Beyond Plug and Pray: Context Sensitivity and *in silico* Design of Artificial Neomycin Riboswitches

Christian Günzel[*1], Felix Kühnl[*†2], Katharina Arnold[1], Sven Findeiß[2], Christina Weinberg[1], Peter F Stadler[‡2,3,4,5,6], and Mario Mörl[§1]

[1]Institute for Biochemistry, Leipzig University, Brüderstraße 34, D-04103 Leipzig, Germany
[2]Bioinformatics Group, Institute of Computer Science, and Interdisciplinary Center for Bioinformatics, Leipzig University, Härtelstraße 16–18, D-04107 Leipzig, Germany
[3]Max-Planck-Institute for Mathematics in the Sciences, Inselstraße Leipzig, D-04103 Leipzig, Germany
[4]Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria
[5]Facultad de Ciencias, Universidad National de Colombia, Sede Bogotá, Colombia
[6]Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

May 11, 2020

1

Abstract

Gene regulation in prokaryotes often depends on RNA elements such as riboswitches or RNA thermometers located in the 5' untranslated region of mRNA. Rearrangements of the RNA structure in response, e. g., to the binding of small molecules or ions control translational initiation or premature termination of transcription and thus mRNA expression. Such structural responses are amenable to computational modeling, making it possible to rationally design synthetic riboswitches for a given aptamer. Starting from an artificial aptamer, we construct the first synthetic transcriptional riboswitches that respond to the antibiotic neomycin. We show that the switching behavior *in vivo* critically depends not only on the sequence of the riboswitch itself, but also on its sequence context. We therefore developed *in silico* methods to predict the impact of the context, making it possible to adapt the design and to rescue non-functional riboswitches. We furthermore analyze the influence of 5' hairpins with varying stability on neomycin riboswitch activity. Our data highlight the limitations of a simple plug-and-play approach in the design of complex genetic circuits and demonstrate that detailed computational models significantly simplify, improve, and automate the design of transcriptional circuits. Our design software is available under a free license on Github.[1]

---

*These authors contributed equally.

†https://orcid.org/0000-0002-0665-6962

‡https://orcid.org/0000-0002-5016-5191

§corresponding author, mario.moerl@uni-leipzig.de

[1]https://github.com/xileF1337/riboswitch_design

1. Introduction

Riboswitches are *cis*-regulatory elements that modulate gene expression in response to small molecules or ions without direct involvement of proteins [1, 2]. They are typically embedded in the 5' untranslated region (UTR) of messenger RNA (mRNA) and usually comprise two overlapping domains: a conserved *aptamer*, presenting the binding site for a specific ligand, and a gene-specific *expression platform* that regulates the expression of the gene located downstream. Riboswitches can adopt mutually exclusive structures depending on the presence or absence of their ligand. The structural reorganization or stabilization of the aptamer upon ligand binding triggers a change of the structure of the expression platform [3]. The latter may affect either translation – e. g., by changing the accessibility of the ribosomal binding site – or transcription, e. g., by modulating an intrinsic terminator causing the premature dissociation of RNA polymerase (RNAP). The relatively small size of riboswitches facilitates their design and optimization [4], which makes them attractive for numerous applications in synthetic biology. The ability to construct orthogonal switches, for instance, makes it possible *in principle* to force a gene to respond to any cell-permeable and non-toxic ligand [5] and thus to create functional biosensors that turn on the expression of a certain reporter gene in response to both intracellular signals and environmental conditions. The practicality of such an approach, however, depends on how easily aptamers and expression platforms can be combined and embedded in an arbitrary sequence context without resorting to a labor-intensive trial-and-error procedure.

The antibiotic neomycin has a generic RNA binding affinity which allowed to engineer cognate aptamers by means of systematic evolution of ligands by exponential enrichment (SELEX) [6, 7]. The aptamer N1 was isolated by Weigand et al. [8] and shown to act as a translational roadblock in presence of neomycin in yeast. N1 binds to its ligand with an extraordinarily high affinity in the lower nanomolar range [9]. Its switching behavior in response to neomycin was analyzed *in vivo* in conjunction with various closing stems that modulate the expression of a green fluorescent protein (GFP) reporter located downstream. The shortest stem, designated M7, exhibited both the highest fluorescence intensity and the best ratio between its *on* and *off* state [9].

Based on the aptamer N1M7, we constructed a riboswitch that controls gene expression in *Escherichia coli* (*E. coli*). To this end, we focused on transcriptional regulation mediated by Rho-independent termination, a mechanism common in prokaryotes. It depends on a specific terminator structure, consisting of a GC-rich stem–loop followed by 7–9 uracil residues referred to as *poly-U stretch* [10, 11].

3

We used the *in silico* pipeline developed previously for designing theophylline and tetracycline riboswitches [12, 13] to integrate the N1M7 aptamer with artificial terminator hairpins. The designed neomycin riboswitches indeed increase the transcription level of the reporter mRNA in presence of neomycin *in vivo*. A detailed analysis of the constructs, however, shows that their function *in vivo* is context-dependent. In particular, we identify a 5' leader hairpin as essential element for the function of these neomycin riboswitches. On the other hand, leader sequences may also abrogate the function by interfering with the riboswitch. We demonstrate here that folding simulations predict the interference and thus can be used to identify functional constructs *in silico*.

2. Materials and Methods

2.1. Cells

For cloning purposes, *E. coli* Top10 cells (Invitrogen) were used. Riboswitch-controlled expression of the enhanced green fluorescent protein (eGFP) was tested by *in vivo* fluorescence measurements using the neomycin-resistant *E. coli* SQ171, derived from *E. coli* MG1655 [14] (obtained from Kurt Fredrick, Columbus, OH).

2.2. Plasmid construction

As reporters, the enhanced green fluorescent protein gene (*egfp*) or the β-galactosidase gene (*bgaB*) were inserted into plasmid pRSF1030Tp [15] by restriction and ligation reactions (using the *Nco*I and *Xba*I site for *egfp* or *Age*I and *Kpn*I site for *bgaB*). Riboswitch constructs were synthesized and amplified by *de novo* PCRs. For *de novo* fragments longer than 90 bp, the construct was split in two or more overlapping fragments and extended by overlap extension PCR with flanking primers. The riboswitch sequences were inserted downstream of the *araBAD* promotor and upstream of *egfp* or *bgaB* through restriction and ligation reactions or mutagenesis PCR (for all LM constructs). All PCR reactions were conducted according to the protocol of HiFi DNA polymerase (PCR Biosystems). For overlap extension PCR, the flanking primers were added after 10 cycles, followed by amplification for additional 25 cycles. Plasmid isolations were performed according to the GeneJET Plasmid Miniprep Kit protocol (Thermo Scientific) for low-copy-number plasmids. All constructs were verified by sequencing.

2.3. *In vivo* measurements of fluorescence intensity

The *in vivo* eGFP fluorescence measurements were performed in *E. coli* SQ171. Precultures were incubated in 5 ml LB medium with $200\,\mu g\,ml^{-1}$ trimethoprim over night at $37\,°C$ and 200 rpm. 100 µl of the precultures were used to inoculate fresh cultures (5 ml), which were incubated at $37\,°C$ and 200 rpm until an OD_{600} of 0.5. Expression of eGFP was induced by addition of L-(+)-arabinose (to 0.1% w/v) in the absence and presence of neomycin ($250\,\mu g\,ml^{-1}$). After incubation for 3 h at $37\,°C$ and 200 rpm, cultures were immediately stored on ice to measure the final OD_{600} and diluted with cold 1× PBS to 1 ml ($OD_{600} = 1.5$). Samples were centrifuged at 8 000 rpm for 5 min, the pellet was washed twice with cold 1× PBS and resuspended in 500 µl 1× PBS (equaling $OD_{600} = 3.0$). eGFP fluorescence was measured from 503 nm to 600 nm with a fluorescence spectrometer (JASCO Deutschland GmbH, model: FP-8500, mode: emission, ex bandwidth: 2.5 nm, em bandwidth: 5 nm, response: 1 s, sensitivity: low, ex wavelength: 488 nm).

As a positive control, we combined a 5' adenosine residue following the transcription start site (TSS) with the poly-U stretch of length eight and the terminator downstream region (TDR) of our riboswitch constructs. The poly-U stretch is known to promote translation [16] and, thus, makes the control comparable to the full riboswitch sequences.

2.4. RNA extraction and northern blots

Total RNA was isolated from *E. coli* SQ171 cells grown in the presence or absence of 275 µM neomycin until $OD_{600} = 0.5$ in 20 ml LB medium supplied with trimethoprim ($200\,\mu g\,ml^{-1}$) and arabinose (0.1%). Total RNA was isolated using TRIzol™ (Invitrogen) according to the manufacturer's protocol.

Northern blot analysis was performed as described by Köhrer and Raj-Bhandary [17] with modifications. 10 µg of total RNA were mixed with 3× denaturing sample buffer (10 mM Tris–HCl, pH 7.6, 80% formamide (v/v), 0.25% bromophenol blue and 0.25% xylene cyanol) and incubated for 1 min at $90\,°C$ before electrophoretic separation on a 1% agarose gel containing 1× Tris-borate-EDTA (TBE) (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0).

The gel (7 × 8 cm) was run at 70 V in 1× TBE for 1.5 h. RNA was trans-ferred onto BrightStar™ Positively Charged Nylon Membranes (Ambion®) at $2\,mA\,cm^{-2}$ in a semi-dry electroblotter (Thermo Fisher Scientific) for 45 min using 1× TBE buffer. RNA fixation by cross-linking was performed by irra-diation at 254 nm UV light for 60 s ($120\,mJ\,cm^{-2}$ total) in a Stratalinker® UV
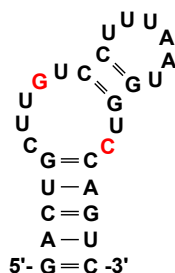
Figure 1: Sequence and secondary structure of the aptamer N1M7 [8]. Note that the minimum free energy structure predicted by the *ViennaRNA* package contains an additional GC pair at position 9 and 22 (colored red, energy difference $\Delta\Delta G = -0.7\,\mathrm{kcal\,mol^{-1}}$).

