De novo design of a synthetic riboswitch that regulates transcription termination

Manja Wachsmuth[1,8], Sven Findeiß[2,3,8], Nadine Weissheimer[1], Peter F Stadler[3-7], Mario Mörl[1*]

[1] University of Leipzig, Institute for Biochemistry, Leipzig, Germany

[2] University of Vienna, Research group Bioinformatics and Computational Biology, Vienna, Austria

[3] University of Vienna, Institute for Theoretical Chemistry, Vienna, Austria

[4] University of Leipzig, Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, Leipzig, Germany

[5] Max Planck Institute for Mathematics in the Science, Leipzig, Germany

[6] Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany

[7] Santa Fe Institute, Santa Fe, USA

[8] The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

* To whom correspondence should be addressed. Tel: +49(0)341-9736-911; Fax: +49(0)341-9736-919; Email: moerl@uni-leipzig.de

ABSTRACT

Riboswitches are regulatory RNA elements typically located in the 5'-untranslated region of certain mRNAs and control gene expression at the level of transcription or translation. These elements consist of a sensor and an adjacent actuator domain. The sensor usually is an aptamer that specifically interacts with a ligand. The actuator contains an intrinsic terminator or a ribosomal binding site for transcriptional or translational regulation, respectively. Ligand binding leads to structural rearrangements of the riboswitch and to presentation or masking of these regulatory elements. Based on this modular organization, riboswitches are an ideal target for constructing synthetic regulatory systems for gene expression. While riboswitches for translational control have been designed successfully, attempts to construct synthetic elements regulating transcription have failed so far. Here, we present an *in silico* pipeline for the rational design of synthetic riboswitches that regulate gene expression at the transcriptional level. Using the well-characterized theophylline-aptamer as sensor, we designed the actuator part as RNA sequences that can fold into functional intrinsic terminator structures. In the biochemical characterization, several of the designed constructs show ligand-dependent control of gene expression in *E. coli*, demonstrating that it is possible to engineer riboswitches not only for translational but also for transcriptional regulation.

INTRODUCTION

Synthetic biology has become a rapidly developing branch of life sciences. At its heart is the rational design and construction of biosynthetic and regulatory networks, pathways and their constituent components. As RNA plays a central role in a diverse set of regulatory pathways, efforts in synthetic biology have produced novel RNA components capable of regulating gene expression *in vivo* {Isaacs 2006 #71}. Riboswitches are a particularly attractive type of gadget for synthetic biology applications {Wieland 2008 #106}{Topp 2010 #107}. These naturally occurring RNA elements determine the expression state of a gene in response to an external signal (in most cases a small molecule that is specifically recognized) through a conformational change, regulating either transcription or translation. The fundamental architecture of a riboswitch combines a sensor (aptamer domain) for the external stimulus and an actuator that influences gene expression. Both elements can in principle be freely combined. Depending on the response to the trigger molecule, gene expression is either enhanced (ON switch) or repressed (OFF switch). In transcription-regulating riboswitches, the actuator element contains a possible intrinsic transcription terminator structure consisting of a stem-loop element followed by a U-rich region. Depending on the bound or unbound state of the sensor, this terminator structure is either masked or formed, leading to activation or inhibition of gene expression, respectively.

While structural features of the conserved aptamer domains are well described, the organization of the regulatory platforms is in many cases unclear. In particular, transcription terminator elements can vary substantially, not only as isolated elements but also in the context of a specific riboswitch {Breaker 2011 #69}{Hoon 2005 #50}{Lesnik 2001 #51}. Hence, the design of transcription regulating riboswitches is a difficult task. This was demonstrated by Fowler et al., who intended

to design transcriptional ON switches by fusing a theophylline aptamer to a *B. subtilis* intrinsic terminator {Fowler 2008 #131}. Detailed analysis revealed, however, that the regulatory mechanism of the selected riboswitches was independent of the terminator structure and was probably targeting translation. The sole artificial combination of aptamer and terminator elements that showed transcriptional regulation consists of two naturally occurring components, where the aptamer region of the *Streptococcus mutans* THF riboswitch was fused to the intrinsic terminator region of the *Bacillus subtilis* metE riboswitch. This resulted in an artificial ON switch that controls transcription termination upon folinic acid binding {Trausch 2011 #27}. Accordingly, currently available riboswitches consisting of synthetic components generated by rational and/or evolutionary approaches exclusively regulate translation {Hanson 2005 #2}{Ogawa 2011 #45}{Beisel 2009 #117}{Sinha 2010 #112}{Muranaka 2009 #35}{Wieland 2009 #113}.

We used the well characterized theophylline aptamer as a sensor element to construct synthetic riboswitches that regulate gene expression at the transcriptional level. This aptamer has been successfully incorporated in different regulatory mechanisms for gene expression *in vivo*, demonstrating its functionality within the cell. Besides representing a sensor platform in translational riboswitches {Desai 2004 #1}{Topp 2007 #34}{Suess 2004 #3}, it was used to modulate mRNA cleavage catalysed by yeast RNase III {Babiskin 2011 #141} as well as RNA splicing {Kim 2005 #142}{Thompson 2002 #143} and to control non-coding RNA activity {Qi 2012 #146}. Fusing this aptamer to a computer-designed spacer element followed by an aptamer-complementary sequence, which together form a transcriptional terminator structure, several synthetic riboswitches were generated. Biochemical analysis revealed that these elements are indeed able to efficiently regulate transcription of a reporter gene. Hence, while several *in vivo* screening approaches have failed so far,

our data show that a strategy based on *in silico* design is successful in constructing functional transcriptional riboswitches in *E. coli*.

MATERIAL AND METHODS

Riboswitch design

The rational design algorithm starts from the known sequence and secondary structure of the aptamer and constructs the sequence of the remaining part of the riboswitch, i.e. the spacer and the terminator. It proceeds by proposing a large set of candidates that are then evaluated with respect to their secondary structure and properties of their folding paths which are estimated using the RNAfold program of the Vienna RNA package {Lorenz 2011}. At the proposal stage it incorporates some knowledge of sequence and structure of terminator elements for the given biological system in which the construct is supposed to function, here *E. coli*. In particular, we used the fact that the terminator hairpin must have a minimal size and that it is followed by a U-stretch. First, a spacer database containing randomly generated sequences with lengths between 6 and 20 nucleotides is created. The size of this database is user defined. For each position k of the 3'-part of the aptamer we form the sequence _ that is complementary to the subsequence from k to the end of the aptamer. For the theophylline aptamer, values of k ranging from position 21 to 32 were used. The sequences of the aptamer, a spacer from the database, a complementary sequence _, and the U-stretch are concatenated to form a complete riboswitch candidate.

