

# **Design, implementation and characterization of synthetic riboswitches in *Saccharomyces cerevisiae***

Vom Fachbereich Biologie der Technischen Universität Darmstadt

zur

Erlangung des akademischen Grades

eines Doctor rerum naturalium

genehmigte Dissertation von

Christopher Schneider, M. Sc.

aus Siegen

1. Referentin: Prof. Dr. Beatrix Süß

2. Referent: Prof. Dr. Heinz Köppl

Tag der Einreichung: 14.03.2017

Tag der mündlichen Prüfung: 04.04.2017

Darmstadt 2017

D 17

## Danksagung

Frau Prof. Dr. Beatrix Süß gilt mein ausdrücklicher Dank für die überaus freundliche Aufnahme in ihren Arbeitskreis und die Möglichkeit weiterhin an der Entwicklung von RNA-Genschaltern im Rahmen meiner Doktorarbeit zu forschen. Ich möchte mich ausserdem sehr für die Integration in den LOEWE Schwerpunkt CompuGene als Doktorand, die hervorragende Betreuung und die Möglichkeit zur Teilnahme an Tagungen und Konferenzen im Rahmen des SFB902 bedanken.

Ich danke Herrn Prof. Dr. Heinz Köppl zum Einen für die freundliche Übernahme des Zweitgutachtens und zum Anderen für die exzellente Kooperationsarbeit, die zu ROC'n'Ribo führte.

Jascha, Leo und Sven möchte ich ganz besonders für die tolle Zusammenarbeit und die überaus hilfreichen Erklärungen zur Modelierung danken. Ihr habt diese Arbeit maßgeblich bereichert. Dank!

Natürlich möchte ich mich auch bei meinen Kollegen für eine gute Zusammenarbeit und eine nette Zeit im Labor bedanken.

Mama, Papa und Bengel: Eure immerwährende und bedingungslose Unterstützung, Euer grenzenloses Verständnis, Eure steten Ermutigungen und Eure nie versiegende Zuversicht haben mir alles im Leben ermöglicht und machen mich sehr glücklich und über alle Maßen dankbar.

Nini: Bei dir kann ich sein wie ich bin und so haben wir gemeinsam so vieles geteilt. Danke, dass du immer an meiner Seite warst.

Janina Atanasov danke ich sehr für das Korrekturlesen meiner Arbeit.

Lotta und Rocky danke ich, dass sie so lieb und geduldig während der Bearbeitung dieser Thesis waren und dass sie nicht allzu häufig über die Tastatur liefen...

Parts of this thesis have been or will be published in:

Schneider, C., and Suess, B. (2016) **Identification of RNA aptamers with riboswitching properties.** *Methods* 97, 44–50

Schneider C., Bronstein L., Diemer J., Koepl H., Suess B. (2017) **ROC'n'Ribo: Characterizing a riboswitching expression system by modeling single-cell data.** *ACS Synth Biol.*, revised

## Table of contents

### Table of contents

1	Zusammenfassung .....	1
1	Summary .....	3
2	Introduction .....	5
2.1	Synthetic biology.....	5
2.2	Yeast – A patriarchal and tame cell factory for synthetic biology applications.....	6
2.3	RNA-engineered gene regulation .....	9
2.3.1	Synthetic RNA regulatory mechanisms .....	9
2.3.2	Natural riboswitches: Occurrence, structure and function.....	12
2.3.3	Synthetic riboswitches .....	16
2.4	RNA synthetic genetic networks.....	27
2.4.1	Synthetic parts characterization .....	29
2.4.2	Synthetic RNA logic gates .....	29
3	Scope.....	31
4	Results & Discussion .....	32
4.1	Project I: Identification of RNA aptamers with riboswitching properties .....	32
4.1.1	Results .....	32
4.1.2	Discussion.....	50
4.2	Project II: Tetracycline-Dimers: A massive approach towards the <i>in silico</i> prediction of riboswitch performance.....	55
4.2.1	Results .....	55
4.2.2	Discussion.....	60
4.3	Project III: ROC'n'Ribo: Characterizing a riboswitching expression system by modeling single-cell data .....	62
4.3.1	Results .....	62
4.3.2	Discussion.....	80
5	Methods .....	86
5.1	General protocols for molecular cloning .....	86
5.1.1	PCR .....	86
5.1.2	Colony-PCR .....	87
5.1.3	Oligonucleotide Hybridization and 5'-Phosphorylation .....	88
5.1.4	Fill-in reaction .....	88
5.1.5	Restriction digest .....	89
5.1.6	Gibson cloning reaction .....	89
5.1.7	Ligation .....	90
5.1.8	Transformation of chemo-competent E.coli in 96-well format .....	90
5.1.9	Transformation of electro-competent E.coli.....	90

## Table of contents

5.1.10	General protocol for yeast homologous recombination .....	90
5.1.11	Transformation of competent yeast cells .....	91
5.1.12	Agarose gel electrophoresis .....	91
5.1.13	DNA isolation – Mini preparation from Top 10 cells .....	91
5.1.14	DNA isolation – Mini preparation from yeast .....	91
5.1.15	DNA isolation – genomic DNA preparation from yeast (Bust'n'grab).....	91
5.1.16	DNA quantification .....	91
5.1.17	DNA purification – Ethanol, isopropanol and butanol .....	92
5.1.18	Sequence analysis.....	92
5.1.19	Preparation of chemo-competent E.coli .....	92
5.1.20	Preparation of electro-competent E.coli .....	92
5.1.21	Preparation of chemo-competent yeast.....	92
5.1.22	Yeast glycerol stocks .....	93
5.2	Adapted protocols for molecular cloning and reporter gene analysis .....	93
5.2.1	<i>In vivo</i> Screening of RNA aptamers in <i>S. cerevisiae</i> .....	93
5.2.2	Characterization of a riboswitching expression system.....	97
5.2.3	Tc-Dimers .....	99
6	Material .....	100
6.1	Chemicals, instrumentation and consumables .....	100
6.2	Prokaryotic and eukaryotic cell strains .....	103
6.3	Buffers and solutions .....	104
6.4	Oligonucleotides .....	105
6.5	Plasmids .....	109
6.5.1	Basic plasmids .....	111
6.5.2	Construction of plasmids .....	115
7	Appendix.....	119
7.1	Abbreviations .....	119
7.2	Units.....	120
7.3	Prefixes.....	120
7.4	Nucleobases.....	120
8	References .....	121
9	Talks and Poster Presentations .....	129
10	Publications .....	130
11	Curriculum Vitae.....	131
12	Ehrenwörtliche Erklärung.....	132

### 1 Zusammenfassung

Eine der treibenden Kräfte in der Synthetischen Biologie ist das Bestreben programmierbare Zellsysteme nach ihren natürlichen Vorbildern zu konstruieren. Um dies zu erreichen müssen Systeme entwickelt werden, die sensitiv auf Umwelteinflüsse reagieren und die empfangenen Signale prozessieren. Entsprechende Sensoren und Reglereinheiten können gemäß natürlich vorkommenden RNA-Schaltern, so genannten Riboswitches, generiert werden. Die Etablierung von Systemen, die mehrere verschiedene Signale verarbeiten können, erfordert zunächst die Verfügbarkeit geeigneter Sensoren, z.B. RNA Aptamere. Viele der derzeit verwendeten synthetischen Riboswitches können nur wenige verschiedene Liganden wahrnehmen, die dementsprechend als Input genutzt werden können. Die Erweiterung des Repertoires an nutzbaren Liganden ist ein unabdingbarer Schritt hin zur Implementierung entsprechender Regelkreise. Die Eigenschaft *in vitro*-generierter Aptamere auch gleichzeitig als Regulatoren in Hefen verwendet werden zu können, wurde genutzt um die aktuelle Literatur auf geeignete synthetische und natürliche Aptamere zu durchsuchen. Identifizierte Kandidaten wurden hinsichtlich ihrer Eignung die Expression eines Reportergens in Anwesenheit ihrer spezifischen Liganden zu verringern getestet. Diese Regulation auf translationaler Ebene basiert auf der Blockade des Translationsinitiationskomplexes, während dieser die mRNA auf der Suche nach dem Translationsstartpunkt abtastet. Zwar konnten verschiedenste Basalexpressionen ermittelt werden, jedoch blieb die Entdeckung eines regulatorischen Phänotyps aus. Diese Ergebnisse bestätigen zum einen, dass *in vitro* hochevolvierte Aptamere nicht in der Lage sind ihre Struktur dahingehend zu ändern, dass sie eine Genregulation herbeiführen können und zum anderen, dass natürliche Aptamerdomänen eine dedizierte Expressionsplattform für ihre Aktivität benötigen. Im Falle des *in vitro* generierten Neomycin-Aptamers ermöglichte ein *in vivo* Screening Ansatz jedoch Identifikation von Aptameren, die regulatorische Aktivität aufwiesen. Gemäß dieses Beispiels wurde ein *in vivo* Screening mit Ciprofloxacin-bindenden Aptamere durchgeführt. Auf diese Weise konnte ein Aptamer identifiziert werden, das eine 2-fache Genregulation zeigte.

In einem weiteren Ansatz zur Verbesserung bereits verwendeter Riboswitches und mit dem Ziel Regel für die Riboswitch-Konstruktion zu ermitteln, wurden Dimere des Tetrazyklin-bindenden Riboswitches in einem Reporterassay untersucht, der den bereits beschriebenen Regulationsmechanismus auf translationaler Ebene verwendet. Bereits die Modifikation eines der beiden Riboswitches genügte um eine große Bandbreite an Basalexpressionen und Schalfaktoren zu erhalten. Die von mehr als 100 Konstrukten erhobenen Daten wurde im Anschluss genutzt um ein Computermodell zu trainieren. Nachdem Werte für die Parameter Basalexpression und Schalfaktor festgesetzt worden waren, die für eine Verbesserung der regulatorischen Aktivität nötig sind, klassifizierte das Computermodell entsprechende Sequenzen aus einem randomisierten Sequenzraum, die im Anschluss *in vivo* getestet wurden. Diese Ergebnisse spiegelten teilweise die erste Datenaufnahme wider, zeigten jedoch zugleich, dass das Computermodell die erlernten Sequenz-Aktivitäts-Beziehungen wiedergeben konnte. Überdies konnten regulatorisch inaktive Sequenzen praktisch komplett eliminiert und die Häufigkeit von Schaltern mit gewünschten Phänotypen erhöht werden.

Der beschriebene Ansatz zur Verbesserung von existierenden Riboswitches wurde auf Einzelnukleotidebene durchgeführt. In einer dritten Studie sollte die regulatorische Leistung eines logischen NOR Gatters bestehend aus Neomycin- und Tetrazyklin-bindenden Riboswitches auf Ebene des Logikgatters untersucht werden. Ein Deskriptor für die regulatorische Leistung des beschriebenen Gatters sollte angewendet werden um seine

## Zusammenfassung & Summary

Funktion auf Einzelzellebene zu bestimmen, da die korrekte Funktionalität von (RNA-) Gattern eine unabdingbare Voraussetzung für ihre Verwendung in einem genetischen Schaltkreis darstellt. In diesem Zusammenhang wurde die Empfänger-Operator-Charakteristik (*Receiver-Operator-Characteristics*, ROC) Analyse als geeigneter Deskriptor identifiziert, um die Variabilität zwischen Einzelzellen, die die vollständige Trennung von zwei Logikzuständen durch Überlappung der beteiligten Populationen erschweren, zu berücksichtigen. Vervollständigt durch die transiente Messung der Induktions- und Repressionskinetiken des konstruierten NOR-Gatters mittels Durchflusszytometrie und Mikrofluidik-basierter Fluoreszenzspektroskopie, konnten die aufgenommenen Daten zur Charakterisierung der regulatorischen Leistung hinsichtlich Geschwindigkeit und Genauigkeit der Logikverarbeitung genutzt werden. Die meisten Studien verwenden nur Endpunktmessungen einer Bulkkultur, in der diese Daten nicht zu differenzieren und dementsprechend nicht zugänglich sind. Anhand der transienten Einzelzelldaten wurde ein hierarchisches stochastisches Modell konstruiert und kalibriert, das die *in silico* Vervollständigung der Gattercharakterisierung im Gleichgewichtszustand der beteiligten Komponenten über verschiedene Induktionslevel und Repressorkonzentrationen ermöglichte. General können solche Daten genutzt werden um experimentell unzugängliche Leistungsdaten zu erhalten. Darüber hinaus erlaubte das Modell die Computer-gestützte Untersuchung eines möglichen Neudesigns des Gatters, das nach experimenteller Umsetzung die Vorhersage des Modells bestätigte. Dies hebt sowohl die Eignung des experimentellen Aufbaus geeignete Daten zu generieren, als auch die Anwendbarkeit von computergestützten Ansätzen den korrekten *in vivo* Phänotyp zu bestimmen, hervor.

### 1 Summary

The driving force of synthetic biology is the desire to build programmable cellular systems *de novo* modeled on their natural counterparts. To achieve this, environmental signals must be sensed and integrated to generate an appropriate output. Sensors and actuators can be created according to naturally occurring riboswitches. The implementation of systems that can process many inputs initially requires the availability of the corresponding sensor domains, called aptamers. Many of the currently used synthetic riboswitches rely on only a handful of ligands as their external input. It is therefore desirable to expand the toolbox by adding other ligand specificities. Harnessing the potential of *in vitro* selected aptamers to function as *in vivo* regulators in yeast, the current literature was screened for suitable synthetic and natural aptamers. Candidates were then assayed for their ability to reduce reporter gene expression at the translational level in presence of their cognate ligands, essentially functioning as a roadblock to the scanning ribosome in the 5'-UTR. The results revealed a wide range of basal expressions, but no switching phenotypes. Since the aptamers had not been adapted to the expression system, the results confirmed either the inability of most highly *in vitro* evolved aptamers to exhibit the structural changes necessary for riboswitching or the requirement of natural aptamers for a dedicated expression platform. In case of the *in vitro* generated neomycin aptamer, an *in vivo* screening yielded aptamers with regulatory activity. Following this example, ciprofloxacin-binding aptamers were subjected to the screening and one aptamer was found that exhibited 2-fold regulation. This aptamer was then subjected to further analysis. In an approach to improve existing riboswitches and derive rules for riboswitch design, dimers of the tetracycline riboswitch were set to control reporter expression by the described regulatory mechanisms. Modifying one riboswitch was sufficient to obtain a range of basal expressions levels and switching factors that could then be used to train a computational model based on data from more than 100 different constructs. After setting desired output parameters with respect to basal expression and switching factors the model computed corresponding sequences that were tested in a second iterative approach. These results partly mirrored the first data recording, but also showed that the model could reproduce the sequence-activity relationships learned from the first recording. Moreover, non-regulating sequences were almost completely erased and the abundance of switches with desired switching factors was increased.

Leaving the refinement of riboswitches at the nucleotide level, a third study was conducted to elucidate the performance of a logic NOR gate constructed from the neomycin and tetracycline riboswitches at the device level. A descriptor for the performance of this RNA logic gate was sought to be applied to assess its functionality at the single-cell level, since the correct functionality of (RNA) devices in each single cell is a prerequisite for their adaption to a genetic circuit. To this end the Receiver-Operator-Characteristics (ROC) analysis was identified as a suitable performance measure to account for cell-to-cell variability compromising the full separation of two Boolean states by overlapping populations. Complemented by the transient induction and repression kinetics of the constructed NOR gate recorded by flow cytometry and microfluidics-based fluorescence microscopy, data was derived that allowed mapping the performance of the NOR gate with respect to speed and accuracy of the logic operation. In most studies only steady state data from bulk cultures are collected and these information are lost.

A hierarchical stochastic model was built and calibrated with the transient single-cell data that enabled the *in silico* completion of the device characterization at the steady state level for various induction levels and repressor concentrations. In general, such data can then be used to simulate the experimentally inaccessible

## Zusammenfassung & Summary

gate performance. Additionally, the model facilitated the computational investigation of a possible re-design of the gate for which the model prediction proved to be correct *in vivo*, highlighting both the appropriateness of the experiment design to obtain meaningful data and the applicability of *in silico* approaches to predict the correct *in vivo* phenotype.

## 2 Introduction

### 2.1 Synthetic biology

Essentially a spin-off from molecular biology, synthetic biology set out to build things that do not exist yet – as a discipline in its making for nearly two decades. Having survived the prime years of orientation-seeking in the late 1990's, this new research field has established itself as a discipline of applied science and is widely acknowledged today <sup>1</sup>. Among the many definitions and envisioned goals, the forward engineering of genetic parts into cellular systems with defined functionality represents a prominent example <sup>2</sup>. In the early stages the field had greatly benefited from the introduction and refinement of cloning strategies, automated DNA sequencing, high-throughput techniques (HTS) and the elucidation of microbial genomes and subsequently focused on the development of simple genetic devices constructed from parts like promoters and transcription factors. Quickly, parts and devices were connected to the first synthetic genetic circuits that could perform logic functions in analogy to electrical circuits <sup>3,4</sup>. Although nowhere near to computing digital behavior, these approaches demonstrated the feasibility of such designs in bacterial cells. The driving organism *Escherichia coli* facilitated these advances by its ease of genetic manipulation and the well-studied and accessible regulatory mechanisms that provided the necessary parts to begin with. Genetic oscillators were built in proof-of-concept studies, tapping the potential of iterative design-built-test-debug approaches that remain until today as a necessity to tune the genotype towards the desired phenotype <sup>5-7</sup>. Computational modeling was introduced as a means to aid in the design of small circuitries and autoregulatory negative- and positive-feedback modules emerged <sup>8-10</sup>. Practically all designs were implemented with transcriptional parts, because tools for the regulation at the translational level had not been developed yet <sup>6,11</sup>.

From roughly 2003 on the overall scope had begun to increase and as a matter a new long term goal of whole-genome engineering manifested <sup>12,13</sup>. Realizing that it was still a far way to go, efforts were put on the standardization of parts and the indexing of complicated cellular systems into hierarchical abstractions <sup>14,15</sup>. Novel regulatory mechanisms that were derived from RNA parts expanded the possibilities of circuit construction to the post-transcriptional and translational level <sup>16,17</sup>. Cell population-wide engineering by quorum-sensing set new standards of what was feasible to manipulate <sup>18-20</sup>. A breakthrough in metabolic engineering was achieved by the heterologous production of precursors to artemisinin – a potent antimalarial drug – in bacteria that was later fully implemented in yeast and eventually commercialized <sup>21-23</sup>. These progressions were made possible by further improvement in the modeling and construction of higher order genetic circuits that could integrate and process various in- and outputs <sup>24-26</sup>. Additionally, HTS DNA assembly methods and ever decreasing costs of gene-synthesis further accelerated the field <sup>2,27</sup>. Advances in the reprogramming of fundamental cellular processes that operate orthogonally across several species like new RNA regulatory tools derived from the Clustered-Regulatory-Interspaced-Short-Palindromic-Repeats (CRISPR) immune system from bacteria and the expansion of the genetic code by unnatural amino acids provided further opportunities in the framework of cellular engineering <sup>28-30</sup>. Another milestone in the field was accomplished by the first total synthesis and assembly of a mini genome in a bacterial cell <sup>31</sup>. Capitalizing on this approach also the synthesis of the first synthetic eukaryote – yeast SC2.0 – was initiated and two chromosome arms have already been constructed

<sup>32</sup>.

## Introduction

Despite the many advances and spectacular highlights, many fundamental problems remain. For once, part characterization is still very poor and thus they may not simply be transferred from one study to another without compromising their functionality<sup>2,33</sup>. Hence, parts have to be tweaked anew and circuits correspondingly re-designed, unnecessarily increasing time and labor<sup>34</sup>. Secondly, and closely related to the first issue, as parts characterization remains confounding, the elucidation of context effects, be it neighboring sequences or interfering metabolic pathways, pose a significant obstacle to the overall progress<sup>35</sup>. Several studies have tried to alleviate this problem by reducing genetic complexity or insulating expressed parts from the context<sup>36–38</sup>. Still, as especially most developed circuits are not used by others than the creators, they experience slowed improvement or even premature death and vanish from the synthetic biology (synbio) toolbox. A rethinking of traditional practices is wishful to build on the work of others and to foster synbio approaches on their way forward.

### 2.2 Yeast – A patriarchal and tame cell factory for synthetic biology applications

Mankind has benefited from yeast for several millennia by its easy-to-culture and non-toxic nature<sup>39,40</sup>. The predominant strain *Saccharomyces cerevisiae* has been used for baking and brewing ever since and it appears that we are now paying tribute to this microbe by constructing the first synthetic eukaryote – yeast strain SC2.0<sup>32,41</sup>. *S. cerevisiae* features many desirable properties that paved its way for world-wide industrial and scientific use and ultimately the re-engineering of its genetic chassis for synbio applications<sup>42</sup>. The proficiency in DNA recombination and the occurrence of tried and tested selection markers enable the straight-forward introduction, expression and analysis of heterologous genes and the effective and targeted manipulation of the yeast genome itself<sup>43–46</sup>. Its ability to grow to high cell density, all the while tolerant to a wide range of carbon sources, and availability of extensive Omics data (genome, proteome, metabolome) facilitate the exploitation of *S. cerevisiae* as host and model organism not only for food and biotech industry, but also for academic purposes in the field of synthetic biology. Especially here, applied basic research is conducted to unravel and expand the possibilities this organism offers for developing new synthetic regulatory pathways and circuits as reviewed by Ellis *et al.*<sup>47</sup>. Being a lower eukaryote, *S. cerevisiae* has typically served as a link between Synbio research in bacteria, foremost *E. coli*, and higher eukaryotes represented by several human cell lines like HEK-293 and HeLa cells since the Synbio field emerged nearly 20 years ago<sup>48</sup>. Interestingly, nowadays least studies report the introduction of engineered genetic devices in yeast compared to massive approaches using more simply structured *E. coli* or the rather complicated biology of human cells that is supposedly more frequently tackled, because of the high demand on promising Synbio applications for human gene therapy<sup>49</sup>. Nevertheless, the paramount value of *S. cerevisiae* to the production of fine chemicals, pharmaceuticals and fuel is reflected by a long history of industrial application<sup>42,50,51</sup> as is its capability for rapid prototyping of expression systems<sup>43,52</sup>. Exemplified by the semi-synthetic production of artemisinic acid, a precursor of the potent antimalarial drug artemisinin<sup>21</sup> and the assembly of entire bacterial genomes<sup>31</sup>, the possibilities to host orthogonal components at a large scale were impressively demonstrated and highlight the potential of *S. cerevisiae* to stably maintain such critical genetic modifications<sup>35,47</sup>.

Advancing Synbio research has generated several gene expression tools especially for yeast metabolic engineering. Among these genetically encoded parts are foremost promoters, terminators, transcription factors and reporters that are coupled to exert control over implemented metabolic pathways<sup>42,47,53</sup>. Figure 2.1 presents

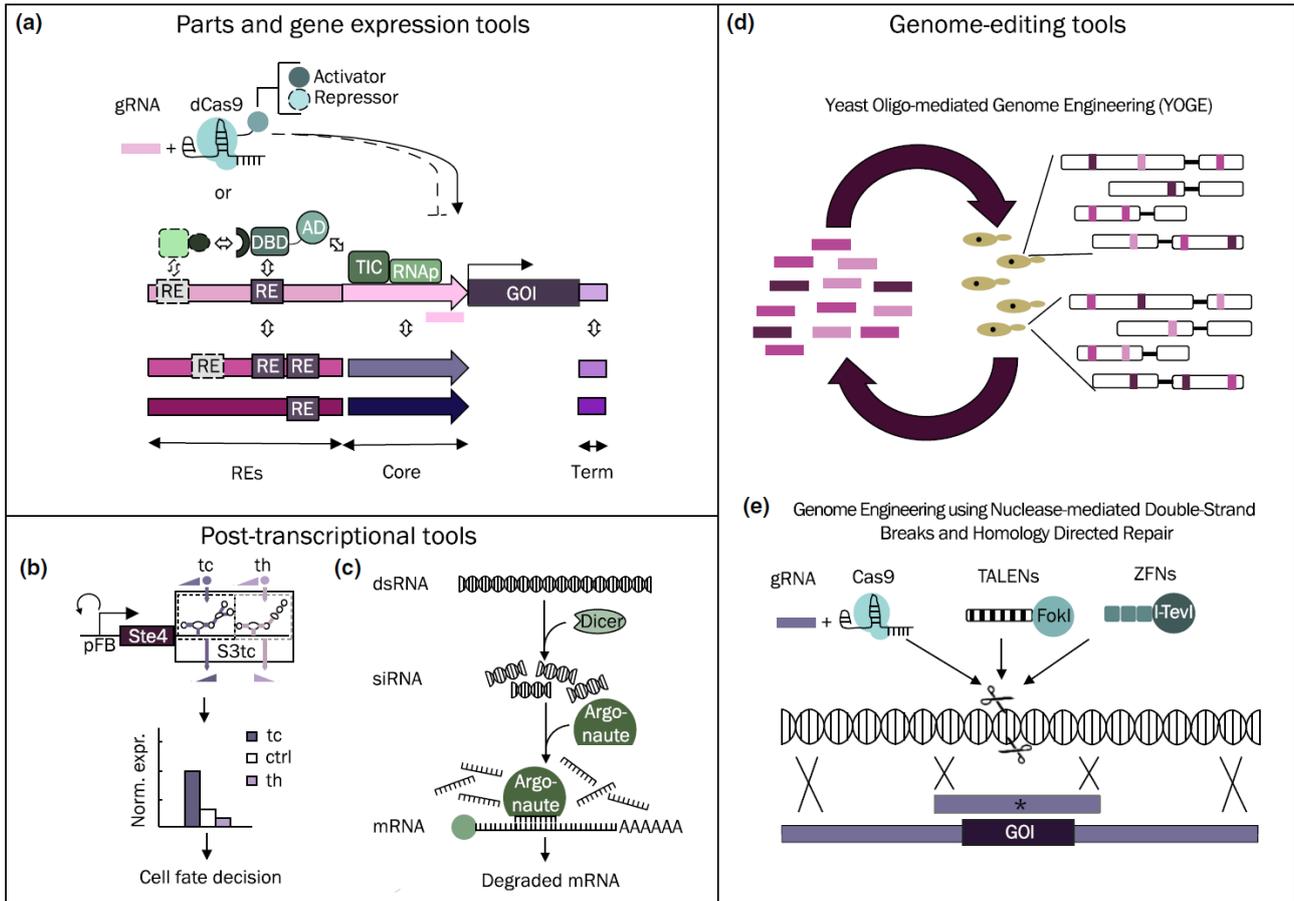
## Introduction

an overview about recently applied tools for improving chemical and fuel production <sup>42</sup>. Promoters and terminators operate as functional units flanking the gene of interest (GOI) and determine its expression level and mRNA stability <sup>54-56</sup>. Promoters can be fine-tuned on the single-nucleotide (nt) level to regulate their activity and promoter-terminator couples may be tailored to match desired expression levels and mRNA half-lives <sup>54,56</sup>. This is especially important for controlling the expression of key intermediates and thus neutralizing flux imbalances that would otherwise compromise pathway performance and product titers <sup>42</sup>. The power and impact of promoters on integer metabolic pathways is ideally complemented by the introduction of regulatory elements facilitating the binding of transcription factors (TF) that in turn modulate promoter activity. Essentially composed of an interchangeable DNA-binding and transcription activation domain, their modular nature allows the re-engineering of TFs to orthogonally regulate synthetic transcriptional expression systems in yeast. Their functionality and aptitude has been even further expanded by the addition of sensor domains or fusions to heterologous enzymes <sup>57-59</sup>. Sensor domains may either be used to create feedback loops in response to metabolite flux <sup>60</sup> or to sense external stimuli targeting only the modified TF and thus exhibit single-gene specificity <sup>57,58,61</sup>. To achieve orthogonal single-gene specificity these and other studies underline the focus on highly programmable zinc-finger TFs and correspondingly modified promoter sequences. In this context Mclsaac *et al.* have designed zinc-finger TFs that may be multiplexed to target several genes with identical promoters or to address multiple promoters specifically with a set of corresponding TFs. Here, the TFs are activated by 17 $\beta$ -estradiol that binds to its cognate receptor domain, the human estrogen receptor, fused to the TF <sup>57,58</sup>. As indicated in Figure 2.1a, not only subdomains of TFs are engineered, but rather fusions with (dys) functional enzymes from heterologous origins are conducted to further enlarge the versatility of TFs. The illustrated hallmark example describes the transcriptional regulation by the CRISPR/Cas system <sup>62</sup>. The fusion of the deactivated Cas9 endonuclease (dCas9) to a TF or a transcription activating or repressing domain provides promoter silencing or activation by simple occupancy or use of the functional domains, respectively <sup>59,63,64</sup>. The promoters are conveniently targeted by the guide RNA (gRNA) inherent to the CRISPR system, allowing the specific regulation of virtually any promoter without or with only minor sequence adaptations to meet the criteria for gRNA recognition.

Promoters and TFs are means to tune gene expression before transcription has been initiated. Co- or posttranscriptional regulation has been achieved by the application of RNA-based tools, so called riboswitches, and RNA interference mechanisms. As riboswitches are subject to a separate section and have not yet been used in metabolic engineering, they will only be described briefly. Riboswitches are comprised of a sensing and actuating domain facilitating control over transcription or translation in response to rearrangements of their global architecture triggered by external stimuli <sup>65</sup>. Figure 2.1b shows an approach by Galloway *et al.* utilizing riboswitches responsive to theophylline and tetracycline to enhance or reduce *STE4* expression level, ultimately determining cell fate <sup>66</sup>. Although riboswitches present the possibility to gain control over gene expression in a tight spatiotemporal manner, they have not been used for biotech applications due to limitations of suitable ligand specificities coupled to robust functionality <sup>42</sup>. As the field is growing rapidly and due to their magnificent engineering capacity, riboswitches are anticipated to largely contribute to the optimization of metabolic pathways <sup>67,68</sup>. RNA interference, another posttranscriptional regulatory mechanism, was heterologously expressed in *S. cerevisiae* after transfer of the required machinery from *S. castellii* <sup>69</sup> (Figure 2.1c). A proof-of-

## Introduction

concept study revealed the aptitude of this portable system to downregulate *ADE3* by siRNA interference to increase levels of itaconic acid<sup>69,70</sup>.



**Figure 2.1 Synthetic biology tools applied to metabolic engineering in *S. cerevisiae*.** (a) Transcription is controlled by engineered promoter and terminator sequences. Synthetic TFs with zinc-finger DNA binding domains are predominantly used to confer single-gene specificity. The CRISPR/Cas system is modified to gain targeted attenuation of transcription initiation by adaption of gRNA and Cas9 endonuclease. (b) RNA regulators, exemplified with riboswitches sensitive to tetracycline and theophylline, are increasingly used to confer co- and posttranscriptional gene regulation, although they have not yet been applied to metabolic engineering. Here the stability of the mRNA was influenced by the ligand-dependent removal of the poly-A tail, determining cell fate. (c) Molecular components of RNA interference (Dicer and Argonaute) can be transferred to *S. cerevisiae* to orthogonally introduce mRNA silencing and thus facilitate gene knockdown. (d,e) Genome-editing tools allow the targeted and multiplex integration of different DNA inserts into the desired locus using Yeast Oligo-mediated Genome Engineering (YOGE), CRISPRi and Transcription Activator-like Endonucleases. Adopted from<sup>42</sup>.

The modification of the genome is another important branch of genetic tools available in different facets (Figure 2.1d, e). Multiplexed yeast oligo-mediated genome engineering (YOGE) is a fast method to simultaneously conduct editing of multiple genomic loci with varying donor templates. The technique yields 0.2-2% insertion efficiency without applied selection pressure and can thus be used to confer moderate screening approaches of library sizes between  $10^2$  and  $10^5$  recombinants per locus per cycle<sup>71</sup>. The method relies entirely on the high proficiency of *S. cerevisiae* to perform homologous DNA repair. This feature is also exploited by the CRISPR interference (CRISPRi)-based technique for genome manipulations. The targeted double-strand break facilitates high efficiency recombinations with any insert DNA homologous to the integration site. Moreover, the design of the gRNA allows integration in many user-specific loci<sup>45,72,73</sup>.

## Introduction

Besides the genetic tools presented in Figure 2.1 also metabolic tools are accessible to efficiently facilitate metabolic flux through a constructed pathway. The spatio-temporal organization of multienzymatic cascades by compartmentalization and scaffolds has frequently been used to increase titer and rate constants, as they physically connect the participating components <sup>74,75</sup>.

Thus, further Synbio research on *S. cerevisiae* can readily capitalize on its inherent cellular, metabolic and genetic properties and concentrate on the design, characterization and *in vivo* assembly of genetic parts needed to *de novo* construct and spatiotemporally regulate higher order expression networks.

### 2.3 RNA-engineered gene regulation

RNA-engineered systems offer simple and versatile control over gene expression in many organisms <sup>76–78</sup>. Realizing their potential as highly programmable synthetic parts, they are beginning to seriously compete with the established engineered transcription factor-promoter couples that have dominated as regulators of gene expression so far <sup>48,79,80</sup>. RNA parts promote the rapid design and implementation of gene regulators due to their simple make up of only four different nucleotides, their superior structural flexibility at low sequence loads and the emergence of increasingly sophisticated computational methods to facilitate *in silico* modeling and structure prediction <sup>76,81–84</sup>. Additionally, RNA parts can be configured to respond to a plethora of external inputs, rendering RNA-derived regulation suitable to quantitatively tune the expression profile of the hosting system <sup>77,85,86</sup>. The indispensability of RNA as a component of larger, protein dominated systems, such as CRISPR/Cas, has further expanded its versatility and applicability to tailor conditional gene expression systems <sup>59,62,76</sup>.

Driven by these compelling benefits, several different synthetic RNA regulatory mechanism inspired by the natural RNA templates have been prototyped to date <sup>76,77,86</sup>.

#### 2.3.1 Synthetic RNA regulatory mechanisms

Synthetic RNA-based regulation has particularly resurfaced during the last several years <sup>76,77,80</sup>. The old pacemakers of the field, small trans-activating riboregulators and classical riboswitches, have been refurbished and complemented by new teammates. As a result, the current toolbox is geared towards controlling gene expression at the levels of transcription and translation in bacteria as well as transcription, translation and splicing in yeast and mammals (reviewed in <sup>76,77,87</sup>). The following chapter deals with regulatory mechanisms other than those exerted by riboswitches. Their contribution to the toolbox is described in 2.3.3 Synthetic riboswitches.

Figure 2.2 illustrates two major classes of promising state-of-the-art design principles applied in *E. coli*, *S. cerevisiae* and mammalian cells. The class of synthetic riboregulators emerged in 2004 as a direct mimic of small bacterial RNAs (sRNA) (Figure 2.2a, b) <sup>88</sup>. Natural sRNAs were the first trans-acting RNAs discovered that specifically regulate gene expression by an anti-sense mechanism <sup>89</sup>. They facilitate binding to their target sequence with a 6-8 nucleotide (nt) seed region, base-pairing in the direct vicinity of the ribosome binding site (RBS) further upstream to a ribosome stand-by site or a translational enhancer sequenced of the respective mRNA and thus control translation. Their mode of action is not only limited to inhibit translation, as they were additionally shown to activate translation by disruption of inhibitory secondary structures in the 5'-UTR <sup>90,91</sup>. The latter process was harnessed in several approaches to regulate gene expression in bacteria, since the used RNA molecule allowed for high sequence variability during the design process according to computable RNA-

## Introduction

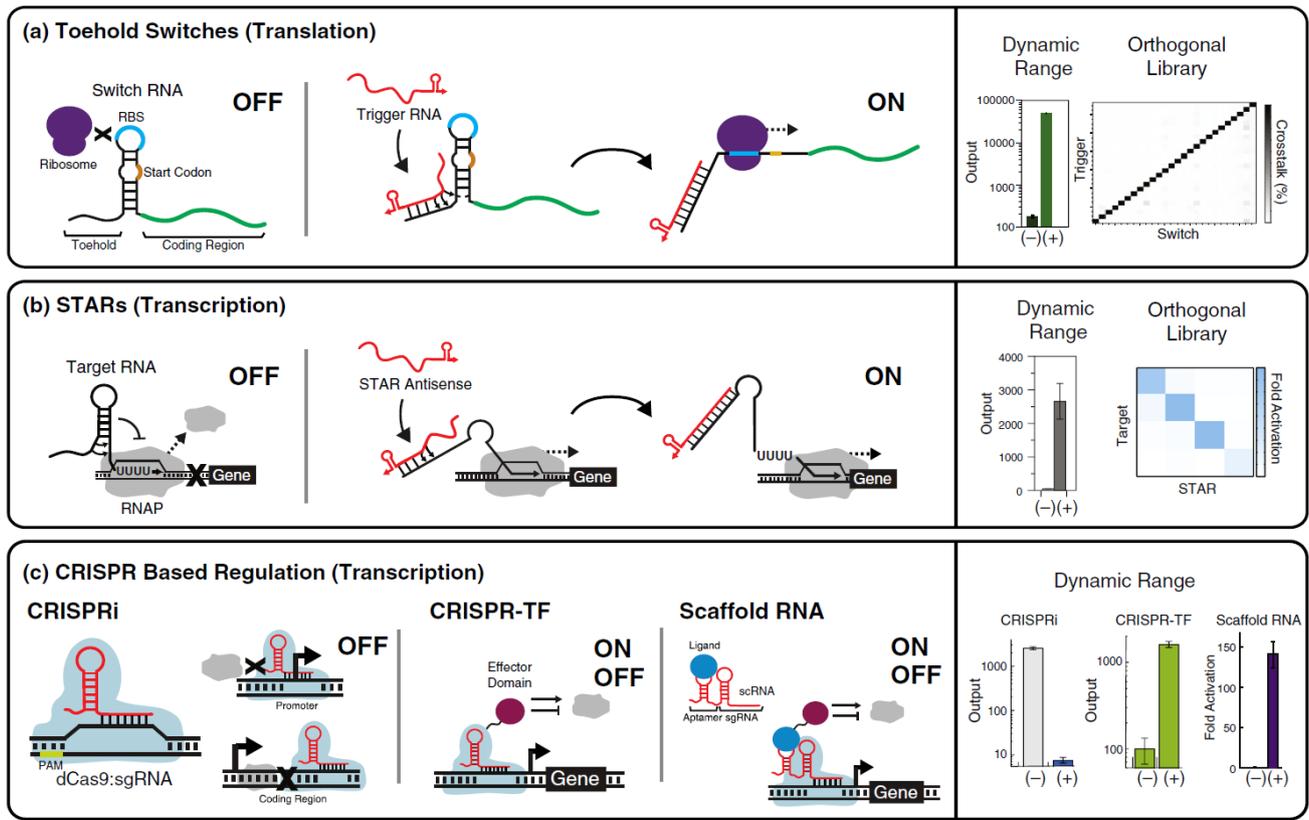
RNA interactions<sup>16,25,92,93</sup>. Further, RNA-RNA interactions are tunable over a wide range of the desired effect and the response times are very short. The prototypes constructed by Isaacs *et al.* were composed of a cis-repressive RNA (crRNA) and a trans-activating RNA (taRNA) counterpart. The crRNA prevented docking of the ribosome to the RBS by sequestering the SD sequence in a hairpin structure. Stem-loop formation was promoted by a 25 nt cis-repressive sequence. The taRNA contained a 26 nt region of large complementarity to the crRNA and was capable to unzip the hairpin, resulting in translation initiation by the ribosome. This system showed minimized leakiness of reporter gene expression in the absence of the taRNA and elevated reporter levels up to 100-fold of the repressed state<sup>16</sup>. This system was extensively employed and modified in successive investigations to yield a genetic switchboard in which several orthogonal taRNA-crRNA pairs were used to independently regulate the expression of three different genes. These works used promoters that responded to environmental signals as sensing units<sup>25</sup>.

As mentioned before, a wide sequence space for the design of taRNAs and their respective counterparts can be utilized. To access the possibilities such sequence variability offers, Rodrigo *et al.* showed that *de novo* computation of small RNA circuits is quite feasible, if certain rationales are imposed on the algorithm calculating possible candidate structures. Kinetic considerations as well as the free energies of the folded and unfolded states of the single RNAs and the proposed complex showed that RNA interactions are predominantly influenced by the energy of complex formation and activation of the educts<sup>93</sup>. The latest and most extensive approach by Green *et al.* is displayed in Figure 2.2a. They picked up the “toehold” principle introduced by Zhang *et al.* and re-designed the riboregulator prototypes from Isaac *et al.*<sup>94</sup>. The combination of both principles turned out to be very powerful by means of dynamic ranges and high orthogonality between the different RNA parts. The taRNA, or trigger RNA, grasps a nucleotide stretch (toehold) of the mRNA upstream of the crRNA structure occluding the RBS and is thus able to efficiently unfold the crRNA and promote translation. A set of 168 toehold switches was designed using NUPACK, a suite of computational tools that enable the user to design RNA sequences that will fold to specified structures<sup>95</sup>. Extended use of NUPACK was then used to identify a set of 26 highly orthogonal switches that were further narrowed to 13 variants with dynamic ranges up to 400-fold. The high degree of orthogonality allowed the simultaneously regulation of 12 genes in the same cell. All discussed approaches targeted translation and were designed as ON switches. Recently, Chappell *et al.* engaged in the design of small transcription-activating RNAs (STARs) (Figure 2.2b)<sup>96</sup>. They rely on RNA hairpins that, if formed, terminate transcription. In analogy to the crRNA-taRNA pairs, trans-acting STAR antisense RNAs are used to reverse the hairpin formation by complementary hybridization. STARs were also shown to exhibit high orthogonality and high dynamic ranges up to 94-fold. The fusion to each other and to transcriptional repressors extended their inherent activating functionality by an OFF switch characteristic allowing to build RNA-only logic devices and genetic circuits. Additionally, STAR design rules were inferred, facilitating their *de novo* construction and paving the way for the application of recently developed libraries of transcriptional terminators.

Summarizing the class of riboregulators, it should be noted that they are insensitive to external stimuli. This requirement must in turn be met by the promoter driving mRNA expression. As demonstrated some years ago with a riboregulator-aptamer pair, it may only be a matter of time until sensor domains are incorporated into the new and refined toehold regulators to enable a fully independent functionality of these riboregulators from other

## Introduction

genetic parts like promoters. However, as they are currently only applicable to bacteria, their portability to eukaryotes has yet to be investigated.



**Figure 2.2 New and refined synthetic RNA regulatory mechanisms of gene expression.** (a) Riboregulators termed toe-hold switches control translation in bacteria. The crRNA occludes the RBS and inhibits ribosome binding. The trigger RNA invades the hairpin structure and liberates the RBS, switching gene expression from OFF to ON. (b) Riboregulators termed STARs exploit the natural transcription termination mechanism in bacteria. The STAR antisense RNA unfolds the terminator hairpin and transcription continues, resulting in an ON state of gene expression. (c) CRISPR-based control over gene expression is constituted by CRISPRi (blockade of transcription by gRNA/Cas9), TFs fused to Cas9 (activation or inhibition of transcription) and tandem constructs of an aptamer and a gRNA (triggered transcription activation or inhibition by external inputs). Riboregulators and CRISPR tools exhibit high dynamic ranges and orthogonality. Graph adopted from <sup>76</sup>.

Turning to the second major class of novel RNA regulatory mechanisms, the versatility of RNA itself is impressively demonstrated by its marriage with functional protein components (Figure 2.2c). Here, the bacterial immune system CRISPR, originally discovered in *Streptococcus pyogenes*, serves as a highly adaptive starting platform <sup>62</sup>. Essentially, it consists of an RNA component, the small gRNA, and a protein component, the endonuclease Cas9. Evolved to defend the bacterial cell against foreign phage DNA, the gRNA specifically targets complementary DNA sequences next to the defined proto-spacer adjacent motif (PAM), directing endonucleolytic degradation by Cas9. Taking full advantage of this system, CRISPR interference was developed as a mechanism for the knockdown of target genes in bacteria <sup>28,29,97</sup>. The catalytic activity of Cas9 can either be deleted or replaced and thus gave rise to a multitude of heterologous protein fusions. These can subsequently be combined with re-programmed gRNAs to two-component mini-systems. Figure 2.2c highlights the use of CRISPRi to block transcription by a dead Cas9 <sup>28,29</sup>, the fusion of a transcription activating domain to Cas9 promoting transcription in eukaryotes <sup>63,64</sup> and the elegant combination of an aptamer with the gRNA

## Introduction

facilitating the sensitivity to external inputs <sup>59</sup>. In other studies the system was further expanded for optogenetic control mechanisms <sup>98,99</sup> and in an *in situ* imaging platform <sup>100</sup>. The regulatory efficiency achieved by CRISPR-based tools far exceeded those obtained with common RNA regulation alone. Only the toehold switches previously described exhibited comparable activity of up to several hundred-fold changes between ON and OFF states.

Both, riboregulators and CRISPR-engineered tools, offer high degrees of modular flexibility and applicability to different genetic contexts. Their regulatory power has set new standards for RNA-based gene expression control. Advances in computational engineering have allowed their *de novo* design and implementation mainly in bacteria, but also in yeast and mammalian cells. A current drawback of especially riboregulators is their limited ability to respond to external stimuli. While this feature might be the next to be added in near future, it has always been inherent to another class of RNA regulators – riboswitches. Riboswitches uniquely unite the ability to sense environmental signals and react *in cis* by a corresponding modulation of gene expression.

### 2.3.2 Natural riboswitches: Occurrence, structure and function

The importance of natural RNA regulation and the feasibility of RNA mimicry has been mostly neglected until the late discovery of such regulatory elements in prokaryotes in 2002 <sup>65,101</sup>. The term riboswitch unites all RNA-based regulatory elements that are sensitive to an external input, typically a cellular metabolite, and transmit this input signal by a rearrangement of their global architecture into an altered expression of the controlled gene. Since the sequences of sensor and actuator are subdomains of a mutual architecture, riboswitches constitute an ideal scaffold to design synthetic derivatives with different ligand specificities and actuation profiles that operate *in cis*.

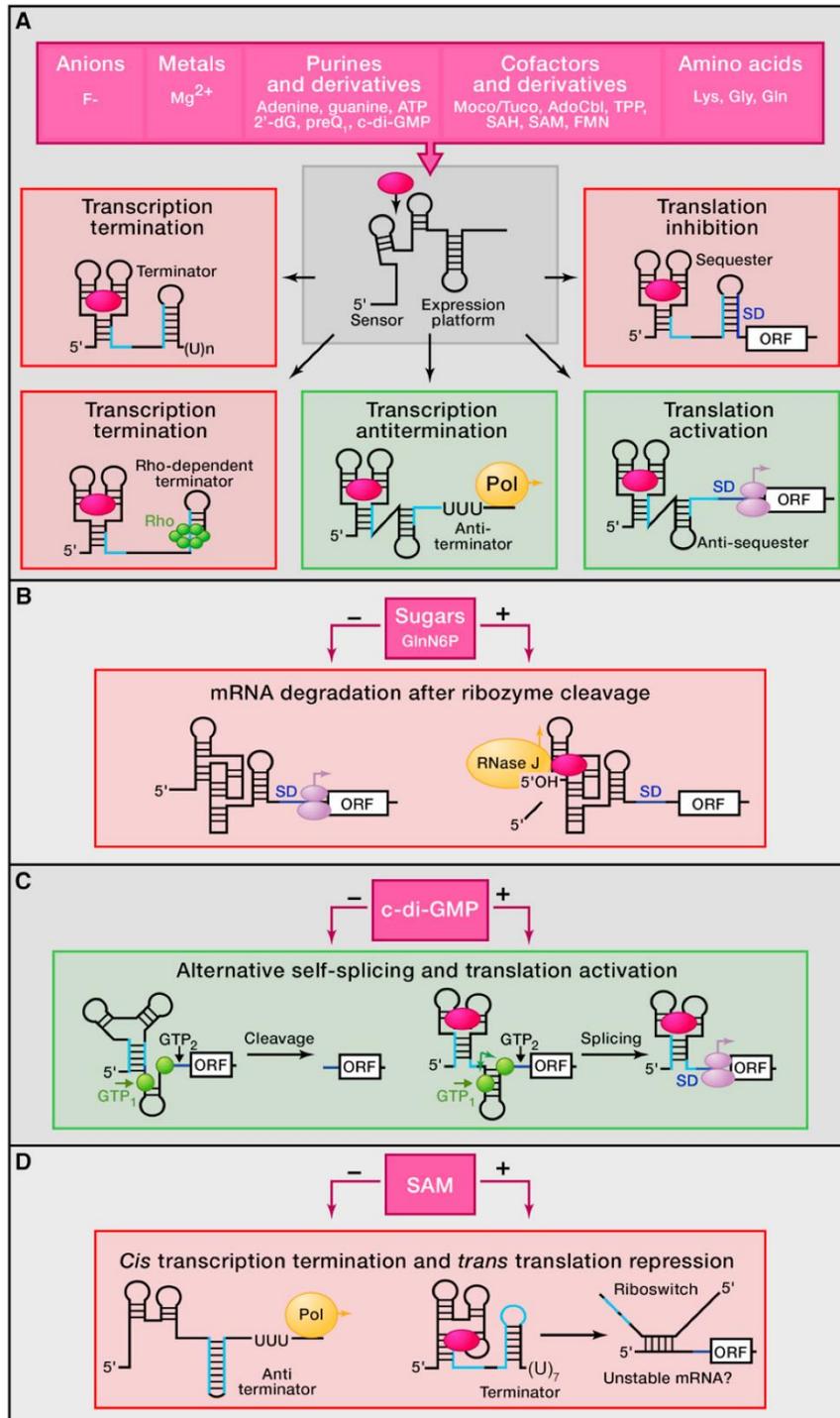
Riboswitches have been discovered in all kingdoms of life. Initially, researchers tried to elucidate the regulatory mechanism of vitamin B<sub>1</sub>-, B<sub>2</sub>-, and B<sub>12</sub>-biosynthesis. After having searched in vain for protein-based feedback control, it was speculated on the involvement of conserved mRNA sequences to act as regulators. Eventually, the direct interactions of three vitamin derivatives with their respective mRNAs were shown and confirmed the regulatory role of RNA in metabolite biosynthesis <sup>101,102</sup>.

In bacteria, the vast majority of riboswitches were found to exclusively control either transcription termination or translation initiation. Their regulatory function is for instance embedded in the control of metabolic pathways of cellular substrates as vitamins, amino acids and nucleotide analogs (Figure 2.3) <sup>65,103</sup>. Here, they contribute to a stable steady state of educts and products consumed and synthesized during the biogenesis of a particular metabolite, thus controlling its flux and abundance. Despite the limited set of four nt as building blocks, the sensor domain, also called aptamer, has evolved to sense many different ligands (Figure 2.3A). While this highly specific ligand-binding site has remained conserved throughout evolution, the actuating domain, also termed expression platform, exhibits greater sequence variability to control the downstream coding sequence. The received external signal is transduced by a conformational change of the aptamer, in turn leading to conformational rearrangements of the expression platform. In this way the ligand triggers the “riboswitching” between two mutually exclusive conformations of the expression platform oppositely affecting gene expression. Transcription controlling riboswitches typically flip between Rho-independent terminator and anti-terminator hairpins, causing the RNA Polymerase II (RNA Pol II) to stall and fall off or continue mRNA elongation, respectively. Recently, also Rho-dependent transcriptional termination has been observed and appears to be a

## Introduction

widespread mechanism as some riboswitches lack Rho-independent terminators or RBS-sequestering hairpins<sup>104</sup>. RBS occlusion prevents the 70S ribosome from accessing the translation initiation site. In a similar manner to terminator-anti-terminator switching, ligand binding releases or obscures the RBS and translation is turned ON or OFF.

A third mechanism found in bacteria relies on ligand induced cleavage within the 5'-UTR of the mRNA<sup>105</sup>. These riboswitches are composed of an aptamer and a ribozyme (also termed aptazyme) and facilitate the endonucleolytic degradation by RNase J from the resulting 5'-OH end turning gene expression OFF at the translational level<sup>106</sup>.



**Figure 2.3 Regulatory mechanisms of riboswitches in bacteria.**

Prokaryotes feature various riboswitching mechanisms to regulate gene expression involving transcription, translation, splicing and mRNA stability.

**A** Different metabolites (pink) evoke inhibition (red insets) or activation (green insets) of transcription (left, middle) or translation (right) by structural rearrangements of sensor and expression platform (grey inset). The alternatively folded terminator-anti-terminator sequences are highlighted in light blue. Shine-Dalgarno (SD) constitutes the RBS and is colored dark blue. RNA Pol II (Pol), open reading frame (ORF), transcription termination complex Rho and the 70S ribosome (light purple) are indicated. 2'-dG, 2'-deoxyguanosine; preQ<sub>1</sub>, pre-queuosine-1; Moco/Tuco, molybdenum and tungsten cofactors; SAH, S-adenosyl-L-homocysteine.

**B** Glucosamine-6-phosphate (GlnN6P) induces the endonucleolytic degradation of the target mRNA by RNase J after cleavage of the mRNA by the riboswitch.

**C** the cyclic di-guanosyl-5'-monophosphate (c-di-GMP) riboswitch controls alternative self-splicing of the target mRNA and confers the fusion of distant RBS parts to a functional unit, promoting translation.

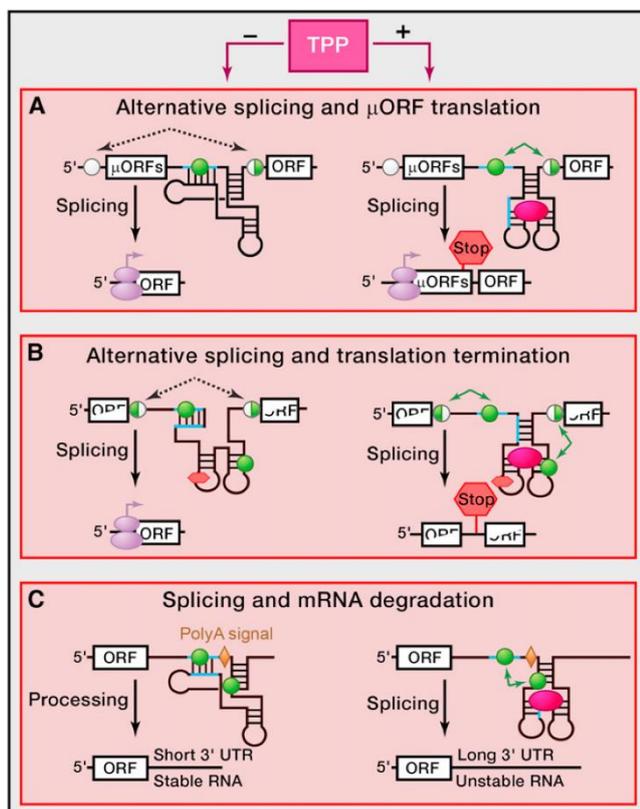
**D** S-adenosyl-methionine (SAM) serves as ligand for transcription termination *in cis* and translation repression *in trans*.

Adopted from<sup>103</sup>.

## Introduction

The only member currently known, the *glmS* riboswitch, senses glucosamine-6-phosphate and related compounds and thus confers regulation over a wide array of metabolic states (Figure 2.3B) <sup>107</sup>. Another rather unconventional mechanism is exerted by the cyclic di-guanosyl-5'-monophosphate (c-di-GMP) riboswitch, that is essentially involved in self-splicing of group I introns to join to distantly located parts of the RBS, thus acting on translation (Figure 2.3C) <sup>108</sup>. Typically acting *in cis*, the described mechanisms were complemented by the discovery of a *trans*-regulating riboswitch sensitive to S-adenosylmethionine (SAM). The SAM riboswitch confers *cis*-regulation over transcription termination of its downstream gene and subsequently functions as a small RNA to *trans*-repress translation of a distant gene (Figure 2.3D) <sup>109</sup>.