Crosslinker (Artisan Technology Group). The membrane was pre-hybridized for 14 h to 20 h at 37 °C in 6× saline–saline sodium citrate buffer (SSC) containing 10× Denhardt's solution and 0.5% SDS, followed by hybridization for 12 h to 24 h at 37 °C (for *bgaB*) or 60 °C (for *egfp*) in 6× SSC containing 0.1% SDS with 5' radioactively labeled oligonucleotide probes for *bgaB* mRNA[2], *egfp* mRNA[3] or 5S ribosomal RNA (rRNA)[4]. Probes were radioactively labeled with γ-^{32}P-ATP (Hartmann analytic), unincorporated nucleotides were removed with NucAway™ Spin Columns (Invitrogen). The hybridized membrane was washed twice at room temperature with 6× SSC, 4× SSC and 2× SSC (each wash step 10 min). Signals were detected using a Typhoon phosphor imager (GE Healthcare Life Sciences, model: 9410). For quantitation of mRNA levels, northern blot signals of three independent colonies per construct (with and without neomycin) were analyzed using ImageQuant 8.1 (GE Healthcare Life Sciences). Membranes were stripped of radiolabelled probes by incubating in 0.5% SDS at room temperature for 1 h.

2.5. The N1M7 aptamer

The starting point for our *in silico* design is the sequence of the well-characterized neomycin aptamer N1M7, i. e., the aptamer N1 combined with the closing stem M7 consisting of five base pairs [8], cf. Fig. 1. The aptamer binds neomycin with a dissociation constant of $K_d = (9.2 \pm 1.3)\,\mathrm{nM}$, corresponding to a stabilizing Gibbs free energy contribution of $\Delta G = (-11.4 \pm 0.1)\,\mathrm{kcal\,mol^{-1}}$ [9].

---

[2]5'-GGAGCAATAACTACTTTGTATTTTG-3'

[3]5'-GCCATGTGTAATCCCAGCAGCTGTTACAAACTCAAGAAGG-3'

[4]5'-TTCTGAATTCGGCATGGGGTCAGGTGG-3'

2.6. Probabilities of RNA conformations

RNA molecules can adopt a huge number of different secondary structures. In thermodynamic equilibrium, the probability to observe a particular structure x depends on its Gibbs free energy $\Delta G(x)$, which can be computed using the nearest-neighbor energy model [18], which specifies sequence-dependent stabilizing contributions for base pair stacking and destabilizing loop terms. The Boltzmann weight of x is defined as $\mathrm{Z}[x] := \exp(-\Delta G(x)/(RT))$, where R is the universal gas constant and T is the absolute temperature. The *partition function* $\mathrm{Z}[Y]$ of a *set* of secondary structures Y is the sum of all the structures' Boltzmann weights, i. e., $\mathrm{Z}[Y] := \sum_{y \in Y} \mathrm{Z}[y]$. Writing $\mathrm{Z} := \mathrm{Z}[X]$ for the partition function of the full ensemble of the RNA, the equilibrium probability of a given structure x is $\Pr[x] = \mathrm{Z}[x]/\mathrm{Z}$, and, more generally, $\Pr[Y] = \mathrm{Z}[Y]/\mathrm{Z}$ for a set of structures Y. Partition functions of the entire set of structures X as well as certain subsets Y (such as the set structures containing a specific hairpin loop, or the set of structures containing a particular base pair) can be computed by dynamic programming. We used the *ViennaRNA* package [19, 20] because it provides a versatile way to handle constraints [21] and thus to compute partition functions for a set Y of structures with desired structural features, e. g., a terminator hairpin or a binding pocket for a certain ligand.

2.7. Simulation of cotranscriptional folding

While initial folding intermediates of RNA form at time-scales of tens of microseconds, the formation of native hairpins appears at millisecond time-scales [22, 23, 24, 25], and the refolding of secondary structure elements may take even longer. In comparison, RNA is transcribed by *E. coli* RNAP with a rate of $30\,\mathrm{nt\,s^{-1}}$ to $90\,\mathrm{nt\,s^{-1}}$ [26, 27]. Thus, RNA folding forms intermediate structures long before the entire molecule is transcribed, i. e., while only part of the transcript is available to form structures. The structures formed initially thus may refold as transcription proceeds. This process of cotranscriptional folding [28] plays an important role in particular for *transcriptional* riboswitches, since incomplete, metastable intermediate structures may be quite different from the thermodynamic ground state [29]. Cotranscriptional folding can be assessed either by stochastic sampling of folding trajectories [30, 31, 32] or by analyzing the energy landscapes for each elongation step [33]. We opted for the latter method.

For each length of the nascent transcript, we enumerated all secondary structures in an energy band above the ground state with *RNAsubopt* [34], also a component of the *ViennaRNA* package. Then we used *Barriers* [35] to

produce a coarse-grained representation of the energy landscape comprising the low energy minima as well as the saddle points between them. *Barriers* assigns each structure to a basin of attraction. The individual landscapes were then integrated using *BarMap* [33]. In brief, *BarMap* determines the correspondence of energy basins in the landscapes of consecutive transcriptional elongation steps. This allows to efficiently simulate the folding dynamics. To determine the energy thresholds for the enumeration of structures with *RNAsubopt* in the individual landscapes, we used the quality scores for both the enumeration and the simulation provided by *BarMap-QA* [36]. We set a simulation stop time of 1 000 au (arbitrary time units) for *BarMap*. Simulation results were plotted with Grace [37]. Candidate riboswitches were screened to satisfy the following criteria: (i) the leader sequence does not interfere with the formation of the aptamer's binding pocket, (ii) the binding pocket is significantly populated during the transcription of the spacer and the 5' part of the terminator, and (iii) the terminator hairpin dominates the structure ensemble as soon as it is fully transcribed. For a precise description of the preparation steps, quality metrics, and the post-processing, we refer to [36]. To allow for easy reproduction, the entire simulation pipeline was packaged into a self-contained and publicly available Docker image.[5]

2.8. Design of transcriptional riboswitches

The design process consists of two separate phases. First, the riboswitch itself is designed from the given aptamer sequence. Second, for a set of designed riboswitch constructs, a decoupling leader (dL) is constructed, which isolates the riboswitches from the original leader (oL) sequence found upstream on the plasmid. This second step may become necessary if oL has the potential to interfere with the structure formation of the designed riboswitches (cf. Section 3.1). The full process is depicted as a flowchart in Fig. 2.

To engineer the switch, a software pipeline similar to the one described by Wachsmuth et al. [12] including updated filter rules has been used. Briefly, this software is given the aptamer sequence and its ligand-binding structure as input, generates riboswitch candidates according to a specified pattern (*generation step*), and then applies a set of filters to remove unsuitable candidates (*filtering step*).

The generation step appends a random spacer sequence of varying length to the aptamer sequence. Then, the reverse complement of the last k nucleotides of the aptamer sequence is appended, forming a stable hairpin loop with the

---

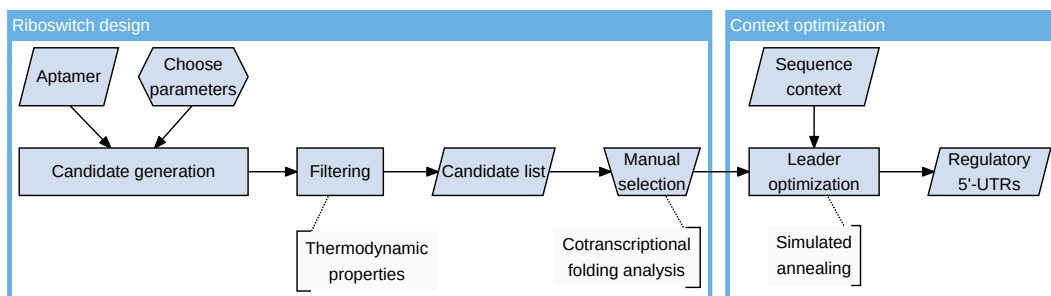[5]https://www.bioinf.uni-leipzig.de/Software/BarMap_QA/

Figure 2: Flowchart of the design process used to create neomycin-dependent riboswitches and to include them into regulatory 5' UTRs. Starting from a given aptamer sequence and structure, riboswitch candidates are generated and selected as described in the text.

aptamer and disrupting the formation of the ligand-binding structure. The value of k varies from two to half of the aptamer length among the candidates. In addition, a U-stretch consisting of eight uracil residues is added, completing the structure of an intrinsic terminator [38].

To remove potentially faulty constructs, the filtering step applies a set of filters as described in Wachsmuth et al. [12], removing each sequence that violates any of the filter rules. The rules ensure correct terminator formation in the minimum free energy (MFE) structure and scan for possibly interfering intermediate structures by the means of thermodynamic folding simulations of a set of subsequences of the full switch. Additionally, we added a probability-based filter ensuring that, for the full sequence, the terminator structure forms with a probability of at least 95%. Also, it was ensured that at least two base pairs of the terminator hairpin can form despite the presence of the binding-competent aptamer structure. These *seed base pairs* facilitate the rapid formation of the terminator hairpin when the ligand is not present since the zippering of a helix happens at a higher rate than the nucleation of the first base pair [22, 23]. Finally, promising candidates resulting from this selection process where analyzed using cotranscriptional folding simulations as described in the previous section.

2.9. Designing a decoupling leader sequence

To prevent the predicted interference of the original leader located upstream in the 5' UTR with the designed riboswitches, a sequence insert that effectively decouples leader and riboswitch has been designed. To this end, we defined an objective function F as described below and optimized a sequence with

9

respect to F. The sequence with the best score was then cloned into the vector immediately upstream of the riboswitch, cf. Fig. A.1 in Appendix A. To keep the required experimental effort as low as possible, only one optimized leader suitable for all riboswitch constructs was designed.