Each of these constructs is then evaluated by folding simulations. In the current implementation, folding paths are represented as a sequence of secondary structures computed for sub-sequences starting at the 5'-end. We used individual transcription steps of 5 to 10 nucleotides to simulate co-transcriptional folding with

varying elongation speed. Secondary structures are computed by RNAfold, a component of the Vienna RNA Package {Lorenz 2011}, with parameter settings -d2 --noPS --noLP. If one of the transcription intermediates forms base pairs between aptamer and spacer, it is likely that this will interfere with the ligand-binding properties; hence such a candidate is rejected. On the other hand, if the full length transcript does not contain a single hairpin structure composed of the aptamer 3'-part, the spacer and the sequence _ complementary to the 3'-part of the aptamer, the construct cannot form a functional terminator and is also rejected. For each construct that passes these filters, additional features are calculated that are used as further selection criteria. The energy difference between the MFE (minimum free energy) fold of the full length construct and a structure where the aptamer region is constrained to form the ligand-binding competent fold is calculated. This parameter provides a rough measure for the stabilizing effect required to prevent terminator formation when the ligand is bound. Furthermore, a z-score of the structure downstream of the constrained aptamer fold is estimated as follows: The stability (MFE) of the candidate sequence is compared to the mean μ and standard deviation σ of the folding energies computed for a set of 1000 di-nucleotide shuffled sequences. A positive value of the z-score ($z = (MFE - \mu)/\sigma$) indicates a structure less stable than expected by chance. This is used to estimate whether stable local structures downstream of the constrained aptamer fold might interfere with transcription.

The six experimentally tested synthetic riboswitch constructs are summarized in Figure 2A. Secondary structures and free energy values of individual sequences were calculated using RNAfold 2.0.7.

Given the dissociation constant K_d, the binding energy is given by: $\Delta G = RT\ln K_d$. With T = 298 K, R = 1.98717 cal/K and the experimentally determined value K_d =

0.32 µM {Jenison 1994 #46} for the theophylline aptamer, we obtain ΔG = -8.86 kcal/mol for the stabilization of the aptamer structure.

Chemicals

Oligonucleotides were purchased from biomers.net and dNTP were obtained from Jena Biosciences. LB medium was purchased from Applichem, theophylline and caffeine were obtained from Sigma-Aldrich. Ampicillin and Arabinose were purchased from Carl Roth.

Plasmid construction

Riboswitch constructs were generated by overlap extension PCR and fused to bgaB reporter in pBAD2_bgaB {Klinkert 2012 #137} by QuickChange site directed mutagenesis (Stratagene). To analyze riboswitch performance in a second promoter/reporter gene system, riboswitch RS10 was subcloned into the 5'-untranslated region of plasmid pBSU0 {Klinkert 2012 #137} using an SpeI restriction site, leading to plasmid pBSU0_RS10. Plasmid constructs were isolated using GeneJET™ Plasmid Miniprep Kit (Fermentas). Correctness of individual assemblies was verified by sequencing.

Quantitative analysis for β-galactosidase activity and GFP fluorescence

Enzymatic activity of the expressed β-galactosidase was determined at 40°C {Waldminghaus 2007 #49}. All activity tests were performed in three independent experiments.

For analysis of GFP expression, Top10 cells holding pBSU0_RS10 were grown over night in the presence of 0, 0.5, 1 or 2 mM theophylline. GFP emission was measured in Sorensen's phosphate buffer (pH 8.0) at 509 nm with excitation wavelength of 488

nm. Fluorescence intensity was normalized by OD600 and measured for all theophylline concentrations in three individual experiments.

Northern Blot analysis

E. coli Top10 cells were grown in the presence or absence of 2 mM theophylline until OD_{600} of 0.5-0.6 in 25 ml LB medium supplied with ampicillin and arabinose. Total RNA was extracted according to J. Lawrence (pers. communication; http://rothlab.ucdavis.edu/protocols/D50.html).

Northern Blot was performed using the NorthernMax Formaldehyd-Based System for Northern blots (Ambion) according to manufacturer's instructions. 5 µg total RNA were separated on 1 % agarose formaldehyde gel and transferred to BrightStar™-Plus Membrane (Ambion).

DNA oligonucleotides (reverse complementary to target sequences) were radioactively labeled using γ-^{32}P-ATP (Hartmann Analytic). The primer sequences used are:

bgaB: 5'-GTTCGATCTTGCTCCAACTG-3'

23S rRNA: 5'-ACGACGGACGTTAGCACCCG-3'

For quantification of the mRNA levels northern blot analysis was repeated twice. Intensities of bgaB signals were normalized using the 23S rRNA signals. The intensity of the mRNA level in the absence of theophylline was defined as 100 %.

RESULTS

We devised an *in silico* selection pipeline based on the structure prediction methods implemented in the RNAfold program of the Vienna RNA package {Lorenz 2011 #135} to rationally design and select transcriptional riboswitches. Following the general paradigm of inverse folding algorithms {Andronescu 2004 #79}{Busch 2006

#162}{Hofacker 1994 #78}{Zadeh 2011 #164}, it operates by iteratively proposing candidate constructs and evaluating them with respect to the design goals. Each designed construct is composed of (i) the theophylline aptamer TCT8-4 that selectively binds its target at a K_d of 0.32 µM and strongly discriminates against the structurally closely related caffeine {Jenison 1994 #46}, (ii) a spacer sequence, (iii) a sequence complementary to the 3'-part of the aptamer and (iv) a U-stretch (Fig. 1 A). The rational model behind such a construct is to terminate transcription of a host gene if the target molecule theophylline is absent. This is achieved by the formation of a terminator structure composed of the aptamer 3'-part and elements (ii), (iii) and (iv). Ligand binding stabilizes the aptamer structure. Thus, the 3'-end of the aptamer sequence is unavailable for the formation of the terminator. As a consequence, the complete sequence of the host gene is transcribed and subsequently translated. Accordingly, a functional construct represents a transcriptional ON riboswitch.