In eukaryotes the validated and predominant example for a riboswitching RNA regulatory mechanism controls splicing (Figure 2.4). More precisely, the thiamine pyro-phosphate (TPP) riboswitch controls alternative splicing, i.e. the rejoining of exons in different combinations after exclusion of the intron. TPP presence above a certain threshold triggers the liberation of a previously sequestered splice-site that can now be used to generate a different gene product. Depending on the species, the activation of the TPP riboswitch causes the synthesis of aberrant peptides (fungi, Figure 2.4A) <sup>110</sup>, premature translation termination (algae, Figure 2.4B) <sup>111</sup> or destabilized transcripts (plants, Figure 2.4C) <sup>112</sup>.



**Figure 2.4 The TPP riboswitch in eukaryotes.**

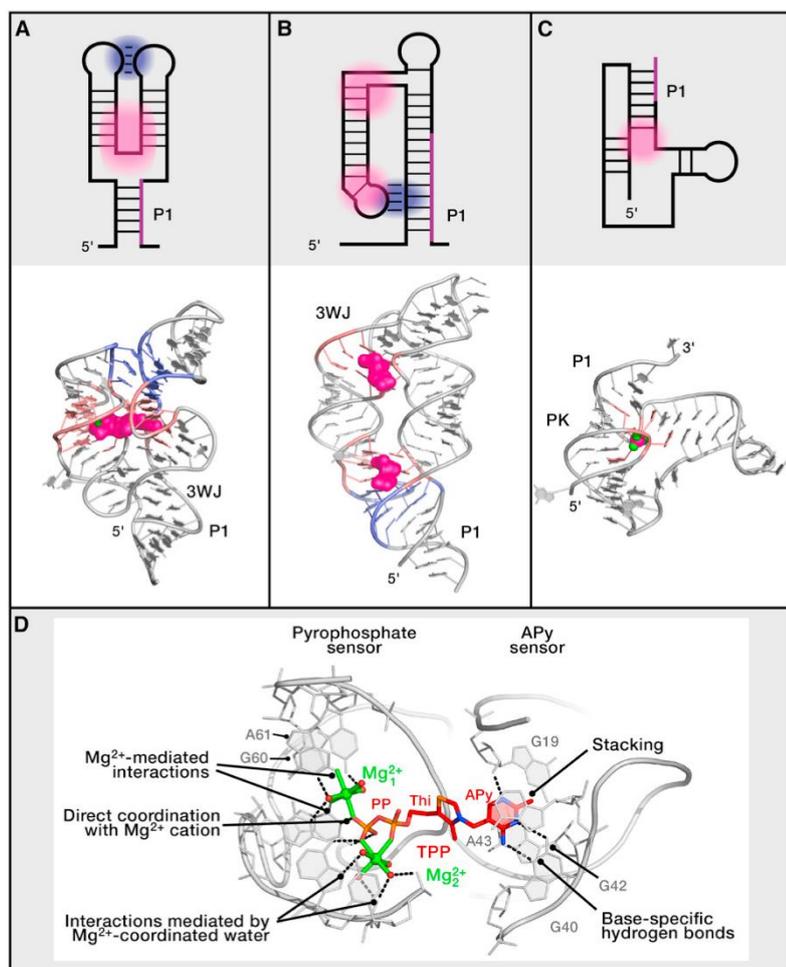
This riboswitch is heavily involved in (alternative) splicing. TPP (pink) presence above a certain threshold triggers the liberation of a previously sequestered splice-site that can then be used to generate a different gene product. Depending on the species, the activation of the TPP riboswitch causes the synthesis of aberrant peptides in fungi by inclusion of the micro open reading frame ( $\mu$ ORF) and a premature stop codon (A), premature translation termination in algae by inclusion of a stop codon (B) or destabilized transcripts in plants by removal of the poly adenylation signal (yellow rhomb) (C). Green circles and arrows illustrate the alternative splice sites. White circles represent the splice sites used in absence of TPP. Orange rhombs in B and missing in A show the translational stop codon. The 70S ribosome is colored in light purple. Adopted from <sup>103</sup>.

Riboswitches realize their exceptional ability to sense a specific ligand by a highly adapted set of different architectures. Practically all major riboswitch classes have been thoroughly investigated and many metabolite-sensing domains are thus available at atomic resolution. As proposed elsewhere, riboswitches may be classified according to their structural principles to accommodate the ligand and confer signal transduction <sup>103,113</sup>. Thus, the first class is composed of riboswitches constructed from multihelical junctions (Figure 2.5A, B) and the second class is engaged in pseudoknot formations (Figure 2.5C). Pseudoknots, or knot-shaped conformations,

## Introduction

are usually formed by a stack of two helices connected by short loops that may involve extra elements like hairpins. Pseudoknot-governed riboswitches as the SAM-II and fluoride switches appear rather compact in structure and pocket the ligand at the interface of the knot-like junction <sup>114,115</sup>.

In contrast, multihelical junctions can be constituted by Y-shaped, long-distance interactions through kissing loops that enclose the ligand sensing domain and stabilize the P1 stem that is part of the expression platform of the TPP or purine riboswitches (Figure 2.5A) <sup>116,117</sup>. Alternatively, long-range, back-folded interactions between a loop of a hairpin and a distant helix determine the overall conformation of riboswitches as that of the THF riboswitch binding to two molecules THF simultaneously, eventually affecting the folding of the P1 stem (Figure 2.5B) <sup>118</sup>.



**Figure 2.5 Common riboswitch architectures and ligand recognition.**

Folding regimes are classified according to participation of multihelical junctions or pseudoknots.

**A, B** Multihelical junctions can either be Y-shaped and contain kissing loop interactions (**A**) or formed by long-range, back-folded interactions (**B**). The ligand is enclosed the junction or incorporated into tertiary interactions of long hairpins.

**C** pseudoknots, or knot-shaped conformations, are usually formed by a stack of two helices connected by short loops that may involve extra elements like hairpins. Pseudoknot-governed riboswitches as the SAM-II and fluoride switches appear rather compact in structure and pocket the ligand at the interface of the knot-like junction. Ligand binding sites are highlighted in pink, long-distance interactions in blue.

**D** RNA-ligand interactions at molecular resolution. Specific hydrogen-bonding, electrostatic and  $\pi$ - $\pi$  stacking interactions as well as cation- or water-mediated ligand coordination facilitate specific recognition of the cognate ligand in concert with global shape complementarity.

Graph adopted from <sup>103</sup>.

The modular constitution of binding, transmitting and actuating motifs leads to the recurrence of several motifs among different riboswitches and allowed even the inverse phenomenon of different motif configurations that recognize the same ligand (SAM-riboswitches) <sup>103,115,119</sup>. This underlines the extraordinary ability of RNA to encompass virtually any conceivable combination of RNA structure-ligand pairs that is only further broadened by the chemistry of these interactions. Metabolites are not identified by a common recognition feature, but rather by adapted recognition patterns (Figure 2.5D). A dominant mechanism to ensure ligand compatibility is shape complementarity that facilitates the tight encapsulation and correct positioning of the cognate ligand and often very closely related compounds and blocks non-cognate ligands by steric hindrance. These ligand pockets are equipped with conserved nucleotides forming specific hydrogen bonds or electrostatic interactions to

heteroatoms on the edges of the ligand. Planar ligands or ligand moieties are sandwiched between purines in  $\pi$ - $\pi$  stacking interactions. A common feature is the occurrence of mono- or divalent metal ions such as  $K^+$  or  $Mg^{2+}$  compensating negatively charged functional groups (phosphate, carboxylate or fluoride) or mediating direct and water-dependent coordination of RNA-ligand interactions.

The global mechanism of ligand-RNA interaction is governed by a combination of conformational capture and induced fit. The aptamer and expression platform fluctuate between different ligand-free ground states and a ligand-bound conformation. The ligand captures only the ligand-bound conformation and induces minor or major conformational changes dependent on the riboswitch. Whereas translational riboswitches are typically in thermodynamic equilibrium, i.e. ligand binding shifts the equilibrium towards the bound conformation, transcriptional riboswitches are never allowed to reach equilibrium as ligand binding triggers a once-in-a-lifetime decision on whether the terminator or anti-terminator is formed. The ligand kinetically traps a metastable conformation of the aptamer, seizing decision-making on the overall folding instead of the minimum free energy (MFE) <sup>65</sup>. Both mechanisms are exploited in approaches to re-engineer riboswitches according to synthetic designs.

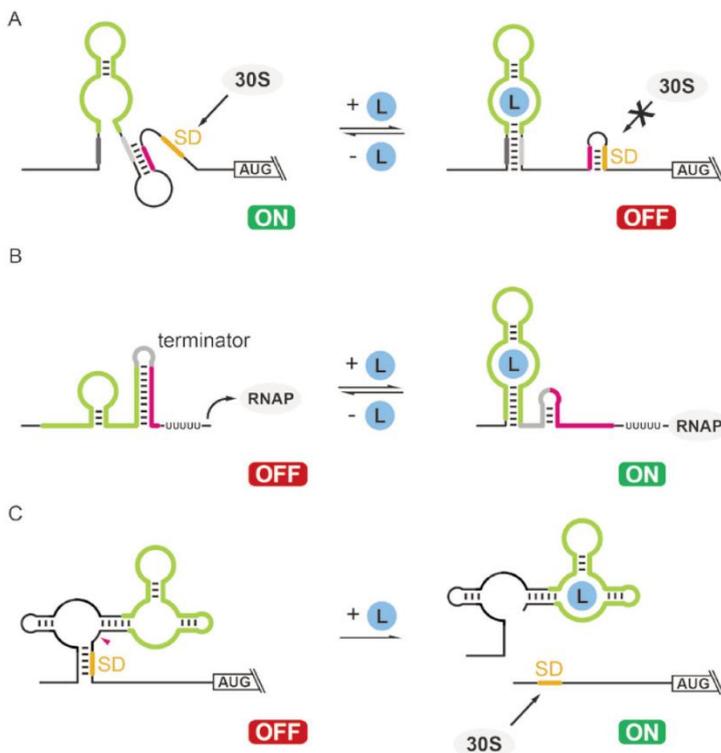
### 2.3.3 Synthetic riboswitches

The design and implementation of riboswitches presents a unique opportunity to manipulate any reporter device *in cis*, executing tight temporal and spatial control at low metabolic costs. Their nucleotide composition can be configured with ease, facilitating access to a multitude of different convertible RNA architectures triggered by virtually any ligand of choice. This short abstract highlights again the important differences between synthetic riboswitches and the previously discussed synthetic riboregulators and emphasizes their rich and highly functional heritage from their natural precursors. Their property to respond to external stimuli *in cis* and their modular composition constitute the key features that make their engineering amenable to self-contained plug-and-play approaches. Practically all such studies may be summarized as proof-of-principle experiments, in which the feasibility and applicability of the constructed synthetic parts and devices was demonstrated in at first trial-and-error style and later enhanced by rational design principles. On the road to true plug-and-play designs still many hurdles must be overcome <sup>78,86,87</sup>.

Synthetic riboswitching mechanism implemented in bacteria essentially mimics those of their natural counterparts (Figure 2.6). Many studies have capitalized on the seemingly straightforward modification of the mRNA transcript.

Translation repression was foremost achieved by the insertion of the theophylline aptamer flanked by some randomized nucleotides to generate an expression platform that allows the alternative hybridization with the RBS (Figure 2.6A) <sup>120,121</sup>. Addition of theophylline captures the ligand bound conformation, engaging the synthetic expression platform in the formation of the aptamer and a hairpin sequestering the RBS. Screening for regulatory active riboswitches yielded switches with dynamic ranges of up to 11-fold. These riboswitches were placed upstream of the *cheZ* gene to control migration and chemotaxis *E. coli* <sup>122</sup>. In other follow-up studies the broad and robust applicability and portability of these riboswitches to diverse bacterial species was demonstrated with regulatory activities of up to 200-fold <sup>123,124</sup>. Theophylline-sensitive aptamers were also designed to exhibit complementarity to slightly different regulatory mechanisms, such as helix slipping and relocations of the SD sequence <sup>125,126</sup>.

## Introduction



**Figure 2.6 Synthetic riboswitching mechanisms in prokaryotes.**

**A** Translation is controlled by the insertion of aptamers with synthetic expression platform in the 5'-UTR of a reporter gene. In the absence of the ligand the 70S can bind to the RBS and start translation. Ligand presence promotes the rearrangement of the expression platform by a stabilization of the aptamer, the RBS is sequestered and translation inhibited.

**B** Transcription termination is regulated by two competing hairpins, the terminator and anti-terminator. Ligand binding triggers the formation of the anti-terminator and transcription continues.

**C** The introduction of aptazymes (aptamer + ribozyme) facilitates ligand dependent control over RBS accessibility. Here, ligand addition stabilizes the conformation of the aptazyme and thus activates its self-cleavage and the RBS is released. Adopted from <sup>78</sup>.

Riboswitch designs harnessing transcription termination have been provided during the past years (mechanism exemplified in (Figure 2.6B)). Again, the theophylline aptamer served as a prototype in a *de novo* design of transcriptional ON switches <sup>127</sup>. Followed by a hairpin complementary to the aptamer and a poly-U stretch, these riboswitches resembled their natural counterparts and could elevate reporter expression by 6-fold upon theophylline addition. A true precursor mimicry was designed by the coupling of the *in vitro* selected theophylline and tetracycline aptamers and the natural purine, SAM, FMN and lysine aptamers to three natural expression platforms <sup>128,129</sup>. Interestingly, the design avoided the insertion of a transmitting module to connect aptamer and expression platform allowing the re-engineering of natural riboswitches without extensive screening steps. Further re-designs of natural purine riboswitches facilitated ligand specificities orthogonal to the cell and yielded translational and transcriptional switches <sup>130</sup>.

Figure 2.6C depicts a rare natural regulatory mechanism that has nonetheless been proven to be very effective in synthetic designs. In contrast to synthetic translational OFF switches, the use of aptazymes allows the construction of translational ON switches. As mentioned earlier, aptazymes are composed of an aptamer domain fused via a communication module to a ribozyme. In bacteria, mainly the theophylline- and TPP-responsive hammerhead ribozymes (HHR) have been used to control downstream reporter expression <sup>131</sup>, tRNA abundance <sup>132,133</sup> and 16S rRNA availability <sup>134</sup>. Another approach combined the taRNAs known from the class of riboregulators with the HHR to yield small RNA triggered mRNA cleavage and a activation of gene expression <sup>135</sup>.

## Introduction

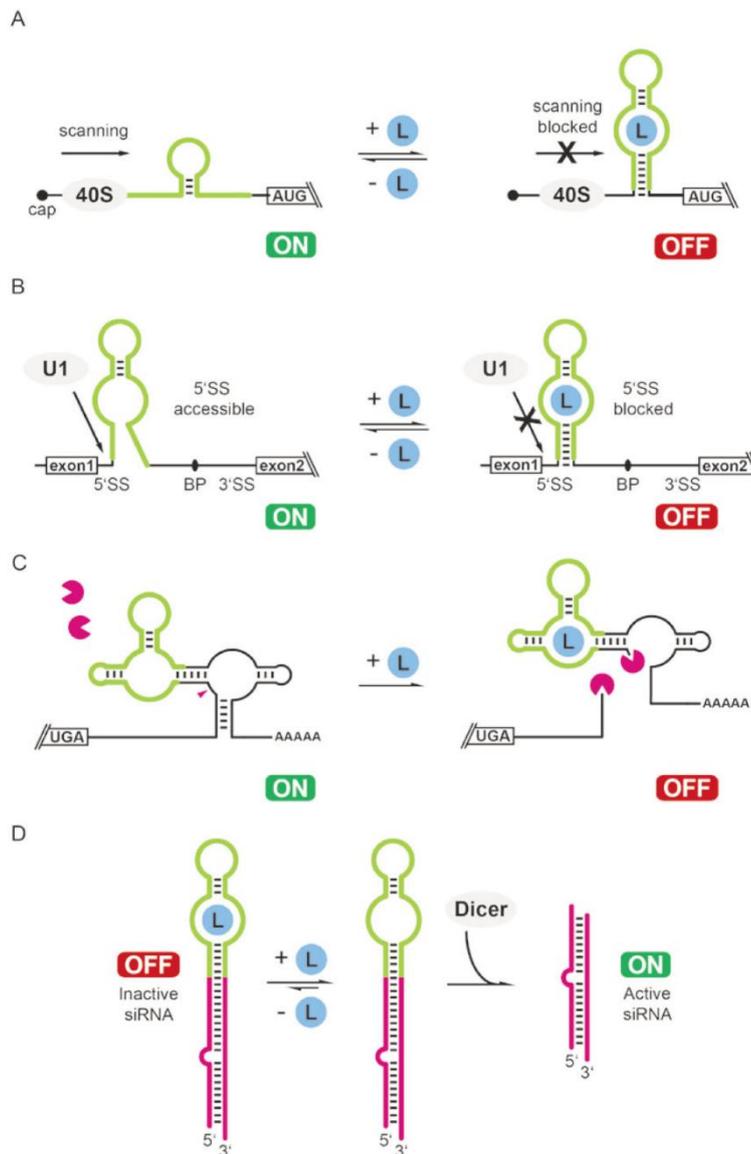
In eukaryotes the focus shifts towards the development of regulatory mechanisms for translation, splicing and RNA interference. One of the first mechanisms mimicked translational control in bacteria even before the actual riboswitching mechanisms had been discovered. *In vitro* generated aptamers were set in the 5'-UTR upstream of different reporter genes in baker's yeast and Chinese hamster ovary cells and used to control a downstream gene (Figure 2.7A). The aptamer interferes with the scanning of the small 43S pre-initiation complex for the start codon. In the absence of the ligand the aptamer is largely unstructured. Presence of the ligand stabilizes the aptamer folding and suppresses 43S scanning. The mechanism is discussed in more detail in 4.1.2. The ligands, a Hoechst dye <sup>136</sup> and tetramethylrosamine <sup>137</sup>, were shown to downregulate gene expression up to 10-fold. Later, aptamers binding to tetracycline and neomycin were demonstrated to confer similar regulation <sup>138,139</sup>. As both aptamers are subject to this study and have been extensively characterized in follow-up experiments, they are described in chapter 2.3.3.2. A different design was brought to eukaryotes by the concept of *trans-activating* RNAs and generated programmable ON and OFF "antiswitches" by use of the theophylline aptamer in yeast <sup>17</sup>. The ligand triggers the formation of a trans-acting antiswitch flanked by an antisense strand complementary to a sequence in the 5'-UTR of a GFP reporter gene. Binding of the antisense strand inhibits gene expression.

Figure 2.7B illustrates the introduction of synthetic control over splicing. In analogy to the TPP riboswitch, the theophylline- and tetracycline-binding aptamers have been used to control constitutive and alternative splicing by insertion of the aptamers close to either the 5'- or 3'-splice site or the internal branch point in yeast and mammalian cells <sup>140-142</sup>.

The regulation of mRNA stability constitutes an efficient possibility to interfere with gene expression. In eukaryotes mRNA stability is strongly influenced by the integrity of the poly-A tail. Insertion of aptazymes into the 3'-UTR of reporter genes facilitates direct control over mRNA fate by immediate cleavage of the target mRNA in absence or presence of the cognate ligand (Figure 2.7C). The communication module between aptamer and ribozyme allows the selection of ON and OFF switches, i.e. inhibition or activation of mRNA cleavage, respectively. The HHR presents the most frequently used ribozyme and later other ribozymes as the HDV <sup>143</sup> or twister ribozymes <sup>144</sup> were tested as well. Ligand sensitivities range from theophylline<sup>145,146</sup> and TPP <sup>144</sup> to tetracycline <sup>24,147,148</sup>, neomycin <sup>149</sup> and guanine <sup>143</sup>. Since the discovery of RNA interference had demonstrated the importance of posttranscriptional regulation in eukaryotes, the underlying mechanism was fully elucidated <sup>150</sup>. RNA interference (RNAi) is a conserved biological mechanism present in eukaryotes that uses small non-coding RNAs to post-transcriptionally modulate gene expression. The two classes of small RNAs, short interfering (si) RNAs and micro (mi) RNAs, play a key role in this process. Whereas miRNAs are encoded in the genome, transcribed as primary miRNAs by RNA Pol II and processed by the RNases III Drosha and Dicer, the siRNA precursors are of exogenous origin and cleaved by Dicer to produce active siRNAs. After processing, both 21 to 25 nt RNAs are transferred to the RNA induced silencing complex (RISC) to function as a guide targeting mRNAs for either translational repression or mRNA degradation. RNAi is an often used mechanism to sequence-specifically knock-down gene expression with synthetic siRNAs. The engineering of siRNA precursors (short hairpin, shRNAs) enable the ligand-dependent regulation of their processing. In this context the terminal loop structure of an shRNA was replaced with the theophylline-binding aptamer, generating a switchable shRNA precursor that facilitated the regulation of GFP expression <sup>151</sup>. The developed approach was further used to control the level of endogenously expressed human serum albumin in mammalian cells <sup>152</sup>.

## Introduction

In a different study ligand-dependent synthetic pri-miRNAs were generated by inserting various aptamers into the basal segment of the precursor. With this design Drosha processing was successfully targeted and it demonstrated the ligand-dependent regulation of GFP expression <sup>153</sup>.



**Figure 2.7 Synthetic riboswitching mechanisms in eukaryotes.**

**A** By inhibition of the scanning of the 43S (or 40S, if only the small ribosomal subunit is considered) pre-initiation complex for the start codon aptamers can regulate gene expression in the 5'-UTR of a target gene. Ligand addition stabilizes the aptamer folding and blocks scanning.

**B** Aptamers inserted close to either the 5'- or 3'-splice site or the internal branch point influence splicing in yeast and mammals by the accessibility of these sites to the spliceosome. Here, the intron is retained and prevents the fusion of exon1 and exon2.

**C** In eukaryotes aptazymes determine mRNA stability in the 3'-UTR by ligand-induced cleavage or maintenance of the poly-A tail. Here, the ligand stabilizes the active conformation of the aptazyme and the poly-A tail is cleaved off triggering rapid mRNA degradation.

**D** RNA interference. The replacement of the terminal loop of the precursor of a small non-coding RNA with a ligand (L)-sensing domain (green) results in an engineered, ligand-controllable hybrid molecule. In the presence of the cognate ligand processing of the modified precursor is blocked by the formation of a stable aptamer structure, preventing Dicer from substrate cleavage. In the absence of ligand the precursor is processed to its mature and active variant by Dicer. Adopted from <sup>78</sup>.

### 2.3.3.1 Design and engineering of synthetic riboswitches

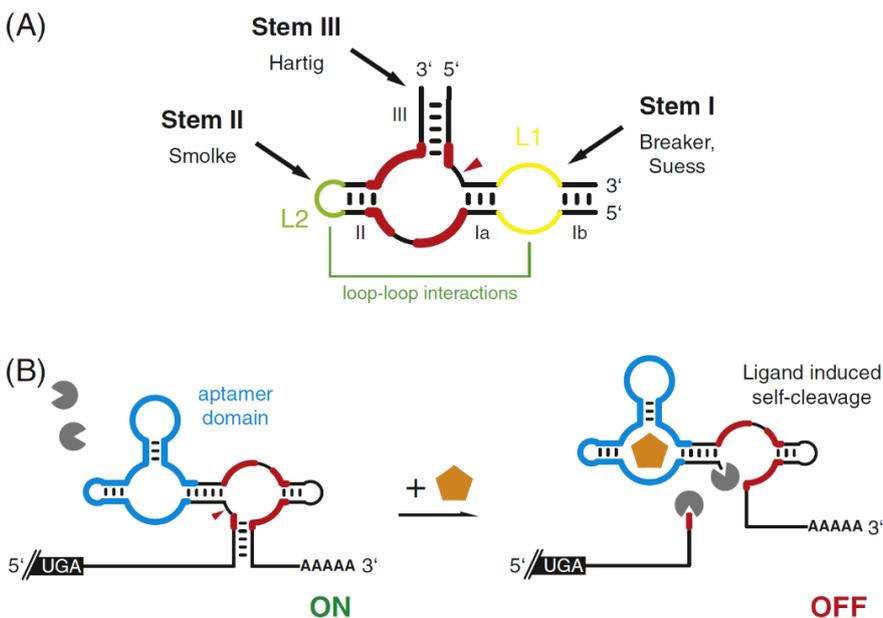
The means by which *in vivo* active synthetic riboswitches have been generated until today can be narrowed down to two fundamental strategies. In general, the sensor domains, i.e. RNA aptamers, are either obtained by an *in vitro* selection process called Systematic Evolution of Ligands by Exponential enrichment (SELEX) or simply by using (modified) natural aptamers. Now, to turn them into riboswitches, they must be connected to some kind of expression platform. Depending on the organism and regulatory mechanism sought to implement, the aptamer, the connecting module and/or the expression platform itself are adapted to exhibit a concerted functionality. Following the first strategy, this is done by randomization and extensive screening and selection steps for the desired phenotypes. The second strategy has evolved over the recent years and is constituted by

## Introduction

rational *de novo* designs facilitated by major advances in the understanding of RNA folding, riboswitch kinetics and thermodynamics and sequence-structure-activity relationships. Among both strategies, several approaches were successfully pursued and are further detailed.

Aptamers for non-natural ligands are selected *de novo* by SELEX (for a short description on the process see chapter 5.1.2.1.1). Although the technology is powerful, only about 60 aptamers have been selected to bind small molecules and only a handful of these has found application as sensor domains of riboswitches<sup>77,78</sup>. A major drawback is associated with the experimental conditions of the selection protocol and the actual goal of the selection. Typically, SELEX is performed to obtain aptamers with maximum affinity to their ligands and use them solely as sensors. The applied selection pressure leaves little room for RNA sequences that exhibit different ligand-free conformations that are highly favored over the ligand-bound conformation that is unfortunately the only conformation trapped by the ligand and retained. Thus, the stringency inherent to the selection excludes aptamers with potential regulatory properties (this phenomenon is further discussed in chapter 4.1.2). Additionally, *in vivo* features as co-transcriptional folding or the occurrence of RNA-binding proteins cannot be simulated or concentrations of crucial cations such as Mg<sup>2+</sup> or K<sup>+</sup> are present in up to 10-fold excess, distorting binding constants. Solving of the latter issue was demonstrated to alleviate the impact of the SELEX protocol on the identification of regulatory active aptamers<sup>154</sup>. Over the years different SELEX approaches have been designed and include *in vivo* SELEX and efforts to enhance the method as such reducing cost, labor, time and failure rate. The combination of SELEX with Next-Generation-Sequencing (NGS) constitutes a milestone in the improvement of the method as now data on each selection and amplification step is accessible, giving insights into enriched and lost sequence and structure motifs (reviewed in<sup>77</sup>).

SELEX-derived aptamers have found application in bacteria, yeast and mammalian cells<sup>33</sup>. The aptamer for theophylline turned out to be very robust, versatile and portable. As indicated in the previous chapter numerous riboswitches respond to this ligand.



**Figure 2.8 Aptazymes as riboswitches.**

**A** The *Schistosoma mansonii* HHR consists of a bulged three-way junction and three corresponding stems. Different groups attached aptamers to different stems. In its tertiary conformation loop-loop interactions between stem II and I enable the formation of the conserved catalytic core (red parts) cleaving off the 5'-stretch of stem III (red arrow). **B** Regulatory mechanism of aptazyme cleavage. Placed in the 3'-UTR in eukaryotes or 5'-UTR in prokaryotes, aptazymes control mRNA stability by cleavage of the poly-A tail or liberation of the RBS, respectively. Both, ON->OFF and OFF->ON switches, have been constructed. Adopted from<sup>87</sup>.

## Introduction

To make broader use of natural or *in vitro* generated aptamers they are fused to a communication module that in turn is connected to a ribozyme serving as expression platform. The resulting aptazyme is thus composed of three interdependent modules. To produce a working riboswitch nucleotides of the aptamer, parts of the closing stem P1, and the ribozyme, here parts of stem I, II or III of the HHR, are usually randomized. The resulting sequence pool is then inserted in the 5'-UTR (prokaryotes) or 3'-UTR (eukaryotes) of a reporter gene and subsequent *in vivo* or *in vitro* screening and selection steps allow the identification of aptazymes with desired phenotypes. The HHR proved to be of exceptional value as all three stems are amenable for a fusion with an aptamer, either constraining or promoting the catalytic conformation of the HHR upon ligand binding (Figure 2.8A). Different screening and selection systems have been applied to yield ON and OFF switches for several ligands. Still, the only way to obtain a regulatory active aptazyme is to perform many iterative test cycles to find a suitable communication module, a resource-consuming endeavor that has not yet been replaced by an *in silico* method computing ligand-free and -bound conformations and predicting the resulting phenotype.

A similar cumbersome approach was established to generate the riboswitches for neomycin and tetracycline that control reporter expression by blocking ribosome scanning in *S. cerevisiae*<sup>138,155</sup>. Both aptamers were obtained by SELEX and only an extensive *in vivo* screening facilitated their application as riboswitches (extensively reviewed and discussed in the following chapter). Both riboswitches as such hold a special position, because they do actually not contain a separate expression platform. Their structure is very compact and gene expression is modulated by the stability of their global folding and especially that of the closing stem, the “mini-actuator”. Without ligand both riboswitches adopt structurally loose conformations that enable ribosome scanning. The presence of their cognate ligand captures a transient, highly stable conformation that inhibits translation. Although no explicit communication module is required here, the need for iterative screening prevails as again a ligand-free and -bound conformation must be identified that first promote and then suppress gene expression. Applied positive and negative selection pressure can aid in the discovery of a switching window, separating ON and OFF states to a maximum feasible level and minimizing the trade-offs between the requirement of loose and rigid conformations exhibited by a single RNA architecture<sup>149</sup>.

Recent studies demonstrated that aptamers that had already been implemented in active synthetic or natural riboswitches could be fused to natural expression platforms without the need for a communication module<sup>129,156</sup>. Controlling transcription, these chimeric riboswitches support the hypothesis that the ability to generate functional synthetic riboswitches is co-determined by the identity of the aptamer sequence and hence unmodified aptamers derived from standard SELEX display poor building blocks. Nonetheless, before the chimeric riboswitches could be coupled in a “mix-and-match” design, the natural expression platforms had to be decoupled from their original aptamers as natural riboswitches exhibit some overlap between sensor and actuator domains. Once decoupled by the introduction of a tried and tested sequence motif, the natural expression platforms accepted various aptamers and produced ON and OFF transcriptional switches. This approach equipped existing expression platforms with completely different aptamers to obtain new sensor domains.

Another study pursued practically the same goal, namely to confer new ligand specificities to existing expression platforms, but with a different engineering effort<sup>130</sup>. By chemical derivatization and genetic selection natural riboswitches were engineered to respond to derivatives of the formerly cognate ligands. Both approaches, the replacement of the complete aptamer and the modification of an existing aptamer addressed an important

## Introduction

bottleneck in the generation of novel synthetic riboswitches; the availability of only a limited number of ligands orthogonal to the cellular metabolism. Importantly, they circumvented the need to screen and select for riboswitching activity by the use of functional riboswitches from the start. Though the approaches were rationally designed, true *de novo* design of riboswitches regulating transcription and translation driven by *in silico* modeling was presented by others. As described, transcriptional control is achieved by the switching between mutually exclusive terminator and anti-terminator structures. During the riboswitch synthesis in bacteria the RNA Pol II is forced to halt specific sequence motifs. This pausing allows the aptamer to fold, bind the ligand and transmit the signal to the expression platform that in turn must have adopted the conformation for termination or anti-termination. The time-window available to this once-in-a-life-time-decision is dictated by RNA Pol II speed and the kinetics of RNA folding and ligand binding. The mechanistic basis for this sequential event have been unraveled for numerous natural riboswitches and can thus be applied to build computational models. Accounting for thermodynamic and kinetic aspects, the folding paths of synthetic riboswitches composed of the theophylline aptamer were calculated<sup>83,127</sup>. It was important that the model also considered the inverse folding problem, i.e. that any predicted sequence can adopt the conformation of the mutually exclusive structures of terminator or anti-terminator. Simulations yielded several transcriptional ON riboswitches in *E. coli*. During the *in vivo* testing optimizations were only possible by rational evaluation of the genetic context, information the *in silico* model had not provided. Besides context, also structure-stabilizing effects by ligand binding and the formation of pseudoknots were not computed. This highlights the vast number of contributing factors that must be considered in such *in silico* design.

The development of a statistical thermodynamic model in a recent study predicted the sequence-structure-function relationship for translation-regulating riboswitches<sup>84</sup>. Since this regulatory mechanism does supposedly predominantly rely on the thermodynamic equilibrium between ON and OFF states, kinetic parameters play a comparably negligible role. Based on six different natural and synthetic aptamers this model automatically computed designs for 62 riboswitches and demonstrated that computational models built, trained and calibrated on thermodynamic data provide facile access to the construction of functional riboswitches.

### 2.3.3.2 The neomycin- and tetracycline-binding riboswitches

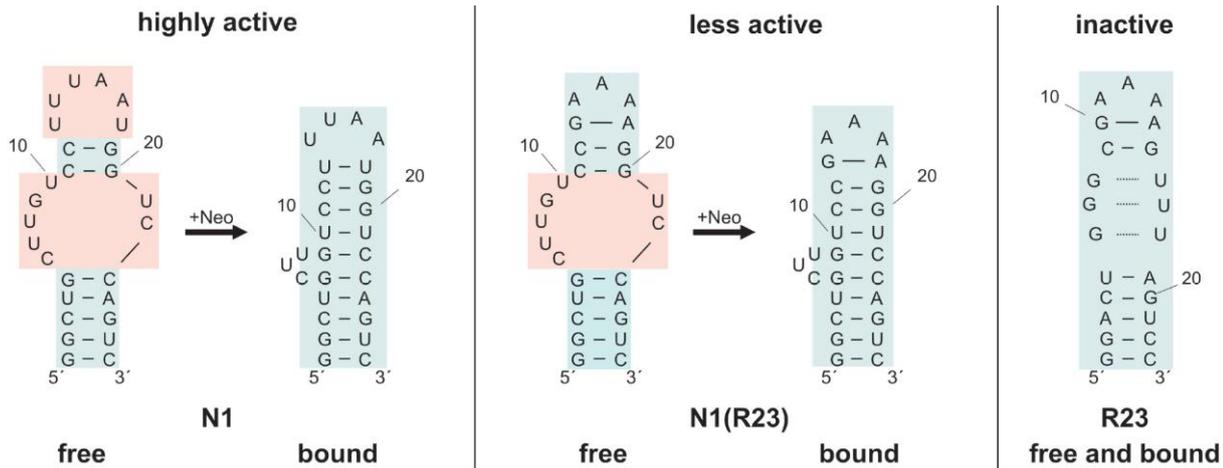
The riboswitches for neomycin (Neo) and tetracycline (Tc) applied in *S. cerevisiae* and are among the best studied synthetic riboswitches to date. Both original aptamers were obtained by SELEX<sup>157,158</sup>. While the Tc aptamer cb32 conferred 6-fold *in vivo* regulation of a *GFP+* reporter gene without further modifications<sup>155</sup>, the Neo aptamer N1 was not identified before an *in vivo* screening revealed its riboswitching properties and subsequent modifications yielded riboswitch M4 exhibiting 8-fold regulation<sup>138</sup>. Both riboswitches confer their regulatory activity in the 5'-UTR of the downstream gene. The as 5'-scanning interference described regulatory mechanism exploits the mechanism of eukaryotic translation initiation. The 40S small ribosomal subunit attaches to the 5'-m<sup>7</sup>GTP cap inherent to all correctly transcribed eukaryotic mRNAs<sup>159</sup>. After binding, eukaryotic initiation factors and other cofactors assemble to the 43S scanning-competent pre-initiation complex. During scanning for the start codon in 3'-direction the 43S complex eventually encounters the riboswitch inserted proximal to the mature start codon (Figure 2.10A). Now, the first key feature exhibited by regulatory active aptamers, the occurrence of loose, vastly unstructured ligand-free ground states, determines the ability of the 43S complex to continue scanning and initiate translation at the start codon. Neo and Tc riboswitch fulfill this

## Introduction

criterion as they allow ~40% and ~30% basal gene expression relative to riboswitch-unobstructed gene expression (see Figure 4.3A). The second key feature for regulatory activity is inherent to the aptamer structure and decides whether ligand addition will have any effect on downstream gene expression. To confer riboswitching activity the aptamer must feature a pre-formed ligand-free ground state that can be captured by the ligand and thus stabilized. The increased structural stability blocks the advancing 43S complex to a higher degree than the loose structure without ligand and represses gene expression. The following text describes the detailed mechanistic characteristics of both riboswitches.

The originally published neo aptamer R23, though exhibiting low nanomolar (nM) affinity and high selectivity for neo, was completely inactive *in vivo*. Thus, an *in vivo* screening of the enriched aptamer pool was conducted to search for regulatory active aptamers. The screening system is detailed in chapter 5.1.2.1. Briefly, the aptamer pool is inserted upstream of a plasmid-borne *GFP+* reporter gene and transformed yeast cells are screened for neo-dependent *GFP+* repression. This way 10 individual sequences were identified with the desired phenotype. Subsequent analysis of aptamer N1 revealed that its sequence was not among the aptamer sequences previously shown to bind neo, indicating that this aptamer was underrepresented in the enriched SELEX pool. Nonetheless, N1 binds in a 1:1 stoichiometry to neo with a dissociation constant of 8 nM and displays equal ligand selectivity. Structural probing data unraveled a putative ligand binding pocket and nucleotides mandatory for regulatory activity<sup>138</sup>. A detailed mechanistic insight into the sequences and structural motifs responsible for ligand accommodation and the switching phenotype could be gained by nuclear magnetic resonance (NMR) spectroscopy. The regulatory activity is governed by the occurrence of temperature-dependent ligand-free ground states that are in thermodynamic equilibrium. These ground states are constituted by an ensemble of open conformations that are entropically favored at ambient temperatures over a compact, highly structured conformation that closely resembles the ligand-bound state (Figure 2.9). The inactive aptamer R23 does not exhibit a regulatory activity, as it lacks certain structural elements that are a prerequisite to fluctuate between loose and stabilized conformations.

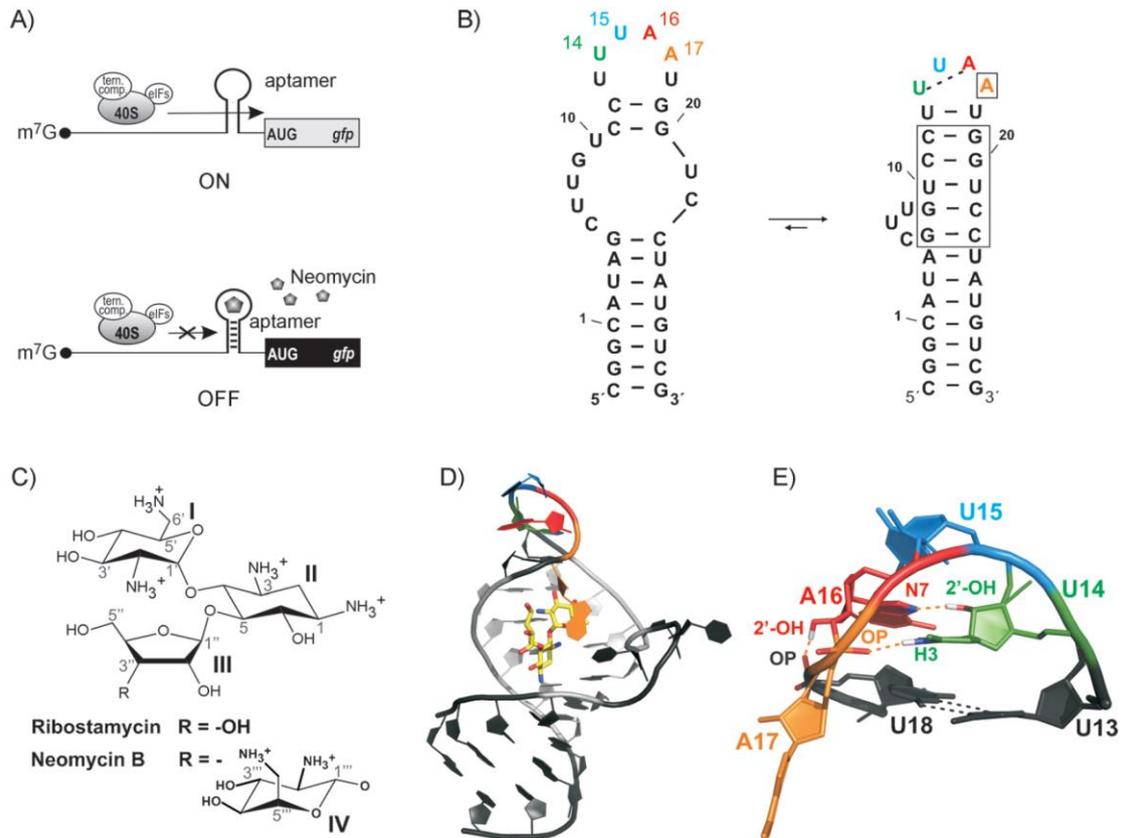
## Introduction



**Figure 2.9 Model of ligand binding and global refolding of the neo riboswitch.** The aptamer identified by *in vivo* screening (N1) exhibits riboswitching properties most likely due to stabilization of a ligand-bound conformation by neo, tightly clamping upper and lower helix together and promoting energy-releasing interactions of the terminal loop. The free conformation, that is highly favored in the absence of neo, fluctuates in a thermodynamic equilibrium with the bound conformation promoted by the switching element 5'-CUU. This interplay between loosely structured and subsequent stabilization confers *in vivo* activity. The aptamer R23 originally identified by SELEX is completely inactive as it lacks the necessary structural prerequisites to fluctuate between an open and closed conformation. Gene expression is always repressed to the same extent. The hybrid N1(R23) contains the asymmetric bulge including the switching element from N1, but the terminal loop of R23. Thus it may display diverse ground states and allow neo-dependent expression control, but since not all structural features are met, it is less active than N1. Unstructured parts are shaded light red.

In the N1 aptamer the folding toward the ligand-bound conformation is restrained by a short 5'-CUU-bulge motif that does not participate in ligand-binding, but promotes the separation of the closing stem and the upper mini hairpin, resulting in an open conformation. Ligand binding to a transient ligand-“bound” conformation stabilizes coaxial stacking of lower and upper helix, forming a long  $\alpha$ -helical element that sandwiches neo between G5:C23 from the lower helix and G9:C22 from the upper, newly formed helix. While the 5'-CUU motif, also termed the switching element, is bulged out, neo acts as a clamp and not only supports the continuous helix, but also promotes U10:U21 wobbling and stabilization of the apical loop by U13:U18 wobbling, stacking interactions between U15:A16 and hydrogen bonding between U14:A16 and U14:A17(P). A17 is looped out and undergoes hydrophobic interactions with neo (Figure 2.10B, D, E) <sup>160,161</sup>. This conformational capture of the preformed bound conformation from the conformational ensemble by neo results in the release of enthalpic energy that more than compensates the entropic penalty incurred by the transition from the entropically favored open, ligand-free conformation to the ligand-bound conformation. The switching element 5'-CUU is a key determinant to allow the ground state to be largely unstructured, hence promote translation in absence of neo. Presence of neo subsequently shifts the thermodynamic equilibrium of the ground states toward the highly structured, ligand-bound conformation that poses a roadblock to the scanning ribosome and ultimately prevents translation. As indicated by the ligand-stabilized interactions in the apical loop, this region was demonstrated to accommodate improving mutations on the *in vivo* activity of the M4 riboswitch, upgrading its regulatory performance from 8 to 10-fold <sup>162</sup>. This increase in performance can be attributed to an enrichment of the preformed conformation that is trapped by neo, because the apical loop exhibits pronounced involvement in the overall stability of the ligand-bound conformation.

## Introduction



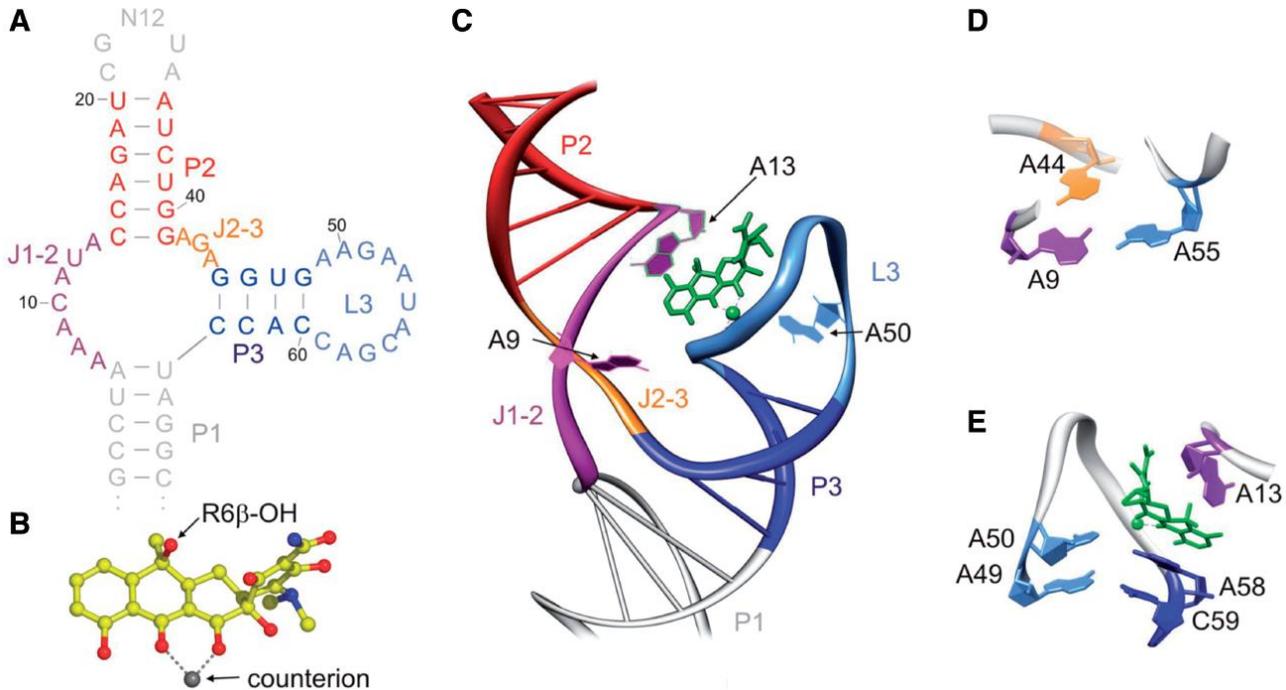
**Figure 2.10 Mechanistic insights into the neo riboswitch M4.** **A** The riboswitch is placed upstream and start codon-proximal of a *GFP* reporter gene. In the absence of neo, the scanning ribosomal complex is only partly constrained, whereas neo presence increases the roadblock and reduces scanning significantly. Ternary complex composed of eIF2, GTP, and Met-tRNA<sup>Met</sup> (tern. comp.); eIFs=eukaryotic initiation factors. **B** Secondary structure of the M4 riboswitch. The closing stem is composed of 8 bp and the terminal loop highlighted in colored letters. The thermodynamic equilibrium between the open and closed structure is indicated. Nucleotides involved in neo binding are boxed. The dashed line represents hydrogen bonding between U14 and A16. **C** Due to its superior spectral resolution, but nearly identical binding profile, ribostamycin is typically used for spectroscopic analyses. **D** NMR tertiary structure of ribostamycin bound to M4 (PDB ID: 2KXM). The continuous  $\alpha$ -helix is shaded in grey and nucleotides involved in ligand binding are light grey. Nucleotides of the terminal loops are color-coded and correspond to **B** and **E**. **E** Single nucleotide resolution of the terminal loop. The loop is closed by the non-canonical U13:U18 bp, forming a U-turn that consists of a loop out A17. Relevant hydrogen bonds are highlighted by dashed lines. Adopted from <sup>162</sup>.

The neo riboswitch has not only been used in the 5'-scanning mechanism, but was also fused to a HHR yielding highly functional aptazymes <sup>149</sup>. This further underlines that a riboswitch design is successful, if the aptamer domain has been shown to work in such a context.

The tc riboswitch was equally well characterized by using electron paramagnetic resonance (EPR) and fluorescence spectroscopy, since the fluorescence of tc can be used to determine thermodynamic and kinetic parameters. The affinity of aptamer and ligand was determined to 0.8 nM in a 1:1 stoichiometric ratio <sup>163</sup>. The aptamer assumes an h-shaped tertiary conformation, binding tc at the center of an irregular triple helix formed by the joining regions J1-2, J2-3 and the loop L3. (Figure 2.11A, C, D). Tc accommodation was demonstrated to be governed by a 2-step regimen <sup>164</sup>. The aptamer exists in two ligand-free ground states and, similar to the binding mode of neo, tc traps one of the two conformations in a first, reversible binding step. Initial association of the aptamer-tc complex is mediated by contacts between tc and the nucleotides A13, A58 and A50 that are mandatory for a fast binding kinetics and a stable conformation preventing the dissociation of tc (Figure 2.11E)

## Introduction

<sup>164</sup>. Since the conformation to which tc binds is already extensively preformed in the absence of tc at physiological  $Mg^{2+}$  concentrations ( $\sim 0.5\text{-}1\text{mM}$ ), no major conformational changes occur in the second binding step in which the J1-2 caps tc by stacking interaction with A13 and water- $Mg^{2+}$ -mediated hydrogen bonds of nucleotide A50 that had already coordinated tc in the first step (Figure 2.11E) <sup>164-166</sup>.

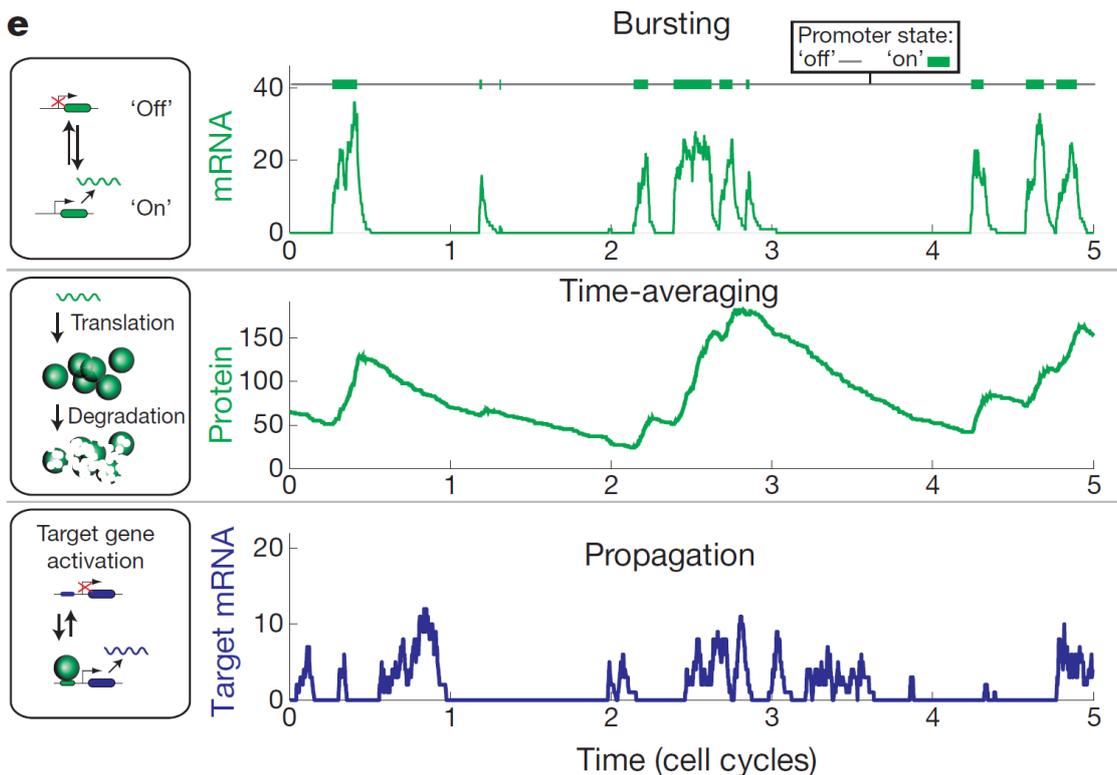


**Figure 2.11 The tc riboswitch AN32sh.** **A, C** Secondary structure and crystal structure <sup>167</sup>, respectively. Structural motifs are color-coded, tc is shown as stick-model and magnesium as green ball. The h-shaped tertiary conformation binding the tc- $Mg^{2+}$ -complex at the center of an irregular triple helix formed by the joining regions J1-2, J2-3 and the loop L3 is trapped in a 2-step binding mechanism by tc. **B** Ball and stick model of tc and magnesium emphasizing the coordination of the cation. **D** The basis of the triple helix if formed by A9, A44 and A55. **E** Nucleotides A13 and A50 are crucial for tc binding and stabilization of the bound conformation of the riboswitch. In absence of tc, especially A13 and the loop L3 are further spaced, destabilizing the tertiary conformation. Adopted from <sup>164</sup>

In accordance with the 2-step binding regime, translation can be effectively regulated with the tc aptamer. The thermodynamic equilibrium between the two ground states allows ribosome scanning, as tc is required to stably maintain the preformed conformation at physiological  $Mg^{2+}$  concentrations and thus to shift the thermodynamic equilibrium towards the bound conformation that massively interferes with translation. Hence, the tc riboswitch has found numerous applications for conditional gene expression. Approaches range from its introduction to the 5'-scanning mechanism as monomer, dimer or trimer over its use for controlling pre-mRNA splicing to its fusion to HHRs in different set ups. The multiple insertion of tc riboswitches in the 5'-UTR of a *GFP+* reporter gene even resulted in the amplification of regulatory activity up to 37-fold and was used to control essential genes in yeast <sup>168</sup>. As part of HHRs the tc aptamer conferred ON and OFF regulation in yeast and mammals <sup>24,147,169</sup>. Control over pre-mRNA splicing was introduced as a novel means over gene control. Analogous to the 5'-scanning design the riboswitch was used as such an inserted close to the 5'-splice site. Splice regulation was found to depend on aptamer stability and exact positioning and yielded switches with up to 33-fold regulatory activity.