Let ℓ denote the original leader, and $R = \{r_1, \ldots, r_n\}$ be a set of n riboswitches. The goal is to construct an insert $x_m \in \{A, U, G, C\}^m$ of length m that minimizes the objective function $F(x_m \mid \ell, R)$ measuring the interference of the upstream sequence ℓx_m with each of the riboswitches, subject to a constant length m. A natural measure for the isolation of ℓx_m and the riboswitches is the probability

$$p_{\text{unpaired}}(\ell x_m, R) = \prod_{r \in R} \frac{Z[\ell x_m] \cdot Z[r]}{Z[\ell x_m r]},$$

that all base pairs occur either in the upstream part ℓx_m of the sequence (structures in $Z[\ell x_m]$) or within the riboswitches (structures in $Z[r]$), but not between the two substructures. Using $\Delta G(x) = -RT \ln Z[x]$, it can be expressed equivalently in terms of Gibbs free energies, with the added benefit of numerical stability. We therefore use the following objective function:

$$F(x_m \mid \ell, R) = \sum_{r \in R} \Delta G(\ell x_m) + \Delta G(r) - \Delta G(Z[\ell x_m r]) \longrightarrow \min.$$

The optimization of x_m was carried out using a standard simulated annealing procedure [39], starting at a random sequence and applying single nucleotide mutations to the insert to generate new candidates. A proposal sequence x_m' was *always* accepted if it performed better than the current state x_m (i. e., if $F(x_m') < F(x_m)$), and otherwise with a probability of $\exp((F(x_m) - F(x_m'))/\tau)$, where the "annealing temperature" τ was slowly decreased with time. More specifically, we set the initial temperature to $\tau_0 = 1000$ and, after each mutation, cooled it down by letting $\tau_{n+1} = 0.97 \cdot \tau_n$. We used the rejection of $3m$ consecutive proposals as stopping criterion, where $3m$ is the number of neighbor sequences which can be obtained by applying a single point mutation to the candidate insert of length m. This ensures that, on average, about two thirds of its neighbors are sampled before terminating the optimization run. During the optimization, m was fixed because longer sequences generally have a higher potential to form stable structures fulfilling the objective, which could cause degenerate optimization runs with candidates of ever-growing sequence length.

3. Results

Many of our results are presented as bar plots of fluorescence intensity. The numerical values and raw data used to generate these can be found in Appendix B, Table B.1. Sequence data is available as a supplementary spreadsheet file.

3.1. Design of artificial neomycin riboswitches

Using computational predictions of ligand-induced differences in RNA secondary structure formation, we designed a set of transcription-regulating riboswitches based on the well-characterized neomycin aptamer N1M7, Fig. 1. These were evaluated *in vivo* in *E. coli* strain SQ171 using eGFP as reporter. This strain is neomycin-resistant, as all endogenous rrn operons were deleted and replaced by a plasmid-borne version carrying the mutated neomycin target site A1408G [14].

The plasmid used for assaying our riboswitch constructs contained a leader sequence we termed *original leader (oL)* immediately downstream of the promoter. Since its effect on transcription was unknown, we did not attempt to delete it. *In silico* analysis of the constructs suggested, however, that the oL sequence interferes with the structure formation of the N1M7 aptamer and thereby prevents the correct folding of its ligand binding pocket, Fig. 3. To remedy this issue, we designed a decoupling leader (dL) with a length of 15 nt that reduces the probability that base pairs form between oL and the riboswitch domain to less than 1%. It does so by forming the stable hairpin LH1, i. e., LH1 = oL + dL (cf. Fig. A.1).

Four candidates prefixed with LH1 were selected for in-depth analysis, Fig. A.2. Fluorescence measurements of reporter gene expression (cf. Fig. 4A–B) showed that LH1-N1M7-C and LH1-N1M7-D function as *on*-switches, i. e., the ligand causes up-regulation of the reporter. In contrast, LH1-N1M7-A remains in a permanent *off* state while LH1-N1M7-B exhibits a permanent *on* state. As a representative example, northern blots in the presence and absence of neomycin were used to validate that LH1-N1M7-D regulates the amount of transcripts in the cell, Fig. 4C–E. Blotting experiments targeting *egfp* mRNA resulted in a low signal-to-noise ratio. Hence, we replaced the reporter gene by the well-established β-galactosidase (bgaB) reporter that we have used successfully for detection in northern blot analyses of theophylline-dependent riboswitches [12].
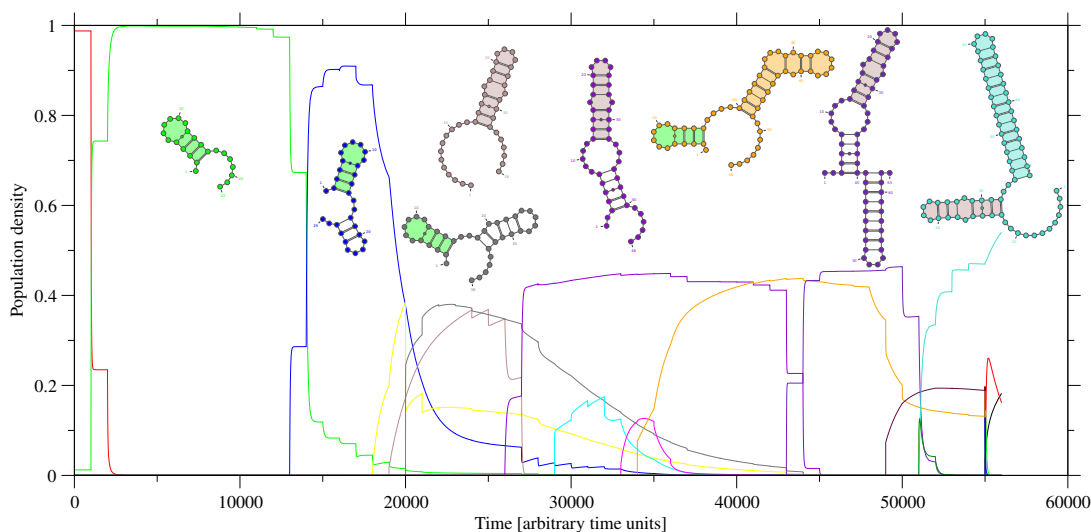
11

Figure 3: Influence of 5' the original leader (oL), which is predicted to interfere with the used aptamer *in silico*, on the kinetics of riboswitch N1M7-D. The individual curves show the population of different macrostates during the elongation of the transcript. The most stable structure of highly populated states is shown in the respective color. Important sub-structures are shaded.

In essence, oL forms two distinct, stable structures: a small hairpin of four base pairs (green) that is compatible with the aptamer's binding pocket (orange), and a larger hairpin (brown) incorporating an interior loop that is extended later (light and dark purple) and interferes with the aptamer region. Both structural shapes are equally populated with about 40%, which means that the aptamer is not available for binding the ligand in about every other transcribed molecule. The terminator (cyan) forms immediately after it has been transcribed. This explains the permanent *off* state that has been observed for Lpl-N1M7-D experimentally, cf. Fig. A.4.

3.2. Riboswitch N1M7-D requires a 5' hairpin structure

Instead of designing a decoupling sequence for oL with a complicated optimization method as described in the previous section, the more obvious approach seems to simply remove oL from the plasmid and place the riboswitch candidates directly downstream of the TSS represented by an adenosine residue. It turns out, however, that the removal of all leader sequences from the functional riboswitch LH1-N1M7-D yields a *non-functional* construct termed N1M7-D,
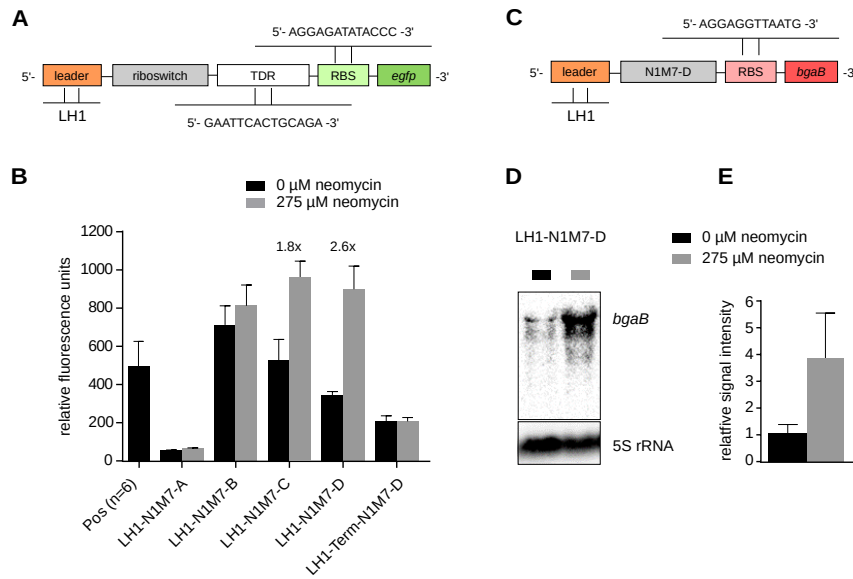
12

Figure 4: Design and analysis of N1M7 riboswitch constructs. (*A*) Schematic overview of the riboswitch constructs for the eGFP assay. The riboswitches are flanked upstream by LH1 and downstream by the terminator downstream region (TDR), ribosomal binding site (RBS) and the enhanced green fluorescent protein gene (*egfp*). (*B*) Fluorescence intensity of the measured N1M7 constructs in the absence (black) and presence (grey) of neomycin. The positive control, a transcript consisting of a single adenosine residue followed by a poly-U stretch and the TDR, shows the fluorescence intensity without leader and riboswitch sequence. The terminator efficiency was verified with construct LH1-Term-N1M7-D. Here, the 5' part of the aptamer sequence that is not overlapping with the intrinsic terminator was deleted, such that the terminator forms irrespective of the presence of neomycin. If not indicated otherwise, measurements were performed using three independent replicates. (*C*) Schematic presentation of a neomycin riboswitch construct for northern blot analysis. Downstream of the riboswitch, the construct carries the β-galactosidase reporter gene (*bgaB*) including the RBS published in Wachsmuth et al. [12]. Due to the design strategy, this construct does not carry a TDR. (*D*) Northern blot with 10 μg of total RNA of *E. coli* strain SQ171 per lane regulated by LH1-N1M7-D, in the absence and presence of neomycin. 5S rRNA was used as internal standard. (*E*) Northern blot quantification of *bgaB* expression of three independent samples.
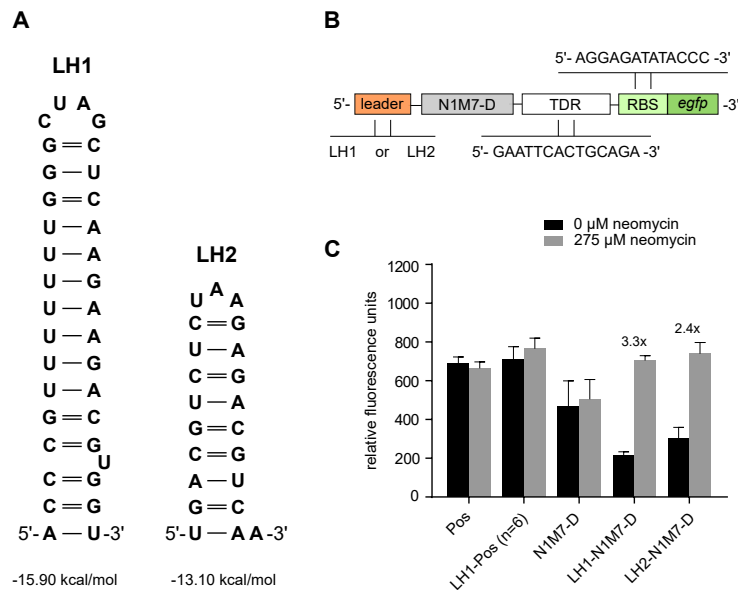
13

Figure 5: Fluorescence intensity of the riboswitch N1M7-D in conjunction with two different leader hairpins LH1 and LH2. (A) Sequence and secondary structure of LH1 and LH2. (B) Schematic overview of the constructs used in panel C. (C) Fluorescence intensity of the two leaders LH1 and LH2 placed upstream of N1M7-D. Legend as described in Fig. 4.

which is no longer sensitive to neomycin (Fig. 5C). The riboswitch function was rescued by inserting a second, computationally designed leader hairpin LH2 (Fig. 5A) resulting in the functional riboswitch LH2-N1M7-D, Fig. 5C.