The functional role of the spacer sequence is twofold. First, it has to ensure that the binding-competent aptamer structure is formed and has sufficient time during transcription to sense the target molecule. Second, it has to provide a functional part within the terminator structure. These requirements translate to specific constraints on both composition and length of a functional spacer sequence. The spacer must not form base pairs with the aptamer sequence during transcription in order to avoid potential inhibition of target molecule sensing. This implies a structural constraint. The design algorithm starts from randomly generated spacer sequences with a length between 6 and 20 nt and employs folding simulations on the concatenation of aptamer and spacer to remove candidates that violate the described constraints. The reverse complement of the 3'-part of the aptamer followed by an eight nucleotide long U-stretch is appended to the valid aptamer-spacer constructs, completing the possible intrinsic terminator structure. The co-transcriptional folding paths of these

complete constructs are simulated using RNAfold predictions {Lorenz 2011 #135} and construct rejection is done based on predicted secondary structures that violate certain constraints of the riboswitch candidate.

Based on the design principles described above, twelve theophylline sensing riboswitches have been constructed. From these predictions, six riboswitch constructs were selected, based on the following criteria of their computationally determined folding paths: (a) the unbound full length transcript disrupts the aptamer and folds into a stable terminator structure; (b) no stable secondary structure is formed downstream of the aptamer/theophylline complex; (c) constructs with a calculated energy difference between bound and unbound state that deviates considerably from the binding energy of the aptamer/theophylline complex (-8.86 kcal/mol) were rejected. The designed constructs were assembled by overlap-extension PCR and cloned into plasmid pBAD2_bgaB between the ara promoter and the reporter gene for β-galactosidase from *B. stearothermophilus* {Klinkert 2012 #137} (Fig. 1B, 2A). Transcription was induced by arabinose in the presence or absence of 2 mM theophylline, and β-galactosidase activity of cell extracts was quantified using ONPG (ortho-nitrophenyl-β-D-galactopyranoside). Three of the constructs turned out to stimulate reporter gene expression by two-fold (RS1 and RS3) or three-fold (RS10) (Fig. 2B). Caffeine, as a control, could not activate the constructs. Fusing riboswitches RS2 and RS4 to the reporter gene resulted in a permanent active state of gene expression, independent of theophylline or caffeine. In contrast, in the RS8 construct, the reporter gene showed a rather low level of expression, resembling an inactive state under all tested conditions. As RS10 had the best theophylline-dependent activation ratio (Fig. 2B), it was chosen for further investigations. To demonstrate the general functionality of such synthetic riboswitches in a different expression system, RS10 was inserted into plasmid

pBSU0. It was positioned downstream of the constitutive promoter P*rrn* and upstream of the green fluorescent protein (GFP) coding sequence {Klinkert 2012 #137}. As in the original construct, reporter gene activity was induced upon the addition of theophylline. The analysis of the resultant relative GFP fluorescence intensity showed a threefold increase in the presence of 1-2 mM theophylline, indicating a similar regulatory profile of RS10 as in the original construct with β-galactosidase expression (Fig. 2C).

To optimize the functional riboswitches, we aimed for a reduction of the background expression level in the absence of theophylline. First, we replaced the tetraloop sequence of the original RS10 construct by GAAA which is frequently found in natural terminators {d'Aubenton Carafa 1990 #139}. Second, as a G-C base pair forms a stable basis for a tetraloop structure, we mutated the terminal A-U base pair in the terminator stem to G-C, leading to construct RS10loop (Fig. 3A). This modification led to an increased efficiency in termination, as the background activity of the riboswitch was reduced from 40 to about 15 Miller units (MU) (Fig. 3B). However, theophylline-induced gene expression activity was also lowered compared to the original construct. Obviously, this optimization of the terminator element stabilized its structure, resulting in an inhibition of a ligand-induced refolding of the upstream aptamer element required for ongoing transcription.

Furthermore, it was described that a stretch of eight U residues closely upstream of the Shine-Dalgarno sequence strongly enhances ribosome attachment and translation, while shorter U stretches have no such effect {Zhang 1992 #134}. As our riboswitches are located immediately upstream of the RBS, it is conceivable that the terminator U-stretch of eight residues (U_8) stimulates translation in a similar way, leading to an unspecific increase in background activity. Indeed, a construct *bgaB*U_8, consisting of the U_8 stretch and the reporter gene only (Fig. 3A), showed a dramatic

and constitutive increase in reporter gene activity up to 100 MU, compared to the original construct bgaB with 31 MU (Fig. 3B). To prove that this effect depends on the close proximity of the U_8 stretch and the ribosomal binding site, a 19 nt region from the original pBAD multiple cloning site was inserted upstream of the reporter gene without a riboswitch, leading to construct *bgaB*shift (Fig. 3A). In contrast to the U_8 insertion, this construct did not show a significant increase in gene activity compared to the original *bgaB* construct (Fig. 3B). Accordingly, we inserted the 19 nt spacer element downstream of RS10, leading to construct RS10shift (Fig. 3A). In the activity test, this riboswitch showed a strongly reduced background signal of approximately 10 MU, resulting in a 6.5-fold activation of gene expression upon theophylline addition (Fig. 3B). To demonstrate the generality of this observation, we inserted the same 19 nt spacer region between RS3 and the ribosomal binding site. Similar to RS10shift, the RS3shift construct also showed an increased ON/OFF rate (5.7-fold), due to a reduced background activity in the absence of the ligand (18 MU compared to 54 MU in RS3; Fig. 3A, B).

To further verify the functionality of the selected riboswitches, individual components were examined in the expression system. Deleting the 5'-part of the aptamer sequence led to isolated terminator structures T1 to T10 that showed efficient repression of the reporter gene activity (Fig. 4A, B). As a positive control, the terminator of the attenuator region of the pyrBI operon from *E. coli* was used {Turnbough 1983 #147}. This natural regulatory element showed a termination efficiency comparable to that of the synthetic terminators, indicating that all generated terminator elements are functional (Fig. 4B).

The observed terminator-dependent reduction of reporter gene activity is a first indication that the synthetic riboswitches indeed control transcription, as terminator function is an essential prerequisite for a riboswitch-mediated transcriptional

regulation. Hence, the integrity of the terminator structure should be essential for regulation, and a deletion of this element should lead to a constitutive active state. To test this, we removed the whole 3'-part of the terminator region of RS10, i.e. the complementary 3'-part of the hairpin and the U-stretch, resulting in RS10ΔT (Fig. 4A). This deletion led to a complete destruction of riboswitch function, and the construct exhibited permanent gene expression, independent of theophylline (Fig. 4B).