## 2.4 RNA synthetic genetic networks

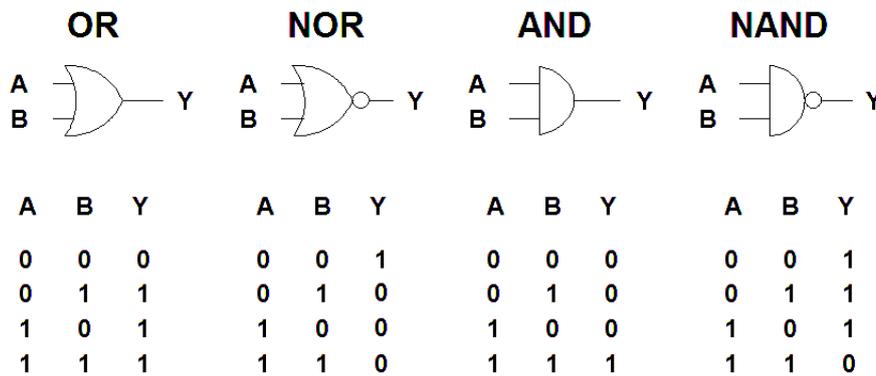
Information inside cells is transmitted and processed by genetic circuits that allow the simultaneous integration of different input signals, typically resulting in a single output. Inputs may come in form of metabolites, transcription factors or other environmental factors like ionic strength, light and temperature. The timely and sufficient processing of these signals is fundamental to the correct functioning of the metabolic state and thus cell viability. Long since it has been known that cell populations, even clonal cell populations, exhibit heterogeneous phenotypes with respect to practically all expressed genes. These “noisy”, inhomogeneous expression patterns are the result of stochastic fluctuations of the involved molecular components in their level and activity<sup>170</sup>. On a global scale noise could be quite advantageous to the population or even the overall species, as phenotypic variability between single cells fosters a wide range of differentiation strategies and thus fast adaption to rapidly changing environmental conditions. On the single-cell level it allows the coordination of larger set of genes, because time-variable fluctuations of one component can mask and compensate for fluctuations of another (Figure 2.12 upper and middle panel). More precisely, short, burst-like availabilities of a transcription factor (mRNA spikes) are buffered by the variable slow degradation rates of the produced protein – they are averaged over time. In turn, if the protein was again a transcription factor, this results in the prolonged variable activation of another downstream gene. If more of such patterns are overlaid, stable expression patterns and a phenomenon called cell memory can occur over cell-cycle timescales.



**Figure 2.12 Classification of noise in a short transcriptional cascade.** Bursting describes the irregular amplification of mRNA whose genetic information is subsequently amplified by many translations rounds, giving rise to sudden increases of the gene product. On the protein level these burst decay with variable rates over time, thus averaging protein levels. Time-averaging directly influences information propagation since the variably decaying signal intensity create downstream bursts on their own that are higher in frequency but generally lower in amplitude. Compared to the initial bursts the transmitted information appears no “smeared” over time.

## Introduction

Although recognized as beneficial to wild type populations, synthetic biologists rather view noise as an impediment to the predictable functionality of designed genetic circuits. Here noise manifests as intrinsic and extrinsic to the studied phenotype. Intrinsic noise refers to variable rate constants of underlying processes such as promoter activity, transcription, partly translation and protein and mRNA half-lives of the respective gene. Extrinsic noise stems from processes that affect gene expression at large, such as the availability of the general and specific transcription and translation machinery. The differentiation between these two can be assessed by monitoring the expression of the equally controlled reporter genes. Uncorrelated fluctuations hint toward intrinsic noise, whereas correlated fluctuations point toward extrinsic influences. Noise as a cause of malfunctions of devices cannot be inferred from bulk measurements as these inherently average any phenotype and thus blend out the significant effects of cell-to-cell variations. To study populations with respect to their heterogeneity at the single-cell level corresponding techniques must be used. Time-lapse recordings by fluorescence spectroscopy in a microfluidic environment have emerged as powerful means to investigate noise phenomena. Flow cytometry is equally suited, but allows only snapshots of the population over time, as a specific cell can only be analyzed once. Efforts to determine and reduce noise have led to the construction and refinement of various synthetic genetic circuits and networks in the past years. Designed to follow Boolean logic, it is generally sought to have them process input signals in a digital-like manner (exemplified in Figure 2.13). Logics are named according to the desired signal integration. A NOR gate for instance will only yield an output, if none of both input signals are present. As described, phenotypic variations bedevil the realization of digital behavior.



**Figure 2.13** Examples of logic gates implementing Boolean function. Logic function can be represented by symbols (here distinctive shape) and complementing truth tables that detail the output according to the input signal. All gates integrate two inputs to a single output. Connecting several gates, greater circuits can be build that process and transmit complex information. Adopted from <https://i.stack.imgur.com/U92yS.png>.

Nonetheless, the assembly of logic devices to higher order genetic circuits is a prerequisite to establish *de novo* designed metabolic pathways *in silico*. Hence, it is desirable to have building blocks that may be plugged together in a manner that is feasible to compute. RNA regulatory mechanism can be implemented at a low sequence load and metabolic burden as they are compact and need only to be transcribed, respectively. Proteins in turn consume much of the produced cellular energy by their translation. RNA regulators have proven high modularity and portability across species. The may be applied to control virtually any gene expression at short timescales if acting *in-cis* and their precise signal propagation is determined by their fast degradation rates, reducing time-averaging and subsequently noise propagation.

### 2.4.1 Synthetic parts characterization

Before plugging parts to devices and devices to systems, both, parts and devices, need to be thoroughly characterized. Since parts characterization could be extended seemingly endlessly, as there are so many factors to consider, some physicochemical, thermodynamic and kinetic parameters are routinely determined, depending on the part or device. For RNA parts these are (i) structure that is determined by chemical probing, x-ray crystallography or other spectroscopic techniques, (ii) ligand interaction that is assessed by binding and dissociation rates through spectroscopy and calorimetric assays, (iii) physicochemical properties as ion and temperature dependence. The quantification of device performance constitutes a key determinant that follows not standard at all. As most studies are conducted for the proof-of-principle, reporter gene assays with fluorescent proteins as output are frequently used. Since each part or device is characterized in a specifically set up expression system, including instrumentation and environmental factors, absolute comparisons of device performances from different studies with respect to their regulatory activity are rather difficult. Therefore relative measures of output levels were proposed that quantify the actual reporter relative to an internal control, typically another fluorescent protein. Moreover, as context-dependencies influence the characterization, the process itself measures the supposedly independent functionality of the studied part rather imprecisely.

Another important issue is constituted by phenotypic noise. To characterize a device in the context of cell-to-cell variability single-cell measurements can aid to unravel its reliable performance across a population and point out parameters that allow efficient tuning. The time-resolved tracking of the involved molecular components is an absolute necessity to resolve transient processes and states. Advances in mRNA and protein tagging allow the reconstruction of time traces and the subsequent determination of *in vivo* kinetic parameters like synthesis and degradation rates and the cellular localization of the involved components. Importantly, this way RNA-ligand interactions can be mapped spatiotemporally, allowing insights and conclusions about their gene regulatory mechanism.

As RNA devices are further incorporated in increasingly complex synthetic networks, the in-depth studies of noise, spatiotemporal response, evolutionary robustness and orthogonality between device and host organism will need to be conducted.

### 2.4.2 Synthetic RNA logic gates

Natural RNA-based logic operations have been observed in the context of riboswitch discovery and characterization. The *glmS* riboswitch from *B. subtilis* processes two inputs, glucosamine-6-phosphate as activating ligand, and glucose-6-phosphate as inhibiting ligand, the modulate its activity<sup>107</sup>. In *B. clausii* two independent riboswitches sensitive to S-adenosylmethionine or vitamin B<sub>12</sub> integrate a NOR logic to control expression of the downstream gene. Presence of either ligand leads to transcription termination.

Synthetic RNA circuits have been constructed from riboswitches, riboregulators and CRISPR-derived parts<sup>76-78,86</sup>. In an early study, aptamers sensitive to theophylline and TPP were set to control translation in the 5'-UTR of a reporter gene in bacteria<sup>171</sup>. Translation was only active in the presence of both ligands (AND gate) or in the absence of both or either ligand (NAND gate). In a follow-up study a transcriptional OFF switch was combined with a translational ON switch that both bound to TPP, but with different sensitivities. Thus, acting as a band pass filter, only intermediate concentrations of TPP promoted gene expression<sup>172</sup>.

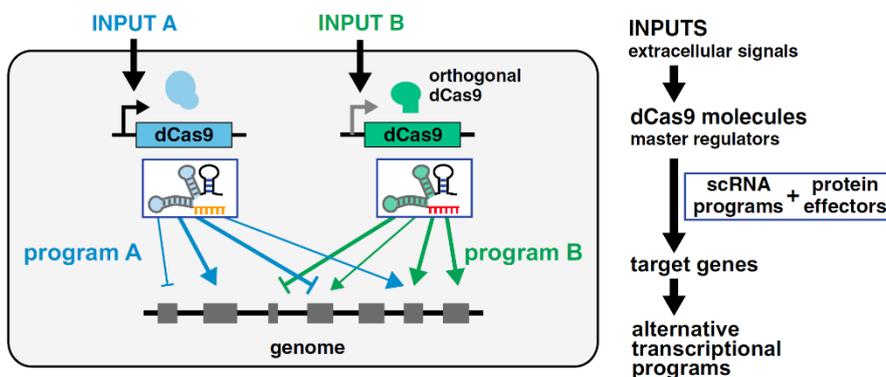
## Introduction

The introduction of aptazymes enabled the construction of various logics in yeast. Using the HHR as a workhorse, aptazymes sensing TPP and theophylline were combined to AND, NOR and ANDNOT gates <sup>173</sup>. This study nicely demonstrated the incorporation of an amber suppressor tRNA to implement the desired logic. Further work on HHRs implemented a series of theophylline and tetracycline-responsive riboswitches by realizing novel ribozyme architectures. Deploying stems I and II, ribozymes with aptamers attached to a single stem or both stem were generated. The single-stem attachment featured even a “backpack-like” organization with one aptamer stacking on top of the other. Ultimately, AND, NOR and NAND gates were assayed for their ability to regulate gene expression in the 3'-UTR of the reporter gene, affecting mRNA stability.

Not only the HHR, but also the twister ribozyme could be engineered to yield AND, NAND, OR, NOR and two different ANDNOT gates that are sensitive to TPP and theophylline. Although signal integration resulted only in quite moderate output regulation, it impressively demonstrated the modularity and plug-and-play accessibility of RNA-derived regulation <sup>144</sup>.

Turning from riboswitches to riboregulators, the previously described toehold switches gained their ligand sensitivities from inducible promoters and were used to construct a layered 4-input AND gate <sup>92</sup>. The complexity of this design in terms of the number of regulatory levels exceeds that of the engineered ribozymes, because here four initial, orthogonal inputs controlled the output of two RNAs that were again used as input to regulate GFP levels. Despite a dramatic performance decrease relative to the devices of which the 4-input system was composed, the signal was sufficiently propagated to yield a 6-fold upregulation of GFP.

A special highlight in the design of synthetic circuits was the marriage of RNA and protein regulators giving birth to sophisticatedly engineered CRISPR tools in yeast. Using catalytically inactive Cas9 as an inducible master regulator, scaffold riboregulators (see Figure 2.2C) recruiting transcription activation domains by their aptamer component addressed various modified promoters of a metabolic pathway to direct its flux and control the output. Less a common logic gate and more a hierarchical, programmable control system, the power of a functional RNA-protein-dependent regulation was demonstrated. RNA scaffolds can be sequence-engineered to specifically bind to almost any target site. Equipped with basically any aptamer from a pool of aptamers for virtually any ligand they can be multiplexed to activate or repress transcription genome wide. Since there is only one inducible master regulator, a defined cellular program can be started by a single input. As dCas9 and scaffold RNA can be designed to exhibit orthogonality to other dCas9-scaffold RNA pairs, different programs could be run at the same time, facilitating the paralleled or time-shifted control over whole cell re-programming (Figure 2.14).



**Figure 2.14 Envisioned application of CRISPR tools to implement cellular programs.** Running two orthogonal cellular programs (blue and green) in parallel that are triggered by two distinct inputs A and B is facilitated by the master regulators dCas9. A master regulator in turn has its own set of molecular slaves within one program. The molecular slaves address different target themselves. By this hierarchical system few input signals can trigger a vast downstream cascade.

### 3 Scope

The generation of riboswitches with novel ligand selectivity constitutes a major bottleneck in their broader application as regulators of conditional gene expression. The aptamer domain can be created against virtually any target molecule and therefore it is highly desirable to use them as sensory units for the conditional control of gene expression. However, it is assumed that aptamers are a priori not suited to regulate gene expression, because the applied *in vitro* selection strategy is designed to identify RNA sequences exhibiting high affinity to their target and not to select for a gene-regulatory activity. At this stage it is key to accomplish the transformation from an aptamer sequence to an RNA sequence capable of switching gene expression from an ON state to an OFF state – a riboswitch. A previously identified putative quality of *in vitro* selected aptamers was sought to be exploited, featuring the inherent combination of aptamer domain and expression platform in a single compact structure to control gene expression in yeast. The basic concept of the system was to interfere with ribosomal scanning and the initiation of translation at the start codon by inserting a road block in the 5'-UTR of the corresponding mRNA, but leaving mRNA levels undisturbed. This approach should then be extended by application of an *in vivo* screening to identify *in vitro* generated ciprofloxacin aptamers with riboswitching properties.

Mechanistic insights into riboswitch function are only beginning to foster the understanding of how to convert aptamers into riboswitches or how to efficiently improve existing riboswitches. Advancing *in silico* riboswitch design and refinement, dimers of the well-characterized tetracycline riboswitch were anticipated to aid in the construction and training of a computational model that would subsequently compute riboswitches exhibiting an increased regulatory activity. Besides improvements at the nucleotide level to enhance performance, a thorough device characterization at the cellular level is needed to dissect the influence of phenotypic variability on device performance. Until today, no standard performance descriptor has been established and thus most devices are solely characterized by their ability to switch gene expression based on end point data averaged over a whole population. These bulk data cannot account for cell-to-cell variability and no information on the transient shift from ON to OFF levels is obtained. Engaging these issues, a novel logic NOR device constructed from the neomycin and tetracycline riboswitches was sought to serve as a model to investigate the regulatory performance with respect to cell-to-cell variability and the separation of two Boolean states at the single-cell level. Transient, single-cell data was supposed to calibrate a hierarchical stochastic model and allow a performance mapping of the logic device based on ROC analysis. Using ROC analysis a performance measure should be derived that assessed the correct functionality of the gate at the single-cell level – a prerequisite for its further adaption to a genetic circuitry. Not being specific for riboswitch devices, this performance indicator could then be applied to other logic devices and facilitate their comparable analysis.

## 4 Results & Discussion

### 4.1 Project I: Identification of RNA aptamers with riboswitching properties

#### 4.1.1 Results

##### 4.1.1.1 A literature-based screen for aptamers with putative riboswitching properties

In a first approach, RNA aptamers already published by various research groups were selected from the literature. It was sought to assess their potential as riboswitches in the aforementioned 5'-UTR design without or with only minor modifications by mutations or truncations of the closing stem of each aptamer (see chapter 2.3.3.2). Truncations of aptamer sequences taken from natural riboswitches were necessary to isolate only the sensing domain and an adjacent closing stem. With the exception of the M4-aptamer (ligand: neomycin), the AN32sh-aptamer (tetracycline) and the MG-4 aptamer (tetramethylrosamine), none of the other aptamers had been shown to work as riboswitches in yeast employing the 5'-UTR design. Aptamers closely matching most of the following specifications were chosen for the experiments. (i) Primarily, aptamers were selected according to their novelty to identify already published sequences whose gene-regulatory potential has not been investigated yet. These aptamers were intended to serve as alternative riboswitches apart from the long since established aptamers sensing neomycin, tetracycline and theophylline. (ii) Besides their novelty, sequence length was of greater importance since sequences smaller than approximately 70 nt allow for a faster re-design to yield enhanced functionality and a determination of the actual ligand-binding domain can be performed at lower efforts, if such data are not already available. (iii) The upper threshold for an important parameter of aptamer-ligand interaction, the dissociation constant  $K_D$ , was set to the low  $\mu\text{M}$  range, because as a general rule a lower  $K_D$  reduces the effective concentration of the supplemented ligand. This in turn diminishes the risk of unspecific side-effects and adverse toxicity issues caused by the ligand. (iv) Permeability of most ligands for the yeast cellular membrane is undetermined and therefore the highest non-toxic concentration was chosen. This requires a good solubility of the ligand in water or water-ethanol mixtures. (v) A last, but not less significant criterion was the costs to purchase or synthesize the ligand of interest to make the system eventually available to a broader public.

Table 4.1 summarizes all aptamers and their ligands tested in this study. Aptamers were clustered into different blocks that refer to the substance class of their respective ligand. The length of the cloned aptamers is given in nucleotides (nt), molecular weight of the ligand as relative molecular weight ( $M_r$ ) and the dissociation constant ( $K_d$ ) of the aptamer-ligand complex is displayed in  $10^{-9}$  mol/L (nM). Dimethylindol red and atrazine were supplemented in two different concentrations and the aptamers MG-4, TMR1 and TMR2 were tested with the non-cognate ligand rhodamine B. Most aptamers were obtained by an *in vitro* SELEX and hence the original publications are tabulated. Two aptamer domains were isolated from naturally occurring riboswitches, the thiM riboswitch from *E.coli* and the xpt-pbuX riboswitch from *B. subtilis*. The secondary structures as predicted by RNAfold are listed in Figures 4.1(#1-22) along annotations specifying the detailed modifications of some aptamers<sup>174,175</sup>.

## Results & Discussion

#	Aptamer	Length [nt]	Ligand	MW [g/mol]	Substance class	K <sub>d</sub> [nM]	c(ligand) [μM]	Reference
1	AN32sh	69	Tetracyclin	444	antibiotic	1	250	158
2	M4	34	Neomycin	615	antibiotic	9	100	138,157
3	J6sl	27	Tobramycin	468	antibiotic	5/9	100	176
4	X1sl	26	Tobramycin	468	antibiotic	12	100	176
5	D1	20	Tobramycin	468	antibiotic	220	100	177
6	Strepto- Tag	45	Streptomycin	582	antibiotic	<100 0	100	178–180
7	8-1-1	22	Kanamycin B	484	antibiotic	180	100	181
8	Dano-I	32	Danofloxacin	357	antibiotic	3	100	182
9	MG-4	38	Malachite green (MG)	365	dye	800	100	137,183,184
9	MG-4	38	Rhodamine B (tetraethylrhodamine)	479	dye	n.d.	100	137,183,184
10	TMR1	38	Tetramethylrhodamine	343	dye	85	100	154
10	TMR1	38	Rhodamine B (tetraethylrhodamine)	479	dye	n.d.	100	154
11	TMR2	46	tetramethylrhodamine	343	dye	70	100	154
11	TMR2	46	Rhodamine B (tetraethylrhodamine)	479	dye	n.d.	100	154
12	TMRV1	28	Tetramethylrhodamine	343	dye	n.d.	100	154
13	TMRV2	30	Tetramethylrhodamine	343	dye	n.d.	100	154
14	13.2	60	DMHBI	300	dye	464	100	185
15	Apt-1	49	Dimethylindol red	484	dye	87	50;250	186
16	SRB-2	52	Sulforhodamine B	559	dye	80	100	187
17	N40 region	49	Atrazine	216	various	<200 0	100;300	188
18	Aptamer 21	53	HAP (heteroaryl- dihydropyrimidine)	360	various	50	100	189
19	thiM riboswitch	72	TPP (Thiamine pyrophosphate)	424	metabolite	20	500	101
20	ag.06	54	L-Arginine	174	metabolite	330	10000	190
21	xpt-pbuX riboswitch	67	Guanine	151	metabolite	5	500	191
22	mTCT8-4	33	Theophylline	180	metabolite	100	1000	120

**Table 4.1 List of all tested aptamers and their assigned ligands.** The denominations were adopted from the original references. The indicated aptamer lengths refer to the actually cloned sequence and are illustrated as predicted secondary structure in Figure 4.1. The molecular weight (MW) of each ligand, their tested concentrations (c (ligand)) and their dissociation constant (K<sub>d</sub>) are provided. All aptamers are categorized with respect to the substance class of their ligands, reflected by the different shades of blue.

Only small molecules were chosen as ligands, because they are easily handled and administered to the growth medium. Proteins or even cellular components constitute other classes of ligands that are either impractical to use or their abundance may not be controlled, rendering a switchable gene expression impossible. Three major classes could be identified, each featuring aptamers that had already been employed as riboswitches or that are naturally occurring riboswitches in bacteria. (i) Antibiotics are purchased at a low cost and toxic side effects can be controlled by the appropriate concentration. Their

## Results & Discussion

corresponding aptamers are short in sequence and accessible to *in silico* modeling. Although antibiotics do permeate through the cellular membrane of *S. cerevisiae*, the intracellular concentration is unknown and thus, empirically determined concentrations derived from earlier publications on the neomycin- and tetracycline- binding riboswitches were adopted<sup>138,155</sup>. (ii) Dyes represent an interesting ligand class as their uptake inside the cell could either be monitored by phenotype – rhodamines cause the cells to become visibly pink – or their concentration may be determined in a cytoplasmic extract. They are widely used for fluorescence-based assays and many are easily available by purchase. (iii) Metabolite-sensing aptamers were either derived from riboswitches identified in bacteria (TPP, guanine) or stem from a SELEX against the respective compound (theophylline, L-arginine). The addition of L-arginine required a pH adjustment with hydrochloric acid before the experiment was conducted. The cytoplasmic concentration for free guanine could not be found in the literature (Yeast metabolome data base (YMDB), <http://www.ymdb.ca>), but the purine metabolism (KEGG pathway database, <http://www.genome.jp/kegg/pathway/map/ec00230.html>) suggests its rapid conversion to guanosine, guanosine monophosphate (GMP), xanthine or deoxyguanosine. Thus no free guanine was expected to obscure a putative regulatory activity of the corresponding aptamer.

Supplied by the purine and vitamin B6 metabolism, the thiamine metabolism generates thiamine pyrophosphate (TPP) that in turn is a cofactor for several enzymes involved in numerous metabolic pathways (YMDB, <http://www.ymdb.ca/compounds/YMDB00381>). The intracellular concentration of thiamine could not be retrieved. The proteinogenic amino acid L-arginine is endogenously anabolized from ornithine in the cytoplasm of *S. cerevisiae* and catabolized in the context of the urea cycle by arginase. Free intracellular concentrations of L-arginine vary greatly depending on the composition of the used culturing medium (YMDB, <http://www.ymdb.ca/compounds/YMDB00592>). Here L-arginine was not supplemented to cultures with cells transformed with the corresponding aptamer. The effect on cell growth was found to be negligible as L-arginine is readily synthesized from glutamate supplemented at 1 mM.

In a first step a toxicity assay was conducted (Table 4.2). As a measure of toxicity the optical density (OD<sub>600</sub>) of the cell suspension was determined in the presence of the respective ligand. Therefore yeast cells were transformed with pWHE601 and toxicity was assayed in a growth test. Samples were drawn every hour for 8 h and at 24 h. The standard doubling time was determined to ~105 min. Malachite green (MG) was found to be too toxic to the cells to use it as a ligand (doubling time ~215 min) and since the other original, non-toxic ligand tetramethylrosamine could neither be purchased nor synthesized in time, the aptamer was tested with a structurally related fluorophore, rhodamine B. Besides MG, DIR and atrazine diminished cell growth to an equal extent at the highest concentration of 250 µM or 300 µM supplemented, increasing doubling time to ~130 min, but both were regarded not toxic. Apart from these three compounds, no other ligand caused impaired cell growth at the indicated working concentration that was also supplemented for the subsequent reporter gene assays (Table 4.1).

The absorption spectra of all antibiotics, atrazine, HAP and all metabolites revealed no spectral interferences with GFP fluorescence emission. The dyes, as expected, either emitted fluorescence overlapping with GFP (DMHBI) or absorb light emitted by GFP (all rhodamine derivatives and DIR). Nonetheless, this potential quenching effect was regarded a second order problem, since without

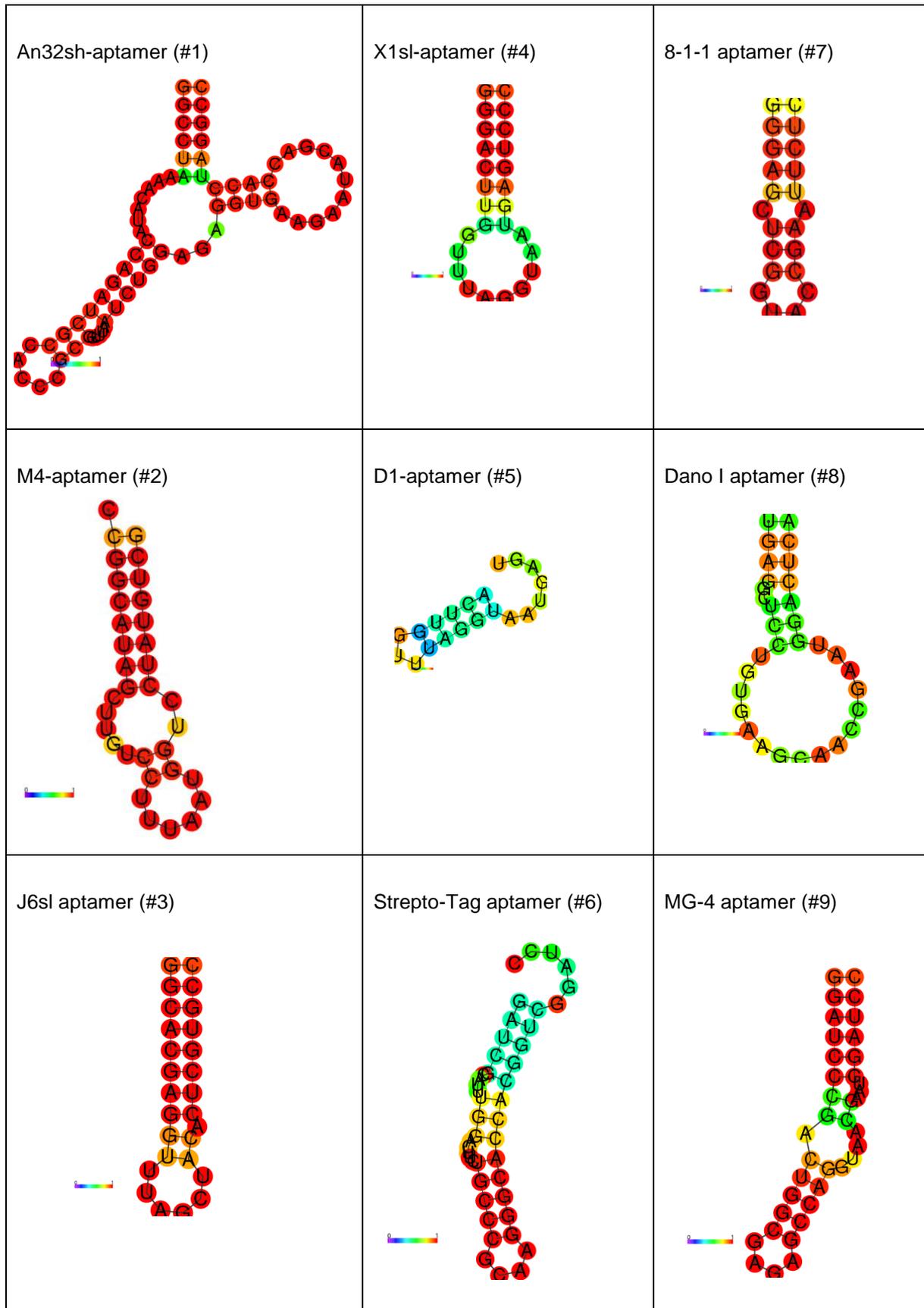
## Results & Discussion

sufficient basal expression the corresponding aptamer could not be used in any case. In addition, a change of the reporter assay to a  $\beta$ -galactosidase-based system could solve any spectral issues.

Ligand	Doubling time in min
untreated	100
Tetracycline	102
Neomycin	105
Streptomycin	108
Kanamycin B	108
Danofloxacin	108
Rhodamine B	110
Sulforhodamine B	108
TMR	107
DMHBI	108
HAP	103
Thiamine	108
L-arginine	110
Guanine	106
Theophylline	108
DIR	133 at 250 $\mu$ M
MG	216
Atrazine	135 at 300 $\mu$ M

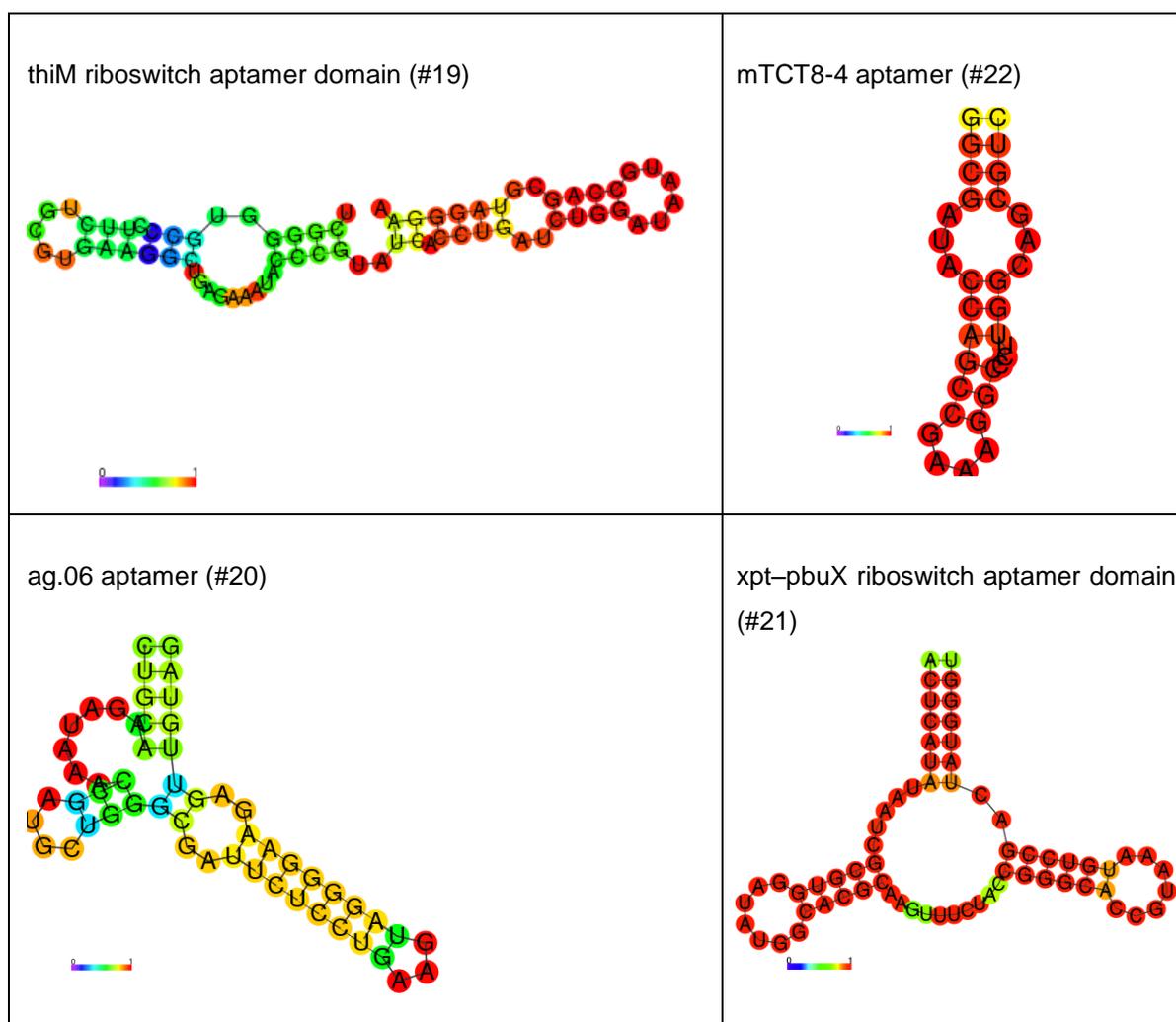
**Table 4.2 Growth test to assess toxicity.** Yeast cells were transformed with vector pWHE601 and cultured in SCD-Ura medium. Separate cultures were prepared with all different ligands and grown for 24 h. Samples were drawn as indicated, diluted and their optical density measured at 800 nm. Ligand concentrations equaled those used for reporter gene assays. For further information on pWHE601 refer to the reporter gene assay and chapter 6.5.1.1. Doubling time was calculated by the division of the  $\ln(2)$  by the growth constant. The growth constant was calculated as the slope of the exponential fit of the logarithmized  $OD_{600}$  values during exponential growth. Experiments were performed in one triplicate. The standard error of the mean was below 5% for all triplicates.

Results & Discussion





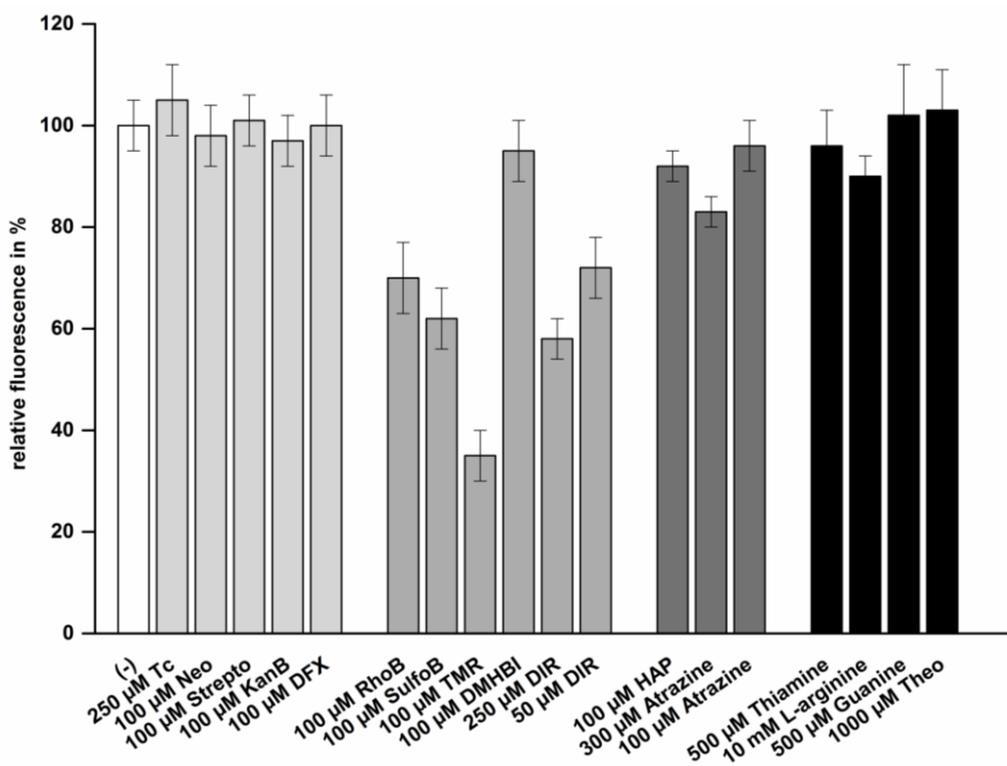
## Results & Discussion



**Figures 4.1 Secondary structures of all tested aptamers as predicted by Vienna RNAfold** <sup>174</sup>. All aptamers are numbered according to table 4.1. The secondary structure exhibiting the minimum folding energy (MFE) of all energetically possible secondary structures was computed and is illustrated. The color code from violet to red represents the increasing probability of each nucleotide to be unpaired or paired, respectively. The following aptamers were modified as described. #04: **X1sl aptamer**: the closing stem was truncated by 4 nt. #08: **Dano-I aptamer**: the minimal motif of 32 nt was selected. #09: **MG-4 aptamer**: the minimal motif was chosen for experiments. #12: **TMRV1 aptamer** and #13: **TMRV2 aptamer**: a redesigned version of TMR1 aptamer to reduce reporter gene depression by the aptamer in absence of the ligand and to allow for ligand-based gene regulation. #14: **13.2 aptamer**: the minimal motif shown in the supplementary data was chosen for experiments. #16: **SRB-2 aptamer**: the minimal motif was used for experiments. #17: **N40 aptamer**: the N40 region was isolated as the atrazine-binding domain of a larger riboswitch-forming sequence and thus selected for further studies. The N40 sequence is truncated 5' by 6 nt to yield a reasonable stem length. #18: **aptamer 21**: this aptamer was taken from data given in the supplement material of the original publication, because it showed superior binding affinity to the ligand #3 that was also picked from the supplementary data and available through custom synthesis. #19: **thiM riboswitch**: for the experiments only the aptamer domain of the thiM riboswitch was cloned. #20: **ag.06 aptamer**: the published sequence contained two subsequent closing stems. Stem 1 was deleted and stem 2 shortened and modified resulting in a 6 nt closing stem. #21: **xpt-pbuX riboswitch**: only the aptamer domain of the xpt-pbuX riboswitch was used and cloned. #22: **mTCT8-4 aptamer**: the published aptamer was taken and the stem modified according to another study <sup>125</sup>.

## Results & Discussion

All aptamer sequences were ordered as forward (fwd) and reverse (rev) primers that were either directly hybridized and ligated with pWHE601\* using *AgeI* and *NheI* restriction sites or amplified in a PCR using fwd and rev primer as mutual templates and subsequently prepared for ligation with the same restriction enzymes and linearized vector. Vector pWHE601\*<sup>192</sup> was modified from pWHE601 that had been introduced and employed in previous studies<sup>138,155</sup>. Both vectors bear a *GFP+* gene expressed from an *ADH1* promoter, whereas in pWHE601\* the start codon was deleted and had thus to be introduced alongside the aptamer sequences in the 5'-UTR of the reporter gene. Vector pWHE601 served as a positive control and pWHE601\* as a negative control. After cloning and sequencing, yeast strain RS453 $\alpha$  was transformed with plasmids harboring the new aptamer sequences in the 5'-UTR of the *GFP+* reporter gene. Cells were cultured in two independent triplicates for 48 h in absence and presence of the ligands assigned to each corresponding aptamer (see Table 4.1) and analyzed in bulk by fluorescence spectroscopy (Figure 4.3)<sup>138</sup>. Yeast cells expressing the neomycin- and tetracycline-binding aptamers were used as technical controls, whereas yeast transformed with pWHE601 were taken as a positive control expressing *GFP+* unimpaired by a 5'-riboswitch. In a preliminary experiment the positive control was treated separately with all different ligands to determine unspecific effects on the fluorescence output (Figure 4.2). Only DIR caused a drop of fluorescence to 50% relative to the untreated positive control, but it was not determined, whether this was due to quenching or other unspecific side effects. Interestingly, no other red dye was found to quench or otherwise diminish GFP fluorescence.



**Figure 4.2 Ligand influence on GFP+ expression from pWHE601.** Non-specific effects on gene expression were tested by supplementation of each ligand to the growth medium of yeast cells bearing pWHE601. GFP+ expression of each culture supplemented with a different ligand is presented relative to an untreated culture. The negative control pWHE601\* was equally treated and used to subtract background fluorescence. Cells were cultured in two independent triplicates for 48h and measured in bulk by fluorescence spectroscopy on a Fluorolog instrument after normalization to cell growth and subtraction of fluorescence background.

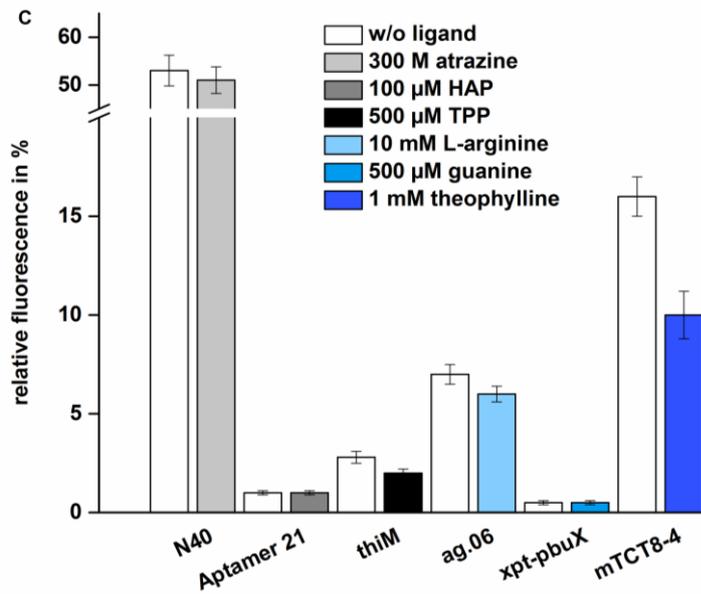
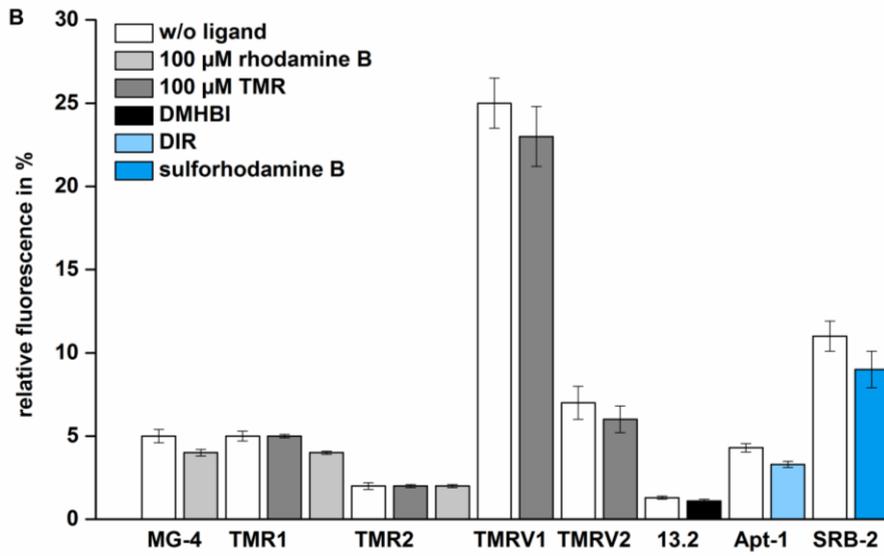
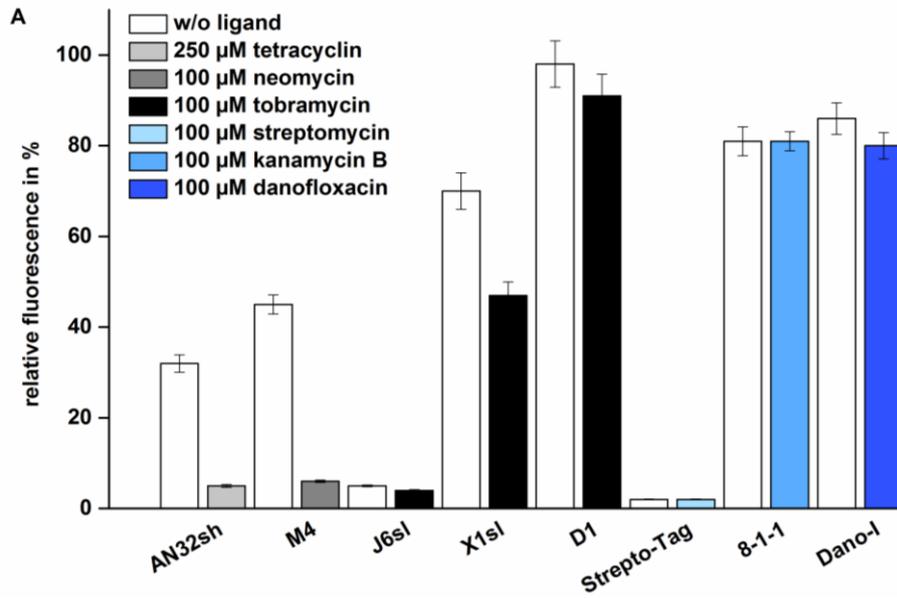
## Results & Discussion

The results of the reporter assay divided all aptamers into three groups. Those strongly suppressing *GFP+* expression in absence of their ligands, those with no effect on translation and those interfering moderately with gene expression. Strong suppression, that is, low to very low basal expression was observed for J6sl (tobramycin), the Strepto-tag (streptomycin), MG-4 (malachite green), TMR1, TMR2, TMRV2 (all tetramethylrhodamine), 13.2 (DMHBI), Apt-1 (DIR), SRB-2 (sulforhodamine B), Aptamer 21 (HAP), thiM-aptamer (TPP), ag.06 (L-arginine) and xpt-pubX aptamer (guanine). No or only slight suppression could be seen for X1sl, D1 (both tobramycin), 8-1-1 (kanamycin B) and Dano-I (danofloxacin). TMRV1 (tetramethylrhodamine), N40-region (atrazine) and mTCT8-4 (theophylline) repressed gene expression ranging from 50% to 16%. Addition of the assigned ligands resulted in no further decrease of gene expression for barely any of the newly introduced aptamer sequences. X1sl (tobramycin) and mTCT8-4 (theophylline) conferred slight regulatory activity with switching factors of about 1.5-fold and did so with moderate to high basal expression levels. To make use of them as riboswitches though, they would need to be extensively improved to reach switching factors greater than 5-fold.

In many cases, reduced basal gene expression may be correlated with increasing length and stability of the closing stem as predicted by RNAfold (Figure 4.1#6, 14). This was especially pronounced for aptamer 13.2 (DMHBI, #14) and the Strepto-tag (streptomycin, #6) that both feature comparably long closing stems. Vice versa, the gain of basal expression by a reduction of the closing stem's length and stability was showcased by the transition from the J6sl- to X1sl- and ultimately to the D1-aptamer all binding tobramycin as well as the rational truncation of the TMR aptamers (tetramethylrhodamine). Short closing stems in combination with more elaborate global architectures found with the SRB-2, N-40 or AN32sh aptamers, promote intermediate basal expression profiles. It should be noted, however, that all these relations are rather a rule of thumb and may not be applied to all aptamers, especially since their tertiary interactions are mostly unknown, but contribute significantly to the final 5'-roadblock opposing the scanning movement of the 43S complex and their ability to confer *in vivo* regulation as will be discussed later.

Almost all ligands were tested at a single concentration with the exception of atrazine and DIR that were also supplemented in lower concentrations, at 100  $\mu$ M and 50  $\mu$ M, since they had previously been found to reduce the growth rate of yeast cells (Table 4.2). The results proved to be no different compared to the originally used concentrations (data not shown). The technical controls M4 (neomycin) and AN32sh (tetracycline) showed the expected reduction of gene expression, ensuring correct sample processing and culturing conditions. The aptamers J6sl and X1sl had been tested by Dr. Julia Weigand in previous experiments and the reporter assay was repeated in this study.

## Results & Discussion

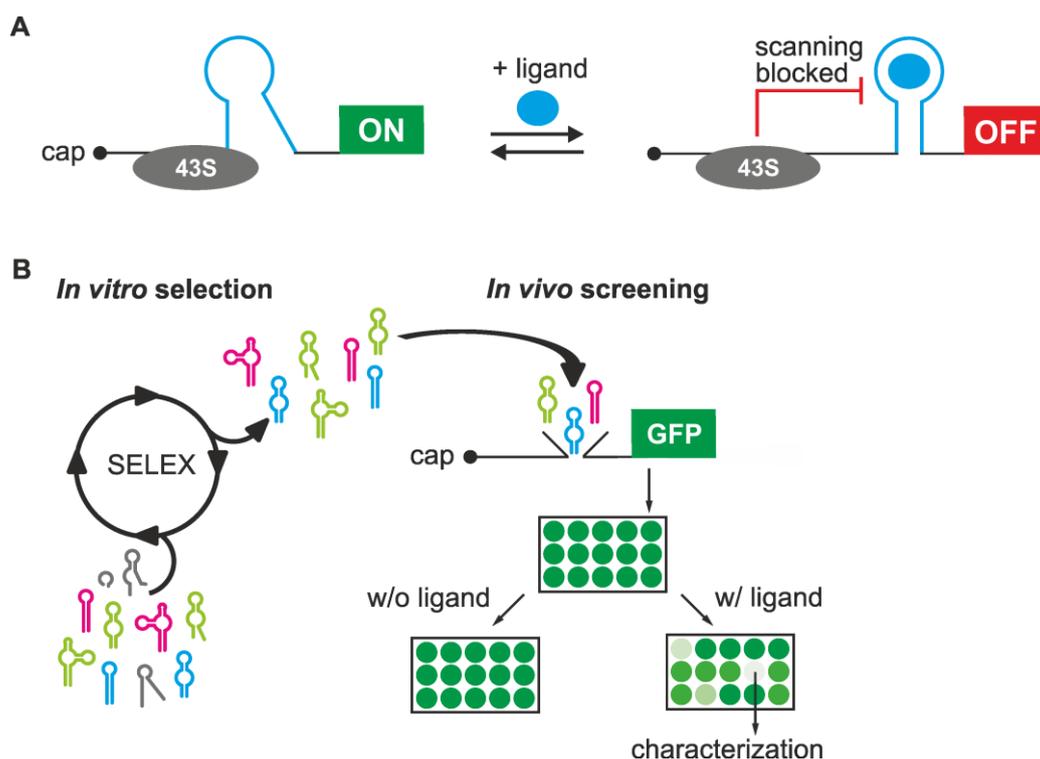


## Results & Discussion

**Figure 4.3 GFP reporter gene assay.** The assay results are grouped by substance class of the ligand. **A** Antibiotics, **B** dyes and **C** metabolites and other small molecules. *GFP+* expression of each culture supplemented with a different ligand is presented relative to the positive control supplemented with the same ligand. The negative control pWHE601\* was equally treated and used to subtract background fluorescence. Cells were cultured in two independent triplicates for 48h and measured in bulk by fluorescence spectroscopy on a Fluorolog instrument after normalization to cell growth and subtraction of fluorescence background.

### 4.1.1.2 *In vivo* screening for Ciprofloxacin-binding riboswitches

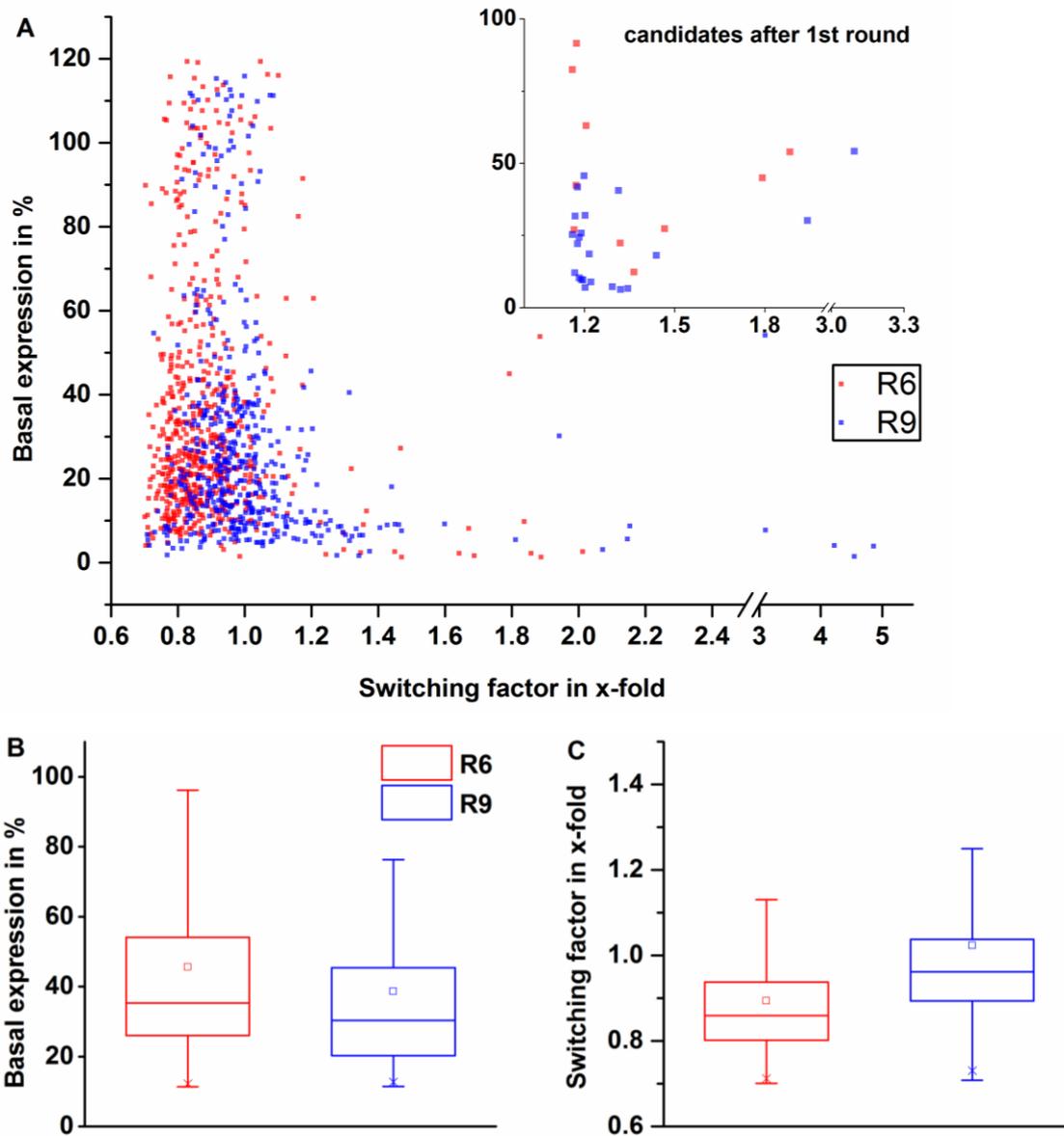
In a second approach, an *in vivo* screening for the identification of ciprofloxacin (CFX)-binding riboswitches was conducted. As described in chapter 5.1.2.1, RNA aptamer pools from SELEX rounds 6 (R6) and 9 (R9) were cloned in front of the *GFP+* reporter gene using homologous recombination up- and downstream of the established *AgeI* and *NheI* restriction sites present in pWHE601\*<sup>192</sup>. Dr. Florian Groher designed the initial pool according to Paige *et al.*<sup>185</sup> and generated the SELEX protocol. The SELEX itself was performed by Dr. Cristina Bofill-Bosch. Two separate reactions with respect to the two different pools were performed and used to transform yeast strain RS453 $\alpha$ . Cells were spread on agar plates that were subsequently assayed for fluorescence phenotypes by fluorescence microscopy. From about 1300 picked colonies 57% originated from R6 and 43% from R9.



**Figure 4.4 *In vivo* screening for CFX binding riboswitches.** **A** The 5'-scanning interference constitutes the regulatory mechanism deployed for the *in vivo* screening. Briefly, in absence of CFX the 43S complex is not or partially hindered by the aptamer structure. Presence of the ligand stabilizes the aptamer and interferes with ribosomal scanning to a greater extent, repressing translation of *GFP+*.  $P_{ADH1}$  and  $T_{ADH1}$  drive and terminate transcription, respectively. **B** The *in vitro* selection for CFX-binding RNA sequences yields moderately and highly enriched RNA pools. The pools are inserted upstream of the *GFP+* reporter gene. Subsequent screening steps in 96-well format in absence and presence of CFX are performed to identify CFX aptamers with riboswitching properties that are then further characterized.

