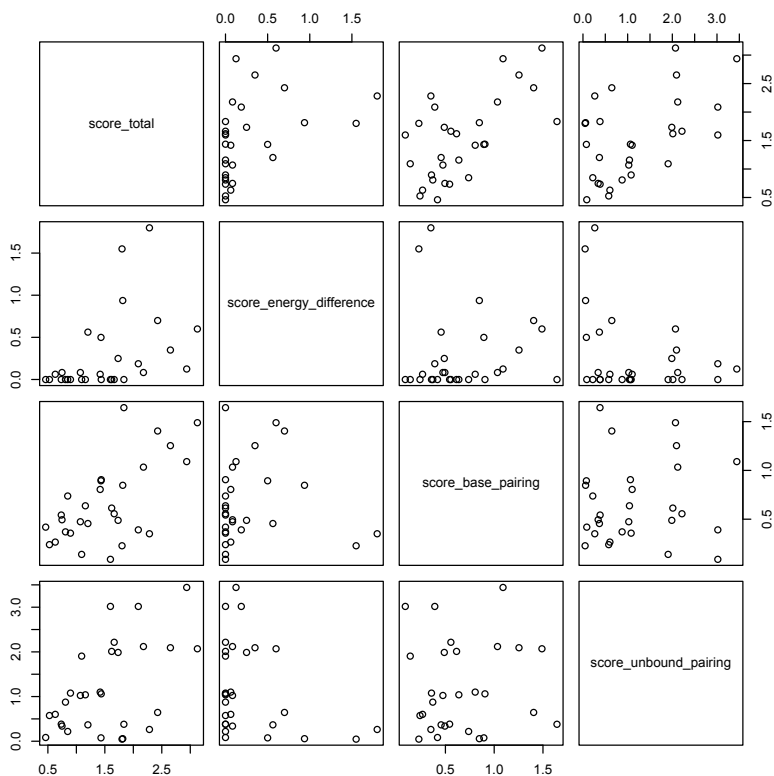
Figure B.7: Pairwise plot of the score$_{\text{total}}$, score$_{\Delta\Delta G}$, score$_{\text{switching}}$, and score$_{\text{unpaired}}$ of all designed riboswitches in this study.
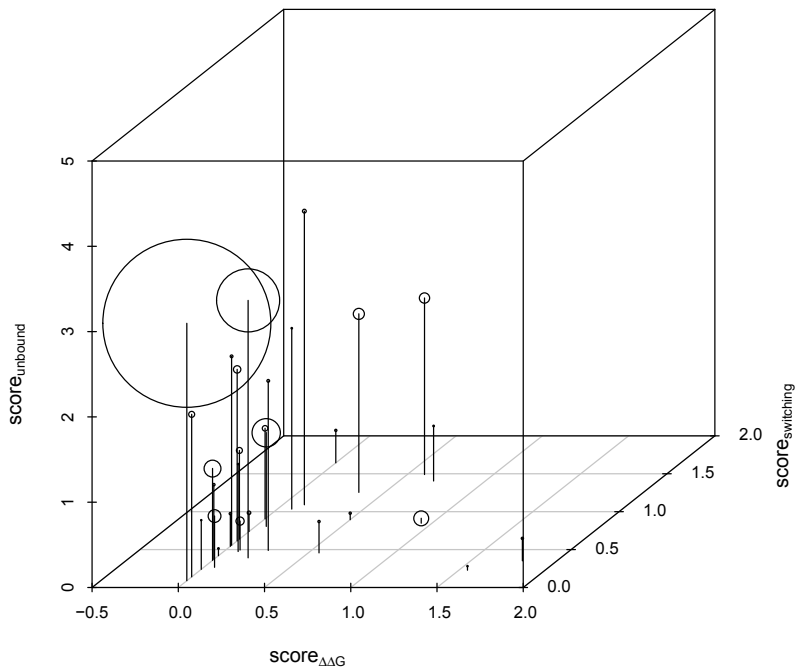
Figure B.8: 3D scatter plot of the $score_{\Delta\Delta G}$, $score_{switching}$, and $score_{unpaired}$ of all designed ribo-switches in this study. The size of the circles indicates the activation ratio (AR) of the designed riboswitches.
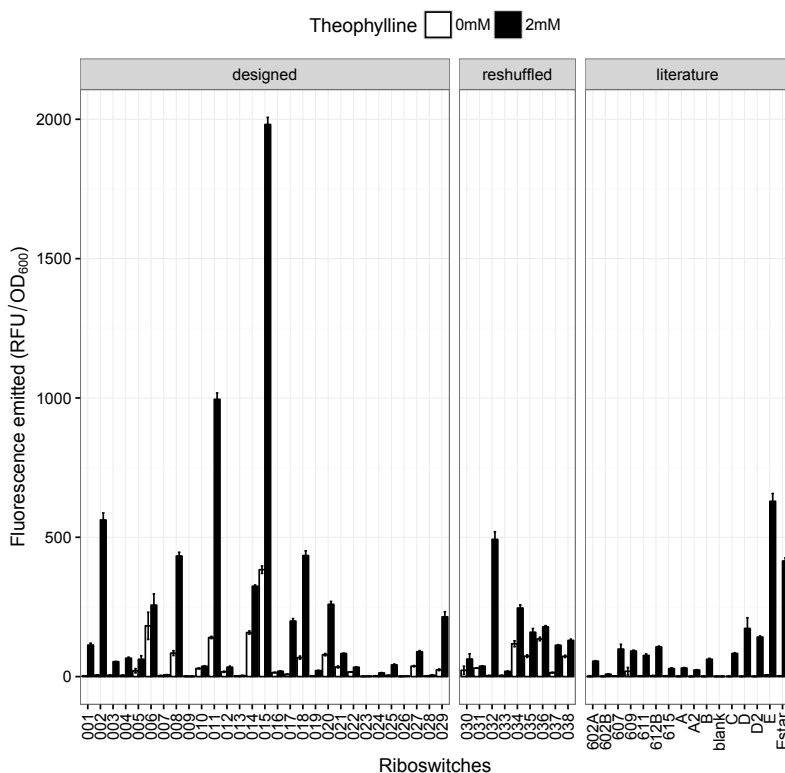
Figure B.9: Barplot of the gene expression modulation of the constructed riboswitches by measuring fluorescence emitted in the presence of theophylline (0 mM and 2 mM). All automated designed riboswitches are displayed on the left and previously described riboswitches are shown on the right. The riboswitches in the middle are constructed by reshuffling the regions up- and downstream of the aptamer of designed riboswitches. Both the designed, reshuffled, and the previously described riboswitches are expressed as depicted in Figure 4.3c. See Figure 4.4 for the activation ratios (ARs) of the constructed riboswitches.
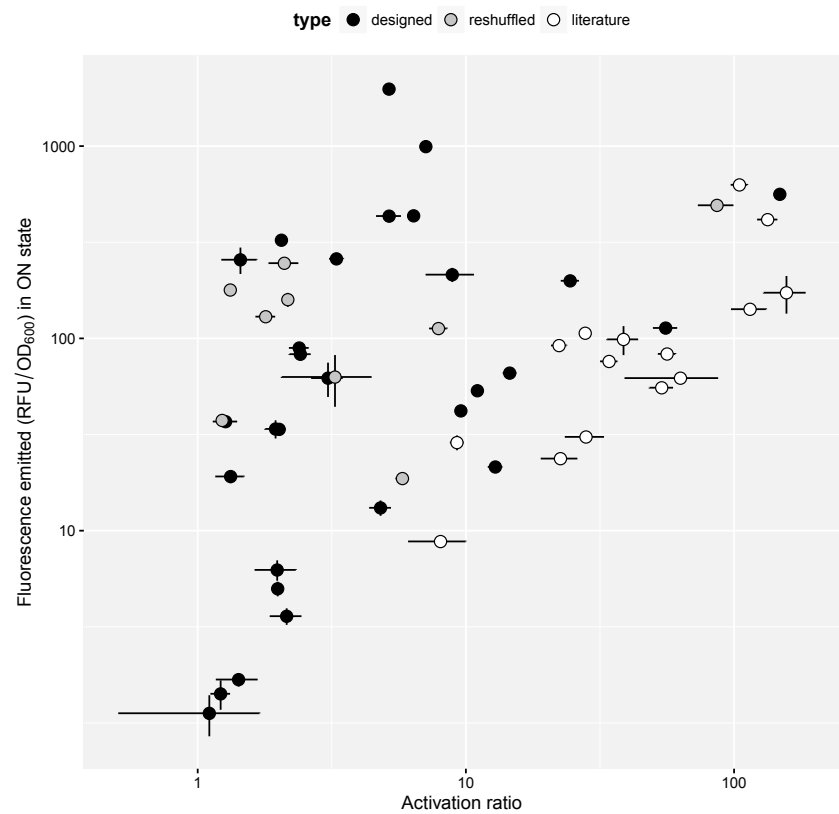
Figure B.10: Plot of the activation ratio (AR) of all riboswitches against the fluorescence emitted (RFP/OD$_{600}$) in the ON state (in the presence of 2 mM theophylline).
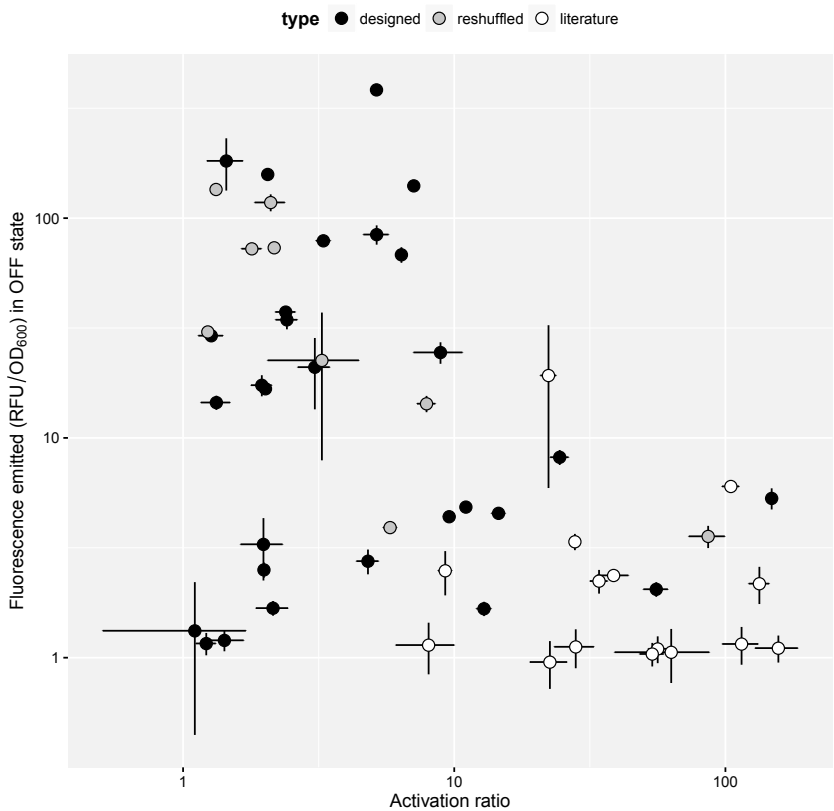
Figure B.11: Plot of the activation ratio (AR) of all riboswitches against the fluorescence emitted (RFP/OD$_{600}$) in the OFF state (in the absence of theophylline).
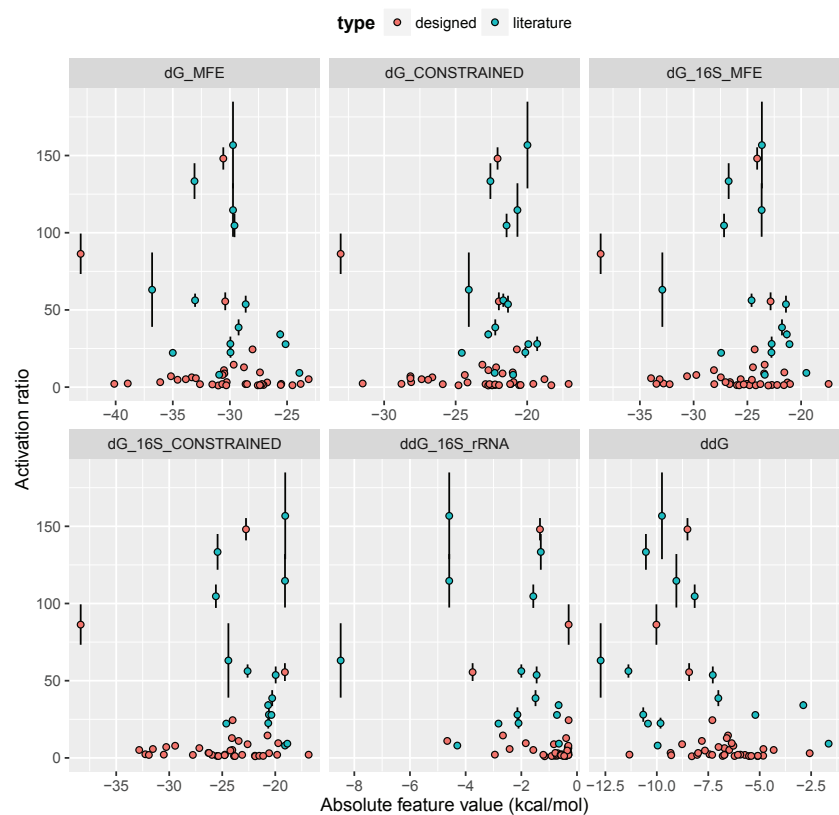
Figure B.12: Plot of the activation ratio (AR) of all riboswitches against the absolute values of 6 riboswitch features. dG_MFE and dG_CONSTRAINED represent the ensemble free energy of the minimum free energy (MFE) and constrained configuration respectively. ddG is the difference between dG_MFE and dG_CONSTRAINED. The ensemble free energy of the hybridization of the 16S rRNA to the riboswitch in the MFE and the constrained configuration (represented by dG_16S_MFE and dG_16S_CONSTRAINED respectively). The thermodynamics of this dimer complex was calculated using RNAcofold[257]. ddG_16S_rRNA is defined as the difference between dG_16S_MFE and dG_16S_CONSTRAINED.

Figure B.13: Plot of the activation ratio (AR) of all riboswitches against maximal AR predicted by Borujeni *et al.* [279].
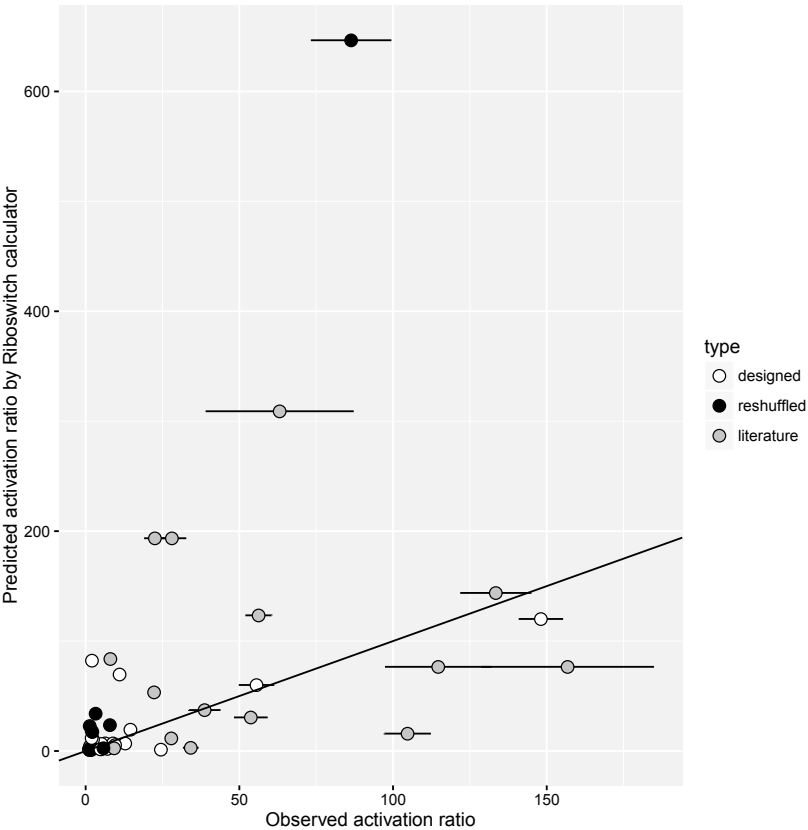
B.3 Supplementary Tables

Table B.1: Overview of all plasmids used in this study.