Because the essential control elements for translation, in particular the ribosomal binding site, are still present in RS10ΔT as well as the corresponding terminator construct T10, these findings are a clear indication that the RS10 riboswitch regulates at the transcriptional level. To confirm this observation, Northern blot analysis was performed for RS10 and the terminator-alone construct T10. Total RNA was prepared from corresponding *E. coli* TOP10 cells grown in the presence or absence of theophylline. Hybridization was performed using bgaB-specific oligonucleotide probes. To make sure that identical amounts of RNA were analysed, a probe for 23S rRNA was used as an internal control. While the intensity of this control signal was identical in each preparation, a full length mRNA of the RS10-bgaB construct was detected only in the presence of theophylline (Fig. 4C). Quantitative analysis of the transcript of the riboswitch 10 construct revealed a 2,7-fold increase in mRNA levels in the presence of theophylline. This is in good agreement with the levels of active protein measured in the β-galactosidase assay.

The T10 construct, lacking the aptamer domain, however completely inhibited transcription regardless of theophylline. Taken together, this result is in agreement with the corresponding β-galactosidase activity tests and clearly shows that RS10 regulates at the transcriptional level.

DISCUSSION

We have successfully applied a combination of *in silico*, *in vitro* as well as *in vivo* experiments to generate theophylline-dependent riboswitches that activate transcription upon binding of the target molecule. This strategy is a promising alternative to riboswitch selection based on randomized RNA sequence libraries. The design of such libraries is crucial for the success of selection and requires features that differ depending on whether an ON or an OFF switch is to be designed {Muranaka 2009 #148}{Topp 2008 #37}. Furthermore, the screening approach with randomized RNA sequences was applied successfully only to riboswitches acting on translation {Desai 2004 #1}{Lynch 2007 #32}{Topp 2008 #4}{Nomura 2007 #109} {Fowler 2008 #131}. For such elements, the sole requirement is that the alternative structures lead to presentation or masking of the ribosomal binding site, which comprises a short sequence of four to six nucleotides {Steitz 1975 #149}. In transcription-controlling riboswitches, on the other hand, the alternative conformations have to promote or inhibit the formation of a larger and much more complex structure, i.e. an intrinsic terminator consisting of a hairpin element followed by a U-rich single-stranded region. Obviously, the screening of such a high sequence complexity with defined structural requirements is a very difficult task for a selection approach and corresponding efforts for designing transcriptional regulators failed so far {Fowler 2008 #131}. Hence, our rational design seems to be more promising for the design of such synthetic regulators.

The possibilities of comparing the synthetic riboswitches to natural riboswitches are very limited , as almost all of those elements represent OFF-switches based on a very different regulation principle. Here ligand binding induces transcription termination while our synthetic riboswitches cause ligand-induced transcription. Furthermore, regulated changes in expression vary dramatically (between 7 and

1.200-fold) in these riboswitches {Winkler 2005 #160}.The only well characterized transcriptional ON-switch that activates gene expression upon ligand recognition is the adenine-sensing *ydhL* riboswitch of *B. subtilis* {Mandal 2004 #47}. Here, the ligand is bound at a K_d value of 300 nM adenine (K_d of the theophylline aptamer: 320 nM theophylline {Jenison 1994 #46}), and the ON/OFF rate is about 10-fold compared to 6.5-fold as observed for RS10shift. For both natural (*ydhL*) and synthetic (RS10, RS3) riboswitches, similar external concentrations of the ligand molecules (0.5 - 2 mM theophylline versus 0.74 mM adenine) were used. The comparably lower regulation efficiency of the synthetic elements could be explained by the rather inefficient and slow theophylline uptake, so that equilibrium is not reached {KOCH 1956 #161}. Altogether, the regulatory power of the synthetic ON-switches is comparable to that of a natural ON-switch.

In the case of other synthetic RNA-based regulators, theophylline-dependent translational riboswitches and the transcription-controlling elements presented here also show a similar efficiency in regulation (8-fold versus 6.5-fold) {Suess 2004 #3} {Desai 2004 #1}. However, these constructs are transcribed from different promoters (*B. subtilis* XylA, *E. coli* IS10) that can also influence termination efficiency {Goliger 1989 #150}. Furthermore, it was shown that sequences downstream of the U-stretch can have an impact on termination efficiency {d'Aubenton Carafa 1990 #139}{Lesnik 2001 #51}, while additional U-rich elements enhance translation {Zhang 1992 #151}. Sequences immediately downstream of the start codon, described as downstream boxes, can also act as translational enhancers, as it was demonstrated for several *E. coli* as well as bacteriophage genes {Sprengart 1996 #152}. Besides these and other sequence-specific features, there are further properties that greatly influence the expression of a given gene. Secondary structure elements and position effects of the above mentioned regulatory regions within a translational unit play a crucial role for

productivity of protein synthesis {Kozak 2005 #154}{Makrides 1996 #153}. As these factors have a great impact on the efficiency of riboswitches in general, it is virtually impossible to compare the regulatory power of the elements. On the other hand, this plethora of features influencing bacterial gene expression allows a further optimization of the individual constructs, eventually leading to higher ON/OFF rates.

Although these data indicate that the nucleic acid sequence context can have a great impact on the efficiency of such synthetic riboswitches, RS10 shows a similar theophylline-dependent regulation in a different genetic background consisting of a constitutive promoter and GFP as a reporter (Fig. 2C). A likely reason for the different ON/OFF rates of RS10 observed for the pBAD/bgaB system (6.5-fold) and the pBSU/GFP system (3-fold) are the specific properties of GFP as a reporter protein. After translation, GFP undergoes a series of maturation steps, leading to a considerable delay in fluorescence. Furthermore, in rapidly dividing cells, GFP will be diluted rather fast, leading to a reduced fluorescence signal in the growing culture {Leveau 2001 #157}. Hence, the absolute ON/OFF rates of the two reporter systems cannot be directly compared. Nevertheless, RS10 is active in both systems, indicating a functional generality of the design approach used in this study. Furthermore, the P*rrn* promoter based transcription in the pBSU/GFP system indicates that arabinose used as an inducer in pBAD/bgaB has no influence on the conformational switch of the regulatory constructs but only activates transcription. This is further supported by the fact that in all individual assays based on the pBAD/bgaB system, arabinose was added for transcription induction, leading to a general normalization for the presence of arabinose in these experiments. Hence, any effect of arabinose on the individual riboswitch constructs can be ruled out.