## Results & Discussion

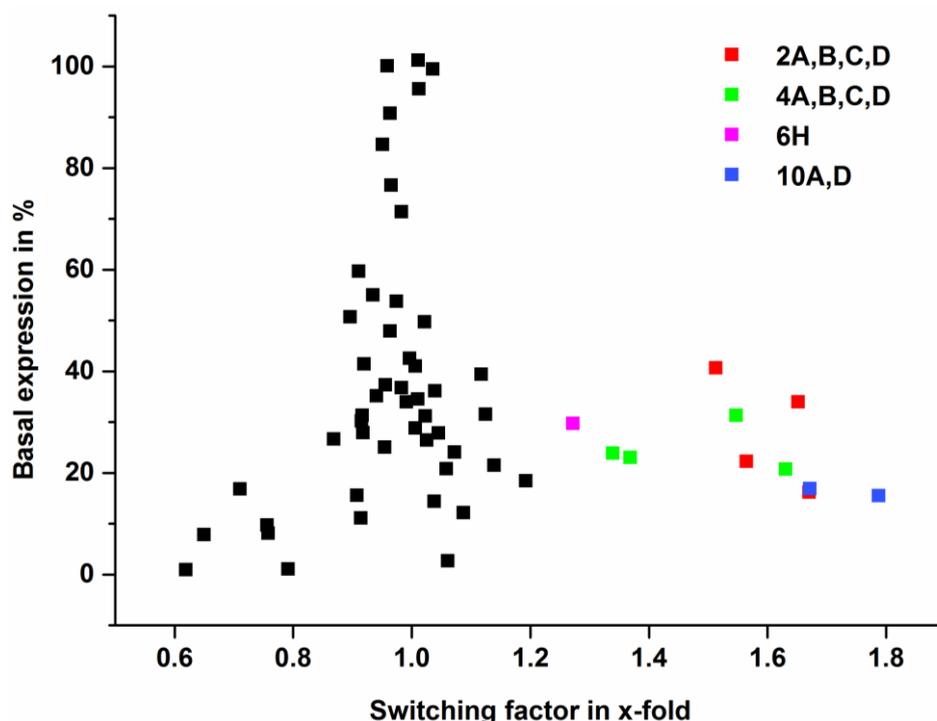
During the first screening round cells were selected with respect to their basal expression (BE) and switching factor (SF). As indicated by the highlighted subset in Figure 4.5A, only cells with a BE greater than 5% and SFs greater 1.1 were further processed. It is evident that over 95% of all assayed cells did not exhibit a switching behavior, but all levels of basal gene expression could be found, accumulating from 10% to 50% with respect to both SELEX rounds (Figure 4.5). The share of relative gene expression and switching factors among all different colonies for both SELEX rounds is presented in Figure 4.5B, C.



**Figure 4.5 Results of the first screening round.** A shows the distribution of 1300 phenotypes from R6 (red) and R9 (blue) previously identified by fluorescence spectroscopy and assayed in 96-well format in absence and presence of 100  $\mu$ M CFX. All phenotypes were classified with respect to basal expression and switching factors. Basal expression was calculated relative to GFP fluorescence exhibited by the equally treated positive control. Background fluorescence was subtracted but the fluorescence output was not normalized to cell growth. The inset depicts the candidates selected for the next screening round. The boxplots B and C illustrate the statistic distribution of all phenotypes identified from R6 (red) and R9 (blue) with respect to basal expression (B) and switching factors (C). The whiskers represent the upper and lower 25% of the population, whereas the mean is symbolized by a circle and the median by the horizontal line within the box that itself harbors 50% of the population.

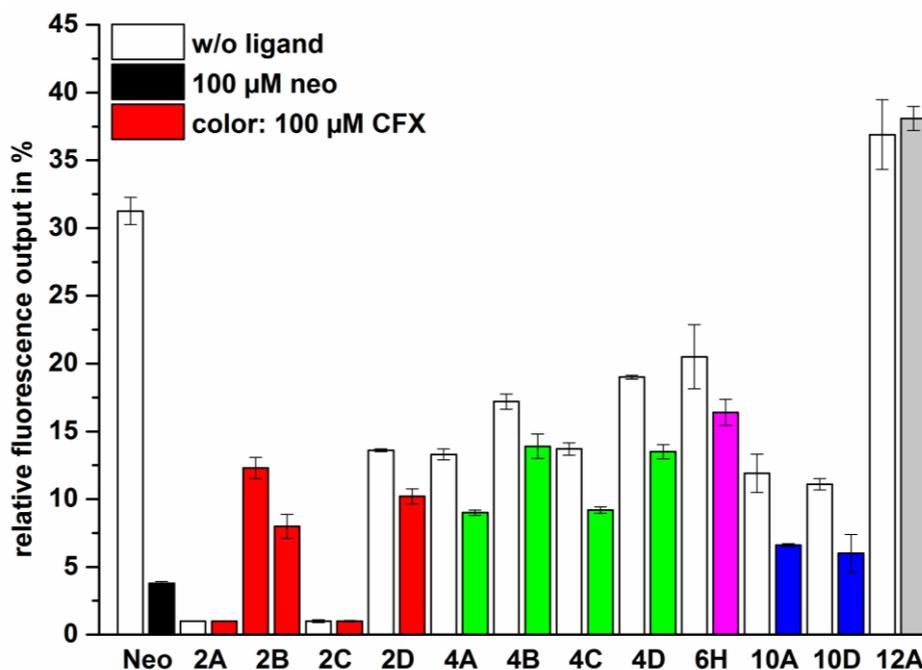
## Results & Discussion

The data suggested a lower overall BE for R9, but simultaneously higher SFs. This supports the fact that less colonies from R9 were identified by fluorescence microscopy compared to R6, because a more enriched SELEX pool harbors a higher concentration of aptamers with a presumably stronger structure depressing basal gene expression. Furthermore this finding is mirrored by the distribution of the 2% sequences sufficing the selection criteria and that were therefore subjected to the next screening round (inset Figure 4.5A). The values of these criteria were chosen based on the overall assay results and will be discussed later. Before the second screening was conducted, 32 colonies had been spread for single clones and thus, 4 clones of each regulatory active colony identified in the first round were analyzed (refer to 5.1.2.1 for more details). The results are shown in Figure 4.6. The graph displays 48% of a total of 128 screened clones – 52% emitted fluorescence on background level and were discarded. Although each four clones were derived from one colony, basal expressions and switching factors may split up as is observed for clones 2A, B, C and D or even 6H. Clones 2A-D vary mainly in their basal expression, whereas 6H was the only clone from colony 6 that continued to exhibit a switching phenotype. 117 out of 128 clones lost this phenotype through the propagation from first to second screening and in total, 4 out of 32 colonies from the first screening passed the second screening. Again, all clones after the singularization and 2<sup>nd</sup> screening were selected based on basal expression and switching factor levels.



**Figure 4.6 Second round of the *in vivo* screening.** The distribution of 61 clones derived from selected candidates from the first screening and assayed in 96-well format in absence and presence of 100  $\mu$ M CFX is shown. All phenotypes were classified with respect to basal expression and switching factors. Basal expression was calculated relative to GFP fluorescence exhibited by the equally treated positive control. Background fluorescence was subtracted and fluorescence output normalized to cell growth. Colored clones were subjected to further processing and analysis specified in Figure 4.7

## Results & Discussion



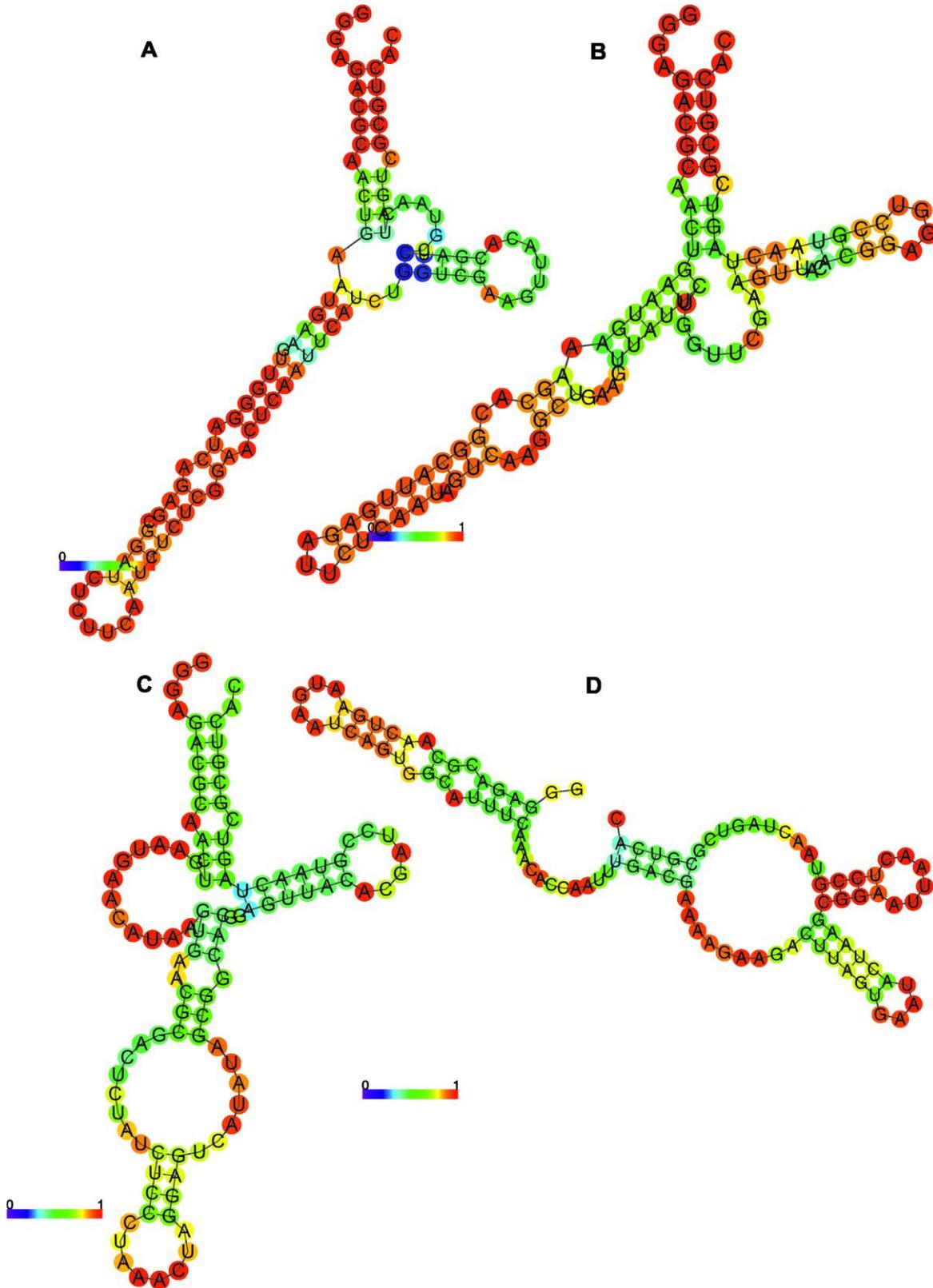
**Figure 4.7 Analysis of candidates from the second screening.** The GFP reporter assay was conducted after the plasmids had been propagated through *E.coli* and retransformed into yeast. Cells were grown in 1.5 ml cultures in 24-well plates in two independent triplicates in absence and presence of 100  $\mu$ M CFX for 24 h and analyzed in bulk with a Fluorolog instrument. The neomycin-sensitive riboswitch M4 and clone 12A were used as controls and treated equally. Relative fluorescence output was calculated as described for Figure 4.6. The color code is adopted from Figure 4.6.

The plasmids of all 11 remaining clones were isolated, propagated through *E.coli* to ensure single plasmid uptake and isolation, and retransformed into yeast cells that were assayed by fluorescence spectroscopy on a Fluorolog instrument (Figure 4.7). The M4-aptamer (neomycin) was used as a technical control to ensure correct culturing and processing of the cells. Clone 12A contains an aptamer sequence unresponsive to CFX, but with a moderate basal expression of 36%. It was chosen here and in later experiments to confirm that even very low switching factors of less than 2 imply a CFX-induced switching phenotype in this assay. Clones 2A and 2C lost their ability to confer CFX-dependent gene downregulation, whereas all other clones kept it, although with diminished basal expression. This reduction can be traced back to the altered experimental set up with different culturing conditions, background levels and instrumental equipment. The disappearance of switching factors is a common phenomenon after propagation of the isolated plasmids through *E.coli*, underlining the ability of yeast cells to take up more than one plasmid.

All clones were sequenced and interestingly clones 2B, 2D and 4A, B, C, D turned out to harbor the same aptamer sequence. Clones 6H, 10A/D and 12A differed in their sequence identity. To confirm the phenotypes now measured with the new experimental set up, the analysis for clones 2B, 4C, 6H 10A and 12A was repeated and found to match the previous results. The predicted secondary structures of the unique sequences 2B, 6H, 10A and 12A are displayed in Figure 4.8. Sequence 2B features a very long terminal hairpin with two symmetrical bulges. Sequence 6H adopts a very similar conformation, but features two asymmetrical bulges in the terminal hairpin. Both structures share a motif resembling a three-way junction. Sequence 10A features again a three-way motif and an internal loop followed by the

## Results & Discussion

short terminal hairpin. However, overall base pairing probabilities are very low and the actual conformation might be completely different.

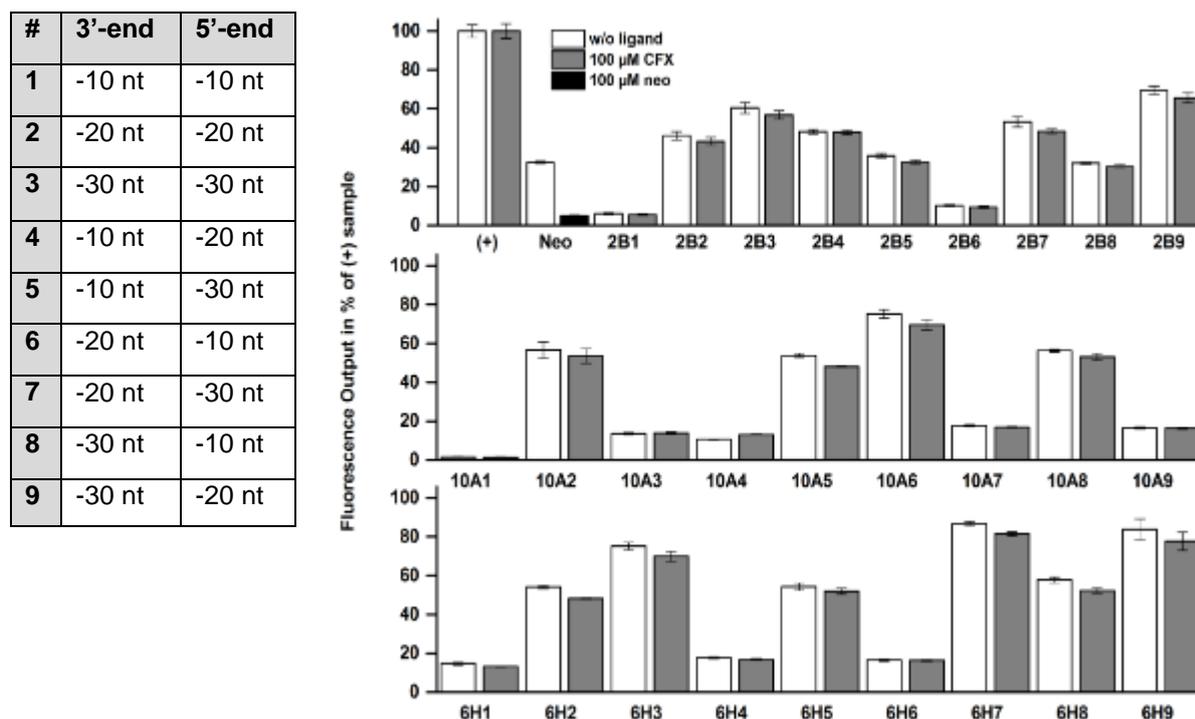


**Figure 4.8** Secondary structures of the unique sequences 2B, 6H, 10A and 12A as predicted by RNAfold. A 2B, B 6H, C 10A, D 12A. The displayed sequences contain the flanking primer sites and thus represent the full length construct cloned for the *in vivo* screening. The 5'-end of each sequence is always displayed on the left hand side. Base pairing probabilities are encoded by the heat map showing the probability of being paired or unpaired on a scale from 0 (low) to 1 (high).

## Results & Discussion

The regulatory inactive sequence 12A assumes a split conformation with respect to the 3'- and 5'-end which might explain its conferred phenotype.

Accounting for the sequence identities, only clones 2B, 6H and 10A were further processed and subjected to extensive truncations of the 5'- and 3'-ends. Truncations aimed at the reduction of sequence length and thus to shed presumably unnecessary sequence load. Furthermore, functional truncations increase the accessibility of a sequence to more intricate modifications towards a better performing riboswitch. The ultimate 20 nt of the 5'- and 3'-ends of the randomized inserts containing the aptamer sequence are constituted by the primer binding sites used to amplify the pool during SELEX and to attach overhangs for the homologous recombination of the aptamer pools with the vector pWHE601\* in yeast. These sequences are principally dispensable, although they might play a role in ligand binding by partially forming the ligand binding pocket. Truncations were performed by the successive reduction of 10 nt blocks from the 5'- and 3'ends in 3 steps, yielding nine different combinations of truncated sequences – 3 evenly matched cutbacks and 6 with either a 3'- or a 5'-unmatched overlap of 10 to 20 nt (Figure 4.9). Thus, truncations (partially) removed the flanking primer sites and cut into the randomized region composing the putative riboswitch sequence.



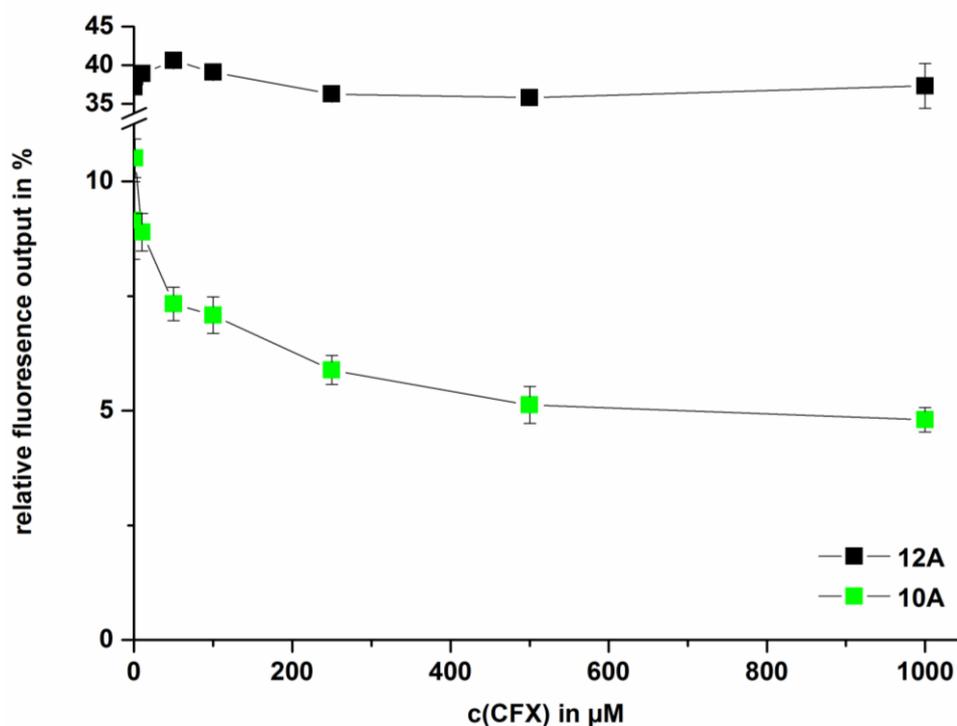
**Figure 4.9 GFP reporter assay of truncated sequences from clones 2B, 10A and 6H.** A The table shows the conducted truncations of the 3'- and 5' ends that are consecutively numbered and match the numbering of the relative fluorescence output plotted in B. Cells were grown in 1.5 ml cultures in 24-well plates in two independent triplicates in absence and presence of 100  $\mu$ M CFX for 24 h and analyzed in bulk with a Fluorolog instrument. Basal expression was calculated relative to GFP fluorescence exhibited by the equally treated positive control. Background fluorescence was subtracted and fluorescence output normalized to cell growth. The neomycin-sensitive riboswitch M4 was used as controls and treated equally.

After the GFP reporter assay had been performed it became evident that, although truncations resulted in greatly varied basal expressions, the switching phenotype had vanished. As truncations had been rather extensive, parts of the ligand binding site could have been destroyed. Unfortunately, no

## Results & Discussion

information could be deduced on basal expression as a function of the preceding 5'-UTR length or the calculated stability of the closing stem, as no correlation between these parameters is apparent.

In an effort to elicit at least some information from the sequences with regard to the ligand binding core and an improvement of the switching profile, clone 10A was further investigated as it exhibits the best switching factor of 1.8-fold and forms a defined closing stem. At first, the dose dependence of the putative riboswitch on increasing concentrations of CFX was tested to further unravel the riboswitching profile of this sequence. Clone 12A was again used as a control to warrant the dose responsiveness of clone 10A, since 12A was not expected to confer downregulation of *GFP+* in the presence of increasing concentrations CFX (Figure 4.10). The experiment verified sequence 10A as a riboswitch sensitive to CFX with a half-maximal inhibitory CFX concentration of 50  $\mu\text{M}$  and capable to decrease gene expression about 2.3-fold in presence of 500-1000  $\mu\text{M}$  CFX. Even in presence of 1000  $\mu\text{M}$  CFX sequence 12A showed no measurably effect on gene expression.

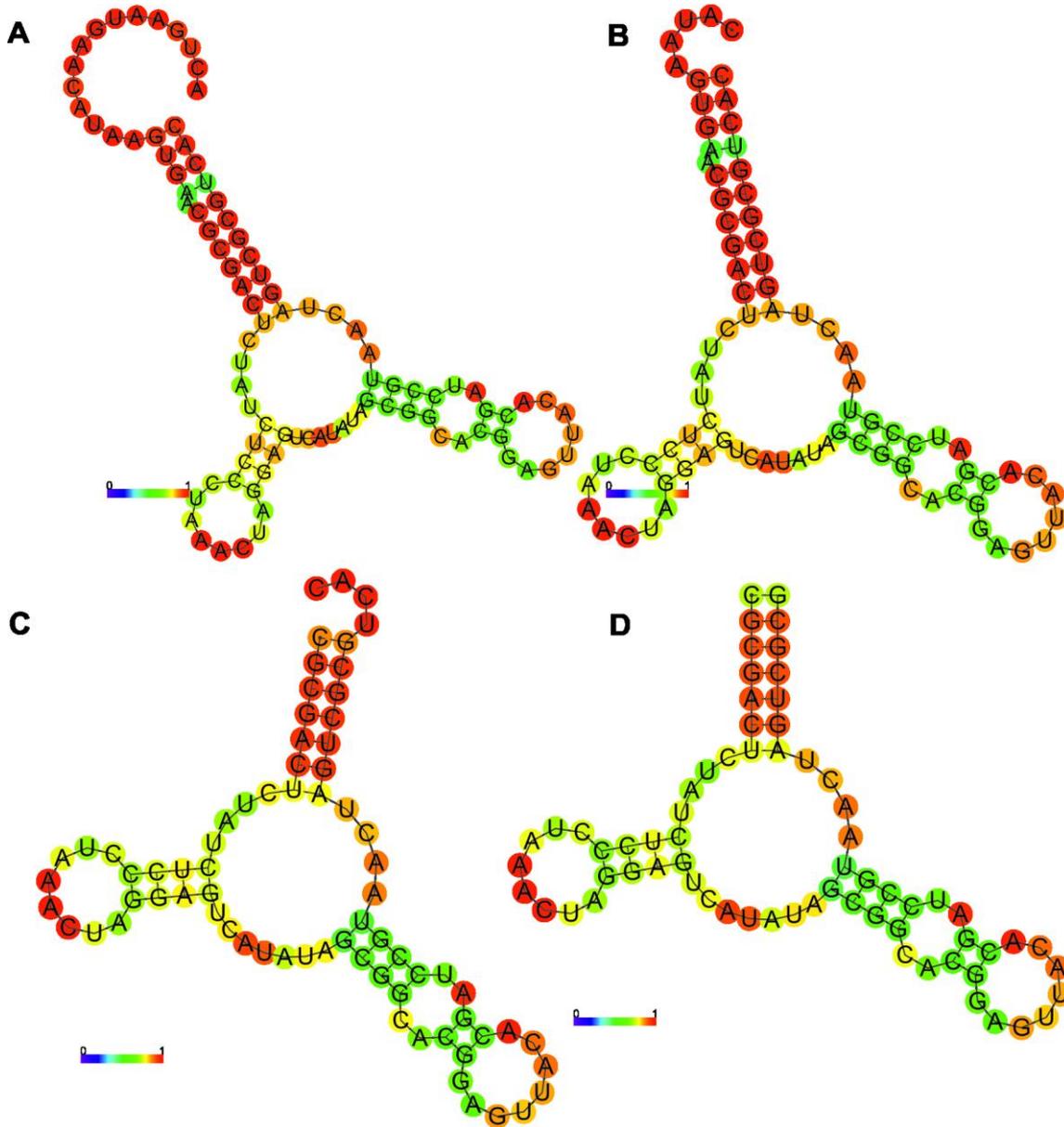


**Figure 4.10 Dose-dependence of sequence 10A on c(CFX).** The dose-dependency of sequence 10A on increasing concentrations of CFX was assessed. Sequence 12A was used as a control. Cells were grown in 1.5 ml cultures in 24-well plates in two independent triplicates in absence and presence of increasing concentrations of CFX for 24 h and analyzed in bulk with a Fluorolog instrument. Basal expression was calculated relative to GFP fluorescence exhibited by the equally treated positive control. Background fluorescence was subtracted and fluorescence output normalized to cell growth. The neomycin-sensitive riboswitch M4 was used as a control and treated equally.

Truncation experiments were focused either on the conservation of the 11 nt closing stem or the introduction of low-impact modifications, because it was unknown to what extent the ligand binding site, that was thought to reside at the core of the bulged three-way junction, extended into this stem region. Constructs 10A(-)10\_0 and 10A(-)20\_0 were truncated 10 and 20 nt from the 5'-end whereas the 3'-end was kept unaltered (Figure 4.11A, B). This way the complete 11 nt closing stem was conserved and only the flapping 5'-end reduced. Constructs 10A(-)30\_0 and 10A(-)30\_4 were cut 30 nt from the 5'-end

## Results & Discussion

and the latter additional 4 nt from the 3'-end, resulting in a closing stem of 7 nt with a 4 nt 3'-flap or an even hybridization terminus (Figure 4.11C, D).

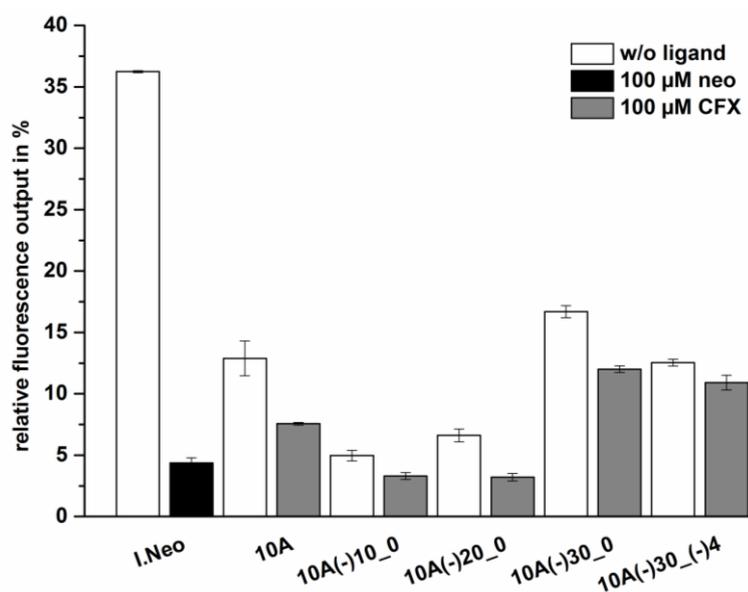


**Figure 4.11 Predicted secondary structures of the truncated versions of sequence 10A by Vienna RNAfold. A** Construct 10A(-)10\_0, **B** construct 10A(-)20\_0, **C** construct 10A(-)30\_0 and **D** construct 10A(-)30\_4. The 5'-end of each sequence is always displayed on the left hand side. Base pairing probabilities are encoded by the heat map showing the probability of being paired or unpaired on a scale from 0 (low) to 1 (high).

The switching phenotype could be maintained for the first three constructs with varying basal expressions (Figure 4.12). Although the sequence load was reduced, basal expression was diminished by the 10 nt and 20 nt truncations from the 5'-end. Nonetheless, this is in accordance with an enhanced probability for closing stem formation by the reduction of the 5'-end that in turn had undergone only weak intra-strand interactions before the truncation and thereby lowered the hybridization efficiency of the actual closing stem. Only the 30 nt truncation restored or elevated *GFP+* expressions relative to construct 10A in the absence of CFX. Again, this coincided with a reduced probability for stem formation

## Results & Discussion

by a truncation of the closing stem length from 11 to 7 nt. Since the switching phenotype was abrogated by only a minor 4 nt truncation of the 3'-end, the closing stem might play a role in the formation of the binding pocket. It was not elucidated, whether this was due to the stem structure itself or tertiary interactions of the stem with other secondary structures of the three-way junction. Further, the extent to which the upstream 5'-UTR affects truncation steps and thus closing stem integrity is unknown. In summary, the experiments highlighted the occurrence of an improved switching factor of 2.1-fold found for the 20 nt truncation and underline the difficulty of rational design operations without sufficient data on the ligand binding site and global folding constraints imposed by up- and downstream framing sequences and truncations.



**Figure 4.12 Truncation experiments of sequence 10A.** A small set of defined truncations of sequence 10A were assessed in the GFP+ reporter gene assay. Cells were grown in 1.5 ml cultures in 24-well plates in two independent triplicates in absence and presence of 100 μM CFX for 24 h and analyzed in bulk with a Fluorolog instrument. Basal expression was calculated relative to GFP fluorescence exhibited by the equally treated positive control. Background fluorescence was subtracted and fluorescence output normalized to cell growth. The neomycin-sensitive riboswitch M4 was used as a control and treated equally.

### 4.1.2 Discussion

The eukaryotic translation machinery is controlled by a variety of mechanisms and stimuli. Importantly for the *in vivo* screening system, local mRNA organization, structure and stability play key roles to facilitate ribosome assembly and translation initiation. The basic concept of this system is to interfere with ribosomal scanning and the initiation of translation at the start codon by inserting a road block in the 5'-UTR of the mRNA. To do so, select aptamers were cloned into the 5'-UTR of the GFP+ reporter gene and assayed for their ability to regulate gene expression in a ligand-dependent fashion. Despite a few contrary reports, none of the tested aptamers was found to act as a riboswitch in the employed set up<sup>136,137</sup>. Because it was sought to determine the potential of readily available and fully evolved aptamers to act as riboswitches, the experiments were performed in a trial and error style, without a detailed adaption of the aptamer sequences to the genetic context. This is in accordance with the two previously mentioned publications by Werstuck & Green and Grate & Wilson<sup>136,137</sup>. In these studies the

## Results & Discussion

authors explored the same putative *in vivo* property of aptamers and successfully used the MG-4 aptamer to control the yeast cell cycle. However, only a morphological phenotype was assayed and western blotting was performed to support the phenotypic appearance. Basal expression levels were not presented, although a 10-fold-change of the Clb2 protein after addition of the MG-analog tetramethylrosamine was calculated by chemoluminescence levels after western blotting. These results underline the need for uniform reporter assays to directly compare two studies. In this study the MG-4 aptamer depressed *GFP+* expression to 3% leaving only marginal room for a switching activity, even if the cognate ligand tetramethylrosamine had been supplied and it is not clear at what expression levels the original aptamer-ligand pair operated.

Instead, it is very likely that a major issue in the generation of 5'-UTR riboswitches was encountered here. Originating from presumably maximally enriched SELEX rounds, all aptamers were selected for the highest possible affinity towards their cognate ligand. Thus, an aptamer with high affinity for its target is very likely strongly pre-structured and may therefore interfere extensively with basal gene expression rendering it inactive for transformation into a riboswitch. The link between high affinity and extensive pre-structuring is given by the selection pressure inherent to the SELEX approach. Only those aptamers may be subjected to the next selection round that could withstand the stringent conditions under which selection was performed. This means, the closer the actual folding resembles the perfect folded aptamer the higher the affinity for the target. Accordingly, they strongly interfere with gene expression leaving no room for a further regulation. On the other hand, aptamers with very little effect on basal gene expression, X1sl and D1 (both tobramycin), 8-1-1 (kanamycin B) and Dano-I (danofloxacin), do not necessarily exhibit low affinity towards their ligand, but are rather small or too weakly structured to pose an obstacle to the 43S scanning complex. Moreover, most SELEX protocols contain unphysiological concentrations of  $Mg^{2+}$ , enhancing aptamer-ligand interactions and the overall folding. The dependence of the SELEX on the  $Mg^{2+}$  concentration was shown to impact low-affinity binders to a greater extent and that high-affinity binders could also be generated with physiological  $Mg^{2+}$  concentrations <sup>193</sup>.

Generally, the ability of an aptamer to either undergo extensive structural rearrangements upon ligand binding or to display an ensemble of different ligand-free ground state conformations in thermodynamic equilibrium, are prerequisites to confer ligand-dependent regulation of gene expression. One of these conformations is then captured by the ligand and stabilized. Several experiments in the past have shown that the ligand exclusively binds to an already pre-folded structure, typically energetically disfavored and short-lived, inducing only minor conformational rearrangements <sup>160,161,163,164,166</sup>. This conformational capture mechanism is mandatory for *in vivo* activity as it allows for the ligand-free ground states to be largely unstructured, hence to promote ribosome scanning, and on the other hand to spontaneously fluctuate into the folded conformation which is specifically bound and trapped by the ligand, hence to prevent ribosome scanning. For the neomycin-dependent aptamer a switching element could be identified that destabilizes the ligand-free ground state, weakening the global fold of the riboswitch and resulting only in a moderate reduction of basal gene expression <sup>160</sup>. In turn, addition of the ligand shifts the thermodynamic equilibrium towards the entropically disfavored, folded conformation which is captured by the ligand. The complex is stable due to the beneficial release of enthalpic energy compensating the entropic penalty. Therefore the neomycin-sensitive riboswitch gains its *in vivo* activity from the fact that the ligand-free conformations are loose and rather unstructured and above all highly

## Results & Discussion

populated in comparison to the folded conformation that can only steadily exist when captured by the ligand. This ensures relatively high gene expression values without ligand and a decrease of gene expression upon ligand addition.

The *in vivo* activity of the tetracycline-dependent riboswitch is governed by a transiently folded loose tertiary structure with high ligand affinity that is captured by the ligand as well. Since the ligand-free conformations fluctuate on biological relevant time-scales and do mostly not exhibit affinity to tetracycline, basal gene expression values are greater than 20%. A transiently occurring tertiary structure serves as a scaffold for ligand docking, resulting in only minor conformational changes towards the ligand-bound conformation that impairs expression from the controlled gene<sup>163,164,166</sup>. Although some of the aptamers selected for this study exhibit conformational rearrangements from ligand-free to ligand-bound state, they did not exhibit a riboswitching phenotype within the 5'-scanning mechanism<sup>178,184,194</sup>.

Aptamer domains pruned off a natural riboswitch might not work, because the natural design requires an expression platform that receives the aptamer-mediated signal for a regulatory activity. Recent studies have highlighted the feasibility of a substitution of natural aptamers for other natural or synthetic aptamers that were known to exhibit a regulatory phenotype. Connected to the expression platform of the original aptamer, the regulatory property of the riboswitch as such could be preserved<sup>128,129</sup>. Additionally, it has been shown in several studies that aptamers derived from natural riboswitches can be fused to different expression platforms by an empirically designed communication module<sup>134,143,144</sup>. The artificial hybrid made from aptamer and mostly a ribozyme displays high regulatory activity in pro- and eukaryotic organisms, but operates by a completely different mechanism than the 5'-UTR design. Considering both examples, the aptamers are basically suited to be incorporated into engineered riboswitches, but needed to be completely re-designed with regard to their structure to function in the pursued 5'-UTR design. The mTCT8-4 aptamer binding to theophylline represents such a proven super-versatile and portable RNA structure that even conferred weak *in vivo* activity in this study. The aptamer itself is SELEX-derived and has, besides other applications, been used in bacteria to control the access to the ribosome binding site (RBS) by coupling to a short expression platform<sup>121,125,195</sup>. The design principal is similar to the 5'-UTR design, but runs in a quite different environment. It was included in this study to investigate whether its flexible nature would allow a direct transferability to the 5'-scanning mechanism. Much for the same reasons, the aptamer domain N40 of the atrazine-sensing riboswitch was chosen. Unfortunately, the publication was retracted shortly after this study had been performed and it is unknown, whether the atrazine aptamer domain actually belongs to the group of aptamers with a regulatory profile.

In conclusion, high affinity aptamers from an *in vitro* selection or isolated from an existing riboswitch, be it natural or engineered, might not directly be used as a 5'-UTR riboswitch in yeast. In the light of a conformational capture mechanism, any aptamer sought to be used as a 5'-UTR riboswitch must feature two or more ligand-free ground states in thermodynamic equilibrium, whereupon the favored conformation must be weakly structured to allow for sufficient basal expression. The ligand will then selectively trap a more strongly structured conformation and thus gene expression is decreased. The greater the difference between the ligand-free and ligand-bound conformations the better the regulatory *in vivo* activity expressed as switching factor. Moreover it has been found that variations of the closing

## Results & Discussion

stem were sufficient to adjust basal expression and switching factors <sup>138</sup>. Albeit this could suggest the adaption of an aptamer's closing stem as a guideline to transform these RNA structures into 5'-UTR riboswitches, modifications of the closing stem are not enough. The strategy to convert an aptamer into a riboswitch is governed by trade-offs between the presence of a loosely pre-structured RNA and the ability to efficiently fold into a stable global architecture. Hence, it must allow a sufficient sequence space to accommodate structural motifs that facilitate an energetic separation of the ligand accessible conformation from the ground states, such as the switching motif identified in the neomycin riboswitch <sup>160,161</sup>.

The *in vivo* screening for CFX-sensitive riboswitches thus contained two different SELEX pools to suffice these criteria. A moderately (R6) and a highly (R9) enriched pool are generally needed, because only from a broad spectrum of sequences candidates can be identified that satisfy the selection criteria of medium to high basal expression and simultaneously exhibit the conformational space required to reduce it. Sequence space itself is a product of the pool diversity which is given by the length of the randomized region, but drastically diminished as the enrichment of binding aptamers proceeds <sup>196</sup>, and the transformation efficiency in yeast. Initially, cells were selected for their GFP fluorescence by microscopy to ensure (i) correctly recombined plasmids and (ii) the occurrence of aptamer sequences that incorporated no premature out-of-frame start codons and whose global architecture promoted scanning of the 43S complex for the original start codon. Aptamer pools enriched over many rounds are normally extensively pre-structured and lose their competence to fluctuate between loose and ridged conformations eventually resulting in low basal gene expression and no switching activity. With 57% of the selected colonies stemming from R6 both SELEX rounds impacted basal expression levels to an almost equal extent. Accounting for the results after the first screening, regulatory active colonies however, were found to originate rather from R9, as a regulatory activity generally follows lower basal expression levels, since they supposedly imply sufficiently pre-structured aptamer sequences.

The values for basal expression and switching factors were determined based on the overall assay results. As such, the thresholds set for both aptamer properties were mainly dictated by the experimental set up. Theoretically, the basal expression cannot be smaller than 1, because expression without ligand is divided by expression with ligand and the latter is supposed to be reduced. Nonetheless, these values are inherently connected to the expression levels of the positive controls and the background. If these values are influenced by CFX in the assay and this is not correspondingly mirrored by the aptamer-containing cells, values lower than 1 may be obtained. In addition, the first screening round was not corrected for cell growth, because the original set up required to culture the cells statically. As this was seen to have a negative impact on the assay itself, the second screening was performed with cells cultured in a platform shaker and expression levels were corrected for the OD<sub>600</sub>. The influence of these factors was also reflected by the cells transformed with the neomycin riboswitch. Normally exhibiting switching factors between 8- and 10-fold, the assay produced factors of only ~3-fold. The lower threshold for basal expression was set to 5%. This values still allows to detect a regulatory activity as it was several fold higher than background fluorescence.

After the second screening all plasmids were isolated and propagated through *E.coli* to ensure that yeast cells were only transformed with a single plasmid. The ability of yeast cells to simultaneously maintain more than one plasmid may result in the occurrence of two or more different aptamers during

## Results & Discussion

homologous recombination and thus corrupting assay results. Therefore some clones lost their regulatory phenotype after that “purification” step. Because only one *E.coli* colony had been processed, only one of the possible multiple plasmids was propagated. At this point, aptamers displaying *in vivo* activity might have been lost.

Truncations of the regulatory active clones 2B, 6H and 10A yielded a variety of different basal expressions, but abrogated their switching phenotype completely. Since it is unknown which nucleotides form the ligand binding pocket and how secondary and tertiary interactions contribute to the overall folding, the extensive truncations performed to remove presumably unnecessary sequence load had most probably destroyed crucial interactions conferring regulatory activity. For the same reasons no correlation between the stability and length of the closing stems and basal expression levels could be deduced. Contrarily, more detailed, low-impact modifications of the aptamer sequence of clone 10A allowed such a correlation and a better switching factor could be gained; but, as described above, only at the cost of a lower basal expression. Further investigations of that sequence led to the discovery of a dose-dependent switching phenotype supporting the identification of a CFX-sensitive riboswitch. Although unearthed from 1300 fluorescent colonies, aptamer 10A exhibited only a 2-fold reduction of gene expression that is of no use to conditionally control a gene of interest. Foremost, a detailed analysis of the ligand binding core would aid to improve the sequence length and enable modifications enlarging the “switching window” limited by basal and ligand-repressed gene expression. Recently, such an *in vivo* selection system was established in baker’s yeast <sup>149</sup>. Here and in a similar approach, riboswitch identification was achieved by applying positive and negative selective pressure to generate a functional window in which the riboswitch is supposed to operate <sup>171,197</sup>. Such tailored gene expression profiles may not only be yielded by selection pressure, but can also be obtained through sole selection from the whole population by FACS-based approaches <sup>198</sup>.

The presented *in vivo* screening system would undoubtedly benefit from the introduction of an automatic and unbiased pre-selection to avoid the screening and processing of phenotypes that suffice either only the criterion for basal expression or exhibit switching factors that are unfit to be used, because they operate in the lower 5%-range of gene expression.

### 4.2 Project II: Tetracycline-Dimers: A massive approach towards the *in silico* prediction of riboswitch performance

A major milestone in eukaryotic riboswitch design would be the construction and calibration of an *in silico* model from rich *in vivo* data to assess sequence-structure-activity relationships (SSAR). In this study two tetracycline-binding riboswitches were set in tandem to control the expression of a *GFP+* reporter gene by the 5'-scanning interference to ultimately feed a computational model with the gathered *in vivo* data. It was known from preliminary experiments that variations of the closing stem of the start codon proximal riboswitch (3'-Tc) yield a sufficient range of different basal expressions (BE) and switching factors (SF) to calibrate the model. A tandem construct was chosen, because in later experiments it was also sought to determine the influence of the spacer unit between both riboswitches on their SSAR, as a context-dependent performance constitutes another major issue with the application of riboswitches. Besides, it was previously demonstrated that Tc-dimers or even trimers dramatically enhance the regulatory performance compared to the monomer<sup>168</sup>. The model was built and calibrated by Sven Jager from the group of Prof. Hamacher.

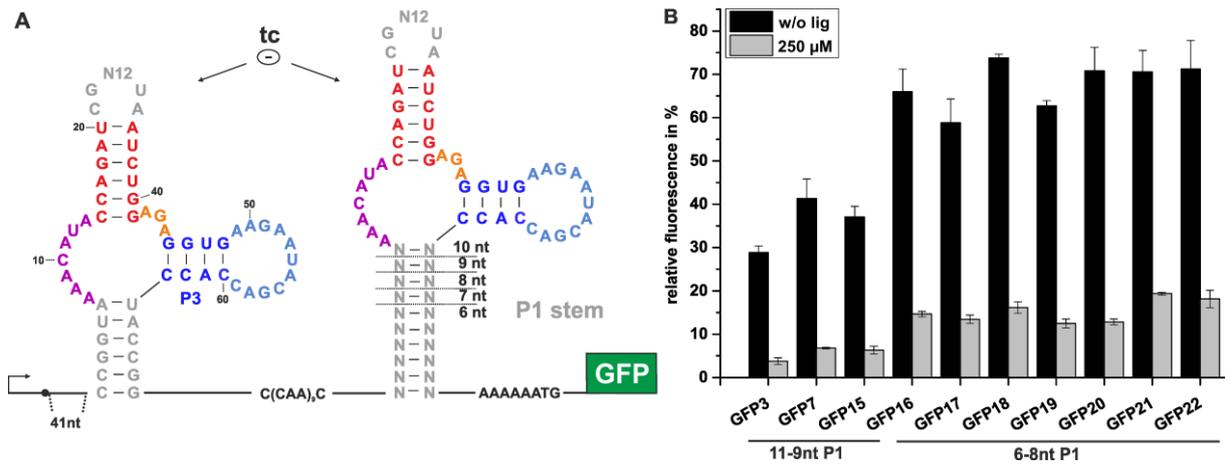
#### 4.2.1 Results

##### 4.2.1.1 Parameter identification for the design of a first set of tandem tc-riboswitches to train the model

The constructed GFP reporter system is depicted in Figure 4.13A. The two tetracycline-binding (tc) riboswitches are placed start-codon proximal in the 5'-UTR of the *GFP+* reporter gene. Both riboswitches are spaced to a total of 27+2 nt including nine CAA-spacer units. The CAA-spacer units generally constitute an unstructured insulation module to allow the riboswitches to fold independently. As a starting experiment to assess parameters relevant to design a large set of different tandem constructs the possible range of BE and SF was probed. While the 5'-Tc was kept constant, the closing stem of the 3'-Tc was randomly mutated with stem lengths varying between 6 and 11 nt. These eight new constructs (GFP15-GFP22) were tested alongside two constructs (GFP3, 7) previously designed by Dr. Julia Weigand and Lara Gorini (group of Prof. Suess) in a GFP reporter gene assay.

Thus, three independent triplicates of yeast cultures each transformed with a single construct were cultured in absence and presence of 250  $\mu$ M tetracycline (tc) for 24h in 24-well plates and their fluorescence determined. Vectors pHWE601\* and pWHE601 were used as negative and positive controls, respectively, and treated equally (Figure 4.13B). As a general rule, longer stems result in lower BE and higher SFs. Based on these results lower and upper limits for P1 stem length and free energy of the 3'-Tc were determined, since only the closing stem of this riboswitch was to be modified. The minimum free energy (MFE) of all 3'-Tcs was calculated by RNAfold<sup>174,175</sup>.

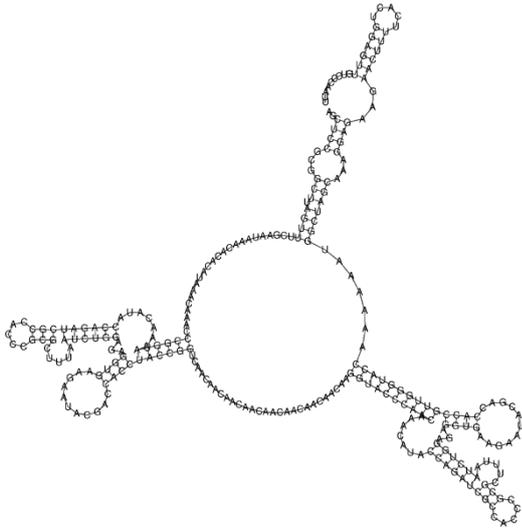
## Results & Discussion



**Figure 4.13 Tc-dimer design and GFP reporter assay.** **A** Tc-dimers were constructed from the tc-binding riboswitch AN32sh<sup>155</sup>, spaced by 29 nt and placed start codon proximal in the 5'-UTR of the *GFP+* reporter gene. The closing stem variations of the 3'-Tc refer to the design of a first set of 96 different 3'-Tcs. Addition of tc represses gene expression as shown in **B** and according to the 5'-scanning interference. Ten different Tc-dimers with respect to the 3'-Tc were assayed for GFP fluorescence. The length of the P1 stems is indicated. Yeast cells were transformed with plasmids bearing the different Tc-dimers and cultured in 1.5 ml 24-well plates in absence and presence of 250  $\mu$ M tc for 24 h. Expression was measured in bulk normalized to cell growth, corrected for background fluorescence and calculated relative to the positive control pWHE601 that was equally treated.

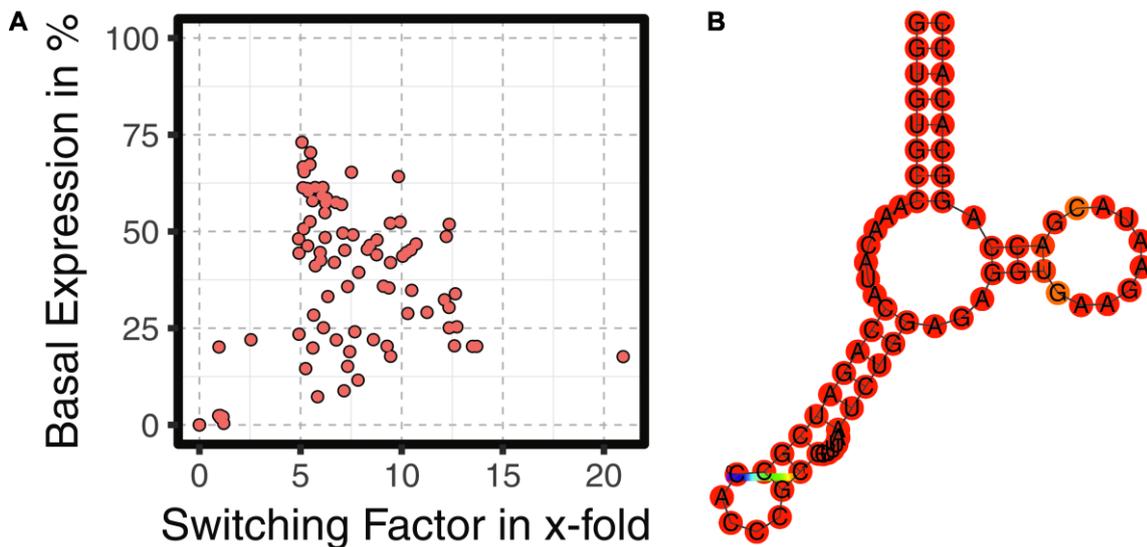
A first set of 96 different tandem constructs was designed considering the following criteria for the 3'-Tc. (i) The stem length should be between 6 nt and 10 nt and (ii) the MFE between -20 and -30 kcal/mol to yield fluorescence outputs in the range of ~20% to ~70% BE as observed in Figure 4.13B. Lower or higher BEs were disregarded, because this either leads to riboswitches operating at a very low expression levels and a further reduction of gene expression is impractical (<20%) or to riboswitches that cannot efficiently respond to tc presence as their BE is too high (>60%) to allow the occurrence of a sufficiently pre-formed conformation. Additionally, it was deduced from the experiments pictured in Figure 4.13B, that MFEs between -20 and -30 kcal/mol would yield functional riboswitches. Ultimately, this meant for the design space that five blocks with regard to stem lengths of 6 nt to 10 nt were pre-set and the ~20 sequences within each block exhibited MFEs between -20 and -30 kcal/mol. Sequences were generated by a randomization algorithm (developed by Sven Jager), trimmed to compute a high sequence entropy (nucleotide composition within a given sequence, here the P1 stem) and clustering diversity between the five blocks (inter, Levenstein distance) and the within the block (intra, Hemming distance). Predicted sequences were then subjected to a folding test, checking for folding constraints of the secondary structure of the tandem constructs flanked by 40 nt upstream and 150 nt downstream (see example in Figure 4.14).

## Results & Discussion



**Figure 4.14 Secondary structure of a Tc-dimer flanked by 40 nt upstream and 150 nt downstream.** The folding algorithm used by RNAfold was used to predict this secondary structure.

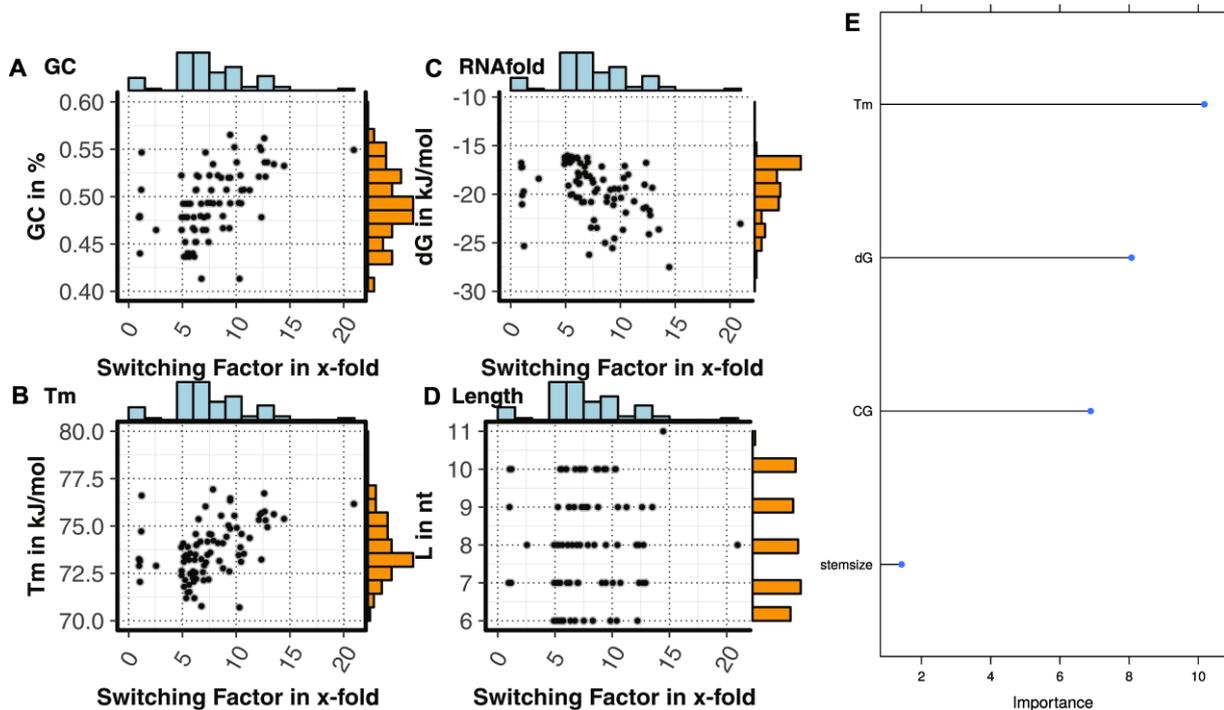
Moreover, only closing stems devoid of the triplet ATG were selected to circumvent issues with premature translation initiation. Constructs were ordered as two primer sets, one forward and one reverse, and PCR amplified from the hybridized core of the 3'-Tc. Fwd and rev primers introduced the restriction sites *AgeI* and *NheI* that were used to ligate the constructs with vector GFP3 cut with the identical enzymes. After passage through *E.coli* and sequencing, yeast strain RS453 $\alpha$  was transformed with plasmids bearing a single construct and the negative and positive controls. Cells were cultured as indicated for the experiments shown in Figure 4.13B and assayed by flow cytometry, making use of an integrated auto sampler. Each construct was analyzed in two independent duplicates and the results are shown in Figure 4.15A.



**Figure 4.15 Classification of 100 Tc-dimers.** **A** Tc-dimers were classified according to basal expression and switching factors. Yeast cells were transformed with plasmids bearing the different Tc-dimers and cultured in 1.5 ml 24-well plates in absence and presence of 250  $\mu$ M tc for 24 h. Expression was measured by flow cytometry and calculated relative to the positive control pWHE601 (basal expression). The switching factors were calculated as the ratio of relative gene expression in absence and presence of tc. The Experiments were performed in two independent duplicates. **B** MFE structure of the best performing riboswitch from the first set as calculated by RNAfold.