Name	Content (5'→3')	Backbone
pBlank	-	pSB6A1
pRS_1	pFAB39 - RS_1 - *mKate2* - rnpB T1	pSB6A1
pRS_2	pFAB39 - RS_2 - *mKate2* - rnpB T1	pSB6A1
pRS_3	pFAB39 - RS_3 - *mKate2* - rnpB T1	pSB6A1
pRS_4	pFAB39 - RS_4 - *mKate2* - rnpB T1	pSB6A1
pRS_5	pFAB39 - RS_5 - *mKate2* - rnpB T1	pSB6A1
pRS_6	pFAB39 - RS_6 - *mKate2* - rnpB T1	pSB6A1
pRS_7	pFAB39 - RS_7 - *mKate2* - rnpB T1	pSB6A1
pRS_8	pFAB39 - RS_8 - *mKate2* - rnpB T1	pSB6A1
pRS_9	pFAB39 - RS_9 - *mKate2* - rnpB T1	pSB6A1
pRS_{10}	pFAB39 - RS_{10} - *mKate2* - rnpB T1	pSB6A1
pRS_{11}	pFAB39 - RS_{11} - *mKate2* - rnpB T1	pSB6A1
pRS_{12}	pFAB39 - RS_{12} - *mKate2* - rnpB T1	pSB6A1
pRS_{13}	pFAB39 - RS_{13} - *mKate2* - rnpB T1	pSB6A1
pRS_{14}	pFAB39 - RS_{14} - *mKate2* - rnpB T1	pSB6A1
pRS_{15}	pFAB39 - RS_{15} - *mKate2* - rnpB T1	pSB6A1
pRS_{16}	pFAB39 - RS_{16} - *mKate2* - rnpB T1	pSB6A1
pRS_{17}	pFAB39 - RS_{17} - *mKate2* - rnpB T1	pSB6A1
pRS_{18}	pFAB39 - RS_{18} - *mKate2* - rnpB T1	pSB6A1
pRS_{19}	pFAB39 - RS_{19} - *mKate2* - rnpB T1	pSB6A1
pRS_{20}	pFAB39 - RS_{20} - *mKate2* - rnpB T1	pSB6A1
pRS_{21}	pFAB39 - RS_{21} - *mKate2* - rnpB T1	pSB6A1
pRS_{22}	pFAB39 - RS_{22} - *mKate2* - rnpB T1	pSB6A1
pRS_{23}	pFAB39 - RS_{23} - *mKate2* - rnpB T1	pSB6A1
pRS_{24}	pFAB39 - RS_{24} - *mKate2* - rnpB T1	pSB6A1
pRS_{25}	pFAB39 - RS_{25} - *mKate2* - rnpB T1	pSB6A1
pRS_{26}	pFAB39 - RS_{26} - *mKate2* - rnpB T1	pSB6A1
pRS_{27}	pFAB39 - RS_{27} - *mKate2* - rnpB T1	pSB6A1
pRS_{28}	pFAB39 - RS_{28} - *mKate2* - rnpB T1	pSB6A1
pRS_{29}	pFAB39 - RS_{29} - *mKate2* - rnpB T1	pSB6A1
pRS_{30}	pFAB39 - RS_{30} - *mKate2* - rnpB T1	pSB6A1
pRS_{31}	pFAB39 - RS_{31} - *mKate2* - rnpB T1	pSB6A1
pRS_{32}	pFAB39 - RS_{32} - *mKate2* - rnpB T1	pSB6A1
pRS_{33}	pFAB39 - RS_{33} - *mKate2* - rnpB T1	pSB6A1
pRS_{34}	pFAB39 - RS_{34} - *mKate2* - rnpB T1	pSB6A1
pRS_{35}	pFAB39 - RS_{35} - *mKate2* - rnpB T1	pSB6A1
pRS_{36}	pFAB39 - RS_{36} - *mKate2* - rnpB T1	pSB6A1
pRS_{37}	pFAB39 - RS_{37} - *mKate2* - rnpB T1	pSB6A1

pRS_{38}	pFAB39 - RS_{38} - *mKate2* - rnpB T1	pSB6A1
pRS_{A2}	pFAB39 - RS_{A2} - *mKate2* - rnpB T1	pSB6A1
pRS_{B}	pFAB39 - RS_{B} - *mKate2* - rnpB T1	pSB6A1
pRS_{C}	pFAB39 - RS_{C} - *mKate2* - rnpB T1	pSB6A1
pRS_{D2}	pFAB39 - RS_{D2} - *mKate2* - rnpB T1	pSB6A1
pRS_{E}	pFAB39 - RS_{E} - *mKate2* - rnpB T1	pSB6A1
pRS_{E*}	pFAB39 - RS_{E*} - *mKate2* - rnpB T1	pSB6A1
pRS_{A}	pFAB39 - RS_{A} - *mKate2* - rnpB T1	pSB6A1
pRS_{D}	pFAB39 - RS_{D} - *mKate2* - rnpB T1	pSB6A1
pRS_{602B}	pFAB39 - RS_{602B} - *mKate2* - rnpB T1	pSB6A1
pRS_{615}	pFAB39 - RS_{615} - *mKate2* - rnpB T1	pSB6A1
pRS_{611}	pFAB39 - RS_{611} - *mKate2* - rnpB T1	pSB6A1
pRS_{612B}	pFAB39 - RS_{612B} - *mKate2* - rnpB T1	pSB6A1
pRS_{602A}	pFAB39 - RS_{602A} - *mKate2* - rnpB T1	pSB6A1
pRS_{607}	pFAB39 - RS_{607} - *mKate2* - rnpB T1	pSB6A1
pRS_{609}	pFAB39 - RS_{609} - *mKate2* - rnpB T1	pSB6A1

Table B.2: Overview of important DNA sequences

Name	Sequence (5'→3')	Ref.
pFAB39	AAAAAGAGTATTGACTTCAGGAAAATTTTTCTGATACTTACAGCCAT	[23]
mKate2	ATGGTTAGCGAGCTGATCAAAGAAAACATGCACATGAAAGCTGATCAAAGAAAACATGCACATGA AACTGTATATGGAAGGCACCGTGAATAACCACCACTTTAAATGTACCAGCGAAGGTGAAAGCTGA TCAAAGAAAACATGCACATGAAACTGTATATGGAAGGCACCGTGAATAACCACCACTTTAAATGT ACCAGCGAAGGTGAAGGTAAACCGTATGAAGGCACCCAGACCATGCGTATTAAAGCAGTTGAAG GTGGTCCGCTGCCGTTTGCATTTGATATTCTGGCAACCAGCTTTATGTATGGCAGCAAAACCTTT ATTAACCATACCCAGGGTATCCCGGATTTTTTCAAACAGAGCTTTCCGGAAGGTTTTACCTGGGA ACGTGTTACCACCTATGAAGATGGTGGTGTTCTGACCGCAACCCAGGATACCAGTCTGCAGGATG GTTGTCTGATTTATAATGTGAAAATTCGCGGTGTGAACTTTCCGAGCAATGGTCCGGTTATGCAG AAAAAAACCCTGGGTTGGGAAGCAAGCACCGAAACCCTGTATCCGGCAGATGGTGGTCTGGAAG GTCGTGCAGATATGGCACTGAAACTGGTTGGTGGTGGTCATCTGATTTGCAATCTGAAAACCACC TATCGTAGCAAAAAACCGGCAAAAAATCTGAAAATGCCTGGCGTGTATTATGTTGATCGTCGTCT GGAACGTATTAAAGAGGCAGATAAAGAAACCTATGTGGAACAGCATGAAGTTGCAGTTGCACGTT ATTGTGATCTGCCGAGCAAACTGGGTCACCGCTGATAA	[252]
rnpB T1	TCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCAG AAAGATGAATGACTGTC	[24]

Table B.3: Overview of all *de novo* designed riboswitches.

Name	Sequence (5'→3')
RS_1	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCGCTCAACGGGCAACAAG
RS_2	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCAGCTGACGAGGACAACAAG
RS_3	AGGAAACCTTAATGCGTGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCATGCTGACGGCAACAAG
RS_4	ACCCCTTCAGGGAAACTGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCGCAACAACGGCAACAAG
RS_5	CTGGTGGGGAACATTTAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTAGGCCAGAGGACAAGATG
RS_6	CCGAGGAGAGATCGGTGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGAAAGAAAAGGACAAGATG
RS_7	GCTGTAGAGCACTCAAAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTAACTAGCACGGCGTCAAG
RS_8	GTGGGAAACTTGATGGAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCAGATCAATGGGTAGAGATG
RS_9	GGCGGATGAGCCTCCATGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCAAGGCATCATATGCATTACC
RS_{10}	CAGCTGGAGAATTATAGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTCGGCTGGAAGGGATTAAG
RS_{11}	TTCCAAGACTGGGAGAAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCAACACATGAGGGTATAGATG
RS_{12}	GATCCATCTATTTATCAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTAGTAGCCGGCATACAGACG
RS_{13}	GGGAAACCCGAGTTGTTGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGGATTAATGTTTCCCGGGGTACAAG
RS_{14}	AAGTGTTGGGTGGATTTGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGAAGGAGAGGGACAAAGGTATTAGG
RS_{15}	AGTAAGTGGTACTATTGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTTTTAAACGCAAACGGGGTATTGAG
RS_{16}	TCTCACTGTGAGCAACGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCATTCACCTTCCGAGGGTATGCAGG
RS_{17}	AAATCCTGTCAGGTTATGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCATTTCGTAACGGAGTTCGAA
RS_{18}	AGGATCAAGCTATAAAAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGCTGTGATCGAGGTATTAAG
RS_{19}	TATTTTGGTAGTCCCCGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGCTATTATCAGCACAGGACG
RS_{20}	CGAGAAGTCTAATTATCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGGAGCTGTGGCAAGATGATA
RS_{21}	TACTTATACGCCCGTGTGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGATGGGAACGCAACGACGACACCAC
RS_{22}	AGGCCACACTACAAAAGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTACTACCAATCAACGGCTCAAGATG
RS_{23}	CACTGATATCTCTGATAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTTGCGGAGGGCAGTGAGATG

RS_{24} TAACAAAAGACCGCCGTGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCC
TAGGGTGTCTGGATTATC

RS_{25} GAACCCATATTTGTTAGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCGA
CTGTTCGGGTGTATTAGG

RS_{26} CAATTTCCATCGAGTGAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTTA
CTGGTGTCCCTTATTTG

RS_{27} GGTTGTCATGCGTAGCGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCG
CGCGTCGCGGCAGGGGTATTAGG

RS_{28} GAGGAATAACGCCTCCGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCC
TTCTGCTAAATCAGCACAAGACG

RS_{29} GCCGATGACAAGACCACGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTG
ACATAACACCCAAGAAGCACAAG

RS_{30} CGAGAAGTCTAATTATCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTAC
TACCAATCAACGGCTCAAGATG

RS_{31} GGTTGTCATGCGTAGCGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTG
ACATAACACCCAAGAAGCACAAG

RS_{32} CGAGAAGTCTAATTATCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTT
GCGGAGGGCAGTGAGATG

RS_{33} CACTGATATCTCTGATAGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCAT
TCACCTTCCGAGGGTATGCAGG

RS_{34} GAGGAATAACGCCTCCGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCT
ATGGACCCAAGGGGCTTAAGATG

RS_{35} TATTTTGGTAGTCCCCGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTTG
CGGAGGGCAGTGAGATG

RS_{36} TATTTTGGTAGTCCCCGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCG
CGCGTCGCGGCAGGGGTATTAGG

RS_{37} GGCGGATGAGCCTCCATGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCT
ATGGACCCAAGGGGCTTAAGATG

RS_{38} AGTAAGTGGTACTATTGGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCT
ATGGACCCAAGGGGCTTAAGATG

Table B.4: Overview of all previously developed riboswitches.

Name	Sequence (5'→3')	Ref.
RS_{A2}	ATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGAGAA GGGGCAACAAG	283
RS_B	ATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCGCTGCG CAGGGGGTATCAACAAG	283
RS_C	ATACGACTCACTATAGGTACCTGATAAGATAGGGGTGATACCAGCATCGTCTTGATGCCCTTGGCA GCACCAAGGGACAACAAG	283
RS_{D2}	ATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTAA GGTAACAACAAG	283
RS_E	ATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTAA GGAGGTAACAACAAG	283
RS_{E*}	ATACGACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTAA GGAGGCAACAAG	283
RS_A	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGAGAAGGG GCAACAAG	195
RS_D	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTAAGGTA ACAACAAG	195
RS_{602B}	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCGAAGATG GCAACAAG	195
RS_{615}	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCCACTTCT ACAACAAG	195
RS_{611}	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCGGGCGCA GCAACAAG	195
RS_{612B}	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTGGGACA ACAACAAG	195
RS_{602A}	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCAAAGGAT GCAACAAG	195
RS_{607}	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCTATGATG GCAACAAG	195
RS_{609}	GACTCACTATAGGTACCGGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCCTGCCGGGCG GGCAACAAG	195

Table B.5: Overview of all inputs for the automated design of translational riboswitches. The XXXXXX in the sequence constraint represents the place for the aptamer sequence, which is the theophylline aptamer (GGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACC)[185].

Name	Sequence constraints (5'→3') $n_{switching}$ (|) and $n_{unpaired}$ (x) constraints	$\Delta\Delta G$
RS$_1$	GACUCACUAUAGGUACCXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNCAACAAGAUGGUUAGCGAG ....||.||||||xxxxxxxxx	-8
RS$_2$	GACUCACUAUAGGUACCXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNCAACAAGAUGGUUAGCGAG ....||.||||||xxxxxxxxx	-8
RS$_3$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNCAACAAGAUGGUUAGCGAG ....||.||||||xxxxxxxxx	-8
RS$_4$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNCAACAAGAUGGUUAGCGAG ....||.||||||xxxxxxxxx	-8
RS$_5$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNAGGANNNNNNNNNAUGGUUAGCGAG ....|||||||||xxxxxxxxx	-6
RS$_6$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNAGGANNNNNNNNNAUGGUUAGCGAG ....|||||||||xxxxxxxxx	-6
RS$_7$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||...xxxxxxxxx	-8
RS$_8$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||...xxxxxxxxx	-8
RS$_9$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||xxxxxxxxx	-6
RS$_{10}$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||xxxxxxxxx	-6
RS$_{11}$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||.....xxxxxxxxx	-6
RS$_{12}$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....||||||||||.....xxxxxxxxx	-6
RS$_{13}$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||xxxxxxxxx	-6
RS$_{14}$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....||||||||||xxxxxxxxx	-6
RS$_{15}$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||.....xxxxxxxxx	-6
RS$_{16}$	NNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||||.....xxxxxxxxx	-6
RS$_{17}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||xxxxxxxxx	-8
RS$_{18}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||||xxxxxxxxx	-8
RS$_{19}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||.....xxxxxxxxx	-8
RS$_{20}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||||.....xxxxxxxxx	-8
RS$_{21}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....||||||||||||xxxxxxxxx	-8
RS$_{22}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||.....xxxxxxxxx	-8
RS$_{23}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||xxxxxxxxx	-10
RS$_{24}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||||xxxxxxxxx	-10
RS$_{25}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....||||||||||||xxxxxxxxx	-10
RS$_{26}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||.....xxxxxxxxx	-10
RS$_{27}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....||||||||||||xxxxxxxxx	-10
RS$_{28}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....|||||||||.....xxxxxxxxx	-10
RS$_{29}$	NNNNNNNNNNNNNNNNNNNNNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXNNNNNNNNNNNNNNNNNNNNNNAUGGUUAGCGAG ....||||||||||||.....xxxxxxxxx	-10

Table B.6: Overview of the reshuffled riboswitches (RS_{30} through RS_{38}) with the origin of the up- and downstream part of the fixed aptamer region (the theophylline aptamer (GGU-GAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACC)[185]). Three downstream regions (indicated with a *) originate from a riboswitch that was only designed *in silico* and that was not constructed due to technical difficulties in the cloning of the plasmid.

Name	Upstream of aptamer	Downstream of aptamer
RS_{30}	RS_{20}	RS_{22}
RS_{31}	RS_{27}	RS_{29}
RS_{32}	RS_{20}	RS_{23}
RS_{33}	RS_{23}	RS_{16}
RS_{34}	RS_{28}	*
RS_{35}	RS_{19}	RS_{23}
RS_{36}	RS_{19}	RS_{27}
RS_{37}	RS_9	*
RS_{38}	RS_{15}	*

*RS only designed, not constructed due to technical difficulties

Appendix C

Development of Neu5Ac-responsive biosensors based on the transcriptional regulator NanR
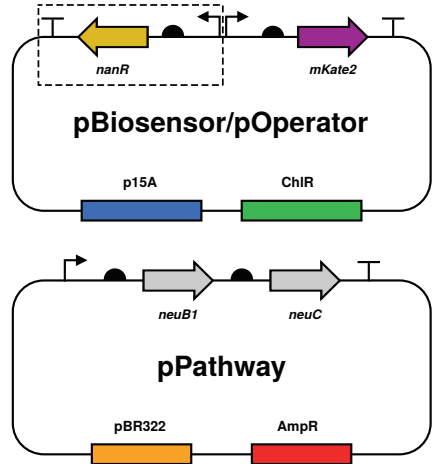
C.1 Supplementary Figures

Figure C.1: Schematic overview of the plasmid types used in this study. The pBiosensor (pNB$_1$-pNB$_4$ and pEB$_1$-pEB$_4$) and pOperator (pO$_1$-pO$_8$) plasmids (p15A origin of replication and chloramphenicol resistance marker, originating from pACYCDuet[306]), are used for mKate2[252] expression with and without NanR expression, respectively. The pOperator plasmids do not contain the part between the dotted rectangle. The pPathway plasmids (pPathway$_{low}$, pPathway$_{medium}$, and pPathway$_{high}$) comprise a medium-copy vector (pBR322 origin of replication and ampicillin resistance marker, originating from pSB6A1[249]) used for the expression of the biosynthetic pathway of Neu5Ac.

Figure C.2: Schematic overview of the biosensor constructs used in this study on a medium-copy vector (p15A origin of replication and chloramphenicol resistance marker, originating from pACYCDuet(Novagen)). These plasmids were used to express the transcription factor (TF) NanR[307] and the reporter gene *mKate2*[252]. On all biosensor plasmids except pNB₁, NanR was expressed using BBa_J23114[308] as promoter, BBa_B0031[308] as ribosome binding site (RBS), and rnpB T1[24] as terminator and mKate was expressed using various promoters and the rnpB T1[24] terminator. A) Plasmids pNB₁ through pNB₄ contain native intergenic regions for mKate2 expression. pNB₁ contains the native NanR cluster of *E. coli* (including the first 13 residues of NanA) as depicted in Figure 5.2. B) Plasmids pEB₁ through pEB₄ contains engineered promoters and UTRs for mKate2 expression. Details of all regulatory DNA sequences depicted here are listed in Supplementary Table C.3. Details on all coding DNA sequences (CDSs) are listed in Supplementary Table C.4.