To rule out that the rescue is the consequence of a changed TSS, or a direct interaction of the leader hairpin with the riboswitch domain, we tested two short unstructured leader sequences U1 and U2, 12 nt and 14 nt in length, Fig. 6A. In the computational design of U1 and U2, base pairing interactions with the riboswitch domain were avoided. Both constructs, U1-N1M7-D and U2-N1M7-D, were functional with a fold change of 1.6 and 1.4 (respectively), however, their fluorescence intensity was significantly reduced, Fig. 6C. As with N1M7-D, it resembled the *off* state of the functional switch. Next, we re-added the leader hairpin LH1 to the 5' ends of these impaired riboswitches. As a result, the original fluorescence activity was restored for both LH1-U1-N1M7-D and LH1-U2-N1M7-D, Fig. 6B.
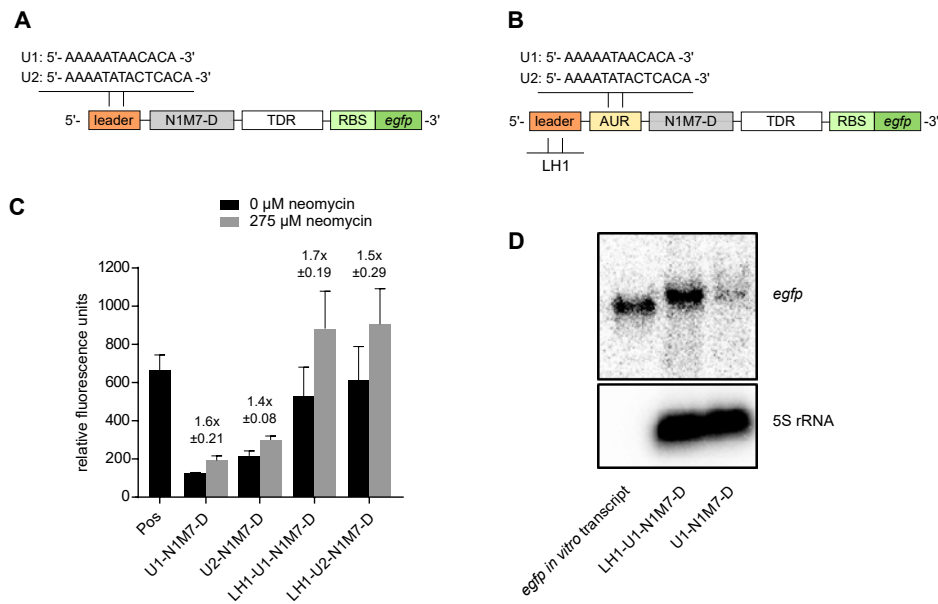
14

Figure 6: Fluorescence intensity of the N1M7-D riboswitch with short unstructured leaders. (*A*) Schematics showing the position of the unstructured sequences U1 and U2 as leaders upstream of N1M7-D and (*B*) as aptamer upstream region (AUR) between LH1 and N1M7-D. (*C*) Fluorescence intensity of the constructs shown in (*A*) and (*B*). Legend as in Fig. 4. (*D*) Northern blot of 10 µg of total RNA per lane from *E. coli* strain SQ171 regulated by U1-N1M7-D or LH1-U1-N1M7-D, in the presence of 275 µM neomycin. Herein, *egfp* was used as reporter gene and 5S rRNA as internal standard. A 1100 nucleotide (nt) *egfp in vitro* transcript was used as positive control. Note that the band of *in vivo egfp* is located about 100–200 nt higher due to the additional rrnB T1 and T2 terminator length, which was not a factor in size calculation for the *in vitro* transcribed control.

3.3. Destabilizing the leader hairpin decreases reporter gene expression

Since unstructured leader regions abrogated the switching behavior, we investigated whether the stability of the leader hairpin has a systematic effect on the riboswitch. To this end, we first introduced destabilizing point mutations into the leader hairpin LH1. We obtained two new leaders LM1 and LM2 (Fig. 7A), each lacking two base pairs. The destabilization was reverted by compensatory mutations in LM1C and LM2C, restoring the stem of LH1. We found that
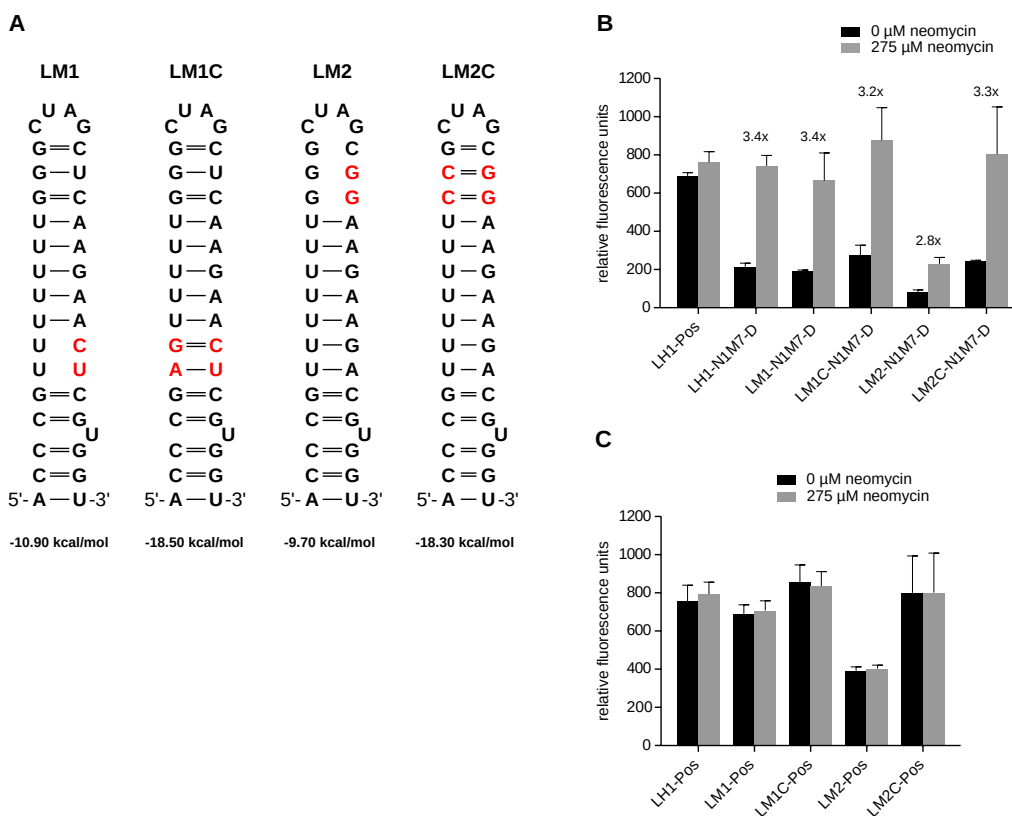
15

A

LM1	LM1C	LM2	LM2C

-10.90 kcal/mol   -18.50 kcal/mol   -9.70 kcal/mol   -18.30 kcal/mol

B

C

Figure 7: Impact of point mutations in the leader hairpin LH1. Fluorescence intensity of the modified leader LM1, LM1C, LM2 and LM2C (A) in conjunction with N1M7-D (B) or Pos (C), along with the controls LH1-N1M7-D or LH1-Pos. Legend as described in Fig. 4. Mutation details and minimum free energies are given in Table B.2.

none of these mutations significantly affected the functionality of the riboswitch N1M7-D. The activation ratios of these constructs range between 2.8-fold to 3.4-fold, comparable to LH1-N1M7-D, cf. Table B.2.

There is, however, a trend in the overall fluorescence intensity: most stable hairpins lead to a higher eGFP expression, both in presence and absence of neomycin (Fig. 7B). Constructs lacking the riboswitch domain, i.e., those consisting of a leader hairpin and the positive control (Fig. 7C), show the same trend in the overall fluorescence intensity.

Similarly, we destabilised the alternative leader hairpin LH2, obtaining hairpins LM3, LM4 and LM5, Fig. A.3A. The construct LM3-N1M7-D, containing the least stable leader hairpin with an MFE of $\Delta G = -5.5\,\text{kcal}\,\text{mol}^{-1}$, exhibits a constitutive *on* state (Fig. A.3B) but still shows a moderate acti-

16

vation rate of about 1.5. Two compensatory mutations restore the hairpin (LM3C, $\Delta G = -13.9 \, \mathrm{kcal \, mol^{-1}}$) to almost the same folding energy as LH2 ($\Delta G = -13.1 \, \mathrm{kcal \, mol^{-1}}$) and rescue the function of the riboswitch. LM4-N1M7-D, with $\Delta G = -10.1 \, \mathrm{kcal \, mol^{-1}}$, shows almost no difference in the activation rate or fluorescence intensity compared to LH2-N1M7-D or LM3C-N1M7-D. The fluorescence intensity of LM5-N1M7-D was decreased compared to LM4-N1M7-D, while the activation ratio virtually remained unchanged.