For transcription-controlling riboswitches, a kinetic control is discussed, where the

riboswitch does not reach a thermodynamic equilibrium with its ligand, but is driven by the kinetics of RNA folding, ligand interaction and RNA synthesis {Wickiser 2005 #9}{Wickiser 2005 #8}{Rieder 2007 #18}. Such a competition between ligand binding and speed of transcription seems to be a common regulatory feature. Several transcriptional riboswitches force the polymerase to pause downstream of the terminator element, giving the sensor sufficient time for ligand binding and, consequently, deciding whether to proceed or terminate transcription {Haller 2011 #25}. In our riboswitches, sophisticated structural features such as transcriptional pause sites {Wickiser 2005 #8} or the energetic contribution of ligand binding to the stability of the riboswitch conformation {Gouda 2003 #132} were not yet included in the design principles. The K_d value used to estimate the binding energy between target molecule and aptamer depends on experimental conditions, e.g. salt concentration and temperature, which are uncertain for the theophylline-aptamer used here. Hence, the resulting ΔG value represents only a rough estimate and cannot be directly integrated into the pipeline at the moment. The current design strategy is based on minimum free energy evaluations. This works as long as the unbound aptamer structure folds into the MFE conformation as it is the case for the used theophylline-aptamer. It needs to be replaced, however, if alternative structures are adapted preferentially. The usage of advanced co-transcriptional folding simulations and explicit predictions of alternative conformations will become necessary at that point.

As transcriptional riboswitches require an overlapping region between sensor (theophylline aptamer) and actuator (terminator), it is not possible to simply fuse a naturally occurring terminator to the aptamer. Natural intrinsic terminators consist of a GC-rich hairpin followed by a U-rich region that readily dissociates from the DNA-template {d'Aubenton Carafa 1990 #139}{Martin 1980 #155}. As the hairpin element

does not show a high conservation at the sequence level, this part of the terminator can be easily manipulated without destroying its regulatory function {Wilson 1995 #156}. As a consequence, it is possible to design artificial terminators simply by adjusting the 5'-half of the hairpin element to the sequence required for proper base pairing with the 3'-part of the aptamer. This is demonstrated by the fact that all designed terminator structures significantly shut down gene expression and represent therefore functional elements, comparable to natural terminators. Obviously, the simple descriptors used for our *in silico* construction are sufficient to predict efficient termination of the reporter system.

As shown in figure 2, not every generated terminator led to a functional riboswitch. Nevertheless, the calculated free energy values of the individual terminators are in good agreement with functionality of the constructs (Fig. 5). The terminator structure of RS8 showed the highest stability (-29.00 kcal/mol), suggesting that it is formed even in the presence of theophylline, and the bound aptamer structure cannot compete with the more stable terminator hairpin. Correspondingly, RS8 exhibits a constitutive low gene expression level. In contrast, terminators of RS2 and RS4, with constitutive gene expression, are the least stable structures with -19.70 kcal/mol and -20.60 kcal/mol, respectively. Here, the binding-competent structure of the aptamer is probably more stable and remains intact even in the absence of theophylline, interfering with terminator formation. The stabilities of terminators in functional riboswitches RS1, RS3 and RS10 were between -21.00 and -25.80 kcal/mol, allowing a stable formation of the ligand-bound aptamer structure in the presence of theophylline. In agreement with this correlation are the regulatory features observed for RS10loop, where the optimization of the terminator hairpin leads to an increased stability. Due to this more robust terminator structure (ΔG = -24.8 kcal/mol), RS10loop has a reduced background activity compared to RS10. This is correlated

with a lowered ON rate, as the ligand-bound aptamer structure now has to compete with a more stable terminator element. Yet, this interpretation is rather speculative, as the stability of a terminator structure is not necessarily correlated with termination efficiency.

Although the adjusted loop sequences are found in most of the naturally occurring intrinsic terminators {d'Aubenton Carafa 1990 #139}, one cannot transfer common features of stand-alone terminators into transcriptional riboswitches. Obviously, the requirements for structural rearrangements between ON and OFF states do not allow optimal terminator structures but favor a compromise between efficient termination and the possibility of riboswitch refolding upon ligand binding. As energy values for most of the available aptamers are not known, it is not possible to predict a priori which of the designed terminators are suited for a certain aptamer domain to form a functional riboswitch. Here, biochemical analyses are required to verify a ligand-dependent regulation of transcription.

A great advantage compared to translational riboswitches is the fact that transcriptional riboswitches can in principle be combined as tandem copies in a single transcriptional unit, potentially leading to a rather tight regulation of gene expression or Boolean NOR gene control logic {Breaker 2012 #6}{Fowler 2008 #131}. Similar tandem arrangements are also described for several natural riboswitches, leading among other effects to an enhanced digital character of regulation {Welz 2007 #158}. Furthermore, transcriptional riboswitches could be used as control elements for the release of non-coding regulatory RNAs that influence expression of certain target genes. However, in order to produce a functional regulatory RNA free of potentially interfering flanking sequences, additional cleavage sites in the primary transcript might be required.

Taken together, our *in silico* design to construct functional synthetic riboswitches

acting at the level of transcription is a first promising approach to generate and optimize sophisticated regulatory RNA control elements. The implication of further parameters like transcriptional pause sites and ligand binding behavior will lead to a more efficient regulatory power, comparable to that of the natural counterparts. Hence, fully understood synthetic elements have the potential to develop into useful and versatile building blocks in synthetic biology.

ACKNOWLEDGEMENTS

The authors thank H. Betat, B. Klinkert and P. Kerpedjiev for valuable discussions and G. Domin for cloning RS8. Thanks to M. Mießler and S. Bonin for expert technical support. This work was supported by the Deutsche Forschungsgemeinschaft and the Christiane-Nüsslein-Vollhard-Stiftung.