CAATTGTGTTGAACTCCGTGTCAAAAGAAAACGGTCAATCCCATAAACGGCAGATTGAAA
ACAACGATGTTATATTTTTTTGCAAGGCTATTTATGGTGCGGATGTCGTGTTTTTAATTGT
AGGTGAGGTGATTTTTCATTAAAAAATATGCGCTTATGATTATTTTGTAAGAACACATTC
ATAATATTCATAATGCTCGTGAATAGTCTTATAAATAATTCAAACGGGATGTTTTTATCT
GCGTTACATTAATTTTTCGCAATAGTTAATTATTCCGTTAATTATGGTAATGATGAGGCA
CAAAGAGAAAACCCTGCCATTTTCCCCTACTTTCAATCCTGTGATAGGATGTCACTGATG
ATGTTAATCACACTGACCTTACAGAATGGGCCTTATGAACGCATTTGATTCGCAAACCGA
AGATTCTTCACCTGCAATTGGTCGCAACTTGCGTAGCCGCCCGCTGGCGCGTAAAAAACT
CTCCGAAATGGTGGAAGAAGAGCTGGAACAGATGATCCGCCGTCGTGAATTTGGCGAAGG
TGAACAATTACCGTCTGAACGCGAACTGATGGCGTTCTTTAACGTCGGGCGTCCTTCGGT
GCGTGAAGCGCTGGCAGCGTTAAAACGCAAAGGTCTGGTGCAAATAAACAACGGCGAACG
CGCTCGCGTCTCGCGTCCTTCTGCGGACACTATCATCGGTGAGCTTTCCGGCATGGCGAA
AGATTTCCTTTCTCATCCCGGTGGGATTGCCCATTTCGAACAATTACGTCTGTTCTTTGA
ATCCAGTCTGGTGCGCTATGCGGCTGAACATGCCACCGATGAGCAAATCGATTTGCTGGC
AAAAGCACTGGAAATCAACAGTCAGTCGCTGGATAACAACGCGGCATTCATTCGTTCAGA
CGTTGATTTCCACCGCGTGCTGGCGGAGATCCCCGGTAACCCAATCTTCATGGCGATCCA
CGTTGCCCTGCTCGACTGGCTTATTGCCGCACGCCCAACGGTTACCGATCAGGCACTGCA
CGAACATAACAACGTTAGTTATCAACAGCATATTGCGATCGTTGATGCGATCCGCCGTCA
TGATCCTGACGAAGCCGATCGTGCGTTGCAATCGCATCTCAACAGCGTCTCTGCTACCTG
GCACGCTTTCGGTCAGACCACCAACAAAAAGAAATAATGCCACTTTAGTGAAGCAGATCG
CATTATAAGCTTTCTGTATGGGGTGTTGCTTAATTGATCTGGTATAACAGGTATAAAGGT
ATATCGTTTATCAGACAAGCATCACTTCAGAGGTATTTATGGCAACGAATTTACGTGGCG
TAATGGCTGCACTCCTGGTTAGCGAGCTGATCAAAGAAAACATGCACATGAAACTGTATA
TGGAAGGCACCGTGAATAACCACCACTTTAAATGTACCAGCGAAGGTGAAGGTAAACCGT
ATGAAGGCACCCAGACCATGCGTATTAAAGCAGTTGAAGGTGGTCCGCTGCCGTTTGCAT
TTGATATTCTGGCAACCAGCTTTATGTATGGCAGCAAAACCTTTATTAACCATACCCAGG
GTATCCCGGATTTTTTTCAAACAGAGCTTTCCGGAAGGTTTTTACCTGGGAACGTGTTACCA
CCTATGAAGATGGTGGTGTTCTGACCGCAACCCAGGATACCAGTCTGCAGGATGGTTGTC
TGATTTATAATGTGAAAATTCGCGGTGTGAACTTTCCGAGCAATGGTCCGGTTATGCAGA
AAAAAACCCTGGGTTGGGAAGCAAGCACCGAAACCCTGTATCCGGCAGATGGTGGTCTGG
AAGGTCGTGCAGATATGGCACTGAAACTGGTTGGTGGTGGTCATCTGATTTGCAATCTGA
AAACCACCTATCGTAGCAAAAAACCGGCAAAAAATCTGAAAATGCCTGGCGTGTATTATG
TTGATCGTCGTCTGGAACGTATTAAAGAGGCAGATAAAGAAACCTATGTGGAACAGCATG
AAGTTGCAGTTGCACGTTATTGTGATCTGCCGAGCAAACTGGGTCACCGCTGATAACCAT
GGGCTAGCGGTTTGAAGGGTATTGGTCGGTCAGTTTCACCTGATTTACGTAAAAACCCGC
TTCGGCGGGTTTTTTGCTTTTTGGAGGGGCAGAAAGATGAATGACTGTCCTTTTATTGGAGA

Figure C.3:   Detailed overview of pNB₁ with the pNanR + untranslated region (UTR), NanR coding DNA sequence (CDS), pNanA + NanA UTR, nanA₁₃ₐₐ (first 13 residues of *nanA* CDS), *mKate2*²⁵² CDS and, rnpB T1 terminator.

```
CAATAAAAGGACAGTCATTCATCTTTCTGCCCCTCCAAAAGCAAAAACCCGCCGAAGCGG
GTTTTTACGTAAATCAGGTGAAACTGACCGACCAATACCCTTCAAACCGCTAGCCCATGG
TTATTATTTCTTTTTGTTGGTGGTCTGACCGAAAGCGTGCCAGGTAGCAGAGACGCTGTT
GAGATGCGATTGCAACGCACGATCGGCTTCGTCAGGATCATGACGGCGGATCGCATCAAC
GATCGCAATATGCTGTTGATAACTAACGTTGTTATGTTCGTGCAGTGCCTGATCGGTAAC
CGTTGGGCGTGCGGCAATAAGCCAGTCGAGCAGGGCAACGTGGATCGCCATGAAGATTGG
GTTACCGGGGATCTCCGCCAGCACGCGGTGGAAATCAACGTCTGAACGAATGAATGCCGC
GTTGTTATCCAGCGACTGACTGTTGATTTCCAGTGCTTTTGCCAGCAAATCGATTTGCTC
ATCGGTGGCATGTTCAGCCGCATAGCGCACCAGACTGGATTCAAAGAACAGACGTAATTG
TTCGAAATGGGCAATCCCACCGGGATGAGAAAGGAAATCTTTCGCCATGCCGGAAAGCTC
ACCGATGATAGTGTCCGCAGAAGGACGCGAGACGCGAGCGCGTTCGCCGTTGTTTATTTG
CACCAGACCTTTGCGTTTTAACGCTGCCAGCGCTTCACGCACCGAAGGACGCCCGACGTT
AAAGAACGCCATCAGTTCGCGTTCAGACGGTAATTGTTCACCTTCGCCAAATTCACGACG
GCGGATCATCTGTTCCAGCTCTTCTTCCACCATTTCGGAGAGTTTTTTACGCGCCAGCGG
GCGGCTACGCAAGTTGCGACCAATTGCAGGTGAAGAATCTTCGGTTTGCGAATCAAATGC
GTTCATAAGGCCCATCTAGTAGGTTTCCTGTGTGACTCTAGTAGCTAGCATTGTACCTAG
GACTGAGCTAGCCATAAACTCTAGAAGCGGCCGCGAATTCTTGAAGACGAAAGGGCCTCG
TGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTG
GCAGCTCCAGCCTACACCAATTGTGCCACTTTAGTGAAGCAGATCGGCATTATAAGCTTTC
TGTATGGGGTGTTGCTTAATTGATCTGGTATAACAGGTATAAAGGTATATCGTTTATCAG
ACAAGCATCACTTCAGAGGTATTTATGGCAACGAATTTACGTGGCGTAATGGCTGCACTC
CTGGTTAGCGAGCTGATCAAAGAAAACATGCACATGAAACTGTATATGGAAGGCACCGTG
AATAACCACCACTTTAAATGTACCGGCGAAGGTGAAGGTAAACCGTATGAAGGCACCCAG
ACCATGCGTATTAAAGCAGTTGAAGGTGGTCCGCTGCCGTTTGCATTTGATATTCTGGCA
ACCAGCTTTATGTATGGCAGCAAAACCTTTATTAACCATACCCAGGGTATCCCGGATTTT
TTCAAACAGAGCTTTCCGGAAGGTTTTACCTGGGAACGTGTTACCACCTATGAAGATGGT
GGTGTTCTGACCGCAACCCAGGATACCAGTCTGCAGGATGGTTGTCTGATTTATAATGTC
AAAATTCGCGGTGTGAACTTTCCGAGCAATGGTCCGGTTATGCAGAAAAAAACCCTGGGT
TGGGAAGCAAGCACCGAAACCCTGTATCCGGCAGATGGTGGTCTGGAAGGTCGTGCAGAT
ATGGCACTGAAACTGGTTGGTGGTGGTCATCTGATTTGCAATCTGAAAACCACCTATCGT
AGCAAAAAACCGGCAAAAAATCTGAAAATGCCTGGCGTGTATTATGTTGATCGTCGTCTG
GAACGTATTAAAGAGGCAGATAAAGAAACCTATGTGGAACAGCATGAAGTTGCAGTTGCA
CGTTATTGTGATCTGCCGAGCAAACTGGGTCACCGCTGATAACCATGGGCTAGCGGTTTG
AAGGGTATTGGTCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTT
GCTTTTGGAGGGGCAGAAAGATGAATGACTGTCCTTTTATT
```

Figure C.4: Detailed overview of a DNA construct using native promoters (with untranslated region (UTR)) containing NanR operators (pNB$_2$ through pNB$_4$). As an example, the insert of pNB$_2$ is displayed with the rnpB T1 terminator, NanR coding DNA sequence (CDS), BBa_B0031[308] ribosome binding site (RBS), pNanA + NanA UTR, nanA$_{13AA}$ (first 13 residues of *nanA* CDS) and, *mKate2*[252] CDS.

CAATAAAAGGACAGTCATTCATCTTTCTGCCCCTCCAAAAGCAAAAACCCGCCGAAGCGG
GTTTTTACGTAAATCAGGTGAAACTGACCGACCAATACCCTTCAAACCGCTAGCCCATGG
TTATTTATTTCTTTTTGTTGGTGGTCTGACCGAAAGCGTGCCAGGTAGCAGAGACGCTGTT
GAGATGCGATTGCAACGCACGATCGGCTTCGTCAGGATCATGACGGCGGATCGCATCAAC
GATCGCAATATGCTGTTGATAACTAACGTTGTTATGTTCGTGCAGTGCCTGATCGGTAAC
CGTTGGGCGTGCGGCAATAAGCCAGTCGAGCAGGGCAACGTGGATCGGCCATGAAGATTGG
GTTACCGGGGATCTCCGCCAGCACGCGGTGGAAATCAACGTCTGAACGAATGAATGCCGC
GTTGTTATCCAGCGACTGACTGTTGATTTCCAGTGCTTTTGCCAGCAAATCGATTTGCTC
ATCGGTGGCATGTTCAGCCGCATAGCGCACCAGACTGGATTCAAAGAACAGACGTAATTG
TTCGAAATGGGCAATCCCACCGGGATGAGAAAGGAAATCTTTCGCCATGCCGGAAAGCTC
ACCGATGATAGTGTCCGCAGAAGGACGCGAGACGCGAGCGCGTTCGCCGTTGTTTATTTG
CACCAGACCTTTGCGTTTTAACGCTGCCAGCGCTTCACGCACCGAAGGACGCCCGACGTT
AAAGAACGCCATCAGTTCGCGTTCAGACGGTAATTGTTCACCTTCGCCAAATTCACGACG
GCGGATCATCTGTTCCAGCTCTTCTTCCACCATTTCGGAGAGTTTTTTACGCGCCAGCGG
GCGGCTACGCAAGTTGCGACCAATTGCAGGTGAAGAATCTTCGGTTTGCGAATCAAATGC
GTTCATAAGGCCCATCTAGTAGGTTTCCTGTGTGACTCTAGTAGCTAGCATTGTACCTAG
GACTGAGCTAGCCATAAACTCTAGAAGCGGCCGCGAATTCTTGAAGACGAAAGGGCCTCG
TGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTG
GCAGCTCCAGCCTACACCAATTGAGCTTTCTGTATGGGGTGTTGCTTAATTGATCTGGTA
TAACAGGTATAAAGGTATATCGTTTATACTAGAGAAAGAGGAGAAATACTAGATGGTTAG
CGAGCTGATCAAAGAAAACATGCACATGAAACTGTATATGGAAGGCACCGTGAATAACCA
CCACTTTAAATGTACCAGCGAAGGTGAAGGTAAACCGTATGAAGGCACCCAGACCATGCG
TATTAAAGCAGTTGAAGGTGGTCCGCTGCCGTTTGCATTTGATATTCTGGCAACCAGCTT
TATGTATGGCAGCAAAACCTTTATTAACCATACCCAGGGTATCCCGGATTTTTTCAAACA
GAGCTTTCCGGAAGGTTTTACCTGGGAACGTGTTACCACCTATGAAGATGGTGGTGTTCT
GACCGCAACCCAGGATACCAGTCTGCAGGATGGTTGTCTGATTTATAATGTGAAAATTCG
CGGTGTGAACTTTCCGAGCAATGGTCCGGTTATGCAGAAAAAAACCCTGGGTTGGGAAGC
AAGCACCGAAACCCTGTATCCGGCAGATGGTGGTCTGGAAGGTCGTGCAGATATGGCACT
GAAACTGGTTGGTGGTGGTCATCTGATTTGCAATCTGAAAACCACCTATCGTAGCAAAAA
ACCGGCAAAAAATCTGAAAATGCCTGGCGTGTATTATGTTGATCGTCGTCTGGAACGTAT
TAAAGAGGCAGATAAAGAAACCTATGTGGAACAGCATGAAGTTGCAGTTGCACGTTATTG
TGATCTGCCGAGCAAACTGGGTCACCGCTGATAACCATGGGCTAGCGGTTTGAAGGGTAT
TGGTCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGG
AGGGGCAGAAAGATGAATGACTGTCCTTTTATT

Figure C.5: Detailed overview of a DNA construct using engineered promoters containing NanR operators (pEB$_1$ through pEB$_4$). As an example, the insert of pEB$_1$ is displayed with the rnpB T1 terminator, NanR coding DNA sequence (CDS), BBa_B0031[308] ribosome binding site (RBS), BBa_J23114 promoter[308], pNanA promoter, BBa_B0034[308] RBS, and, *mKate2*[252] CDS.

```
TCTCCGGGAGAGCTCGATATCCCGGGCGGCCGCCTTCATTCTATAAGTTTCTTGACATCT
TGGCCGGCATATGGTATAATAGGGCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGG
AGATATACATAATGAAAGAGATTAAGATCCAGAATATTATCATCAGCGAAGAGAAAGCGC
CGCTGGTGGTGCCGGAAATTGGCATTAACCCACAACGGCAGCCTGGAACTGGCTAAGATTA
TGGTGGATGCCGCGTTCAGCGCCGGTGCGAAGATCATTAAACATCAGACGCATATCGTTG
AAGATGAGATGAGCAAGGCGGCGAAGAAGGTGATTCCTGGCAACGCCAAGATTAGCATCT
ATGAGATCATGCAGAAATGCGCGCTTGATTATAAAGATGAACTGGCGCTGAAAGAATATA
CCGAGAAGTTAGGTCTGGTCTATCTGTCGACGCCATTCTCGCGCGCAGGTGCCAACCGTC
TGGAAGATATGGGCGTGTCTGCCTTCAAGATTGGTTCCGGTGAATGTAATAATTATCCAC
TGATCAAGCATATTGCCGCATTCAAGAAGCCGATGATTGTCAGCACCGGCATGAACAGCA
TTGAATCTATCAAACCGACCGTTAAGATTCTGCTGGATAATGAGATTCCGTTCGTTCTGA
TGCACACCACCAATCTGTATCCGACGCCGCATAACCTGGTTCGCCTGAACGCGATGCTGG
AGCTGAAGAAGGAGTTCTCCTGTATGGTTGGCCTGAGCGATCATACCACCGATAACCTCG
CCTGTCTTGGCGCGGTGGTTCTCGGCGCATGCGTGCTTGAACGTCACTTCACCGACAGCA
TGCATCGCAGCGGTCCGGATATCGTCTGCTCGATGGATACCAAGGCACTGAAGGAACTGA
TTATTCAGAGCGAGCAGATGGCGATTATTCGCGGCAATAACGAATCCAAGAAGGCCGCCA
AGCAGGAACAGGTGACCATCGACTTCGCGTTCGCTTCGGTGGTCAGTATTAAGGACATCA
AGAAAGGCGAAGTGCTGTCAATGGACAACATCTGGGTGAAGCGTCCAGGCTTAGGCGGCA
TCAGTGCGGCAGAATTCGAGAACATTCTCGGTAAGAAGGCTCTGCGCGATATTGAGAATG
ATGCGCAGCTGAGCTATGAAGACTTCGCCTGATAAGAATTCGTTTAGAGCTCTAAATAAG
GAGGAATAACCATGGTGAAGAAGATCCTGTTCATTACCGGCTCCCGCGCCGACTACAGCA
AAATTAAATCGCTGATGTATCGCGTGCAGAATAGCAGCGAGTTTGAGCTCTATATCTTCG
CCACCGGGATGCACCTGTCGAAAAACTTCGGCTACACCGTGAAGGAGCTGTATAAAAATG
GCTTTAAAAACATCTACGAGTTCATTAACTACGATAAATATTATCAGACCGACAAAGCGC
TGGCGACCACCATTGATGGCTTCTCGCGCTATGCCAACGAACTGAAACCGGATCTGATCG
TGGTGCACGGCGATCGCATTGAACCGCTGGCAGCGGCGATTGTCGGCGCGCTGAATAATA
TCCTGGTGGCGCACATCGAAGGCGGCGAGATTTCCGGCACCATCGACGATAGCCTCCGCC
ACGCCATCAGCAAGCTCGCGCATATTCATCTGGTTAACGATGAATTTGCCAAACGCCGCC
TGATGCAGCTGGGCGAAGATGAGAAAAGCATTTTTATTATTGGCTCGCCGGACCTGGAAC
TGCTGAACGACAATAAAATCTCCCTGAGCGAAGCGAAGAAATACTACGACATCAATTACG
AAAACTACGCCCTGTTGATGTTCCATCCGGTGACGACCGAAATCACCAGCATCAAGAATC
AGGCGGATAACCTGGTCAAAGCCCTGATTCAGTCGAACAAAAACTATATTGTGATTTATC
CGAACAATGATCTCGGTTTTGAATTGATTCTGCAAAGCTATGAAGAATTCAAAAATAACC
CGCGCTTTAAGCTGTTCCCGAGCCTGCGCTTCGAGTATTTCATCACGCTGCTCAAGAACG
CCGATTTTATCATCGGCAACAGCTCCTGCATTCTGAAAGAGGCGCTGTACCTGAAAACCG
CGGGCATTCTGGTGGGCAGCCGCCAGAACGGCCGCCTCGGCAATGAAAATACCCTGAAGG
TGAACGCGAACTCCGACGAAATTCTCAAAGCAATCAACACCATCCATAAAAAACAGGATT
TGTTCAGCGCGAAACTGGAGATCCTCGACAGCAGTAAACTCTTCTTTGAATATCTGCAGA
GCGGCGACTTCTTCAAACTGTCCACCCAGAAAGTGTTCAAGGACATCAAGTGATAAATTG
CCGTTCCGACTGATTGGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAG
ACTGGGGCGTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGTCCTTATGTGGTG
```

Figure C.6: Detailed overview of pPathway (pPathway$_{medium}$ as an example) with the p14 promoter[107], the T7 ribosome binding site (RBS)[310], the neuB1 coding DNA sequence (CDS)[309,311], the designed intergenic region[306], the neuC CDS[309,311] and, rrnB T1 terminator[24].