In summary, we observe that (i) a sufficiently stable leader hairpin appears to be required for functional constructs and (ii) the stability of the leader hairpin correlates with the constructs' fluorescence intensity.

3.4. Upstream sequence context may impair riboswitch function

As mentioned before, we predicted that oL interferes with the aptamer N1M7 using *in silico* simulations. To verify this finding experimentally, we inserted oL between the leader hairpins and the riboswitch domains, resulting in the constructs LH1-oL-N1M7-D and LH2-oL-N1M7-D. In addition, we analysed oL-N1M7-D without any leader hairpin. All three constructs showed no neomycin-dependent regulation of eGFP expression *in vivo*, cf. Fig. A.4, even though the corresponding constructs without oL between LH1 and N1M7-D (namely LH1-N1M7-D, LH1-U1-N1M7-D, and LH1-U2-N1M7-D) were functional, cf. Section 3.2.

The expression level of LH1-oL-N1M7-D or LH2-oL-N1M7-D resembles the *off* state of the functional switch LH1-N1M7-D with a fluorescence intensity of 200–400 RFU (relative fluorescence units). Without a leader hairpin upstream of oL, the fluorescence reached an intensity of about 150 RFU. These low fluorescence intensities combined with the disturbed switching behaviour is in accordance with the predicted interaction of the oL with the neomycin binding pocket of the aptamer sequence N1M7. According to the simulations, this interaction not only prevents the formation of the neomycin binding pocket, but also facilitates the emergence of the terminator hairpin.

4. Discussion

In this work, we combined a synthetic neomycin aptamer with computationally designed terminator hairpins that abrogate transcription in the absence of the aptamer-specific ligand. The design process followed our earlier successful construction of theophylline- and tetracycline-dependent, transcription-regulating

17

riboswitches [12, 13]. Here, we employed the neomycin aptamer N1M7, exhibiting an extraordinarily high binding affinity of $K_d = (9.2 \pm 1.3)\,\mathrm{nM}$ [9], as ligand sensor. Of the four computationally designed constructs that were tested *in vivo*, LH1-N1M7-C and LH1-N1M7-D were functional neomycin riboswitches. Northern blot analysis proved that LH1-N1M7-D acts as a transcriptional regulator, as intended. This demonstrates that our *in silico* approach can be successfully applied to novel aptamers.

The function of our artificial neomycin riboswitches depends on a stable 5' hairpin structure. A plausible explanation for this requirement is an increased resistance to mRNA degradation: the half-life of a mutant *ompA* transcript increases 3–4-fold when such a hairpin is added upstream of its single-stranded 5' UTR sequence [40]. ROSE elements – RNA thermometers in the 5' UTR controlling many small heat shock genes in *E. coli* – consist of at least two consecutive hairpins, and only the last one is thermosensitive and regulating the mRNA's translation rate [41]. In gram-negative bacteria including *E. coli*, mRNA degradation is carried out by the degradosome, a complex of several proteins including Ribonuclease E (RNase E), supported by RNA 5' pyrophosphohydrolase (RppH) [42, 43]. Stable 5' structures are known to effectively obstruct the enzymatic activity of RppH and, thus, of the degradosome, explaining mRNA stabilization [44]. Our data show that the effect of the 5' hairpin depends on its thermodynamic stability rather than its sequence, lending support to the hypothesis that stable 5' hairpins play a general protective role in the process of mRNA degradation. An increased life-time of the mRNA would also explains the observed correlation between thermodynamic stability and fluorescence intensity. In the positive control, we observed no dependence of fluorescence intensity on the leader hairpin. However, this construct forms a very stable hairpin (instead of the flexible aptamer structure), which likely explains its stability.

While the leader hairpin is crucial for our neomycin riboswitch, no corresponding structure was required for similar theophylline or tetracycline riboswitches [12, 13], cf. Fig. A.5 for an overview of their thermodynamic ground states. We suspect that this is the consequence of differences in aptamer stability. While the neomycin aptamer has an MFE of $\Delta G = -6.2\,\mathrm{kcal\,mol^{-1}}$ [9], both the theophylline aptamer used by Wachsmuth et al. [12] ($\Delta G = -12.1\,\mathrm{kcal\,mol^{-1}}$) and the tetracycline aptamer used by Domin et al. [13] ($\Delta G = -19.8\,\mathrm{kcal\,mol^{-1}}$) are considerably more stable. We therefore hypothesize that, in the presence of the ligand, a highly stable aptamer–ligand complex sufficiently protects the transcript from degradation. In absence of the ligand, the terminator hairpin forms rapidly and leaves the transcript with an unstructured 5' end, facilitating its quick degradation and thus contributing to a high activation ratio. The

observed change in fluorescence activity upon ligand binding is therefore a result of two independent effects: an increased transcription rate due to the suppression of the terminator hairpin formation, and a reduced transcript degradation due to the stabilization of the aptamer structure located at its 5' end.

Synthetic biology is concerned with solving sophisticated design problems with minimal expenditure of time, effort and money and thus strives to construct complex genetic circuits as a combination of fully modular, independently optimized components. Natural biological systems, however, have not evolved to adhere to the paradigm of engineering. It is well-documented, therefore, that the embedding of engineered circuits into living organisms faces a multitude of problems, mostly from unintended and often unexpected interaction with the natural context. Orthogonal systems attempt to avoid or at least minimize this problem [45]. Some quite spectacular success stories, e. g., the construction of re-engineered versions of secondary metabolite biosynthetic pathways [46] or the assembly of artificial operons [47] have raised the hope for a "plug-and-play" synthetic biology that allows the usage of pre-fabricated components in a combinatorial fashion. Even within an orthogonal system, however, it appears that this goal will remain elusive in many cases, at least in the strict sense of devising context-independent components [48]

Our data show that already conceptually very simple devices such as transcriptional riboswitches are more than a simple concatenation of building blocks. This is not surprising, given that the function of a transcriptional riboswitch is not just determined by the properties of aptamer and terminator, but depends on a delicate balance of mutually exclusive structural features and a carefully orchestrated time course of folding and refolding during its transcription [49]. As a consequence, prefabricated modules require adaptation and partial re-design to properly function in combination. In the case of RNA devices, the intrinsically global nature of RNA structure formation, in which base pairs are not restricted to local interaction, explains the imperfect modularity. At the same time, computational models of RNA folding are capable of capturing the undesired side effects and make it possible to algorithmically optimize novel designs based on existing components.

In previous work, for example, we observed that only one of two terminator hairpins functioned properly in the synthetic theophylline riboswitch RS10 [50]. Using a cotranscriptional folding algorithm, we found that the inactive construct harbored an intermediate structure that likely acted as a kinetic trap, delaying the formation of the terminator hairpin sufficiently to render it inactive.

The adaptation of the interaction between the modular components is, by itself, not necessarily sufficient, as the presented work shows. The neomycin switches depend on an obligatory 5' leader hairpin, presumably because the riboswitch itself is not sufficiently stable in itself to protect the mRNA from degradation. In addition, it needs to be adapted to the sequence context of the transcript to avoid interference of the leader sequence with the riboswitch domain – again an effect that can be captured by the RNA secondary structure model and remedied by adapting the constructs *in silico*.

Taken together, we conclude that complex RNA devices cannot be engineered by simply combining pre-optimized components in a plug-and-play fashion. Modular components can be combined and put into a functional context, however, with the help of *in silico* simulation techniques. We can therefore do better than "plug and pray": computationally, we can adapt the modules to function in context and optimize the constructs as a whole. We are confident that the reliability of the computational predictions will continue to improve as our understanding of the individual components and their underlying mechanisms of action evolves.

Acknowledgements

Disclosure of Interest

The authors report no conflict of interest.

List of Abbreviations

This work makes use of the following abbreviations:

au   arbitrary time units

AUR   aptamer upstream region

bgaB	β-galactosidase
bgaB	the β-galactosidase gene
dL	decoupling leader
E. coli	*Escherichia coli*
eGFP	the enhanced green fluorescent protein
egfp	the enhanced green fluorescent protein gene
GFP	green fluorescent protein
LH	leader hairpin
LM	mutated leader
MFE	minimum free energy
mRNA	messenger RNA
oL	original leader
Pos	positive control
RBS	ribosomal binding site
RFU	relative fluorescence unit
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase E	Ribonuclease E
RppH	RNA 5' pyrophosphohydrolase
rRNA	ribosomal RNA
SELEX	systematic evolution of ligands by exponential enrichment
SSC	saline sodium citrate buffer
TBE	Tris-borate-EDTA
TDR	terminator downstream region

TSS   transcription start site

U   unstructured region

UTR   untranslated region

List of Figures

List of Tables

References

[1] Ali Nahvi, Narasimhan Sudarsan, Margaret S. Ebert, Xiang Zou, Kenneth L. Brown, and Ronald R. Breaker. "Genetic control by a metabolite binding mRNA". In: *Chemistry & Biology* 9.9 (2002), p. 1043. ISSN: 10745521 (cit. on p. 3).

[2] Alexander Serganov and Evgeny Nudler. "A decade of riboswitches". In: *Cell* 152.1-2 (2013), pp. 17–24. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.12.024 (cit. on p. 3).

[3] Ronald R. Breaker. "Riboswitches and the RNA world". In: *Cold Spring Harbor perspectives in biology* 4.2 (2012). ISSN: 1943-0264. DOI: 10.1101/cshperspect.a003566 (cit. on p. 3).

[4]   Casey C. Fowler, Eric D. Brown, and Yingfu Li. "A FACS-based approach to engineering artificial riboswitches." In: *ChemBioChem* 9.12 (2008), pp. 1906–1911. DOI: 10.1002/cbic.200700713 (cit. on p. 3).

[5]   Maja Etzel and Mario Mörl. "Synthetic Riboswitches: From Plug and Pray toward Plug and Play". In: *Biochemistry* 56.9 (Mar. 2017), pp. 1181–1198. ISSN: 0006-2960. DOI: 10.1021/acs.biochem.6b01218 (cit. on p. 3).

[6]   Andrew D. Ellington and Jack W. Szostak. "*In vitro* selection of RNA molecules that bind specific ligands". In: *nature* 346.6287 (1990), p. 818 (cit. on p. 3).

[7]   Craig Tuerk and Larry Gold. "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase". In: *Science* 249.4968 (1990), pp. 505–510 (cit. on p. 3).