REREFERNCES

FIGURE LEGENDS

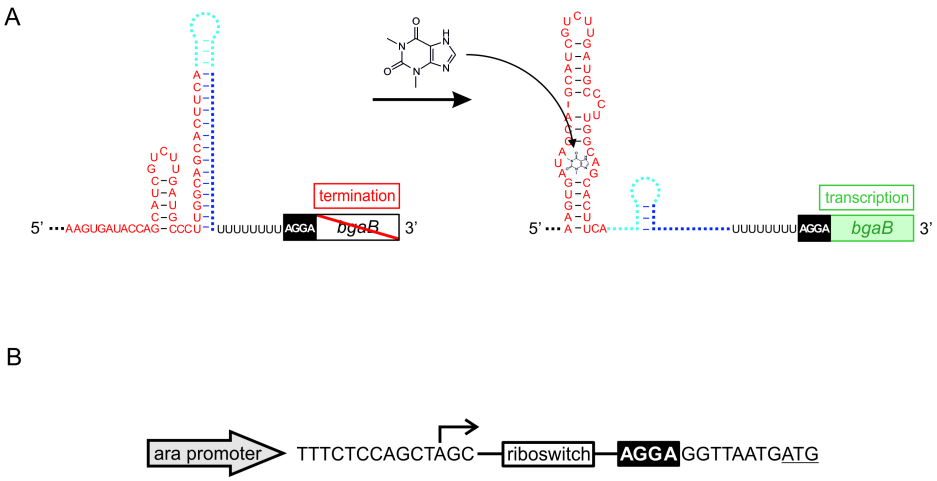


Fig.1

Fig. 1. Riboswitch constructs. (A) Design strategy for theophylline-dependent riboswitches controlling transcription. The sequence of the TCT8-4 theophylline aptamer (red) was fused to a short spacer region (cyan) followed by a sequence complementary to the 3'-part of the aptamer (blue) and a U-stretch (black). The Shine-Dalgarno-Sequence (black box) and the open reading frame of the reporter gene bgaB are located immediately downstream of this construct. Upon theophylline binding, terminator structure formation should be inhibited and transcription should proceed, resulting in expression of the reporter gene. **(B)** The final riboswitch constructs are transcribed from the arabinose promoter of plasmid pBAD. The transcription start point is indicated by the angled arrow. The Shine-Dalgarno-Sequence (AGGA) of the reporter gene for β-galactosidase is located immediately downstream of the riboswitch insert. The start codon of the reporter gene is underlined.

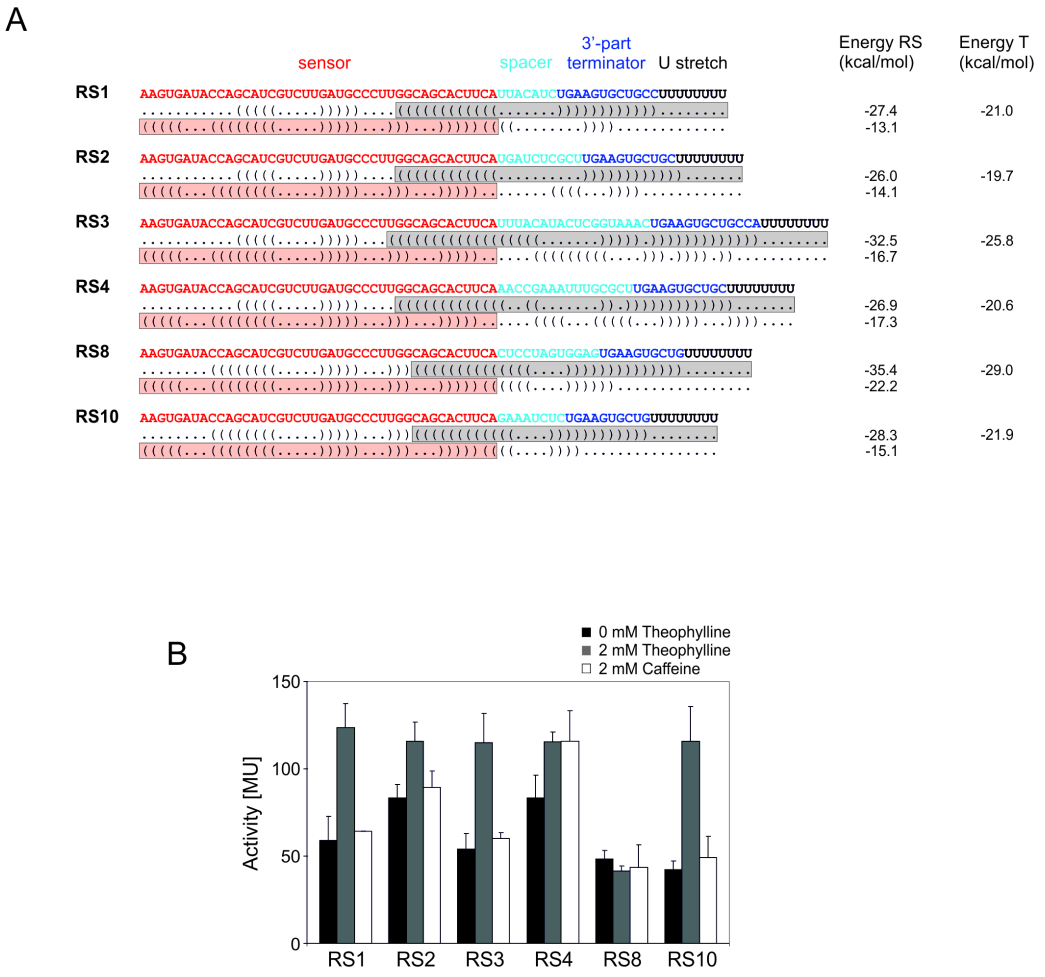
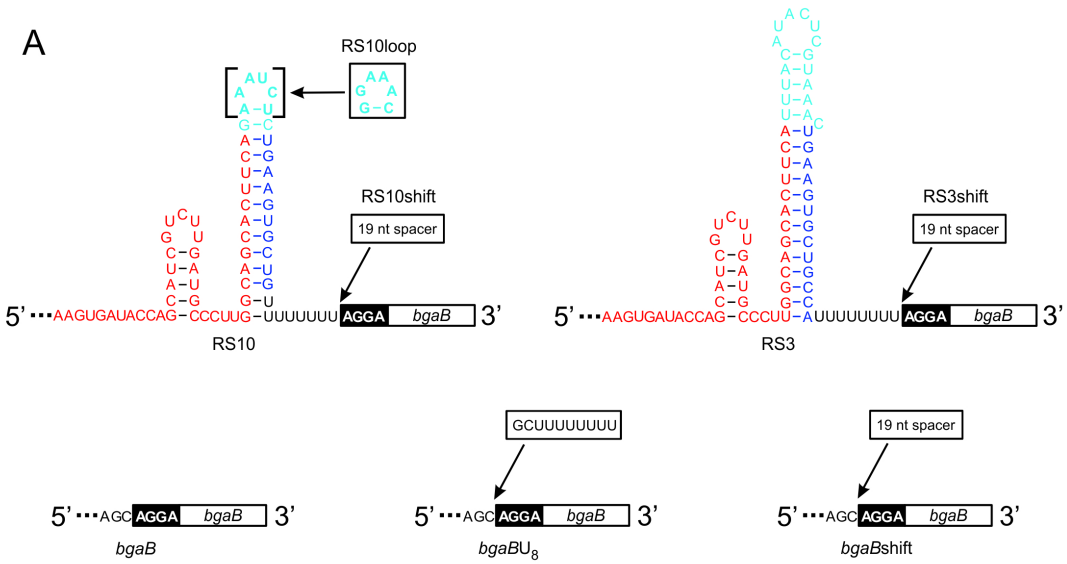