A

$$Fru\text{-}6\text{-}P \xrightarrow{glmS} GlcN\text{-}6\text{-}P \xrightarrow{glmM} GlcN\text{-}1\text{-}P \xrightarrow{glmU} \textbf{UDP-GlcNAc}$$

central carbon metabolism

GlcNAc-6-P ⬅✖ ManNAc-6-P ⬅✖ **ManNAc**

nanE   *nanK*   *neuC*

nagA

PEP   pyruvate

neuB1   *nanA*

Neu5Ac

B

p14   T7 RBS   NeuB1   intergenic RBS   NeuC   rnpB T1
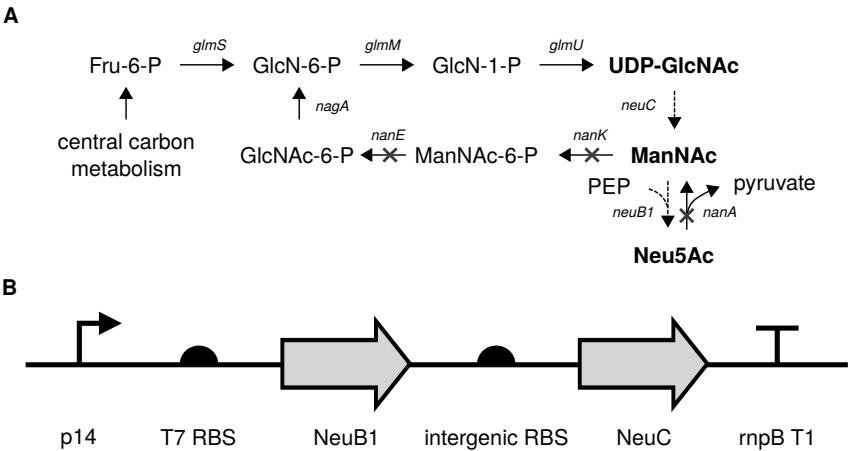
Figure C.7: Schematic overview of the *N*-acetylneuraminic acid (Neu5Ac) pathway. A) An overview of the introduced metabolic pathway (in bold) for the production of Neu5Ac and the gene deletions performed to prevent Neu5Ac degradation. B) The genetic organisation of the genes *neuB1* and *neuC*, expressed on the pPathway$_{medium}$ in a single operon, which was controlled by the promoter p14[107]. The translation of the NeuB1 and the NeuC protein is controlled by the T7 RBS and the intergenic RBS, respectively.
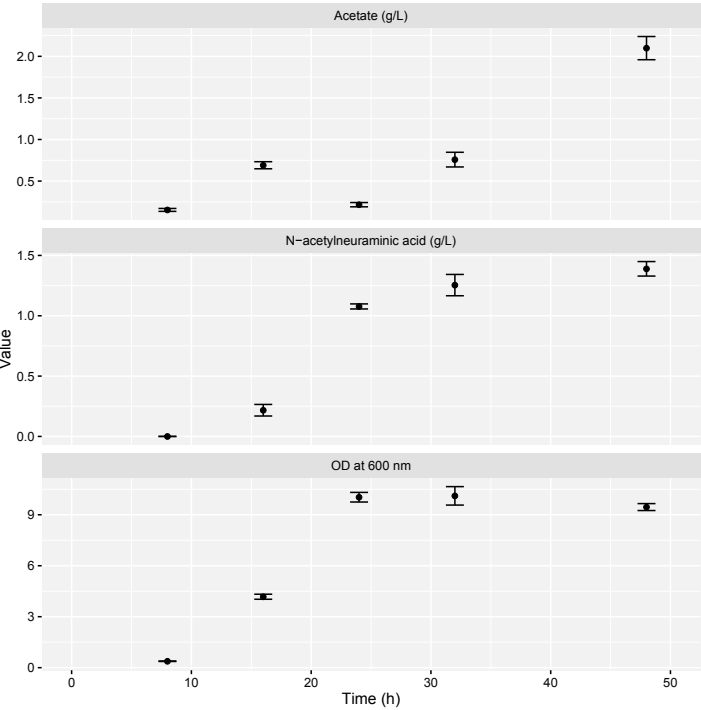
Figure C.8: Plot of optical density (OD) and product (acetic acid and *N*-acetylneuraminic acid (Neu5Ac)) concentration over time during 24-well DWP cultivation of *E. coli* K-12 MG1655 Δ*nanATEK* containing the pPathway$_{medium}$ plasmid, which used defined medium with glucose as sole carbon source. The pPathway$_{medium}$ plasmid encodes the biosynthetic pathway from UDP-*N*-acetylglucosamine (UDP-GlcNAc) to Neu5Ac.
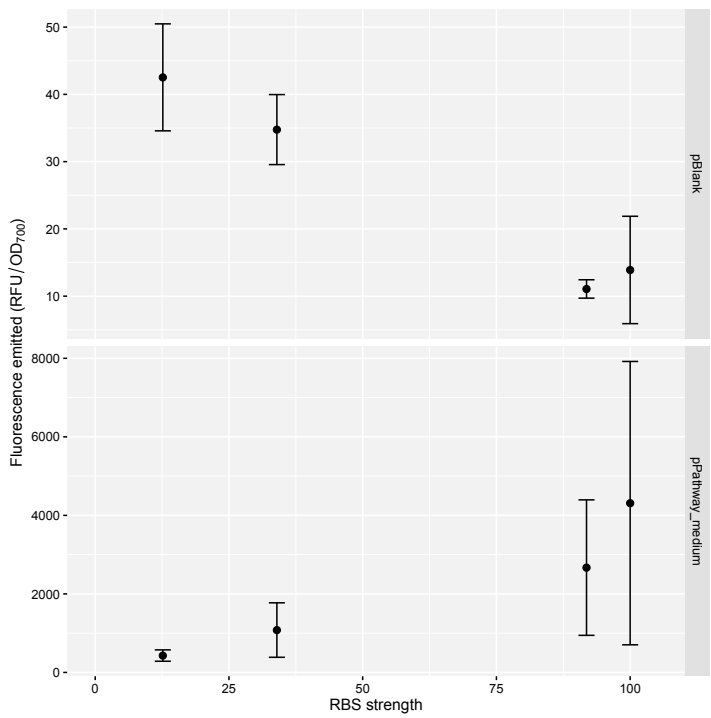
Figure C.9: Plot of the fluorescence emitted by the pEB$_2$ variants in the absence (pBlank) and the presence (pPathway$_{medium}$) of Neu5Ac against the previously reported strength of the RBS used for NanR expression[308]. Error bars represent standard deviation (n=3).
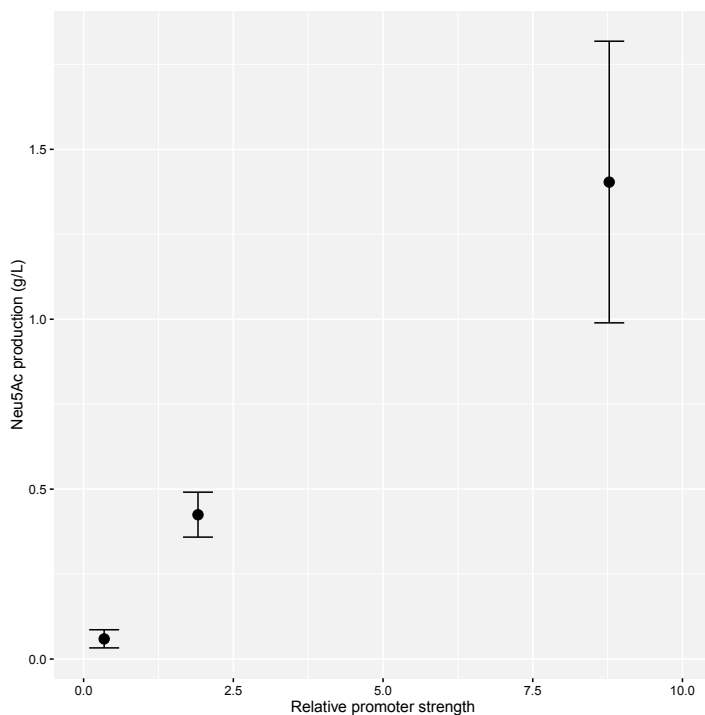
Figure C.10: Plot of the amount of extracellular *N*-acetylneuraminic acid (Neu5Ac) produced (in shake flask) against the relative strength of the promoter used to control the Neu5Ac producing operon. The pPathway plasmids used are pPathway$_{low}$, pPathway$_{medium}$, and pPathway$_{high}$, which contain the promoters p14, pFAB46* and pJ23105, respectively. These promoters have a relative strength of 1.91, 8.77, and 0.34, respectively. Error bars represent standard deviation (n=3).
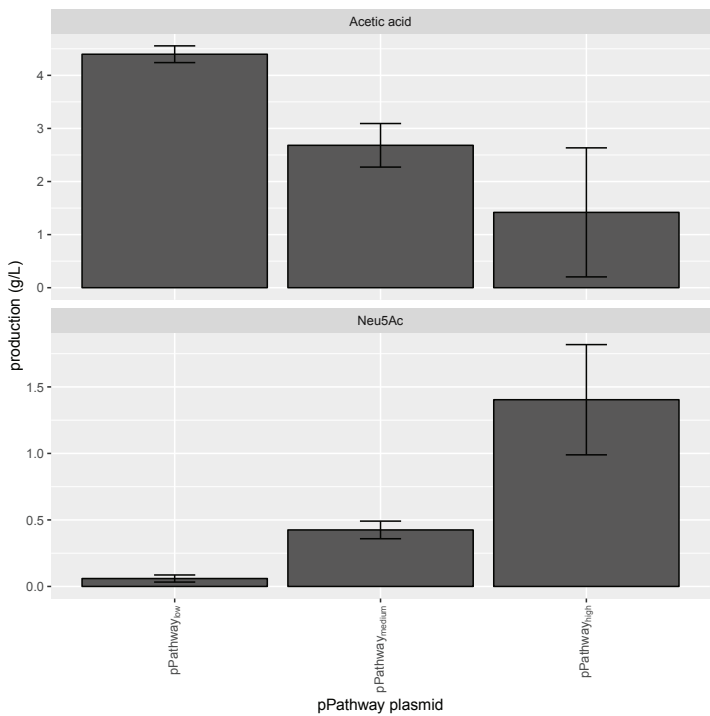
Figure C.11: Bar plot of the *N*-acetylneuraminic acid (Neu5Ac) and acetic acid production of the three constructed pPathway plasmids (pPathway$_{low}$, pPathway$_{medium}$, and pPathway$_{high}$) in *E. coli* K-12 MG1655 Δ*nanATEKR*. Neu5Ac and acetic acid were the only detectable metabolites produced. Error bars represent standard deviation (n=3).

C.2   Supplementary Tables

Table C.1: Overview of all primers used for gene deletions

Name	Sequence (5'→3')
Fw_nanATEK_homology	TAATGCGCCGCCAGTAAATCAACATGAAATGCCGCTGGCTCCGTGGCAGTGTAGGCTG GAGCTGCTTC
Rv_nanATEK_homology	CCAACAACAAGCACTGGATAAAGCGAGTCTGCGTCGCCTGGTTCAGTTCACATATGAA TATCCTCCTTAG
Fw_nanATEK_control	GTCGCCCTGTAATTCGTAAC
Rv_nanATEK_control	CTTTCGGTCAGACCACCAAC
Fw_nanR_homology	CAACACCCCATACAGAAAGCTTATAATGCGATCTGCTTCACTAAAGTGGCAGTGTAGG CTGGAGCTGCTTC
Rv_nanR_homology	CAATCCTGTGATAGGATGTCACTGATGATGTTAATCACACTGACCTTACAGACATATGA ATATCCTCCTTAG
Fw_nanR_control	CGATGCCCTGCTGAATATTG
Rv_nanR_control	TTTATGGTGCGGATGTCGTG
Fw_nanRATEK_homology	TAATGCGCCGCCAGTAAATCAACATGAAATGCCGCTGGCTCCGTGTAGGCTGGAGCTG CTTC
Rv_nanRATEK_homology	CAATCCTGTGATAGGATGTCACTGATGATGTTAATCACACTGACCTTACAGACATATGA ATATCCTCCTTAG
Fw_nanRATEK_control	CGATGCCCTGCTGAATATTG
Rv_nanRATEK_control	TTTATGGTGCGGATGTCGTG

Table C.2: Overview of all plasmids used in this study

Name	Content (5'→3')	Backbone
pBiosensors		
pNB_1	pNanR - nanR - pNanA + NanA UTR - $nanA_{13AA}$ - $mKate2$ - rnpB T1	p15A
pNB_2	(BBa_J23114 - BBa_B0031 - $nanR$ - rnpB T1)$_{reversed}$ - pNanA + NanA UTR - $nanA_{13AA}$ - $mKate2$ - rnpB T1	p15A
pNB_3	(BBa_J23114 - BBa_B0031 - $nanR$ - rnpB T1)$_{reversed}$ - pYjhB + YjhB UTR - $yjhB_{13AA}$ - $mKate2$ - rnpB T1	p15A
pNB_4	(BBa_J23114 - BBa_B0031 - $nanR$ - rnpB T1)$_{reversed}$ - pNanC + NanC UTR - $nanC_{13AA}$ - $mKate2$ - rnpB T1	p15A
pEB_1	(BBa_J23114 - BBa_B0031 - $nanR$ - rnpB T1)$_{reversed}$ - pNanA - BBa_B0034 - $mKate2$ - rnpB T1	p15A
pEB_2	(BBa_J23114 - BBa_B0031 - $nanR$ - rnpB T1)$_{reversed}$ - pJ23119$_{H10}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
$pEB_{2,1}$	(BBa_J23114 - BBa_B0032 - $nanR$ - rnpB T1)$_{reversed}$ - pJ23119$_{H10}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
$pEB_{2,2}$	(BBa_J23114 - BBa_B0030 - $nanR$ - rnpB T1)$_{reversed}$ - pJ23119$_{H10}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
$pEB_{2,3}$	(BBa_J23114 - BBa_B0034 - $nanR$ - rnpB T1)$_{reversed}$ - pJ23119$_{H10}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
pEB_3	(BBa_J23114 - BBa_B0031 - $nanR$ - rnpB T1)$_{reversed}$ - pJ23119$_{H35,1}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
pEB_4	(BBa_J23114 - BBa_B0031 - $nanR$ - rnpB T1)$_{reversed}$ - pJ23119$_{H35,2}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
pOperators		
pO_1	pNanA + NanA UTR - $nanA_{13AA}$ - $mKate2$ - rnpB T1	p15A
pO_2	pYjhB + YjhB UTR - $yjhB_{13AA}$ - $mKate2$ - rnpB T1	p15A
pO_3	pNanC + NanC UTR - $nanC_{13AA}$ - $mKate2$ - rnpB T1	p15A
pO_5	pNanA - BBa_B0034 - $mKate2$ - rnpB T1	p15A
pO_6	pJ23119$_{H10}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
pO_7	pJ23119$_{H35,1}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
pO_8	pJ23119$_{H35,2}$ - BBa_B0034 - $mKate2$ - rnpB T1	p15A
Other		
$pPathway_{low}$	pJ23105 - T7 RBS - $neuB1$ - intergenic RBS - $neuC$ - rrnB T1	pBR322
$pPathway_{medium}$	p14 - T7 RBS - $neuB1$ - intergenic RBS - $neuC$ - rrnB T1	pBR322
$pPathway_{high}$	pFAB46* - T7 RBS - $neuB1$ - intergenic RBS - $neuC$ - rrnB T1	pBR322