## Results & Discussion

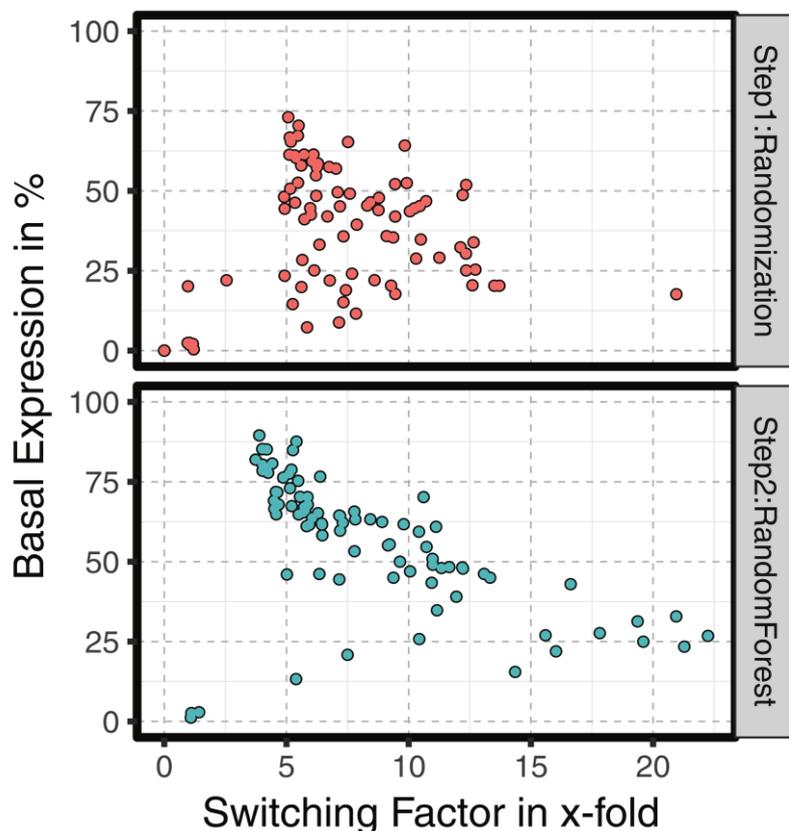
In Figure 4.15A the performance of 100 constructs is plotted according to BE and SF. Values for BE and SF range mainly between 20% and 70% (averaging at 34%) and 5-fold to 15-fold (averaging at 6.4-fold), respectively. Some constructs display no regulatory phenotype, whereas one construct exhibits a ~23-fold switching at ~25% BE. This construct is shown in Figure 4.15B. It features a 8 nt closing stem, a MFE of -24.4 kcal/mol and a GC content of 75%. The results support the overall design of the first set, providing a wide range of BEs and SFs to train the model. Moreover, already before the training with *in vivo* data, the model had predominantly generated sequences sufficing the criterion set for BE levels. To assess putative correlations between the SF and the P1 parameters melting temperature ( $T_m$ ), MFE, stem length, GC-content (GC%) and folding time, the parameter values were calculated and plotted against the SFs of each construct (Figure 4.16A-E). The structure prediction heuristic employed by RNAfold was used to compute the MFE (Figure 4.16C). No correlation was found for stem length, as basically all SFs were yielded with all different stem length (Figure 4.16D). The GC-content, contributing to stem stability and thus to  $T_m$  and MFE, showed a weak correlation to the SF, but a tendency for higher GC% to allow a broader range of SFs; also the highest SF of ~23-fold (Figure 4.16A). MFE and  $T_m$ , two closely related parameters showed the same correlation to SFs as basal expression levels, namely that higher SF are obtained with lower MFE and thus lower basal expression. Though the correlation was also weak, it confirmed the tendency observed for GC%. The importance of each parameter for riboswitch performance is summarized in Figure 4.16E.



**Figure 4.16 Influence of various inherent folding parameters on switching factors.** The different graphs show the SF plotted against the GC% (A), the melting temperature  $T_m$  (B), the MFE as computed by RNAfold (C), and the stem length (D). E rates all parameters according to their importance for the SF.

#### 4.2.1.2 Design of a second set of Tc-dimers

Considering the low correlation between SF and BE (data not shown) and the calculated parameters, the model is not a simple linear equation, but must rather account for all parameters to compute a different relationship. A random forest was now trained with the generated *in vivo* data to compute a better connection between all parameters and SF and BE. The random forest describes the analysis of many thousands of decision trees interrogating the parameter set and reconciling the parameter values with their corresponding BE and SF. In such way the values of all parameters are linked by their mutual BE and SF. Essentially, this represents the training of the random forest that was based on the empirically *in vivo* determined dependence of SF and BE on the *in silico* calculated parameters. The random forest was then used to classify random sequences sufficing the criteria of BE > 20% and SF > 8 and all other criteria proposed for the first data set. To reduce solution space it was important that BEs and SFs spanned a wide range of different values. Ultimately, a second set of 96 different 3'-Tcs was generated. Only sequences exhibiting predicted BEs > 20% and SFs > 8 were selected, cloned, subjected to the described reporter assay and analyzed.



**Figure 4.17 Comparison of riboswitch performance as obtained by randomized mutagenesis of the P1 stem or computed using a random forest.** Different Tc-dimers are plotted with respect to BE and SF. Results from the first set generated by **randomized mutagenesis** of the P1 stem are shown in the upper panel. The results from the second set based on the trained **random forest** are depicted in the lower panel. Experimental procedures and calculations were performed as described for Figure 4.15.

Results of the experiments of the second dimer set are displayed in the lower panel of Figure 4.17. These results partly mirrored the first data recording, but also showed that the model can reproduce the sequence-activity relationships that it had learned from the first recording. Moreover, non-regulating sequences were almost completely erased and the abundance of switches with desired switching factors increased to a mean SF of 8.3-fold compared to the previously obtained mean SF of 6.4-fold and an average BE of 56%.

## Results & Discussion

The relationship between BE and SF became now more apparent than in the first set. Higher BEs, due to a higher MFE and thus highly destabilized ground states, correlate with lower SFs. Vice versa, high SFs are only obtained with BEs < ~30%. In total, the model was able to improve the correlation between BE, SF and the parameter set. Unfortunately, there is still a significant abundance of Tc-dimers exhibiting forbidden values of BE (> 60%) and SF (< 5-fold), suggesting ample room for a refinement of the model.

### 4.2.2 Discussion

The redesign of functional natural or artificial riboswitches towards higher basal gene expression and better switching factors still poses a fundamental challenge. Many approaches are based on permutation, selection and screening, following base pairing and simple secondary structure folding rules and considering structural data from advanced spectroscopy like NMR or fluorescence spectroscopy. While these approaches have led to the construction of several riboswitches, the design-built-test cycle is generally cumbersome and resource-intensive<sup>33,48,77,86</sup>. Recently, several rational design principles were proposed for the construction of riboswitches in bacteria. Capitalizing on the availability of many resources on structure, regulatory mechanisms and advancing computational modeling capacities, they demonstrated the applicability of *in silico* designs for facilitating *in vivo* activity. While the substitution of natural aptamer domains<sup>129,156</sup> or the mutation of natural binding sites<sup>130</sup> to generate different ligand specificities are promising approaches to retain existing regulatory activity, two other approaches on the *de novo* construction of transcriptional<sup>83,127</sup> and translational<sup>84</sup> riboswitches are similar to the scope of this study. The regulatory mechanism of transcriptional riboswitches was thought to rely mostly on kinetic aspects like folding traps of the terminator structure, but it was shown that also the stability, a thermodynamic property, of the terminator hairpin inflicted major pressure on its performance. Moreover, it was demonstrated that dimers or trimers of the used theophylline riboswitch enhanced performance in a linear dose-dependent manner. Before the conduction of the presented study, similar behavior had also been reported for dimers and trimers composed of the tc-binding riboswitch in yeast<sup>168</sup>. In the present study the serial dimer arrangement led again to improved regulatory activity as compared to the monomer. Additionally, since the dimer is spaced by a defined number of nucleotides, this arrangement also allows the assessment of context dependencies in terms of spacer identity in future experiments.

Returning to the modeling of the transcriptional riboswitches, it was important that the model would also consider the inverse folding problem, i.e. that any predicted sequence can adopt the conformation of the mutually exclusive structures of terminator or anti-terminator. In the present study BE and SF account for the interplay between ON and OFF states. While it was sought to obtain riboswitches with loose ligand-free ground states, thus permitting ribosome scanning and represented by a high (> 20%) BE, the corresponding SF would reveal whether the ligand-free ground states had also comprised a short-lived, entropically unfavored pre-formed conformation that, if bound by tc, would reduce gene expression. It was subsequently discovered that BE and SF were antiproportional, i.e. high BEs (rare occurrence of pre-formed conformations or weakly pre-structured conformations) incurred lower SFs. The sequence exhibiting the best performance after analysis of the first set confirmed this hypothesis as it features a low MFE and high GC content for the 3'-Tc. The model was thus trained with the

## Results & Discussion

empirically derived *in vivo* data to discern a possible correlation between BE, SF and several thermodynamic and kinetic parameters. After the first set had been analyzed no correlation was evident and so the second set was based on model predictions obtained through training and not correlation. In part this mirrors the experimental procedure of the transcriptional riboswitch design as it was reported that during the *in vivo* testing optimizations were only possible by rational evaluations of the genetic context, information the *in silico* model had not provided. Besides context, also structure-stabilizing effects by ligand binding and the formation of pseudoknots were not computed.

The development of a statistical thermodynamic model in a recent study predicted the sequence-structure-function relationship for translation-regulating riboswitches in *E. coli*<sup>84</sup>. The model was built under the assumption of thermodynamic equilibrium between the ribosome, mRNA and ligand. Additionally, aptamer structure, affinity, mRNA sequence context and macromolecular crowding were considered. Interestingly, the model did not compute changes of mRNA stability, although it accounted for the switching free energy. The latter is exactly the parameter the SF describes in the present study and seems as such suited for model building and training. Sequence context was included as a design parameter and was used to test whether it would result in misfolded riboswitches rather than whether it would exhibit performance constraints. The influence of ribosome binding energetics on riboswitch performance with respect to ligand-free and ligand-bound conformers was not regarded. As the 5'-scanning mechanism relies crucially on mRNA-ribosome interactions, such parameters should be incorporated into the model.

Co-transcriptional folding seems to be another parameter that could improve *in silico* modeling of translational riboswitches. As a kinetic parameter it was thought to dominate the regulatory mechanism of transcriptional riboswitches. Both studies on tc riboswitches in yeast<sup>139</sup> and the discussed design of translational riboswitches in *E. coli*<sup>84</sup> have found a higher regulatory activity when the ligand was supplied before or co-transcriptionally compared to after transcription had been induced. Supplementation before transcription yielded the highest activity. This may be explained by the occupancy of the mRNA with translating ribosomes, which is especially pronounced in bacteria, as translation itself occurs co-transcriptionally. Once subjected to the translation machinery, the spatiotemporal chances of structural rearrangements are distinctly limited, compromising the regulatory activity. If included in the model, these data could set boundaries to riboswitch performance, an upper threshold of efficiency that cannot be overcome due to factors extrinsic to the SSAR.

Other limitations could be inherent to the system, such as whether high BE and high SF could ever be reached. At the moment the model is trained and can compute a classification with respect to both parameters. More data could result in a regression, e.g. by regression trees obtained by machine learning, yielding explicit BEs and SFs and the model could then be used to simulate the maximum *in vivo* accessible performance space. Another drawback is the current inability to sufficiently consider pseudoknots due to limitations of the computing time and insufficient validation of the developed algorithms. This is especially important in the case of the tc riboswitch that displays an H-type pseudoknot-like tertiary interaction between the interhelical junction of J1-2 and the loop L3 (see Figure 2.11).

Turning towards benefits, the developed method exhibits high portability to other biological systems. This feature can be attributed to the universal approach itself. First, some element, motif or sequence

must be identified that is susceptible to randomization for machine learning. Next, rationales must be inferred from preliminary experiments, pointing towards logic, outer boundaries of the randomization and to confirm a direct dependence of the sought re-design on the investigated parameters. Then the generated data is subjected to train a random forest and improve the model. The model is subsequently used to classify random motifs, elements or sequences according the desired solution space. At last, another data set is generated and the first feedback loop closed. The more iterative cycles are performed the more accurate the model will evolve and may eventually even be calibrated on rich *in vivo* data to correctly compute output parameter.

### **4.3 Project III: ROC'n'Ribo: Characterizing a riboswitching expression system by modeling single-cell data**

#### **4.3.1 Results**

RNA-engineered systems offer simple and versatile control over gene expression in many organisms. In particular, the design and implementation of riboswitches presents a unique opportunity to manipulate any reporter device *in cis*, executing tight temporal and spatial control at low metabolic costs. Assembled to higher order genetic circuits, such riboswitch-regulated devices may efficiently process logical operations. The following study presents an approach that facilitates the comprehensive single-cell characterization of a novel RNA-controlled devices constructed from the neomycin- and tetracycline-sensitive riboswitches and implemented in *S. cerevisiae*. By coupling two synthetic riboswitches that both block translation initiation by the same mechanism, representing a NOR operation, it was initially sought to improve the dynamic range between ON and OFF state of reporter gene expression (Figure 4.18). With its orthogonal ligands however, the devices could also realize single-input NOT operations. After a preliminary fluorescent reporter assay to assess the general gate performance, one gate was subjected to more extensive analyses (Figures 4.18B and 4.19).

##### **4.3.1.1 Construction of repressor gates operating by NOT and NOR logic**

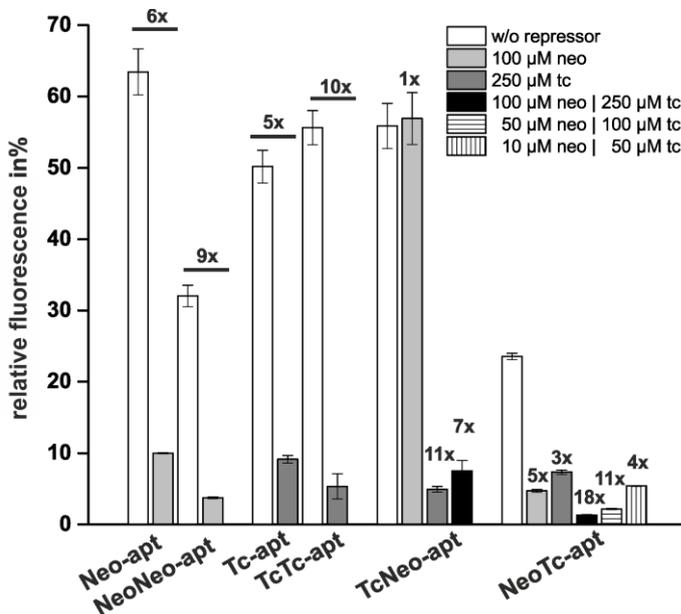
Initially, tandem devices composed of the neomycin- (aptamer M4)<sup>138</sup> and tetracycline- (aptamer cb32-ANsh)<sup>155</sup> binding riboswitches were constructed and inserted in the 5'-UTR of the *GFP+* reporter gene expressed from pWHE601\* by the *ADH1* promoter and terminator (Figure 4.18A)<sup>199</sup>. The resulting dimers capable to transduce only NOT logic are represented by two copies of the M4 (termed NeoNeo) or AN32sh riboswitch (TcTc), respectively (devices not shown). The two NOR gates, 5'-NeoTc and 5'-TcNeo may either be operated with both ligands (NOR gate) or with a single ligand, then constituting NOT gates. The two riboswitches were separated by multiple CAAA-spacer units to minimize global folding constraints between the two RNA architectures<sup>192</sup>. This strategy has also been used in other studies, since CAA(A) repeats do not form any secondary structures and may thus be seen solely as an insulation module. As both riboswitches operate by the conformational capture mechanism, it is crucial for their functionality to facilitate free fluctuations between their ground-state conformations<sup>200</sup>. To assess gate functionality in absence and presence of the repressors fluorescence of the *S. cerevisiae* strain RS453 $\alpha$  transformed with the reporter devices was recorded in a standard bulk measurement after 24 h (Figure 4.18B and 4.19).



## Results & Discussion

Supplementation of both ligands resulted in a concerted response with a switching factor of 18-fold. Moreover, this gate displayed a dose-dependent reduction of its switching behavior on decreasing repressor concentrations. On a first level, the device could thus be used to only moderately inhibit gene expression, whereas its application as a NOR gate led to a multiplied decrease of residual reporter expression to about 2%, yielding an increased level of regulation.

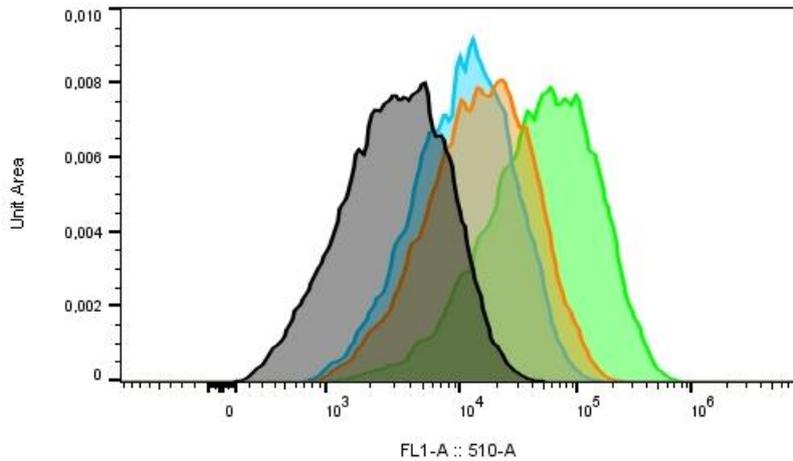
From these data it can be concluded that the 5'-riboswitch contributes more to the switching efficiency than its 3'-orthogonal counterpart. This supports the mechanism of gene regulation pursued here – inhibition of ribosome scanning in 3'-direction. Whether neomycin or tetracycline constitutes the stronger repressor cannot be determined as the single riboswitches perform comparably. In addition, a NOR operation might only be achieved using the 5'-NeoTc gate, as the inverse version 5'-TcNeo showed no regulation at the addition of neomycin. Especially the switching profile of the 5'-TcNeo gate hints at structural interferences between the two parts that have not been resolved yet.



**Figure 4.19 GFP reporter assay of all constructed gates.** To assess the effect of different combinations of the Neo and Tc riboswitches the depicted variants were analyzed by fluorescence spectroscopy in a bulk measurement after 24 h. Cells were cultured in the absence and presence of neomycin (neo) and/or tetracycline (tc) at the indicated concentrations. Output fluorescence was corrected for background fluorescence, normalized to cell growth and is displayed relative to the positive control that is not riboswitch-controlled but otherwise treated equally. SFs are indicated. A dose dependence experiment was performed with the 5'-NeoTc riboswitch, investigating the dependence of switching efficiency on ligand concentration. Experiments were performed in triplicates of at least two independent biological samples.

Accounting for its sequential switching phenotype processing a NOR operation efficiently, the 5'NeoTc gate was further characterized at the single-cell level. To determine the phenotypic distribution with respect to the fluorescence output, flow cytometry experiments with steady-state cultures were conducted (Figure 4.20). Population analysis of the constitutively expressed 5'-NeoTc gate highlighted the broad phenotype arising in the fluorescence profile of cells harboring 2μ-replicated plasmids. Extraction of the mean fluorescence values allowed rebuilding the previously calculated basal expression and switching factors (compare Figure 4.18B).

## Results & Discussion



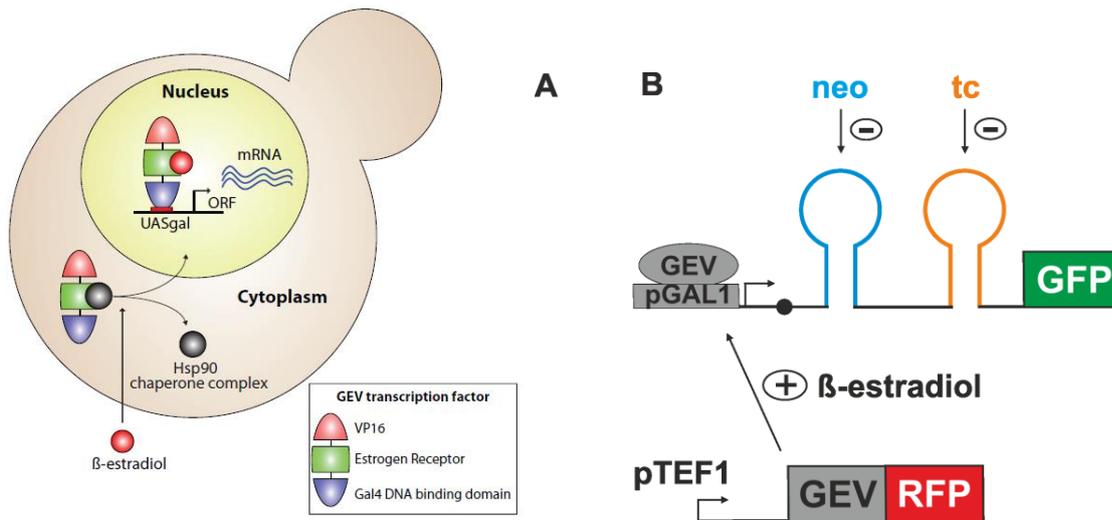
**Figure 4.20 Single-cell analysis of the constitutively expressed 5'-NeoTc gate.** The histogram shows the normalized abundance of each GFP fluorescence level for the unexpressed and repressed populations of the 5'-NeoTc gate analyzed by flow cytometry after 24 h. 10,000 events were recorded per population. The x-axis is plotted biexponentially. The different repressed states are indicated by color: unexpressed (lime green), repressed by neo (blue), tc (orange), neo and tc (black). The presented data is an average replicate of at least three independent replicates. Samples were prepared as described in Figure 4.19.

Besides the high cell-to-cell variability conferred by 2 $\mu$  plasmids, they impose an increased metabolic burden on the host cell, especially when equipped with strong promoters that drive the overexpression of reporter genes<sup>44,204</sup>. Genetic circuits of any complexity constructed from several 2 $\mu$  plasmids will therefore exhibit severe noise propagation and ultimately broad distributions of phenotypes, rendering performance on the single-cell level rather poor<sup>205</sup>. In concert with bulk measurements at the steady state level from outgrown cultures of the constitutively expressed *GFP+*, the expression system as a whole was not suited to assess the transient dynamics of the 5'NeoTc device and to facilitate an increased functionality at the single-cell level.

### 4.3.1.2 Introduction of the GEV system for inducible transcriptional regulation

The expression system was prepared for kinetic analysis by adopting the GEV system developed for rapid, tunable and single-gene specific conditional gene expression in yeast<sup>206</sup>. GEV is a chimeric transcription factor composed of a VP16 transcription activation domain, a Gal4 DNA binding site and, most importantly, a  $\beta$ -estradiol-receptive domain. The transcription factor itself is constitutively expressed from an ARS/CEN plasmid using the *TEF1* promoter and *CYC1* terminator. Additionally, the RFP derivative mCherry is C-terminally coupled. Transcription of  $P_{GAL1}$ -controlled genes is switched on by supplementation of nanomolar concentrations of  $\beta$ -estradiol ( $\beta E$ ). A scheme of the complete inducible system encoded on two different plasmids is illustrated in Figure 4.21.

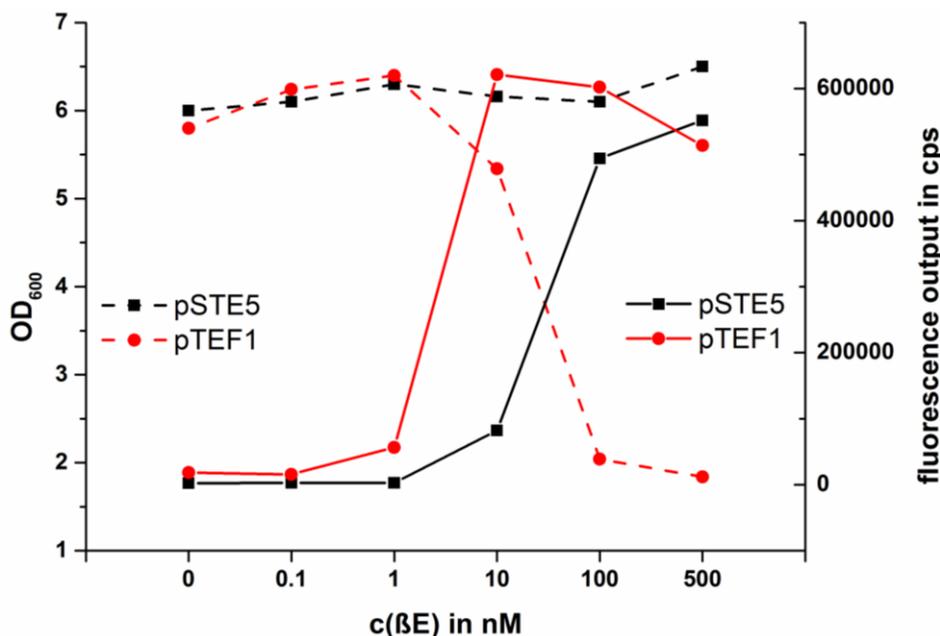
## Results & Discussion



**Figure 4.21 Scheme of the inducible expression system.** **A** The GEV transcription factor is composed of a Gal4 DNA-binding domain, the human estrogen receptor and a transcription activation domain VP16. In absence of  $\beta$ -estradiol GEV is bound to HSP90 in the cytoplasm. The presence of  $\beta$ -estradiol activates GEV and a GEV-dimer locates into the nucleus and induces gene expression of the GAL family. Adopted from <sup>206</sup>. **B** The complete system is encoded on two plasmids. GEV is C-terminally coupled to mCherry and its expression is driven by the constitutive *TEF1* promoter. Supplementation with  $\beta$ -estradiol activates GEV and induces transcription of the 5'-NeoTc gate controlling *GFP*<sup>+</sup> expression. Different Boolean states can be realized by repressor addition.

The promoter of the original plasmid containing the 5'NeoTc device was substituted for  $P_{GAL1}$  to introduce the binding site for GEV, resulting in an extension of the 5'-UTR from 43 nt to 75 nt upstream of the beginning of the 5'-NeoTc gate (compare Figure 4.18A). The positive control pWHE601 was modified accordingly (pWHE601- $P_{GAL1}$ ). In a first test, it was sought to measure the induction rate alongside the  $OD_{600}$  to determine cellular growth. Cellular growth was measured, because it was known from the original and further studies that GEV expression driven by a strong promoter like  $P_{TEF1}$  would reduce cell growth in presence of  $\beta E$  due to off-target gene activation or repression <sup>206,207</sup>. Hence, a titration experiment was conducted to check the influence of increasing concentrations of  $\beta$ -estradiol on the growth rate. In a parallel approach,  $P_{TEF1}$  was replaced by the weak  $P_{STE5}$  to verify that a decreased expression rate of GEV would leave cell growth undisturbed in presence of  $\beta E$ . Fluorescence and  $OD_{600}$  of yeast cells transformed with pWHE601-  $P_{GAL1}$  and either the GEV- $P_{STE5}$  or GEV- $P_{TEF1}$  plasmid in absence and presence of increasing concentrations of  $\beta E$  were recorded after 24 h. Figure 4.22 highlights the defective growth of yeast cells in presence of  $\beta E$  at concentrations greater than 10 nM, if GEV expression was driven by  $P_{TEF1}$ .  $P_{STE5}$ -driven expression of GEV had no impact on cellular growth at any of the tested concentrations of  $\beta E$ , but the maximum fluorescence values reached at each titrated concentration were generally shifted to higher concentration for the  $P_{STE5}$  system. The  $P_{TEF1}$  system is thus more sensitive to lower concentrations of  $\beta E$ . On the downside, owing to the comparably high fluorescence of uninduced cells, the  $P_{TEF1}$  system exhibits much lower induction efficiencies of 28-fold to 35-fold as compared to the over 200-fold in the  $P_{STE5}$ -based system at 100 and 500 nM  $\beta E$ . To circumvent any adverse effects on cell growth and accounting for its superior induction efficiency, the  $P_{STE5}$  system was selected to conduct further analyses.

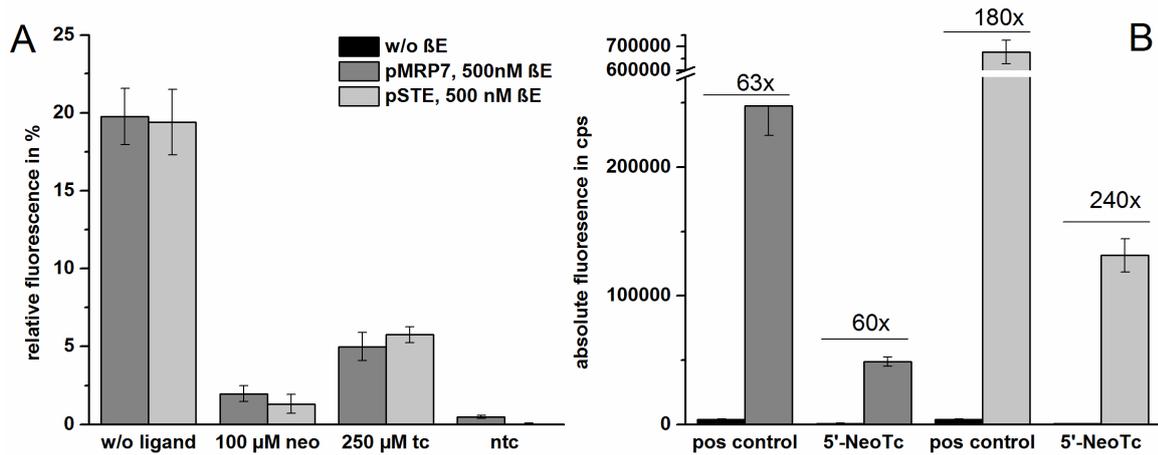
## Results & Discussion



**Figure 4.22 Induction test of the positive control.** Yeast cells transformed with plasmids expressing GEV by either *STE5* or *TEF1* promoter and pWHE601 were cultured in absence or presence of increasing concentrations of  $\beta$ E for 24 h and assayed in bulk for GFP fluorescence output and cell growth. Fluorescence output was normalized to cell growth, corrected for background fluorescence and is plotted in counts per second as computed by the software of the Fluorolog instrument. Optical density was determined at 600 nm using diluted samples.

In a subsequent GFP+ reporter gene assay the 5'-NeoTc gate was tested as part of the now inducible expression system. Alongside  $P_{STE5}$ -driven GEV expression another supposedly weak promoter,  $P_{MRP7}$ , was assayed for its aptitude to serve as an alternative promoter.  $P_{MRP7}$  had been used previously in a similar approach to circumvent the growth deficiencies imposed by  $P_{TEF1}$  and was found to exhibit very low leakiness in absence of  $\beta$ E, but to provide ample expression to reach high induction in presence of  $\beta$ E<sup>207</sup>. Figure 4.23A shows the relative fluorescence output of the 5'-NeoTc gate in absence and presence of the repressors neomycin and tetracycline as compared to GFP+ expression from pWHE601 (positive control), induced by 500 nM  $\beta$ E in concert with either  $P_{STE5}$ -GEV or pMRP7-GEV. Fluorescence data obtained from bulk measurements was recorded after 24h from outgrown cultures. Both expression systems operated equally with respect to basal gene expressions of 20% and switching factors exerted by the 5'-NeoTc gate. Additionally, expression levels in absence of the repressors were in the same range as provided by the constitutively expressed 5'-NeoTc gate. Repression by neomycin alone and both ligands was more pronounced for the inducible system with switching factors of ~12-fold and ~50-fold, respectively. The tetracycline induced repression is found at the same level with a ~3-fold off-switching in both, the constitutively and induced set up. Noteworthy, the expression of the 5'-NeoTc gate, if induced by  $P_{STE5}$ -GEV, was basically zero after addition of both repressors and thus, no reasonable switching factor could be calculated. Such differences in the calculated switching factors were a result of the underlying background levels. The higher the background that was subtracted, the greater its influence on low (repressed) expression levels. Here, the inducible system exhibited higher background fluorescence compared to the previous constitutively expressed system driven by  $P_{ADH1}$ . The background was constituted by the negative control pWHE601\* that features GFP+, but no stat codon and no 5'-riboswitch.

## Results & Discussion



**Figure 4.23 GFP reporter assay of the 5'-NeoTc gate and induction test of the 5'-NeoTc gate and pWHE601.** **A** Yeast cells transformed with plasmids expressing GEV by either *STE5* or *MRP7* promoter and the 5'-NeoTc gate were cultured in absence or presence of indicated repressor concentrations and induced by 500 nM  $\beta$ E for 24 h and assayed in bulk for GFP fluorescence output. Fluorescence output was normalized to cell growth, corrected for background fluorescence and is plotted relative to pWHE601. **B** the induction of pWHE601 and the 5'-NeoTc gate by variable GEV levels was tested. Induction factors were calculated from cultures used in **A**. **A** and **B** share the same legend. Experiments were performed in two independent triplicates.

Figure 4.23B emphasizes the influence of both promoters,  $P_{STE5}$  and  $P_{MRP7}$ , on the induction efficiency of the 5'-NeoTc gate and pWHE601. The absolute fluorescence output was normalized to cell growth and corrected for background fluorescence with and without  $\beta$ E to clearly discern the differences between the two promoters. The  $P_{STE5}$ -driven system provides 3- to 4-fold higher *GFP+* expression levels for pWHE601 (180-fold) and for the 5'-NeoTc gate (240-fold). Residual fluorescence measured for all uninduced cells containing pWHE601 is less than 1% relative to the induced state, whereas it is even further reduced for the 5'-NeoTc gate owing to generally decreased basal expressions in presence of a 5'-riboswitch. In this direct comparison to an already tested promoter, the  $P_{STE5}$  performed even better and confirmed the previous results.

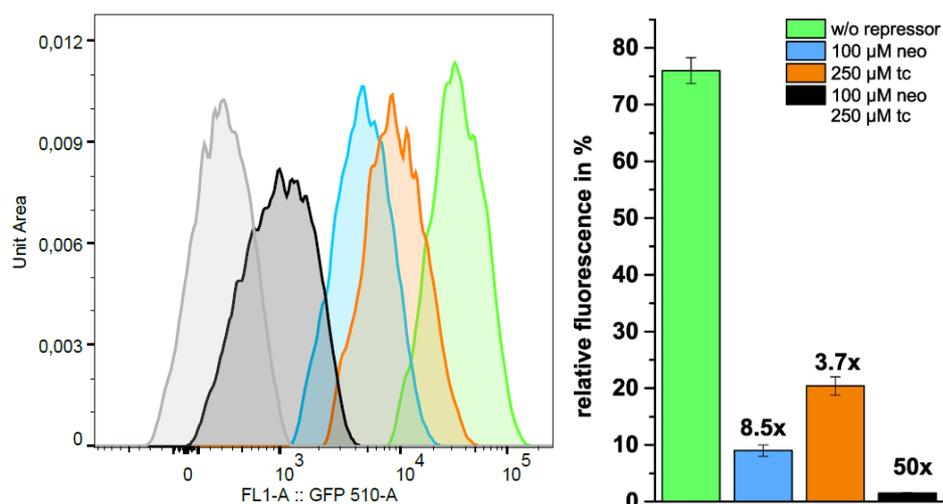
### 4.3.1.3 Genomic integrations

After the successful adoption of the GEV system, the 5'-NeoTc gate as well as positive and negative controls were integrated into the *ade2* locus situated on yeast chromosome XV. A genomic integration was thought to reduce phenotypic heterogeneity and by that to improve the switching performance with respect to a better separation of the individual (un)repressed populations. The CRISPRs/Cas9-based system CASEMBLR was employed to facilitate genomic integration<sup>45</sup>. This system allows for the marker-free integration of any insert DNA into a locus of choice and is highly effective. The only constraint is posed by the design of the small gRNA targeting the locus of choice, as it must meet the criteria necessary for target DNA recognition and processing by the Cas9 endonuclease. Most importantly the target site must be unique to the yeast genome to avoid off-target integrations or DNA damage leading to genome instability. In this study all integration sites were selected based on already tried and verified target loci<sup>45</sup>. The 5'-NeoTc gate was amplified from its plasmid backbone pWHE601\* as a single integration cassette of ~1700 nt flanked up- and downstream by 100 nt DNA stretches homologous to the *ade2* locus as pictured in Figure 6.12. Thus, yeast strain RS453 $\alpha$  was transformed with the linear dsDNA integration cassette and plasmids encoding Cas9 endonuclease and the gRNA

## Results & Discussion

targeting the *Ade2* locus. Correct integrations were identified by yeast colony PCR or a PCR on previously isolated genomic DNA from various colonies. Once the phenotype had been confirmed in a preliminary reporter gene assay, integrations were verified by sequencing of the integration cassette and the up- and downstream flanking regions.

The single-cell analysis was initially performed by flow cytometry. Yeast cells containing a chromosomally integrated 5'-NeoTc gate were transformed with the GEV-*P<sub>STE5</sub>* plasmid and subsequently cultured for 24 h in absence and presence of 500 nM  $\beta$ E and the repressors to perform a steady state measurement from an outgrown culture providing first insights into the phenotypic distribution. Figure 4.24 shows the GFP fluorescence of gated (see chapter 5.1.2.2.4 for gating protocol) populations in an overlay histogram. Plots were normalized to unit area displaying event counts for each fluorescence value (channel) relative to all events within the gate, allowing for a direct comparison of all histograms. Additionally, fluorescence channels were plotted biexponentially, revealing events beyond zero and hence avoiding a cut-off at zero. Besides, low fluorescent values are highly compressed on the "linear" log-scale, facilitating a compact presentation without long leads or tails for populations between 0 and 1000.



**Figure 4.24 Single-cell analysis of the inducible 5'-NeoTc gate.** The histogram shows the normalized abundance of each GFP fluorescence level for the unrepressed and repressed populations of the 5'-NeoTc gate analyzed by flow cytometry. 10,000 events were recorded per population that was sampled from cultures grown in 24-well plates for 24 h in absence and presence of indicated repressor concentrations and 500nM  $\beta$ E. The x-axis is plotted biexponentially. The presented data is an average replicate of at least three independent replicates. Extracted mean fluorescence values from the histograms are shown in the column plot. The switching factors are given in x-fold for each repressor. Fluorescence was calculated relative to the positive control pWHE601 that was also analyzed by flow cytometry.

Uninduced cells and induced cells expressing the 5'-NeoTc gate are shaded light grey and lime green, respectively, and mark the range of 120-fold induction of *GFP+* expression that may be reached with this inducible expression system. Notably, populations of fully induced (lime green) and fully repressed (black) cells are completely baseline-separated, demonstrating an efficient processing of the NOR operation at the single-cell level. The neomycin repressed population (light blue) overlaps slightly with the fully induced, whereas, reflecting its poor switching factor, the tetracycline repressed population (orange) superimposes the fully induced population to a larger extent. Thus, switching factors and phenotypic distributions are major parameters contributing to the correct and efficient functionality of the

## Results & Discussion

5'-NeoTc gate at the single-cell level. Expression of GEV was monitored by mCherry fluorescence and found to exhibit the same level and phenotypic distribution among all investigated conditions and populations.

Figure 4.24 shows also the extracted mean fluorescence values plotted relative to the chromosomally integrated positive control pHWE601. With 76%, basal expression far exceeded that displayed by the plasmid-borne system. Surprisingly, switching factors were comparable across substantially unlike expression systems (plasmid-borne and chromosomal expression of the NOR gate) and experimental set ups (fluorescence spectroscopy from bulk and flow cytometry), mainly owing to different background levels and basal expressions. A bulk measurement of the integrated 5'-NeoTc gate confirmed that the major difference observed with respect to basal expression and switching factors was due to the reduction of gene copy number by genomic integration, since the bulk measurement recorded on a Fluorolog instrument mirrored the results obtained by flow cytometry.

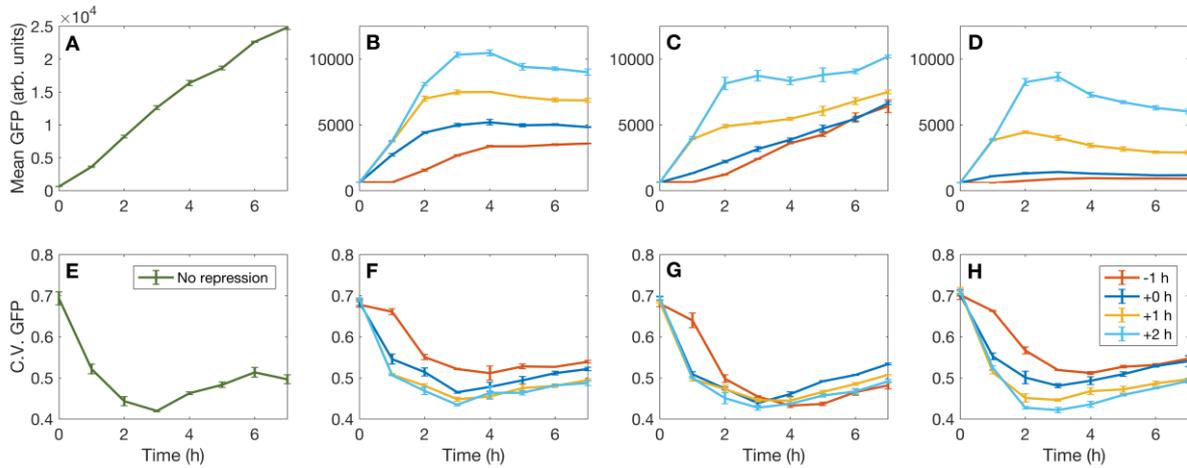
Now, that the performance of the inducible 5'-NeoTc gate had been improved at the single-cell level by a reduction of its gene copy number, the transient dynamics of the fluorescence phenotypes after induction and repression were sought to be assessed. As a start, exponentially growing yeast cells bearing the integrated 5'-NeoTc gate and the GEV- $P_{STE5}$  plasmid were repressed at different time points (-1h, 0h, +1h, +2h) relative to the addition of 500 nM  $\beta E$  to measure whether the timing of repression had any influence on gate performance. Samples were drawn during a time course of 7 h after induction and analyzed by flow cytometry. The analysis revealed that the transient induction driven by  $P_{STE5}$  exhibited only a 3-fold increase of fluorescence after 2 h that was far too low to clearly discern induced from uninduced populations. The plasmid-borne expression of the 5'-NeoTc gate yielded a 10-fold induction of  $GFP+$  expression after 2h, suggesting that higher copy numbers of  $GFP+$  could partially compensate very low GEV availability. To overcome this problem, the  $P_{TEF1}$ -driven expression of GEV was considered and evaluated again. It was hypothesized that with such a strong promoter and a GEV activation by 10 nM  $\beta E$  being sufficient to reach maximum expression levels and avoid impaired cell growth, the transient characterization of the integrated 5'-NeoTc gate would be feasible, because very high levels of GEV would overcompensate the occurrence of only a single  $GFP+$  copy due to a permanent occupancy of the  $P_{GAL1}$ .

### 4.3.1.4 Transient characterization of the 5'-NeoTc gate

Yeast cells bearing the integrated 5'-NeoTc gate were transformed with the GEV- $P_{TEF1}$  plasmid and processed as described for the abandoned GEV- $P_{STE5}$  approach. Figure 4.25A-D show the mean fluorescence of samples drawn over 7 h of continuous culturing induced by 10 nM  $\beta E$ , supplemented with neomycin and/or tetracycline and measured by flow cytometry. Onset of GFP expression was fast enough to clearly discern induced cells from uninduced cells after 1 h by an 8-fold increase of the mean fluorescence value (Figure 4.25A, 4A). By 2 h, the induced populations shifted to a near complete baseline separation from background fluorescence (13-fold induction) and showed an almost linear increase of fluorescence over the time course of the experiment (Figure 4.25A). The device featured a high repression efficiency for the dual-input, as indicated by the -1 h and 0 h experiments, where only a 2-fold increase of the mean fluorescence over background level was observed for 7 h. Importantly, onset kinetics of the repression cascade triggered by ligand addition showed an immediate effect on gene

## Results & Discussion

expression within the first hour after ligand addition, irrespective of the time point after induction (Figure 4.25D). Thus, the device processed a NOR operation within a short time scale, enabling low-delay repressive operations for genetic circuit design.



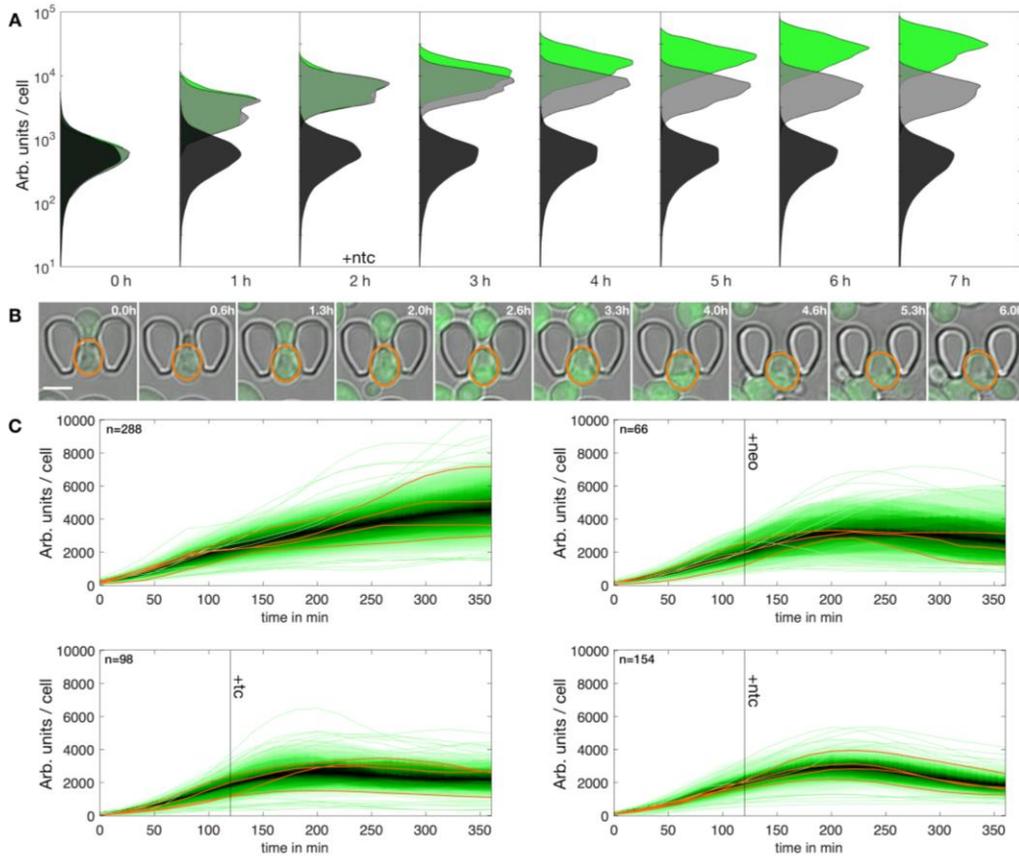
**Figure 4.25 Gate characterization on the single cell level.** A-D Fluorescence means over time for unrepresed cells **A** and cell repressed with 100 M neomycin **B**, 250 M tetracycline **C** and both ligands **D**. E-H show the corresponding coefficients of variation (CV) summarizing the width of the cell population. Trajectories are color-coded to show different times of repression; 1 h of repression before induction (red), simultaneous repression and induction (dark blue), repression after 1 h of induction (yellow) and repression after 2 h (light blue). 10 nM  $\beta$ -estradiol were used for all experiments. Samples were drawn every hour for 7 h from cultures grown in 12-well plates. Data was averaged over three technical replicates as indicated by the error bars and is representative for three independent replicates. Fluorescence background was not subtracted.

The single-input data suggests different repression kinetics for the neomycin- and tetracycline-inputs, respectively (Figures 4.25B, C). The device exhibited a progressive decrease of repression efficiency for the tetracycline-input, reflecting a fast onset followed by a gradual reduction of repression with advancing induction (Figure 4.25C). The differences in the onset of repression for both ligands are best seen by comparing the slopes of the graph between 1 h and 2 h for the +1 h experiments and between 2 h and 3 h for the +2 h experiments. For the single-input these slopes are parallel with respect to the different time points of repression. This is also observed for the dual-input experiments and hints towards a timing-independent application of the device as stated above (Figures 4.25B-D). The comparison of these slopes for the two single-inputs reveals the 4-fold faster onset that was achieved with tetracycline. Interestingly, the onsets (slopes) of the +1 h and +2 h experiments were very similar for the tetracycline-input (750 arbitrary units/h (arb. units/h)) and the dual-input (800 arb. units/h). This points towards a predominant role of the tetracycline-sensitive riboswitch in the fast repression kinetics of the NOR gate that were complemented by the more efficient repression by neomycin in the long run (see Figure 12 for comparison). Figures 4.25E to H show the coefficients of variation (CV) for all experimental setups to account for cell-to-cell variability. Although the CVs changed overall less than 2-fold, the CVs decrease upon induction indicates a stabilizing effect on phenotypic heterogeneity. This effect correlates mainly with the disappearance of all background fluorescence that could be traced back to autofluorescence of the cells and  $\beta$ -estradiol independent GEV activity broadening the phenotypic distribution of uninduced cells, but vanishing upon induction among all other induced cells, ultimately reducing the CV.

In Figure 4.26A the population dynamics exemplified by the dual-input repression for the +2 h experiment are illustrated. In accordance to the CV analysis, cell-to-cell variability was reduced by

## Results & Discussion

induction, turning cells towards an ON-state at varying time scales reflected in a smooth population shift. Repression at +2 h caused the population shift to halt as further translation was blocked efficiently by the 5'-NeoTc gate. Although flow cytometry is a powerful technique to analyze a bulk culture at the single-cell level, it does not allow a single cell to be tracked over time to determine the characteristics of a transient response on the single-cell level. To link the snapshot data of a population recorded by flow cytometry to its single-cell transient dynamics the 5'-NeoTc gate was analyzed in a +2 h experiment on a microfluidic chip by time-lapse fluorescence microscopy by Jascha Diemer (Prof. Koepl) (Figure 4.26B).



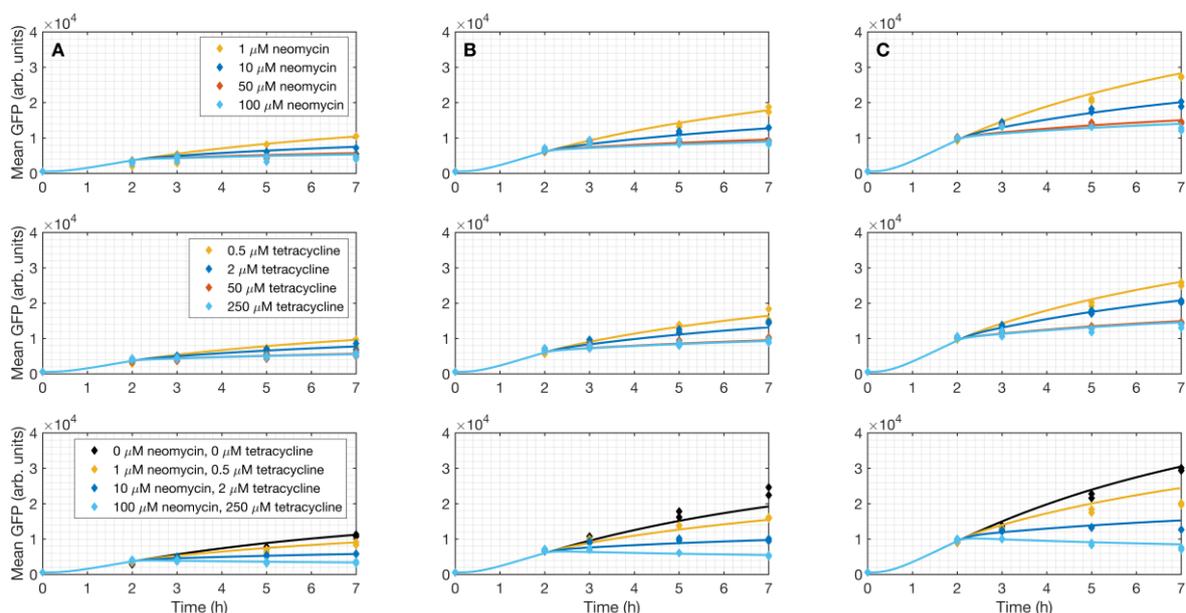
**Figure 4.26 Qualitative population dynamics of the 5'-NeoTc gate.** **A** Histogram of time-lapse cytometry data from a +2 h experiment. Uninduced cells (black) show a broader distribution than cells induced at 0 h (green) and cells induced at 0 h and repressed at +2 h (gray). **B** The picture series shows a representative cell trapped on the microfluidic chip and recorded over time by fluorescence microscopy. Scale bar 5  $\mu\text{m}$ . **C** Density plots of segmented single-cell traces. Random examples for individual traces are highlighted in orange. Induction with 10 nM  $\beta$ -estradiol, repression with 100  $\mu\text{M}$  neomycin and/or 250  $\mu\text{M}$  tetracycline. Color becomes darker the more cell traces approaching the median. For cytometry samples were drawn every hour for 7 h from cultures grown in 12-well plates. Plotted data are representative for at least three independent replicates. From the same inoculum cells were taken to conduct microfluidic measurements and analyzed data is representative for at least three independent replicates. Fluorescence background was not subtracted in both experimental setups.

Single cells harvested from an actively growing culture were trapped on the chip, cultured in presence of 10 nM  $\beta\text{E}$  and the repressors (100  $\mu\text{M}$  neomycin and/or 250  $\mu\text{M}$  tetracycline) and their emitted fluorescence was recorded every 0.2 h for 6 h. In addition to the density profile, single-cell traces were randomly picked to demonstrate their coherently developing phenotypes over time (Fig. 4.26C). That is, a highly expressing cell emits a stronger fluorescence signal after several hours of repression than a

## Results & Discussion

lowly expressing cell, thus their dynamic fluorescent phenotypes develop coherently. In total, the live-cell imaging data provided further backing for the considered time-lapse flow cytometry dataset.

Having identified that the repression kinetics and efficiencies are independent of the preceding induction period, gate performance at different induction levels was assessed and its dose dependence on different repressor concentrations determined (Figure 4.27). Exponentially growing yeast cultures were sampled at indicated time points (dots), whereas an *in silico* model calibrated on these data by Leo Bronstein (Prof. Koepl) reproduced the full time-resolved response of the 5'-NeoTc device (lines). Induction with progressively increasing levels of  $\beta$ -estradiol led to a proportional increase of the GFP signal. All concentrations of neomycin and tetracycline were titrated in preliminary experiments to determine the optimal concentrations for model calibration. The two highest concentrations of neomycin (50  $\mu$ M and 100  $\mu$ M) and tetracycline (50  $\mu$ M and 250  $\mu$ M) were used to determine the saturation level of full repression.