Fig.2

Fig. 2. Riboswitch constructs and activity tests. (A) The theophylline aptamer represents the sensor (red), followed by a variable spacer sequence (cyan) and the

3’-part of the terminator (blue), forming a hairpin structure with a subsequence of the sensor domain. The terminator element is completed by a stretch of eight U residues (black). Below each construct, dot bracket notations of the secondary structures with terminator or antiterminator formation are indicated. The overlapping situation of sensor (theophylline aptamer, red box) and actuator element (terminator element, grey box) is required for a mutual exclusion of the two alternative structures. For each construct, calculated energy values of complete riboswitch elements (RS) as well as isolated terminators (T) are indicated. **(B)** Activity tests of the reporter enzyme β-galactosidase. Activities are indicated in Miller units (MU). All synthetic riboswitches were tested in the absence (black bars) and presence (grey bars) of 2mM theophylline. As a control, the structurally related caffeine was offered (white bars). **(C)** Activity test of RS10 under the control of the P*rrn* promoter and GFP as reporter gene. Relative fluorescence signals are indicated as a function of theophylline concentration. The riboswitch shows a similar regulation profile as in **(B)**, indicating a general functionality of the construct.


A
RS10loop
RS10shift
19 nt spacer
5'···AAGUGAUACCAG–CCCUUG–UUUUUUUU AGGA bgaB 3'
RS10
RS3shift
19 nt spacer
5'···AAGUGAUACCAG–CCCUU–AUUUUUUUU AGGA bgaB 3'
RS3
5'···AGC AGGA bgaB 3'
bgaB
GCUUUUUUUU
5'···AGC AGGA bgaB 3'
bgaBU8
19 nt spacer
5'···AGC AGGA bgaB 3'
bgaBshift

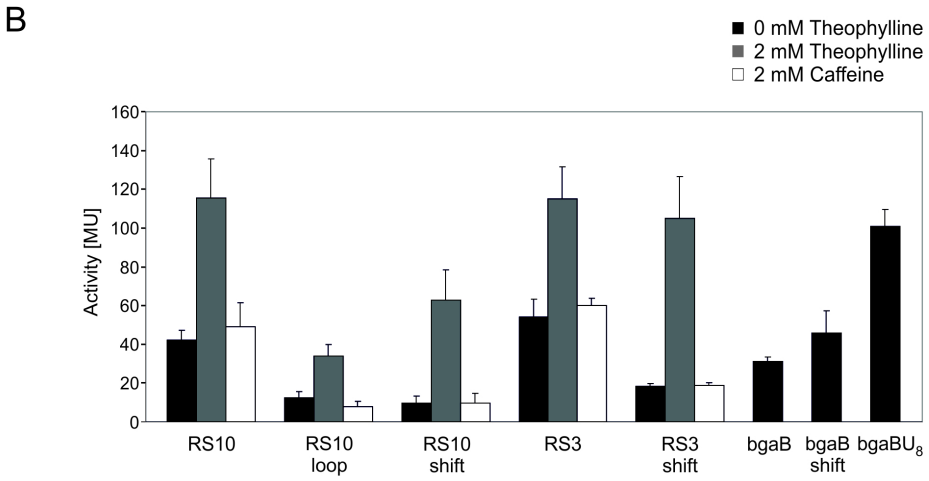
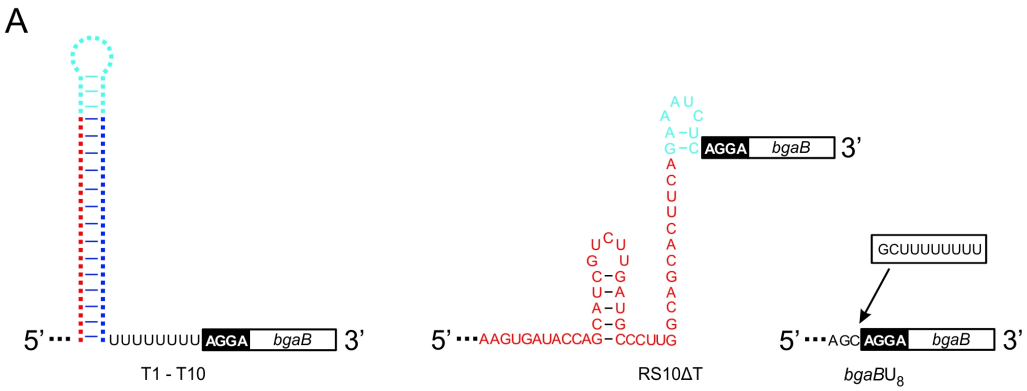
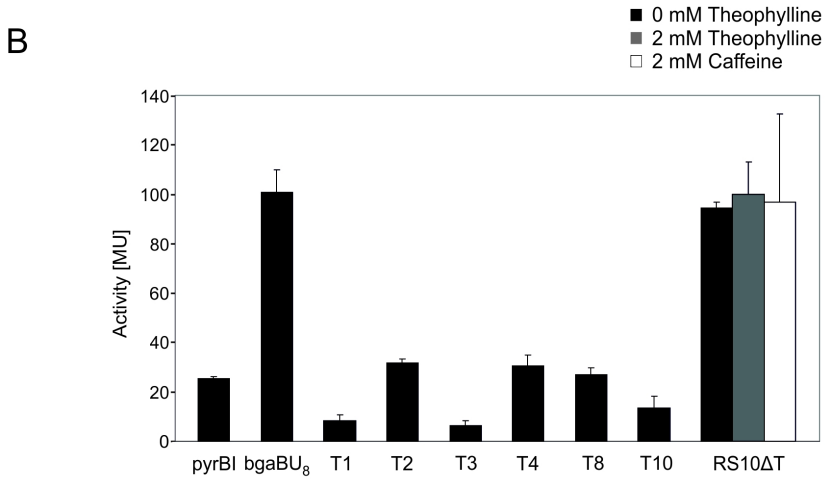