Table C.3: Overview of all regulatory DNA sequences

Name	Sequence (5'→3')	Reference
BBa_B0030	ATTAAAGAGGAGAAA	308
BBa_B0031	TCACACAGGAAACC	308
BBa_B0032	TCACACAGGAAAG	308
BBa_B0034	AAAGAGGAGAAA	308
BBa_J23114	TTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGC	308
p14	CTTCATTCTATAAGTTTCTTGACATCTTGGCCGGCATATGGTATAATAGGG	107
pJ23119$_{H10}$	GATTCGTTACCAATTGACAGCTACTGGTATAACAGGTATAAAGGTAGC	This study
pJ23119$_{H35,1}$	GATTCGTTCTGGTATAACAGGTATAAAGGTACTAGGTATAATGCTAGC	This study
pJ23119$_{H35,2}$	GATTCGTTCTGGTATGACAGGTATAAAGGTACTAGGTATAATGCTAGC	This study
pNanA	AGCTTTCTGTATGGGGTGTTGCTTAATTGATCTGGTATAACAGGTATAAAGGTATATCGTTTA	301
pNanA + UTR	TGCCACTTTAGTGAAGCAGATCGCATTATAAGCTTTCTGTATGGGGTGTTGCTTAATTGATCTGGTATAACAGGTATAA AGGTATATCGTTTATCGACAAGCATCACTTCAGAGGTATTT	301
pNanC + UTR	CGTTTTTCCCTTATAATTACAGACGCGCACTAGCTGTGCTGGGTTAGCAATAATCCAACATTTTTATCCTGATTATGTTT ATAAAAGCGAACGTTTGCTTAATAACTAACGTAAGTGGACCAGTTCTTCTGAGTGAACTTAAATGGAGTAGCAACTGTT AATTATAATAACGTTGCCATACAAACGCCATGCTCTTGCATGCTATGTACCTTTATATATTTATCAATCGGATACGAATC TAATAATAACCCTCATCATTAATGCAACTAATCTTATCTATATATCATGTGATATGTTCTGTAACAAGTAAAATCAACATA AAAAATTCAGCAAAATTTAATATTCGTTCAACTAGCCGTCTCACATATGACATTATTAGTCAACTCCATTTGCCACAAAA TGGCTATTTCATGGAGCGCAGTAATTTCTTAGTGGCTATTATCATGCTAATTCTTAGGTCCCTAGCGATTATTCCTGCTG ATTCGTCAGCTTACGCCATGAGATTGTTACATTCCTTTTTATCCTAGAGATATGGGGGAACGCAGGAGATGATTTTTTCA TCTACATTCGTTGGTAAATATAGCTTTGGTCACTACGACTTAAACCTTCCTCAATCACAGCCATTCGCCGGGCGGATACGG CAAACAAACTAAGTGAAACATCATCCTGAAAACACCAACATCAACAAGCCTCTCCAGATCGACTTCAGAAGTGACCAGT TACAAGCCACAAACCAGGAAACATATTTCATGATGAGAATTATGCTCAGCCTGATGGCGGGTAAACACTGATTTTCACC AAGCTCACTGAGTTAGCTTGCAAAGCTCCCCTTGTTTAGACTCTTAATAAAATTATTTATAAAACAAATAGTTAAACTAAG ATCTTTAGTTTTTGATGACCACCGCAAGTGTTCGTCTGGCTTCACATGGCATCTTCCTCTTAGAAAAAGATCGACATATT TTGTGACACGAATTGCAAATCTGGTTTTGTTGTATGGATTGCGTGATTTTTGATCTGGTATAACAGGTATAAAGGTGCA CCAAGATAGTCAATGAGACAGGGCATCTCGCAATCTATGGCAAACATCACTTCAGTTCTTTCTCATCGGGTGATGAAAA CGCACTTCAGTCTGAAAGGAATATGAAAATGAGATCAACAGACATTCTATTTTATGACTCTGGGTAAAATGGATTGAGT AAGTGATATAGCTTACGAACATTCAAATCAATTAAACATCAGAAGAGATTTTATACTCAGGTATTTAATCTGGATCTCTG TTTATTTAAATAATGTGAAAAGAGATTTTTCACAGGAGACCTTATACAAAAAAATATAAAATACAGCTACCGGTTGCCAA AGACACTATAAGCCTGGCAAAAAAAATATTACACAACATAAATGCTAATTGTTTATGCGGGCTTTGTATTGCTTTCTGTAT CCTACAAATGAGTGAAATTT	301
pNanR + UTR	TGTTGAACTCCGTGTCAAAAGAAAACGGTCAATCCCATAAACGGCAGATTGAAAACAACGATGTTATATTTTTTGCAAG GCTATTTATGGTGCGGATGTCGTGTTTTTAATTGTAGGTGAGGTGATTTTTCATTAAAAAATATGCGCTTATGATTATTT TGTAAGAACACATTCATAATATTCATAATGCTCGTGAATAGTCTTATAAATAATTCAAACGGGATGTTTTTTATCTGCGTT ACATTAATTTTTCGCAATAGTTAATTATTCCGTTAATTATGGTAATGATGAGGCACAAAGAGAAAACCCTGCCATTTTCC CCTACTTTCAATCCTGTGATAGGATGTCACTGATGATGTTAATCACACTGACCTTACAGA	301
pYjhB + UTR	CCGGAAACAAGTGAGCGTTTCCGGATTCTTACACAGCCACTTGATCGGTCAACTGATCCTTAACTGATCGGCATTAATC TTGGTTCTGGTGTTTGTAACAAACTATCAGCTACAAAAATATGCTCAATTTGTGACATCAGTACAAAACGCGTTTTGTT ATGTGGATTGCTTGTTTTTTGATCTGGTATAACAGGTATAAAGGTATACAGAAAAGCAAAGAAATACTGCAAAGGAAAA CAGCTATAACGTAAGCTAAAGTAATAACCTCTCAGTCTTTCCTCATTTGACGAAGGGAGTTTTATTCAACCTGAACGGAC TACGAAAATGAGCACAATGAATAAGTCAATTTTTGAAGCAGGGTTGAACTGTGCCAGTGGTGGAATATCTGGCGAACAT TGTTTTTTGTTGGTGACCCAAAACTGTAGGTGGGTATCAGTTTTATCTTTCATAGAGTGAAATATGTTAAGAAGAAATG GAGGAAAAGATTGTACTGATTAGGTATTGATAACAATCAATAGTACTGGCGTATTTGAAGACAATATAATTATTTCTGG ATATTGTTGAGGCTCCCTAATATTTACTTTAAGGGCTATATTAGAATAACACAGGAAACAAAT	301
rrnB T1	GGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTG AACGCTCTCCTG	24
rnpB T1	TCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGACT GTC	24
T7 RBS	CCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATA	310
intergenic region	GAATTCGTTTAGAGCTCTAAATAAGGAGGAATAACC	306

Table C.4: Overview of all coding DNA sequences used. The NeuB1 and NeuC CDSs were designed for expression in *E. coli* K-12 MG1655 using the COOL algorithm[311].

Name	Sequence (5'→3')	Ref.
nanA$_{13AA}$	ATGGCAACGAATTTACGTGGCGTAATGGCTGCACTCCTG	301
nanC$_{13AA}$	ATGAAAAAGGCTAAAATACTTTCTGGCGTATTATTACTG	301
nanR	ATGGGCCTTATGAACGCATTTGATTCGCAAACCGAAGATTCTTCACCTGCAATTGGTCGCAACTTGCGTAGCCGCCCGCTGGCGCG TAAAAAACTCTCCGAAATGGTGGAAGAAGAGCTGGAACAGATGATCCGCCGTCGTGAATTTGGCGAAGGTGAACAATTACCGTCT GAACGCGAACTGATGGCGTTCTTTAACGTCGGGCGTCCTTCGGTGCGTGAAAGCGCTGGCAGCGTTAAAACGCAAAGGTCTGGTGC AAATAAACAACGGCGAACGCGCTCGCGTCTCGCGTCCTTCTGCGGACACTATCATCGGTGAGCTTTCCGGCATGGCGAAAGATTTC CTTTCTCATCCCGGTGGGATTGCCCATTTCGAACAATTACGTCTGTTCTTTGAATCCAGTCTGGTGCGCTATGCGGCTGAACATGCC ACCGATGAGCAAATCGATTTGCTGGCAAAAGCACTGGAAATCAACAGTCAGTCGCTGGATAACAACGCGGCATTCATTCGTTCAGA CGTTGATTTCCACCGCGTGCTGGCGGAGATCCCCGGTAACCCAATCTTCATGGCGATCCACGTTGCCCTGCTCGACTGGCTTATTG CCGCACGCCCAACGGTTACCGATCAGGCACTGCACGAACATAACAACGTTAGTTATCAACAGCATATTGCGATCGTTGATGCGATC CGCCGTCATGATCCTGACGAAGCCGATCGTGCGTTGCAATCGCATCTCAACAGCGTCTCTGCTACCTGGCACGCTTTCGGTCAGAC CACCCAACAAAAAGAAATAA	301
neuB1	ATGAAAGAGATTAAGATCCAGAATATTTATCATCAGCGAAGAGAAAGCGCCGCTGGTGGTGCCGGAAATTGGCATTAACCACAACG GCAGCCTGGAACTGGCTAAGATTATGGTGGATGCCGCGTTCAGCGCCGGTGCGAAGATCATTAAACATCAGACGGCATATCGTTGA AGATGAGATGAGCAAGGCGGCGAAGAAGGTGATTCCTGGCAACGCCAAGATTAGCATCTATGAGATCATGCAGAAATGCGCGCTT GATTATAAAGATGAACTGGCGGCTGAAAGAATATACCGAGAAGTTAGGTCTGGTCTATCTGTCGACGCCATTCTCGCGCGCAGGTGC CAACCGTCTGGAAGATATGGGCGTGTCTGCCTTCAAGATTGGTTCCGGTGAATGTAATAATTATCCACTGATCAAGCATATTGCCG CATTCAAGAAGCCGATGATTGTCAGCACCGGCATGAACAGCATTGAATCTATCAAACCGACCGTTAAGATTCTGCTGGATAATGAG ATTCCGTTCGTTCTGATGCACACCACCAATCTGTATCCGACGCCGCATAACCTGGTTCGCCTGAACGCGATGCTGGAGCTGAAGAA GGAGTTCTCCTGTATGGTTGGCCTGAGCGATCATACCACCGATAACCTCGCCTGTCTTGGCGCGGTGGTTCTCGGCGCATGCGTGC TTGAACGTCACTTCACCGACAGCATGCATCGCAGCGGTCCGGATATCGTCTGCTCGATGGATACCAAGGCACTGAAGGAACTGATT ATTCAGAGCGAGCAGATGGCGATTATTCGCGGCAATAACGAATCCAAGAAGGCCGCCAAGCAGGAACAGGTGACCATCGACTTCG CGTTCGCTTCGGTGGTCAGTATTAAGGACATCAAGAAAGGCGAAGTGCTGTCAATGGACAACATCTGGGTGAAGCGTCCAGGCTT AGGCGGCATCAGTGCGGCAGAATTCGAGAACATTCTCGGTAAGAAGGCTCTGCGCGGATATTGAGAATGATGCGCAGCTGAGCTAT GAAGACTTCGCCTGA	309,311
neuC	ATGGTGAAGAAGATCCTGTTCATTACCGGCTCCCGCGCCGACTACAGCAAAATTAAATCGCTGATGTATCGCGTGCAGAATAGCAG CGAGTTTGAGCTCTATATCTTCGCCACCGGGATGCACCTGTCGAAAAACTTCGGCTACACCGTGAAGGAGCTGTATAAAAATGGCT TTAAAAACATCTACGAGTTCATTAACTACGATAAATATTATCAGACCGACAAAGCGCTGGCGACCACCATTGATGGCTTCTCGCGCT ATGCCAACGAACTGAAACCGGATCTGATCGTGGTGCACGGCGATCGCATTGAACCGCTGGCAGCGGCGATTGTCGGCGCGCTGAA TAATATCCTGGTGGCGCACATCGAAGGCGGCGAGATTTCCGGCACCATCGACGATAGCCTCCGCCACGCCATCAGCAAGCTCGCG CATATTCATCTGGTTAACGATGAATTTGCCAAACGCCGCCTGATGCAGCTGGGGCGAAGATGAGAAAAGCATTTTTATTATTGGCTC GCCGGACCTGGAACTGCTGAACGACAATAAAAATCTCCCTGAGCGAAGCGAAGAAATACTACGACATCAATTACGAAAACTACGCCC TGTTGATGTTCCATCCGGTGACGACCGAAATCACCAAGCATCAAGAATCAGGCGGATAACCTGGTCAAAGCCCTGATTCAGTCGAAC AAAAACTATATTGTGATTTATCCGAACAATGATCTCGGTTTTGAATTGAATCTCGCAAAGCTATGAAGAATTCAAAAATAACCCGCGC TTTAAGCTGTTCCCGAGCCTGCGCTTCGAGTATTTCATCACGCTGCTCAAGAACGCCGATTTTATCATCGGCAACAGCTCCTGCATT CTGAAAGAGGCGCTGTACCTGAAAACCGCGGGCATTCTGGTGGGCAGCCGCCAGAACGGCCGCCTCGGCAATGAAAATACCCTGA AGGTGAACGCGAACTCCGACGAAATTCTCAAAGCAATCAACACCATCCATAAAAAACAGGATTTGTTCAGCGCGAAACTGGAGATC CTCGACAGCAGTAAACTCTTCTTTGAATATCTGCAGAGCGGCGACTTCTTCAAACTGTCCACCCAGAAAGTGTTCAAGGACATCAA GTGA	309,311
mKate2	ATGGTTAGCGAGCTGATCAAAGAAAACATGCACATGAAAGCTGATCAAAGAAAACATGCACATGAAACTGTATATGGAAGGCACCG TGAATAACCACCACTTTAAATGTACCAGCGAAGGTGAAAGCTGATCAAAGAAAACATGCACATGAAACTGTATATGGAAGGCACCG TGAATAACCACCACTTTAAATGTACCAGCGAAGGTGAAGGTAAACCGTATGAAGGCACCCAGACCATGCGTATTAAAGCAGTTGAA GGTGGTCCGCTGCCGTTTGCATTTGATATTCTGGCAACCAGCTTTATGTATGGCAGCAAAACCTTTATTAACCATACCCAGGGTATC CCGGATTTTTTCAAACAGAGCTTTCCGGAAGGTTTTACCTGGGAACGTGTTACCACCTATGAAGATGGTGGTGTTCTGACCGCAAC CCAGGATACCAGTCTGCAGGATGGTGGTGTCTGATTTATAATGTGAAAATTCGCGGTGTGAACCTTTCCGAGCAATGGTCCGGTTATGC AGAAAAAAACCCTGGGTTGGGAAGCAAGCACCGAAACCCTGTATCCGGCAGATGGTGGTCTGGAAGGTCGTGCAGATATGGCACT GAAACTGGTTGGTGGTGGTGGTCATCTGATTTGCAATCTGAAAAACCACCTATCGTAGCAAAAAACCGGCAAAAAAATCTGAAAATGCCTG GCGTGTATTATGTTGATCGTCGTCTGGAACGTATTAAAGAGGCAGATAAAGAAACCTATGTGGAACAGCATGAAGTTGCAGTTGCA CGTTATTGTGATCTGCCCGAGCAAACTGGGTCACCGCTGATAA	252
yjhB$_{13AA}$	ATGGCAACAGCATGGTATAAACAAGTTAATCCACCACAA	301

Bibliography

[1] Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–32 (2013).

[2] Yim, H. *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chemical Biology* **7**, 445–452 (2011).

[3] Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).

[4] Pirie, C. M., De Mey, M., Prather, K. L. J. & Ajikumar, P. K. Integrating the protein and metabolic engineering toolkits for next-generation chemical biosynthesis. *ACS Chemical Biology* **8**, 662–672 (2013).

[5] Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74 (2010).

[6] Nielsen, J. *et al.* Engineering synergy in biotechnology. *Nature Chemical Biology* **10**, 319–322 (2014).

[7] Agapakis, C. M. Designing synthetic biology. *ACS Synthetic biology* **3**, 121–128 (2013).

[8] Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–8 (2014).

[9] Gibson, D. G. Programming biological operating systems: genome design, assembly and activation. *Nature Methods* **11**, 521–526 (2014).

[10] Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–6 (2010).

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

[12] Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Molecular Cell* **43**, 880–891 (2011).

Bibliography

[13] Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136**, 615–628 (2009).

[14] Papenfort, K. & Vogel, J. Regulatory RNA in bacterial pathogens. *Cell Host Microbe* **8**, 116–127 (2010).

[15] Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17–24 (2013).

[16] Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).

[17] Lucks, J. B., Qi, L., Whitaker, W. R. & Arkin, A. P. Toward scalable parts families for predictable design of biological circuits. *Current Opinion in Microbiology* **11**, 567–573 (2008).

[18] Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **110**, 14024–14029 (2013).

[19] Cheng, A. A. & Lu, T. K. Synthetic biology: an emerging engineering discipline. *Annual Review of Biomedical Engineering* **14**, 155–178 (2012).

[20] Rodrigo, G. & Jaramillo, A. AutoBioCAD: full biodesign automation of genetic circuits. *ACS synthetic biology* **2**, 230–236 (2012).

[21] Clancy, K. & Voigt, C. A. Programming cells: towards an automated Genetic Compiler. *Current Opinion in Biotechnology* **21**, 572–581 (2010).

[22] Registry of Standard Biological Parts (2014). URL http://parts.igem.org/.

[23] Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods* **10**, 354–60 (2013).

[24] Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Research* **41**, 5139–5148 (2013).

[25] Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* **27**, 946–950 (2009).

[26] Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods* **10**, 659–664 (2013).

[27] Choi, S., Song, C. W., Shin, J. H. & Lee, S. Y. Biorefineries for the production of top building block chemicals and their derivatives. *Metabolic Engineering* (2015).

[28] Biggs, B. W., De Paepe, B., Santos, C. N. S., De Mey, M. & Ajikumar, P. K. Multivariate modular metabolic engineering for pathway and strain optimization. *Current Opinion in Biotechnology* **29**, 156–162 (2014).

[29] Jones, J. A., Toparlak, Ö. D. & Koffas, M. A. Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Current Opinion in Biotechnology* **33**, 52–59 (2015).

[30] Jullesson, D., David, F., Pfleger, B. & Nielsen, J. Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology advances* **33**, 1395–1402 (2015).

[31] Benner, S. A. & Sismour, A. M. Synthetic biology. *Nature Reviews. Genetics* **6**, 533–543 (2005).

[32] Voigt, C. A. Genetic parts to program bacteria. *Current Opinion in Biotechnology* **17**, 548–557 (2006).

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

[34] Purnick, P. E. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature Reviews. Molecular Cell Biology* **10**, 410–422 (2009).

[35] Mukherji, S. & van Oudenaarden, A. Synthetic biology: understanding biological design from synthetic circuits. *Nature Reviews. Genetics* **10**, 859–871 (2009).

[36] Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nature Reviews. Genetics* **11**, 367–379 (2010).

[37] Woolston, B. M., Edgar, S. & Stephanopoulos, G. Metabolic engineering: past and future. *Annual review of chemical and biomolecular engineering* **4**, 259–288 (2013).

[38] Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. *Nature Biotechnology* **24**, 545–554 (2006).

[39] Davidson, E. A. & Ellington, A. D. Synthetic RNA circuits. *Nature Chemical Biology* **3**, 23–28 (2006).

[40] Benenson, Y. Synthetic biology with RNA: progress report. *Current Opinion in Chemical Biology* **16**, 278–284 (2012).

[41] Kang, Z. *et al.* Small RNA regulators in bacteria: powerful tools for metabolic engineering and synthetic biology. *Applied Microbiology and Biotechnology* **98**, 3413–3424 (2014).

[42] Chappell, J. *et al.* The centrality of RNA for engineering gene expression. *Biotechnology Journal* **8**, 1379–1395 (2013).

[43] Qi, L. S. & Arkin, A. P. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nature Reviews. Microbiology* **12**, 341–354 (2014).

[44] Win, M. N., Liang, J. C. & Smolke, C. D. Frameworks for programming biological function through RNA parts and devices. *Chemistry & Biology* **16**, 298–310 (2009).