[8]   Julia E. Weigand, Martin Sanchez, Ewald-Bernd Gunnesch, Sabrina Zeiher, Reneé Schroeder, and Beatrix Suess. "Screening for engineered neomycin riboswitches that control translation initiation". In: *RNA* 14.1 (2008), pp. 89–97. DOI: 10.1261/rna.772408 (cit. on pp. 3, 6).

[9]   Julia E. Weigand, Sina R. Schmidtke, Tristan J. Will, Elke Duchardt-Ferner, Christian Hammann, Jens Wohnert, and Beatrix Suess. "Mechanistic insights into an engineered riboswitch: a switching element which confers riboswitch activity". In: *Nucleic Acids Research* 39.8 (2011), pp. 3363–3372. ISSN: 0305-1048. DOI: 10.1093/nar/gkq946 (cit. on pp. 3, 6, 18).

[10] Ivan Gusarov and Evgeny Nudler. "The Mechanism of Intrinsic Transcription Termination". In: *Molecular Cell* 3.4 (1999), pp. 495–504. DOI: 10.1016/S1097-2765(00)80477-3 (cit. on p. 3).

[11] Jason M. Peters, Abbey D. Vangeloff, and Robert Landick. "Bacterial transcription terminators: the RNA 3'-end chronicles". In: *Journal of Molecular Biology* 412.5 (2011), pp. 793–813. ISSN: 1089-8638. DOI: 10.1016/j.jmb.2011.03.036 (cit. on p. 3).

[12] Manja Wachsmuth, Sven Findeiß, Nadine Weissheimer, Peter F. Stadler, and Mario Mörl. "*De novo* design of a synthetic riboswitch that regulates transcription termination". In: *Nucleic Acids Research* 41.4 (Feb. 2013), pp. 2541–2551. DOI: 10.1093/nar/gks1330 (cit. on pp. 4, 8, 9, 11, 13, 18, 33).

[13] Gesine Domin, Sven Findeiß, Manja Wachsmuth, Sebastian Will, Peter F. Stadler, and Mario Mörl. "Applicability of a computational design approach for synthetic riboswitches". In: *Nucleic Acids Res* 45.7 (Apr. 2017), pp. 4108–4119. ISSN: 0305-1048. DOI: 10.1093/nar/gkw1267 (cit. on pp. 4, 18, 33).

[14] Selwyn Quan, Ole Skovgaard, Robert E. McLaughlin, Ed T. Buurman, and Catherine L. Squires. "Markerless *Escherichia coli* rrn Deletion Strains for Genetic Determination of Ribosomal Binding Sites". In: *G3 (Bethesda, Md.)* 5.12 (2015), pp. 2555–2557. ISSN: 2160-1836. DOI: 10.1534/g3.115.022301 (cit. on pp. 4, 11).

[15] Vandana Chakravartty and John E. Cronan. "A series of medium and high copy number arabinose-inducible *Escherichia coli* expression vectors compatible with pBR322 and pACYC184". In: *Plasmid* 81 (2015), pp. 21–26. ISSN: 1095-9890. DOI: 10.1016/j.plasmid.2015.03.001 (cit. on p. 4).

[16] Jiren Zhang and Murray P. Deutscher. "A uridine-rich sequence required for translation of prokaryotic mRNA". en. In: *PNAS* 89.7 (Apr. 1992), pp. 2605–2609. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.89.7.2605 (cit. on p. 5).

[17] Caroline Köhrer and Uttam L. RajBhandary. "The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs". In: *Methods (San Diego, Calif.)* 44.2 (2008), pp. 129–138. ISSN: 1095-9130. DOI: 10.1016/j.ymeth.2007.10.006 (cit. on p. 5).

[18] Douglas H. Turner and David H. Mathews. "NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure." In: *Nucleic Acids Res* 38.Database issue (2010), pp. D280–D282. DOI: 10.1093/nar/gkp892 (cit. on p. 7).

[19] Ivo L. Hofacker, Walter Fontana, Peter F. Stadler, L. Sebastian Bonhoeffer, Manfred Tacker, and Peter Schuster. "Fast Folding and Comparison of RNA Secondary Structures". en. In: *Monatshefte für Chemie / Chemical Monthly* 125.2 (Feb. 1994), pp. 167–188. ISSN: 0026-9247, 1434-4475. DOI: 10.1007/BF00818163 (cit. on p. 7).

[20] Ronny Lorenz, Stephan H. Bernhart, Christian Höner zu Siederdissen Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. "ViennaRNA Package 2.0". In: *Algorithms for Molecular Biology* 6.1 (Nov. 2011), p. 26. DOI: 10.1186/1748-7188-6-26 (cit. on pp. 7, 29).

[21] Ronny Lorenz, Ivo L. Hofacker, and Peter F. Stadler. "RNA folding with hard and soft constraints". In: *Algorithms for Molecular Biology* 11.1 (2016). DOI: `10.1186/s13015-016-0070-z` (cit. on p. 7).

[22] Dietmar Pörschke, Olke C. Uhlenbeck, and Francis H. Martin. "Thermodynamics and kinetics of the helix–coil transition of oligomers containing GC base pairs". In: *Biopolymers: Original Research on Biomolecules* 12.6 (1973), pp. 1313–1335 (cit. on pp. 7, 9).

[23] Srividya Mohan, Chiaolong Hsiao, Halena VanDeusen, Ryan Gallagher, Eric Krohn, Benson Kalahar, Roger M. Wartell, and Loren D. Williams. "Mechanism of RNA Double Helix-Propagation at Atomic Resolution". In: *The Journal of Physical Chemistry B* 113.9 (Mar. 5, 2009), pp. 2614–2623. ISSN: 1520-6106. DOI: `10.1021/jp8039884` (cit. on pp. 7, 9).

[24] Philip C. Bevilacqua and Joshua M. Blose. "Structures, kinetics, thermodynamics, and biological functions of RNA hairpins". In: *Annu Rev Phys Chem* 59 (2008), pp. 79–103. DOI: `10.1146/annurev.physchem.59.032607.093743` (cit. on p. 7).

[25] Artem V. Melnykov, Rajesh K. Nayak, Kathleen B. Hall, and Alan Van Orden. "Effect of loop composition on the stability and folding kinetics of RNA hairpins with large loops". In: *Biochemistry* 54 (2015), pp. 1886–1896. DOI: `10.1021/bi5014276` (cit. on p. 7).

[26] Ulla Vogel and Kaj F. Jensen. "The RNA chain elongation rate in *Escherichia coli* depends on the growth rate." In: *Journal of Bacteriology* 176.10 (May 1, 1994), pp. 2807–2813. ISSN: 0021-9193, 1098-5530. DOI: `10.1128/jb.176.10.2807-2813.1994` (cit. on p. 7).

[27] J. Ryals, R. Little, and Hans Bremer. "Temperature dependence of RNA synthesis parameters in *Escherichia coli*." In: *Journal of Bacteriology* 151.2 (Aug. 1, 1982), pp. 879–887. ISSN: 0021-9193, 1098-5530 (cit. on p. 7).

[28] Daniel Lai, Jeff R. Proctor, and Irmtraud M. Meyer. "On the importance of cotranscriptional RNA structure formation". In: *RNA* 19 (2013), pp. 1461–1473. DOI: `10.1261/rna.037390.112` (cit. on p. 7).

[29] Benjamin Lutz, Michael Faber, Abhinav Verma, Stefan Klumpp, and Alexander Schug. "Differences between cotranscriptional and free riboswitch folding". In: *Nucleic Acids Res* 42 (2014), pp. 2687–2696. DOI: `10.1093/nar/gkt1213` (cit. on p. 7).

[30]   Christoph Flamm, Walter Fontana, Ivo L. Hofacker, and Peter Schuster. "RNA folding at elementary step resolution". In: *RNA* 6.3 (2000), pp. 325–338. DOI: 10.1017/S1355838200992161 (cit. on p. 7).

[31]   A. Xayaphoummine, T. Bucher, and H. Isambert. "Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots". In: *Nucleic Acids Research* 33.suppl-2 (July 2005), W605–W610. ISSN: 0305-1048. DOI: 10.1093/nar/gki447 (cit. on p. 7).

[32]   Michael Geis, Christoph Flamm, Michael T. Wolfinger, Andrea Tanzer, Ivo L. Hofacker, Martin Middendorf, Christian Mandl, Peter F. Stadler, and Caroline Thurner. "Folding Kinetics of Large RNAs". In: *Journal of Molecular Biology* 379.1 (2008), 160–173. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2008.02.064 (cit. on p. 7).

[33]   Ivo L. Hofacker, Christoph Flamm, Christian Heine, Michael T. Wolfinger, Gerik Scheuermann, and Peter F. Stadler. "BarMap: RNA folding on dynamic energy landscapes". In: *RNA* 16.7 (July 1, 2010), pp. 1308–1316. ISSN: 1355-8382. DOI: 10.1261/rna.2093310 (cit. on pp. 7, 8).

[34]   Stefan Wuchty, Walter Fontana, Ivo L. Hofacker, and Peter Schuster. "Complete suboptimal folding of RNA and the stability of secondary structures." In: *Biopolymers* 49.2 (Feb. 1999), pp. 145–65. DOI: 10.1002/(SICI)1097-0282(199902)49:2$<$145::AID-BIP4$>$3.0.CO;2-G (cit. on p. 7).

[35]   Christoph Flamm, Ivo L. Hofacker, Peter F. Stadler, and Michael T. Wolfinger. "Barrier Trees of Degenerate Landscapes". In: *Zeitschrift für Physikalische Chemie* 216.2/2002 (Jan. 2002). DOI: 10.1524/zpch.2002.216.2.155 (cit. on p. 7).

[36]   Felix Kühnl, Peter F. Stadler, and Sven Findeiß. "Assessing the Quality of Cotranscriptional Folding Simulations". In: ed. by Ronny Lorenz. submitted. Berlin: Springer, 2019 (cit. on p. 8).

[37]   The Grace contributors. *Grace – a WYSIWYG 2D plotting tool for the X Window System and M*tif.* [Online; accessed 10-September-2019]. 2015. URL: http://plasma-gate.weizmann.ac.il/Grace/ (cit. on p. 8).