B
0 mM Theophylline
2 mM Theophylline
2 mM Caffeine
Activity [MU]
0
20
40
60
80
100
120
140
160
RS10
RS10 loop
RS10 shift
RS3
RS3 shift
bgaB
bgaB shift
bgaBU8

Fig.3

Fig. 3. Optimized RS10 and RS3 variants and control constructs. (A) In RS10loop, the tetraloop sequence of the terminator hairpin was adjusted to the frequently found sequence GAAA, including a closing G-C base pair. In RS10shift, a spacer sequence of 19 nt was inserted between the U_8 stretch of the terminator and the Shine-Dalgarno-Sequence (black box). The same 19 nt spacer was inserted in RS3shift at the identical position. In control bgaB, the upstream untranslated region of the reporter gene did not carry any riboswitch or terminator inserts. Furthermore, the U_8 stretch alone was placed immediately upstream of the ribosomal binding site and the 19 nt spacer sequence region was inserted into the original bgaB control construct, leading to bgaBU8 and bgaBshift, respectively. **(B)** β-galactosidase activity test of the constructs described in (A). Activities are indicated in Miller units (MU). The tetraloop adaptation (RS10loop) in the terminator element led to a reduction of the background activity, but lowered also the expression of the reporter gene. Increasing the distance between the riboswitch U_8 stretch and the ribosomal binding site of the reporter gene in RS10shift and RS3shift increased the ON/OFF rate substantially, compared to the original constructs (for comparison, the regulation profiles of RS3 and RS10 shown in Fig. 2B are again indicated in this panel). In the controls, the U_8 stretch immediately upstream of the ribosomal binding site led to a rather high gene expression (bgaBU_8), while increasing the distance between these sequences reduced the β-galactosidase activity (bgaBshift).


A
5'
UUUUUUUU
AGGA
bgaB
3'
T1 - T10
AAGUGAUACCAG
CCCUUG
RS10ΔT
GCUUUUUUUU
AGC
bgaBU8


B
0 mM Theophylline
2 mM Theophylline
2 mM Caffeine
Activity [MU]
140
120
100
80
60
40
20
0
pyrBI
bgaBU8
T1
T2
T3
T4
T8
T10
RS10ΔT

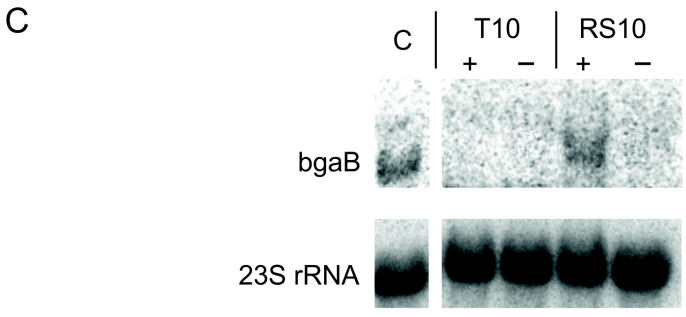

C
C
T10
RS10
+
–
+
–
bgaB
23S rRNA

Fig.4

Fig. 4. Functional analysis of synthetic terminator elements. **(A)** The isolated terminators of the tested riboswitch constructs were cloned upstream of the reporter gene. In a control construct, the 3'-complementary sequence of terminator 10 and the U_8 stretch was deleted, leading to RS10ΔT. As shown in Fig. 3, construct bgaBU_8 represents the positive control. **(B)** β-galactosidase activity of terminator constructs. While the positive control bgaBU_8 (activation data from Fig. 3B are presented) as well as the terminator deletion RS10ΔT show a strong enzyme activity, all terminators reduce the β-galactosidase activity at an efficiency comparable to that of the natural terminator from the pyrBI operon in *E. coli* (pyrBI). The terminator elements of the functional riboswitches 1, 3 and 10 show the strongest activity reduction. **(C)** Northern blot analysis of RS10. Left: Upper panel: While the terminator-alone construct T10 showed no transcription of the reporter gene, RS10 allowed bgaB gene transcription exclusively in the presence (+) of theophylline. C, control expression of bgaB using pBAD2-bgaB without riboswitch inserts. Lower panel: 23S rRNA was used as an internal standard for normalizing the presented lanes. Right: Quantification of bgaB-transcript in RS10 construct with and without theophylline.

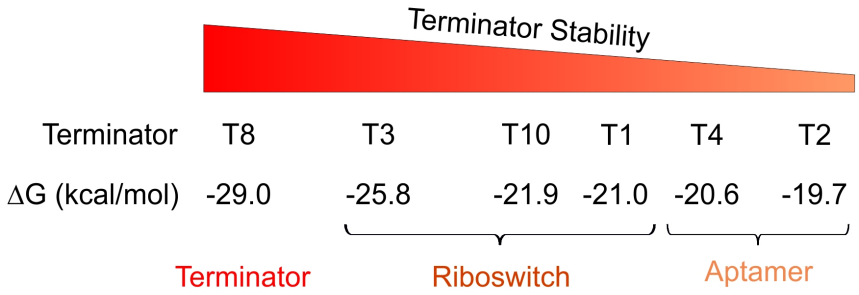

Terminator Stability
Terminator
T8
T3
T10
T1
T4
T2
ΔG (kcal/mol)
-29.0
-25.8
-21.9
-21.0
-20.6
-19.7
Terminator
Riboswitch
Aptamer

Fig.5

Fig. 5. A possible correlation between terminator stability and riboswitch function. Among all tested terminators, T8 has the lowest free energy value and forms a stable structure that obviously cannot be disrupted upon ligand binding. As a consequence, RS8 is locked in an OFF state. In contrast, T2 and T4 form a much weaker hairpin, shifting the equilibrium towards the aptamer structure that interferes with terminator formation. The functional riboswitches RS1, RS3 and RS10 show intermediate stabilities of their terminator elements that allow the desired ligand-dependent rearrangement of the constructs, switching between ON and OFF states. The stabilized terminator hairpin of RS10loop is located between the terminators of RS8 and the functional RS3. Although this stabilized hairpin reduces the background transcription rate considerably, it obviously impedes the structural rearrangement induced by ligand binding and allows only a moderate level of gene expression.