[45] Wittmann, A. & Suess, B. Engineered riboswitches: expanding researchers toolbox with synthetic RNA regulators. *FEBS Letters* **586**, 2076–2083 (2012).

[46] Chang, A. L., Wolf, J. J. & Smolke, C. D. Synthetic RNA switches as a tool for temporal and spatial control over gene expression. *Current Opinion in Biotechnology* **23**, 679–688 (2012).

[47] Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D. & Arkin, A. P. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences* **108**, 8617–8622 (2011).

[48] Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).

[49] Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B. & Arkin, A. P. Rationally designed families of orthogonal RNA regulators of translation. *Nature Chemical Biology* **8**, 447–454 (2012).

[50] Liang, J. C., Bloom, R. J. & Smolke, C. D. Engineering biological systems with synthetic RNA molecules. *Molecular Cell* **43**, 915–926 (2011).

[51] Farasat, I. *et al.* Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular systems biology* **10**, 731 (2014).

[52] Pardee, K. *et al.* Paper-based synthetic gene networks. *Cell* **159**, 940–954 (2014).

[53] Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160**, 339–350 (2015).

[54] Yang, J. *et al.* Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. *Nature Communications* **4**, 1413 (2013).

[55] Carothers, J. M. & Stevens, J. T. Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. *ACS Synthetic biology* (2014).

[56] Michener, J. K. & Smolke, C. D. High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metabolic Engineering* **14**, 306–316 (2012).

[57] Brewster, R. C., Jones, D. L. & Phillips, R. Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. *PLoS Computational Biology* **8**, e1002811 (2012).

[58] de Smit, M. H. & Van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences* **87**, 7668–7672 (1990).

[59] Cruz, J. A. & Westhof, E. The dynamic landscapes of RNA architecture. *Cell* **136**, 604–609 (2009).

[60] Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).

[61] Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nature Chemical Biology* **9**, 18–20 (2013).

[62] Shapiro, B. A., Yingling, Y. G., Kasprzak, W. & Bindewald, E. Bridging the gap in RNA structure prediction. *Current Opinion in Structural Biology* **17**, 157–165 (2007).

[63] Lai, D., Proctor, J. R. & Meyer, I. M. On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461–1473 (2013).

[64] Bida, J. & Das, R. Squaring theory with practice in RNA design. *Current Opinion in Structural Biology* **22**, 457–466 (2012).

[65] Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

[66] Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F. & Mörl, M. *De novo* design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research* **41**, 2541–2551 (2013).

[67] Rodrigo, G., Landrain, T. E., Majer, E., Daròs, J.-A. & Jaramillo, A. Full design automation of multi-state RNA devices to program gene expression using energy-based optimization. *PLoS Computational Biology* **9**, e1003172 (2013).

[68] Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).

[69] Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols* **4**, 206–223 (2009).

[70] Na, D. *et al.* Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nature Biotechnology* **31**, 170–174 (2013).

[71] Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12678–12683 (2005).

Bibliography

[72] Lee, J. Y. *et al.* Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in *Escherichia coli*. *Nucleic Acids Research* **36**, e102–e102 (2008).

[73] Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Research* gks404 (2012).

[74] Politz, M. C., Copeland, M. F. & Pfleger, B. F. Artificial repressors for controlling gene expression in bacteria. *Chemical Communications* **49**, 4325–4327 (2013).

[75] Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature Chemical Biology* **10**, 99–105 (2014).

[76] Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).

[77] Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* **41**, 7429–7437 (2013).

[78] Yang, Y., Lin, Y., Li, L., Linhardt, R. J. & Yan, Y. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. *Metabolic Engineering* **29**, 217–226 (2015).

[79] Liu, C. C. *et al.* An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nature Methods* **9**, 1088–1094 (2012).

[80] Chappell, J., Takahashi, M. K. & Lucks, J. B. Creating small transcription activating RNAs. *Nature Chemical Biology* **11**, 214–220 (2015).

[81] Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: de-novo-designed regulators of gene expression. *Cell* (2014).

[82] Shao, Y., Wu, Y., Chan, C. Y., McDonough, K. & Ding, Y. Rational design and rapid screening of antisense oligonucleotides for prokaryotic gene modulation. *Nucleic Acids Research* **34**, 5660–5669 (2006).

[83] Stefan, A., Schwarz, F., Bressanin, D. & Hochkoeppler, A. Shine-Dalgarno sequence enhances the efficiency of *lacZ* repression by artificial anti-*lac* antisense RNAs in *Escherichia coli*. *Journal of Bioscience and Bioengineering* **110**, 523–528 (2010).

[84] Nakashima, N. & Tamura, T. Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Research* **37**, e103–e103 (2009).

[85] Solomon, K. V., Sanders, T. M. & Prather, K. L. A dynamic metabolite valve for the control of central carbon metabolism. *Metabolic Engineering* **14**, 661–671 (2012).

[86] Kim, J. Y. & Cha, H. J. Down-regulation of acetate pathway through antisense strategy in *Escherichia coli*: Improved foreign protein production. *Biotechnology and Bioengineering* **83**, 841–853 (2003).

[87] De Lay, N., Schu, D. J. & Gottesman, S. Bacterial Small RNA-based Negative Regulation: Hfq and its Accomplices. *Journal of Biological Chemistry* (2013).

[88] Man, S. *et al.* Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Research* **39**, e50–e50 (2011).

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

[90] Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).

[91] Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

[92] Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233–239 (2013).

[93] Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

[94] Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology* **31**, 397–405 (2013).

[95] Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **8**, 2180–2196 (2013).

[96] Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).

[97] Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synthetic biology* **2**, 604–613 (2013).

[98] Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).

[99] Cress, B. F. *et al.* CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in *E. coli*. *ACS Synthetic biology* (2015).

[100] Lv, L., Ren, Y.-L., Chen, J.-C., Wu, Q. & Chen, G.-Q. Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P (3HB-co-4HB) biosynthesis. *Metabolic Engineering* **29**, 160–168 (2015).

Bibliography

[101] Mojica, F., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).

[102] Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).

[103] Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).

[104] Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116–1121 (2013).

[105] Yadav, V. G., De Mey, M., Giaw Lim, C., Kumaran Ajikumar, P. & Stephanopoulos, G. The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metabolic Engineering* **14**, 233–241 (2012).

[106] Pfleger, B. F., Pitera, D. J., Smolke, C. D. & Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology* **24**, 1027–1032 (2006).

[107] De Mey, M., Maertens, J., Lequeux, G. J., Soetaert, W. K. & Vandamme, E. J. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnology* **7**, 34 (2007).

[108] De Mey, M. *et al.* Promoter knock-in: a novel rational method for the fine tuning of genes. *BMC Biotechnology* **10**, 26 (2010).

[109] Brewster, R. C. *et al.* The transcription factor titration effect dictates level of gene expression. *Cell* **156**, 1312–1323 (2014).

[110] Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3-end chronicles. *Journal of Molecular Biology* **412**, 793–813 (2011).

[111] Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).

[112] Studer, S. M. & Joseph, S. Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Molecular Cell* **22**, 105–115 (2006).

[113] Na, D., Lee, S. & Lee, D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC systems biology* **4**, 71 (2010).

[114] Seo, S. W. *et al.* Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metabolic Engineering* **15**, 67–74 (2013).

[115] Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research* **42**, 2646–2659 (2014).

[116] Na, D. & Lee, D. RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics* **26**, 2633–2634 (2010).

[117] Coussement, P., Maertens, J., Beauprez, J., Van Bellegem, W. & De Mey, M. One step DNA assembly for combinatorial metabolic engineering. *Metabolic Engineering* **23**, 70–77 (2014).

[118] Oliver, J. W., Machado, I. M., Yoneda, H. & Atsumi, S. Combinatorial optimization of cyanobacterial 2, 3-butanediol production. *Metabolic Engineering* **22**, 76–82 (2014).

[119] Lin, Z. *et al.* Metabolic engineering of *Escherichia coli* for the production of riboflavin. *Microbial cell factories* **13**, 104 (2014).

[120] Smanski, M. J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nature Biotechnology* **32**, 12411249 (2014).

[121] Ng, C. Y., Farasat, I., Maranas, C. D. & Salis, H. M. Rational design of a synthetic Entner–Doudoroff pathway for improved and controllable NADPH regeneration. *Metabolic engineering* **29**, 86–96 (2015).

[122] Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A. & Arkin, A. P. RNA processing enables predictable programming of gene expression. *Nature Biotechnology* **30**, 1002–1006 (2012).

[123] Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. & Voigt, C. A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nature Biotechnology* **30**, 1137–1142 (2012).

[124] Mutalik, V. K. *et al.* Quantitative estimation of activity and quality for collections of functional genetic elements. *Nature Methods* **10**, 347–353 (2013).

[125] Carrier, T. A. & Keasling, J. Engineering mRNA stability in *E. coli* by the addition of synthetic hairpins using a 5 cassette system. *Biotechnology and Bioengineering* **55**, 577–580 (1997).

[126] Carrier, T. A. & Keasling, J. Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnology Progress* **15**, 58–64 (1999).

[127] Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology* **22**, 841–847 (2004).

[128] Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R. & Collins, J. J. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences* **107**, 15898–15903 (2010).

[129] Callura, J. M., Cantor, C. R. & Collins, J. J. Genetic switchboard for synthetic biology applications. *Proceedings of the National Academy of Sciences* **109**, 5850–5855 (2012).

[130] Gallagher, R. R., Patel, J. R., Interiano, A. L., Rovner, A. J. & Isaacs, F. J. Multilayered genetic safeguards limit growth of microorganisms to defined environments. *Nucleic Acids Research* gku1378 (2015).

[131] Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Molecular Systems Biology* **9** (2013).

[132] Rodrigo, G., Landrain, T. E. & Jaramillo, A. De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proceedings of the National Academy of Sciences* **109**, 15271–15276 (2012).

[133] Klauser, B. & Hartig, J. S. An engineered small RNA-mediated genetic switch based on a ribozyme expression platform. *Nucleic Acids Research* **41**, 5542–5552 (2013).

[134] Shen, S. *et al.* Dynamic signal processing by ribozyme-mediated RNA circuits to control gene expression. *Nucleic Acids Research* **43**, 5158–5170 (2015).

[135] Takahashi, M. K. & Lucks, J. B. A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucleic Acids Research* **41**, 7577–7588 (2013).

[136] Kiani, S. *et al.* CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nature Methods* **11**, 723–726 (2014).

[137] 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 (2014).

[138] Liu, Y. *et al.* Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. *Nature Communications* **5** (2014).

[139] Nielsen, A. A. & Voigt, C. A. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular Systems Biology* **10** (2014).

[140] Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical/wires/'. *Nature* **469**, 212–215 (2011).

[141] Brophy, J. A. & Voigt, C. A. Principles of genetic circuit design. *Nature methods* **11**, 508–520 (2014).

[142] Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Reviews. Genetics* **8**, 776–790 (2007).

[143] Marchisio, M. A. & Stelling, J. Automatic design of digital synthetic gene circuits. *PLoS Comput. Biol* **7**, e1001083 (2011).

[144] Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300–304 (2006).

[145] Hussein, R. & Lim, H. N. Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Research* **40**, 7269–7279 (2012).

[146] Briner, A. E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular Cell* **56**, 333–339 (2014).

[147] Didovyk, A., Borek, B., Hasty, J. & Tsimring, L. Orthogonal modular gene repression in *E. coli* using engineered CRISPR/Cas9. *ACS Synthetic biology* (2015).

[148] Liu, Y. *et al.* Modular pathway engineering of *Bacillus subtilis* for improved N-acetylglucosamine production. *Metabolic Engineering* **23**, 42–52 (2014).

[149] Bayer, T. S. & Smolke, C. D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology* **23**, 337–343 (2005).

[150] Beisel, C. L., Bayer, T. S., Hoff, K. G. & Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Molecular Systems Biology* **4** (2008).

[151] Tu, K. C., Long, T., Svenningsen, S. L., Wingreen, N. S. & Bassler, B. L. Negative feedback loops involving small regulatory RNAs precisely control the *Vibrio harveyi* quorum-sensing response. *Molecular Cell* **37**, 567–579 (2010).

[152] Svenningsen, S. L., Waters, C. M. & Bassler, B. L. A negative feedback loop involving small RNAs accelerates *Vibrio cholerae*s transition out of quorum-sensing mode. *Genes and Development* **22**, 226–238 (2008).

[153] Beisel, C. L. & Storz, G. The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Molecular Cell* **41**, 286–297 (2011).

[154] Kang, T. *et al.* Reverse engineering validation using a benchmark synthetic gene circuit in human cells. *ACS synthetic biology* **2**, 255–262 (2013).

[155] Hu, C. Y., Varner, J. D. & Lucks, J. B. Generating effective models and parameters for RNA genetic circuits. *bioRxiv* 018358 (2015).

[156] Schallmey, M., Frunzke, J., Eggeling, L. & Marienhagen, J. Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. *Current Opinion in Biotechnology* **26**, 148–154 (2014).

[157] Eggeling, L., Bott, M. & Marienhagen, J. Novel screening methodsbiosensors. *Current Opinion in Biotechnology* **35**, 30–36 (2015).

[158] Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology* **4**, a003566 (2012).

[159] Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: structures and mechanisms. *Cold Spring Harbor perspectives in biology* **3**, a003533 (2011).

[160] Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nature Reviews. Molecular Cell Biology* **5**, 451–463 (2004).

[161] Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology* **15**, 342–348 (2005).

[162] Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. *Annual Review of Microbiology* **59**, 487–517 (2005).

[163] Suess, B. & Weigand, J. E. Engineered riboswitches. *RNA Biol.* **5**, 24–29 (2008).

[164] Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

[165] Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

[166] Lemke, E. A. & Schultz, C. Principles for designing fluorescent sensors and reporters. *Nature Chemical Biology* **7**, 480–483 (2011).

[167] Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX - a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **24**, 381–403 (2007).

[168] Janssen, K. P., Knez, K., Spasic, D. & Lammertyn, J. Nucleic acids for ultra-sensitive protein detection. *Sensors* **13**, 1353–1384 (2013).

[169] Vanbrabant, J., Leirs, K., Vanschoenbeek, K., Lammertyn, J. & Michiels, L. reMelting curve analysis as a tool for enrichment monitoring in the SELEX process. *Analyst* **139**, 589–595 (2014).

[170] Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA interactions. *Science* **282**, 296–298 (1998).

[171] Bauer, G. & Suess, B. Engineered riboswitches as novel tools in molecular biology. *Journal of Biotechnology* **124**, 4–11 (2006).

[172] Grate, D. & Wilson, C. Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer–ligand complex. *Bioorganic and Medicinal Chemistry* **9**, 2565–2570 (2001).

[173] Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E. & Suess, B. Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular Microbiology* **49**, 1627–1637 (2003).

[174] Suess, B. *et al.* Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Research* **31**, 1853–1858 (2003).

[175] Harvey, I., Garneau, P. & Pelletier, J. Inhibition of translation by RNA-small molecule interactions. *RNA* **8**, 452–463 (2002).

[176] Buskirk, A. R., Landrigan, A. & Liu, D. R. Engineering a ligand-dependent RNA transcriptional activator. *Chemistry & Biology* **11**, 1157–1163 (2004).

[177] Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *Journal of the American Chemical Society* **126**, 13247–13254 (2004).

[178] Lynch, S. A., Desai, S. K., Sajja, H. K. & Gallivan, J. P. A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chemistry & Biology* **14**, 173–184 (2007).

[179] Topp, S. & Gallivan, J. P. Random walks to synthetic riboswitches - A high-throughput selection based on cell motility. *Chembiochem* **9**, 210–213 (2008).

[180] Lynch, S. A. & Gallivan, J. P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Research* **37**, 184–192 (2009).

[181] Nomura, Y. & Yokobayashi, Y. Reengineering a natural riboswitch by dual genetic selection. *Journal of the American Chemical Society* **129**, 13814–13815 (2007).

[182] Muranaka, N., Sharma, V., Nomura, Y. & Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in *Escherichia coli*. *Nucleic Acids Research* **37**, e39–e39 (2009).

[183] Weigand, J. E. *et al.* Screening for engineered neomycin riboswitches that control translation initiation. *RNA* **14**, 89–97 (2008).

[184] Buskirk, A. R., Kehayova, P. D., Landrigan, A. & Liu, D. R. *In vivo* evolution of an RNA-based transcriptional activator. *Chemistry & Biology* **10**, 533–540 (2003).

[185] Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425–1429 (1994).

[186] Sharma, V., Nomura, Y. & Yokobayashi, Y. Engineering complex riboswitch regulation by dual genetic selection. *Journal of the American Chemical Society* **130**, 16310–16315 (2008).

[187] Dixon, N. *et al.* Reengineering orthogonally selective riboswitches. *Proceedings of the National Academy of Sciences* **107**, 2830–2835 (2010).

[188] Dixon, N. *et al.* Orthogonal riboswitches for tuneable coexpression in bacteria. *Angewandte Chemie International Edition* **51**, 3620–3624 (2012).

[189] Robinson, C. J. *et al.* Modular riboswitch toolsets for synthetic genetic control in diverse bacterial species. *Journal of the American Chemical Society* **136**, 10615–10624 (2014).

[190] Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* (2014).

[191] Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression *in vivo*. *Nucleic Acids Research* **32**, 1610–1614 (2004).

[192] Beisel, C. L. & Smolke, C. D. Design principles for riboswitch function. *PLoS Computational Biology* **5**, e1000363 (2009).

[193] Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences* **104**, 14283–14288 (2007).

[194] Babiskin, A. H. & Smolke, C. D. Engineering ligand-responsive RNA controllers in yeast through the assembly of RNase III tuning modules. *Nucleic Acids Research* **39**, 5299–5311 (2011).

[195] Mishler, D. M. & Gallivan, J. P. A family of synthetic riboswitches adopts a kinetic trapping mechanism. *Nucleic Acids Research* **42**, 6753–6761 (2014).

[196] Ogawa, A. Rational design of artificial riboswitches based on ligand-dependent modulation of internal ribosome entry in wheat germ extract and their applications as label-free biosensors. *RNA* **17**, 478–488 (2011).

[197] Qi, L., Lucks, J. B., Liu, C. C., Mutalik, V. K. & Arkin, A. P. Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic Acids Research* **40**, 5775–5786 (2012).