[38]   Ananya Ray-Soni, Michael J. Bellecourt, and Robert Landick. "Mechanisms of Bacterial Transcription Termination: All Good Things Must End". In: *Annual Review of Biochemistry* 85.1 (June 2016), pp. 319–347. ISSN: 0066-4154, 1545-4509. DOI: 10.1146/annurev-biochem-060815-014844 (cit. on p. 9).

[39] Scott Kirkpatrick, C. Daniel Gelatt, and Mario P. Vecchi. "Optimization by simulated annealing". In: *science* 220.4598 (1983), pp. 671–680 (cit. on p. 10).

[40] Sheri A. Emory, Philippe Bouvet, and Joel G. Belasco. "A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*." en. In: *Genes Dev.* 6.1 (Jan. 1992), pp. 135–148. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.6.1.135 (cit. on p. 18).

[41] Stefanie S. Krajewski and Franz Narberhaus. "Temperature-driven differential gene expression by RNA thermosensors". In: *Biochimica et Biophysica Acta – Gene Regulatory Mechanisms*. Riboswitches 1839.10 (Oct. 2014), pp. 978–988. ISSN: 1874-9399. DOI: 10.1016/j.bbagrm.2014.03.006 (cit. on p. 18).

[42] George A. Mackie. "Ribonuclease E is a 5'-end-dependent endonuclease". In: *Nature* 395.6703 (1998), pp. 720–723. ISSN: 0028-0836. DOI: 10.1038/27246 (cit. on p. 18).

[43] Xunqing Jiang, Alexis Diwa, and Joel G. Belasco. "Regions of RNase E Important for 5'-End-Dependent RNA Cleavage and Autoregulated Synthesis". In: *Journal of Bacteriology* 182.9 (2000), pp. 2468–2475. ISSN: 0021-9193. DOI: 10.1128/JB.182.9.2468-2475.2000 (cit. on p. 18).

[44] Atilio Deana, Helena Celesnik, and Joel G. Belasco. "The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal". In: *Nature* 451.7176 (2008), p. 355. ISSN: 0028-0836. DOI: 10.1038/nature06475 (cit. on p. 18).

[45] Jordan K. Villa, Yichi Su, Lydia M. Contreras, and Ming C. Hammond. "Synthetic Biology of Small RNAs and Riboswitches". In: *Regulating with RNA in Bacteria and Archaea*. Vol. 6. 3. American Society of Microbiology, 2019, pp. 527–545. DOI: 10.1128/microbiolspec.RWR-0007-2017 (cit. on p. 19).

[46] Marnix H. Medema, Rainer Breitling, Roel Bovenberg, and Eriko Takano. "Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms". In: *Nature Rev Microbiol* 9 (2011), pp. 131–137. DOI: 10.1038/nrmicro2478 (cit. on p. 19).

[47] Patrick Basitta, Lucia Westrich, Manuela Rösch, Andreas Kulik, Bertolt Gust, and Alexander Kristian Apel. "AGOS: A Plug-and-Play Method for the Assembly of Artificial Gene Operons into Functional Biosynthetic Gene Clusters". In: *ACS Synth. Biol.* 6 (5 2017), pp. 817–825. DOI: 10.1021/acssynbio.6b00319 (cit. on p. 19).

[48] E. Karamasioti, C. Lormeau, and J. Stelling. "Computational design of biological circuits: putting parts into context". In: *Mol. Systems Design Eng.* 2 (4 2017), pp. 410–421. DOI: `10.1039/c7me00032d` (cit. on p. 19).

[49] Giulio Quarta, Ken Sin, and Tamar Schlick. "Dynamic Energy Landscapes of Riboswitches Help Interpret Conformational Rearrangements and Function". In: *PLOS Computational Biology* 8.2 (Feb. 16, 2012), e1002368. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1002368` (cit. on p. 19).

[50] Manja Wachsmuth, Gesine Domin, Ronny Lorenz, Robert Serfling, Sven Findeiß, Peter F. Stadler, and Mario Mörl. "Design criteria for synthetic riboswitches acting on transcription". In: *RNA Biol* 12.2 (Feb. 2015), pp. 221–231. DOI: `10.1080/15476286.2015.1017235` (cit. on p. 19).

[51] Peter Kerpedjiev, Stefan Hammer, and Ivo L. Hofacker. "Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams". In: *Bioinformatics* 31.20 (Oct. 15, 2015), pp. 3377–3379. ISSN: 1367-4803. DOI: `10.1093/bioinformatics/btv372` (cit. on p. 33).

A. Supplementary Figures
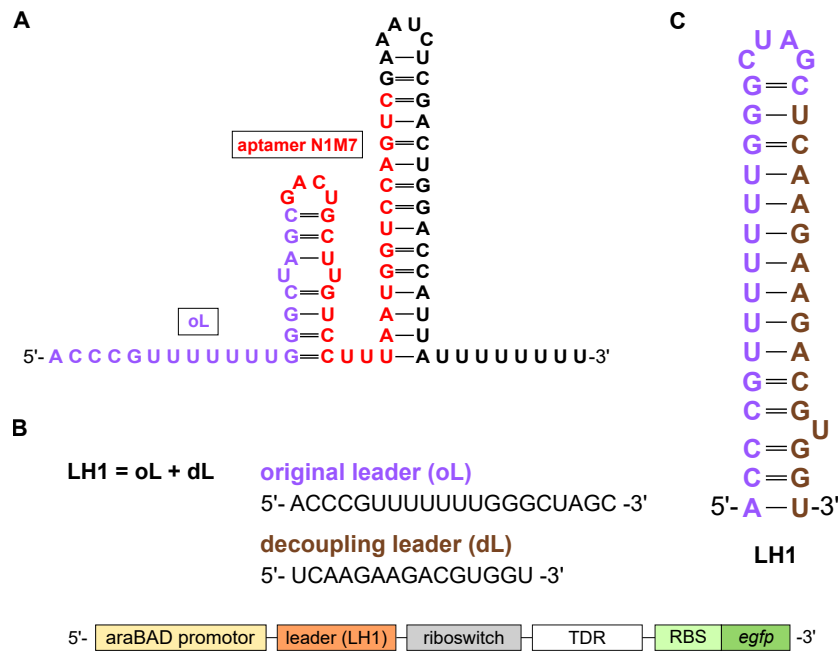
Figure A.1: A decoupling leader for riboswitches. (*A*) Secondary structure of riboswitch N1M7-D with the original leader (oL) as computed by free energy minimization [20]. (*B*) Schematic of riboswitch-surrounding elements in the plasmid pRSF1030Tp-SD-eGFP. A decoupling leader was inserted downstream of oL to protect the riboswitch from leader-dependent misfolding by forming the stable hairpin LH1 depicted in (*C*).

Figure A.2: Secondary structure of the riboswitch constructs with the aptamer N1M7 used for the eGFP assay in Fig. 4.

A

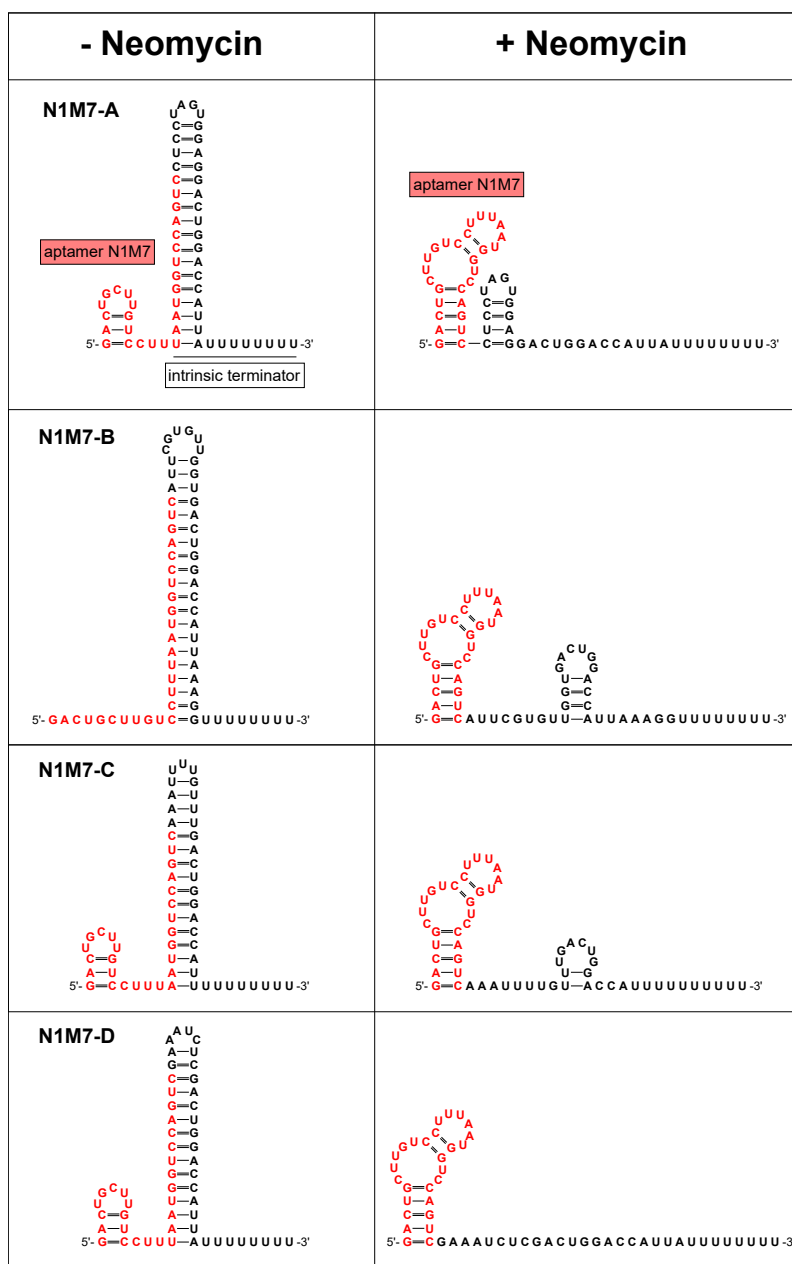| LM3	LM3C	LM4	LM5
U^AA	U^AA	U^AA	U^AA
C=G	C=G	C　C	C　C
U—A	U—A	U　U	U　U
C=G	C=G	C=G	C　C
U—A	U—A	U—A	U—A
G　G	C=G	G=C	G=C
C　C	G=C	C=G	C=G
A—U	A—U	A—U	A—U
G=C	G=C	G=C	G=C
5'-U—AA-3'	5'-U—AA-3'	5'-U—AA-3'	5'-U—AA-3'
-5.50 kcal/mol	**-13.90 kcal/mol**	**-10.10 kcal/mol**	**-5.30 kcal/mol**