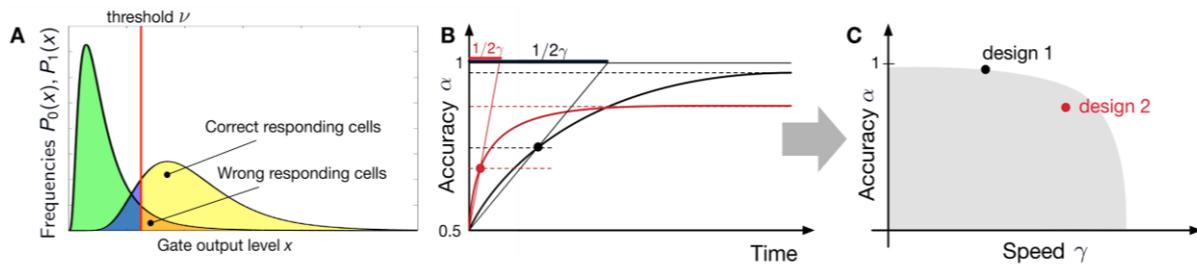


**Figure 4.27 Matrix of dose-response profiles at varying induction levels.** NeoTc gate induced with **A** 2.5 nM  $\beta$ -estradiol, **B** 5 nM  $\beta$ -estradiol and **C** 10 nM  $\beta$ -estradiol and repressed with indicated ligand concentrations after 2 h induction. Fluorescence output was recorded by flow cytometry. Mean fluorescence values (dots) are plotted over the time course of the experiment from samples drawn at the indicated time points. The calibrated computational model reproduces the full time-resolved response of the device (lines) for all gate induction levels and for all combinations of repressor concentrations. Cultures were grown in 12-well plates. Plotted data is representative for at least five independent replicates. Fluorescence background was not subtracted.

The gate featured an equal repression onset and efficiency for all three induction levels indicated by the slopes within the hour of repression and the expression levels after 7 h relative to the time point of repressor addition. Consistent with the precedent results, these experiment series highlighted the high repression efficiency after addition of both ligands. For 100  $\mu$ M neomycin and 250  $\mu$ M tetracycline translation was fully blocked and the observed signal decay is congruent with the half-life of the reporter GFP. By supplementing different repressor concentrations the full dose-response profile of the gate could be mapped out and its performance was computed with respect to the three inductor concentrations as presented in the following sections.

#### 4.3.1.5 Performance characterization

As of now, data analysis was rather focused on a qualitative description of the performance of the 5'-NeoTc gate at the single-cell level. Cell-to-cell variability is directly observed in histograms as the width of the population on the x-axis plotted as fluorescence output. Overlaps of populations with different Boolean states upon repressor addition indicate a malfunction of the logic gate at the single-cell level, since for correct functionality and high performance not only a separation of the mean, as in bulk measurements typically expressed by the switching factor, but rather a separation of the complete population is sought (Figure 4.28). Hence, driven by the need to quantitatively characterize the performance of the 5'-NeoTc gate, a new measure, the ROC curve, describing receiver-operator-characteristics and accounting for cell-to-cell variability was introduced (Figure 4.29A, B). The principal idea is summarized in Figure 4.28. Two overlapping populations in two different Boolean states are divided by an arbitrarily chosen threshold determining false and true positive as well as false and true negative cells (Figure 4.28A). For instance, cells of the repressed population (lime green) that are beyond the threshold to the right (orange) are false positive cells, because they show an ON state instead of the supposed OFF state. These subpopulations decrease the performance of a logic gate at the single-cell level.



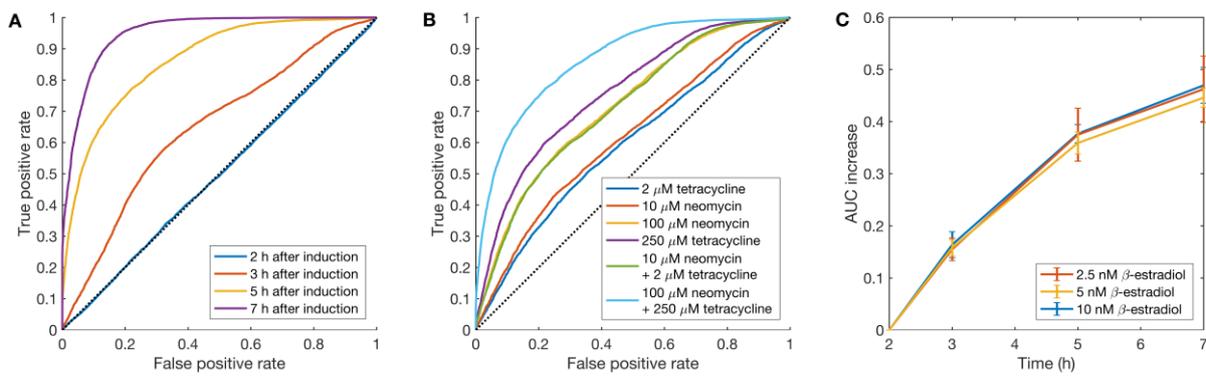
**Figure 4.28** Illustrating performance characterization using the ROC curve. **A** Empirical distributions for the ON state (right) and the OFF state (left) of a gate across a cell population. Generally, both distributions are divided into two parts by a choice of decision threshold. Corresponding to each, there are cells in the OFF state producing output levels above (false positive rate, orange) and cells in the ON state correctly producing output levels above that threshold (true positive rate, yellow). **B** The curves represent the continuous evolution of the areas under the ROC curve (AUCs) over time. The accuracy of the gate is thus computed as the achieved asymptotic AUC, whereas the speed of the gate is determined by the time it takes until its AUC reaches half-maximum. **C** Different gate designs can be located in the two-dimensional performance space, where trade-offs for a given gate architecture define fundamental performance constraints that cannot be overcome for that architecture (gray-shaded area).

The lower the threshold in Figure 4.28A is set the more true positives are derived, but on cost of the appearance of more false positives. Plainly spoken, if the red line (threshold) shifts left more yellow cells appear to be true positives, but also the orange area (false positives) increases. The ROC curve displays the evolution of the true positive rate over decreasing threshold levels, i.e. increasing rates of false positives (Figure 4.29A, B). The ideal classifier, or here the 5'-NeoTc gate, would produce a ROC curve that runs in parallel with the ordinate at a rate of zero of the false positives and keeps a value of 1 for the true positive rates regardless of increasing false positive rates. A random classifier in turn, always shows a 1:1 output, i.e. if the false negative rate increases by 0.1 the true positive rate increases also by 0.1. To gain a measure of the evolution of the true positive rate at any threshold value, the area under the ROC curve (AUC) is computed. An ideal classifier exhibits an AUC of 1 and all other AUCs are calculated relative to the ideal classifier (Figure 4.29C). Thus the AUC accounts for the total overlap of

## Results & Discussion

two populations, in which one is in the OFF state and one in the ON state. From a log-normal distribution, as can be approximated for the populations measured by flow cytometry here, the AUC can simply be regarded as the difference of the means divided by the square root of the sum of the variances of two populations. In that way the AUC accounts for the separation of the mean values, the actual switching factors, and more importantly for the width of each population, i.e. its cell-to-cell variability.

Figures 4.29A, B show the resulting performance of the gate for specific repressor concentrations and for a combination of the repressors for different induction levels over the course of a 7 h experiment. Correspondingly, Figure 4.29C depicts the time course of AUC during the course of induction. The ROC and AUC characterization was solely done on the experimental data. Figure 4.29C indicates that gate performance within the observed time interval is independent of the gate induction levels that were considered, suggesting that the AUC characterizes the gate as such and is not affected by its expression level.

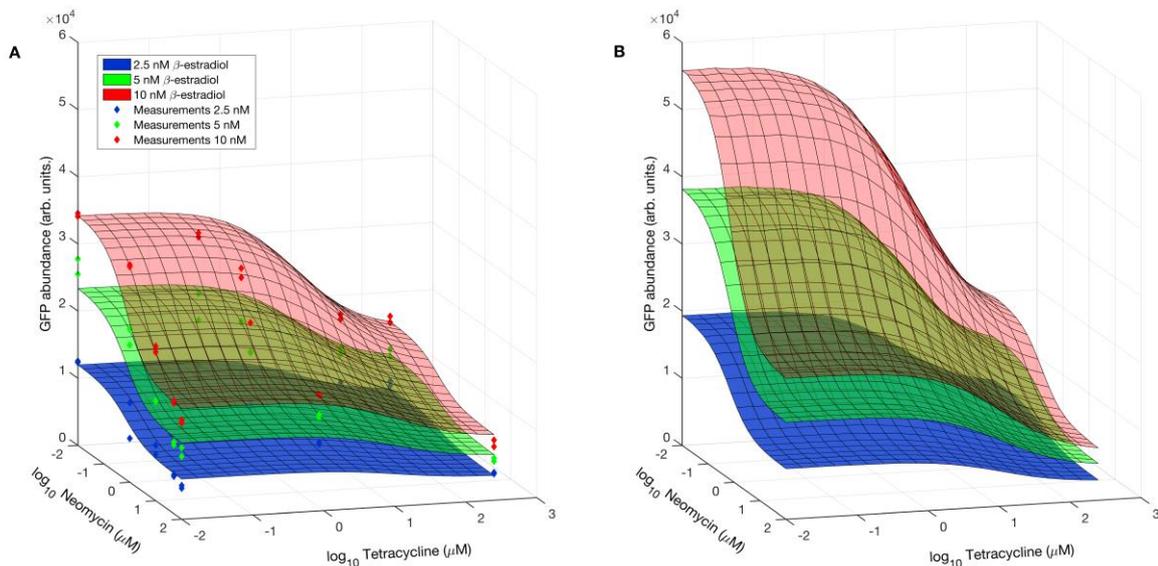


**Figure 4.29 ROC analysis of flow cytometry data.** **A** ROC curve evolution over time for full induction (10 nM  $\beta$ -estradiol) and maximum repression (100  $\mu$ M neomycin and 250  $\mu$ M tetracycline) initiated after 2 h of induction. **B** ROC curves at different concentrations of repressors (added after 2 h), induced with 10 nM  $\beta$ -estradiol and measured after 5 h. In **A** and **B**, the dotted diagonal line shows the performance of a hypothetical, completely random (i.e., dysfunctional) gate. **C** Increase of AUC relative to its value at the time of repressor addition (100  $\mu$ M neomycin and 250  $\mu$ M tetracycline, 2 h after induction) over time for different inducer concentrations.

The slope of the AUC accordingly provides a measure of how fast the gate becomes functional; in particular the slope or speed is defined as the inverse time it takes to reach half-maximum performance. Based on this observation two orthogonal features of the gate that describe (i) the accuracy of the gate and (ii) the speed or responsiveness of the gate were extracted. Every gate design can then be mapped into a two-dimensional performance space as indicated in Figure 4.28B, C. For a given gate design only a certain region of the performance space will be accessible through parameter variations (e.g. rate constants) and some parameterization thereof will lead to an optimization of that gate design.

#### 4.3.1.6 *In silico* completion of device characterization

Although the short and transient experimental characterization of the gate did not provide a full account of the gate's behavior, it was directly geared toward providing minimal, yet informative data for translation of the system to the *in silico* domain. Having a computational model that accurately predicts the repression dynamics and its cell-to-cell variability, the model was now used to complete the device characterization *in silico*. In particular, it was employed to map out the dose-response curve of the repressor and the dose-response surface of the NOR gate at arbitrary time points and especially at steady state phase. In such a way, the device's output over the time course of simulated prolonged exponential growth was emulated; data that were beyond the experimentally feasible time window. Figure 10 shows the dose-response of the NOR gate at 7 h (Fig. 4.30A) and at steady state (Fig. 4.30B) for different induction levels of the gate. The experimental data at 7 h aligned remarkably well with model predictions. As it was previously observed, different induction levels do not influence the performance of the gate and thus computed dose-response surfaces superimpose each other, only distanced by the respective induction strength. The model predicted a residual fluorescence output for maximum repressed cells at the higher induction levels of 5 nM (3-fold) and 10 nM  $\beta$ E (6-fold) relative to the starting level of uninduced cells at  $\sim$ 500 arb. units, suggesting the continuous translation of a small number of mRNAs escaping any repressor interaction. Additionally, total fluorescence induced by 10 nM  $\beta$ E was expected to increase by 1.5 - 2-fold, comparable to the induction levels reached with the  $P_{STE5}$  system after 24 h.

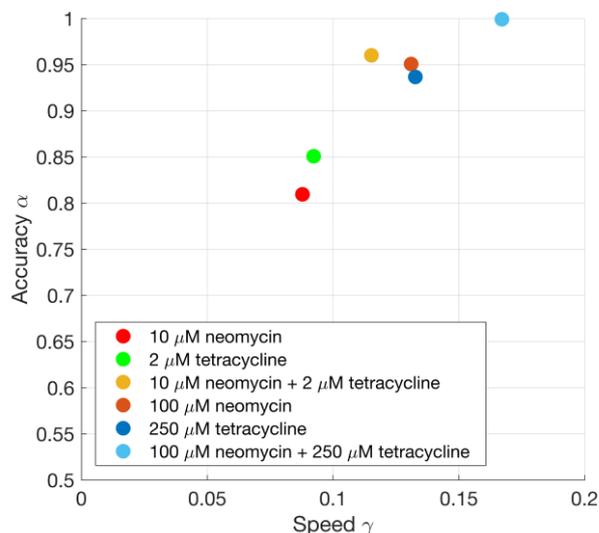


**Figure 4.30 Dose-response surfaces at various induction levels** of the gate, showing the dependence of gate output on the concentrations of repressors. The system was measured until 7 h after induction. The model, calibrated from the measured data, interpolates the missing parts of the dose-response surface. **A** Measured data at 7 h (colored diamonds) and interpolated dose-response surface. **B** the unobserved dose-response after 24 h is predicted using the model since this time point was beyond the possible sampling interval that requires prolonged exponential growth.

From the steady state predictions different NOT gate configurations (single input) were located in the proposed two-dimensional performance space of accuracy and speed. Figure 4.31 shows the coordinates of the NOT/NOR gate at various repressor concentrations in the two dimensional

## Results & Discussion

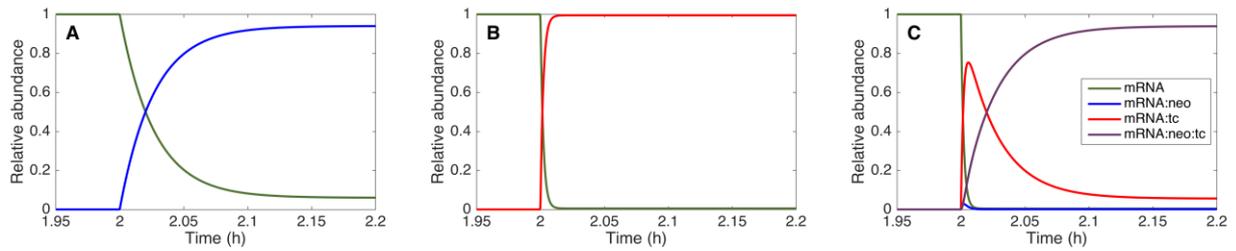
performance space. It is evident that addition of higher repressor concentrations lead to a faster response of the gate and a better separation of the two Boolean states with respect to cell-to-cell variations and the fold-change between them. The performance space mapped here can then be used to characterize the gate design as shown in Fig. 4.28C, by simply setting a core coordinate for the gate design itself and defining an area around it determined by the tested repressor concentrations to visualize the performance constraints that cannot be overcome by the design of the 5'-NeoTc gate.



**Figure 4.31 NOT gate performance** characterized through position in the performance space for various combinations of repressor concentrations. Inducer concentration was 10 nM  $\beta$ -estradiol. Repressors were added at stationarity of the system.

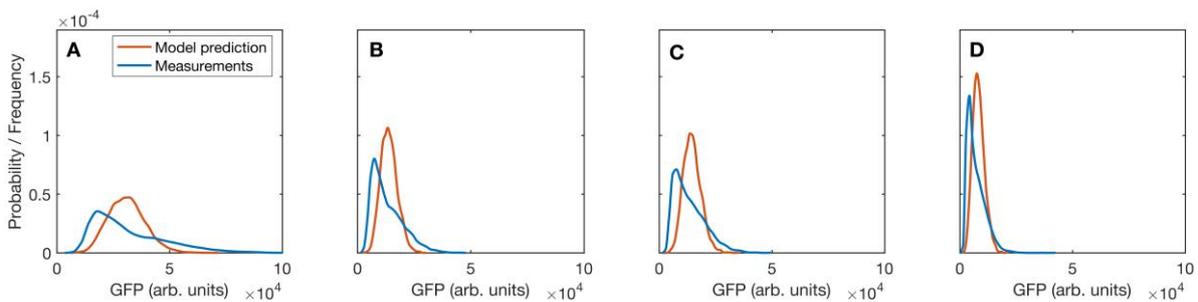
The reporter measurements presented in Figure 4.25 and 4.27 imply differences in the repression kinetics of Tc and Neo and a closer look at the concurrent events for the case of a double-repression would be desirable. However, due to the stability and hence low-pass filtering of the reporter, quantifications based solely on the experimental data were unfeasible. The model-based reconstruction provided mRNA estimates at their comparably faster time-scales, allowing a better dissection of the temporal sequence of events at the onset of repression after 2 h induction. By this only newly transcribed mRNAs in the presence of repressors are captured and used to calibrate the model. Figure 4.32 shows the evolution of occupancies for the different mRNA subpopulations (i.e. free, single-bound, double-bound) at the onset of repression. In accordance to protein levels (Figures 4.25 and 4.27), free mRNA abundances decrease faster in presence of tetracycline as compared to neomycin (Figure 4.32A, B). Increase of double-bound mRNAs is slowed by transient occupancy of the mRNA with tetracycline alone as neomycin exhibits relatively slow association kinetics (Figure 4.32C). For the same reason, no single-bound mRNAs by neomycin are found in the presence of tetracycline. The overall abundance of mRNAs encoding the NeoTc gate may be assumed to be unaffected by the integration of the two riboswitches near the start codon as has been shown in earlier studies<sup>137,139</sup>.

## Results & Discussion



**Figure 4.32** Relative abundance of mRNA subpopulations after addition of repressors at +2 h. **A** 100 nM neomycin. **B** 250 nM tetracycline. **C** 100 nM neomycin and 250 nM tetracycline. Due to the faster binding kinetics of tetracycline, the majority of mRNA molecules binds tetracycline immediately after addition. Subsequently, the balance progressively shifts to double-bound mRNA due to the slower binding kinetics of neomycin.

The model also allowed the investigation of the noise in GFP expression which is left after removing the effect of cell size. Hence, an *in silico* homogenization was performed. The resulting distributions are shown in Figure 4.33 for several combinations of inducer and repressor concentrations. It indicates that a considerable variability due to stochastic gene expression remains after such homogenization that is inherent to the system and cannot be decreased by a further reduction of gene copy number for instance.

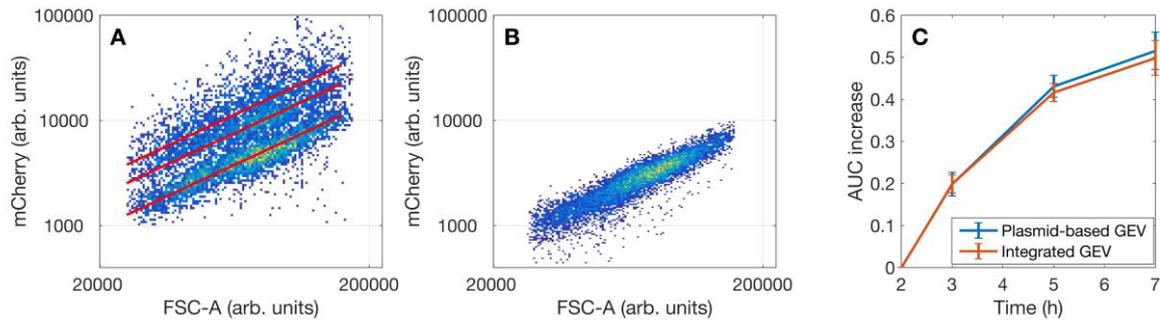


**Figure 4.33** Predicted GFP abundance for fixed cell size after 7 h of induction. **A** No repressors. **B** 100  $\mu$ M neomycin. **C** 250  $\mu$ M tetracycline. **D** 100  $\mu$ M neomycin + 250  $\mu$ M tetracycline. Repressors were added after 2 h induction with 10 nM  $\beta$ -estradiol.

### 4.3.1.7 Computer-aided redesign

Having general performance measures in place, it was sought to assess whether performance could be further improved by a redesign of the inducible expression system, for instance by changing specific reaction rate constants that are amenable to experimental modification. As the width of the distribution and in particular the distributional overlap between the Boolean states are critical to achieve high AUC it was specifically investigated, whether reduction of extrinsic noise resulted in improved gate performance. Although ARS/CEN plasmids were used for the expression of GEV that should be ideally maintained as a single copy within the cell, recorded data and previous publications suggest that cells do not always carry just a single plasmid, but multiple<sup>52</sup>. In particular, Figure 4.34A provides evidence for a considerable population of cells carrying two and three plasmids. This plasmid heterogeneity leads to a broadened distribution of the GEV transcription factor.

## Results & Discussion



**Figure 4.34 Influence of GEV variability on gate performance.** GEV vs. forward scatter for plasmid-based GEV **A** and for chromosomally integrated GEV **B**. For the plasmid-based case, at least two subpopulations are clearly discernable. The distance on the mCherry-axis between the red lines corresponds to doubling or tripling of GEV abundance, which is the simplest assumption for the behavior when 2 or 3 plasmid copies are present in a cell. For chromosomally integrated GEV, only the subpopulation corresponding to a single copy of the plasmid remains. **C** AUC increase over time for plasmid-based and for integrated GEV for induction at 2.5 nM  $\beta$ -estradiol and repression with 100  $\mu$ M neomycin and 250  $\mu$ M tetracycline. The differences between these two are within the margin of error. Data for plasmid-borne GEV expression are derived from cultures processed in Fig. 4.27. Data for integrated GEV expression were obtained equally.

Generally, through noise propagation a broadened GEV distribution will give rise to a broadened distribution of the gate output, and hence, deteriorating the gate's performance. Using the model, the sensitivity of the gate's AUC with respect to changes in GEV heterogeneity was queried. Practically, the AUC was computed for different standard deviations of GEV heterogeneity in the model and it was found that variability in GEV abundance had little effect on the AUC and hence did not contribute much to the output variability. Thus, the model suggested that redesigns targeting a reduction of GEV variability through, for instance, a chromosomal integration of GEV, would not yield a performance increase. To test this prediction, the plasmid-derived GEV expression cassette was inserted into the *His3* locus using CASEMBLER and the above single-cell device characterization was repeated. Figure 4.34B shows the reduction in GEV heterogeneity after integration. In terms of performance the device behaved the same under the same conditions (Figure 4.34C). Thus, the integration of GEV and thereby the reduction of its noisy expression had no influence on the AUC increase of the gate. On top of the AUC analysis the model also provided an explanation for this counterintuitive insensitivity of gate performance with respect to GEV variability. A plain saturating effect through GEV at the promoter would be incompatible with the observed gradual dose dependency of gate output with respect to different induction levels (e.g. Figure 4.30) and the model indeed determined that the insensitivity arises because  $\beta$ -estradiol is the limiting factor, while GEV is in excess for a very broad range of values, exceeding the threshold needed for maximum GFP+ expression even with a single gene copy.

### 4.3.2 Discussion

The natural design goal for logic-based devices is to require correct functionality of the gate at the single-cell level and not just at the bulk level. The x-fold-change at the bulk level is a poor descriptor of the gate performance, because cell-to-cell variability in the device's behavior is not considered, although it plays a key role for efficient operation at the single-cell level. Hence, apart from the fold-change between the Boolean states, the overlap between the single-cell distributions associated with these two states is crucial. Importantly, time course measurements of device behavior during induction and repression provide insight into the kinetics of the device, such as the time interval it requires until it equilibrates to a new expression profile. Characterizing the device kinetics on the single-cell level moreover, allows to capture the heterogeneity in the onset of a response to the inducer and repressor directly visible in live cell imaging and at least discernable in flow cytometry data through transient fluctuations of the measured distributions. Resolving the transient dynamics of a device is also important for building and calibrating computational models for it as the characterization at steady state provides only insufficient data. Considering the requirements for logic gates and the ideal assessment of their performance, the ROC analysis of transient single-cell data was introduced as a new measure for gate performance accounting for both, the x-fold change and the distributional overlap of (un)repressed populations bearing an RNA-based logic gate. Initially, four different logic gates constructed from the tetracycline- and neomycin-binding riboswitches had been constructed, but only one, the 5'-NeoTc gate, was further analyzed and yielded the data for ROC analysis and model calibration. Among two exclusive NOT gates, the 5'-NeoTc gate was the only one of two NOR gates that exhibited the phenotypic behavior expected from a NOR operation. The output of its inverse counterpart, the 5'-TcNeo gate, suggested extensive malfunction possibly due to yet unresolved structural interferences. Further analysis of the transient performance of the 5'-NeoTc gate at the single-cell level by flow cytometry and time-lapse fluorescence microscopy using a microfluidic chip unraveled its fast and efficient repression kinetics. The absolute output processed by the gate could be fine-tuned with different inducer concentrations, whereas signal processing itself exhibited high robustness and hence the performance of the gate was found to be independent of the induction level.

Besides the designed NOR gate composed of riboswitches, recent work on the construction of RNA-based logic gates in bacteria has shown the versatility and aptitude of RNA-based regulation for the development of small genetic circuits<sup>92,144</sup>. The RNA toehold switches generated by Collins and co-workers feature high orthogonality in *E.coli* and have been yielded by a combination of *in silico* computation and rational design. Felletti *et al.* have engineered the twister ribozyme in *E.coli* to sense different ligands mainly by permutation experiments guided by rational design principles from previous works<sup>131–133,149,208</sup>. The performance of any logic gate presented in these two studies appeared to be generally lower compared to the single-input counterparts, underlining the need for connectable RNA parts without compromised functionality, i.e. performing better than the single parts of which they are composed. In an attempt to rank the performance of their logic gates, Felletti *et al.* employed a metric derived from probability binning developed to analyze single-cell data that was recorded by flow cytometry<sup>209</sup>. The method uses a statistic parameter  $T(X)$  that provides a measure of the probability by which two populations, e.g. the unrepressed and a repressed, are different and is also capable of ranking different populations with respect to their difference. An empirical threshold has to be determined beyond

## Results & Discussion

which populations are regarded as different. Although being rather similar to the ROC analysis, ROC analysis is independent of a threshold to summarize the performance of a given logic gate. Moreover, if transient fold-changes are recorded, the ROC analysis can be used to compute the speed with which the gate reaches its maximum accuracy. A particular gate design can subsequently be mapped into a whole design space determined by accuracy and speed allowing direct comparisons to other designs and their performance to elucidate the best possible design for a given operation (Fig. 6B, C). These analyses are beyond what is feasible with the approach of probability binning. Collins and co-workers did not consider phenotypic noise at all to calculate switching factors and did so solely by extraction of the mean fluorescence from the corresponding histograms. Besides, the discussed studies used steady state measurements that only allowed to assess the equilibrium of all involved rate constants, and hence all transient information needed to dissect transitions from one Boolean state into another was lost.

In the present study the model built and calibrated on the transient single-cell data was not only employed to perform ROC analysis, but also to determine the gate's full dose response profile at steady state. In such a way, the device's output over the time course of simulated prolonged exponential growth was emulated; data that were beyond the experimentally feasible time window as these growth phase may only be maintained up to about 8 h in synthetic complement medium supplemented with all necessary nutrients. To solve this issue of limited exponential growth an alternative approach, the EnPresso® growth systems, had been developed by BioSilta. Generally applied to extend exponential growth for higher protein yields of a continuous fed-batch culture of the yeast *Pichia pastoris*, the so called EnBase® technology could also be transferred to reach prolonged exponential growth for a reporter gene assay. The main idea is a controlled release of glucose to the medium by an enzymatic degradation (called the glucose-releasing agent, GRA) of long polysaccharide chains. The growth rate may be regulated by the concentration of the GRA. *S. cerevisiae* has been cultivated without problems, but growth protocols must be developed by the user (personal communication with Antje Neubauer, BioSilta). In particular, the GRA must be titrated to adapt its concentration to the timing of the overall experiment. Polysaccharide (administered as tablet) and GRA may be supplemented to the medium at any time. Additionally, cultures must be well aerated and the complete experimental set up would have to be changed from 3ml culture in 12-well plates to 2ml culture in 24 deep-well plates to allow for vigorous, high-amplitude shaking in a different incubator.

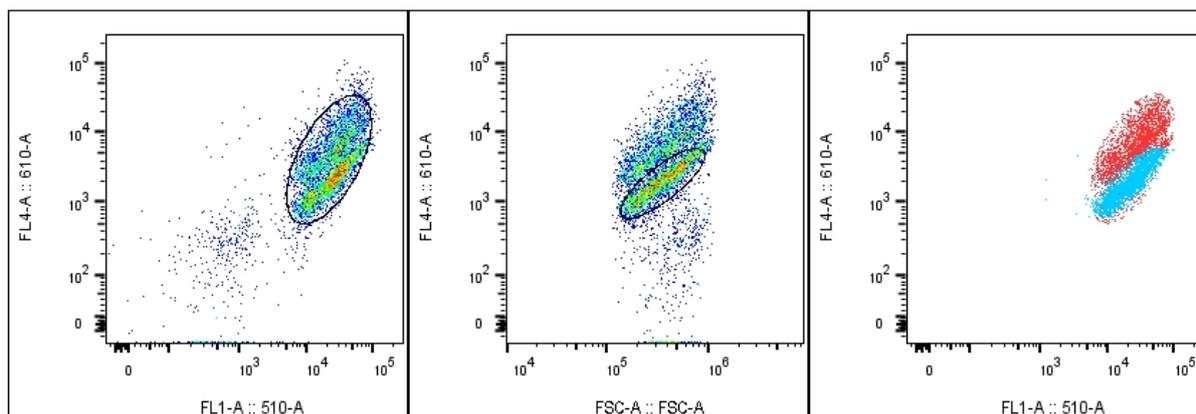
With the simulated dose-response surface of the 5'-NeoTc gate, the model predicted a residual fluorescence output for maximum repressed cells at the higher induction levels of 3-fold (5 nM  $\beta$ E) and 6-fold (10 nM  $\beta$ E) relative to the starting level of uninduced cells at ~500 AU, suggesting the continuous translation of a small number of mRNAs escaping any repressor interaction, consistent with steady state measurement of outgrown cultures. Considering the mechanism by which the 5'-NeoTc gate inhibits gene expression, it may be hypothesized that mRNA already transcribed in the absence of any repressor is densely occupied by ribonucleoprotein particles (RNP) and especially 43S pre-initiation complexes actively scanning the 5'-UTR for the start codon. As a consequence the mRNA is not accessible for the repressors and additionally the riboswitch structure is continuously compromised and unable to form the entropically disfavored conformation that can be captured by the ligand. On the contrary, the -1 h experiment illustrated in Figure 4.25 and a previous study suggest that a pre-incubation with repressors leads to a lower residual gene expression after repression. This is presumably because neomycin and

## Results & Discussion

tetracycline have already permeated through cell membrane and nucleus, encountering the newly transcribed mRNA before it is exposed to RNPs and the ribosomal pre-initiation complex<sup>139</sup>. Transient data of the +1h and +2h experiments shown in Figures 4.25 and 4.26 further support this mechanism, since the increase of GFP fluorescence is stopped within the hour of ligand addition, leaving only already transcribed mRNAs to be translated. A further decrease of fluorescence over time can be attributed to a progressive degradation of already translated GFP in concert with a gradual capture of 5'-NeoTc gates in their ligand binding conformation by the repressors, thus inhibiting gene expression.

The model was further used for a computational redesign of the complete expression system, exploring the influence of extrinsic noise on the distributional overlap of the different Boolean states. The GEV transcription factor presented an easy to monitor extrinsic noise factor as it is coupled to mCherry whose fluorescence is proportional to the concentration of GEV. The theory of noise propagation suggested that variability of GEV would result in increased variability of GFP. Interestingly, the model computed the counterintuitive alternative, a negligible influence of GEV variability on the phenotypic heterogeneity of GFP. The plasmid-borne copy number of *GEV* was subsequently reduced from 1-3 copies to a single copy by a genomic integration into the *His3* locus. Indeed, reduction of GEV variability did not influence the AUC analysis (Fig. 14). This finding is plausible, if the gating protocol providing the data for model building and calibration is considered. After the gating of a single-cell population, this population is displayed in a bivariate plot composed of both fluorescence channels. Figure 4.35 shows bivariate plots for an induced population bearing the integrated 5'-NeoTc gate and plasmid-borne GEV expression after 7 h. The circled subpopulation in plot A (that may itself be divided in at least two subpopulations) is subject to gating and will be used to build and calibrate the model, because all cells of that population exhibit a positive correlation between GFP (FL1-510) and mCherry (FL4-610). The other subpopulation is composed of cells that do not express GEV-mCherry and thus no GFP. Plot B shows a bivariate of mCherry and the forward scatter (FCS-A). Acknowledging the FSC as a measure for cell size, the plot correlates mCherry fluorescence to cell size. At least two subpopulations are clearly discernable that have the same FSC values, but differ in their fluorescence signal. A similar plot has been shown in Figure 34. Now, it can be verified in plot C, a bivariate overlay of the gated populations from A (red) and B (light blue), that the two subpopulations from B are identical with the two subpopulations discernable in the gated population of plot A. This in turn confirms the prediction of the model, because "stacked" subpopulations from plot B will also "stack" in plot A, thus only broadening the mCherry signal, but not GFP fluorescence. A main contributor to the width of a histogram is consequently identified as cell size. A reduction of GEV variability by reducing its gene copy number leads to a single subpopulation, for that matter the circled population in plot B, that, if plotted against GFP will not change the GFP signal and thus phenotypic noise cannot be reduced (compare Figure 4.34). It seems likely that a threshold for the concentration of GEV exists, that, once reached, provides enough GEV to fully induce GFP, irrespective of how much the threshold was exceeded.

## Results & Discussion

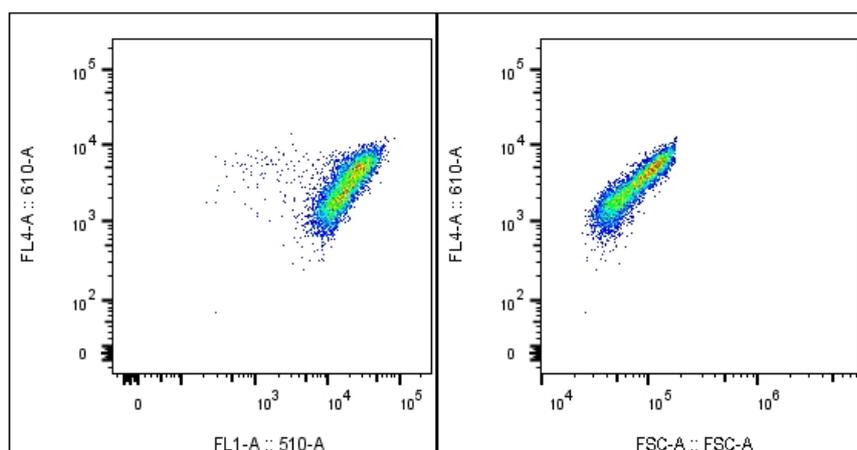


**Figure 4.35 Population analysis of cells expressing GEV from plasmid. A** The whole population as recorded by flow cytometry in a bivariate plot of mCherry (FL4) and GFP (FL1) fluorescence. The drawn gate marks the red subpopulation displayed in **C**. **B** The same population as displayed in **A**, but plotted as mCherry (FL4) fluorescence and forward scatter area signal (FSC-A). The gate denotes the light blue subpopulation in **C**. **C** An overlap of both gates from **A** and **B** in a bivariate plot of mCherry (FL4) and GFP (FL1) fluorescence. Data is exemplary and originates from cultures processed in Fig. 4.25.

In conclusion, a reduction of extrinsic noise did not lead to an increase of the actual gate performance as also all repressed populations were not affected by fluctuating GEV level. This suggests that noise propagation is not so much of an issue in this constructed expression system. A closer look at the data and a comparison with data recorded from the integrated GEV factor however, reveal a fundamental benefit at last. As immediately apparent from the bivariate plots in Figure 4.36 with data generated according to Figure 4.35, *GFP+* expression driven by the integrated GEV factor results in a single compact population. No subpopulation exist that does not exhibit mCherry and thus no GFP fluorescence. The FL4:FSC-A plot confirms the occurrence of a single gene copy of *GEV*. Although cell-to-cell variability of the GFP phenotype remains unchanged compared to Figure 4.35, the whole population as such is solely composed of cells exhibiting a uniform expression pattern. If GFP was replaced by a transcription factor, signal transduction would not be corrupted by non-responding cells – an inevitable event that, in contrast to the *in silico* domain, cannot be corrected *in vivo*. Since the model was built and calibrated from the gated data mimicking the GFP phenotype gained by a *GEV* integration, no improvement upon *GEV* integration was predicted.

Another benefit that arose with *GEV* integration was the 3-fold reduction of background fluorescence emitted by uninduced cells, increasing the induction-fold accordingly. The model also uncovered another extrinsic parameter that would improve the induction-fold of the system. The *GAL1* promoter is not saturated with activated GEV, since a dose dependency of gate output with respect to different induction levels was observed (e.g. Figure 4.30). The model determined that  $\beta$ -estradiol is the limiting factor of maximum GFP expression, a feature that would have remained hidden as it was experimentally not accessible due to growth deficiencies in the presence of higher  $\beta$ E level. Considering a volume of 100 fL per yeast cell (cell radius 3 $\mu$ m) and 10 nM  $\beta$ E, the approximate intracellular number of  $\beta$ E, if all  $\beta$ E was available as ligand for GEV, is 600. This number is surely to be reduced by the amount of  $\beta$ E remaining at or within the bilayered cell membrane, associated to lipids as its logP in an n-octanol:water mixture is 4, exhibiting the same hydrophobicity as polyaromatic hydrocarbons<sup>210</sup>. Increasing quantities of  $\beta$ E could thus promote a higher transcription rate.

## Results & Discussion



**Figure 4.36 Population analysis of cells expressing GEV from the *HIS3* locus.** In analogy to Fig. X, **A** shows the whole population as recorded by flow cytometry in a bivariate plot of mCherry (FL4) and GFP (FL1) fluorescence. **B** The same population as displayed in **A**, but plotted as mCherry (FL4) fluorescence and forward scatter area signal (FSC-A). Data was obtained from cultures presented in Fig. 14.

The GEV system had been further developed over the years to completely eliminate off-target induction of other pGAL1-driven genes, abolishing any defect on cell growth<sup>57,58</sup>. This was achieved by the fusion of engineered artificial transcription factors (AF) containing zinc-finger DNA-binding domains with the human estrogen receptor and the VP16 activation domain, both former constituents of the GEV factor. Thus, the system could still operate with  $\beta$ E, but induced only specifically modified *GAL1* promoters. The possibility to multiplex the response to  $\beta$ E by the simultaneous induction of multiple promoters with a single AF or the employment of different and orthogonal AF-promoter pairs opens the door to a new level of transcriptional control over gene expression. In concert with the here and elsewhere developed RNA-based logic gates, the specific characterization of such RNA devices with respect to extrinsic noise contributing to signal transduction and amplification could be assessed. As both components of the system, the inducing and the inducible, are fast-acting and highly efficient, a transient analysis on biologically relevant time scales is straightforward. In combination with the EnBase<sup>®</sup> technology even very long time lapse measurements could be conducted without the need to account for distorting metabolic adaptations of the growing yeast cells. As the results of the reduction of gene copy numbers implied a “cleansing” of non-responding cells from the total population, the adoption of this finding will further diminish losses during signal propagation. Consequently, other variable extrinsic factors like the abundance of components of the transcription and translation machinery and their impact on the performance of an artificially engineered genetic circuit can be studied unobscured from malfunctions inherent to the expression system. Additionally, intrinsic noise stemming from variable synthesis and degradation rates inherent to the studied device can be elucidated more easily by modifications and redesigns, now that they are entangled from excessive systemic errors. Albeit not presented in the results chapter, a second 5'-NeoTc gate with RFP as a reporter, but otherwise isogenic to the already integrated 5'-NeoTc gate, was also chromosomally integrated. This dual-color set up of equal gates was designed to better dissect extrinsic noise from intrinsic noise. Preliminary experiments revealed that GFP and RFP signals correlated well at steady state, suggesting minor influences of intrinsic noise, but

## Results & Discussion

rather a mutual dependence on extrinsic parameters as discussed. A preliminary transient characterization showed that GFP levels increased on faster timescales, but the phenotypic variabilities of GFP and RFP behaved equally. For one, this supports the finding of the steady state measurement. Secondly, since gene expression strongly depends on the local chromatin structure, the GFP device could respond faster due to issues with accessibility of the locus. A flipped integration, i.e. 5'-NeoTc-GFP moved from *ade2* to *his3* and 5'-NeoTc-RFP from *his3* to *ade2* was not yet conducted, but would shed some light on the matter.

## 5 Methods

### 5.1 General protocols for molecular cloning

#### 5.1.1 PCR

PCR was used to (i) amplify a defined sequence from a vector or other double-stranded oligonucleotide sequence with or without novel 5'-flanking parts and to (ii) confirm the correct ligation of a vector and insert after subsequent transformation into *E.coli*. To work-up the PCR (i) a column purification or (ii) a gel-purification or (iii) an alcohol precipitation was performed.

To amplify a distinct part of a vector or oligonucleotide primers were designed to meet the criteria listed in Table 5.1 and, if necessary, 5'-flanked with a new oligonucleotide sequence. Typically, these amplicons comprised two 5'-flanking restriction sites allowing ligation with a complementarily cut plasmid or up to 50 nt stretches homologous to the target vector in a Gibson cloning or yeast homologous recombination.

The applied general protocols are summarized below.

**Table 5.1 Criteria for primer design**

#	Parameter	value
1	<b>GC in % range</b>	50 – 60
2	<b>T<sub>m</sub> in °C range</b>	55 – 70
3	<b>T<sub>hyb</sub> in °C</b>	55
4	<b>Stability 5' vs 3' in kcals</b>	>0
5	<b>3'-Dimer matches in nt</b>	<4
6	<b>3'-GC clamp</b>	≥1 ultimate or penultimate 3'-nucleotide

These criteria were set manually in Clone Manager. Other pre-set criteria were adopted from the software.

**Table 5.2 PCR mix**

Concentration	Reagent	Volume [μl]	Final Concentration
100%	H <sub>2</sub> O	31.9	-
5x	Q5-Buffer	10	1x
100%	DMSO	1.5	3%
10 mM	dNTPs	1	200 μM
20 ng/μl	Vector	0.5	2 ng/μl
100 μM	Primer-fw	0.3	600 nM
100 μM	Primer-rev	0.3	600 nM
2 U/μl	Q5 Polymerase	0.5	1 U
	<b>Endvolume</b>	<b>50</b>	

Table 5.3 Thermocycler program

#	Step	Duration [s]	Temperature [°C]
1	Initial Denaturation	30	98
2	Denaturation	10	98
3	Hybridisation	15-60	58-72
4	Extension	2 s/100 nt	72
5	Final Extension	180	72
6	Cool-down	∞	8

} 35 cycles

Temperatures for primer hybridization and durations for primer extension were adjusted to target length and primer-template hybridization temperatures.

### 5.1.2 Colony-PCR

By means of Colony-PCR, several bacterial clones were screened by amplification of a distinct sequence targeted by an appropriate set of primers. The reaction mix was visualized on an agarose gel. Appearance of the desired band on the gel implied the insertion or deletion of a particular sequence of the vector used.

A bacterial colony was picked with a sterile toothpick from an LB-agarose plate and suspended in 20-50 µl H<sub>2</sub>O for colonies with a diameter <1 mm. 1 µl was taken for PCR.

Table 5.4 Reaction mix

Concentration	Reagent	Volume [µl]	Final Concentration
100%	H <sub>2</sub> O	6.8	-
10x	Taq-Buffer	1	1x
10 mM	dNTPs	0.2	200 µM
10 µM	Primer-fwd	0.4	400 nM
10 µM	Primer-rev	0.4	400 nM
5 U/µl	Taq-Polymerase	0.2	1 U
-	Bacterial colony	1	-
<b>Endvolume</b>		<b>10</b>	

Table 5.5 Thermocycler program

#	Step	Duration [s]	Temperature [°C]
1	Initial Denaturation	600	95
2	Denaturation	10	95
3	Annealing	30	48
4	Extension	6 s/100 nt	72
5	Cool-down	∞	4

} 24 cycles

Extension time is varied according to target sequence.

## Methods

### 5.1.3 Oligonucleotide Hybridization and 5'-Phosphorylation

To clone short oligonucleotides (<80nt) without a template for PCR the complementary single strands were ordered either pre-cut to match the restriction site of the target vector or blunt to serve as a template for PCR. The double-stranded products were 5'-phosphorylated, if further subjected to a sticky end ligation. The protocol is displayed in table X.

**Table 5.6 Reaction mix hy**

Concentration	Reagent	Vol [ $\mu$ l]	Final concentration
10x	T4-Polynucleotide Kinase Buffer	5	1x
100 $\mu$ M	Oligo fwd	1.5	1.5 pmol/50 $\mu$ l
100 $\mu$ M	Oligo rev	1.5	1.5 pmol/50 $\mu$ l
10 mM	ATP	5	1 mM
10 U/ $\mu$ l	T4-Polynucleotide Kinase (T4-PNK)	1	0.2 U/ $\mu$ l
100%	H <sub>2</sub> O	ad 50	

The mix without T4-PNK was incubated 2 min at 98 °C and then cooled at 0.1 °C/s to 25 °C. T4-PNK was added, the reaction mixed and incubated for 1 h at 37 °C. Phosphorylated double-stranded oligonucleotides were used without further purification for ligation or as template in a PCR.

### 5.1.4 Fill-in reaction

To ligate a linear vector with incompatible sticky ends, blunt ends were created by a fill-in reaction at 12 °C for 15 min. The reaction was stopped by adding EDTA to a final concentration of 10 mM and incubation at 75 °C for 20 min. To shift the equilibrium of 5'→3'-polymerisation versus 3'→5'-exonucleolytic degradation towards the fill-in, low temperature, low enzyme concentration and high dNTP concentration was chosen.

**Table 5.7 Reaction mix**

Concentration	Reagent	Vol [ $\mu$ l]	Final concentration
10x	NEB4	5	1x
1mg/ml	BSA	5	100 $\mu$ g/ml
-	Linear vector	500 ng	-
2 mM	dNTPs	2.5	100 mM
3 U/ $\mu$ l	T4-DNA Polymerase	0.3	0.02 U/ $\mu$ l
100%	H <sub>2</sub> O	ad 50	

### 5.1.5 Restriction digest

#### Type IIM<sup>(165)</sup> - DpnI digest

After PCR the template vector was digested at 37°C for 1h with Dpn I, a type IIM restriction enzyme, with cleavage specificity to (hemi-) methylated DNA. The PCR product remained undigested and was isolated by column purification, gel extraction or alcohol precipitation. 1 µl 20U/µl Dpn I was added directly to the PCR mix and simultaneously removed by PCR purification.

#### Type IIP<sup>(165)</sup> (orthodox)

Plasmids and inserts that were subject to ligation were incubated for 1 h or up to 24h at 37°C with two restriction enzymes type IIP. A digested vector was purified by gel extraction. PCR products were purified by column purification or alcohol precipitation. The enzymes were inactivated at 80°C for 20 min or excluded in the purification step. The general reaction volume was 50 µl.

**Table 5.8 Reaction mix**

Concentration	Reagent	Vol [µl]	Final concentration
10x	CutSmart Buffer or as indicated	5	1x
-	Vector or	2-5 µg	
-	Insert	Purified PCR reaction	
20 U/µl	Restriction enzyme	1-2	0.4-0.8 U/µl
100%	H <sub>2</sub> O	ad 50	

### 5.1.6 Gibson cloning reaction

Gibson cloning was employed to fuse one or more inserts with a target vector taking advantage of the introduced 40 nt homology between inserts and vector by a previous PCR. A 20 µl Gibson reaction contained 0.02-0.5 pmol DNA fragments in a total of 5 µl that were mixed with 15 µl home-made Gibson master mix (Ref openwetware) and incubated at 50 °C for 1 h. PCR fragments and linearized vector were purified as described. A downscale of the 5x ISO mix and Gibson master mix was prepared. After incubation the mix was purified by BuOH precipitation and dissolved in 10 µl water. 1.5 µl were taken to transform electro-competent Top10 cells.

**Table 5.9 Gibson master mix**

Concentration	Reagent	Vol [µl]	Final concentration
5x	ISO buffer	320	1x
10 U/µl	T5 Exonuclease	0.64	0.004 U/µl
2 U/µl	Phusion Polymerase	20	0.025 U/µl
40 U/µl	Taq Ligase	160	0.1 U/µl
100%	H <sub>2</sub> O	ad 1600	

## Methods

Typically, 300 µl Gibson master mix were prepared and stored at -20 °C. Up to 10 freeze-thaw cycles are possible. Here, up to 4 cycles were tested without a compromised activity. 5x ISO mix was stored at -20 °C in 500 µl aliquots.

### 5.1.7 Ligation

The fusion of an insert and a vector backbone cut with identical restriction enzymes was carried out at RT for 45 min. The ratio vector : insert was chosen with respect to insert size and resulted in typical ratios of 1:3 – 1:5. Always 25 ng Vector were used for ligation. The reaction was purified by BuOH precipitation and dissolved in 10 µl water. 1.5 µl were taken for transformation of electro-competent Top10 cells.

**Table 5.10 Reaction mix**

Concentration	Reagent	Vol [µl]
10x	T4-DNA Ligation Buffer	2
-	Linear vector	25 ng
-	Insert	q.s.
400 U/µl	T4-DNA Ligase	1
100%	H <sub>2</sub> O	ad 20

### 5.1.8 Transformation of chemo-competent E.coli in 96-well format

Cells were thawed on ice. 100 µl cells were added to the ligation mix and incubated 1 h on ice. Cells were heat-shocked at 42°C for 45s in a PCR cycler. Cells were then incubated 10 min on ice. Subsequently, the mixture was pipetted in 1 ml LB in 1.5 ml reaction tubes and incubated at 37°C for 1 h in a 37°C incubator. Reaction tubes were inverted occasionally using the racks as sandwich cover. Cells were concentrated by centrifugation at 8000 rpm for 15 s und plated on LB plates containing 100 µg/ml ampicillin. Plates were incubated at 37 °C overnight.

### 5.1.9 Transformation of electro-competent E.coli

50 µl of electro-competent E. coli Top 10 cells were thawed on ice, mixed with 1.5 µl purified ligation or Gibson mix, transferred to a sterile electroporation cuvette pre-chilled on ice and electro-shocked at 1800 V for 5ms. 450 µl sterile SOC-medium at 37 °C were added and plated on an LB-agarose dish supplemented with 100 µg/ml ampicillin under aseptic conditions before incubation at 37 °C overnight.

### 5.1.10 General protocol for yeast homologous recombination

To facilitate a ligation reaction exploiting the highly efficient DNA repair mechanisms in yeast a purified and linearized vector and one or more inserts were transformed into yeast. Inserts and vector bear a total of 40 nt sequence homology at the junction of two parts to facilitate an efficient recombination event. The molar ratio insert : vector was set to 1 : 5 – 1 : 10.

### **5.1.11 Transformation of competent yeast cells**

The Frozen Yeast EZ II Transformation Kit from Zymo Research was chosen to conduct all yeast transformations. The procedure was carried out according to the manufacturer's instructions.

### **5.1.12 Agarose gel electrophoresis**

For DNA extraction and analysis high quality agarose was utilized. After casting gels with agarose concentrations ranging from 1% to 3%, samples were mixed with agarose-loading buffer and pipetted into pre-formed gel slots alongside an adequate molecular length marker. Gels were run in 1x TAE buffer at 6 V/cm for 35 – 60 min. Next, gels were stained with 0.5 µg/ml ethidium bromide in water for 1 – 5 min. DNA was visualized with 254 nm light for analytical purposes and 366 nm for extraction. Afterwards the gel was digested and the DNA extracted using the Qiagen Gel extraction Kit.

### **5.1.13 DNA isolation – Mini preparation from Top 10 cells**

All plasmids were isolated with the Qiagen Plasmid MiniPrep Kit. Usually, pDNA was eluted in 55 µl and stored at -20 °C for further use.

### **5.1.14 DNA isolation – Mini preparation from yeast**

The following steps were taken from a user-developed protocol published by Qiagen. All steps were performed at RT. 4 ml saturated yeast culture were harvested by centrifugation at 13000 rpm for 1 min. Cells were resuspended in 250 µl P1 Buffer. 150 µl glass beads (diameter 0.25 – 0.5 mm) were added and vortexed for 5 min. Glass beads were allowed to settle, the supernatant was transferred to another reaction tube. 350 µl N3 Buffer were supplemented and incubated for 5 min. From here the standard protocol for a plasmid mini preparation according to the Qiagen Plasmid Mini Prep Kit was followed. The DNA was eluted in 20 µl and 2 µl were taken for subsequent transformation of Top 10 cells.

### **5.1.15 DNA isolation – genomic DNA preparation from yeast (Bust'n'grab)**

In general gDNA was extracted by 2 heat-freeze cycles. First 1-2 ml saturated yeast culture were harvested by centrifugation at 13000 rpm for 1 min. Cells were resuspended in 200 µl STET Buffer in PCR reaction tubes and subjected to two rounds of freezing at -80 °C and heating at 98 °C. The supernatant was extracted with 1 Vol. phenol/chloroform/isoamyl alcohol (25:24:1). The aqueous phase was again extracted with 1 Vol. chloroform. DNA was precipitated by isopropanol and dissolved in 100 µl water. The concentration was measured and the samples stored at -20 °C. For a subsequent PCR reaction 1ng/µl PCR reaction were taken.

### **5.1.16 DNA quantification**

The Nanodrop ND-1000 was used to measure the concentration of 2 µl DNA solution at 260 nm employing Lambert-Beer's law assuming only marginal contamination by RNA and/or genomic DNA.

## Methods

### 5.1.17 DNA purification – Ethanol, isopropanol and butanol

1. EtOH: 1/10 vol. 3 M sodium acetate was added to the DNA solution and mixed. 2 – 3 vol. ethanol were added and mixed. The mixture was incubated at 4 °C for 15 – 20 h, centrifuged at 4 °C at 13000 rpm for 30 min and washed one time with 70% ethanol followed by centrifugation for 15 min again at 13000 rpm and 4 °C. DNA was dried at RT for 20 min, resolved in H<sub>2</sub>O and kept at -20 °C.
2. 2-Propanol: see EtOH protocol. Instead of 2 – 3 vol. EtOH 1 vol. isopropanol was used.
3. 1-Butanol: 10 vol. BuOH were added and thoroughly mixed. The precipitate was collected by immediate centrifugation at 13000 rpm for 10 min. A 70% EtOH wash was performed prior to dissolving the DNA pellet in water.

### 5.1.18 Sequence analysis

To confirm sequence integrity and successful cloning single clones were picked from a LB-agarose plates supplemented with 100 µg/ml ampicillin and incubated in LB-medium containing ampicillin at 100 µg/ml for about 15h. After mini preparation, pDNA was analyzed by Seqlab laboratories, Goettingen.

### 5.1.19 Preparation of chemo-competent E.coli

The desired strain, typically Top10, was taken from the frozen stock and streaked out on an LB plate to yield single colonies. 4 ml LB were inoculated with a single colony and incubated at 37 °C, 150 rpm for 16 h. 300 ml pre-warmed LB medium were inoculated to an OD<sub>600</sub> of 0.1 – 0.2 and grown to an OD<sub>600</sub> of 0.6 – 0.8. Cells were cooled to 4°C on ice and harvested by centrifugation with 8000 rpm for 20 min at 4°C. Cells were washed with ice-cold 150 ml 0.1 M CaCl<sub>2</sub> and resuspended in 25 ml 0.1 M CaCl<sub>2</sub>. Sterile glycerol was used to adjust the cell suspension to 15% glycerol. 1ml aliquots were frozen at -80°C.