[198] Dotu, I. *et al.* Complete RNA inverse folding: computational design of functional hammerhead ribozymes. *Nucleic Acids Research* **42**, 11752–11762 (2014).

[199] Dietrich, J. A., McKee, A. E. & Keasling, J. D. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annual Review of Biochemistry* **79**, 563–590 (2010).

[200] Eckdahl, T. T. *et al.* Programmed Evolution for Optimization of Orthogonal Metabolic Output in Bacteria. *PLoS ONE* **10**, e0118322 (2015).

[201] Lee, S.-W. & Oh, M.-K. A synthetic suicide riboswitch for the high-throughput screening of metabolite production in *Saccharomyces cerevisiae*. *Metabolic Engineering* **28**, 143 – 150 (2015).

[202] Xu, P., Li, L., Zhang, F., Stephanopoulos, G. & Koffas, M. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proceedings of the National Academy of Sciences* **111**, 11299–11304 (2014).

[203] Carothers, J. M., Goler, J. A., Juminaga, D. & Keasling, J. D. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* **334**, 1716–1719 (2011).

[204] Zhou, L.-B. & Zeng, A.-P. Exploring lysine riboswitch for metabolic flux control and improvement of L-lysine synthesis in Corynebacterium glutamicum. *ACS Synthetic biology* (2015).

[205] Chudakov, D. M., Matz, M. V, Lukyanov, S. & Lukyanov, K. A. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews* **90**, 1103–1163 (2010).

[206] Song, W., Strack, R. L. & Jaffrey, S. R. Imaging bacterial protein expression using genetically encoded RNA sensors. *Nature Methods* **10**, 873–875 (2013).

[207] Passaris, I. *et al.* Isolation and Validation of an Endogenous Fluorescent Nucleoid Reporter in *Salmonella Typhimurium*. *PloS one* **9** (2014).

[208] Meech, S. R. Excited state reactions in fluorescent proteins. *Chemical Society Reviews* **38**, 2922–2934 (2009).

[209] Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333**, 642–646 (2011).

[210] Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311 (2011).

[211] Wroblewska, L. *et al.* Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nature Biotechnology* (2015).

[212] Lapique, N. & Benenson, Y. Digital switching in a biosensor circuit via programmable timing of gene availability. *Nature Chemical Biology* (2014).

[213] Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nature methods* **10**, 1219–1224 (2013).

Bibliography

[214] Song, W., Strack, R. L., Svensen, N. & Jaffrey, S. R. Plug-and-play fluorophores extend the spectral properties of Spinach. *Journal of the American Chemical Society* **136**, 1198–1201 (2014).

[215] Rogers, T. A., Andrews, G. E., Jaeger, L. & Grabow, W. W. Fluorescent monitoring of RNA assembly and processing using the split-Spinach aptamer. *ACS synthetic biology* **4**, 162–166 (2014).

[216] Paige, J. S., Nguyen-Duc, T., Song, W. & Jaffrey, S. R. Fluorescence imaging of cellular metabolites with RNA. *Science* **335**, 1194–1194 (2012).

[217] Nakayama, S., Luo, Y., Zhou, J., Dayie, T. K. & Sintim, H. O. Nanomolar fluorescent detection of c-di-GMP using a modular aptamer strategy. *Chemical Communications* **48**, 9059–9061 (2012).

[218] Kellenberger, C. A., Wilson, S. C., Sales-Lee, J. & Hammond, M. C. RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *Journal of the American Chemical Society* **135**, 4906–4909 (2013).

[219] Strack, R. L. & Jaffrey, S. R. New approaches for sensing metabolites and proteins in live cells using RNA. *Current Opinion in Chemical Biology* **17**, 651–655 (2013).

[220] Strack, R. L., Song, W. & Jaffrey, S. R. Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria. *Nature Protocols* **9**, 146–155 (2014).

[221] You, M., Litke, J. L. & Jaffrey, S. R. Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proceedings of the National Academy of Sciences* **112**, E2756–E2765 (2015).

[222] Kellenberger, C. A., Chen, C., Whiteley, A. T., Portnoy, D. A. & Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic di-AMP. *Journal of the American Chemical Society* (2015).

[223] Leisner, M., Bleris, L., Lohmueller, J., Xie, Z. & Benenson, Y. Rationally designed logic integration of regulatory signals in mammalian cells. *Nat. Nanotechnol.* **5**, 666–670 (2010).

[224] Benenson, Y. RNA-based computation in live cells. *Current Opinion in Biotechnology* **20**, 471–478 (2009).

[225] Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnol.* **25**, 795–801 (2007).

[226] Klumpp, S., Scott, M., Pedersen, S. & Hwa, T. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences* **110**, 16754–16759 (2013).

[227] Myhrvold, C. & Silver, P. A. Using synthetic RNAs as scaffolds and regulators. *Nature structural & molecular biology* **22**, 8–10 (2015).

[228] Lee, H., DeLoache, W. C. & Dueber, J. E. Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering* **14**, 242–251 (2012).

[229] Savage, D. F., Afonso, B., Chen, A. H. & Silver, P. A. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* **327**, 1258–1261 (2010).

[230] Agapakis, C. M., Boyle, P. M. & Silver, P. A. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nature Chemical Biology* **8**, 527–535 (2012).

[231] Srere, P. A. The metabolon. *Trends in Biochemical Sciences* **10**, 109–110 (1985).

[232] Conrado, R. J. *et al.* DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Research* **40**, 1879–1889 (2012).

[233] Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **27**, 753–759 (2009).

[234] Zhai, Y. *et al.* NahK/GlmU fusion enzyme: characterization and one-step enzymatic synthesis of UDP-*N*-acetylglucosamine. *Biotechnology Letters* **34**, 1321–1326 (2012).

[235] Bunka, D. H. & Stockley, P. G. Aptamers come of age - at last. *Nature Reviews. Microbiology* **4**, 588–596 (2006).

[236] Delebecque, C. J., Silver, P. A. & Lindner, A. B. Designing and using RNA scaffolds to assemble proteins *in vivo*. *Nature Protocols* **7**, 1797–1807 (2012).

[237] Sachdeva, G., Garg, A., Godding, D., Way, J. C. & Silver, P. A. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic Acids Research* **42**, 9493–9503 (2014).

[238] Geraldi, A., Bui, L. M. & C, K. S. A novel RNA scaffold system for the enhancement of the *in vivo* solubilization of recombinant proteins in *Escherichia coli*. *The Sixth International Meeting on Synthetic Biology (Synthetic Biology 6.0)* (2013).

[239] Cho, S. H., Haning, K. & Contreras, L. M. Strain engineering via regulatory noncoding RNAs: not a one-blueprint-fits-all. *Current Opinion in Chemical Engineering* **10**, 25–34 (2015).

[240] Cleto, S., Jensen, J. K., Wendisch, V. F. & Lu, T. K. Corynebacterium glutamicum metabolic engineering with CRISPR interference (CRISPRi). *ACS Synthetic Biology* (2016).

[241] Peters, G., Coussement, P., Maertens, J., Lammertyn, J. & De Mey, M. Putting RNA to work: Translating RNA fundamentals into biotechnological engineering practice. *Biotechnology advances* **33**, 1829–1844 (2015).

Bibliography

[242] Hoynes-OConnor, A. & Moon, T. S. Development of design rules for reliable antisense RNA behavior in E. coli. *ACS Synthetic Biology* (2016).

[243] Shachrai, I., Zaslaver, A., Alon, U. & Dekel, E. Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth. *Molecular cell* **38**, 758–767 (2010).

[244] Gorochowski, T. E., van den Berg, E., Kerkman, R., Roubos, J. A. & Bovenberg, R. A. Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of Escherichia coli. *ACS synthetic biology* **3**, 129–139 (2013).

[245] Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature methods* **12**, 415–418 (2015).

[246] Takahashi, M. K. *et al.* Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription–Translation (TX-TL) Systems. *ACS Synthetic biology* (2014).

[247] Nakashima, N., Tamura, T. & Good, L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia coli. *Nucleic acids research* **34**, e138–e138 (2006).

[248] Johnson, E. & Srivastava, R. Volatility in mRNA secondary structure as a design principle for antisense. *Nucleic Acids Research* **41**, e43–e43 (2013).

[249] Lee, T. S. *et al.* BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *Journal of biological engineering* **5**, 1 (2011).

[250] Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research* gkq810 (2010).

[251] Lerner, C. G. & Inouye, M. Low copy number plasmids for regulated low-level expression of cloned genes in Escherichia coli with blue/white insert screening capability. *Nucleic acids research* **18**, 4631 (1990).

[252] Shcherbo, D. *et al.* Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J* **418**, 567–574 (2009).

[253] Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3**, e3647 (2008).

[254] Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PloS one* **4**, e6441 (2009).

[255] Hecht, A., Endy, D., Salit, M. L. & Munson, M. S. When Wavelengths Collide: Bias in Cell Abundance Measurements due to Expressed Fluorescent Proteins. *ACS synthetic biology* (2016).

[256] Hofacker, I. L. *et al.* Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie/Chemical Monthly* **125**, 167–188 (1994).

[257] Bernhart, S. H. *et al.* Partition function and base pairing probabilities of RNA heterodimers. *Algorithms for Molecular Biology* **1**, 3 (2006).

[258] Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 1 (2011).

[259] Wuchty, S., Fontana, W., Hofacker, I. L., Schuster, P. *et al.* Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**, 145–165 (1999).

[260] Mückstein, U. *et al.* Thermodynamics of RNA–RNA binding. *Bioinformatics* **22**, 1177–1182 (2006).

[261] Grönmping, U. R package FrF2 for creating and analyzing fractional factorial 2-level designs. *Journal of Statistical Software* **56**, 1–56 (2014).

[262] Mevik, B.-H., Wehrens, R. *et al.* The pls package: principal component and partial least squares regression in R. *Journal of Statistical software* **18**, 1–24 (2007).

[263] Dayal, B., MacGregor, J. F. *et al.* Improved PLS algorithms. *Journal of chemometrics* **11**, 73–85 (1997).

[264] Laursen, B. S., Sørensen, H. P, Mortensen, K. K. & Sperling-Petersen, H. U. Initiation of protein synthesis in bacteria. *Microbiology and Molecular Biology Reviews* **69**, 101–123 (2005).

[265] Strobel, E. J., Watters, K. E., Loughrey, D. & Lucks, J. B. RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Current opinion in biotechnology* **39**, 182–191 (2016).

[266] Kleckner, N. Regulating Tn10 and IS10 transposition. *Genetics* **124**, 449 (1990).

[267] Meyer, S., Chappell, J., Sankar, S., Chew, R. & Lucks, J. B. Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. *Biotechnology and bioengineering* **113**, 216–225 (2016).

[268] de Smit, M. H. & van Duin, J. Control of Translation by mRNA Secondary Structure in *Escherichia coli*: A Quantitative Analysis of Literature Data. *Journal of Molecular Biology* **244**, 144–150 (1994).

[269] Jonsson, J., Norberg, T., Carlsson, L., Gustafsson, C. & Wold, S. Quantitative sequence-activity models (QSAM)tools for sequence design. *Nucleic acids research* **21**, 733–739 (1993).

[270] Bonde, M. T. *et al.* Predictable tuning of protein expression in bacteria. *Nature methods* (2016).

Bibliography

[271] Watters, K. E., Abbott, T. R. & Lucks, J. B. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic acids research* **44**, e12–e12 (2016).

[272] Takahashi, M. K. *et al.* Using in-cell SHAPE-Seq and simulations to probe structure–function design principles of RNA transcriptional regulators. *RNA* **22**, 920–933 (2016).

[273] Kim, T., Folcher, M., Baba, M. D.-E. & Fussenegger, M. A Synthetic Erectile Optogenetic Stimulator Enabling Blue-Light-Inducible Penile Erection. *Angewandte Chemie International Edition* **54**, 5933–5938 (2015).

[274] Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature biotechnology* **34**, 78–83 (2016).

[275] Bacchus, W., Aubel, D. & Fussenegger, M. Biomedically relevant circuit-design strategies in mammalian synthetic biology. *Molecular systems biology* **9**, 691 (2013).

[276] Meinhardt, S. *et al.* Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic acids research* **40**, 11139–11154 (2012).

[277] Taylor, N. D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nature methods* **13**, 177–183 (2016).

[278] Berens, C., Groher, F. & Suess, B. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnology journal* **10**, 246–257 (2015).

[279] Borujeni, A. E., Mishler, D. M., Wang, J., Huso, W. & Salis, H. M. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic acids research* gkv1289 (2015).

[280] Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P. *et al.* Optimization by simmulated annealing. *science* **220**, 671–680 (1983).

[281] Koch, A. L. The metabolism of methylpurines by *Escherichia coli* I. Tracer studies. *Journal of Biological Chemistry* **219**, 181–188 (1956).

[282] Bolivar, F. *et al.* Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. *Gene* **2**, 95–113 (1977).

[283] Topp, S. *et al.* Synthetic riboswitches that induce gene expression in diverse bacterial species. *Applied and environmental microbiology* **76**, 7881–7884 (2010).

[284] Watters, K. E., Strobel, E. J., Angela, M. Y., Lis, J. T. & Lucks, J. B. Cotranscriptional folding of a riboswitch at nucleotide resolution. Tech. rep., Nature Research (2016).

[285] Philp, J. C., Ritchie, R. J. & Allan, J. E. Biobased chemicals: the convergence of green chemistry with industrial biotechnology. *Trends in biotechnology* **31**, 219–222 (2013).

[286] Chen, Y. & Nielsen, J. Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. *Current opinion in biotechnology* **24**, 965–972 (2013).

[287] Mahr, R. *et al.* Biosensor-driven adaptive laboratory evolution of l-valine production in Corynebacterium glutamicum. *Metabolic engineering* **32**, 184–194 (2015).

[288] Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Current opinion in biotechnology* **42**, 84–91 (2016).

[289] De Paepe, B., Peters, G., Coussement, P., Maertens, J. & De Mey, M. Tailor-made transcriptional biosensors for optimizing microbial cell factories. *Journal of Industrial Microbiology & Biotechnology* 1–23 (2016).

[290] Kelly, C. L. *et al.* Synthetic chemical inducers and genetic decoupling enable orthogonal control of the rhaBAD promoter. *ACS synthetic biology* **10**, 1136–1145 (2016).

[291] Shis, D. L., Hussain, F., Meinhardt, S., Swint-Kruse, L. & Bennett, M. R. Modular, multi-input transcriptional logic gating with orthogonal LacI/GalR family chimeras. *ACS synthetic biology* **3**, 645–651 (2014).

[292] Varki, N. M. & Varki, A. Diversity in cell surface sialic acid presentations: implications for biology and disease. *Laboratory investigation* **87**, 851–857 (2007).

[293] Büll, C., Stoel, M. A., den Brok, M. H. & Adema, G. J. Sialic acids sweeten a tumor's life. *Cancer research* **74**, 3199–3204 (2014).

[294] Carlson, S. E. & House, S. G. Oral and intraperitoneal administration of N-acetylneuraminic acid: effect on rat cerebral and cerebellar N-acetylneuraminic acid. *The Journal of nutrition* **116**, 881–886 (1986).

[295] Wang, B. Sialic acid is an essential nutrient for brain development and cognition. *Annual review of nutrition* **29**, 177–222 (2009).

[296] von Itzstein, M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418–423 (1993).

[297] Dreitlein, W. B., Maratos, J. & Brocavich, J. Zanamivir and oseltamivir: two new options for the treatment and prevention of influenza. *Clinical therapeutics* **23**, 327–355 (2001).

[298] Bondioli, L. *et al.* Sialic acid as a potential approach for the protection and targeting of nanocarriers. *Expert opinion on drug delivery* **8**, 921–937 (2011).

Bibliography

[299] Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).

[300] Wang, B. *et al.* Dietary sialic acid supplementation improves learning and memory in piglets. *The American journal of clinical nutrition* **85**, 561–569 (2007).

[301] Keseler, I. M. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research* **41**, D605–D612 (2013).

[302] Novichkov, P. S. *et al.* RegPrecise 3.0–a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC genomics* **14**, 1 (2013).

[303] Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. Diversity of microbial sialic acid metabolism. *Microbiology and molecular biology reviews* **68**, 132–153 (2004).

[304] Kalivoda, K. A., Steenbergen, S. M. & Vimr, E. R. Control of the Escherichia coli sialoregulon by transcriptional repressor NanR. *Journal of bacteriology* **195**, 4689–4701 (2013).

[305] Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**, 6640–6645 (2000).

[306] Biggs, B. W. *et al.* Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 201515826 (2016).

[307] Kalivoda, K. A., Steenbergen, S. M., Vimr, E. R. & Plumbridge, J. Regulation of sialic acid catabolism by the DNA binding protein NanR in Escherichia coli. *Journal of bacteriology* **185**, 4806–4815 (2003).

[308] Registry of Standard Biological Parts. `http://parts.igem.org/` (2016).

[309] Fierfort, N. & Samain, E. Genetic engineering of Escherichia coli for the economical production of sialylated oligosaccharides. *Journal of biotechnology* **134**, 261–265 (2008).

[310] Tabor, S. Expression using the T7 RNA polymerase/promoter system. *Current protocols in molecular biology* 16–2 (1990).

[311] Chin, J. X., Chung, B. K.-S. & Lee, D.-Y. Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics* **30**, 2210–2212 (2014).

[312] Condemine, G., Berrier, C., Plumbridge, J. & Ghazi, A. Function and expression of an N-acetylneuraminic acid-inducible outer membrane channel in Escherichia coli. *Journal of bacteriology* **187**, 1959–1965 (2005).

[313] Severi, E. *et al.* Sialic acid mutarotation is catalyzed by the Escherichia coli β-propeller protein YjhT. *Journal of Biological Chemistry* **283**, 4841–4849 (2008).

[314] Steenbergen, S. M., Jirik, J. L. & Vimr, E. R. YjhS (NanS) is required for Escherichia coli to grow on 9-O-acetylated N-acetylneuraminic acid. *Journal of bacteriology* **191**, 7134–7139 (2009).

[315] Deutscher, J. The mechanisms of carbon catabolite repression in bacteria. *Current opinion in microbiology* **11**, 87–93 (2008).

[316] Kang, J. *et al.* Engineering of an N-acetylneuraminic acid synthetic pathway in Escherichia coli. *Metabolic engineering* **14**, 623–629 (2012).