B
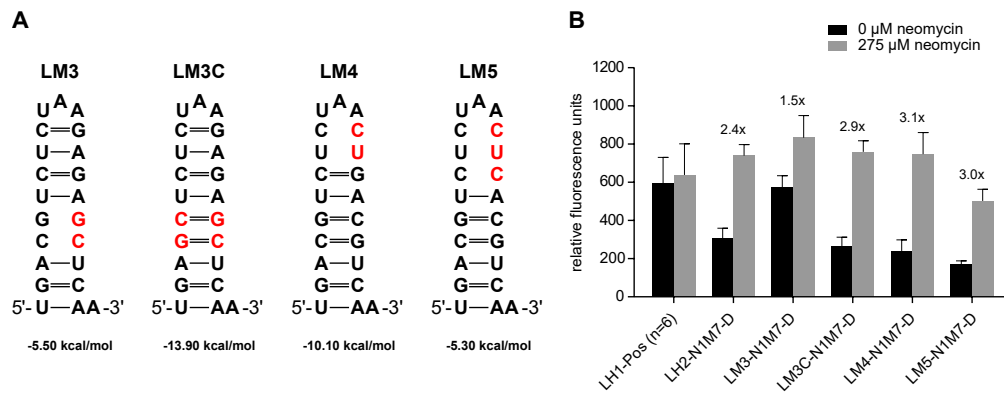
Figure A.3: Impact of point mutations in the leader hairpin LH2. Fluorescence intensity of the modified leaders LM3, LM3C, LM4 and LM5, folded into their predicted secondary structure (*A*), in conjunction with N1M7-D (*B*). LH2-N1M7-D and LH1-Pos were used as control. Legend as described in Fig. 4.

Figure A.4: Influence of a sequence predicted to interfere with N1M7-D by *in silico* methods. The fluorescence intensity of N1M7-D with original leader (oL) in conjunction with LH1 or LH2 is shown. Legend as described in Fig. 4.

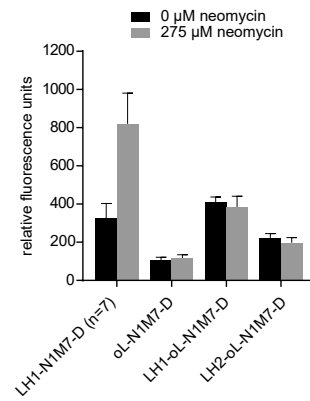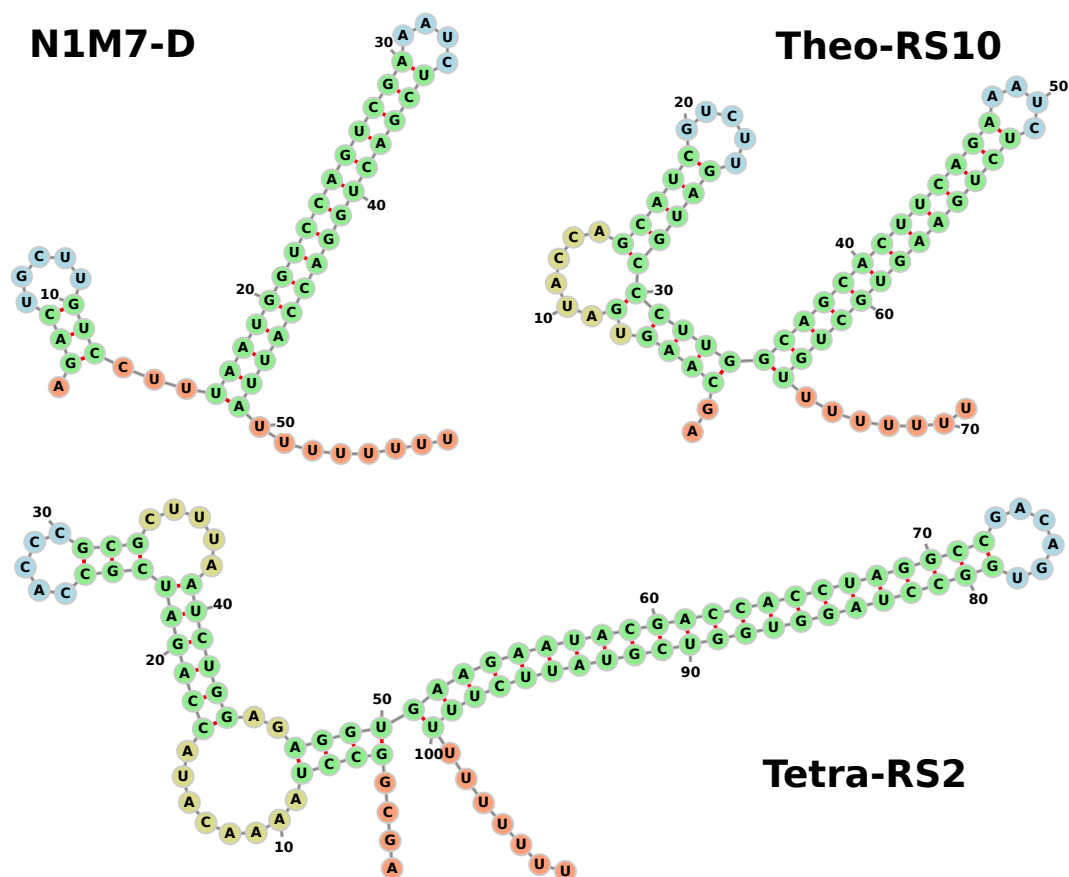Figure A.5: Minimum free energy secondary structures of the neomycin riboswitch N1M7-D, the theophylline riboswitch RS10 [12], and the tetracycline riboswitch RS2 [13] drawn with *forna* [51]. Colors mark different structural features (hairpin loops in blue, interior loops and bulges in light brown, stems in green, and exterior loops in red). All riboswitches display a long terminator hairpin followed by an 8 nt poly-U stretch at their 3' end. In contrast to the other constructs shown, N1M7-D exhibits a remarkably small structure at its 5' end.

B. Supplementary Data

Table B.1: Overview of all constructs analysed in this work and which figures they appear in, along with their corresponding fluorescence intensity data (mean, standard deviation (SD), and the number of repetitions (N)). Sequence data can be found in a supplementary spreadsheet file.

		- neomycin			+ neomycin		
		mean	SD	N	mean	SD	N
Figure 4B	Pos	493.32	133.42	6			
	LH1-N1M7-A	54.77	2.96	3	63.86	4.61	3
	LH1-N1M7-B	712.27	100.31	3	812.05	109.73	3
	LH1-N1M7-C	528.68	108.21	3	963.45	83.19	3
	LH1-N1M7-D	342.28	21.67	3	899.39	121.33	3
	LH1-Term-N1M7-D	207.54	29.29	3	207.92	19.49	3
Figure 5C	Pos	703.90	33.10	3	676.98	33.92	3
	LH1-Pos	724.84	65.92	6	779.09	56.16	6
	N1M7-D	467.18	132.07	3	504.81	101.70	3
	LH1-N1M7-D	214.95	18.29	3	704.86	24.39	3
	LH2-N1M7-D	304.37	54.78	3	739.21	57.94	3
Figure 6C	Pos	662.93	79.58	3			
	U1-N1M7-D	121.50	4.84	3	189.22	24.45	3
	U2-N1M7-D	213.61	25.92	3	296.02	21.88	3
	LH1-U1-N1M7-D	525.68	153.12	3	880.53	195.61	3
	LH1-U2-N1M7-D	609.89	176.20	3	903.36	185.90	3
Figure 7B	LH1-Pos	691.41	19.95	3	762.86	55.21	3
	LH1-N1M7-D	214.95	18.29	3	704.86	24.39	3
	LM1-N1M7-D	193.40	4.10	3	665.74	144.60	3
	LM1C-N1M7-D	275.94	51.94	3	876.00	171.67	3
	LM2-N1M7-D	82.61	10.98	3	229.10	34.66	3
	LM2C-N1M7-D	244.70	3.76	3	803.76	248.15	3
Figure 7C	LH1-Pos	758.27	84.34	3	795.33	63.61	3
	LM1-Pos	690.96	46.96	3	707.80	50.35	3
	LM1C-Pos	855.49	90.97	3	835.45	76.11	3
	LM2-Pos	392.54	19.53	3	403.31	18.43	3
	LM2C-Pos	800.75	192.71	3	801.90	206.77	3
Figure A.3	LH1-Pos	595.96	134.81	6	638.34	163.03	6
	LH2-N1M7-D	304.37	54.78	3	739.21	57.94	3
	LM3-N1M7-D	574.30	60.17	3	833.45	116.35	3
	LM3C-N1M7-D	266.00	46.67	3	759.79	57.27	3
	LM4-N1M7-D	240.69	57.47	3	747.31	112.81	3
	LM5-N1M7-D	170.10	18.16	3	502.81	60.70	3
Figure A.4	LH1-N1M7-D	321.31	78.80	7	818.07	161.59	7
	oL-N1M7-D	102.61	16.15	3	114.33	17.41	3
	LH1-oL-N1M7-D	408.52	25.88	3	382.05	56.92	3
	LH2-oL-N1M7-D	218.60	23.81	3	194.26	27.54	3

Table B.2: Effects of mutations on the stability of leader hairpins LH1 and LH2. The stability is given as minimum free energy (MFE) of the entire (mutated) hairpin. Activities of the resulting constructs are shown in Fig. 7 and Fig. A.3.

Hairpin	Mutations	MFE ($\mathrm{kcal\,mol^{-1}}$)
LH1	none	-15.9
LM1	G28C, A29U	-10.9
LM1C	G28C, A29U, U6A, U7G	-18.5
LM2	U21C, C22G,	-9.7
LM2C	U21G, C22G, G13C, G14C	-18.3
LH2	none	-13.1
LM3	C17G, G18C	-5.5
LM3C	C17G, G18C, C3G, G4C	-13.9
LM4	G13C, A14U	-10.1
LM5	G13C, A14U, G15C	-5.3