### 5.1.20 Preparation of electro-competent E.coli

The desired strain, typically Top10, was taken from the frozen stock and streaked out on an LB plate to yield single colonies. 4 ml LB were inoculated with a single colony and incubated at 37 °C, 150 rpm for 16 h. 300 ml pre-warmed LB medium were inoculated to an OD<sub>600</sub> of 0.1 – 0.2 and grown to an OD<sub>600</sub> of 0.6 – 0.8. Cells were cooled to 4°C on ice and harvested by centrifugation with 8000 rpm for 20 min at 4°C. Cells were washed with ice-cold 150 ml 10% glycerol (v/v) twice. Cells were resuspended in 10% glycerol, 100 µl aliquots were prepared and frozen at -80°C.

### 5.1.21 Preparation of chemo-competent yeast

The Frozen Yeast EZ II Transformation Kit from Zymo Research was chosen to prepare competent yeast cells. The procedure was carried out according to the manufacturer's instructions.

### 5.1.22 Yeast glycerol stocks

4 ml YPD or SCD full medium were inoculated with a single yeast colony and cultured for 20h. Sterile glycerol was added to a final concentration of 15% in a total of 1 ml. This aliquot was incubated at 30 °C for 30 min and finally stored at -80 °C.

## 5.2 Adapted protocols for molecular cloning and reporter gene analysis

### 5.2.1 *In vivo* Screening of RNA aptamers in *S. cerevisiae*

#### 5.2.1.1 SELEX – pool design and selection strategy

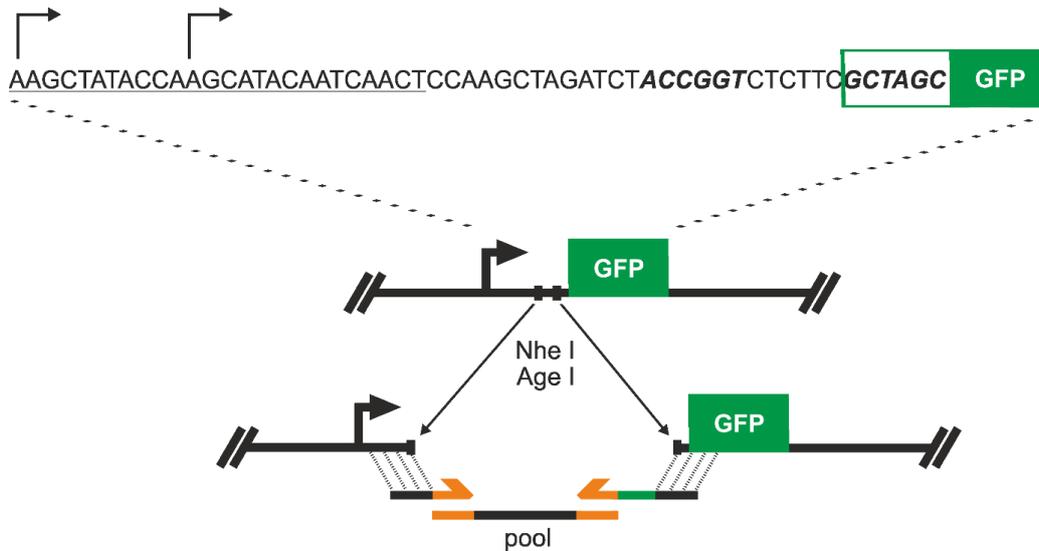
SELEX – Systematic Evolution of Ligands by Exponential enrichment, is a process used to identify RNA or DNA sequences competent to bind ligands of choice<sup>211</sup>. A SELEX was not conducted in this study and therefore the general procedure for an RNA SELEX is described briefly. A randomized pool of DNA is generated by DNA synthesis. The pool is transcribed to RNA and incubated with the desired ligand. The ligand is coupled to a solid state, a column, or to a bead suspended in solution. By incubation and several washing steps unspecific binders are discarded and all sequences bound to the ligand are eluted. A reverse-transcription yields DNA that is amplified by a PCR and the cycle concludes. Up to 10 or more cycles may be performed to narrow the pool down to the best binders.

Here, it must be pointed out that for the design of the initial pool a rather long randomized region of 50-90 nt was considered that is optimal for the identification of small molecule-binding aptamers<sup>211,212</sup>. Long nucleotide stretches exhibit a greater structural flexibility and may probe the limited binding sites on a small molecule more efficiently<sup>154</sup>. To bias the evolution of RNA sequences towards a stable folding, structural elements as hairpins can be considered in the pool design<sup>213,214</sup>. This strategy however, as it compromises structural flexibility, had not proven to be beneficial to obtain small molecule binding aptamers that exhibit riboswitching properties. The number of selection rounds depends on the target molecule, but typically 5-10 rounds of enrichment are performed yielding pools with moderate ( $\mu$ M) to high (nM) affinity binding profiles.

#### 5.2.1.2 RNA-library cloning for *in vivo* screening

The SELEX approach from 2.1 yielded RNA aptamer pools from different rounds of selection differing in their enrichment of sequences that exhibit binding towards the target molecule. Here, two pools were used, the pool from round 9 with presumably strong aptamer target interaction and the pool from round 6 with average binding properties. Copy DNA pools from rounds 6 and 9 were PCR-amplified with a forward (fwd) and reverse (rev) primer. These primers attach 5'-overhangs that are homologous to the 5' UTR upstream of the green fluorescent reporter gene (*gfp*) and 3'-overhangs to a 5' part of the coding sequence of the reporter gene itself (for details see Figure 5.1). After PCR amplification of the selected pools with the correct primers the PCR product was used without further purification, nonetheless the correct length of the product was confirmed by gel electrophoresis.

## Methods



**Figure 5.1 SELEX pool cloning by homologous recombination in yeast.** The cloning site of vector pWHE601\* upstream of the *GFP+* reporter gene is depicted. Upper panel: The two proposed TSS of the *ADH1* promoter are indicated by arrows and the ultimate 3'-sequence of the promoter is underlined. The *AgeI* and *NheI* restriction sites are highlighted in bold and italicized. The *GFP+* reporter gene is emphasized in green, including the *NheI* restriction site. The vector is linearized using both restriction sites. The SELEX pool is amplified with primers (orange arrows) attaching overhangs (black extension of orange arrows) homologous to the sequence stretch upstream of *AgeI* and downstream of *NheI*. The rev primer introduces the Kozak-Start sequence (green line).

The expression of GFP from the 2 $\mu$  yeast shuttle vector pWHE601 is constitutively driven by a  $P_{ADH1}$  promoter with a 5' UTR of 34 to 44 nt since two transcription start sites are described<sup>155</sup>. The template vector, termed pWHE601\*, used in this study is a derivative of the aforementioned pWHE601 where the start codon has been deleted (also known as I.BB – lab internal communication). For homologous recombination it was prepared by cutting with two restriction enzymes that produce incompatible ends – *NheI* and *AgeI*. The restriction sites reside immediately upstream of the reporter gene to promote an efficient recombination event between promoter sequence and *gfp* gene<sup>155</sup>. The start codon including the yeast Kozak sequence 5'-AAA-start-G-3' was introduced with the 5' overhang of the rev primer<sup>201,202</sup>. This ensures that only yeast cells that have undergone an accurate homologous recombination will be selected for the *in vivo* screening step as incorrect recombined plasmids or survivors of the restriction digest cannot code for the fluorescent reporter and therefore remain colorless. That said, to minimize transformation of uncut plasmids, a gel extraction of the cut plasmids after the restriction digest was performed.

Overall, one transformation mix for a homologous recombination contains around 100 fmol gel purified vector, a 10:1 molar excess of the PCR-amplified inserts and competent yeast cells according to the protocol used. Transformation was carried out according to the protocol of the Frozen Yeast Transformation Kit II from Zymo Research or a lithium acetate-based method<sup>215</sup>. The transformation mix was plated on up to 40 petri dishes with a diameter of 9.2 cm. This assured a moderate colony density that simplified the selection of clones for the *in vivo* screening. Agar plates were composed of minimal medium (0.2% [w/v] yeast nitrogen base, 0.55% [w/v] ammonium sulfate, 2% [w/v] glucose, 12  $\mu$ g/mL adenine, MEM amino acid mix). The transformed yeast cells were cultured for 3-4 days at 28-30°C.

## Methods

### 5.2.1.3 Selection – Fluorescence Microscopy

The selection step was accomplished employing a standard stereo microscope (M50, Leica Microsystems) equipped with a setup to visualize the fluorescence of interest. Petri dishes were scanned by eye for fluorescent colonies. Controls have to be transformed alongside the pool to determine the minimal and maximal fluorescence. The negative control accords to the vector backbone pWHE601\* unable to express GFP and the positive control, pWHE601, constitutively expressing GFP from the same backbone vector, but containing a start codon and Kozak sequence. In addition it is helpful to have an already working riboswitch available to get an idea which fluorescence levels are to be expected approximately. Fluorescent colonies were marked with a pen on the bottom of the petri dish and later picked into a 96-well plate as described the next section. Colonies were select spanning the whole spectrum of fluorescence intensities. Only those colonies displaying no or a very weak fluorescence were omitted.

### 5.2.1.4 *In vivo* Screening

The screening step was carried out in transparent 96-well plates using the Tecan plate reader Infinite M200 Pro. Colonies chosen from plate in 2.3 were incubated as pre-culture in 200  $\mu$ l minimal medium for 16-20 h at 30°C using 96-well format, U-bottom, on a plate shaker Titramax 1000 (Heidolph *instruments*). The pre-culture was then transferred to fresh medium and grown for additional 24 h in the absence and presence of the target molecule. 20  $\mu$ l are drawn each time to inoculate 2 triplicates for each independently grown colony. The first triplicate is cultured in absence of the ligand, the second triplicate in presence of the target molecule. These cultures are used to determine whether addition of the ligand reduces fluorescence levels and thereby indicates the presence of a working riboswitch transcribed by the corresponding yeast cells.

Fluorescence was measured from the top at an emission wavelength of 510 nm after excitation at 482 nm at an optical density ( $OD_{600}$ ) of 0.6-0.9 in 96-well flat-bottom plates. Cultures were diluted accordingly in PBS and have to be shaken up shortly prior to analysis. In addition to fluorescence, the  $OD_{600}$  was recorded in parallel to normalize the fluorescence signal for cell growth. Yeast cells transformed with the positive and negative control as well as a working riboswitch were included in each plate and measured alongside the picked colonies. The determined fluorescence signal was normalized to an  $OD_{600}$  of 1, the fluorescence of the negative control was subtracted and the signal is displayed as relative fluorescence with the positive control set to 100% signal intensity. The riboswitch control was used to check whether a systematic error had occurred, since the fluorescence values including the dynamic change of fluorescence of cells expressing this construct was known.

Cells in wells displaying a dynamic change of the fluorescent signal were spread out on agar plates and singularized. This separation step (singularization) after the first screening round was important, because sometimes a partitioned fluorescence distribution within a given colony was observed that is due to an uptake of multiple plasmids at once. Four individual colonies from one original candidate were incubated and subjected to a second round of screening as described above.

## Methods

Typically, only few colonies were identified to show a change in fluorescence signal intensity upon ligand addition. The dynamic range of the signal change was rather low and a threshold needed to be set above which candidates were chosen. Once candidates had been verified in the second screening round, the plasmids were isolated and transformed into *E. coli*. Isolated plasmids should then be retransformed into yeast and measured to confirm their switching activity. Ultimately, the plasmids were sequenced.

### 5.2.1.5 Fluorescence spectroscopy reporter assay

In general, reporter plasmids containing a fluorescent reporter protein with or without a preceding riboswitch in the 5'-UTR were transformed into yeast and gene expression was measured by fluorescence spectroscopy on a Fluorolog FL3-22 instrument from Horiba. All experiments were performed in two biological triplicates. Yeast cultures were grown to saturation in 4ml SCD minimal medium at 30 °C and used to inoculate a fresh culture in a 1:100 dilution. These cultures were grown in 1.5 ml SCD minimal medium in 24-well plates for 24 h in absence and presence of the cognate ligand. A positive and negative control were treated equally and used to calculate basal expression and switching performance of the reporter constructs. Initially, 500 µl culture were harvested by centrifugation at 13000 rpm for 30 s and resuspended in 1 ml PBS. 500 µl were transferred to 1.5 ml PBS in a 2 ml four-way cuvette to yield a 1:8 dilution of the original optical density about 6 at 600 nm of. This suspension was also taken to measure the OD<sub>600</sub>. GFP was excited at 482 nm and fluorescence was measured from 500 nm to 520 nm. Fluorescence values at 510 nm were taken for calculations. Since a plateau in the emission spectrum emerges from 508 nm to 512 nm, errors are kept at a minimum. Entrance and exit slits were set to allow a width of 2 nm for the path of light. Dark current was automatically subtracted and fluorescence recorded in counts per second. To calculate the basal expression of yeast cells expressing a riboswitch in the 5'-UTR, all fluorescence values were corrected for cell growth, background fluorescence emitted by the negative control was subtracted and the fluorescence values of the constructs were divided by the fluorescence values of the positive control. This protocol was followed for samples treated with and without ligand, such that by the division of untreated by treated conditions an x-fold reduction of gene expression could be determined.

### 5.2.1.6 *In silico* rational design

The secondary structure of the identified sequences was simulated deploying online tools such as the Vienna RNA package <sup>174</sup>. Based on these predictions it was now necessary to track down the minimal motif that is essential for ligand binding and *in vivo* activity. This was done by carefully evaluating the *in silico* generated structural data and a sequence alignments from truncated versions <sup>138,216</sup>. For the design of primers that will truncate the sequence step by step it was found to be critical to start with short steps of about 5 nt, as it is completely unknown where the binding site and the structural information for *in vivo* activity are located. Truncated sequences were cloned using standard cloning techniques. After cloning, the plasmids were sequenced and the regulatory effect was measured after transformation into yeast.

## 5.2.2 Characterization of a riboswitching expression system

**P<sub>ADH1</sub>-driven device:** A modified version of the plasmid pWHE601 was used as the backbone vector to clone the riboswitch parts downstream of the *ADH1* promoter, generating the NOR and NOT devices shown in Figures 4.18 and 6.9. Plasmids were transformed into the electrocompetent *E. coli* strain TOP10 grown in LB medium supplemented with 100 µg/ml ampicillin.

**P<sub>GAL1</sub>-driven device:** The plasmid pSP212-P<sub>TEF1</sub>-GLC-T<sub>CYC1</sub> expresses the GEV transcription factor that is C-terminally coupled to mCherry and was transformed into competent yeast containing the NeoTc device integrated into the *ade2* locus.

### 5.2.2.1 Genome manipulations

Genotype RS453α                      *mata ade2-1 trp1-1 can1-100 leu2-3 his3-1 ura3-52*

Genotype CEN.PK111-27B        *mata leu2-3 trp1-1*

Genomic integrations were performed by the CRISPR/Cas9-based method *CASEMBLR*<sup>45</sup>. The yeast shuttle plasmids p414-P<sub>TEF1</sub>-Cas9-T<sub>CYC1</sub> (plasmid #83946)<sup>217</sup> and p426-pSNR52-gRNA.CAN1.Y-T<sub>SUP4</sub> (plasmid #43803)<sup>72</sup> were ordered from *addgene.org*. p414 (Trp<sup>-</sup>) expresses the Cas9 endonuclease and p426 (URA marker) harbors the gRNA expression cassette that was transferred to p425 (*Leu*<sup>-</sup>) as the *URA3* locus presented a putative integration site and the CEN.PK111-27B strain is Ura<sup>+</sup>. Gibson assembly was used to simultaneously ligate the gRNA expression cassette amplified in one part from p426 (broken in two by a *Cla* I restriction site) with one of the three gRNAs targeting the *ade2*, *his3* and *ura3* loci and the linearized p425 plasmid, excluding the original gRNA. Initially, as described in the publication from Jakoucinas *et al.*, yeast strain CEN.PK 111-27B was used for integrations into the *ade2* locus, resulting in null mutants accumulating a red pigment that absorbs green fluorescence light. Experiments with GFP as a reporter protein were thus found to be impracticable. The strain was therefore switched to RS453α that, too, is a null mutant with respect to the *ade2* gene, but passaged long enough to have a white phenotype due to mutations of enzymes in the upstream *de novo* biosynthesis of purine nucleotides, thus preventing the occurrence of a red pigment. For integrations into the *his3* and *ura3* loci strain RS453α was taken as well. DNA parts for integration were assembled by PCR, purified, and EtOH precipitated to increase their concentration. Parts featured 50 nt homology upstream and downstream to the neighboring part. Overlaps to the genomic target sites were found to be at least 100 nt for a integration without extra homology block to confer correct assembly. Using an extra homology block, that is, a 300-500 nt dsDNA block identical to the upstream and downstream integration site, 50 nt overlap to such a block are sufficient to yield correct integrations. 2-4 pmol/part were mixed with 0.5 µg p414-Cas9 and 1 µg p425-gRNA. Transformations were conducted according to the instructions provided with the Frozen-EZ Yeast Transformation II Kit from Zymo Research and spread on appropriate SCD drop-out agar plates (see below for details on growth medium and protocol). As may be deduced from the genotype of this strain, neither integration site could be exploited for a phenotypic screening to detect locus destruction. However, it was found, in accordance with the developers of the method, that integration efficiency as confirmed by yeast colony PCR was between 80-100 % for both loci, using primer pairs within and upstream or downstream of the integration sites.

### 5.2.2.2 Cultivation of yeast

The Frozen-EZ Yeast Transformation II Kit was used for the preparation of competent yeast cells from strain RS453 $\alpha$  and their subsequent transformation on appropriate SCD drop-out plates (0,2% YNB w/o AA (Difco), 0,15% Yeast Synthetic Drop-out Medium w/o uracil, leucine and tryptophan (Sigma Aldrich), 2% dextrose (Carl Roth) and 1,8% agar (Carl Roth). Uracil, leucine and tryptophan were added as indicated on the complete formula of the Yeast Synthetic Drop-out Medium.

After 3 d growth at 30°C in a humidified incubator, single colonies were cultured in 5 ml SCD drop-out medium for 24 h, then diluted to an OD<sub>600</sub> of approx. 0.05 (flow cytometry) or 0.01 (microfluidic measurements) and grown for additional 10 h (flow cytometry) or 5 h (microfluidic measurements) in 5 ml SCD drop-out medium to obtain actively growing cells in early to mid-log phase for the subsequent experiments.

Cells for bulk measurements were grown in 3 ml SCD drop-out medium without repressors or supplemented with 100  $\mu$ M neomycin and /or 250  $\mu$ M tetracycline in 12-well plates for 24 h and analyzed in two independent triplicates. Fluorescence of all cultures was OD<sub>600</sub>- and background-corrected and is displayed relative to a positive control that is not riboswitch-controlled, but treated equally.

### 5.2.2.3 Cytometry

All cytometry measurements were performed on a CytoFlex S instrument from Beckman Coulter equipped with a 488 nm laser and a 561 nm laser for excitation of GFP and mCherry, respectively. Emission light was bandpass-filtered at 510/20 nm or 610/20 nm, respectively.

Cells taken from the 10 h pre-culture were split to 12-well plates at an OD<sub>600</sub> of 0.05 and cultured for 10 h at 30 °C in 3 ml SCD drop-out medium. This protocol generates yeast cultures that are in the mid-log phase at the time point of induction and grow until late log phase over the time course of the experiment. 4 separate experiments with regard to the timing of repression at -1 h, 0 h, +1 h and +2 h relative to the induction with 10 nM  $\beta$ -estradiol at 0 h were performed. 10  $\mu$ l - 30  $\mu$ l samples were drawn hourly for 7 h (Figure 3). Experiments to investigate dose-dependencies of the gate on repressor and inducer concentrations were conducted with 2.5, 5 or 10 nM  $\beta$ -estradiol and 2, 10, 50 and 100  $\mu$ M neomycin and 0.5, 2, 10, 250  $\mu$ M tetracycline. Here, repressors were added +2 h relative to induction and 10  $\mu$ l - 30  $\mu$ l samples were drawn at 0 h, 2 h, 3 h, 5 h and 7 h.

### 5.2.2.4 Gating

10,000 events were recorded for each sample. Doublets were removed by drawing a gate in the FCS-H and FSC-A dot plot. Doublets are events displaying larger signal integration values (FSC-A) compared to other events with the same FSC-H value. This subpopulation was then subjected to a dot plot fluorescence gating (FL1-510 and FL2-610) to remove non-responsive cells. Non-responders are cells emitting no mCherry fluorescence and thus no GFP fluorescence. The final subset was plotted over time.

### 5.2.2.5 Time-lapse Microscopy

For *in vivo* time-lapse experiments, cells from the 5 h pre-culture were trapped on a PDMS-based microfluidic chip, inspired by the Alcatraz chip <sup>218</sup>. Single yeast cells were kept in place in a continuous flow for several hours and images were recorded every 10 min at 3 different focal planes. Imaging was performed on a Nikon Eclipse TI with a 100x NA 1.45 objective, ORCA Flash4.0 camera (Hamamatsu) with SpectraX light engine (Lumencor) at 470/24 nm for GFP and 575/25 nm for mCherry. Cells were kept at 30°C with an incubator box (life imaging services). Concentrations of inducer and repressors were used according to the cytometry measurements: 10 nM  $\beta$ -estradiol, 100  $\mu$ M neomycin and 250  $\mu$ M tetracycline. Cells were induced for 2 h, then exposed to single or both repressors. Single cell traces were extracted with a FIJI/Matlab script.

### 5.2.3 Tc-Dimers

#### 5.2.3.1 Cloning

Standard cloning protocols as described in chapter 5.1.1 were applied and scaled up accordingly. All reactions were performed in PCR reaction tubes and 96-well format. Primers were ordered as 96-well plates containing all forward and all reverse primers in two separate plates. As PCR template a dsDNA stretch of the 3'-aptamer was used and amplified with fwd and rev primers attaching new stems and thereby generating individual new aptamers. PCR reactions were precipitated by isopropanol/sodium acetate employing a table-top centrifuge equipped with a rotor for PCR stripes and placed in a fridge. The complete PCR reaction was digested with Age I and Nhe I alongside the vector pGFP3 for 16 h at 37 °C. Restriction enzymes present in the insert digests were heat-inactivated at 80 °C for 30 min and stored at -20 °C. The digest of pGFP3 was gel-purified from a 1% agarose gel at 7000 nt to exclude enzymes and uncut plasmid. Singly cut plasmids could not be excluded, because restriction of pGFP3 reduces the vector size only by 120 nt, which cannot be separately detected on a 1% agarose gel due to the broad band pattern of the migrating linearized dsDNA. Inserts and linearized pGFP3 were ligated at RT for 1 h in 96 separate reactions and used to transform chemo-competent Top10 cells.

Note: the heat shock was performed in a PCR thermocycler for 60s to allow the reaction to heat up from 0 °C and reach 42 °C. After heat shock and cooling the transformation mixture was transferred to 1.5 ml reaction tubes filled with 1 ml LB. Cells were spread with glass beads on LB-Amp agar plates.

## 6 Material

The materials and instruments that were utilized in this work are listed below. The following section contains the lists for chemicals and reagents (Table 6.1), instrumentation (Table 6.2), kits and commercially available systems (Table 6.3), enzymes and proteins (Table 6.4), protein standards and DNA ladders (Table 6.5), consumables (Table 6.6), antibodies (Table 6.7), cell strains (Table 6.8), buffers and solutions (Table 6.9), oligonucleotides (Table 6.10) and plasmids (Table 6.11).

All buffers and solutions were produced using deionized water or aqua valde purificata. If necessary, buffers and solutions were sterilized by autoclaving at 121°C and 2 bar for 20 min. Oligonucleotides were ordered from Sigma-Aldrich, Munich (desalted or RP1 purified). Cloned sequences were analyzed by Seqlab, Göttingen using Sanger-Sequencing and capillary electrophoresis.

### 6.1 Chemicals, instrumentation and consumables

**Table 6.1** List of employed chemicals and reagents.

Chemicals and reagents	Manufacturer
Adenine	Roth, Karlsruhe
Agar	Oxoid, Heidelberg
Agarose peqGold Universal	Peqlab, Erlangen
Ammonium sulfate	Roth, Karlsruhe
Ampicillin	Roth, Karlsruhe
Bromophenol blue	Roth, Karlsruhe
1-Butanol	Roth, Karlsruhe
Calcium chloride	Roth, Karlsruhe
Chloroform, p.a.	Roth, Karlsruhe
Deoxynucleotide triphosphate (dNTP)	Peqlab, Erlangen
Dimethyl sulfoxide	Peqlab, Erlangen
Dulbecco's Phosphate buffered Saline (PBS)	Life Technologies, USA
Dithiothreitol (DTT)	Roth, Karlsruhe
Ethanol absolute, p.a.	Merck, Darmstadt
Ethanol, denatured	VWR, Darmstadt
Ethidium bromide	Roth, Karlsruhe
Ethylenediaminetetraacetic acid (EDTA)	Roth, Karlsruhe
Glucose, water-free	Roth, Karlsruhe
Glycerol, p.a.	Roth, Karlsruhe
Hydrochloric acid (HCl)	Roth, Karlsruhe
Isopropanol, p.a.	VWR, Darmstadt

## Material

Leucine	Roth, Karlsruhe
MEM Amino acids, 50X	Sigma Aldrich
Magnesium chloride	Roth, Karlsruhe
Nicotinamide adenine dinucleotide (NAD)	Roth, Karlsruhe
PEG-8000	Roth, Karlsruhe
Sucrose	Roth, Karlsruhe
Sodium chloride (NaCl)	Roth, Karlsruhe
Sodium dodecyl sulfate (SDS), pellets	Roth, Karlsruhe
Tris	Roth, Karlsruhe
Tryptone	Oxoid, Heidelberg
Tryptophane	Roth, Karlsruhe
Triton-X-100	Roth, Karlsruhe
Uracil	Roth, Karlsruhe
Xylene cyanole	Roth, Karlsruhe
Yeast extract	Oxoid, Heidelberg
Yeast nitrogen base (w/o ammonium sulfate)	Difco
Yeast synthetic drop-out (-Ura/Leu/Trp)	Sigma Aldrich

**Table 6.2** List of utilized instruments.

<b>Instrument</b>	<b>Manufacturer</b>
Accuracy weighing machine	Acculab, USA
Centrifuges	Heraeus Christ, Osterode
CytoFlex S (flow cytometer)	Beckman-Coulter, Krefeld
Electroporator MicroPulser™	Bio-Rad, Munich
Flourescence Stereo Microscope	Leica Microsystems, Wetzlar
Fluorolog FL3-22	Horiba, Darmstadt
Gel documentation with UV screen	INTAS, Göttingen
Heating block	VWR, Darmstadt
Incubation Shaker Multitron	Infors AG, Bottmingen
Incubator	Heraeus Christ, Osterode
Infinite® M200 plate reader	Tecan Trading AG, Switzerland
Magnetic stirrer IKA RET basic	IKA, Staufen
Milli-Q® water purification system with RNase filter	EMD Millipore, France
NanoDrop® ND-1000 spectral photometer	Peqlab, Erlangen
Scale	Acculab, USA

## Material

T100™ Thermal Cycler	Bio-Rad, Munich
Titramax 100, 1000 (platform shakers)	Heidolph, Schwabach
Inkubator 1000	Heidolph, Schwabach
Thermomixer comfort	Eppendorf AG, Hamburg

**Table 6.3** List of kits and commercially available systems.

<b>Kits and commercially available systems</b>	<b>Manufacturer</b>
Frozen-EZ Yeast Transformation II	Zymo Research, USA
QIAquick Gel Extraction	QIAGEN, Hilden
QIAprep® Spin Miniprep Kit	QIAGEN, Hilden
QIAquick PCR Purification Kit	QIAGEN, Hilden

**Table 6.4** List of employed enzymes. Buffers were used as supplied and recommended by the manufacturer unless otherwise stated.

<b>Enzymes and proteins</b>	<b>Manufacturer</b>
Q5 High-Fidelity DNA Polymerase [2 U/μl]	New England Biolabs, USA
Phusion Polymerase	Thermo Fisher Scientific, Bremen
T4 DNA Ligase [400 U/μl]	New England Biolabs, USA
T4 DNA Polymerase [3 U/μl]	New England Biolabs, USA
T4 Polynucleotide kinase [10 U/μl]	New England Biolabs, USA
Taq DNA Polymerase [5 U/μl]	New England Biolabs, USA
Taq DNA Ligase [40 U/μl]	New England Biolabs, USA
T5 Exonuclease [10 U/μl]	New England Biolabs, USA
<b>Restriction endonucleases</b>	
AgeI-HF [20 U/μl]	New England Biolabs, USA
ClaI [10 U/μl]	New England Biolabs, USA
Dpn I [20 U/μl]	New England Biolabs, USA
NheI-HF [20 U/μl]	New England Biolabs, USA
SacI-HF [20 U/μl]	New England Biolabs, USA
SpeI-HF [20 U/μl]	New England Biolabs, USA
XbaI [20 U/μl]	New England Biolabs, USA
[20 U/μl]	New England Biolabs, USA

## Material

**Table 6.5** List of used DNA ladders.

Protein standards/ DNA ladders	Manufacturer
peqGold Ultra Low Range DNA ladder II	PeqLab, Erlangen
peqGold 1 kB DNA ladder	PeqLab, Erlangen

**Table 6.6** List of utilized consumables.

Consumables	Manufacturer
Cellstar® cell culture plates, 12-well/ 24-well	Greiner Bio-One, Austria
Cuvettes, acrylic	Sarstedt, Nürnberg
Dispenser tips	Greiner, Nürtingen
Micro Pulser Electroporation Cuvettes, 0.1 cm gap	Bio-Rad, Munich
Breathseal, gas permeable, sterile	Greiner Bio-One, Austria
96-well U bottom/ flat bottom	Greiner Bio-One, Austria
Nitrile Gloves	Starlab, Hamburg
Petri dishes, plastic	Greiner, Nürtingen
Pipette tips, plastic	Greiner, Nürtingen Starlab, Hamburg Nerbe Plus, Winsen
Pipettes, plastic	Nerbe Plus, Winsen
Polyethylene tubes	Sarstedt, Nürnberg
Reaction tubes (1.5 ml, 2 ml)	Greiner, Nürtingen
Syringes	Becton Dickinson, Heidelberg

## 6.2 Prokaryotic and eukaryotic cell strains

**Table 6.7** List of employed cell strains.

Cells	Genotype/ Description	Reference
<b>Bacterial strains</b>		
<i>E. coli</i> Top10	F- <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) Φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ( <i>araleu</i> )7697 <i>galU galK</i> <i>rpsL</i> (StrR) <i>endA1 nupG</i>	Invitrogen, USA
<b>Eukaryotic cell lines</b>		
RS453α	<i>mata ade2-1 trp1-1 can1-100 leu2-3 his3-1 ura3-52</i>	Hillen Lab
CEN.PK111-27B	<i>mata leu2-3 trp1-1</i>	Entian Lab

## 6.3 Buffers and solutions

Table 6.8 List of buffers and solutions and their composition.

Buffer/ solution	Ingredients	Concentration
5x ISO buffer	Tris-HCl pH 7.5	0.5 M
	MgCl <sub>2</sub>	50 mM
	Each dNTP	1 mM
	DTT	50 mM
	PEG-8000	25% (w/v)
	NAD	5 mM
STET buffer	Sucrose	8% (w/v)
	Tris-HCl, pH 8	50 mM
	EDTA	50 mM
	Triton-X-100	5% (v/v)
YPD	Yeast extract	1% (w/v)
	Peptone	2% (w/v)
	Glucose	2% (w/v)
	Agar for plates	1.8% (w/v)
SCD -Ura	50x MEM Amino acids	1x
	Adenine	12 µg/ml
	Ammonium sulfate	0.55% (w/v)
	YNB	0.2% (w/v)
	Glucose	2% (w/v)
	Agar for plates	1.8% (w/v)
SCD -Ura/Leu/Trp	Yeast synthetic drop-out (- Ura/Leu/Trp)	0.15% (w/v)
	Ammonium sulfate	0.55% (w/v)
	YNB	0.2% (w/v)
	Glucose	2% (w/v)
	Uracil, if needed	0.076 mg/ml
	Leucine, if needed	0.38 mg/ml
	Tryptophan, if needed	0.076 mg/ml
	Agar for plates	1.8% (w/v)
50x TAE	Tris	2 M
	Acetic acid	1 M
	EDTA	50 mM
	pH 8.3	
6x DNA loading dye	Tris-HCl pH 7.6	40 mM
	EDTA	1 mM
	Acetic acid	20 mM

## Material

	Glycerol Bromophenol blue Xylene cyanole	50% (v/v) Spatula point Spatula point
Ampicillin, stock solution	Ampicillin In 70% (v/v) EtOH	100 mg/ml
LB medium	Tryptone Yeast extract NaCl	1% (w/v) 0.5% (w/v) 1% (w/v)
LB-Amp plates	Agar Ampicillin	2% (w/v) 100 µg/ml
SOC-medium	Yeast Extract Tryptone NaCl KCl MgCl <sub>2</sub> MgSO <sub>4</sub> Glucose	0.5% 0.2% 10 mM 2.5 mM 10 mM 10 mM 20 mM

### 6.4 Oligonucleotides

List of oligonucleotides synthesized by Sigma Aldrich for cloning purposes.

**Table 6.9 Primers for Project I: Aptamers from literature**

# stock	Primer name	5'->3' sequence
2 fw	CS1-fw-Atra	CCGGTGGGCTAGCATGAGGCGGGGTAAAATTGCTCCGATAAAAAACGCAAAGTCCTAAA ATGG
2 rev	CS1-r-Atra	CTAGCCATTTTAGGACTTTGCGTTTTTATCGGAGCAATTTACCCCGCCTCATGCTAGC CCA
3 fw	CS1-fw-DMHBI	CCGGTGGGCTATTGCTGGAGGGGCGCCACATTCGGTGGTGGTTGGGTGCGGTGCG CGATAGCTCAAAATGG
3 rev	CS1-r-DMHBI	CTAGCCATTTTGAGCTATCGCCGACCGCACCCAACCACCACCGAAATGTGGCGCCCT CCAGCAATAGCCCA
5 fw	CS1-fw-MG*AgeI	CTCTTCACCGGTGGATCCCGACTGGCGAGAGCCAGGTAACGAATG
5 rev	CS1-r-MG*NheI	CTCTTCGCTAGCCATTTTAGGATCCATTCGTTACCTGGCTCTCGC
6 fw	CS1-fw-StrepT*AgeI	CTCTTCACCGGTGATCGCATTGGACTTCTGCCGCAAG
6 rev	CS1-r-StrepT*NheI	CTCTTCGCTAGCCATTTTGATCCGACCGTGGTGCCTTGCAGGAGCAAGTCCAAATG
7 fw	CS1-fw-TobD1*AgeI	CTCTTCACCGGTACTTGGTTAGGTAATGAG
7 rev	CS1-r-TobD1*NheI	CTCTTCGCTAGCCATTTACTCATTACCTAAACCAAGTACCGG
13 fw	CS1-fw-DIR	CCGGTGCGCCTTGAAGAGCCTGCTTCGGCAGCTGGTGAATGACAGCTATGGCGCTTA AAATGG
13 rev	CS1-r-DIR	CTAGCCATTTTAAGCGCCATAGCTGTCATTCACCAGCTGCCGAAGCAGGCTTTTCAAG GCGCA
14 fw	CS1-fw-TPP*AgeI	CTCTTCACCGGTATCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACC TG
14 rev	CS1-r-TPP*NheI	CTCTTCGCTAGCCATTTCCCTACGCTGGCATTATCCAGATCAGGTGATACGGGTATTT CTCAG
15 fw	CS1-fw-L-Arg	CCGGTCTGCAAGATAAACCGATGCTGGGCGATTCTCCTGAAGTAGGGGAAGAGTTGTA GAAATGG
15 rev	CS1-r-L-Arg	CTAGCCATTTTCTACAACCTTCCCTACTTCAGGAGAATCGCCAGCATCGGTTTATC TTGCAGA
16 fw	CS1-fw-Ade*AgeI	CTCTTCACCGGTACTCATATAACCTCAATAATATGGTTTGGGGTGTCTACCAGGAACC
16 rev	CS1-r-Ade*NheI	CTCTTCGCTAGCCATTTTACCATAATCAGGATTTTACGGTTCCTGGTAGACACCCTCA
17 fw	CS1-fw-Gua*AgeI	CTCTTCACCGGTACTCATATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC C

## Material

17 rev	CS1-r-Gua*NheI	CTCTTCGCTAGCCATTTTACCCATAGTCGGACATTTACGGTGCCCGGTAGAAACTGCG
18 fw	CS1-fw-TMRlg	CCGGTGGGACGCGTAAAGAGTGCTGCTTCGGCAGAGAGGTCTTCGCGTCCCTAAAATGG
18 rev	CS1-r-TMRlg	CTAGCCATTTTAGGGACGCGAAGACCTCTCTGCCGAAGCAGCACTCTTTACGCGTCCCA
19 fw	CS1-fw-TMRsh	CCGGTGGGACGAGAGAGCTGCTTCGGCAGAGAGGTCTCGTCCCTAAAATGG
19 rev	CS1-r-TMRsh	CTAGCCATTTTAGGGACGAGACCTCTCTGCCGAAGCAGCTCTCTCGTCCCA
20 fw	CS1-fw-HAP	CCGGTGCAGCCAACTAGCGAGAGCTTAAATCTCTGAGCCCAGAGGGTTCAGTGCTGCAAATGG
20 rev	CS1-r-HAP	CTAGCCATTTTGCAGCACTGAACCCTCTCGGGCTCAGAGATTTAAGCTCTCGCTAGTTGGCTGCA
21 fw	CS1-fw-Theo	CCGGTGGCGATACCAGCCGAAAGGCCCTTGGCAGCGTCAAATGG
21 rev	CS1-r-Theo	CTAGCCATTTTGACGCTGCCAAGGGCCTTTTCGGCTGGTATCGCCA
22 fw	CS1-fw-SRB	CCGGTGAACCTCGCTTCGGCGATGATGGAGAGGCGCAAGGTTAACCGCCTCAGGTTCAAATGG
22 rev	CS1-r-SRB	CTAGCCATTTTGAACCTGAGGCGGTTAACCTTGCCTCTCCATCATCGCCGAAAGCGAGGTTCA
23 fw	CS1-fw-TMRV1	CCGGTGAGAGAGCTGCTTCGGCAGAGAGGTCTCAAATGG
23 rev	CS1-r-TMRV1	CTAGCCATTTTGAGACCTCTCTGCCGAAGCAGCTCTCTCA
24 fw	CS1-fw-TMRV2	CCGGTGAGAGAGTGTCTGCTTCGGCAGAGAGGTCTCAAATGG
24 rev	CS1-r-TMRV2	CTAGCCATTTTGAGACCTCTCTGCCGAAGCAGCACTCTCTCA
25 fw	CS1-fw-DFX	CCGGTGAGGCTCCTGTGAAGCAACCGAATGGACTCAAAAATGG
25 rev	CS1-r-DFX	CTAGCCATTTTGTAGTCCATTCGGTTGCTTACAGGAGCCTCA
28-fw	CS1-fw-KanB(8-1-1)	CCGGTGGGAGCTCGGTACCGAATTCTCAAATGG
28-rev	CS1-r-KanB(8-1-1)	CTAGCCATTTTGAGAATTCGGTACCGAGCTCCCA

**Table 6.10 Primers for Project I: CFX *in vivo* screening**

# stock	Primer name	5'->3' sequence
42-fw	CS7-fw-HR-pRS453a	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGTGGGAGACGCAACTGAATGAA
42-rev	CS7-r-HR-pRS453a	CAAGAATTGGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTGTGACCGCGACTAGTTACGGA
55-fw	CS9-fw-2B-10	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGTACTGAATGAAGTTGGGATCAGAG
55-rev	CS9-r-2B-10	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTGTAGTTACGGATCGGTGTAACCTCGAC
56-fw	CS9-fw-2B-20	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTGTTGGGATCAGAGCGGATCTC
56-r	CS9-r-2B-20	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTTCGTGTAACCTCGACAGATGA
57-fw	CS9-fw-2B-30	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTGAGCGGATCTTCAATCTCTC
57-r	CS9-r-2B-30	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTTCGACCAGATGAATTGAGTTCC
58-fw	CS9-fw-6H-10	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGTACTGAATGAAAGCACGGCATTTG
58-r	CS9-r-6H-10	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTGTAGTTACGGACCTCCGTGTAAC
59-fw	CS9-fw-6H-20	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGCACGGCATTGAGATTCTC
59-r	CS9-r-6H-20	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCTCCGTGTAACCTCGAACCC
60-fw	CS9-fw-6H-30	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTTGAGATTCTCAATAGTCAAGGCTGAAG
60-r	CS9-r-6H-30	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTACTTCGAACCAGAATAACTTCAGCC
61-fw	CS9-fw-10A-10	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGTACTGAATGAACATAAGTGACGCG
61-r	CS9-r-10A-10	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTGTAGTTACGGATCGGTGTAACCTCCGTC
62-fw	CS9-fw-10A-20	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTGATAAGTGAACGCGACTCTATCTC
62-r	CS9-r-10A-20	GGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTTCGTGTAACCTCCGTCGCG
63-fw	CS9-fw-10A-30	ATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTGCGGACTCTATCTCCCTAAACTAG



## Material

### P<sub>GAL1</sub>

# stock	Primer name	5'→3' sequence
72-fw	CS11-fw-Gal1-prom	GCCAGCAACGCGGCCCGACCTCGACGCATGCTATATTGAAGTACGGATTAGAAGCCG
72-rev	CS11-r- Gal1-prom	CCATTAAAGGACAAGCTAACAGCGTTGTTGACCGGAATCC
87-rev	CS11-r-Gal1-l.BB	CAGTGAAAAGTTCTTCTCCTTTGCTAGCGAAGAGACCGGTAATCCGGGGTTTTTCTCC TTG
88-rev	CS11-r-Gal1-601	ATTGGGACAACCTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATCTTAAGAATCCGGGG TTTTTCTCCTTG

### pGLC modifications

#### P<sub>STE5/MRP7</sub>

# stock	Primer name	5'→3' sequence
99-fw	CS13-fw-mrp7prom	CAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGCCTCCCTTATCAATGAAA CATACG
99-rev	cS13-rev-mrp7prom	TAGAAGACAGTAGCTTCATCCGCGGGAATTCTTTGTAATTATTCTCACCTTAGGTTAAG CTTG
100-fw	CS13-fw-ste5prom	CAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGTAGAAGGCGTATTGCTCA ATAGT
100-rev	CS13-rev-ste5prom	TAGAAGACAGTAGCTTCATCCGCGGGAATTCTTTGTAATTTAAAAGTTGTTCCGCTG TATCCT

### Genomic integrations – insert preparation

#### NOR gate

# stock	Primer name	5'→3' sequence
76-fw	CS13-fw-HR-Ade2	CAGGCGCATAACATAAGTCACAAATATTGTCTTGCCTCGACGCATGCAACTTC
76-rev	CS13-rev-HR-Ade2	GGTTTAGTGTTTTCTTACCCAATTGTAGAGACTATGCATGCCGGTAGAGGTGTG
92-fw	CS13-fw-HR2-Ade	ATTTGATTGCATTTCTGCCAACAACCTTCGCCTTAAGTTGAACGGAGTCCGGAACCTA GCAGGCGCATAACATAAGTCACA
92-rev	CS13-r-HR2-Ade	TGGGCACCATTTACTAAAGAATTAGCAGTCATGATTGTGAGATCTGTTAACGGTTTAGT GTTTTCTTACCCAATTGTAGA
94-fw	CS13-fw-Ade100	ATTTGATTGCATTTCTGCCAACAAC
94-rev	CS13-r-ade100	TGGGCACCATTTACTAAAGAATTAGCAG

#### GLC

# stock	Primer name	5'→3' sequence
160-fw	CS18-fw-GEV1	GAAGCAGGCGGCAGAAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTCCCTCACT AAAGGGAACAAAAGC
161-fw	CS18-fw-GEV2	CATTACGAATGCACACGGTGTGGTGGGCCAGGTATTGTTAGCGGTTTGAAGCAGGC GGCAGAAGAAG
160-rev	CS18-r-GEV1	CTTTGTCGCTCTTCGCAATGTCAACAGTACCCTTAGTATATTCTGCCGCAAATTAAGC CTTCGAG
161-rev	CS18-r-GEV2	CCTTCATCTTCCACCCATGTCTCTTTGAGCAATAAAGCCGATAACAAAATCTTTGTGCG CTCTTCGCAATGTC
162-fw	CS18-fw-GEV-OL	ATGACCTGCTGCTGGAGATG
162-rev	CS18-r-GEV-OL	GTGGCCAAGTGGCTTTGGT

## Material

### 6.5 Plasmids

**Tables 6.13** List of plasmids generated and used in this work.

#### Plasmids for Project I: Aptamers from literature

# stock	Plasmid name	Description
1	I.601	pWHE601, adopted from Dr. Julia Weigand
2	I. BB	pWHE601*, substitution of <i>Afill</i> for <i>AgeI</i> , deletion of start codon
3	I. MG-Apt	Insertion of respective aptamer
4	I. StrepT-Apt	
5	I. TobD1-Apt	
7	I. Atra-Apt	
8	I. TPP-Apt	
9	I. DIR-Apt	
11	I. DMHBI-Apt	
12	I. L-Arg-Apt	
13	I. Ade-Apt	
14	I. Gua-Apt	
15	I. Neo-Apt	Adopted from Dr. Julia Weigand, M4 riboswitch
16	I. Tc-AN32sh-Apt	Adopted from Dr. Julia Weigand, AN32sh riboswitch
17	I. TMRlg-Apt	Insertion of respective aptamer
18	I. TMRsh1-Apt	
19	I. TMRsh2-Apt	
25	I. Theo-Apt	
26	I. HAP-Apt	
27	I. DFX-Apt	
28	I. TMRV1-Apt	
29	I. TMRV2-Apt	
30	I. SRB-Apt	
31	I. KanB(8-1-1)-Apt	

#### Plasmids for Project I: CFX *in vivo* screening

# stock	Plasmid name	Description
	V.pool-R6	The pools from SELEX cloned by homologous recombination in yeast.
	V.pool-R9	
	V.2B	Unique sequences obtained after <i>in vivo</i> screening that display minor riboswitching properties.
	V.6H	
	V.10A	
	V.12A	
--	V.2B1	Truncations of the indicated sequences. Refer to Figure 4.9 for the exact modifications. All stored in PCR-strips and numbered according to plasmid name.
-	V.2B1	
-	V.2B1	
-	V.2B1	
	V.2B1	
-	V.6H1	

## Material

-	V.6H1	
-	V.6H1	
-	V.6H1	
-	V.10A1	
56	V.6144	10A truncations: 10 nt from 5'-end
57	V.6244	10A truncations: 20 nt from 5'-end
58	V.6344	10A truncations: 30 nt from 5'-end
59	V.6370	10A truncations: 30 nt from 5'-end and 4 nt from 3'-end

### Plasmids for Project II: Tc-dimers

# stock	Plasmid name	Description
87	VIII.GFP3	Adopted from Dr. Julia Weigand and Lara Gorini
105	VIII.GFP7	
117	VIII.GFP15	Tc-dimers used to explore the ranges of basal expression and switching factor. Stem length of the 3'-Tc vary between 6-8 nt.
118	VIII.GFP16	
119	VIII.GFP17	
120	VIII.GFP18	
121	VIII.GFP19	
122	VIII.GFP20	
123	VIII.GFP21	
124	VIII.GFP22	
1st set	Plate 1	Plasmids are stored in a 96-well plate and named according to the plate coordinates
2nd set	Plate 2	Plasmids are stored in a 96-well plate and named according to the plate coordinates

### Plasmids for Project III: ROC'n'Ribo

#### 2 $\mu$ plasmid-based NOR gate expression

##### **P<sub>ADH1</sub>**

# stock	Plasmid name	Description
40	III.Neo	3'-Neo with 5'-(CAAA) spacer
40.1	III.NeoNeo	Neo-dimers
40.2	III.NeoTc	5'-TcNeo
40	III.Tc	3'-Tc with 5'-(CAAA) spacer
41.1	III.TcTc	Tc-dimers
41.2	III.TcNeo	5'-NeoTc

##### **P<sub>GAL1</sub>**

# stock	Plasmid name	Description
73	pRS303-gal1	<i>GAL1</i> promoter source
41.6	III.Tc1+C-Neo-gal1	5'-NeoTc with <i>GAL1</i> promoter instead of <i>ADH1</i> promoter
74	I.BB-gal1	pWHE601* with <i>GAL1</i> promoter instead of <i>ADH1</i> promoter
75	I.601-gal1	pWHE601 with <i>GAL1</i> promoter instead of <i>ADH1</i> promoter

## Material

### pGLC modifications

#### P<sub>STE5</sub>/MRP7

# stock	Plasmid name	Description
80	pSP-212-pMRP7-GEV	GEV driven by P <sub>MRP7</sub> without linker between GEV and mCherry
81	pSP-212-pSTE5-GEV	GEV driven by P <sub>STE5</sub> without linker between GEV and mCherry
141	III.GEVlinkerCherry	GEV driven by P <sub>TEF1</sub> with linker between GEV and mCherry, resulted in only a minor increase of mCherry fluorescence

### Genomic integrations – insert preparation

#### NOR gate

# stock	Plasmid name	Description
82	III.pJET-BB-gal1p	Subcloning of pWHE601*-gal into pJET1.2 for PCR amplification
83	III.pJET-601-gal1p	Subcloning of pWHE601-gal into pJET1.2 for PCR amplification
84	III.pJET-5'NTC-gal1p	Subcloning of 5'-NeoTc into pJET1.2 for PCR amplification

#### GLC

# stock	Plasmid name	Description
141	III.GEVlinkerCherry	GEV driven by P <sub>TEF1</sub> with linker between GEV and mCherry, resulted in only a minor increase of mCherry fluorescence. Used as template to amplify two halves of the expression cassette

### CRISPR

# stock	Plasmid name	Description
62	p426-SNR52p-gRNA.CAN1.Y-SUP4t	Addgene plasmid #43803 <sup>72</sup>
63	p414-TEF1p-Cas9-CYC1t	Addgene plasmid #83946 <sup>217</sup>
68	III.p425	Standard yeast shuttle vector, gift from group of Prof. Bertl
69	425-antiAde-gRNA	gRNA expression cassette for the indicated loci. Amplified from #62 in two parts and cloned by Gibson assembly into p425. The exchange of the gRNA was simultaneously performed.
70	425-antiUra-gRNA	
71	425-antiHis-gRNA	

### 6.5.1 Basic plasmids

#### 6.5.1.1 pWHE601, pHWE601\*

Vector pWHE601<sup>155</sup> was used as a positive control for all reporter gene assays in Project I and II and III. For selection in bacteria a  $\beta$ -lactamase gene is encoded conferring resistance against ampicillin. To select for the plasmid in yeast *URA3* is expressed that codes for orotidine-5'-phosphate (OMP) decarboxylase enabling Ura<sup>-</sup> strains to grow on synthetic media without uracil. The vector contains a 2 $\mu$  origin of replication (*ori*) for its replication in yeast cells. The expression cassette is comprised of an *ADH1* promoter and terminator and a *GFP+* reporter gene that is 5'-scarred with the *NheI* restriction site. Vector pWHE601\* represents a derivatives of pWHE601 with the following minor modifications. (i) The *AflII* restriction site was exchanged for the restriction site for *AgeI*, (ii) the start codon was deleted and substituted by CTCTTC now spacing *AgeI* and *NheI*. This vector was used as a negative control in all reporter gene assays and as a basis for introducing all aptamers in Project I via the two described restriction sites as well as the NOR and NOT gates from Project III.

## Material

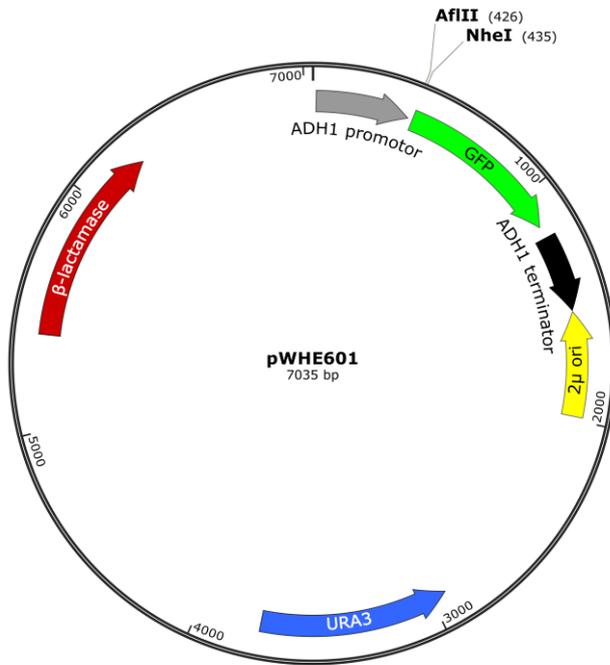


Figure 6.1 Plasmid map of pWHE601.

### 6.5.1.2 pGFP3

The vector pGFP3 constitutes the basis vector for all cloning in Project II and was adopted from Dr. Julia Weigand and Lara Gorini. Derived from pWHE601, it contains all features displayed in Figure 6.1. The vector map is displayed linearly and deprived of the redundant sequence homology to pWHE601 for clarity. The differences are described as follows. (i) The 5'-UTR from the transcription start site (TSS) to the beginning of the first riboswitch measures 41 nt (45 nt for pWHE601). It contains a *SacII* restriction site used for previous cloning experiments to introduce the tc-dimers. The dimers are separated by a 9 copies of a CAAA spacer. The Kozak sequences used here reads AAAAA-ATG and is immediately followed by the *GFP+* reporter gene. The highlighted sequences for stem 5' and stem 3' indicated the positions of randomization for the tc-dimers set 1 and 2.

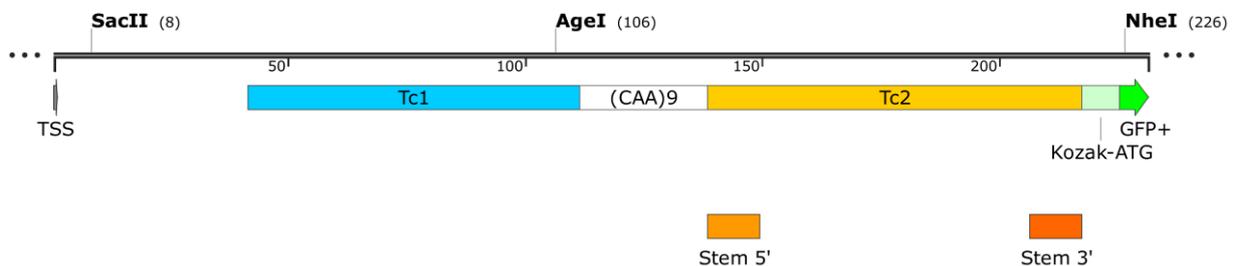


Figure 6.2 Take-out of the plasmid map of pGFP3 that is otherwise isogenic to pWHE601.