[317] Domin, G. *et al.* Applicability of a computational design approach for synthetic riboswitches. *Nucleic Acids Research* gkw1267 (2016).

[318] Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing non-canonical RNA structure. *Nature methods* **7**, 291–294 (2010).

[319] Cruz, J. A. *et al.* RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *Rna* **18**, 610–625 (2012).

[320] Lang, P. T. *et al.* DOCK 6: Combining techniques to model RNA–small molecule complexes. *Rna* **15**, 1219–1230 (2009).

[321] Chushak, Y. & Stone, M. O. In silico selection of RNA aptamers. *Nucleic acids research* gkp408 (2009).

[322] Sousa, F. L. *et al.* AlloRep: A repository of sequence, structural and mutagenesis data for the LacI/GalR transcription regulators. *Journal of molecular biology* **428**, 671–678 (2016).

[323] Zhou, H., Vonk, B., Roubos, J. A., Bovenberg, R. A. & Voigt, C. A. Algorithmic co-optimization of genetic constructs and growth conditions: application to 6-ACA, a potential nylon-6 precursor. *Nucleic acids research* gkv1071 (2015).

[324] Xu, P., Rizzoni, E. A., Sul, S.-Y. & Stephanopoulos, G. Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synthetic Biology* (2016).

[325] Giegerich, R., Voß, B. & Rehmsmeier, M. Abstract shapes of RNA. *Nucleic acids research* **32**, 4843–4851 (2004).

[326] Huthoff, H. & Berkhout, B. Two alternating structures of the HIV-1 leader RNA. *RNA* **7**, 143–157 (2001).

[327] Evers, D. & Giegerich, R. RNA movies: visualizing RNA secondary structure spaces. *Bioinformatics* **15**, 32–37 (1999).

Summary

Summary

Spurred by environmental motives and the limited supply of fossil fuels, chemical in-dustry increasingly relies on microbial engineering to create economically feasible pro-duction processes from renewable sources. Several successful metabolic engineering efforts were driven by enabling technologies from synthetic biology, which allow the development of new microbial cell factories for the production of a molecule of inter-est. One such interesting molecule is *N*-acetylneuraminic acid (Neu5Ac), a sugar moiety with a vital role in various physiological processes, such as, tumor progression, bacterial infection, infant brain development and, immune responses, which results in various applications in pharmaceutical and food industry. However, the development of these applications is hindered by the limited availability of Neu5Ac due to the lack of suffi-cient production technologies. This shortage could be solved by metabolic engineering which allow *ad hoc* rewiring the metabolism of microbes to obtain an optimized biosyn-thetic pathway for the with maximal productivity. This is a daunting task that requires various tools to modulate gene expression, build genetic circuitery, specifically detect molecules throughout the cell. Therefore, the main objective of this PhD research was the development of advanced enabling technologies for metabolic engineering, allowing the construction of genetic circuitry and detection of small molecules.

Recently, the programmable nature of RNA spurred the development of various novel tools based on RNA regulators, which are increasingly employed for various metabolic engineering strategies to maximize productivity. One interesting type of RNA devices to build complex biological systems are riboregulators, which allow rapid control of translation without the need of coexpressed burdensome proteins. However, these de-vices are limited by the lack of clear design principles, hindering their applicability in metabolic engineering. To address this problem, so called translation inhibiting RNAs (tiRNAs) were developed, which are riboregulators that allow programmable control of protein expression on a post-transcriptional level. These tiRNA devices were created by exploring possibly important features using a design of experiments (DOE). The *de novo* developed riboregulators repressed translation up to 6 % of the original protein expression levels, outperforming the dynamic range of previously described riboregu-lators. Moreover, compared to previous efforts, the tiRNA regulators created here are designed from scratch and do not require any naturally occurring chassis to function. To link the properties of tiRNA riboregulators to its performance, a partial least squares (PLS) regression model was constructed, further increasing the programmability of the riboregulator.

Besides gene expression modulation, RNA technology also allows controlling gene ex-

pression based on the presence of small molecules. Recently, various ligand responsive RNA devices, more specifically riboswitches, were previously used in various metabolic engineering strategies as an attractive alternative to their traditional protein counterparts. However, the creation of these riboswitches from *in vitro* selected aptamers typically involves laborious high-throughput screening efforts. To remove this hurdle in metabolic engineering, a computer-aided design approach was developed, allowing the *in silico* screening of riboswitches. To quantify the riboswitch capacities of a specific untranslated region (UTR), an objective function was defined based on previously described riboswitches. Using this objective function, 29 potential riboswitches were computationally designed using a simulated annealing algorithm. Subsequently, these riboswitches were evaluated *in vivo*, yielding functional riboswitches out of the box with 12 out of the 29 created riboswitches activating gene expression more than five fold. However, despite the high probability of yielding functional riboswitches, linking performance to structural or thermodynamic properties remains challenging. Overall, the developed algorithm can help reducing the development times of translational riboswitches, improving their applicability.

The natural complex regulation of the microbial metabolism spurred the development of various metabolic engineering strategies, which often require intracellular detection of small molecules. To this end, various biosensors were created based on naturally occurring transcription factors (TFs), typically having limited possibilities to engineer the desired response curve. As a proof of concept, novel biosensors were created that respond to Neu5Ac based on native and engineered promoters that interact with the TF NanR, which were evaluated using a engineered Neu5Ac producing strain. To allow modular biosensor optimization, a NanR binding site was inserted in a constitutive promoter, which resulted in biosensors composed of defined parts. This enables more reliable engineering of the response curve, further expanding the applicability in metabolic engineering. The increased engineering capabilities by the modular design of biosensors was shown by modulating the response of one of the created biosensors by solely changing the ribosome binding site (RBS) used for NanR expression. Also, when exposed to varying Neu5Ac production levels (up to 1.4 ± 0.4 g/L extracellular Neu5Ac produced) three biosensors emit fluorescence proportional to amount of Neu5Ac produced. This indicates the broad operating range of these biosensors, a critical property of biosensors for various applications in metabolic engineering. Overall, the range of biosensors was further expanded with various functional biosensors capable of detecting Neu5Ac, which can be applied in various metabolic engineering approaches to produce Neu5Ac

with maximal productivity.

Overall, various tools were developed in this doctoral research that enable the reliable optimization of microbial cell factories. Specifically, the forward engineering capacity of translation inhibiting riboregulators and translational riboswitches was improved, which further improves the applicability of the various tools originating from the field of RNA synthetic biology. Additionaly, a various techniques were used to create modular biosensors composed of defined parts, which allows reliable response curve engineering. Moreover, various biosensors were creating to detect Neu5Ac *in vivo*, which was previously impossible and paves the way for various novel metabolic engineering strategies.

Samenvatting

Samenvatting

Onder invloed van ecologische bezorgdheden en de beperkte hoeveelheid fossiele brandstoffen gebruikt de chemische industrie meer en meer microbiële technologie voor de economisch rendabele productie van diverse moleculen uit hernieuwbare bronnen. Verschillende van dergelijke *metabolic engineering* strategieën worden ondersteund door technieken uit de synthische biologie wat de ontwikkeling van nieuwe microbiële celfabrieken voor een bepaalde molecule mogelijk maakt.

Eén van dergelijke interessante moleculen is *N*-acetylneuraminezuur, een suikermolcule met een essentiële rol in verschillende fysiologische processen zoals tumor groei, bacteriële infecties, hersenontwikkeling en immuniteit, waardoor het diverse toepassingen heeft in de farmaceutische en de voedingsindustrie. De ontwikkeling van deze toepassingen wordt echter verhinderd door de beperkte beschikbaarheid van *N*-acetylneuraminezuur wat een gevolg is van een tekort aan afdoende productie technologieën. Deze beperking kan worden opgeheven door gebruik te maken van *metabolic engineering* wat toelaat om het metabolisme *ad hoc* aan te passen om zo een organisme te bekomen met een geoptimaliseerde biosynthetische pathway met een maximale productiviteit. Dergelijke optimalisatie is echter een arbeidsintensief proces dat verschillende technieken vereist om gen expressie aan te passen, genetische ciruits te bouwen en specifieke moleculen te detecteren in de cel. Hierdoor was het hoofddoelstelling van dit doctoraatsonderzoek de ontwikkeling van verschillende ondersteunde technologieën voor *metabolic engineering* welke toelaten om genetische circuits te bouwen en kleine moleculen te detecteren.

De laatste jaren heeft de programmeerbaarheid van RNA de ontwikkeling van diverse nieuwe riboregulator gebaseerde technieken gestimuleerd. Deze technologie wordt meer en meer gebruikt in *metabolic engineering* om de productiviteit te maximaliseren. Eén interessant type van RNA regulatoren zijn riboregulatoren, welke toelaten om translatie te controleren zonder de nood voor coexpressie van stress veroorzakende eiwitten. De ontwikkeling van deze riboregulatoren wordt echter gehinderd door een gebrek aan duidelijke design regels, wat de toepasbaarheid in *metabolic engineering* beperkt. Om dit probleem op te lossen werden zo genaamde tiRNAs ontwikkeld. Deze riboregulatoren laten toe om controleerbaar eiwitexpressie op het translationele niveau te regelen. Deze tiRNA moleculen werden ontwikkeld via een systematische zoektocht door mogelijk interessante eigenschappen via een experimenteel design. Deze *de novo* ontwikkelde riboregulatoren onderdrukken translatie tot 6 % van de originele eiwitexpressieniveaus, wat een verbetering is van de dynamisch gebied van dergelijke riboregulatoren die vroeger werden beschreven. In contrast met deze vorige studies, zijn de tiRNA regulatoren volledig van niets ontworpen en ze dus geen natuurlijk voorkomend chassis

nodig hebben om te functioneren. Om de eigenschappen van de tiRNA riboregulatoren te linken aan hun efficiëntie werd een PLS model ontwikkeld, wat de programmeerbaarheid van deze riboregulatoren verder uitbereidt.

Naast aanpassingen van gen expressie laat RNA technologie ook toe om gen expressie te regelen op basis van de aanwezigheid van kleine moleculen. De laatste jaren zijn dergelijke RNA tools die reageren op kleine moleculen, meer specifiek *riboswitches*, gebruikt in verschillende *metabolic engineering* strategieën als alternatief voor hun traditionele eiwittegenhangers. Echter, de ontwikkeling van *riboswitches* op basis van *in vitro* geselecteerde aptameren vereist typisch arbeidsintensieve screening, wat hun toepasbaarheid sterk limiteert. Om deze hinderpaal in *metabolic engineering* te verwijderen werd een computationeel design algoritme ontwikkeld wat toelaat om *in silico* te screenen voor *riboswitches*. Om de riboswitch capaciteiten van een UTR te kwantificeren werd een doelfunctie opgesteld op basis van vroeger beschreven *riboswitches*. Gebruik makend van deze doelfunctie werden 29 potentiële *riboswitches* computationeel ontworpen gebruik makend van een zoekalgoritme. Vervolgens werden deze *riboswitches in vivo* geëvalueerd, waaruit bleek dat 12 van de 29 ontwikkelde *riboswitches* een activatie ratio hadden van meer dan vijf. Echter, ondanks de hoge waarschijnlijkheid om werkende *riboswitches* te ontwerpen, was het onmogelijk om de riboswitch functionaliteit te linken aan structurele of thermodynamische eigenschappen. Het ontwikkelde algoritme kan helpen om de ontwikkelingstijden te verkorten van translationele *riboswitches*, wat hun toepasbaarheid kan vergroten.

De complexe natuurlijke regulatie van het microbiële metabolisme laat toe om verschillende *metabolic engineering* strategieën te ontwikkelen, welke dikwijls intracellulaire detectie van specifieke moleculen vereisen. Om dit te bekomen werden verschillende biosensoren ontwikkeld gebaseerd op natuurlijk voorkomende transcriptiefactoren, wat typisch de mogelijkheden tot *ad hoc* optimalisatie van de responsecurve beperkt. Als een *proof of concept* werden er verschillende nieuwe biosensoren voor de detectie van *N*-acetylneuraminezuur, gebaseerd op natieve of aangepaste transcriptionele promotoren die interageren met de transcriptiefactor NanR. Vervolgens werden deze ontwikkelde biosensoren geëvalueerd gebruik makende van een *N*-acetylneuraminezuur producerende stam. Om de ontwikkelde biosensoren op een modulaire manier te optimaliseren werd een NanR binding site in een constitutieve promoter geplaatst, wat resulteerde in biosensoren die uit verschillende gedefinieerde onderdelen bestaan. Dergelijke manier van biosensoren ontwikkeling laat toe om op een meer betrouwbare manier de responsecurve aan te passen, wat de toepassingsmogelijkheden van biosensoren in

metabolic engineering verder vergroot. De vergrote mogelijkheden om de responsecurve aan te passen werden aangetoond door de response van één van de ontwikkelde biosensoren aan te passen via het vervangen van de ribosoom binding site die instaat voor de translatie initiatie van NanR. Daarnaast werden drie biosensoren gecombineerd met verschillende productieniveaus van *N*-acetylneuraminezuur (tot 1.4 ± 0.4 g/L extracellulair) waarbij werd aangetoond dat de waargenomen fluorescentie evenredig was met het productieniveau. Dit toont aan dat de biosensoren een brede operationele regio bevatten, een belangrijke eigenschap van biosensoren waardoor ze verschillende toepassingen hebben in *metabolic engineering*. Algemeen werd het aantal beschikbare biosensoren uitgebreid met verschillende functionele biosensoren die intracellulair *N*-acetylneuraminezuur kunnen meten, waardoor ze kunnen helpen bij diverse *metabolic engineering* strategieën om een maximale productie te bereiken.

Algemeen werden er in dit doctoraatsonderzoek verschillende technieken ontwikkeld die een meer betrouwbare optimalisatie van microbiële celfabrieken toe laten. Meer specifiek werden de mogelijkheden verbeterd om translatie inhiberende riboregulatoren en translationele *riboswitches* te ontwerpen, wat de toepasbaarheid van verschillende technieken uit de RNA synthetische biologie vergroot. Daarnaast werden verschillende technieken gebruikt om modulaire biosensoren met vergrootte *engineering* mogelijkheden te ontwikkelen. Bovendien werden biosensoren voor *N*-acetylneuraminezuur ontwikkeld die de vereiste eigenschappen bevatten die toelaten om verschillende nieuwe *metabolic engineering* strategieën toe te passen die vroeger onmogelijk waren.

Curriculum vitae

Gert Peters (September 2, 1988)

Education

2011-2016   **Doctor of Applied Biological Sciences**, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium.
PhD thesis:
Promotors: Marjan De Mey (Ghent University) and Jeroen Lammertyn (University of Leuven)

2009-2011   **Master of Science in Bioscience Engineering**, Chemistry and Bioprocess Technology, Ghent University, Ghent, Belgium.
Master thesis: *The development of a genetically modified* Escherichia coli *strain for the production of sophorose*

2006-2009   **Bachelor of Science in Bioscience Engineering**, Environmental Technology, Antwerp University, Antwerp, Belgium
Bachelor thesis: *The design of an algae photobioreactor*

2000-2006   Science and mathematics, Sint-Maarten, Beveren, Belgium.

Specialized courses

2014   **Summer School in RNA Bioinformatics**
University of Copenhagen, Copenhagen, Denmark.

2013   **Ghent Biobased Economy Summer School**
Ghent University, Ghent, Belgium

2012   **Bioinformatics**, part 1 and 2
Ghent University, Ghent, Belgium

Transferable skills

2016   **Leadership Foundation Course**
Ghent University, Ghent, Belgium

2014   **Advanced Academic English: Writing Skills**
Ghent University, Ghent, Belgium

2012   **Getting Started with High-Performance Computing**, part 1 and 2
Ghent University, Ghent, Belgium

Attended conferences

2016   **17th European Congress on Biotechnology**, 3-6 July, Krakow, Poland
Oral Presentation: *Automated* de novo *design of ligand-responsive RNA devices*

2014   **Metabolic Engineering X**, 15-19 June, Vancouver, Canada
Poster presentation: *Steering prokaryotic gene expression using engineered riboswitches*

2013   **iGem Europe Regional Jamboree**, 11-13 October, Lyon, France
Instructor of team UGent 2013 and iGem judge

2013   **Genome Engineering & Synthetic Biology: Tools and Technologies**, 16-17 September, Ghent, Belgium

2013   **Synthetic Biology 6.0: The Sixth International Meeting on Synthetic Biology**, 9-11 July, London, UK
Poster presentation: *A computational approach to building gene silencing modules*

2012   **Benelux Bioinformatics Conference 12**, 11-12 December, Nijmegen, the Netherlands, Belgium
Poster presentation: *Bioinformatics tools for the design of antisense RNA and aptamers*

Student guidance

Practical exercises General Microbiology (2012-2013)

Instructor iGEM Team UGent (2013).

Project: **A new model for chromosomal evolution: eliminating antibiotic resistance**

Tutor in student theses

2014-2015   **Robbe Christiaens**, M.Sc.: *A balancing act: optimization of sialic acid production via combinatorial engineering*

2014-2015   **Margot Bleys**, M.Sc.: *Simulating aptamers and riboswitches using bioinformatics*

2014-2015   **Maxim Rommelaere**, M.Sc.: *Development of a generic high-throughput screening method based on synthetic riboswitches*

2013-2014   **Ruben Walcarius**, B.Sc.: *Constructie van een synthetische genschakelaar voor de gecontroleerde expressie van een fluorescent proteïne.*

2012-2013   **Louise Delahaye**, M.Sc.: *De ontwikkeling van een genetisch gemodificeerde Escherichia coli stam voor de productie van gesialyleerde oligosachariden*

2011-2012   **Jorn Huylebroeck**, B.Sc.: *Evaluatie en optimalisatie van suikeropnamesystemen in Escherichia coli via metabolic engineering*

Publications

Peters, G., Coussement, P., Maertens, J., Lammertyn, J. & De Mey, M. Putting RNA to work: Translating RNA fundamentals into biotechnological engineering practice. *Biotechnology Advances* **33**, 1829-1844 (2015).

De Paepe, B., Peters, G., Coussement, P., Maertens, J. & De Mey, M. Tailormade transcriptional biosensors for optimizing microbial cell factories. *Journal of Industrial Microbiology & Biotechnology*, (2016).