### 6.5.1.3 pGLC

Vector pGLC (GEV-Linker-mCherry) expresses the chimeric transcription factor (TF) GEV that was used in Project III. GEV is composed of a Gal4 DNA binding domain (Gal4 BD), the human estrogen receptor

## Material

(hER) and a virus particle 16 transcription activation domain (VP16) that is not shown in the map (compare Figure 4.21). The TF is fused to mCherry by a short linker promoting independent folding of both proteins. Transcription is driven by the *TEF1* promoter and terminated by the *CYC1* terminator. For selection in bacteria a  $\beta$ -lactamase gene is encoded conferring resistance against ampicillin. To select for the plasmid in yeast *LEU2* is expressed that codes for beta-isopropylmalate dehydrogenase enabling  $Leu^-$  strains to grow on synthetic media without leucine. The vector contains an autonomous replication sequence (ARS) and a centromeric region (CEN) for replication by mitosis in yeast cells.

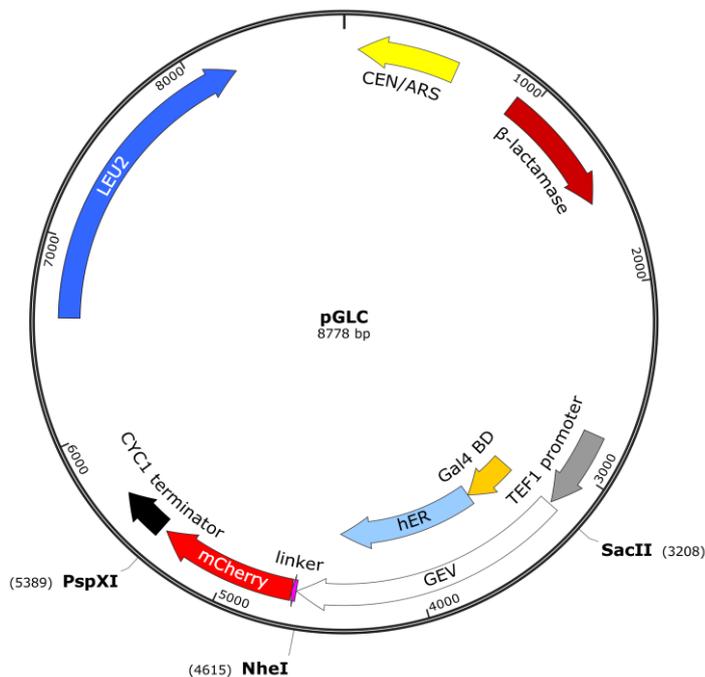


Figure 6.3 Plasmid map of pGLC.

### 6.5.1.4 p425-gRNA

Vector p425-gRNA is the host plasmid for expression of the gRNA targeting *Ura3*, *Leu2* and *His3*. The gRNA expression cassette was cloned from vector p426-P<sub>SNR52</sub>-gRNA.CAN1.Y-T<sub>SUP4</sub> (Addgene plasmid #43803)<sup>72</sup> and inserted into p425 changing *URA3* for *LEU2* (compare chapter 5.2.2.1). The gRNA expression cassette consists of a *SNR52* promoter to recruit RNA Pol III, the actual 20 nt of the gRNA, the structural gRNA recognized by Cas9 and the *SUP4* terminator. The gRNAs were substituted according to the targeted locus when cloned from the Addgene plasmid. Thus, three different p425-gRNA exist. For selection in bacteria a  $\beta$ -lactamase gene is encoded conferring resistance against ampicillin. To select for the plasmid in yeast *LEU2* is expressed that codes for beta-isopropylmalate dehydrogenase enabling  $Leu^-$  strains to grow on synthetic media without leucine. The vector contains a 2 $\mu$  origin of replication (*ori*) for its replication in yeast cells.

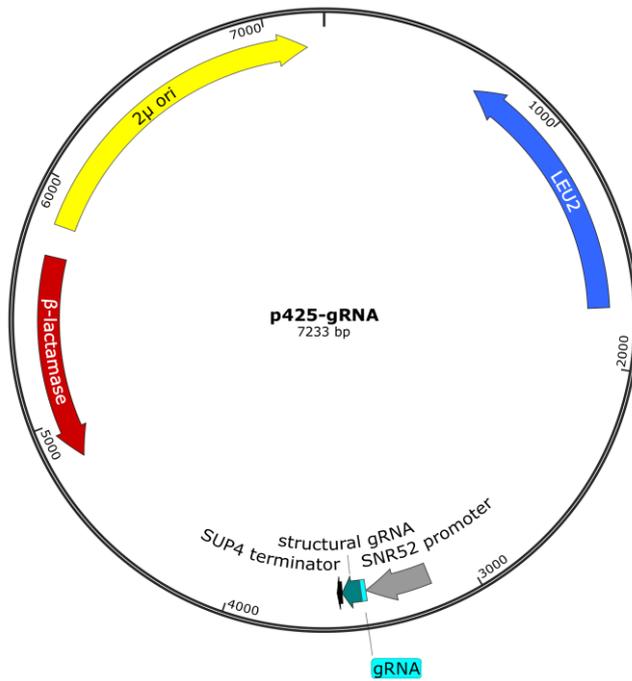


Figure 6.4 Plasmid map of p425-gRNA.

#### 6.5.1.5 p414-Cas9

Vector p414-Ca9 equals vector p414- $P_{TEF1}$ -Cas9- $T_{CYC1}$  (plasmid #83946)<sup>217</sup> purchased from Addgene. The vector expresses the *CAS9* gene coding for the endonuclease Cas9 recruited by the gRNA to introduce a double-strand break within the targeted locus. Expression is driven by *TEF1* promoter and terminated by *CYC1* terminator. For selection in bacteria a  $\beta$ -lactamase gene is encoded conferring resistance against ampicillin. To select for the plasmid in yeast *TRP1* is expressed that codes for phosphoribosylanthranilate isomerase enabling  $Trp^-$  strains to grow on synthetic media without tryptophane. The vector contains an autonomous replication sequence (ARS) and a centromeric region (CEN) for replication by mitosis in yeast cells.

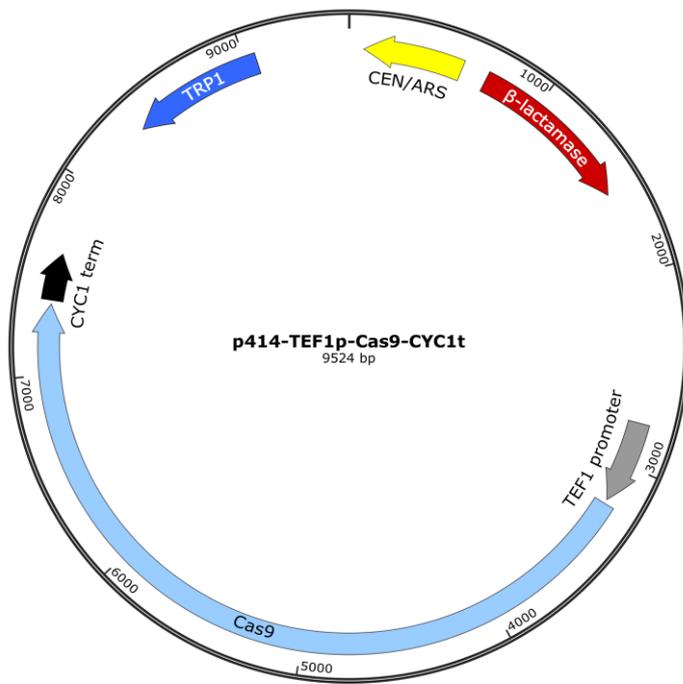


Figure 6.5 Plasmid map of p414-Cas9.

### 6.5.2 Construction of plasmids

All established constructs were precipitated with butanol after ligation and transformed into *E.coli* Top10. Then, the bacterial clones were tested in a CoPCR and sequenced to confirm construct integrity. Described cloning schemes should be regarded as examples for all similarly designed constructs.

#### 6.5.2.1 Project I: Aptamers from literature

Vector pWHE601\* was used for aptamer insertion. Aptamers were cloned with *AgeI* and *NheI* restriction sites. Therefore aptamers were either ordered as self-priming oligonucleotides that needed to be restricted after PCR with the corresponding enzymes or as “pre-cut” dsDNA hybrids that were annealed and 5'-phosphorylated and ligated. Every aptamer featured the Kozak-ATG (AAA-ATG) sequence that was thus introduced alongside the aptamer. The example shows the inserted theophylline aptamer cloned by primer hybridization, 5'-phosphorylation and subsequent ligation. Integration was confirmed by sequencing with GFP-rev.



Figure 6.6 Cloning scheme of various aptamers into pWHE601\*.

### 6.5.2.2 Project I: CFX *in vivo* screening

The cDNA pool from R6 and R9 of the SELEX was cloned by yeast homologous recombination. Vector pWHE601\* was prepared by linearization using *AgeI* and *NheI* restriction sites. The pool was PCR-amplified with fwd and rev primers featuring 40 nt 5'-overhangs homologous to sequence stretches upstream of *AgeI* or downstream of *NheI*, respectively (compare Figure 5.1). The Kozak-ATG sequence was introduced as part of the rev primer. 5'- and 3'-randomization indicate the in total  $N_{64}$  randomized nucleotides. The forced loop describes a sequence motif promoting hairpin formation.

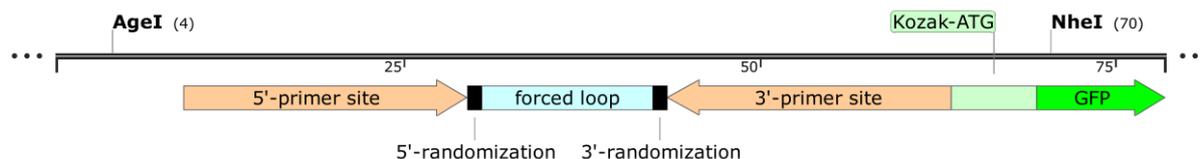


Figure 6.7 Cloning scheme of the homologous recombination of the CFX pool with restricted pWHE601\*.

### 6.5.2.3 Project II: Tc-dimers

For the generation of the different Tc-dimers, only the closing stem of Tc2 was randomized. Therefore restriction sites *AgeI* and *NheI* of vector pGFP3 were used. The core of Tc2 was used as template for aptamer amplification. The fwd primer introduced the 5'-stem, the (CAA)<sub>9</sub>-spacer and the *AgeI* restriction site. The rev primer was 5'-flanked by the 3'-stem, the Kozak-ATG sequence and the *NheI* restriction site. Constructs GFP15-22 were cloned either by amplification from the Tc2 core or by ligation assembly in four parts as indicated by the a, b pairs in the primers list.

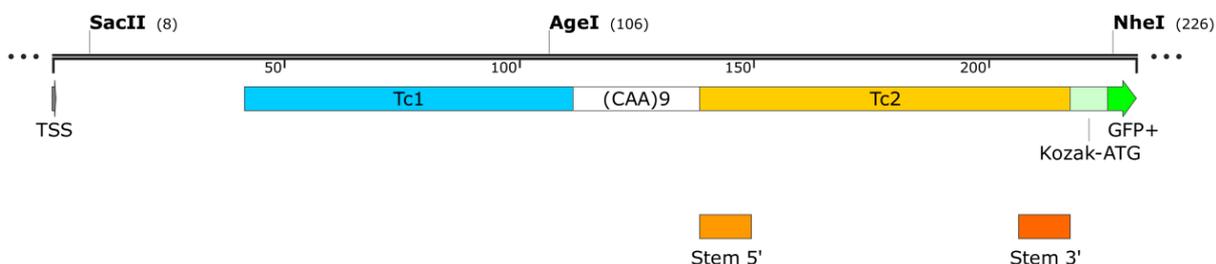


Figure 6.8 Cloning scheme of the Tc-dimers.

### 6.5.2.4 Project III: ROC'n'Ribo

#### 2 $\mu$ plasmid-based NOR gate expression

##### *P<sub>ADH1</sub>*

The NOR gate 5'-NeoTc was generated using pWHE601\* as backbone and *AgeI* and *NheI* restriction sites to insert the Tc aptamer with the CAAA-spacer. The Neo aptamer was cloned by Gibson assembly using *AgeI* site to linearize pWHE601\*-Tc. Primers to amplify the Neo aptamer from its original plasmid were 5'-flanked with sequences homologous to the upstream sequence of *AgeI* and to the CAAA-spacer. The other NOR gate and both NOR gates were prepared analogously.

## Material



Figure 6.9 Cloning scheme of the 5'-NeoTc NOR gate.

### $P_{GAL1}$

The *ADH1* promoter of all NOR and NOT gates as well as pWHE601 and pWHE601\* was substituted for  $P_{GAL1}$ . The target plasmid was linearized with *AgeI* and a Gibson assembly used to fuse the amplified  $P_{GAL1}$  with the linearized vector. The  $P_{GAL1}$  was amplified from another plasmid with primers that had 40 nt 5'-flanking sequences homologous to the upstream sequence of the  $P_{ADH1}$  to exclude the former promoter and to the downstream sequence of the *AgeI* site. Substitution increased the 5'-UTR length upstream of the riboswitch to 75 nt and introduced a second *AgeI* site.

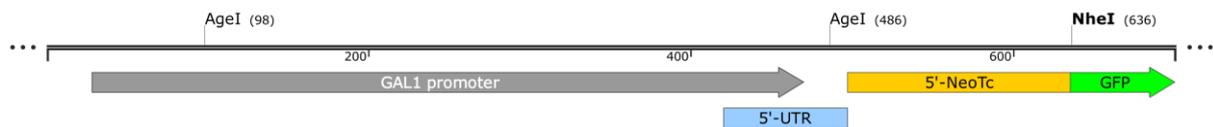


Figure 6.10 Exchange of the *ADH1* promoter for *GAL1* promoter 5'-NeoTc NOR gate.

## pGLC modifications

### $P_{STE5/MRP7}$

The *TEF1* promoter was substituted for  $P_{STE5}$  or  $MRP7$  in pGLC. pGLC was linearized with *SacII* and the new promoters introduced by homologous recombination in yeast or Gibson assembly. Therefore both promoters were amplified from the genomic DNA of yeast strain RS453 $\alpha$  and 5'-flanked with sequences exhibiting 40nt homology to the upstream sequence of the *TEF1* promoter and 40 nt to *GEV*. Thus  $P_{TEF1}$  was eliminated and the new promoter introduced.



Figure 6.11 Exchange of the *TEF1* promoter for *STE5* or *MRP7* promoter.

## Genomic integrations – insert preparation

### NOR gate

All gates and pWHE601 and pWHE601\* were subcloned into a pJET1.2 vector after they had been amplified from their pWHE601\*-backbones with primers attaching 100 nt homology to the *ade2* locus up- and downstream of the integration site targeted by the gRNA. They were amplified from pJET1.2 to yield high concentrations for the genomic integration. The amplicon spans ~1900 nt in case of the insert 5'-NeoTc.

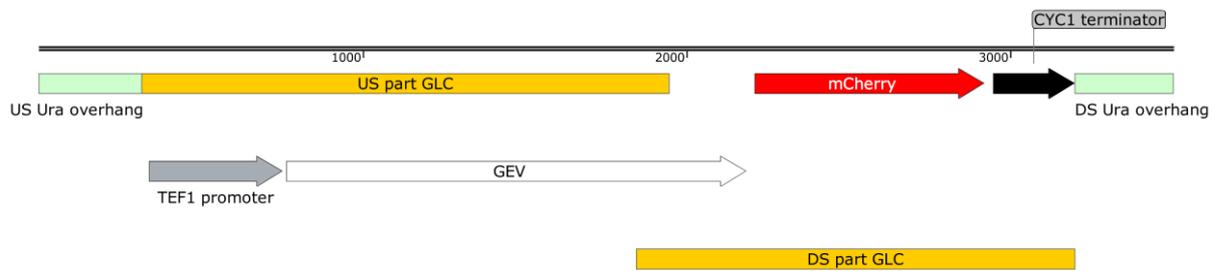
## Material



**Figure 6.12** Cloning scheme of the insert 5'-NeoTc as prepared for genomic integration.

## GLC

The preparation of the genomic integration of GLC was conducted by dividing the fusion protein in two parts (US/DS part GLC) that were separately amplified from pGLC and exhibited 100 nt homology between them and 100 nt homology to the integration site (US/DS Ura). The 100 nt homology to the integration site were attached in two subsequent PCR reactions with two primer sets.



**Figure 6.13** Cloning scheme of the GLC inserts as prepared for genomic integration.

## 7 Appendix

### 7.1 Abbreviations

**Table 7.1** List of abbreviations.

Abbreviations		Abbreviations	
% (v/v)	% (volume/volume)	PCR	polymerase chain reaction
% (wv)	% (weight/volume)	Prom	promoter
ADH1	Alcohol dehydrogenase 1	rev/r	reverse
Amp	ampicillin	RNA	ribonucleic acid
bp	base pair	SCD	Synthetic complete dextrose
CYC1	Cytochrome C1	SDS	Sodium dodecyl sulfate
DMSO	dimethylsulfoxide	shRNA	small hairpin RNA
dNTP	deoxynucleoside triphosphate	siRNA	small interfering RNA
<i>et al.</i>	et alii	STET5	STERile
EtOH	ethanol	TBS	tris buffered saline
fwd/f	forward	tc	tetracycline
LB	lysogeny broth	TEF1	Transcription elongation factor EF-1 alpha
MRP7	Mitochondrial Ribosomal Protein 7	UTR	untranslated region
mRNA	messenger RNA		
neo	neomycin		
nt	nucleotides		
PBS	phosphate buffered saline		

## 7.2 Units

**Table 7.2** List of units.

Units	
°C	degree Celsius
Da	Dalton
g	gram
h	hours
l	liter
M	molar
min	minutes
rpm	rounds per minute
sec	seconds
U	unit
V	volt
v/v	volume per volume
w/v	weight per volume

## 7.3 Prefixes

**Table 7.3** List of used prefixes.

Dimensions	
k	kilo ( $10^3$ )
m	milli ( $10^{-3}$ )
$\mu$	micro ( $10^{-6}$ )
n	nano ( $10^{-9}$ )
p	pico ( $10^{-12}$ )

## 7.4 Nucleobases

**Table 7.4** Abbreviations for nucleobases.

Nucleobases	
A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil
N	A, C, G, T, U

## References

### 8 References

- (1) Alberts, B. (2011) EDITORIAL A Grand Challenge in Biology. *Science* 333, 1248.
- (2) Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014) A brief history of synthetic biology. *Nature Reviews Microbiology* 12, 381–390.
- (3) McAdams, H. H., and Arkin, A. (2000) Gene regulation: Towards a circuit engineering discipline. *Current Biology* 10, 318–320.
- (4) Collins, J. J., Gardner, T. S., and Cantor, C. R. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342.
- (5) Slusarczyk, A. L., Lin, A., and Weiss, R. (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nature Reviews Genetics* 13, 406–420.
- (6) Elowitz, M. B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338.
- (7) Kærn, M., Blake, W. J., and Collins, J. J. (2003) The Engineering of Gene Regulatory Networks. *Annual Review of Biomedical Engineering* 5, 179–206.
- (8) Isaacs, F. J., Hasty, J., Cantor, C. R., and Collins, J. J. (2003) Prediction and measurement of an autoregulatory genetic module. *Proceedings of the National Academy of Sciences of the United States of America* 100, 7714–7719.
- (9) Hasty, J., McMillen, D., Isaacs, F., and Collins, J. J. (2001) Computational studies of gene regulatory networks: in numero molecular biology. *Nature Reviews Genetics* 2, 268–279.
- (10) Becskei, A., and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature* 405, 590–593.
- (11) Basu, R. W. and S. (2002) The device physics of cellular logic gates. *NSC-1: The First Workshop of Non-Silicon ...* 158, 39–41.
- (12) Ball, P. (2004) Synthetic biology: Starting from scratch. *Nature* 431, 624–626.
- (13) Ferber, D. (2004) SYNTHETIC BIOLOGY: Microbes Made to Order. *Science* 303, 158–161.
- (14) Endy, D. (2005) Foundations for engineering biology. *Nature* 438, 449–53.
- (15) Shetty, R. P., Endy, D., and Knight, T. F. (2008) Engineering BioBrick vectors from BioBrick parts. *Journal of biological engineering* 2, 5.
- (16) Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J. (2004) Engineered riboregulators enable post-transcriptional control of gene expression 841–847.
- (17) Bayer, T. S., and Smolke, C. D. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature biotechnology* 23, 337–43.
- (18) You, L., Cox, R. S., Weiss, R., and Arnold, F. H. (2004) Programmed population control by cell–cell communication and regulated killing. *Nature* 428, 868–871.
- (19) Danino, T., Mondragón-Palomino, O., Tsimring, L., and Hasty, J. (2010) A synchronized quorum of genetic clocks. *Nature* 463, 326–330.
- (20) Anderson, J. C., Voigt, C. A., and Arkin, A. P. (2007) Environmental signal integration by a modular AND gate. *Molecular systems biology* 3, 133.
- (21) Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievens, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S., and Newman, J. D. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–32.
- (22) Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943.
- (23) Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology* 21, 796–802.
- (24) Win, M. N., and Smolke, C. D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science (New York, N.Y.)* 322, 456–60.
- (25) Callura, J. M., Cantor, C. R., and Collins, J. J. (2012) Genetic switchboard for synthetic biology applications. *Proceedings of the National Academy of Sciences of the United States of America* 109, 5850–5.
- (26) Brophy, J. A. N., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nature methods* 11, 508–20.
- (27) Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Iii, C. A. H., Smith, H. O., and America, N. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases 12–16.
- (28) Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J. a. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science (New York, N.Y.)* 329, 1355–8.
- (29) Qi, L. S., Larson, M. H., Gilbert, L. a, Doudna, J. a, Weissman, J. S., Arkin, A. P., and Lim, W. a. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–83.

## References

- (30) Chin, J. W. (2014) Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annual Review of Biochemistry* 83, 379–408.
- (31) Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science (New York, N.Y.)* 319, 1215–20.
- (32) Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwertzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S., and Boeke, J. D. (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* 477, 471–476.
- (33) Berens, C., and Suess, B. (2015) Riboswitch engineering - making the all-important second and third steps. *Current opinion in biotechnology* 31C, 10–15.
- (34) Kwok, R. (2010) Five hard truths for synthetic biology. *Nature* 463, 288–290.
- (35) Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnology Journal* 7, 856–866.
- (36) Lou, C., Stanton, B., Chen, Y.-J., Munsy, B., and Voigt, C. a. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nature Biotechnology* 30.
- (37) Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. a, and Arkin, A. P. (2012) RNA processing enables predictable programming of gene expression. *Nature biotechnology* 30, 1002–6.
- (38) Temme, K., Zhao, D., and Voigt, C. A. (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proceedings of the National Academy of Sciences* 109, 7085–7090.
- (39) Mortimer, R. K. (2000) Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome research* 10, 403–9.
- (40) Sicard, D., and Legras, J.-L. (2011) Bread, beer and wine: Yeast domestication in the *Saccharomyces sensu stricto* complex. *Comptes Rendus Biologies* 334, 229–236.
- (41) Annaluru, N., Ramalingam, S., and Chandrasegaran, S. (2015) Rewriting the blueprint of life by synthetic genomics and genome engineering. *Genome Biology* 16, 125.
- (42) K.Jensen, M., and Keasling, J. D. (2014) Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS Yeast Research* n/a-n/a.
- (43) Lee, M. E., Deloache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly.
- (44) Redden, H., Morse, N., and Alper, H. S. (2014) The synthetic biology toolbox for tuning gene expression in yeast. *FEMS yeast research*.
- (45) Jakociunas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., Skjoedt, M. L., Nielsen, A. T., Borodina, I., Jensen, M. K., and Keasling, J. D. (2015) CasEMBLR: Cas9-facilitated multi-loci genomic integration of in vivo assembled DNA parts in *Saccharomyces cerevisiae*. *ACS Synthetic Biology* 150317152158005.
- (46) Jessop-Fabre, M. M., Jakočiūnas, T., Stovicek, V., Dai, Z., Jensen, M. K., Keasling, J. D., and Borodina, I. (2016) EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. *Biotechnology Journal* 11, 1110–1117.
- (47) Blount, B. A., Weenink, T., and Ellis, T. (2012) Construction of synthetic regulatory networks in yeast. *FEBS Letters* 586, 2112–2121.
- (48) Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014) A brief history of synthetic biology. *Nature Reviews Microbiology* 12, 381–390.
- (49) Karlsson, M., and Weber, W. (2012) Therapeutic synthetic gene networks. *Current Opinion in Biotechnology* 23, 703–711.
- (50) Hong, K.-K., and Nielsen, J. (2012) Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cellular and molecular life sciences : CMLS* 69, 2671–90.
- (51) Buchholz, K., and Collins, J. (2013) The roots--a short history of industrial microbiology and biotechnology. *Applied microbiology and biotechnology* 97, 3747–62.
- (52) Redden, H., Morse, N., and Alper, H. S. (2014) The synthetic biology toolbox for tuning gene expression in yeast. *FEMS yeast research* 15, 1–10.
- (53) Siddiqui, M. S., Thodey, K., Trenchard, I., and Smolke, C. D. (2012) Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Research* 12, 144–170.
- (54) Blazeck, J., Garg, R., Reed, B., and Alper, H. S. (2012) Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnology and Bioengineering* 109, 2884–2895.
- (55) Redden, H., and Alper, H. S. (2015) The development and characterization of synthetic minimal yeast promoters. *Nature Communications* 6, 7810.
- (56) Curran, K. A., Karim, A. S., Gupta, A., and Alper, H. S. (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metabolic Engineering* 19, 88–97.
- (57) Mclsaac, R. S., Gibney, P. A., Chandran, S. S., Benjamin, K. R., and Botstein, D. (2014) Synthetic biology tools for programming gene expression without nutritional perturbations in *Saccharomyces cerevisiae*. *Nucleic Acids Research* 42, 1–8.
- (58) Mclsaac, R. S., Oakes, B. L., Wang, X., Dummit, K. a, Botstein, D., and Noyes, M. B. (2013) Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic acids research* 41, e57.

## References

- (59) Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., Tsai, J. C., Weissman, J. S., Dueber, J. E., Qi, L. S., and Lim, W. A. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350.
- (60) Purcell, O., Peccoud, J., and Lu, T. K. (2014) Rule-based design of synthetic transcription factors in eukaryotes. *ACS synthetic biology* 3, 737–44.
- (61) Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N. C., Joung, J. K., and Collins, J. J. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell* 150, 647–658.
- (62) Wiedenheft, B., Sternberg, S. H., and Doudna, J. a. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–8.
- (63) Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–51.
- (64) Farzadfard, F., Perli, S. D., and Lu, T. K. (2013) Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS synthetic biology* 2, 604–13.
- (65) Breaker, R. R. (2012) Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology* 4.
- (66) Galloway, K. E., Franco, E., and Smolke, C. D. (2013) Dynamically reshaping signaling networks to program cell fate via genetic controllers. *Science (New York, N.Y.)* 341, 1235005.
- (67) Michener, J. K., Thodey, K., Liang, J. C., and Smolke, C. D. (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metabolic engineering* 14, 212–22.
- (68) Michener, J. K., and Smolke, C. D. (2012) High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metabolic engineering* 14, 306–16.
- (69) Crook, N. C., Schmitz, A. C., and Alper, H. S. (2014) Optimization of a yeast RNA interference system for controlling gene expression and enabling rapid metabolic engineering. *ACS synthetic biology* 3, 307–13.
- (70) Blazeck, J., Miller, J., Pan, A., Gengler, J., Holden, C., Jamoussi, M., and Alper, H. S. (2014) Metabolic engineering of *Saccharomyces cerevisiae* for itaconic acid production. *Applied Microbiology and Biotechnology* 98, 8155–8164.
- (71) DiCarlo, J. E., Conley, A. J., Penttilä, M., Jäntti, J., Wang, H. H., and Church, G. M. (2013) Yeast oligo-mediated genome engineering (YOGE). *ACS synthetic biology* 2, 741–9.
- (72) DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., and Church, G. M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research* 41, 4336–4343.
- (73) Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S. J., Kristensen, M., Pedersen, L. E., Jensen, M. K., and Keasling, J. D. (2015) Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metabolic Engineering* 28, 213–222.
- (74) Liu, F., Banta, S., and Chen, W. (2013) Functional assembly of a multi-enzyme methanol oxidation cascade on a surface-displayed trifunctional scaffold for enhanced NADH production. *Chemical communications (Cambridge, England)* 49, 3766–8.
- (75) Lee, H., DeLoache, W. C., and Dueber, J. E. (2012) Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering* 14, 242–251.
- (76) Chappell, J., Watters, K. E., Takahashi, M. K., and Lucks, J. B. (2015) A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future. *Current Opinion in Chemical Biology* 28, 47–56.
- (77) McKeague, M., Wong, R. S., and Smolke, C. D. (2016) Opportunities in the design and application of RNA for gene expression control. *Nucleic acids research* 44, 2987–99.
- (78) Berens, C., Groher, F., and Suess, B. (2015) RNA aptamers as genetic control devices : The potential of riboswitches as synthetic elements for regulating gene expression 246–257.
- (79) Qi, L., Lucks, J. B., Liu, C. C., Mutalik, V. K., and Arkin, A. P. (2012) Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic acids research* 40, 5775–86.
- (80) Chappell, J., Takahashi, M. K., Meyer, S., Loughrey, D., Watters, K. E., and Lucks, J. (2013) The centrality of RNA for engineering gene expression. *Biotechnology Journal* 8, 1379–1395.
- (81) Ge, P., and Zhang, S. (2015) Computational analysis of RNA structures with chemical probing data. *Methods (San Diego, Calif.)* 79–80, 60–6.
- (82) Cordero, P., Lucks, J. B., and Das, R. (2012) An RNA Mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics* 28, 3006–3008.
- (83) Wachsmuth, M., Domin, G., Lorenz, R., Serfling, R., Findeiß, S., Stadler, P. F., and Mörl, M. (2015) Design criteria for synthetic riboswitches acting on transcription. *RNA biology* 12, 221–31.
- (84) Borujeni, A. E., Mishler, D. M., Wang, J., Huso, W., and Salis, M. (2016) NAR Breakthrough Article Automated physics-based design of synthetic riboswitches from diverse RNA aptamers 44, 1–13.
- (85) Chang, A. L., Wolf, J. J., and Smolke, C. D. (2012) Synthetic RNA switches as a tool for temporal and spatial control over gene expression. *Current opinion in biotechnology* 23, 679–88.
- (86) Etzel, M., and Mörl, M. (2017) Synthetic Riboswitches – from Plug and Pray towards Plug and Play. *Biochemistry* acs.biochem.6b01218.
- (87) Groher, F., and Suess, B. (2014) Synthetic riboswitches - A tool comes of age. *Biochimica et biophysica acta* 1839, 964–973.
- (88) Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J. (2004) Engineered riboregulators

## References

- enable post-transcriptional control of gene expression. *Nature biotechnology* 22, 841–7.
- (89) Mizuno, T., Chou, M. Y., and Inouye, M. (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proceedings of the National Academy of Sciences of the United States of America* 81, 1966–70.
- (90) Gottesman, S., and Storz, G. (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor perspectives in biology* 3, 1–16.
- (91) Thomason, M. K., and Storz, G. (2010) Bacterial antisense RNAs: how many are there, and what are they doing? *Annual review of genetics* 44, 167–88.
- (92) Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: De-novo-designed regulators of gene expression. *Cell* 159, 925–939.
- (93) Rodrigo, G., Landrain, T. E., and Jaramillo, A. (2012) De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells 15271–15276.
- (94) Zhang, D. Y., Turberfield, A. J., Yurke, B., and Winfree, E. (2007) Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science* 318, 1121–1125.
- (95) Zadeh, J. N., Wolfe, B. R., and Pierce, N. A. (2011) Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of Computational Chemistry* 32, 439–452.
- (96) Chappell, J., Takahashi, M. K., and Lucks, J. B. (2015) Creating small transcription activating RNAs. *Nature Chemical Biology* 11, 214–220.
- (97) Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* 41, 7429–7437.
- (98) Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., and Sato, M. (2015) CRISPR-Cas9-based Photoactivatable Transcription System. *Chemistry & Biology* 22, 169–174.
- (99) Polstein, L. R., and Gersbach, C. A. (2015) A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology* 11, 198–200.
- (100) Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., and Huang, B. (2013) Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* 155, 1479–1491.
- (101) Winkler, W., Nahvi, A., and Breaker, R. R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419, 952–6.
- (102) Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111, 747–56.
- (103) Serganov, A., and Nudler, E. (2013) A decade of riboswitches. *Cell* 152, 17–24.
- (104) Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E. A. (2012) Riboswitch control of Rho-dependent transcription termination. *Proceedings of the National Academy of Sciences of the United States of America* 109, 5376–81.
- (105) Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428, 281–286.
- (106) Collins, J. A., Irnov, I., Baker, S., and Winkler, W. C. (2007) Mechanism of mRNA destabilization by the glmS ribozyme. *Genes & development* 21, 3356–68.
- (107) Watson, P. Y., and Fedor, M. J. (2011) The glmS riboswitch integrates signals from activating and inhibitory metabolites in vivo. *Nature structural & molecular biology* 18, 359–63.
- (108) Chen, A. G. Y., Sudarsan, N., and Breaker, R. R. (2011) Mechanism for gene control by a natural allosteric group I ribozyme. *RNA* 17, 1967–1972.
- (109) Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009) A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in *Listeria monocytogenes*. *Cell* 139, 770–779.
- (110) Cheah, M. T., Wachter, A., Sudarsan, N., and Breaker, R. R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447, 497–500.
- (111) Croft, M. T., Moulin, M., Webb, M. E., and Smith, A. G. (2007) Thiamine biosynthesis in algae is regulated by riboswitches. *Proceedings of the National Academy of Sciences* 104, 20770–20775.
- (112) Bocobza, S., Adato, A., Mandel, T., Shapira, M., Nudler, E., and Aharoni, A. (2007) Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes & Development* 21, 2874–2879.
- (113) Peselis, A., and Serganov, A. (2014) Themes and variations in riboswitch structure and function. *Biochimica et biophysica acta* 1839, 908–918.
- (114) Ren, A., Rajashankar, K. R., and Patel, D. J. (2012) Fluoride ion encapsulation by Mg<sup>2+</sup> ions and phosphates in a fluoride riboswitch. *Nature* 486, 85–9.
- (115) Gilbert, S. D., Rambo, R. P., Van Tyne, D., and Batey, R. T. (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature Structural & Molecular Biology* 15, 177–182.
- (116) Batey, R. T., Gilbert, S. D., and Montagne, R. K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432, 411–415.
- (117) Edwards, T. E., and Ferré-D'Amaré, A. R. (2006) Crystal Structures of the Thi-Box Riboswitch Bound to Thiamine

## References

- Pyrophosphate Analogs Reveal Adaptive RNA-Small Molecule Recognition. *Structure* 14, 1459–1468.
- (118) Trausch, J. J., Ceres, P., Reyes, F. E., and Batey, R. T. (2011) The Structure of a Tetrahydrofolate-Sensing Riboswitch Reveals Two Ligand Binding Sites in a Single Aptamer. *Structure* 19, 1413–1423.
- (119) Montange, R. K., and Batey, R. T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* 441, 1172–1175.
- (120) Jenison, R. D., Gill, S. C., Pardi, a, and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science (New York, N.Y.)* 263, 1425–9.
- (121) Desai, S. K., and Gallivan, J. P. (2004) 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.
- (122) and, S. T., and Gallivan\*, J. P. (2007) Guiding Bacteria with Small Molecules and RNA.
- (123) Topp, S., Reynoso, C. M. K., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., Shen, A., Puri, A. W., Komeili, A., Bertozzi, C. R., Scott, J. R., and Gallivan, J. P. (2010) Synthetic riboswitches that induce gene expression in diverse bacterial species. *Applied and environmental microbiology* 76, 7881–4.
- (124) Rudolph, M. M., Vockenhuber, M.-P., and Suess, B. (2013) Synthetic riboswitches for the conditional control of gene expression in *Streptomyces coelicolor*. *Microbiology (Reading, England)* 159, 1416–22.
- (125) Suess, B., Fink, B., Berens, C., Stentz, R., and Hillen, W. (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic acids research* 32, 1610–4.
- (126) Nomura, Y., and Yokobayashi, Y. (2007) Reengineering a natural riboswitch by dual genetic selection. *Journal of the American Chemical Society* 129, 13814–5.
- (127) Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F., and Mörl, M. (2013) De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic acids research* 41, 2541–51.
- (128) Ceres, P., Trausch, J. J., and Batey, R. T. (2013) Engineering modular “ON” RNA switches using biological components. *Nucleic acids research* 41, 10449–61.
- (129) Ceres, P., Garst, A. D., Marcano-Velázquez, J. G., and Batey, R. T. (2013) Modularity of Select Riboswitch Expression Platforms Enables Facile Engineering of Novel Genetic Regulatory Devices. *ACS Synthetic Biology* 2, 463–472.
- (130) Robinson, C. J., Vincent, H. A., Wu, M.-C., Lowe, P. T., Dunstan, M. S., Leys, D., and Micklefield, J. (2014) Modular Riboswitch Toolsets for Synthetic Genetic Control in Diverse Bacterial Species. *Journal of the American Chemical Society* 136, 10615–10624.
- (131) Wieland, M., and Hartig, J. S. (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angewandte Chemie (International ed. in English)* 47, 2604–7.
- (132) Saragliadis, A., and Hartig, J. S. (2013) Ribozyme-Based Transfer RNA Switches for Post-transcriptional Control of Amino Acid Identity in Protein Synthesis. *Journal of the American Chemical Society* 135, 8222–6.
- (133) Berschneider, B., Wieland, M., Rubini, M., and Hartig, J. S. (2009) Small-molecule-dependent regulation of transfer RNA in bacteria. *Angewandte Chemie (International ed. in English)* 48, 7564–7.
- (134) Wieland, M., Berschneider, B., Erlacher, M. D., and Hartig, J. S. (2010) Aptazyme-mediated regulation of 16S ribosomal RNA. *Chemistry & biology* 17, 236–42.
- (135) Klausner, B., and Hartig, J. S. (2013) An engineered small RNA-mediated genetic switch based on a ribozyme expression platform. *Nucleic acids research* 41, 5542–52.
- (136) Werstuck, G. (1998) Controlling Gene Expression in Living Cells Through Small Molecule-RNA Interactions. *Science* 282, 296–298.
- (137) Grate, D., and Wilson, C. (2001) Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer-ligand complex. *Bioorganic & medicinal chemistry* 9, 2565–70.
- (138) Weigand, J. E., Sanchez, M., Gunnesch, E., Zeiher, S., Schroeder, R., and Suess, B. (2008) Screening for engineered neomycin riboswitches that control translation initiation 89–97.
- (139) Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E. G., and Suess, B. (2003) Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular Microbiology* 49, 1627–1637.
- (140) Weigand, J. E., and Suess, B. (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic acids research* 35, 4179–85.
- (141) Kim, D.-S., Gusti, V., Pillai, S. G., and Gaur, R. K. (2005) An artificial riboswitch for controlling pre-mRNA splicing. *RNA (New York, N.Y.)* 11, 1667–77.
- (142) Kim, D.-S., Gusti, V., Dery, K. J., and Gaur, R. K. (2008) Ligand-induced sequestering of branchpoint sequence allows conditional control of splicing. *BMC molecular biology* 9, 23.
- (143) Nomura, Y., Zhou, L., Miu, A., and Yokobayashi, Y. (2013) Controlling Mammalian Gene Expression by Allosteric Hepatitis Delta Virus Ribozymes. *ACS synthetic biology*.
- (144) Felletti, M., Stifel, J., Wurmthaler, L. A., Geiger, S., and Hartig, J. S. (2016) Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nature Communications* 7, 12834.
- (145) Ausländer, S., Ketzner, P., and Hartig, J. S. (2010) A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Molecular bioSystems* 6, 807–14.
- (146) Win, M. N., and Smolke, C. D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences of the United States of America* 104, 14283–8.

## References

- (147) Beilstein, K., Wittmann, A., Grez, M., and Suess, B. (2014) Conditional Control of Mammalian Gene Expression by Tetracycline-Dependent Hammerhead Ribozymes. *ACS synthetic biology*.
- (148) Wittmann, A., and Suess, B. (2011) Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Molecular bioSystems* 7, 2419–27.
- (149) Klausner, B., Atanasov, J., Siewert, L. K., and Hartig, J. S. (2014) Ribozyme-Based Aminoglycoside Switches of Gene Expression Engineered by Genetic Selection in *S. cerevisiae*. *ACS synthetic biology*.
- (150) Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.
- (151) An, C.-I., Trinh, V. B., and Yokobayashi, Y. (2006) Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA (New York, N.Y.)* 12, 710–6.
- (152) Tuleuova, N., An, C.-I., Ramanculov, E., Revzin, A., and Yokobayashi, Y. (2008) Modulating endogenous gene expression of mammalian cells via RNA-small molecule interaction. *Biochemical and biophysical research communications* 376, 169–73.
- (153) Beisel, C. L., Chen, Y. Y., Culler, S. J., Hoff, K. G., and Smolke, C. D. (2011) Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic acids research* 39, 2981–94.
- (154) Carothers, J. M., Goler, J. a, Kapoor, Y., Lara, L., and Keasling, J. D. (2010) Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic acids research* 38, 2736–47.
- (155) Suess, B. (2003) Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Research* 31, 1853–1858.
- (156) Ceres, P., Trausch, J. J., and Batey, R. T. (2013) Engineering modular “ON” RNA switches using biological components. *Nucleic acids research* 41, 10449–10461.
- (157) Wallis, M. G., von Ahsen, U., Schroeder, R., and Famulok, M. (1995) A novel RNA motif for neomycin recognition. *Chemistry & biology* 2, 543–52.
- (158) Berens, C., Thain, a, and Schroeder, R. (2001) A tetracycline-binding RNA aptamer. *Bioorganic & medicinal chemistry* 9, 2549–56.
- (159) Ramanathan, A., Robb, G. B., and Chan, S. H. (2016) mRNA capping: Biological functions and applications. *Nucleic Acids Research* 44, 7511–7526.
- (160) Weigand, J. E., Schmidtke, S. R., Will, T. J., Duchardt-Ferner, E., Hammann, C., Wöhnert, J., and Suess, B. (2011) Mechanistic insights into an engineered riboswitch: a switching element which confers riboswitch activity. *Nucleic acids research* 39, 3363–72.
- (161) Duchardt-Ferner, E., Weigand, J. E., Ohlenschläger, O., Schmidtke, S. R., Suess, B., and Wöhnert, J. (2010) Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. *Angewandte Chemie (International ed. in English)* 49, 6216–9.
- (162) Weigand, J. E., Gottstein-schmidtke, S. R., Demolli, S., Groher, F., Duchardt-ferner, E., Wçhnert, J., and Suess, B. (2014) Sequence Elements Distal to the Ligand Binding Pocket Modulate the Efficiency of a Synthetic Riboswitch 1627–1637.
- (163) Müller, M., Weigand, J. E., Weichenrieder, O., and Suess, B. (2006) Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic acids research* 34, 2607–17.
- (164) Förster, U., Weigand, J. E., Trojanowski, P., Suess, B., and Wachtveitl, J. (2012) Conformational dynamics of the tetracycline-binding aptamer. *Nucleic acids research* 40, 1807–17.
- (165) Reuss, A. J., Vogel, M., Weigand, J. E., Suess, B., and Wachtveitl, J. (2014) Tetracycline Determines the Conformation of Its Aptamer at Physiological Magnesium Concentrations. *Biophysical Journal* 107, 2962–2971.
- (166) Wunnicke, D., Strohbach, D., Weigand, J. E., Appel, B., Feresin, E., Suess, B., Müller, S., and Steinhoff, H.-J. (2011) Ligand-induced conformational capture of a synthetic tetracycline riboswitch revealed by pulse EPR. *RNA (New York, N.Y.)* 17, 182–8.
- (167) Xiao, H., Edwards, T. E., and Ferré-D’Amaré, A. R. (2008) Structural basis for specific, high-affinity tetracycline binding by an in vitro evolved aptamer and artificial riboswitch. *Chemistry & biology* 15, 1125–37.
- (168) Kötter, P., Weigand, J. E., Meyer, B., Entian, K.-D., and Suess, B. (2009) A fast and efficient translational control system for conditional expression of yeast genes. *Nucleic acids research* 37, e120.
- (169) Wittmann, A., and Suess, B. (2011) Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Molecular bioSystems* 7, 2419–27.
- (170) Eldar, A., and Elowitz, M. B. (2010) Functional roles for noise in genetic circuits. *Nature* 467, 167–73.
- (171) Sharma, V., Nomura, Y., and Yokobayashi, Y. (2008) Engineering complex riboswitch regulation by dual genetic selection. *Journal of the American Chemical Society* 130, 16310–5.
- (172) Muranaka, N., and Yokobayashi, Y. (2010) A synthetic riboswitch with chemical band-pass response. *Chemical Communications* 46, 6825.
- (173) Klausner, B., Saragliadis, A., Ausländer, S., Wieland, M., Berthold, M. R., and Hartig, J. S. (2012) Post-transcriptional Boolean computation by combining aptazymes controlling mRNA translation initiation and tRNA activation. *Molecular BioSystems* 8, 2242.
- (174) Lorenz, R., Bernhart, S. H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algorithms for molecular biology: AMB* 6, 26.
- (175) Andronescu, M., Condon, A., Hoos, H. H., Mathews, D. H., and Murphy, K. P. (2007) Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics (Oxford, England)* 23, i19–28.

## References

- (176) Wang, Y., and Rando, R. R. (1995) Specific binding of aminoglycoside antibiotics to RNA. *Chemistry & biology* 2, 281–90.
- (177) Schoukroun-Barnes, L. R., Wagan, S., and White, R. J. (2014) Enhancing the analytical performance of electrochemical RNA aptamer-based sensors for sensitive detection of aminoglycoside antibiotics. *Analytical chemistry* 86, 1131–7.
- (178) Wallace, S. T., and Schroeder, R. (1998) In vitro selection and characterization of streptomycin-binding RNAs : recognition discrimination between antibiotics . In vitro selection and characterization of streptomycin-binding RNAs : Recognition discrimination between antibiotics 112–123.
- (179) Bachler, M., Schroeder, R., Ahsen, U. Von, Bachler, M., Schroeder, R., and Ahsen, U. W. E. V. O. N. (1999) StreptoTag : a novel method for the isolation of RNA-binding proteins . StreptoTag : A novel method for the isolation of RNA-binding proteins 1509–1516.
- (180) Windbichler, N., and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nature protocols* 1, 637–40.
- (181) Kwon, M., Chun, S., Jeong, S., and Yu, J. (2001) Molecules and In Vitro Selection of RNA against Kanamycin B 11, 303–311.
- (182) Han, S. R., Yu, J., and Lee, S.-W. (2014) In vitro selection of RNA aptamers that selectively bind danofloxacin. *Biochemical and Biophysical Research Communications* 1–6.
- (183) Grate, D., and Wilson, C. (1999) Laser-mediated, site-specific inactivation of RNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America* 96, 6131–6.
- (184) Baugh, C., Grate, D., and Wilson, C. (2000) 2.8 Å crystal structure of the malachite green aptamer. *Journal of molecular biology* 301, 117–28.
- (185) Paige, J. S., Wu, K. Y., and Jaffrey, S. R. (2011) RNA mimics of green fluorescent protein. *Science (New York, N.Y.)* 333, 642–6.
- (186) Constantin, T. P., Silva, G. L., Robertson, K. L., Hamilton, T. P., Fague, K., Waggoner, A. S., and Armitage, B. a. (2008) Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules. *Organic letters* 10, 1561–4.
- (187) Holeman, L. a, Robinson, S. L., Szostak, J. W., and Wilson, C. (1998) Isolation and characterization of fluorophore-binding RNA aptamers. *Folding & design* 3, 423–31.
- (188) Sinha, J., Reyes, S. J., and Gallivan, J. P. (2010) Reprogramming bacteria to seek and destroy an herbicide. *Nature chemical biology* 6, 464–70.
- (189) Lau, J. L., Baksh, M. M., Fiedler, J. D., Brown, S. D., Kussrow, A., Bornhop, D. J., Ordoukhanian, P., and Finn, M. G. (2011) Evolution and protein packaging of small-molecule RNA aptamers. *ACS nano* 5, 7722–9.
- (190) Geiger, a, Burgstaller, P., von der Eltz, H., Roeder, a, and Famulok, M. (1996) RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic acids research* 24, 1029–36.
- (191) Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., and Breaker, R. R. (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* 113, 577–86.
- (192) Schneider, C., and Suess, B. (2016) Identification of RNA aptamers with riboswitching properties. *Methods* 97, 44–50.
- (193) Carothers, J. M., Goler, J. a, Kapoor, Y., Lara, L., and Keasling, J. D. (2010) Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic acids research* 38, 2736–47.
- (194) Jiang, L., and Patel, D. J. (1998) Solution structure of the tobramycin-RNA aptamer complex. *Nature structural biology* 5, 769–74.
- (195) Rudolph, M. M., Vockenhuber, M. P., and Suess, B. (2013) Synthetic riboswitches for the conditional control of gene expression in *Streptomyces coelicolor*. *Microbiology (United Kingdom)* 159, 1416–1422.
- (196) Schütze, T., Wilhelm, B., Greiner, N., Braun, H., Peter, F., Mörl, M., Erdmann, V. a., Lehrach, H., Konthur, Z., Menger, M., Arndt, P. F., and Glöckler, J. (2011) Probing the SELEX process with next-generation sequencing. *PLoS ONE* 6, 1–10.
- (197) Muranaka, N., Sharma, V., Nomura, Y., and Yokobayashi, Y. (2009) An efficient platform for genetic selection and screening of gene switches in *Escherichia coli*. *Nucleic acids research* 37, e39.
- (198) Liang, J. C., Chang, A. L., Kennedy, A. B., and Smolke, C. D. (2012) A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity. *Nucleic acids research* 40, e154.
- (199) Scholz, O., Thiel, A., Hillen, W., and Niederweis, M. (2000) Quantitative analysis of gene expression with an improved green fluorescent protein. p6. *European journal of biochemistry* 267, 1565–70.
- (200) Schneider, C., and Suess, B. (2015) Identification of RNA aptamers with riboswitching properties. *METHODS* 1–7.
- (201) Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L. B., Weinberger, A., and Segal, E. (2013) Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 110, E2792–E2801.
- (202) Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187–208.
- (203) Wittmann, A., and Suess, B. (2012) Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS letters* 586, 2076–83.
- (204) Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly-characterized Yeast Toolkit for Modular, Multi-part Assembly. *ACS Synthetic Biology* 150414151809002.
- (205) Lee, J. W., Gyorgy, A., Cameron, D. E., Pyenson, N., Choi, K. R., Way, J. C., Silver, P. A., Del Vecchio, D., and Collins, J. J. (2016) Creating Single-Copy Genetic Circuits. *Molecular Cell* 63, 329–336.
- (206) McIsaac, R. S., Silverman, S. J., McClean, M. N., Gibney, P. a., Macinskas, J., Hickman, M. J., Petti, a. a., and Botstein,

## References

- D. (2011) Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell* 22, 4447–4459.
- (207) Gao, C. Y., and Pinkham, J. L. (2000) Short Technical Reports. *BioTechniques* 29, 1226–1231.
- (208) Wieland, M., Benz, A., Klauser, B., and Hartig, J. S. (2009) Artificial Ribozyme Switches Containing Natural Riboswitch Aptamer Domains. *Angewandte Chemie* 121, 2753–2756.
- (209) Roederer, M., Moore, W., Treister, A., Hardy, R. R., and Herzenberg, L. A. (2001) Probability binning comparison: A metric for quantitating multivariate distribution differences. *Cytometry* 45, 47–55.
- (210) Takigami, H., Taniguchi, N., and Shimizu, Y. (2011) Sorption and desorption of 17  $\beta$ -estradiol to natural sediment. *Water Science & Technology* 64, 1473.
- (211) Stoltenburg, R., Reinemann, C., and Strehlitz, B. (2007) SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* 24, 381–403.
- (212) McKeague, M., and Derosa, M. C. (2012) Challenges and opportunities for small molecule aptamer development. *Journal of Nucleic Acids* 2012.
- (213) Nutiu, R., and Li, Y. (2005) In vitro selection of structure-switching signaling aptamers. *Angewandte Chemie (International ed. in English)* 44, 1061–5.
- (214) Davis, J. H., and Szostak, J. W. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proceedings of the National Academy of Sciences of the United States of America* 99, 11616–21.
- (215) Wu, S., and Letchworth, G. J. (2004) High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. *BioTechniques* 36, 152–154.
- (216) Hanson, S., Bauer, G., Fink, B., and Suess, B. (2005) Molecular analysis of a synthetic tetracycline-binding riboswitch Molecular analysis of a synthetic tetracycline-binding riboswitch 503–511.
- (217) Stovicek, V., Borodina, I., and Forster, J. (2015) CRISPR–Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae* strains. *Metabolic Engineering Communications*.
- (218) Crane, M. M., Clark, I. B. N., Bakker, E., Smith, S., and Swain, P. S. (2014) A Microfluidic System for Studying Ageing and Dynamic Single-Cell Responses in Budding Yeast. *PLoS ONE* (Moses, A. M., Ed.) 9, e100042.

## 9 Talks and Poster Presentations

2015

**Winter School SFB902**, Obergurgl, Austria

Talk: riboregulation - from sensing to switching –

2015

**33. Rabensteiner Kolleg**, Pottenstein, Deutschland

Talk: synthetic RNA biology - logic gates in yeast –

2015

**International Synthetic and Systems Biology Summer School 2015**, Taormina, Sicily

Poster: Design and Implementation of RNA-regulated Genetic Circuits following Boolean Logic

2016

**LOEWE Schwerpunkt CompuGene, 1. Retreat, Annweiler-Bindersbach**

Talk: Yeast Single Cell Analysis of RNA-engineered Logic Gates

2017

**LOEWE Schwerpunkt CompuGene, 2. Retreat, Annweiler-Bindersbach**

Talk: ROC'n'Ribo: Characterizing a riboswitching expression system by modeling single-cell data

## 10 Publications

Schneider, C., and Suess, B. (2016) **Identification of RNA aptamers with riboswitching properties.** *Methods* 97, 44–50

Schneider C., Bronstein L., Diemer J., Koepl H., Suess B. (2017) **ROC'n'Ribo: Characterizing a riboswitching expression system by modeling single-cell data.** *ACS Synth Biol.*, revised

## 11 Curriculum Vitae

**Name** Christopher Schneider  
**Date of birth** 13.11.1987 in Siegen  
**Address** Darmstaedterstraße 35a  
64807 Dieburg  
Germany

### Professional experience

01/2014 - present **Institute for Biology | Synthetic genetic circuits |  
Technische Universitaet Darmstadt, Germany**  
Graduate student in the Lab of Prof. Dr. Beatrix Suess  
Doctoral thesis: *Design, implementation and characterization of  
synthetic riboswitches in Saccharomyces cerevisiae*  
Degree: Dr. rer. nat.

### Education

10/2011-12/2013 **Master of Science in Life Science**  
Universitaet Konstanz  
Master thesis: *Ribozyme-regulated control of functional RNAs in E. coli*  
Degree: M. Sc.

10/2008 - 08/2011 **Bachelor of Science in Pharmaceutical Sciences**  
Ludwig-Maximilians-Universitaet Muenchen  
Bachelor thesis: *Analysis of the protein interactome of  
cyclonucleotide-gated (CNG-) channels*  
Degree: B. Sc.

08/1998 - 06/2007 **Gymnasium Netphen, Netphen, Germany**  
University-entrance diploma

## 12 Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den 14.03.2017

.....

(Christopher Schneider)